



برآورد اجزای واریانس وزن بدن گوسفند مریوس در تولد و شیرگیری با استفاده از نشانگرهای تک نوکلئوتیدی و دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی

آذر راشدی ده‌صحرایی^۱، *جمال فیاضی^۲، رستم عبداللهی آرپناهی^۳،

جولیوس ون در ورف^۴ و هدایت‌اله روشنفکر^۵

دانشجوی دکتری، آدانشیار و استاد ژنتیک و اصلاح نژاد دام، گروه علوم دامی، دانشگاه کشاورزی و منابع طبیعی رامین خوزستان، آستادیار ژنتیک و اصلاح نژاد دام، پردیس ابوریحان دانشگاه تهران، آستاد دانشکده کشاورزی و علوم روستائی دانشگاه نیوانگلند، آرمیدال، استرالیا

تاریخ دریافت: ۹۶/۰۱/۲۹؛ تاریخ پذیرش: ۹۶/۰۴/۲۷

چکیده

مقدمه: برآورد دقیق اجزای واریانس ژنتیکی و غیرژنتیکی با اطلاعات شجره‌ای و ژنومی، از ملزومات پیش‌بینی صحیح ارزش‌های اصلاحی می‌باشد. دسترسی به آرایه‌های چندشکلی تک‌نوکلئوتیدی با تراکم بالا و افزایش تعداد حیوانات با اطلاعات ژنوتیپی، دقت و صحت برآوردهای مبتنی بر جمعیت را افزایش می‌دهد. انتخاب ژنومی به‌طور بالقوه‌ای قادر است بیشتر واریانس ژنتیکی را توسط نشانگرها توجیه نماید. هدف از مطالعه حاضر، برآورد مؤلفه واریانس ژنتیکی افزایشی برای صفات وزن تولد و وزن شیرگیری در گوسفند مریوس با دو روش حداکثر درست‌نمایی محدود شده و بیزی بود.

مواد و روش‌ها: برای انجام این پژوهش از اطلاعات گوسفندان مریوس استرالیایی که با تراشه نشانگری SNP50k شرکت ایلومینا، تعیین ژنوتیپ شده بودند، استفاده شد. پس از کنترل کیفیت داده‌های فنوتیپی و نشانگری، ۲۱۸۹ فرد و ۴۷۳۴۲ نشانگر برای انجام تجزیه و تحلیل استفاده شدند. صفات مورد بررسی در این تحقیق، وزن تولد (۱۳۳۱ رکورد) و وزن شیرگیری (۲۱۳۶ رکورد) بودند. برای مطالعه رابطه بین فراوانی آللی و مقدار واریانس ژنتیکی افزایشی توجیه شده، SNPها در پنج گروه مختلف از فراوانی آللی کمیاب، با تعداد تقریباً برابر در هر گروه، طبقه‌بندی شدند (۰/۱۸-، ۰/۲۸-، ۰/۱۸-، ۰/۳۶-، ۰/۲۸-، ۰/۴۳-، ۰/۳۶- و ۰/۴۹۹-). تجزیه و تحلیل با دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی با استفاده از تکنیک نمونه‌گیری گیبس و مدل RKHS انجام گرفت.

یافته‌ها: مقدار وراثت‌پذیری ژنومی برآورد شده با همه SNPها در رویکرد حداکثر درست‌نمایی محدود شده برای وزن تولد و وزن شیرگیری به ترتیب برابر ۰/۵۸±۰/۰۷ و ۰/۴۶±۰/۰۵ بود. این مقدار وراثت‌پذیری در آنالیز بیزی و به روش RKHS برای صفات مذکور به ترتیب برابر ۰/۵۸±۰/۰۷ و ۰/۴۶±۰/۰۵ برآورد شد. برآوردهای به‌دست آمده از ۵ گروه مختلف فراوانی آللی کمیاب، در آنالیزهای جداگانه و توأم، با هم متفاوت بود. برای هر دو صفت، در دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی، مقادیر واریانس ژنتیکی افزایشی در تجزیه‌های جداگانه، برای همه گروه‌ها، بیشتر از مقادیر به‌دست آمده در آنالیز توأم بود. در تجزیه و تحلیل مجزای گروه‌های مختلف فراوانی آللی کمیاب، در هر دو رویکرد، مقدار وراثت‌پذیری ژنومی به‌دست آمده، برای گروه‌های مختلف، مشابه بود ولی در تجزیه توأم، بین دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی مقدار واریانس

*نویسنده مسئول: j_fayazi@ramin.ac.ir

ژنتیکی توجیه شده در زیرگروه‌های مختلف فراوانی آللی کمیاب، تفاوت زیادی وجود داشت. در رویکرد حداکثر درست‌نمایی محدود شده، در تجزیه و تحلیل توأم، مقدار وراثت‌پذیری برای گروه ۲ ($MAF=0/18-0/28$) در وزن تولد و برای گروه ۵ ($MAF=0/43-0/499$) در وزن شیرگیری صفر به دست آمد. در رویکرد بیزی برای هیچ‌کدام از گروه‌ها، مقدار وراثت‌پذیری برابر صفر نبود. در مجموع واریانس‌های ژنتیکی پنج گروه مختلف فراوانی آللی کمیاب، در آنالیز جداگانه نسبت به واریانس محاسبه شده به وسیله همه SNPها به صورت همزمان، بسیار بزرگتر بود. اما مجموع این واریانس‌ها در آنالیز توأم، مشابه مقدار به دست آمده از کل SNPها، برای هر دو صفت و در هر دو رویکرد بود.

نتیجه‌گیری: در رویکرد بیزی یک توزیع پیشین مشترک برای واریانس‌ها در نظر گرفته می‌شود، بنابراین به نظر می‌رسد نتایج حاصل از رویکرد بیزی دقیق‌تر و معتبرتر از رویکرد دیگر باشد. اگر چه تعداد SNPها در گروه‌های مختلف، مشابه بود، اما مقدار واریانس ژنتیکی توجیه شده توسط گروه‌های مختلف فراوانی آللی کمیاب، متفاوت بود.

واژه‌های کلیدی: گوسفند، انتخاب ژنومی، چندشکلی تک نوکلئوتیدی، فراوانی آلل کمیاب، رویکرد بیزی، آنالیز توأم

در قرن اخیر پیشرفت‌های چشمگیر در زمینه اصلاح نژاد، با کاربرد تئوری‌های ژنتیک کمی، مبتنی بر مدل بی نهایت ژن‌گاه با اثرات جزئی^۱، حاصل شده است (۱). حالت ایده‌آل برای انتخاب فنوتیپی زمانی است که صفت وراثت‌پذیری بالایی داشته و فنوتیپ برای همه حیوانات قبل از سن تولیدمثل، قابل مشاهده باشد. این حالت ایده‌آل به ندرت پیش می‌آید و بنابراین مؤثر بودن انتخاب فنوتیپی را محدود می‌کند (۶). کشف نشانگرهای SNP^۲، پیشرفت‌های به‌دست آمده در ژنتیک مولکولی، کشف روش‌های نوین توالی‌یابی کل ژنوم موجودات زنده، پیشرفت روش‌های بیوانفورماتیک، دانش رایانه‌ای و... حجم بسیار زیادی از داده‌های مولکولی فراهم آورده و در علم ژنتیک شاخه‌ای به نام ژنومیک ایجاد کرده است. با توسعه تکنولوژی آرایه پلی مورفیسم تک نوکلئوتیدی (SNP) برای صفات پیچیده، می‌توان از این اطلاعات به طور مستقیم در برنامه‌های اصلاحی دام‌ها، از طریق انتخاب ژنومیک استفاده کرد (۳۱). در انتخاب ژنومی اثر هر یک از نشانگرها به عنوان یک عامل تصادفی وارد مدل می‌شوند و ارزش اصلاحی افراد از مجموع اثرات نشانگرها برآورد می‌گردد. هدف از انتخاب ژنومی استفاده همزمان از داده‌های ژنوتیپی در سطح DNA به همراه داده‌های فنوتیپی است، تا بتوان بدون نیاز به صرف زمان و هزینه‌های زیاد، دام‌ها را ارزیابی کرده و دام‌های بهینه را انتخاب نمود (۱۳ و ۳۰). در این روش بر خلاف روش انتخاب به کمک نشانگرهای با تعداد محدود که درصد اندکی از واریانس ژنتیکی صفت توجیه می‌شود، به‌طور بالقوه‌ای بیشتر واریانس ژنتیکی به‌وسیله نشانگرها توجیه می‌شود (۱۲). در توجیه این مسأله می‌توان چنین استدلال نمود که انتخاب ژنومی از نمونه‌گیری مندلی در حالتی که در طول شکل‌گیری گامت رخ می‌دهد استفاده می‌کند و لذا بطور مؤثری ماتریس روابط خویشاوندی را نسبت به ماتریس میانگین خویشاوندی مورد استفاده قرار می‌دهد (۳۳). از قویترین روش‌های برآورد اجزای واریانس، در اصلاح نژاد دام می‌توان روش حداکثر درست‌نمایی محدود شده^۳ و روش بیزی مبتنی بر تکنیک نمونه‌گیری گیبس را نام برد. در روش حداکثر درست‌نمایی محدود شده، پیش فرض اساسی این است که نمونه‌ها از یک جامعه با توزیع نرمال هستند. این روش برای اثر انتخاب در جامعه تصحیح انجام می‌دهد و مؤلفه‌های واریانس را از طریق تکرار و همگرایی به دست می‌آورد (۳ و ۵). در روش بیزی برآورد یک پارامتر از طریق توزیع پسین می‌باشد و امکان به دست آوردن مستقیم میانگین توزیع پسین در بیشتر موارد وجود ندارد. بنابراین نمونه‌هایی با خصوصیات مستقل و یکسان از توزیع پسین تولید می‌شود (۱۶ و ۵).

در سال‌های اخیر مطالعات زیادی با استفاده از تراشه‌های SNP، برای شناسایی تنوع ژنتیکی، انجام گرفته است. مطالعاتی روی میزان تنوع ژنتیکی توجیه شده با SNPها در حوزه انسانی (یانگ و همکاران، ۲۰۱۰؛ لی و همکاران، ۲۰۱۲؛ یانگ و همکاران، ۲۰۱۱؛ لتر، ۲۰۱۱؛ واتسون و همکاران، ۲۰۱۲؛ لی و همکاران، ۲۰۱۳؛ وایت و همکاران، ۲۰۱۴؛ مانولیو و همکاران، ۲۰۰۹، برای طیور گوشتی (عبداللهی آرپناهی و همکاران، ۲۰۱۴) برای صفات تولیدی در گاو (اینارد و همکاران، ۲۰۱۵، اوگوا و همکاران، ۲۰۱۶ و یموتو و همکاران، ۲۰۱۵) صورت گرفته است (۳۸، ۱۹، ۳۹، ۲۱، ۳۴، ۲۰، ۳۵، ۲۲، ۲، ۷ و ۲۴). تقسیم بندی طیف آلی برای محاسبه تنوع ژنتیکی، به وسیله یانگ و همکاران (۲۰۱۰) برای قد انسان، پیمتل ادا و همکاران (۲۰۱۱) برای صفات تولید و ترکیب شیر در گاوهای شیری، یانسن و

1. Infinitesimal Model (IFM)
2. Single Nuclotid Polymorphism
3. Restricted Maximum Likelihood (REML)

مدسن (۲۰۱۲) برای صفات مرتبط با تولید و تناسب اندام در گاوهای شیری، لی و همکاران (۲۰۱۳) برای بیماری آلزایمر، عبداللهی آرپناهی و همکاران (۲۰۱۴) برای صفات تولیدی جوجه گوشتی، اوگاو و همکاران (۲۰۱۶) برای وزن بدن گاو سیاه ژاپنی و برخی محققین دیگر انجام گرفته است (۳۸، ۲۷، ۱۷، ۲۰، ۲ و ۲۴). با توجه به این‌که تاکنون در گوسفند مطالعاتی مشابه موارد ذکر شده انجام نشده، هدف از انجام این مطالعه، مشخص نمودن مقدار تنوع ژنتیکی افزایشی حاصل از کل نشانگرهای SNP برای صفات وزن تولد و وزن شیرگیری در گوسفند مرینوس استرالیا، همچنین طبقه‌بندی نشانگرها بر اساس فراوانی آلی کمیاب^۱ و مشخص نمودن سهم هر گروه در توجیه تنوع ژنتیکی افزایشی با استفاده از دو رویکرد حداکثر درست‌نمایی محدود شده و بی‌زی بود. مقایسه این دو رویکرد آماری در برآورد اجزای واریانس ژنومی از دیگر اهداف این پژوهش بود.

مواد و روش‌ها

داده‌های فنوتیپی: داده‌های استفاده شده در این پژوهش، شامل رکوردهای فنوتیپی ۲۱۸۹ رأس گوسفند مرینوس بوده که شامل ۱۲۲۹ نر و ۹۶۰ ماده از ۱۱ پدر (خانواده ناتنی) بودند. همه حیوانات در ایستگاه مزرعه فالکینر استرالیا^۲ متولد شدند. این ایستگاه در دهه ۱۹۴۰ در ریورینا توسط SCIRO برای تحقیقات کشاورزی تأسیس شد و پس از آن توسط مؤسسات دیگر توسعه یافت (۳۶). آمار توصیفی و کنترل کیفیت داده‌های ژنوتیپی: تعداد SNP ۴۸۵۹۹ اتوزومی از تراشه ۵۰ k گوسفندی، شرکت ایلومینا، پس از حذف کروموزوم‌های جنسی X و Y در این پژوهش مورد استفاده قرار گرفتند. برخی از نشانگرها به دلیل کم بودن فراوانی آلی کمیاب ($MAF < 0.01$) از محاسبات خارج شدند. معیارهای دیگر کنترل کیفیت نرخ ژنوتیپ گمشده (> 0.05) و انحراف از تعادل هاردی-واینبرگ ($p < 10^{-6}$)، بر اساس آزمون کای مربع، بود. پس از کنترل کیفیت، ۴۷۳۴۲ نشانگر برای تجزیه و تحلیل‌های نهایی باقی ماندند. توزیع SNPها در طول ۲۶ جفت کروموزوم همولوگ به شرح زیر بود: میانگین فاصله بین SNPهای مجاور ۵۵/۱۸۸ کیلو جفت باز، بین نزدیک‌ترین SNPها ۴/۵۴۸ کیلو جفت باز و بین دورترین SNPها ۱۱۸۰/۸۸۶ کیلو جفت باز بود. اگر چه بیشتر SNPها روی کروموزوم‌های طویل‌تر قرار داشتند، اما متوسط فاصله بین دو SNP در هر کروموزوم تقریباً مشابه بود. خلاصه‌ای از تجزیه و تحلیل نشانگرهای SNP مورد استفاده، در جدول ۱ نشان داده شده است. ویرایش و کنترل کیفیت داده‌های نشانگری با استفاده از نرم‌افزار PLINK در سیستم عامل لینوکس انجام گرفت (۲۸).

1. Minor Allele Frequency (MAF)
2. Falkiner Memorial Field Station (FMFS)

Table 1. Characteristics of used the marker panel

48599	Total Number of SNPs	تعداد کل SNPها
0	Unknown position SNPs	SNPهای با موقعیت ناشناخته
384	SNPs with MAF <0.01	SNPهای با فراوانی آلی کمیاب (>0.01)
873	SNPs not in Hardy-Weinberg equilibrium ($p < 10^{-6}$)	SNPهایی در انحراف از تعادل هاردی-واینبرگ
0	SNPs with missing rates >0.05	نرخ نشانگر گمشده (>0.05)
47342	Used SNPs	SNPهای مورد استفاده
0	Individuals with genotype missing rate >0.05	افراد با نرخ ژنوتایپ گمشده (>0.05)
55.188	Mean distance (kb)	میانگین فاصله SNPها (کیلو جفت باز)
1180.886	Maximum distance (kb)	بیشترین فاصله (کیلو جفت باز)
4.548	Minimum distance (kb)	کمترین فاصله (کیلو جفت باز)

واریانس ژنومی توجیه شده به وسیله نشانگرها: بعد از ویرایش و کنترل کیفیت داده‌ها، مقدار تنوع توجیه شده به وسیله همه SNPها به طور همزمان، با استفاده از یک مدل مختلط خطی برآورد شد. برای مطالعه رابطه بین فراوانی آلی و مقدار واریانس ژنتیکی افزایشی، SNPها در پنج گروه مختلف (در فواصل چندک های ۰/۲۰، ۰/۴۰، ۰/۶۰ و ۰/۸۰) از فراوانی آلی کمیاب، با تعداد تقریباً برابر در هر گروه، طبقه بندی شدند (۰-۰/۱۸، ۰/۱۸-۰/۲۸، ۰/۲۸-۰/۳۶، ۰/۳۶-۰/۴۳، ۰/۴۳-۰/۴۹ و ۰/۴۹-۰/۴۳). تعداد نشانگرها در گروه‌های فراوانی آلی کمیاب، در مطالعه حاضر به ترتیب برابر ۱۱۳۰۸، ۹۴۳۷، ۹۳۵۰، ۹۲۷۰ و ۹۲۳۴ بود. به منظور برآورد میزان واریانس ژنتیکی افزایشی توجیه شده به وسیله هر یک از پنج گروه فراوانی آلی کمیاب، تجزیه و تحلیل‌ها به صورت آنالیزهای مجزا^۱ و آنالیزهای توأم^۲ انجام گرفت. در آنالیز مجزا، پنج مدل جداگانه هر کدام با یک اثر تصادفی ژنتیکی افزایشی با ساختار واریانس-کواریانس ماتریس خویشاوندی ژنومی منحصر به هر یک از پنج گروه مختلف فراوانی آلی کمیاب، برازش داده شد (به عبارت دیگر پنج مدل مختلف برازش شد و مقدار واریانس ژنتیکی افزایشی توجیه شده برای هر یک از گروه‌های MAF با استفاده از هر مدل به دست آمد). در آنالیز توأم، یک مدل آماری در برگیرنده پنج اثر تصادفی ژنتیکی افزایشی ژنومی مربوط به پنج گروه مختلف فراوانی آلی کمیاب (پنج ماتریس خویشاوندی ژنومی) برازش شد. در نهایت از این مدل، پنج جزء واریانس ژنتیکی افزایشی متمایز که هر کدام به یکی از گروه‌های فراوانی آلی کمیاب اختصاص داشت، به طور همزمان حاصل شد.

آنالیزهای آماری: در این مطالعه برای برآورد اجزای واریانس ژنتیکی از دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی استفاده شد. در رویکرد حداکثر درست‌نمایی محدود شده از روش تجزیه و تحلیل صفات پیچیده ژنوم گسترده^۳، استفاده گردید که به منظور برآورد نسبتی از واریانس فنوتیپی توجیه شده توسط SNPها برای صفات پیچیده، طراحی شده است (۴۰). در رویکرد بیزی برای برآورد اجزای واریانس ژنومی از تکنیک نمونه‌گیری گیبس و مدل

1. Separate Analysis
2. Joint Analysis
3. Genome-wide Complex Trait Analysis (GCTA)

آماري $RKHS^1$ (۸) از طریق بسته $BGLR^2$ در محیط R ، استفاده شد (۲۶). در پژوهش حاضر، طول زنجیره نمونه‌گیری گیس ۱۰۰۰۰۰ تکرار، تعداد تکرارهای قلق‌گیری ۱۰۰۰۰ و فاصله نمونه‌گیری ۵ در نظر گرفته شد. در رویکرد حداکثر درست‌نمایی محدود شده، برای تجزیه و تحلیل مجزای داده‌ها از مدل زیر استفاده شد: لازم به ذکر است که داده‌ها برای اثرات ثابت قبل از انجام آنالیز تصحیح شدند.

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

در این مدل: y : بردار مشاهدات، μ : میانگین، 1 : بردار یکه، $g \sim N(0, G\sigma^2g)$: بردار تصادفی ارزش ژنتیکی افزایشی حیوان، G : ماتریس روابط خویشاوندی افزایشی ژنومی، σ^2g : واریانس ژنومی توجیه شده به وسیله m نشانگر، $e \sim N(0, I\sigma^2e)$: بردار باقیمانده‌های مدل و σ^2e واریانس باقیمانده است، I : ماتریس واحد با ابعادی $n \times n$ و n برابر با تعداد حیوان دارای رکورد است. در این معادله عناصر ماتریس G با استفاده از رابطه زیر محاسبه شدند (۳۸):

$$G_{jj} = \frac{1}{m} \sum_{i=1}^m \frac{(x_{ij} - 2p_i)(x_{ij} - 2p_i)}{2p_i(1-p_i)} \quad (2)$$

در این معادله x_{ij} : تعداد کپی آلل کمیاب برای آلمین نشانگر SNP در ژامین فرد، p_i : فراوانی آلل کمیاب و m تعداد نشانگرهای SNP می‌باشد. در مطالعات به این نوع مدل، مدل بهترین پیش‌بینی ناریب خطی ژنومی^۳ می‌گویند (۳۲). در تجزیه و تحلیل توأم، واریانس ژنتیکی توجیه شده به وسیله SNPها، با استفاده از مدل خطی زیر محاسبه شد:

$$y = 1\mu + \sum_{t=1}^h g_t + e \quad (3)$$

در این معادله $g_t \sim (0, G_t\sigma_{g_t}^2)$ ، $t = 1, 2, \dots, h$. ساختار واریانس-کواریانس اثرات تصادفی به صورت زیر بود.

$$V = \sum_{t=1}^h G_t\sigma_{g_t}^2 + I\sigma_e^2 \quad (4)$$

در این معادله h : تعداد گروه‌های فراوانی آللی کمیاب ($h=5$)، فرض بر این است که هیچ کواریانسی بین گروه‌های مختلف وجود ندارد.

در رویکرد بیزی مدل‌های آماری مشابه با رویکرد حداکثر درست‌نمایی محدود شده بودند با این تفاوت که اثرات تصادفی ژنتیکی افزایشی از ضرب ماتریس خویشاوندی کرنل خطی (K) در α حاصل شدند. در اینجا K یک ماتریس کرنل $n \times n$ بوده که به وسیله متغیرهای پیوسته SNP نمایه شده است و $K \sim XX^T$ بوده که در آن X : ماتریس ژنوتیپ SNPها است. K از ماتریس X مرکزی شده و استاندارد بدست آمد، سپس طبق پیشنهاد ون رادن (۲۰۰۸) و یانگ و همکاران (۲۰۱۰) بر تعداد SNP تقسیم شده است (۳۲ و ۳۸). α بردار ضرایب رگرسیون $RKHS$ است که تابع درست‌نمایی را به صورت زیر حداکثر می‌کند:

$$l(\alpha|\lambda) = (y - K\alpha)'(y - K\alpha) + \lambda\alpha'K\alpha \quad (5)$$

در رابطه بالا $\alpha \sim N(0, K^{-1}\sigma_g^2)$ ، σ_g^2 : واریانس ژنومی افزایشی توجیه شده به وسیله نشانگرها و λ : یک فاکتور تنظیم کننده است. مقادیر ارزش‌های ژنتیکی افزایشی از رابطه $g = K\alpha$ بدست می‌آید. در روش $RKHS$ ابعاد معادلات از $m \times m$ یعنی صدها هزار SNP به تعداد $n \times n$ یعنی برابر با تعداد مشاهدات کاهش می‌یابد. به عنوان مثال در این مطالعه ابعاد ماتریس‌ها برابر ۱۳۳۱ معادله برای وزن تولد و ۲۱۳۶ معادله برای وزن شیرگیری بود. بنابراین می‌توان معادله را به شکل ماتریسی زیر نوشت:

1. Reproducing Kernel Hilbert Spaces
2. Bayesian Generalized Linear Regression
3. Genomic Best Linear Unbiased Prediction (GBLUP)

$$y = 1\mu + G\alpha + e \quad (۶)$$

تمامی اجزای این معادله قبلا در معادله (۱) تعریف شده است.

تجزیه و تحلیل توأم با پنج کرنل افزایشی ($G_1+G_2+\dots+G_5$). در آنالیز توأم پنج گروه مختلف MAF از مدل آماری زیر استفاده شد:

$$y = 1\mu + \sum_{t=1}^n K_t \alpha_t + e \quad (۷)$$

تمامی اجزای معادله (۷) قبلا در معادلات (۳) و (۴) تعریف شده است به استثنای $\alpha_t \sim (0, K_t^{-1} \sigma_{gt}^2)$ و $t=1, \dots, 5$ یک بردار رگرسیون در RKHS برای گروه‌های MAF t ام، K_t : ماتریس خویشاوندی افزایشی ژنومی برای t امین گروه MAF، σ_{gt}^2 : واریانس توجیه شده به وسیله SNPها در t امین گروه MAF است. بنابراین تابع چگالی توأم پنج بردار تصادفی ژنتیکی و باقیمانده به صورت زیر می‌باشد:

$$p(e, g_1, \dots, g_t | \sigma_e^2, \sigma_{g_1}^2, \dots, \sigma_{g_t}^2) = N(e | I\sigma_e^2) \prod_{t=1}^n N(g_t | 0, K_t \sigma_{g_t}^2) \quad (۸)$$

و ماتریس واریانس-کواریانس برابر:

$$V = \sum_{t=1}^n K_t \sigma_{g_t}^2 + I\sigma_e^2 \quad (۹)$$

است. در اینجا فرض بر این است که هیچ کواریانسی بین اثرات SNPها در گروه‌های مختلف وجود ندارد. در هنگام پیاده سازی آنالیز بیزی برای اجزای واریانس توزیع پیشین کای اسکور و برای اثرات محیطی توزیع پیشین یکنواخت در نظر گرفته شد.

نتایج و بحث

صفات مورد بررسی در این تحقیق، وزن تولد (۱۳۳۱ رکورد) و وزن شیرگیری (۲۱۳۶ رکورد) بودند. آمار توصیفی صفات مورد مطالعه در جدول ۲ آورده شده است.

جدول ۲: آماره‌های توصیفی صفات مورد بررسی

Table 2. Descriptive statistics of studied traits

بیشینه	کمینه	انحراف معیار	میانگین	تعداد	صفت (واحد)
Max	Min	SD	Mean	n	Trait (unit)
7.5	1.6	0.95	4.19	1331	وزن تولد (کیلوگرم) Birth weight(kg)
34.8	6.2	4.42	17.44	2136	وزن از شیرگیری (کیلوگرم) Weaning weight(kg)

اجزای واریانس ژنومی برای کل نشانگرها، با استفاده از رویکرد حداکثر درست‌نمایی محدود شده و بیزی در جدول ۳ ارائه شده است. مقدار وراثت‌پذیری ژنومی برآورد شده توسط همه SNPها در هر دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی برای وزن تولد و وزن شیرگیری مشابه و به ترتیب ۰/۵۸ و ۰/۴۶ برآورد شد. در مطالعه‌ای که توسط هویزمن و همکاران (۲۰۰۹) در گوسفندان مرینوس و با استفاده از اطلاعات شجره‌ای انجام شد، مقدار وراثت‌پذیری صفات وزن تولد و وزن شیرگیری به ترتیب ۰/۳۹ و ۰/۷۲ برآورد گردید (۱۵). تفاوت مشاهده شده در اجزای واریانس به دست آمده در روش ژنومی و شجره‌ای می‌تواند به دلیل تفاوت ساختار داده‌ها و مدل‌های

آماري استفاده شده مبتني بر شجره باشد. لي و همكاران (۲۰۱۳) در مطالعه‌اي بيان داشتند كه مقايسه واريانس افزايشي توجيه شده به وسيله همه SNPها و كل تنوع ژنتيكي برآورد شده توسط مطالعات مبتني بر شجره، ساده نيست (۲۰). با اين حال، برآوردهاي حاصل از مطالعات شجره‌اي، به دليل وجود تنوع ژنتيكي غيرافزايشي و تاثير عوامل محيطي، مي‌تواند اريب باشد. يانگ و همكاران (۲۰۱۰) مقدار واريانس ژنتيكي توجيه شده توسط نشانگرها را براي قد انسان ۴۵ درصد گزارش نمودند (۳۸). لي و همكاران (۲۰۱۲) براي بيماران مستعد به شيزوفرنی اين مقدار واريانس توجيه شده را ۲۳ درصد برآورد كردند (۱۹). واتسون و همكاران (۲۰۱۲)، گزارش نمودند كه ۳۰ درصد واريانس ژنتيكي در استعداد ابتلا به اسكلروز متعدد، توسط نشانگرهاي SNP توجيه مي‌شود (۳۴). اين مطالعات نشان دادند كه بخش زيادي از وراثت پذيري براي صفات كمی، توسط SNPهاي رايج موجود در تراشه‌هاي تجاري توجيه مي‌شود. عبداللهي آرپناهي و همكاران (۲۰۱۴)، مقدار واريانس فنوتیپی توجيه شده توسط SNPها را براي صفات وزن بدن، عضله سينه و توليد تخم مرغ روزآزمون به ترتيب ۰/۳۰، ۰/۱۹ و ۰/۳۳ برآورد نمودند (۲). محققين در مدلي با استفاده از پنل SNP، ۵۰k، نشان دادند كه بين ۹۰ تا ۹۲ درصد از تنوع ژنتيكي افزايشي كل براي صفات توليدي و تناسب اندام (۲۴)، ۳۲٪ براي باروري و تقريباً ۸۰٪ براي صفت توليد شير (۱۱)، در گاو شيري با SNPها توجيه مي‌شود. يموتو و همكاران (۲۰۱۵)، در پژوهشي روي گاو سپاه ژاپني، مقدار وراثت‌پذيري QTL را در جمعيت‌هاي با اندازه متفاوت، تقريباً يكسان (۰/۴) برآورد نمودند ولي مقدار خطاي استاندارد وراثت‌پذيري در پژوهش اين محققين با کاهش اندازه جمعيت، افزايش يافت (۳۱).

طرفداران آمار بيزي (بيزيني‌ها)^۱ منتقدان جدي ميزان خطاي استاندارد و فاصله اطمينان در روش‌هاي فراواني‌گرا مانند حداكثر درست‌نمايي محدود شده هستند (۴ و ۹) به طوري كه بيزيني‌ها به جاي واژه فاصله اطمينان، مجموعه اعتبار^۲ يا فاصله اعتبار^۳ را به كار مي‌برند. با توجه به اينكه رويكرد بيزي از توزيع‌هاي پيش آگامي دهنده^۴ استفاده مي‌كند، نتايج دقيق‌تري از آن انتظار مي‌رود (۱۰ و ۲۳). در صورتي كه از تعداد نمونه بيشتري در الگوريتم نمونه‌گيري گيبس استفاده شود و تعداد كمی از نمونه‌ها حذف شود، برآوردهاي معتبري از پارمترها به دست خواهد آمد. بنابراین به نظر مي‌رسد نتايج حاصل از رويكرد بيزي دقيق‌تر و معتبرتر از رويكردهاي ديگر باشد (۲۹).

-
1. Bayesians
 2. Credible set
 3. Credibility Intervals
 4. Informative

جدول ۳: اجزای واریانس ژنومی و پارامترهای ژنتیکی با استفاده از آنالیز تک متغیره و دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی با استفاده از تمام نشانگرها.

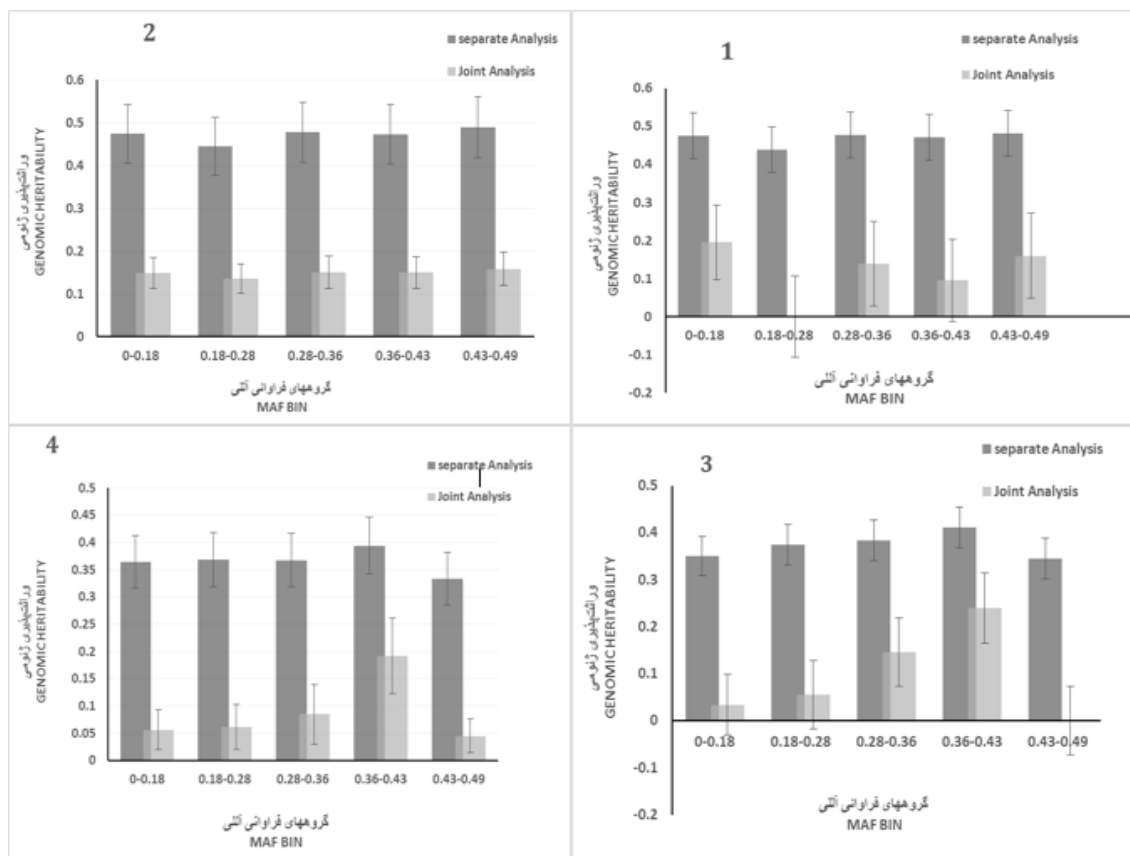
Table 3. Genomic variance components and genetic parameters for univariate analysis using all SNPs via two approaches

وراثت پذیری (خطای معیار) h^2_m (SE)	واریانس باقیمانده (خطای معیار) V_e (SE)	واریانس فنوتیپی (خطای معیار) V_p (SE)	واریانس ژنومی افزایشی (خطای معیار) V_g (SE)	صفت (واحد) Trait (unit)	رویکرد مورد استفاده Approach
0.58 (0.07)	0.33 (0.04)	0.78 (0.04)	0.45 (0.07)	وزن تولد (کیلوگرم) Birth weight(kg)	رویکرد حداکثر درست نمایی محدود شده
0.46 (0.05)	6.25 (0.48)	11.63 (0.44)	5.38 (0.71)	وزن شیرگیری (کیلوگرم) Weaning weight(kg)	REML approach
0.58(0.07)	0.33(0.04)	0.77(0.04)	0.45(0.07)	وزن تولد (کیلوگرم) Birth weight(kg)	رویکرد بیزی
0.46(0.05)	6.22(0.47)	11.49(0.42)	5.27(0.67)	وزن شیرگیری (کیلوگرم) Weaning weight(kg)	Bayesian approach

مقدار وراثت‌پذیری ژنومی مرتبط با گروه‌های مختلف فراوانی آللی کمیاب، برای وزن تولد و وزن شیرگیری، با استفاده از رویکردهای حداکثر درست‌نمایی محدود شده و بیزی در شکل ۱ نشان داده شده است. برآوردهای به دست آمده از پنج گروه مختلف فراوانی آللی کمیاب، در آنالیزهای جداگانه و توأم، با هم متفاوت بود. در آنالیز جداگانه، واریانس‌های مربوط به هر گروه، اریب می‌باشد، زیرا بین SNPها در گروه‌های مختلف، عدم تعادل لینکاژ (LD') وجود دارد. برای هر دو صفت، در هر دو رویکرد حداکثر درست‌نمایی محدود شده و بیزی (مدل RKHS)، مقادیر برآورد شده در آنالیزهای جداگانه، برای همه گروه‌ها، بیشتر از مقادیر به دست آمده در آنالیز توأم بود. در تجزیه و تحلیل جداگانه گروه‌های مختلف فراوانی آللی کمیاب، در هر دو رویکرد، مقدار واریانس ژنومی به دست آمده، برای گروه‌های مختلف، مشابه بود. در آنالیز توأم، بین واریانس ژنتیکی شناسایی شده توسط زیرگروه‌های مختلف، تنوع بالایی وجود داشت. در رویکرد حداکثر درست‌نمایی محدود شده در تجزیه و تحلیل توأم، صفر در فاصله اطمینان برای گروه ۲ در وزن تولد و برای گروه ۵ در وزن شیرگیری به دست آمد. مجموع واریانس‌های ژنتیکی ۵ گروه مختلف فراوانی آللی کمیاب که در آنالیز جداگانه به دست آمد، نسبت به واریانس محاسبه شده توسط همه SNPها به صورت همزمان، بسیار بیشتر بود. اما مجموع این واریانس‌ها در آنالیز توأم، مشابه مقدار به دست آمده از کل SNPها، برای هر دو صفت و در هر دو رویکرد بود. بنابراین، در ادامه تنها نتایج آنالیزهای توأم شرح داده می‌شود. اگر چه تعداد SNPها در گروه‌های مختلف، مشابه بود، اما مقدار واریانس ژنتیکی توجیه شده توسط گروه‌های فراوانی آللی کمیاب، متفاوت بود. وراثت‌پذیری ژنومی وزن تولد در رویکرد حداکثر درست‌نمایی محدود شده در دامنه ۰ تا حدود ۰/۲ در پنج گروه فراوانی آللی کمیاب، متغیر بود. بیشترین مقدار برآورد شده در گروه اول با فراوانی آللی ۰/۱۸-۰ بود. در آنالیز بیزی، بین پنج گروه مختلف، تفاوت چندانی در برآورد وراثت‌پذیری صفت وزن تولد وجود نداشت. برای وزن شیرگیری، الگوی توزیع واریانس ژنتیکی توجیه شده بین پنج گروه فراوانی آللی کمیاب، دارای نوسان بوده و از ۰ در گروه پنجم تا ۰/۲۵ در گروه چهارم متغیر بود (رویکرد حداکثر درست‌نمایی محدود شده). در حالی که

1. Linkage disequilibrium

رویکرد بیزی دامنه‌ای از ۰/۰۵ (گروه پنجم) تا ۰/۲ (گروه چهارم) داشت. بنابراین در هر دو رویکرد کمترین مقدار وراثت‌پذیری مربوط به گروه پنجم و بیشترین آن مربوط به گروه چهارم بود. با توجه به انتخاب در جمعیت‌های دامی، انتظار می‌رود توزیع واریانس ژنتیکی بر حسب فراوانی آللی متغیر باشد. رویکرد بیزی، در گروه‌های مختلف فراوانی آللی کمیاب، یک توزیع پیشین مشترک به واریانس‌ها اختصاص داده می‌شود و انعطاف‌پذیری بیشتری دارد، بنابراین نتایج حاصل از این رویکرد قطعیت بیشتری داشته و قابل اعتمادتر می‌باشند. دیمی غیاث‌آبادی و همکاران (۲۰۱۲) در مطالعه‌ای روی مرغ‌های بومی استان فارس، دو روش آماری بیزی و حداکثر درست‌نمایی محدود شده را مورد مقایسه قرار دادند و گزارش نمودند که روش آنالیز بیزی قدرتمندتر از روش حداکثر درست‌نمایی محدود شده بود. این محققین چنین عنوان کردند که روش بیزی بسیار دقیق و پرهزینه است و پارامترهای ژنتیکی را با دقت بالایی پیش‌بینی می‌کند (۵). طبق نتایج به دست آمده (شکل ۱)، حدود ۸۰ درصد واریانس ژنتیکی وزن تولد و ۸۶ درصد واریانس ژنتیکی وزن شیرگیری توسط SNP‌های رایج با $MAF > 0.18$ ، توجیه می‌شوند و مقدار اندکی از واریانس، مربوط به SNP‌های غیر رایج و واریانت‌های کمیاب بود. عدم تعادل لینکاژ (LD)، بین واریانت‌های علی^۱ با فراوانی پایین و SNP‌های ژنوتیپ شده رایج، پایین است (۳۷)، بنابراین مشاهده شده که SNP‌های رایج با $MAF > 0.2$ ، بخش قابل توجهی از تنوع را در نقاطی که واریانت‌های علی رایج، دارای LD بالایی با این SNP‌ها هستند، توجیه می‌کند. نتایجی مشابه با یافته‌های تحقیق حاضر در جمعیت‌های انسانی نیز گزارش شده است. در یک مطالعه در بیماران شیروزوفرنی، حدود ۷۰٪ کل واریانس مربوط به SNP‌های رایج با $MAF > 0.1$ توجیه شد (۱۹). در مطالعات دیگر روی بیماری آلزایمر، اندومترروز و اسکروز ۹۰ درصد از واریانس ژنتیکی توسط SNP‌های رایج با $MAF > 0.1$ (۲۰) و در استعداد ابتلا به بیماری‌های سرطان، دیابت نوع یک و دیابت نوع دو، بیشترین توجیه واریانس ژنتیکی توسط SNP‌های با $MAF > 0.2$ بوده است (۲۵). در یک تحقیق در گاو سیاه ژاپنی و روی صفت وزن بدن، مقدار نسبتاً بالایی از واریانس ژنتیکی افزایشی، هنگام استفاده از SNP‌های با فراوانی آللی کمیاب که در محدوده ۰/۲-۰/۲۵ و ۰/۳-۰/۲۵ بودند توجیه شد (۲۴). عبداللهی آرپناهی و همکاران (۲۰۱۴) در مطالعه‌ای روی جوجه‌های گوشتی گزارش نمودند که ۷۵ درصد از واریانس ژنومی برای صفات وزن بدن و عضله سینه توسط نشانگرهای با فراوانی کمتر از ۰/۲ توجیه می‌شود که با نتایج به دست آمده در این پژوهش مغایرت داشت (۲). تمرکز روی طیف فراوانی آللی واریانت‌های علی، برای یک صفت خاص، اثراتی را که این واریانت‌ها ممکن است بر صفات دیگر داشته باشند نادیده می‌گیرد. بنابراین اثر خالص یک واریانت علی بر صفات مرتبط با شایستگی، ممکن است از آن‌چه که توسط اثرات آن بر یک صفت خاص پیشنهاد می‌شود، کاملاً متفاوت باشد (۲۰ و ۲). بنابراین نتایج به دست آمده در مطالعه حاضر می‌تواند با پلیوتروپی صفات پیچیده متعدد، که تحت یک سبک معماری پلی‌ژنیک هستند، سازگار باشد. طبق پیشنهاد لی و همکاران (۲۰۱۳)، حجم نمونه بسیار بزرگ، ایده‌آل و پوشش بهتر واریانت‌های با فراوانی پایین مورد نیاز است تا استنتاج قوی‌تر و قابل اعتمادتری به دست آید (۲۰).



شکل ۱: وراثت پذیری ژنومی برآورد شده با SNPها در پنج گروه MAF ۱- برای وزن تولد با رویکرد REML، ۲- برای وزن تولد با رویکرد بیزی، ۳- برای وزن شیرگیری با رویکرد REML، ۴- برای وزن شیرگیری با رویکرد بیزی

Figure 1. Genomic heritability estimates with SNPs for five groups of MAF 1- for birth weight via REML approach, 2- for birth weight via Bayesian approach, 3- for weaning weight via REML approach, 4- for weaning weight via Bayesian approach

نتیجه گیری

طبق نتایج به دست آمده، سهم SNPهای با فراوانی آلی کمیاب مختلف در توجیه واریانس ژنتیکی برای دو صفت وزن تولد و وزن شیرگیری متفاوت است و به طور کلی بخش قابل توجهی از واریانس ژنتیکی، توسط SNPهای با $MAF > 0.20$ توجیه می شود. در این پژوهش، هنگامی که آنالیزهای جداگانه انجام گرفت، اجزای واریانس به دست آمده در هر دو روش حداکثر درست‌نمایی محدود شده و بیزی یکسان بودند ولی در آنالیز توأم و استفاده از مدل پیچیده تر، اجزای واریانس محاسبه شده در دو روش متفاوت بود. برای محاسبه اجزای واریانس ژنومی با استفاده از حداکثر درست‌نمایی محدود شده و نرم افزار GCTA، اگر یک جزء واریانس نزدیک به صفر باشد در نظر گرفته نمی شود و این روش برای به دست آوردن چندین اثر تصادفی به طور همزمان دچار مشکل است. در رویکرد بیزی یک توزیع پیشین مشترک برای واریانس‌ها در نظر گرفته می شود و با توجه به انعطاف پذیری بیشتر روش بیزی، به نظر می رسد نتایج حاصل از رویکرد بیزی دقیق تر و معتبرتر از رویکردهای دیگر باشد و نتایج آن قابل اعتمادتر می باشند.

تشکر و قدردانی

نویسندگان مقاله از همکاری صمیمانه دانشگاه نیوانگلند استرالیا به دلیل در اختیار گذاشتن اطلاعات داده‌ها، از دانشگاه کشاورزی و منابع طبیعی رامین خوزستان بخاطر حمایت‌های مالی و از پردیس ابوریحان دانشگاه تهران به دلیل همکاری صمیمانه، کمال تشکر را دارند.

منابع

1. Abdollahi-Arpanahi, R., Pakdel, A., and Zandi, M.B. 2012. From infinity locus genetic model with minor effects (infinitesimal model) to genomic selection. *Modern Genetics Journal*. 7: 105- 114.
2. Abdollahi-Arpanahi, R., Pakdel, A., Nejati-Javaremi, A., Moradi Shahrabak, M., Morota, G., Valente, B.D., Kranis, A., Rosa, G.J.M., and Gianola, D. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *Journal of Animal Breeding and Genetics*. 131: 183–193.
3. Alijani, S. 2010. Major genes detection in farm animals using statistical Bayesian and molecular methods. PHD Thesis. Tehran University, Karaj, Iran. 142 p.
4. Blasco, A. 2001. The Bayesian controversy in animal breeding. *Journal of Animal Science*. 79: 2023-2046.
5. Deimi Ghias Abadi, P., Alijani, S., Shodja Ghias, J., and Pirani, N. 2012. Comparison of two restricted maximum likelihood (REML) and Bayesian statistical methods for estimating genetic parameter of some economically important traits in Fars native chickens. *Research On Animal Production (Scientific and Research)*. 3: 1-13.
6. Dekkers, J.C.M., and Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*. 3: 22-32.
7. Eynard, S.E., Windig, J.J., Leroy, G., Van Binsbergen, R., and Calus, M. 2015. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genetics*. 16: 24.
8. Gianola, D., and De los Campos, G. 2013. Genome-enabled prediction of complex traits. University of Wisconsin-Medison, May 27th-31st, 2013.
9. Gianola, D., and Fernando, R.L. 1986. Bayesian methods in animal breeding theory. *Journal of Animal Science*. 63: 217-244.
10. Goddard, M. 2010. Introduction to Bayesian Statistics. University of Melbourne and Victorian Institute of Animal Science. Pp: 135-162.
11. Haile-Mariam, M., Nieuwhof, G., Beard, K., Konstatinov, K., and Hayes, B. 2013. Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. *Journal of Animal Breeding and Genetics*. 130: 20–31.
12. Hayes, B.J. 2007. QTL Mapping, MAS and Genomic Selection. A short-course organized by Animal Breeding and Genetics, Department of Animal Science, Iowa State University. 118 pp.
13. Hayes, B., Bowman, P., Chamberlain, A., and Goddard, M. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*. 92: 433–443.
14. Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*. 106: 9362–9367.
15. Huisman, A.E., and Brown, D.J., 2009. Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in Merino sheep. 3. Genetic relationships between ultrasound scan traits and other traits. *Journal of Animal Production Science*. 49: 283-288.

16. Jasori, M., Alijani, S., Pirany, N., Baghernejad, M., and Jafarzadeh, R. 2011. Estimation of genetic parameters of Holstein dairy cattle using Bayesian procedure. 4th Iranian Animal Science Congress. Tehran, Iran. Pp: 3022-3025.
17. Jensen, J., Su, G., and Madsen, P. 2012. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. *BMC Genetics*. 13: 44.
18. Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto-Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. 2012. Other members of the International Sheep Genomics Consortium. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*. 10: e1001258. 1-14.
19. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. 2012. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Journal of National Genetics*. 44: 247-250.
20. Lee, S.H., Harold, D., Nyholt, D.R., Goddard, M.E., Zondervan, K.T., Williams, J., Montgomery G.W., Wray, N.R., and Visscher, P.M. 2013. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Journal of Human Molecular Genetics*. 22: 832-841.
21. Lettre, G. 2011. Recent progress in the study of the genetics of height. *Journal of Human Genetic*. 129: 465-472.
22. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whitemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., and Visscher, P.M. 2009. Finding the missing heritability of complex diseases. *Nature*. 461: 747-753.
23. Mrode, R.A. 1996. Linear models for the prediction of animal breeding values. C.A.B. International. Pp: 187.
24. Ogawa, S., Matsuda, H., Taniguchi, Y., Watanabe, T., Sugimoto, Y., and Iwaisaki, H. 2016. Estimated Genetic Variance Explained by Single Nucleotide Polymorphisms of Different Minor Allele Frequencies for Carcass Traits in Japanese Black Cattle. *Journal of Biosciences and Medicines*. 4: 89-97.
25. Park, J.H., Gail, M.H., Weinberg, C.R., Carroll, R.J., Chung, C.C., Wang, Z., Chanock, S.J., Fraumeni, J.F., and Chatterjee, N. 2011. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proceedings of the National Academy of Sciences of the USA*. 108: 18026-18031.
26. Pérez, P., and De los Campos, G. 2013. BGLR: a statistical package for whole genome regression and prediction. R package version, 1(0.2)
27. Pimentel, E.C.G., Erbe, M., König, S., and Simianer, H. 2011. Genome partitioning of genetic variation for milk production and composition traits in Holstein cattle. *Journal of Frontiers Genetics*. 2: 19.
28. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., and Daly, M.J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Journal of Human Genetics*. 81: 559-575.
29. Rashedi-Dehsahraei, A., Fayazi, J., Vatankhah, M., and Beigi-Nasiri, M.T. 2013. Estimation of (Co) variance components and genetic parameters for growth traits in Lori-Bakhtiari lambs using a Bayesian approach via Gibbs sampling. *Journal of Ruminant Research*. 1: 109-128.
30. Solberg, T.R., Sonesson, A.K., Woolliamas, J.A., and Meuwissen, T.H. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science*. 86: 2447-2454.

31. Uemoto, Y., Sasaki, S., Kojima, T., Sugimoto, Y., and Watanabe, T. 2015. Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. *BMC Genetics*. 16: 134.
32. Van Raden, P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 91: 4414–23.
33. Villanueva, B., Pong-Wong, R., Fernandez, J., and Toro, M.A. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science*. 83: 1747-1752.
34. Watson, C.T., Disanto, G., Breden, F., Giovannoni, G., and Ramagopalan, S.V. 2012. Estimating the proportion of variation in susceptibility to multiple sclerosis captured by common SNPs. *Journal of Scientific Reports*. 2: 770.
35. White, J.D., Allingham, P.G., Gorman, Ch.M., Emery, D.L., Hynd, P., Owens, J., Bell, A., Siddell, J., Harper, G., Hayes, B.J., Daetwyler, H.D., Usmar, J., Goddard, M.E., Henshall, J.M., Dominik, S., Brewer, H., van der Werf, J.H.J., Nicholas, F.W., Warner, R., Hofmyer, C., Longhurst, T., Fisher, T., Swan, P., Forage, R., and Oddy, V.H. 2012. Design and phenotyping procedures for recording wool, skin, parasite resistance, growth, carcass yield and quality traits of the Sheep GENOMICS mapping flock. *Journal of Animal Production Science*. 52: 157–171.
36. Witte, J.S., Visscher, P.M., and Wray, N.R. 2014. The contribution of genetic variants to disease depends on the ruler. *Nature Reviews Genetics*. 15: 765–776.
37. Wray, N.R. 2005. Allele frequencies and the r^2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin. Res. Journal of Human Genetics*. 8: 87–94.
38. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., and Visscher, P.W. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Reviews Genetics*. 42: 565–569.
39. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., Hill, W.G., Landi, M.T., Alonso, A., Letter, G., Lin, P., Ling, H., Lowe, W., Mathias, R.A., Melbye, M., Pugh, E., Cornelis, M.C., Weir, B.S., Goddard, B.S., and Visscher, P.M. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Reviews Genetics*. 43: 519–525.
40. Yang, J., Lee, H., Goddard, M., and Visscher, P. 2014. GCTA: A tool for genome-wide complex trait analysis. Version 1.24, 28 July 2014. University of Queensland.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Ruminant Research, Vol. 5(2), 2017

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

Estimation of variance components for body weight of Merino sheep at birth and weaning using single nucleotide markers and REML and Bayesian approaches

A. Rashedi Dehsahraei¹, *J. Fayazi², R. Abdollaho Arpanahi³, J. Van Der Werf⁴ and H. Roshanfekr⁵

¹Ph.D Student, ²Associat Prof., and ⁵Professor of Genetics and Animal Breeding, Dept. of Animal sciences, Ramin Agriculture and Natural Resources University of Khuzestan, ³Assistant Prof., Dept., of Animal and Poultry Sciences, College of Aburaihan, University of Tehran ⁴Professor, School of Environmental and Rural Sciences, University of New England, Armidale, Australia,

Received: 04/18/2017; Accepted: 07/18/2017

Abstract

Background and objectives: Accurate estimation of variance components using pedigree and genomic data is important for prediction of breeding values. Availability of high density single nucleotide polymorphisms (SNP) arrays and genotyping many individuals resulted in the increase of accuracy of population-based estimates. Genomic selection potentially can explain all genetic variance by markers. The aim of this study was to determine the amount of genomic additive variation for birth weight and weaning weight of Merino sheep by different minor allele frequency (MAF) groups using two statistical approaches i.e. REML and Bayesian.

Material and methods: In this study, data of 2189 Merino sheep, genotyped by 50k Illumina SNP chip were used. After the quality control of genotyping data, 47342 markers remained for subsequent analysis. For birth weight and weaning weight 1331 and 2136 records were available, respectively. To study the association between allele frequency spectrum and captured additive genetic variance, all SNPs were partitioned in five MAF bins with the equal numbers of SNPs (0-0.18, 0.18-0.28, 0.28-0.36, 0.36-0.43 and 0.43-0.499). The analysis was performed using REML and a Bayesian method implemented via Gibbs sampling and RKHS model.

Results: Using all common SNPs in REML approach, estimates of genomic heritability were 0.58 ± 0.07 and 0.46 ± 0.05 for birth weight and weaning weight, respectively. This heritability values in Bayesian analysis and RKHS method for these traits were estimated to be 0.58 ± 0.07 and 0.46 ± 0.05 , respectively. The amount of genetic variation explained by five different MAF groups was different in separate and joint analysis. The estimates from separate analysis were higher than joint analysis for two traits and two (REML and Bayesian) methods. In separate analysis, the genomic heritability was similar for all MAF bins by two approaches. In the joint analysis there were large differences between REML and Bayesian estimates in terms of explaining genetic variation across MAF subsets. For birth weight, SNPs with MAF 0.18-0.28 marked the largest amount of genomic heritability using REML method. In this method, heritability of Weaning weight was zero using SNP with MAF 0.43-0.499. All MAF bins

*Corresponding author: j_fayazi@ramin.ac.ir

contributed to genetic variation in Bayesian method. In separate analysis, sum of genomic variances for five MAF bins was larger than estimated variance by all set of SNPs together, while sum of this variances in joint analysis was same as the value of variance obtained by all SNPs for both traits and both statistical methods.

Conclusion: Variance components obtained by Bayesian method are more realistic, because they have less prediction error variance. Bayesian methods considered common prior distribution for the variance, so results of this method were more reliable than other methods. Although the number of SNPs in different groups was similar, the amount of genetic variance explained by the different MAF groups was different.

Keywords: Genomic prediction, Minor allele frequency, Bayesian approach, Joint analysis, Sheep