



دانشگاه علوم کشاورزی و منابع طبیعی گیلان

نشریه پژوهش در نشخوارکنندگان

جلد هشتم، شماره دوم، ۱۳۹۹

<http://ejrr.gau.ac.ir>

۳۹-۵۶

DOI: 10.22069/ejrr.2020.17367.1719

## روش‌های کارآمد برآورد ارزش اصلاحی ژنومی و مکان‌یابی QTN‌ها در راهبردهای به‌نژادی گاو شیری

حسین عبداللهی<sup>۱</sup>، \* غلامرضا داشاب<sup>۲</sup>، محمد رکوعی<sup>۲</sup> و مهدی سرگلزائی<sup>۳</sup>

<sup>۱</sup>دانشجوی دکتری و <sup>۲</sup>دانشیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، ایران

<sup>۳</sup>دانشیار گروه پاتولوژی، دانشگاه گوتلف، انتاریو، کانادا

تاریخ دریافت: ۹۸/۹/۴؛ تاریخ پذیرش: ۹۹/۴/۱۸

### چکیده

**سابقه و هدف:** در دسترس بودن هزاران تک‌نکلوتید چندشکل که در سراسر ژنوم گسترش یافته، امکان استفاده از اطلاعات نشانگرهای ژنوم گسترده در انتخاب ژنومی برای پیش‌بینی ارزش اصلاحی کل را فراهم نمودند. اثر متقابل بین چند جایگاه ژن، به تغییرات فنوتیپی مرتبط با بیان صفات پیچیده چند ژنی کمک می‌کند. مطالعه حاضر به منظور مقایسه عملکرد مدل‌های مختلف با اثر ژنتیکی افزایشی و غیرافزایشی به ویژه اثر اپیستاتیک در پیش‌بینی اجزای واریانس و ارزش‌های اصلاحی ژنومی و همچنین مکان‌یابی نوکلئوتیدهای کنترل‌کننده صفت کمی در یک جمعیت گاو شیری بر اساس مدل بیز لاسو انجام شد.

**مواد و روش‌ها:** یک جمعیت گاو شیری بر اساس انتخاب چهار مسیر به منظور پیشینه نمودن سرعت پیشرفت ژنتیکی در طی ۱۰ نسل با اندازه مؤثر ۱۰۰ فرد در جمعیت تاریخی شبیه‌سازی شدند. ساختار ژنومی شامل ۳ کروموزوم با طول ۱۰۰ سانتی مورگان فرض شد که بر روی هر کدام ۱۰۰۰ نشانگر ۲ آللی با فراوانی ۰/۵ جانمایی شدند. ۵۰ جایگاه کنترل‌کننده صفت کمی دو آللی با فراوانی برابر که به صورت تصادفی در روی هر کروموزوم قرار داشتند، شبیه‌سازی شدند. برای اثرات آللی QTL از یک توزیع آماری گاما با شیب پارامتر ۰/۴ در نرم‌افزار QMSim استفاده شد و نمونه‌برداری انجام گرفت. اطلاعات فنوتیپی و ژنوتیپی ۱۰ نسل آخر برای تجزیه و تحلیل‌ها استفاده شدند. اثرات نشانگری، مؤلفه‌های واریانس و پارامترهای ژنتیکی با روش بیز لاسو در قالب ۶ مدل آماری برآورد شدند که مدل ۱ شامل تنها آثار افزایشی نشانگرها، مدل ۲ شامل اثرات چندژنی حیوان، مدل ۳ شامل اثرات توأم افزایشی نشانگرها و چندژنی حیوان، مدل ۴ شامل اثرات غالبیت نشانگرها و آثار چندژنی حیوان، مدل ۵ شامل اثرات افزایشی و غالبیت نشانگرها بعلاوه آثار چندژنی حیوان و مدل ۶ شامل اثرات افزایشی، غالبیت و اپیستاتیک نشانگرها بعلاوه آثار چندژنی حیوان می‌باشند. به منظور کنترل خطای اول در برآورد آثار نشانگری از آزمون بنفرونی در سطح احتمال یک درصد استفاده شد.

**یافته‌ها:** نتایج این مطالعه نشان داد که مدل افزایشی (یا مدل چند ژن) به تنهایی نمی‌تواند مقادیر واریانس‌های گمشده یا مخفی را نمایان سازد، به نحوی که افزودن آثار ژنتیکی غیرافزایشی شامل غالبیت و اپیستاتیک نشانگرها منجر به کاهش واریانس باقیمانده شد. همچنین در مدل کامل شامل آثار افزایشی و غیرافزایشی، واریانس ژنتیکی کل افزایش یافت. با اضافه شدن آثار غیرافزایشی در مدل آماری میزان وراثت‌پذیری عام افزایش و وراثت‌پذیری خاص صفت کاهش یافت. صحت برآورد ارزش‌های اصلاحی در

\*نویسنده مسئول: [dashab@uoz.ac.ir](mailto:dashab@uoz.ac.ir)

مدل یک کمترین و در مدل‌های ۳، ۴، ۵ و ۶ مشابه هم بودند. مدل ۵ با کمترین QTN مثبت کاذب و به نسبت بیشترین مثبت واقعی مدل مناسبی در ارزیابی ژنتیکی صفت مورد مطالعه بود. با قرار دادن آثار چندژنی و آثار غیرافزایشی در مدل از تعداد خطاهای مثبت کاذب کاسته شد.

**نتیجه‌گیری:** نتایج این تحقیق نشان داد اجزای واریانس افزایشی و غالبیت نشانگرها بر میزان تنوع ژنتیکی صفت مورد مطالعه سهم کم داشت، اما اثرات چندژنی حیوان و اپیستاتیک بیشترین سهم را در بروز تنوع صفت داشتند. همچنین نتایج نشان داد که سهم عمده جایگاه های ژنی در بروز صفات مربوط به ژن های افزایشی با اثر افزایشی هستند که توزیع نرمال دارند. نادیده گرفتن آثار غیرافزایشی موجب التهاب در واریانس افزایشی صفت می‌شود.

**واژه‌های کلیدی:** اثرات غیرافزایشی، اجزای واریانس، انتخاب ژنومیک، بیز لاسو، گاو شیری،

### مقدمه

سپس بر اساس آثار نشانگری برآورد شده از اطلاعات جمعیت مرجع، ارزش‌های اصلاحی ژنومی جمعیت تأیید پیش‌بینی می‌شود. اهمیت نسبی هر جزء واریانس معمولاً در صفات مختلف متفاوت است. اثر متقابل بین چند جایگاه ژنی اثر اپیستاتیک نامیده می‌شود. آثار اپیستاتیک توجیهی برای وراثت‌پذیری گم شده هستند که معمولاً در مطالعات پیوستگی گسترده ژنوم<sup>۵</sup> در نظر گرفته نمی‌شوند. بررسی آثار اپیستاتیک می‌تواند در درک معماری مراتبی از تعاملات ژنی و مسیرهای ژنتیکی مؤثر باشند (۲۰). واریانس آثار اپیستاتیک در صفات کمی شامل اجزای مختلف از جمله افزایشی در افزایشی، افزایشی در غالبیت، غالبیت در افزایشی و غالبیت در غالبیت هستند (۵). تقسیم‌بندی دقیق این اجزای واریانس اپیستاتیک می‌تواند به شناخت مکانیزم‌های ژنتیکی صفات پیچیده و توسعه برنامه‌های اصلاحی مؤثرتر کمک کند. یکی از مؤلفه‌های اصلی اثر انتخاب ژنومیک، صحت پیش‌بینی است. صحت پیش‌بینی به عنوان همبستگی بین ارزش های اصلاحی ژنومی برآورد شده<sup>۶</sup> و ارزش اصلاحی واقعی تعریف می‌شود که با اندازه‌گیری عملکرد

انتخاب ژنومی، انتخاب به کمک نشانگرها<sup>۱</sup> است که در آن از اطلاعات نشانگرهای ژنتیکی که کل ژنوم را پوشش می‌دهند استفاده می‌شود و جایگاه‌های صفات کمی<sup>۲</sup> در عدم تعادل پیوستگی<sup>۳</sup> با حداقل یک نشانگر شناسایی می‌گردند. این رویکرد به کمک کشف تعداد زیادی تک‌نکلئوتید چندشکلی<sup>۴</sup> با توالی‌یابی ژنومی و روش‌های جدید تعیین ژنوتیپ کارآمد تعداد بسیار زیاد SNP تسهیل شد (۱۲). در مقایسه با شیوه‌های معمول اصلاح نژادی، وارد کردن اطلاعات نشانگرهای ژنتیکی، می‌تواند به میزان قابل توجهی پاسخ به انتخاب را در حیوانات جوان‌تر که سوابق فنوتیپی ندارند، افزایش دهد (۳۱). همچنین بطور بالقوه می‌تواند هزینه‌های برنامه اصلاح‌نژادی را تا ۹۰ درصد کاهش دهد (۴۰). فرآیند کلی انتخاب ژنومی ابتدا ایجاد یک جمعیت مرجع با استفاده از حیوانات با فنوتیپ‌های تک‌رکوردی یا چند رکوردی و تراشه‌های SNP ژنوتیپ شده می‌باشد و تخمین آثار نشانگرهای هر SNP بر پایه اطلاعات جمعیت مرجع است (۴۱).

1. Marker assisted selection (MAS)
2. Quantitative trait loci (QTL)
3. Linkage disequilibrium (LD)
4. Single nucleotide polymorphism (SNP)

5. Genome-wide association (GWAS)
6. Genomic estimation breeding value (GEBV)

رگرسیون حاشیه<sup>۱</sup> بهترین پیش‌بینی نااریب خطی<sup>۲</sup> فرض می‌گردد همه آثار نشانگرها دارای توزیع نرمال و یکنواختی در سراسر ژنوم هستند (۳۱). رگرسیون حاشیه‌ای یک فرآیند مستمری دارد که موجب کاهش انقباض لامبدا می‌گردد و پایدارتر است. با این حال، هیچ ضربی برای صفر تنظیم نمی‌کند، لذا یک مدلی که به راحتی قابل تفسیر باشد را ارائه نمی‌دهد (۴۴). در روش بیس A، فرض می‌گردد که نشانگرها واریانس‌های مختلف دارند و بر اساس توزیع معکوس<sup>۲</sup> مدل‌سازی می‌شوند (۳۱). توزیع پیشین در روش بیس B (۳۱) فرض می‌کند که واریانس نشانگرها برابر با صفر با احتمال P است و مکمل با احتمال (1-P) از یک توزیع معکوس<sup>۲</sup>، با درجه آزادی V و مقیاس پارامتر S پیروی می‌کند. تعریف احتمال P به ساختار ژنتیکی صفات وابسته است، که عامل پیشرفت مدل بیس B است که به‌عنوان مدل بیس Cp شناخته شده است. در بیس Cp ترکیب احتمالی P توزیع پیشین یکنواخت دارد (۱۴). مدل‌های منظم مخلوط (که به عنوان خط الرأس رگرسیون<sup>۳</sup> و یا بهترین پیش‌بینی نااریب خطی ژنومی<sup>۴</sup> شناخته شده است) برای فهم، کاربرد و بسط دادن ساده هستند (۱) و (۴۵)، ولی حداقل برای برخی از صفات، خروج آثار SNP از حالت نرمال وجود دارد که در نتیجه دقت بالاتر روش‌ها با پیچیدگی زیاد توزیع پیشین آثار نشانگرها، مانند رگرسیون غیرخطی یا بیس A است (۱۶) و (۴۶). هیچ نشانه روشن، بر اساس دانش بیولوژیکی، درباره این توزیع پیشین وجود ندارد. بنابراین یک جایگزین محبوب بیس لاسو<sup>۵</sup> است (۸ و ۳۴) که اجازه

فنونیتی تخمین زده می‌شود (۱۲ و ۵۱). سه روش کلی برای ارزیابی صحت پیش‌بینی با استفاده از داده‌های واقعی وجود دارد: (یک) اعتبار زیرمجموعه، (دوم) اعتبار بینابینی و (سوم) اعتبار فرزندان. اعتبار زیرمجموعه از تقسیم تصادفی یک جمعیت واحد از افراد به زیرمجموعه‌های برابر انجام می‌شود. یک نمونه به عنوان مجموعه‌ای از اعتبارسنجی استفاده می‌شود که با استفاده از نمونه‌های باقی‌مانده به‌عنوان مجموعه آموزش پیش‌بینی می‌شود. اعتبارسنجی زیرمجموعه برای ارزیابی دقت پیش‌بینی در گاو، گندم، جو و در میان بسیاری دیگر از گونه‌های دامی و زراعی استفاده شده است (۱۷، ۲۵، ۲۶ و ۳۵). در اعتبارسنجی بین‌گروهی، مجموعه‌های از پیش‌تعریف شده ژنوتیپ‌ها به عنوان جمعیت پرورشی و تأیید تعیین می‌شوند. این مجموعه‌ها می‌توانند همان ژنوتیپ‌های محیط‌های مستقل به عنوان مجموعه داده‌های پرورشی و تأیید یا مجموعه‌ای از خطوط تولید به صورت زمانی تعریف شده باشند که در آن خطوط قدیمی‌تر برای پیش‌بینی خطوط جدیدتر از محیط‌های یکسان یا مستقل استفاده می‌شوند (۲) و (۲۵). اعتبار فرزندان حاکی از آن است که جمعیت پرورشی شامل والدین (یا پدربزرگ و مادربزرگ و غیره) از لاین فرزندان هستند که جمعیت تأیید را تشکیل می‌دهند.

در یک مطالعه شبیه‌سازی در حیوانات گزارش شده است که کاهش در صحت پیش‌بینی با تنزل LD بین نشانگرها و مکان‌های کمی صفت ناشی از نوترکیبی در نسل‌های فرزندان در ارتباط است (۱۳).

روش‌های مختلفی برای برآورد آثار نشانگری و

پیش‌بینی ارزش‌های اصلاحی ژنومی با فرضیات متعدد برای آثار و توزیع نشانگرها توسط محققین مختلف پیشنهاد شده است (۷، ۱۴ و ۳۱). در روش

1. Ridge Regression
2. Ridge Regression- best linear unbiased prediction (RR-BLUP)
3. Bayesian ridge regression (BLR)
4. Genomic BLUP (GBLUP)
5. Bayesian least absolute shrinkage and selection operator (BL)

می‌دهد برای خروج از نرمال (SNPهای با اثر بزرگ) در حالی که هنوز هم انقباض انجام می‌شود، به توانایی پیش‌بینی کمک کند. بیز لاسو براحتی در نمونه برداری گیبس به کار گرفته می‌شود و متغیرهای اضافی  $t_i$  که واریانس  $i$  امین اثر SNP را توصیف می‌کند، به مدل سلسله مراتبی اضافه می‌نماید (۲۴). همچنین، یک اشکال روش‌های بیزی نیاز به تعریف توزیع پیشین است که با روش بیز لاسو مرتفع می‌گردد (۸، ۲۳ و ۳۷). برآورد لاسو برای پارامترهای رگرسیون خطی می‌تواند به عنوان یک برآورد حالت بیزین پسین هنگامی که پارامترهای رگرسیون توزیع لاپلاس اولیه (یعنی نمایی درجه دو) دارند به اجرا در آید. بنابراین، بیز لاسو برآورد فاصله‌ای قابل اعتمادتری (فواصل قابل اعتماد بیزی) را برای هدایت به انتخاب متغیرها فراهم می‌کند (۳۴).

در مطالعه حاضر، در قالب مدل‌های مختلف با آثار ژنتیکی افزایشی و غیرافزایشی و اپیستاتیک با روش بیز لاسو به تجزیه و تحلیل اجزای واریانس ژنتیکی و پیش‌بینی ارزش‌های اصلاحی پرداخت شد. در مرحله بعد به توانایی هر کدام از مدل‌ها در شناسایی نوکلئوتیدهای کنترل‌کننده صفت کمی<sup>۱</sup> مورد بررسی قرار گرفت. همچنین، در این تحقیق از ترکیب مدل‌های با آثار ژنتیکی غیرافزایشی شامل غالبیت و اپیستاتیک و روش آماری بیز لاسو جهت تعیین QTNهای مثبت و منفی واقعی استفاده شد که تاکنون مطالعه‌ای در این زمینه انجام نگرفته است.

### مواد و روش‌ها

یک جمعیت گاو شیری مطابق با الگوی شکل ۱ با استفاده از نرم‌افزار QMSim (۳۹) ایجاد شد. به منظور ایجاد عدم تعادل پیوستگی بین نشانگرها و QTLها و همچنین تعادل رانش - جهش بر مبنای روش

ویلوومن و همکاران (۴۷) اندازه مؤثر جمعیت ۱۰۰ در نظر گرفته شد که شامل ۵۰ رأس نر و ۵۰ رأس ماده بودند. این ساختار جمعیتی در جمعیت تاریخی<sup>۲</sup> برای ۴۰۰۰ نسل با آمیزش تصادفی ادامه یافت (۹) تا عدم تعادل پیوستگی لازم بین نشانگرها و QTLها بوجود آید. در ادامه اندازه جمعیت به ۶۰۰۰ رأس افزایش یافت و برای مطالعات ژنومی استفاده شدند. اندازه جمعیت مذکور برای ۶ نسل بعدی (تا نسل ۴۰۰۶) ثابت در نظر گرفته شد. به منظور حصول به حداکثر سرعت پیشرفت ژنتیکی از این گله تاریخی مطابق شکل ۱ چهار لاین مشتق و انتخاب شدند که به ترتیب گروه ۱: پدر بزرگ مادری، گروه ۲: مادر بزرگ مادری، گروه ۳: پدر بزرگ پدری و گروه ۴: مادر بزرگ پدری را تشکیل دادند. در هر گله گاوها براساس بیشترین مقدار ارزش اصلاحی برآورد شده انتخاب شدند و حذف آنها از گله براساس سن انجام شد. روش تخمین ارزش اصلاحی با استفاده از روش بهترین پیش‌بینی ناریب خطی بود. چهار گروه والدین مجزا که از نسل آخر جمعیت تاریخی (در خط پدری ۱۰۰ نر و ۴۰۰ ماده و در خط مادری ۲۰ نر و ۴۰۰ ماده) برحسب ارزش ارثی واقعی بالا مجزا شدند و برای ۴۰ نسل دیگر باهم آمیزش تصادفی کردند. سپس جمعیت پدری از نسل ۴۰ (در خط پدری ۱۰۰ نر و ۲۰۰ ماده و در خط مادری ۲۰ نر و ۲۰۰ ماده) برای ۱۰ نسل دیگر باهم آمیزش تصادفی دادند و از دهمین نسل آنها جمعیت مورد نظر (۱۰ نر و ۱۰۰ ماده) که برای ده نسل دیگر آمیزش داده شدند، تشکیل شد. آثار آلی و SNP کدها برای ۱۰ نسل آخر ذخیره شدند. جدول ۱ خلاصه ساختار جمعیت شبیه‌سازی شده از جمعیت تاریخی را نشان می‌دهد. ساختار ژنومی شامل ۳ کروموزوم با طول ۱۰۰ سانتی مورگان بودند که بر

### 1. Quantitative trait neocleotide (QTN)

### 2. Historical population (HP)

قرار داشتند شبیه‌سازی شدند. اثر آلی QTL از یک توزیع گاما با شیب پارامتر ۰/۴ در نرم افزار QMSim پیاده سازی شد و نمونه برداری انجام گرفت. نرخ جهش بازگشت کننده برای QTL و نشانگر  $1 \times 10^{-4}$  (۹ و ۳۳) در نظر گرفته شدند.

روی هر کروموزوم ۱۰۰۰ نشانگر دو آلی با فراوانی - های برابر (۰/۵) در جمعیت پایه جانمایی شدند (فراوانی‌های آلی بدلیل عدم وزن دهی به یک آل برابر در نظر گرفته شد). ۵۰ QTL دو آلی با فراوانی برابر که به صورت تصادفی در روی هر کروموزوم

جدول ۱: خلاصه ساختار جمعیت شبیه سازی شده

Table 2. Summary of the simulated population structure

| حذف شده | مادر Dam | حذف شده Culled | پدر Sire | انتخاب شده Selected | ماده Female | انتخاب شده Selected | نر Male | نر % Male % | فرزندان Progeny | جمعیت population                               |
|---------|----------|----------------|----------|---------------------|-------------|---------------------|---------|-------------|-----------------|--|
| 3120    | 16000    | 624            | 1600     | 3120                | 8400        | 624                 | 8040    | 0.49        | 16440           | پدر بزرگ مادری (۱)<br>Sire of dam<br>مادر بزرگ |
| 3120    | 16000    | 624            | 1600     | 3129                | 8400        | 624                 | 8040    | 0.49        | 16440           | مادری (۲)<br>Dam of dam<br>پدر بزرگ پدری       |
| 3120    | 16000    | 1560           | 4000     | 3120                | 8400        | 1560                | 8100    | 0.49        | 16500           | (۳)<br>Sire of sire<br>مادر بزرگ پدری          |
| 3120    | 16000    | 1560           | 4000     | 3120                | 8400        | 1560                | 8100    | 0.49        | 16500           | (۴)<br>Dam of sire<br>مادر (C)                 |
| 360     | 2000     | 72             | 200      | 3120                | 1200        | 72                  | 1020    | 0.46        | 2220            | Dam (D)  |
| 360     | 2000     | 360            | 1000     | 3120                | 1200        | 360                 | 1100    | 0.48        | 2300            | Sire   |
| 1180    | 6000     | 236            | 600      | 1180                | 3100        | 236                 | 3010    | 0.49        | 6110            | تائید Validation                               |

- 1)  $y = Xb + Tg + e$
- 2)  $y = Xb + Za + e$
- 3)  $y = Xb + Za + Tg + e$
- 4)  $y = Xb + Za + Wd + e$
- 5)  $y = Xb + Za + Wd + Tg + e$
- 6)  $y = Xb + Za + Tg + Wd + Sep + e$

در مدل‌های بالا،  $b$  بردار آثار ثابت (شامل اثر جنس)،  $a$  بردار آثار چندژنی حیوان،  $d$  بردار آثار غالبیت جایگاه نشانگری،  $g$  آثار افزایشی نشانگرها،  $ep$  بردار

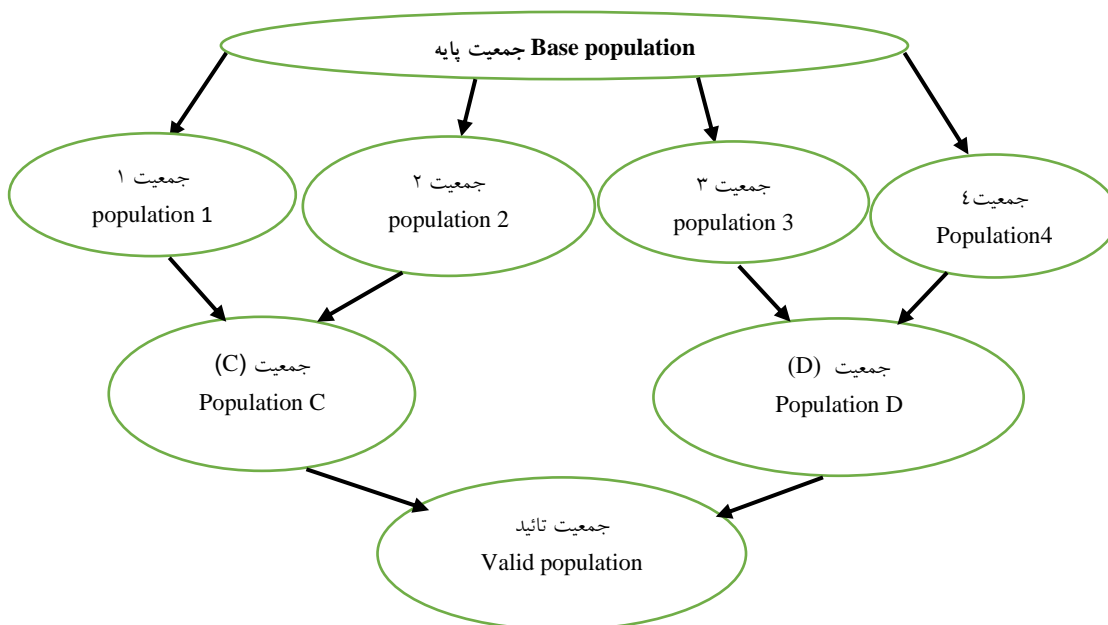
توزیع آثار نشانگری و چندژنی حیوان در نسل آخر جمعیت پایه با وراثت پذیری ۰/۳ معرفی شد و ۵۰ QTL با آثار افزایشی، غالبیت و اپیستاتیک در مجموع ۱۰ درصد واریانس فنوتیپی را شامل شدند.

### مدل آماری

برآورد مؤلفه‌های واریانس و ارزش اصلاحی ژنومی با استفاده از آثار نشانگرها (افزایشی، غالبیت و اپیستاتیک) و در نظر گرفتن آثار چندژنی حیوان در قالب ۶ مدل آماری به شرح ذیل انجام شد:

خطای باقیمانده است. ماتریس‌های  $X, Z, W, T$  و  $S$  به ترتیب ماتریس‌های ارتباط دهنده آثار ثابت، چند ژنی حیوان، غالبیت نشانگرها، افزایشی نشانگرها و آثار اپیستاتیک به رکوردهای فنوتیپی هستند.

آثار متقابل (اپیستاتیک) است، آثار اپیستاتیک به  $AA$  بردار آثار اپیستاتیک افزایشی در افزایشی،  $AD$  بردار آثار اپیستاتیک افزایشی در غالبیت،  $DA$  بردار آثار اپیستاتیک غالبیت در افزایشی،  $DD$  بردار آثار اپیستاتیک غالبیت در غالبیت تقسیم می‌شود و  $e$  بردار



شکل ۱: شماتیک جمعیت شبیه‌سازی شده و جریان ژن از جمعیت پایه

Figure 1. The schematic of simulated population and gene flow from base population

(۲۳) کامل توصیف شده است. برای آثار غالبیت، توزیع مشابه آثار افزایشی در نظر گرفته شد (۴). مدلی که مورد استفاده قرار می‌گیرد عبارت است از:

$$y_i = \eta_i + \epsilon_i \quad (7)$$

$\eta_i$  پیش‌بینی کننده خطی،  $\epsilon_i$  باقیمانده با میانگین صفر و واریانس  $w_i^2 \sigma_{\epsilon}^2$  جای که  $y = \{y_1, y_2, \dots, y_n\}$  و  $\epsilon = \{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$  و  $\eta = \{\eta_1, \eta_2, \dots, \eta_n\}$

پیش‌بینی کننده خطی تابع احتمال شرطی را ارائه می‌دهد که ساختار آن بصورت زیر می‌باشد:

$$y = 1\mu + \sum_{j=1}^J X_j \beta_j + \sum_{i=1}^I u_i \quad (8)$$

در بیز لاسو توزیع پیشین آثار نشانگرها بک توزیع نمایی دوگانه<sup>۱</sup> است. آثار همه نشانگرها به طور مستقل و یکسان به صورت تابع نمایی دوگانه توزیع می‌شود. این توزیع پیشین واریانس یکسان با توزیع پیشین نامعتبر برای همه نشانگرها را طراحی می‌کند، اما دارای دنباله‌های ضخیم‌تر از توزیع پیشین نرمال است (۶ و ۱۰). مدل با دو مؤلفه واریانس (افزایشی و باقیمانده)، مدل بیز لاسو بهبود یافته نامیده می‌شود (۲۳). اجزای عملی این مدل از طریق نمونه‌گیری گیبس، از جمله توزیع‌های شرطی کامل پسین، توسط

## 1. Double- Exponential (DE)

و برای آن توزیع گاما اختصاص یافت. بیز لاسو شکل خاصی از حداقل مربعات جریمه شده است که مجموع مربعات باقیمانده را حداقل می‌کند. در بیز لاسو توزیع پیشین واریانس آثار نشانگرها به صورت زیر است:

$$\text{Var}(\beta_j | \lambda^2, \sigma_\varepsilon^2) = \sigma_\beta^2 = 2 \frac{\sigma_\varepsilon^2}{\lambda^2} \quad (12)$$

بنابراین،  $\frac{\sigma_\beta^2}{\sigma_\varepsilon^2} = \frac{2}{\lambda^2}$  و لامبدا برابر  $\lambda = \sqrt{2 \frac{1-k^2}{k^2} (n^{-1} \sum_i \sum_j X_{ij}^2)}$  می‌باشد.

برای انجام مدل‌های افزایشی و غالبیت از نرم افزار GS3 (22) استفاده شد.

برای بدست آوردن آثار اپیستاتیک از مدل ژنتیکی و روش آماری پیشنهادی زو (49) استفاده شد. آثار افزایشی، غالبیت و اپیستاتیک (Ad:Dom:Epi) ساده‌ترین مدل ژنتیک کمی شامل اثر اپیستاتیک یک مدل دو جایگاهی است که در هر لوکوس دو آلل قرار دارد. عمل ژن اپیستاتیک روی متوسط اثر آلل‌ها و انحراف غالبیت و در نتیجه روی واریانس افزایشی و غالبیت تاثیر می‌گذارد (32).

(13)

$$y = X\beta + \sum_{k=1}^{nm} A_{t_j} a_j + \sum_{k=1}^{nm} D_{t_j} d_j + \sum_{k=1}^{m-1} \sum_{k=k+1}^m aal_k l_k + \sum_{k=1}^{m-1} \sum_{k=k+1}^m adl_k l_k + \sum_{k=1}^{m-1} \sum_{k=k+1}^m dal_k l_k + \sum_{k=1}^{m-1} \sum_{k=k+1}^m ddl_k l_k + e$$

در معادله بالا  $aal_k l_k$ ،  $adl_k l_k$ ،  $dal_k l_k$  و  $ddl_k l_k$  به ترتیب آثار اپیستاتیک  $A \times A$  (افزایشی در افزایشی)،  $A \times D$  (افزایشی در غالبیت)،  $D \times A$  (غالبیت در افزایشی) و  $D \times D$  (غالبیت در غالبیت) بین  $m$  جایگاه ژنتیکی  $k$  و  $k'$  ( $k < k' = 1, \dots, p$ ) هستند.

جایی که  $X\beta$  آثار غیر ژنتیکی را نشان می‌دهد. هنگامی که آثار ژنتیکی بعنوان متغیر تصادفی توزیع نرمال میانگین صفر و یک واریانس معمول در تمام

در معادله بالا  $\mu$  عرض از مبدا،  $X_j$  ماتریس طرح برای پیش‌بینی‌کننده‌ها است،  $X_j = \{x_{ijk}\}$ ،  $\beta_{jk}$  بردار آثار مرتبط با ستون  $X_j$  و  $u_i = \{u_{i1}, \dots, u_{im}\}$  بردار آثار تصادفی است.

با توجه به فرمول بالا توزیع شرطی از داده‌ها به صورت زیر می‌باشد:

(9)

$$p(y|\theta) = \prod_{i=1}^n N(y_i | \mu + \sum_{j=1}^J \sum_{k=1}^{K_j} x_{ijk} \beta_{jk} + \sum_{t=1}^t u_i \cdot \theta_{tE}^2 w_i^2)$$

که در آن  $\theta$  نشان دهنده مجموعه از ناشناخته‌ها که شامل عرض از مبدا، ضریب رگرسیون، دیگر آثار تصادفی و واریانس باقیمانده است. تفاوت اصلی روش‌های بیز در انتخاب توزیع پیشین برای تأثیرات نشانگر است. در روش بیز لاسو (34) توزیع آثار نشانگرها به صورت توزیع نمایی دوگانه هستند. این توزیع مانند توزیع  $t$  دارای دم درازی است، اما تعداد بیشتری آثار کوچک غیر صفر دارد. روش بیز لاسو از توزیع دونمایی برای توزیع آثار QTL (مشروط به پارامتر تنظیم‌کننده لامبدا) استفاده می‌کند. برآوردهای لاسو می‌توانند از مدل توزیع پسین بیزی تحت توزیع پیشین دونمایی مستقل برای آثار QTL‌ها استخراج و استنباط شوند. توزیع دونمایی همچنین می‌تواند به عنوان توزیع مختلطی از توزیع نرمال با واریانس نمونه‌گیری شده از یک توزیع نمایی بیان شود.

$$P(\beta_j | \lambda, \sigma_\varepsilon) = DE \left( \beta_j \mid 0, \frac{\lambda}{\sigma_\varepsilon^2} \right) \quad (10)$$

توزیع پیشین حاشیه‌ای ضرایب رگرسیونی به این صورت است:

$$P(\beta_j | \lambda) = \int N(\beta_j | 0, \sigma_\varepsilon^2 \tau_j^2) \text{EXP}(\tau_j^2 | \lambda^2) \tau_j^2 \quad (11)$$

رابطه بالا توزیع نمایی دوگانه است. این توزیع تراکم بیشتری در صفر دارد که نشان می‌دهد دارای افت شدیدی برای نشانگرهای با اثر نسبتاً کوچک و افت پایین‌تر برای نشانگرهای با اثر قابل ملاحظه می‌باشد. در معادله فوق،  $\lambda$  ناشناخته فرض شده است

برای برآورد مؤلفه‌های واریانس و ارزش‌های اصلاحی ژنومی در قالب مدل ایستاتیک از برنامه نویسی در برنامه R (۳۶) و از بسته‌های آماری *glmnet* (۱۵) و *EBglmnet* (۱۹) استفاده شدند.

همبستگی بین مقادیر واقعی ارزش اصلاحی و ارزش اصلاحی ژنومیکی پیش‌بینی شده به عنوان معیار صحت برآورد ارزش اصلاحی استفاده شد  $r(TBV, GEBV) = r^*$  (۳). جهت کنترل خطای نوع اول در برآورد اثر نشانگرها از شاخص بنفرونی<sup>۱</sup> در سطح احتمال یک درصد استفاده شد. در برآورد اثرات نشانگرها برای هر مدل *QTN*‌هایی که در مجاورت *QTL*‌های صفت مورد مطالعه باشند و اثراتشان از حد آستانه *P* محاسبه شده با روش بنفرونی بالاتر باشند به‌عنوان مثبت واقعی و اگر از حد آستانه کمتر باشد به‌عنوان منفی کاذب می‌باشند و *QTN*‌هایی که در مجاورت هیچ *QTL*‌یی نباشند و اثراتشان از شاخص بنفرونی بالاتر باشد به‌عنوان مثبت کاذب و اگر از شاخص بنفرونی پایین‌تر باشند به‌عنوان منفی واقعی شناخته شدند.

### نتایج و بحث

نتایج تجزیه و تحلیل اجزای واریانس ژنتیکی افزایشی و غیرافزایشی در قالب ۶ مدل مختلف در جدول ۲ ارائه شده است. بیشترین مقدار واریانس ژنتیکی کل اعم از افزایشی و غیرافزایشی مربوط به مدل ۶ بود که دربرگیرنده آثار ژنتیکی افزایشی و غیرافزایشی به ویژه آثار ایستاتیک بودند. هنگامی که آثار ایستاتیک و غالبیت با هم در مدل قرار گرفتند از واریانس باقیمانده کاسته شد و به قسمت واریانس ژنتیکی افزوده شد و این باعث افزایش واریانس ژنتیکی کل شد. بعد از مدل ۶، بیشترین واریانس

نشانگرها یا جفت نشانگرها را دارند یک مدل ترکیب وجود دارد. امید ریاضی  $y$  برابر است با  $E(y) = X\beta$  و واریانس ماتریکس  $y$  برابر است با:

$$Var(y) = K_a \sigma_a^2 + K_d \sigma_d^2 + K_{aa} \sigma_{aa}^2 + K_{ad} \sigma_{ad}^2 + K_{da} \sigma_{da}^2 + K_{dd} \sigma_{dd}^2 + I \sigma^2 \quad (14)$$

جایی  $\sigma^2$  نشان دهنده وریانس خطای باقیمانده است. ماتریکس  $K$  یک ماتریس روابط خویشاوندی است که برای محاسبه ساختار جمعیت استفاده کردیم. ماتریس  $K$  می‌تواند از اطلاعات شجره برای همه ورودی‌ها یا بوسیله نشانگرهای ژنتیکی تخمین زده شود. چگونگی استخراج ماتریس  $K$  از اطلاعات نشانگرها بطور گسترده شرح داده شده است (۲۱) و (۴۲).

برآورد کننده لاسو به صورت زیر بدست آمد:

$$\hat{\beta} = \arg \min \{ (y - X^T \beta)^T \hat{\Sigma}^{-1} (y - X^T \beta) + \lambda_n \sum_{k=1}^m \omega_k |\beta_k| \} \quad (15)$$

$\lambda_n$  پارامتر تنظیم برای کنترل مقدار تنظیم است.  $\omega_n$  مقدار وزنی برای اثر ژنتیکی  $k$  امین است. بطوری‌که هر اثر مقدار مختلف انقباض را دریافت خواهد کرد و  $\hat{\Sigma}$  برآورد کننده ثابت  $\Sigma$  است.

ماتریس واریانس نشان داده شده در معادله ۱۴ می‌تواند به صورت زیر نوشت:

$$Var(y) = (K_a \lambda_a + K_d \lambda_d + K_{aa} \lambda_{aa} + K_{ad} \lambda_{ad} + K_{da} \lambda_{da} + K_{dd} \lambda_{dd} + I) \sigma^2 \quad (16)$$

$$K_a \lambda_a + K_d \lambda_d + K_{aa} \lambda_{aa} + K_{ad} \lambda_{ad} + K_{da} \lambda_{da} + K_{dd} \lambda_{dd} + I) \sigma^2$$

در معادله بالا  $\lambda_a = \sigma_a^2 / \sigma^2$ ،  $\lambda_d = \sigma_d^2 / \sigma^2$  و  $\lambda_{da} = \sigma_{da}^2 / \sigma^2$ ،  $\lambda_{ad} = \sigma_{ad}^2 / \sigma^2$ ،  $\lambda_{aa} = \sigma_{aa}^2 / \sigma^2$  و  $\lambda_{dd} = \sigma_{dd}^2 / \sigma^2$  نسبت‌های واریانس هستند. شش نسبت واریانس چندژنی در یک بردار  $\lambda = [\lambda_a, \lambda_d, \lambda_{aa}, \lambda_{ad}, \lambda_{da}, \lambda_{dd}]$  جمع می‌شود. اگر  $K = [K_a, K_d, K_{aa}, K_{ad}, K_{da}, K_{dd}]^T$  باشد، پس داریم  $Var(y) = V = (K\lambda + I) \sigma^2$

1. Bonferroni
2. Quantitative trait nucleotide



بودند که این مورد انتظار بود، زیرا کامل تر شدن مدل، واریانس خطا را کاهش می‌دهد. نسبتی از واریانس فنوتیپی که بواسطه آثار ژنتیکی ایجاد می‌شود ( $h_x = \sigma_x^2 / \sigma_p^2$ ) تحت عنوان وراثت‌پذیری عام یا ضریب تعیین ژنتیکی صفت در مدل ۶ بیشترین و در مدل ۱ کمترین بود. با اضافه شدن آثار ژنتیکی غیرافزایشی به مدل میزان وراثت‌پذیری عام افزایش یافت که مطابق با نتایج سایر محققین بود (۱۱ و ۴۹). ضریب وراثت‌پذیری خاص که نسبت واریانس ژنتیکی افزایشی به واریانس فنوتیپی یا میزان شباهت رکورد نتاج و والدین در مدل ۱ بیشترین و در مدل ۶ کمترین بود. در مدل‌های ژنوتیپی تخمین واریانس ژنتیکی افزایشی زمانیکه غالبیت و اپیستاتیک در مدل گنجانده می‌شود ممکن است کاهش یابد (۳۳ و ۴۳) که نتایج مشابهی در مطالعه حاضر بدست آمد. وراثت‌پذیری عام و وراثت‌پذیری خاص در مدل ۱ مساوی بودند. بدلیل اینکه سایر آثار غالبیت و اپیستاتیک برآورد نشده است در نتیجه این وراثت‌پذیری‌ها با هم برابر شدند.

سهم واریانس ژنتیکی در مدل‌های مختلف بستگی به معماری و توزیع آثار ژنی دارد و می‌تواند در مطالعات مختلف نتایج مختلفی به همراه داشته باشد. نسبت اجزای واریانس‌های ژنتیکی افزایشی و غیرافزایشی به تفکیک هر مدل در شکل ۲ ارائه شده است. نتایج این مطالعه نشان داد که به کار بردن مدل چندژنی با تعداد زیاد نشانگرهای ژنتیکی در مدل بیز لاسو نسبت زیادی از وراثت‌پذیری را توضیح می‌دهد. در مدل دوم تا چهارم واریانس چندژنی حیوان بر اساس شجره بیشترین بود که نشان دهنده نقش ژن‌های کوچک اثر متعدد با توزیع نرمال بر روی صفت می‌باشد. در مدل ۶ که به منظور توصیف بهتر آثار متقابل نشانگرها بر میزان تنوع

ژنتیکی کل به ترتیب مربوط به مدل ۵ و ۲ بودند و کمترین واریانس ژنتیکی کل از مدل ۱ حاصل شد و مدل‌های ۳ و ۴ تفاوت بسیار اندک با هم داشتند. نتایج این تحقیق تأیید می‌نماید که در کنار آثار ژنتیکی افزایشی، آثار غالبیت نشانگرها منجر به بهبود بیشتر برآوردها شدند، ولی آثار اپیستاتیک تغییرات زیادی در برآوردها به ویژه ژنتیکی کل ایجاد نمود. غالبیت منبع هتروزیس است و باید در طراحی سیستم‌های پیش‌بینی ژنومی و جفت‌گیری قرار گیرد تا عملکرد تولید به حداکثر برسد (۱۸). واریانس غالبیت در مدل ۵ و ۶ بسیار کوچک تخمین زده شد. واریانس غالبیت در مدل ۴ اندکی افزایش یافت که ممکن است بیش از مقدار واقعی واریانس غالبیت باشد که به دلیل همبستگی زیاد بین سایر مؤلفه‌های واریانس پنهان مانده ایجاد می‌شود (۲۸). در مطالعه‌ای با مدل‌های مختلف آثار افزایشی و غیرافزایشی شامل غالبیت نسبت واریانس غالبیت به واریانس ژنتیکی کل و فنوتیپی در صفات مرتبط با رشد به ترتیب ۱۴/۷ و ۳/۸ درصد گزارش شدند (۲۸). مقدار واریانس ژنتیکی افزایشی چندژنی حیوان به واسطه شجره در مدل‌های مختلف تغییرات زیادی نداشت، هر چند که اختلافات جزئی با هم داشتند و هنگامی که در مدل فقط اثر چندژنی حیوان وجود داشت، کمتر از مدل‌هایی بود که آثار افزایشی نشانگرها در مدل وجود دارند و بدین ترتیب اضافه نمودن آثار نشانگری در مدل می‌تواند مقدار واریانس ژنتیکی گمشده یا مخفی را نمایان سازد. البته مقداری تداخل بین آثار افزایشی چندژنی و آثار افزایشی نشانگری و غالبیت وجود دارد که در سایر تحقیقات هم گزارش شده است (۲۹ و ۵۲).

واریانس باقیمانده در مدل ۶ که شامل آثار افزایشی، غالبیت و اپیستاتیک نشانگرها به‌علاوه آثار چندژنی حیوان است، کمترین و در مدل ۱ بیشترین

صفت استفاده شد. در این مدل آثار متقابل به چهار جزء شامل افزایشی × افزایشی، افزایشی × غالبیت، غالبیت × افزایشی و غالبیت × غالبیت طبقه‌بندی شدند، نتایج نشان داد که همانند نتایج (۴) سهم جزء افزایشی × افزایشی بر تنوع صفت بیشترین و غالبیت × غالبیت کمترین بودند. همانند نتایج (۳۲) زمانی که آثار اپیستاتیک در مدل گنجانده شد به واریانس ژنتیکی کل یک جزء دیگر به نام واریانس اپیستاتیک علاوه بر واریانس غالبیت و افزایشی اضافه گردید و این باعث افزایش واریانس ژنتیکی کل و در نتیجه باعث افزایش وراثت پذیری عام شد. آثار اپیستاتیک اغلب برای تنوع صفات پیچیده مهم تلقی می‌شود. با این حال، تجزیه آثار اپیستاتیک از درون ژنوم دشوار است زیرا واریانس اپیستاتیک اغلب با واریانس‌های افزایشی و باقیمانده اشتباه گرفته می‌شود (۵۰). برای جدا کردن واریانس اپیستاتیک از واریانس‌های افزایشی و غالبیت به اطلاعات شجره نیاز است، ولی در این مطالعه از اطلاعات مربوط به نشانگرها برای مدل‌سازی اثرات اپیستاتیک استفاده شد. مشکلی که در آثار اپیستاتیک وجود دارد به ارث رسیدن جز افزایشی در غالبیت یا برعکس غالبیت در افزایشی می‌باشد. از آنجا که رانش ژنتیکی یک فرآیند تصادفی است، انتظار می‌رود نوسانات گسترده‌ای در پارامترهای ژنتیکی کمی از یک نسل (یا جمعیت) به نسل دیگر، حتی برای جمعیت‌هایی با اندازه‌های مؤثر از چند صد باشد. برآورد اجزای واریانس نیز ممکن است بین نمونه‌های مختلف از همان جمعیت متفاوت باشد، به خصوص هنگامی که نمونه‌ها کوچک هستند (۴۸). تخمین‌های تحقیق حاضر بهینه است، چون اندازه نمونه در جمعیت مورد مطالعه مطلوب بود و هم اینکه طبق مطالعه (۳۴) بیز لاسو فواصل قابل

اعتمادتری را برای تخمین وراثت پذیری ارائه می‌دهد. تنها مشکلی که احتمالاً می‌تواند مشکل ساز باشد رانش تصادفی ژن‌های کنترل‌کننده صفت مورد مطالعه در جمعیت پایه است که این امکان وجود دارد که انتخاب در جمعیت‌های بعدی نتوانسته جبران این رانش تصادفی را بکند و در نتیجه برآورد واریانس ژنتیکی افزایشی و در نتیجه وراثت‌پذیری خاص کمتر بدست آمد.

صحت پیش‌بینی ارزش‌های اصلاحی ژنومی با مدل‌های مختلف در جدول ۳ ارائه شده است. نتایج تحقیق نشان داد که مدل‌های ۳، ۴، ۵ و ۶ بیشترین صحت پیش‌بینی ارزش‌های اصلاحی را دارد و مدل ۱ کمترین صحت را داشت. همچنین، نتایج نشان داد با کامل‌تر شدن مدل‌ها و اضافه شدن اثرهای چندژنی حیوان در مدل پایه علاوه بر آثار ژنتیکی افزایشی و غیرافزایشی نشانگرها از جمله غالبیت و اپیستاتیک، صحت برآورد ارزش‌های اصلاحی ژنومی بهبود می‌یابد که با نتایج سو و همکاران (۵۰) همخوانی دارد. در مدل ۱ هنگامی که واریانس غیرافزایشی و واریانس چندژنی بر اساس شجره از مدل برداشته شد موجب کاهش چشمگیر در صحت برآورد ژنتیکی گردید. این می‌تواند به دلیل کاهش مقدار اطلاعات آماری موجود در مدل باشد که در تخمین اجزای واریانس استفاده شد. هنگامی که در مدل آثار اپیستاتیک جانمایی شد تأثیری بر روی صحت برآورد ارزش اصلاحی ژنومی نداشت. صحت پیش‌بینی در روش بیز لاسو بستگی به تعداد  $QTL$  و توزیع اثرات آنها دارد. تعداد ۱۰ الی ۱۰۰  $QTL$  در بیز لاسو می‌تواند عملکرد بهتری داشته باشد و به همین دلیل در اینجا ۵۰  $QTL$  در نظر گرفته شد.

جدول ۲: برآورد مؤلفه‌های واریانس ژنتیکی با استفاده از مدل‌های مختلف

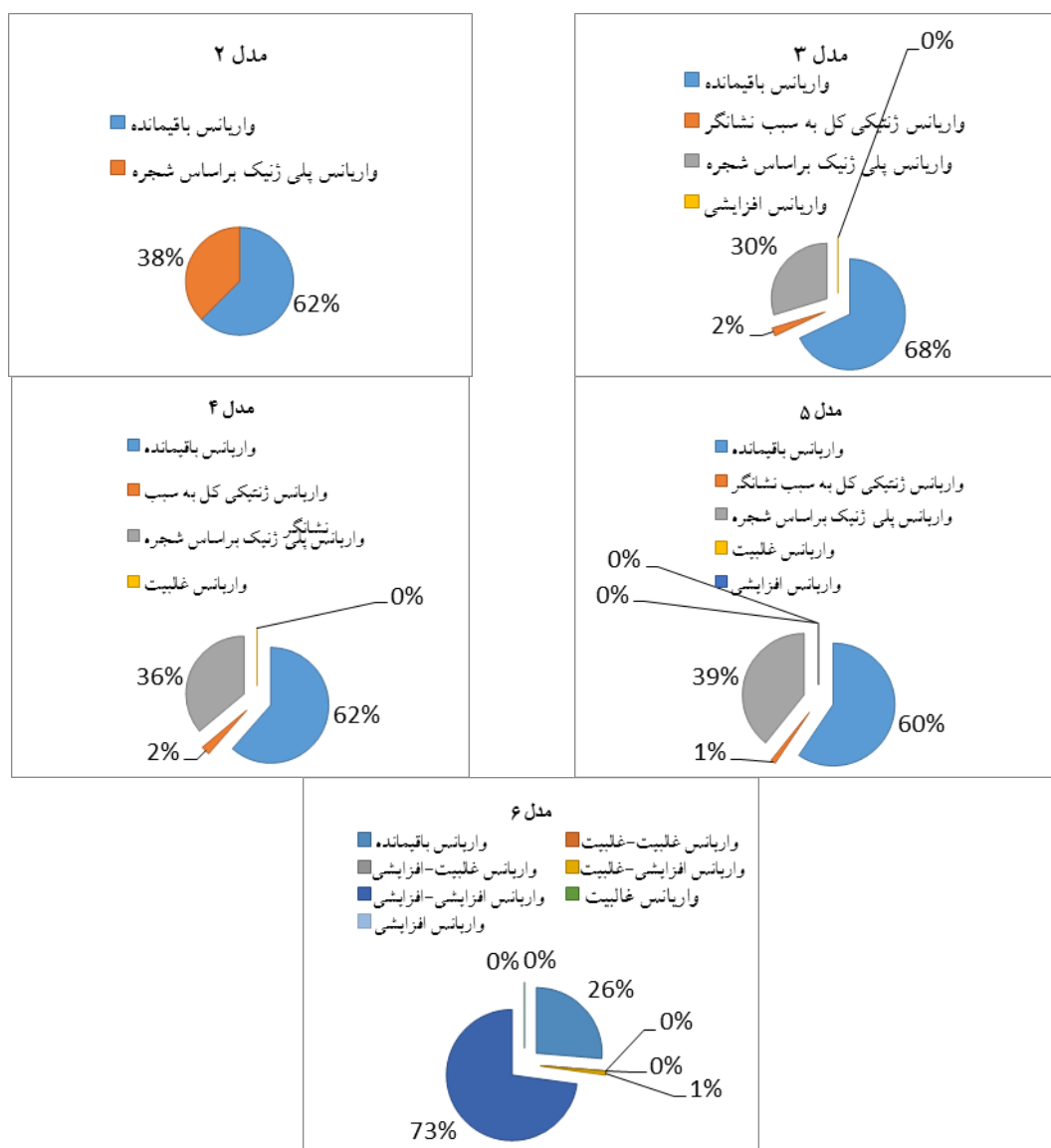
Table 2. Estimation of genetic variance components using different models

| لامبدا<br>lambda | واریانس<br>کل ژنتیکی<br>Total genetic<br>variance | واریانس<br>ژنتیکی کل<br>به سبب<br>نشانگرها<br>Total genetic<br>variance<br>based on<br>markers | واریانس<br>باقیمانده<br>Residual<br>variance | واریانس<br>چند زنی، بر اساس<br>شجره<br>Polygenic variance<br>based on pedigree | واریانس<br>غالبیت-<br>غالبیت<br>Dominant<br>e-<br>dominant<br>e variance | واریانس<br>غالبیت-<br>افزایشی<br>Dominant<br>e-additive<br>variance | واریانس<br>غالبیت<br>افزایشی<br>Additive<br>dominant<br>variance | واریانس<br>افزایشی-<br>افزایشی<br>Additive-<br>additive<br>variance | واریانس اثر<br>SNP غالبیت<br>SNP<br>Dominance<br>variance<br>effect | واریانس اثر<br>SNP افزایشی<br>Additive<br>SNP<br>variance<br>effect | اجزای واریانس<br>Variance<br>component |
|------------------|---|--|--|--|--|---|--|---|---|---|--|
| 5.25             | 0.22  | -  | 0.78   | -  | -  | -   | -  | -   | -   | 0.22  | 1                                      |
| 5.25             | 0.36  | -  | 0.60   | 0.36   | -  | -   | -  | -   | -   | -   | 2                                      |
| 5.25             | 0.303   | 0.022  | 0.64   | 0.28   | -  | -   | -  | -   | -   | 0.000086  | 3                                      |
| 5.25             | 0.31  | 0.025  | 0.63   | 0.28   | -  | -   | -  | -   | 0.000066  | -   | 4                                      |
| 5.25             | 0.385   | 0.012  | 0.56   | 0.37   | -  | -   | -  | -   | 0.0000011   | 0.000047  | 5                                      |
| 5.25             | 0.74  | -  | 0.26   | -  | 0.00002  | 0.00002   | 0.009  | 0.72  | 0.00002   | 0.00002   | 6                                      |

جدول ۳: تخمین پارامترهای ژنتیکی و صحت پیش بینی ارزش‌های اصلاحی ژنومیکی بدست آمده با مدل‌های مختلف

Table 3. Estimation of genetic parameters and accuracy of prediction of genomic breeding values obtained by different models

| صحت پیش بینی ارزش‌های اصلاحی<br>Accuracy of prediction of breeding values | وراثت‌پذیری خاص<br>Hertability in the narrow sense | وراثت‌پذیری عام<br>Hertability in the broad sense | مدل<br>Model |
|---|--|---|--------------|
| 0.42  | 0.22   | 0.22  | 1            |
| 0.83  | -  | 0.375   | 2            |
| 0.84  | 0.000091   | 0.32  | 3            |
| 0.84  | -  | 0.328   | 4            |
| 0.84  | 0.00005  | 0.406   | 5            |
| 0.84  | 0.000003   | 0.739   | 6            |



شکل ۲: نسبت اجزای واریانس ژنتیکی افزایشی و غیر افزایشی در قالب مدل‌های مختلف دربرگیرنده آثار ژنتیکی افزایشی و غیر افزایشی

Figur 2. Ratio of additive and non-additive genetic variance components in different models including additive and non-additive genetic effects

محاسبه متفاوت می‌باشند. علت این تفاوت بستگی به نظر گرفتن آثار تصادفی هم چون آثار افزایشی و غالبیت نشانگرها دارد که در شناسایی QTN های معنی دار تأثیرگذار هستند. مدلی که بتواند تعداد جایگاه‌های بیشتر مثبت واقعی با مثبت کاذب کمتر را تخمین برند مدل دقیق‌تری می‌باشد (۳۰).

در تحقیق حاضر نیز مدل با آثار افزایشی نشانگرها بهترین مدل برای شناسایی QTN های معنی دار بود. هم چنین، مدل مدل ۱ بیشترین تعداد مثبت کاذب را شناسایی کرد، پس نمی‌تواند مدل مناسبی باشد. مدل ۶ کمترین مثبت کاذب را شناسایی کرد، اما به جهت اینکه تعداد مثبت واقعی کمتری را شناسایی می‌نماید، نمی‌تواند مدل مناسبی باشد. مدل ۵ با کمترین QTN مثبت کاذب و به نسبت بیشترین مثبت واقعی مدل مناسبی در ارزیابی ژنتیکی صفت مورد مطالعه می‌باشد.

زمانی که آثار چندژنی از مدل برداشته شد، تعداد QTN های مثبت کاذب افزایش یافت. مدل یک بیشترین تعداد مثبت کاذب را دارد و با قرار دادن آثار چند ژنی حیوان در مدل از تعداد QTN مثبت کاذب کاسته شد (۱۲ و ۲۷) که با مطالعه حاضر مطابقت دارد.

دلیل اینکه با اضافه نمودن آثار چندژنی تعداد مثبت کاذب کاهش می‌یابد این است که در یک مجموعه داده، تمام نشانگرها و QTN از طریق شجرنامه در بین حیوانات در ارتباط هستند. در نتیجه، هر نشانگر می‌تواند به طور اتفاقی، با QTN با فاصله کمی یا حتی با کروموزوم دیگر ارتباط داشته باشد، بنابراین به نظر می‌رسد دارای اثری است که محصول ساختار شجره است.

با این وجود اگر هزاران QTL در تنوع ژنتیکی صفت دخیل باشد طبق یافته‌های (۳۸) هر یک اثر بسیار اندکی بر صحت پیش‌بینی ارزش اصلاحی داشت، مانند یافته‌های که در مورد قد انسان وجود دارد. صحت پیش‌بینی ارزش اصلاحی برآورد شده با مدل‌های ۳، ۴، ۵ و ۶ برابر بودند (اگر چه آثار غالبیت و اپیستاتیک ضعیف تخمین زده شدند) که با نتایج (۴) مطابقت داشت. میزان صحت پیش‌بینی ارزش اصلاحی برآورد شده در صفات مرتبط با توده بدن با مدل‌های مختلف مشمول آثار ژنتیکی افزایشی و غیرافزایشی شامل غالبیت و اپیستاتیک در دامنه ۴۶/۱ تا ۶۹ درصد گزارش شدند که بیش‌ترین میزان آن مربوط به مدل با آثار اپیستاتیک و افزایشی نشانگرها بود (۲۸). یکی دیگر از نتایج مدل‌های مختلف برآورد آثار نشانگرها تعیین تعداد جایگاه‌های مثبت واقعی است.

#### خطاها در تعیین مکان و آثار نشانگرها: تعداد QTN

مثبت واقعی و کاذب و همچنین تعداد QTN منفی کاذب و مثبت در قالب مدل‌های مختلف در جدول ۴ ارائه شده است. ایده انتخاب ژنومی این است که از آزمون معنی‌داری چشم‌پوشی کند و به سادگی اثر همه ژن‌ها یا موقعیت‌های کروموزومی را به‌طور همزمان تخمین بزند (۱۲ و ۳۱). شناسایی QTN های صفات کمی در گونه‌های اهلی به خصوص برای صفاتی که دارای ارزش اقتصادی هستند، موجب می‌شود تا دقت برآورد ارزش اصلاحی دام‌ها افزایش و فاصله نسلی کاهش یابد که نتیجه آن افزایش سرعت پیشرفت ژنتیکی در این صفات است. هم چنین، شناسایی این QTN ها موجب شناخت بیشتر ژن‌ها و علت تنوع صفات مهم دام‌ها می‌شود. از طرف دیگر، تعداد QTN های معنی دار بر اساس مدل و نوع روش آماری مورد

جدول ۴: تعداد QTN مثبت و منفی شناسایی شده در مدل‌های آماری مختلف

**Table 4. Number of positive and negative QTNs identified in different statistical models**

| مدل   | مثبت واقعی    | مثبت کاذب      | منفی واقعی    | منفی کاذب      |
|-------|---------------|----------------|---------------|----------------|
| model | True positive | False positive | True negative | False negative |
| 1     | 61            | 1003           | 1847          | 89             |
| 2     | *             | *              | *             | *              |
| 3     | 30            | 572            | 2278          | 120            |
| 4     | 0             | 0              | 2850          | 150            |
| 5     | 32            | 557            | 2293          | 118            |
| 6     | 12            | 253            | 2600          | 138            |

### نتیجه‌گیری کلی

به‌طور کلی در صفاتی که از الگوی وراثت‌مندلی پیروی نمی‌کنند، عوامل متعددی از جمله محیطی و ژنتیکی (چندژنی، اثر متقابل بین جایگاه‌های ژن) در بروز تنوع نقش دارند. پیش‌بینی اجزای واریانس ژنتیکی اعم از افزایشی و غیرافزایشی شامل آثار غالبیت و اپیستاتیک و همچنین تعیین نشانگرهای مثبت یا منفی جایگاه‌های کنترل‌کننده صفات کمی با استفاده از مدل‌های که تخمین‌های درستی از هریک را بدهد می‌تواند در فرآیند بهبود بخشیدن این صفات از طریق انتخاب کمک‌شایانی را انجام دهد. نتایج مطالعه حاضر نشان داد که هریک از اجزای واریانس افزایشی و غالبیت به‌طور جزئی بر

روی صفت مورد مطالعه تاثیر می‌گذارد و اثرات چندژنی و اپیستاتیک در جمعیت مورد مطالعه بیشترین تاثیر را در بروز صفت دارند و این بدان معنی است که هر جایگاه ژنی بر اساس یک روند افزایشی عمل کرده و در بروز یک دیگر اثر هم‌افزایی دارند. نادیده گرفتن آثار غیرافزایشی تخمین‌های بالایی از واریانس افزایشی را می‌دهد و همچنین باعث برآورد بیشتر تعداد QTN‌های مثبت‌های کاذب در مدل می‌شود. بنابراین، بهتر است برای درک بهتر عمل ژن و برای انتخاب درست استراتژی انتخاب آثار غیرافزایشی و اثر پلی‌ژن در مدل گنجانده شود.

### References

1. Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. and Lawlor, T.J. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*. 93: 743-752.
2. Asoro, F.G., Newell, M.A., Beavis, W.D., Scott, M.P. and Jannink, J.L. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Journal of Plant Genome*. 4: 132-144.
3. Atefi, A., Shadparvar, A.A. and Ghavi Hossein-Zadeh, N. 2018. Accuracy of Genomic Prediction under Different Genetic Architectures and Estimation Methods. *Iranian Journal of Applied Animal Science*. 8(1): 43-52.
4. Azevedo, C.F., de Resende, M.D.V., Silva, F.F. and Viana, J.M.S. 2015. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genetics*. 16: 150-118.
5. Cockerham, C.C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics Research*. 39(6): 859-882.
6. De los Campos, G., Gianola, D. and Rosa, G.J.M. 2009a. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *Journal of Animal Science*. 87: 1883-1887.
7. De los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D. and Callus, M.P.L. 2012. Whole genome regression and prediction

- methods applied to plant and animal breeding. *Genetics*. 193: 327-45.
8. De los Campos, G., Naya, H., Gianola, D., Crossa, J. and Legarra, A. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics Research*. 182: 375-385.
  9. Dehnavi, E., Mahyari, S.A., Schenkel, F.S. and Sargolzaei, M. 2018. The effect of using cow genomic information on accuracy and bias of genomic breeding values in a simulated Holstein dairy cattle population. *Journal of Dairy Science*. 101(6): 5166-5176.
  10. Fang, M., Jiang, D., Li, D., Yang, R., Fu, W. and Pu, L. 2012. Improved LASSO priors for shrinkage quantitative trait loci mapping. *Theoretical and Applied Genetics Research*. 124: 1315-24.
  11. Filho, J.E.A., Guimarães, J.F.R., Silva, F.F., Resende, M.D.V., Muñoz, P., Kirst, M. and Júnior, M. F.R.R. 2019. Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. *Journal of Genomic Prediction*. 9: 2739-2748.
  12. Goddard, M.E. and Hayes, B.J. 2007. Genomic selection. *Journal of Animal Breeding and Genetic*. 124: 323-330.
  13. Habier, D., Fernando, R.L. and Dekkers, J.C.M. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics Research*. 177: 2389-2397.
  14. Habier, D., Fernando, R.L., Kizilkaya, K. and Garrick, D.J. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 12: 186.
  15. Hastie, T. and Qian, J. 2016. Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. <https://CRAN.R-project.org/package=glmnet>.
  16. 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.
  17. Heffner, E.L., Lorenz, A.J., Jannink, J. and Sorrells, M.E. 2010. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science*. 50: 1681-1690.
  18. Hu, G., Gu, W., Cheng, L. and Wang, B. 2019. Genomic prediction and variance component estimation for carcass fat content in rainbow trout using SNP markers. *World Aquaculture Society*. 51: 501-511.
  19. Huang, A. and Liu, D. 2016. EBglmnet: Empirical Bayesian Lasso and Elastic Net Methods for Generalized Linear Models. <https://CRAN.R-project.org/package=EBglmnet>.
  20. Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A. and Heckerman, D. 2008. Efficient control of population structure in model organism association mapping. *Genetics*. 178: 1709-1723.
  21. Kang, M., Zhang, C., Chun, H.W., Ding, C. and Liu, C. 2015. eQTL epistasis: detecting epistatic effects and inferring hierarchical relationships of genes in biological pathways. *Bioinformatics*. 31: 656-664.
  22. Legarra, A., Ricard, A. and Filangi, O. 2011. GS3: genomic selection, Gibbs Sampling, Gauss Seidel (and BayesC $\pi$ ). [http://genoweb.toulouse.inra.fr/~alegarra/g3\\_folde/](http://genoweb.toulouse.inra.fr/~alegarra/g3_folde/).
  23. Legarra, A., Robert-Granie, C., Croiseau, P., Guillaume, F. and Fritz, S. 2011. Improved LASSO for genomic selection. *Genetics Research*. 93: 77-87.
  24. Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., Fritz, S. and Ducrocq, V. 2010. Aptitude of Bayesian lasso for genomic selection. In: 9. World Congress on Genetics Applied to Livestock Production (p. 1-4). Presented at 9. World Congress on Genetics Applied to Livestock Production, Leipzig, DEU (2010-08-01-2010-08-06).
  25. Lorenz, A.J., Smith, K.P. and Jannink, J.L. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science*. 52: 1609-1621.
  26. 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 Research*. 183: 1119-1126.
  27. MacLeod, I.M., Hayes, B.J., Savin, K., Chamberlain, A.J., McPartlan, H. and Goddard, M.E. 2010. Power of dense bovine single nucleotide polymorphisms (SNPs) for genome scans to detect and position quantitative trait loci (QTL). *Journal of Animal Breeding and Genetics*. 127: 133-142.
  28. Mahdavi, M., Dashab, G.R., Vafaya Valleh, M., Rokouei, M. and Sargolzaei, M. 2018. Genomic evaluation and variance component estimation of additive and dominance effects using single nucleotide polymorphism markers in heterogeneous stock mice. *Czech Journal of Animal Science*. 63 (12): 492-506.
  29. Melchinger, A.E., Utz, H.F., Piepho, H.P., Zeng, Z.B. and Schön, C.C. 2007. The Role of epistasis in the manifestation of heterosis: A systems-oriented approach. *Genetics Research*. 177(3): 1815-1825.
  30. Meuwissen, T.H. 2003. Genomic selection: The future of marker assisted selection and animal breeding. In *Marker – assisted*

- selection: a fast track to increase genetic gain in plant and animal breeding? Session II: MAS in animals. 54-59.
31. Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics Research*. 157: 1819-1829.
  32. Momen, M., Mehrgardi, A.A., Sheikhi, A., Kranis, A., Tusell, L., Morota, G., Rosa, G.J.M. and Gianola, D. 2018. Predictive ability of genome-assisted statistical models under various forms of gene action. *Scientific Reports*. 8: 12309.
  33. Muñoz, P.R., Resende, J.M.F.R., Gezan, S.A., Resende, M.D.V., de los Campos, G. and Kirst, M. 2014. Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics Research*. 198: 1759-68.
  34. Park, T. and Casella, G. 2008. The Bayesian Lasso. *Journal of the American Statistical Association*. 103: 681-686.
  35. Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sanchez-Villeda, H., Sorells, M. and Jannink, J.L. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome Research*. 5: 103-113.
  36. R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
  37. Resende, M.F.R., Muñoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, M.J., Jokela, E.J., Martin, T.A., Peter, G.F. and Kirst, M. 2012. Accuracy of genomic selection methods in a standard data set of Loblolly Pine (*Pinus taeda* L.). *Genetics Research*. 190: 1503-1510.
  38. Sanna, S., Jackson, A.U., Nagaraja, R., Willer, C.J., Chen, W.M., Bonnycastle, L.L., Shen, H., Timpson, N., Lettre, G., Usala, G., Chines, P.S., Stringham, H.M., Scott, L.J., Dei, M., Lai, S., Albai, G., Crisponi, L., Naitza, S., Doheny, K.F., Pugh, E.W., Ben-Shlomo, Y., Ebrahim, S.H., Lawlor, D.A., Bergman, R.N., Watanabe, R.M., Uda, M., Tuomilehto, J., Collins, F.S., Smith, G.S., Boerwinkle, E., Cao, A., Boehnke, M., Abecasis, G.R. and Mohlke K.L. 2008. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature Genetics Research*. 40: 198-203.
  39. Sargolzaei, M. and Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25: 680-681. <http://www.aps.uoguelph.ca/~msargol/qmsim/>
  40. Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123: 218-223.
  41. Song, H., Li, L., Zhang, Q., Zhang, S. and Ding, X. 2018. Accuracy and bias of genomic prediction with different de-regression methods. *Animal*. 12(6): 1111-1117.
  42. Stich, B., Mohring, J., Piepho, H.P., Heckenberger, M., Buckler, E.S. and Melchinger, A.E. 2008. Comparison of mixed-model approaches for association mapping. *Genetics Research*. 178: 1745-1754.
  43. Su, G., Christensen, O.F., Ostersen, T., Henryon, M. and Lund, M.S. 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PloS One*. 7: 45293Pp.
  44. Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*. 58: 267-288.
  45. Van Raden, P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 91: 4414-4423.
  46. 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.
  47. Villumsen, T.M., Janss, L. and Lund, M.S. 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 126: 3-13.
  48. Waldmann, P. 2001. Additive and non-additive genetic architecture of two different-sized populations of *Scabiosa canescens*. *Heredity*. 86: 648-657.
  49. Xu, S. 2007. An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometric*. 63: 513-521.
  50. Xu, S. 2013. Mapping quantitative trait loci by controlling polygenic background effects. *Genetics Research*. 195(4): 1209-1222.
  51. Zhang, Y.M., Jia, Z., Jim, M. and Dunwell, J.M. 2019. Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Frontiers in Plant Science*. 10: 1664-462.
  52. Zhou, G., Chen, Y., Yao, W., Zhang, C. and Xie, W. 2012. Genetic composition of yield heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the USA*. 109: 15847-15852.





Gorgan University of Agricultural  
Sciences and Natural Resources

*J. of Ruminant Research*, Vol. 8(2), 2020

<http://ejrr.gau.ac.ir>

## Effective methods for estimating genomic breeding values and QTN mapping in dairy cows breeding strategies

H. Abdollahy<sup>1</sup>, \*G.R. Dashab<sup>2</sup>, M. Rokouei<sup>2</sup> and M. Sargolzaei<sup>3</sup>

<sup>1</sup>Ph.D Student and <sup>2</sup>Associate Prof., Dept. of Animal Sciences, Faculty of Agriculture, University of Zabol, Zabol, Iran. <sup>3</sup>Associate Prof., Dept. of Pathobiology, University of Guelph, Guelph, Canada

Received: 11/25/2019; Accepted: 07/08/2020

### Abstract

**Background and objective:** The availability of thousands of polymorphic single nucleotides (SNPs) spread across the genomes made it possible to use genome-wide marker information to predict total breeding value in the implementation of genomic selection. The interaction between several gene loci contributes to the phenotypic changes associated with the expression of complex polygenic traits. The present study aimed to compare the performance of different models with additive and non-additive genetic effects, especially epistasis effects in predicting variance components and genomic breeding values as well as QTN in a dairy cow population based on Bayes Lasso model.

**Materials and methods:** A population of dairy cattle was simulated based on a choice of four pathways to maximize the rate of genetic progress over ten generations with an effective size of 100 individuals in the base population. The genomic structure was assumed consisting of three chromosomes with a length of 100 cM and On each chromosome were located 1000 markers two allelic markers with a frequency of 0.5. 50 QTL double alleles with equal frequency were randomly assigned to each chromosome. For QTL allelic effects, a gamma distribution with a parameter slope of 0.4 was used in QMSim software and sampling was performed. Phenotypic and genotypic data of the last 10 generations were used for analysis. Marker effects, variance components and genetic parameters were estimated using the Bayes Lasso method in the form of six statistical models, that model 1 includes only the marker additive effects, model 2 includes the polygenic effects of animal, model 3 includes the combined effects of additive marker effects and polygenic effects of animal, model 4 includes the dominance effect of markers and polygenic effects of animal, model 5 includes the dominance and additive effects of markers in addition to the polygenic effects of animal, and finally, model 6 includes the additive, dominance and epistatic effects of markers in addition to the polygenic effects of animal. To control the type I error in estimating marker effects, Bonferroni test was used at 1% probability level.

**Results:** The results of this study showed that the additive model or the polygenic model alone could not show the missing or hidden variance, in such a way that adding non-additive genetic effects including dominance and epistatic effects reduced the residual variance effects. Also in the complete model including additive and non-additive effects, the total genetic variance increased. With the addition of non-additive effects in the statistical model the Heritability in the broad sense was increased and the Heritability in the narrow sense of trait decreased. The accuracy of breeding value estimation in model 1 was the lowest and in models 3 to 6 was the same. Model 5 with the lowest false positive QTN and the highest true positive proportion was

---

\*Corresponding author: dashab@uoz.ac.ir

the appropriate model for genetic evaluation of the studied traits. By including polygenic and non-additive effects in the model, the number of false positive errors was reduced.

**Conclusion:** The results of this study showed that the components of additive and dominance variance markers had a small share in the genetic diversity of the studied trait, but the polygenic effects of animal and epistatic had the largest share in the exploring of genetic diversity of trait. The results also showed that the major contribution of gene loci in the exploring of traits is related to additive genes with an additive effect that have a normal distribution. Ignoring non-incremental effects causes inflammation in the additive variance of the trait.

**Keywords:** Bayes Lasso, Dairy cattle, Genomic selection, Non-additive, Variance components effects