

برآورد اثر لایه بندی جمعیتی در گله های خویشاوند و آمیخته گاوهای شیری به روش مطالعات ارتباطی کل ژنومی با شبیه سازی

مرتضی بیطرف ثانی^۱ - محمد رضا نصیری^{۲*} - علی اصغر اسلمی نژاد^۳ - محمد مهدی شریعتی^۴

تاریخ دریافت: ۱۳۹۲/۱۲/۱۷

تاریخ پذیرش: ۱۳۹۳/۰۹/۰۱

چکیده

لایه بندی جمعیتی^۵ یکی از مسائلی است که می‌تواند در مطالعات ارتباطی ژنومی باعث بروز خطا شود. در تحقیق حاضر از شیوه مطالعه ارتباطی ژنومی (طرح Case-Control) برای شناسایی اثر لایه بندی جمعیتی استفاده شد. جامعه آماری و ژنوم ۱۰۰۰۰ گاو در طی ۱۰۰ نسل با روش تعادل رانش - جهش ژنتیکی شبیه سازی و سپس با استفاده از جامعه مذکور، ۸۰۰ گاو آمیخته و خالص با ۵۰۰۰۰ SNP تعیین ژنوتیپ شده در طول ۳۰ جفت کروموزوم، ایجاد شدند. نتایج نشان داد که هر چه نسبت Case/Control در بین جوامع خویشاوند از عدد یک بیشتر منحرف شود، شاخص آماری لامبدا که نشان دهنده اثر لایه بندی جمعیتی است، افزایش خواهد یافت. به طوری که شاخص لامبدا در مدل ژنتیکی افزایشی با نسبت های ۱، ۰/۷۷ و ۰/۳۳ به ترتیب ۰/۴۲، ۱۱/۳۱ و ۹۷/۷۷ و در مدل ژنتیکی عدم غلبه، به ترتیب ۰/۴۷، ۸/۲۱ و ۵۷/۴۰ برآورد گردید. اثر لایه بندی جمعیتی در بین گروههای مختلف جامعه گاوهای آمیخته وجود نداشت و شاخص لامبدا در مدل‌های ژنتیکی عدم غلبه، غلبه کامل، مغلوبیت، افزایشی و فوق غلبه به ترتیب ۰/۵۵، ۰/۶۶، ۰/۸۹، ۰/۷۶ و ۰/۴۱ برآورد شد. نتایج تحقیق حاضر نشان داد که برای کنترل اثر لایه بندی جمعیتی در مطالعات ارتباطی ژنومی، گروههای case-control نایستی از جوامعی مختلف با اجداد و درجه خویشاوندی متفاوت انتخاب شوند، مگر با نسبت‌های کاملاً یکنواخت در دو جامعه. همچنین، پیشنهاد می‌گردد از گاوهای آمیخته به دلیل عدم وجود عامل ساختار ژنتیکی جمعیتی استفاده گردد.

واژه های کلیدی: لایه بندی جمعیتی، مطالعات ارتباطی، SNP.

مقدمه

که توجیه کننده تنوع صفات مهم اقتصادی در حیوانات اهلی است، نتایج قابل توجهی در دهه‌های اخیر حاصل شده است، اما به خاطر تراکم پایین مارکرها در مطالعات نقشه یابی QTL، برآورد مقدار واقعی تنوع ژنتیکی صفات کمی با این شیوه امکانپذیر نبود. مطالعات ارتباطی کل ژنومی^۶ با استفاده از مارکر های متراکم "چند شکلی تک نوکلئوتیدی (SNP)"، روش جدیدی است که بر این مشکل فائق آمد (۱۶) و یک روش مؤثر برای شناسایی تنوع ژنتیکی مرتبط با احتمال ابتلا بیماری و صفات چند عاملی (پیچیده) است (۲). این مطالعات در ابتدا برای بیماری‌های انسانی استفاده شد و سپس به سرعت گسترش یافت. با تولید و انتشار داده‌های ارزشمند و تعداد زیادی از SNP ها برای ژنوم چندین گونه اهلی، مطالعات ارتباطی کل ژنوم در ژنتیک و اصلاح نژاد دام گسترش یافت (۱). تراکم بالای مارکرهای SNP که در مطالعات ارتباطی کل ژنومی استفاده می‌شود، به اندازه کفایت، عدم تعادل لینکاژی بین مارکر SNP و جهش سببی را باعث

با تکمیل پروژه ژنوم انسان در سال ۲۰۰۳، محققین شروع به بررسی تنوع ژنتیکی بین افراد پرداختند (۱۱). پس از مدت کوتاهی آنها بیشترین تنوع موجود در بین افراد را در SNP ها یافتند؛ در حقیقت ۱۰ میلیون SNP در ژنوم انسان شناسایی گردید. در ادامه پروژه‌هایی به منظور شناسایی و نقشه‌یابی این تنوع ژنتیکی در بین جوامع و نقاط مختلف آغاز گردید (۱۱). حیوانات اهلی منابع ارزشمندی برای مطالعه ساختار ملکولی صفات چند عاملی و پیچیده هستند. اگر چه بواسطه نقشه‌یابی جایگاه‌های صفات کمی (QTL)

۱- دانش آموخته دکتری اصلاح نژاد (ژنتیک کمی) دانشگاه فردوسی مشهد،

۲- استاد ژنتیک و بیوتکنولوژی دانشگاه فردوسی مشهد،

۳- دانشیار گروه علوم دامی دانشگاه فردوسی مشهد،

۴- استادیار گروه علوم دامی دانشگاه فردوسی مشهد.

(Email: nassiry@um.ac.ir

*) نویسنده مسئول:

می شود (۱۶). استفاده از این مطالعات برای صفات مهم اقتصادی در گاو شامل میزان تولید شیر، کیفیت شیر، باروری، رشد، کیفیت گوشت و صفات در ارتباط لاشه در بیش از ۱۰ مقاله گزارش شده است (۱۶). اگر چه اهمیت انتخاب گروهای case و control از یک جامعه در دهه های گذشته مشخص شده است، اما با ظهور مطالعات ارتباطی کل ژنومی این موضوع بیشتر حائز اهمیت شد (۳). لایه بندی جمعیتی، که ناشی از تفاوت ژنتیکی جد مشترک است، مهمترین چالش مطالعات ارتباطی کل ژنومی است و باعث نتایج اریب مرتبط با بیماری یا صفت مورد نظر می‌گردد (۲). آزمون‌های ارتباطی ژنتیکی به شناسایی تفاوت‌های فراوانی آلی در بین گروه‌های case و control می‌پردازد (۳). اشتباه نوع اول در این آزمون‌ها زمانی رخ می‌دهد که به خاطر خطای سنجش، تفاوت‌های معنی‌دار در فراوانی آلی برآورد گردد یا به عبارتی علت واقعی تفاوت معنی‌دار بین فراوانی‌های آلی در دو گروه در ارتباط با بیماری یا صفت چند عاملی نباشد (۳). در طرح‌های case-control ممکن است یکسری ارتباطاتی که مربوط به ساختار جمعیتی است، شناسایی می‌شود (۴)، ۱۰ و ۱۲). دلیل اصلی لایه بندی جمعیتی، آمیزش‌های غیر تصادفی بین گروه‌هاست، که اغلب به دلیل رانش ژنتیکی فراوانی‌های آلی در هر گروه می‌باشد. در جوامع معاصر، به دلیل هم‌آمیزی و اختلاط افراد جوامع مختلف، اجداد متفاوتی وجود خواهد داشت (۴، ۱۰ و ۱۲). بیش از ۱۰ نسل آمیزش تصادفی می‌تواند عامل لایه بندی جمعیتی را حذف کند (۴). فرضیه همگنی جامعه در مطالعات ارتباطی بویژه طرح‌های case-control، می‌تواند به آسانی نقض شود و باعث اشتباه نوع اول و دوم گردد (۴، ۱۰ و ۱۲). هیچ جامعه طبیعی، خواه انسانی یا غیر انسانی، نمی‌تواند در معنی واقعی کلمه از نظر ژنتیکی "همگن" باشد. عوامل جهش، رانش ژنتیکی، مهاجرت و انتخاب باعث شکل‌گیری گروه‌های مختلف و ایجاد الگوهای منحصر به فرد تنوع می‌شود. کاملاً واضح است که میزان تنوع بین گروه‌ها به فاصله زمانی تا جد مشترک بستگی دارد (۳). به دلیل وجود مشکل لایه‌بندی جمعیتی در مطالعات ارتباطی ژنومی طی سالهای اخیر، برآورد آن ضروری به نظر می‌رسد. هدف از این تحقیق برآورد اثر لایه بندی جمعیتی در بین گله‌های خویشاوند و آمیخته گاوهای شیری با مدل‌های پنج‌گانه ژنتیکی به روش کنترل ژنومی^۱ می‌باشد.

مواد و روش‌ها

برای ارزیابی لایه‌بندی جمعیتی و انحراف ژنتیکی از روش GWAS (طرح case-control) و ۸۰۰ گاو شبیه‌سازی شده آمیخته

و همخون استفاده شد. جامعه تاریخی^۲ و ژنوم ۱۰۰۰۰ گاو پس از طی ۱۰۰ نسل با روش تعادل جهش - رانش ژنتیکی شبیه‌سازی گردید (۱۳). وراثت پذیری صفت ۰/۲۰ منظور گردید. لاین‌های با ارزش ژنتیکی بالا (HL^۳) و پایین (LL^۴) پس از ۲۰ نسل انتخاب بر اساس ارزش اصلاحی واقعی و با استفاده از جامعه تاریخی شبیه‌سازی گردید. نسبت جنسیتی والدین (نر به ماده) و نتاج برای هر نسل به ترتیب ۰/۵ و ۰/۰۵ در نظر گرفته شد. چهار ترکیب مختلف case-control از بین لاین‌های HL و LL انتخاب شدند. همچنین، گاو‌های آمیخته از تلاقی لاین‌های HL, LL ایجاد و دو گروه تصادفی case-control از بین آنها ایجاد گردید (جدول ۱).

۵۰۰۰۰ SNP روی ۳۰ کروموزوم برای هر گاو شبیه‌سازی شد (جدول ۲). زمانیکه منشاء گروه‌های case و control از جوامع مختلف با درجات متفاوت خویشاوندی است، روش کنترل ژنومی برای پیمایش ارتباطات مارکرها تحت فرضیه صفر و برآورد لایه بندی جمعیتی با استفاده از آماره لامبدا و نمودار Q-Q plot استفاده می‌گردد. آماره لامبدا که از تقسیم میانه مقادیر کای مربع مشاهده شده بر میانه مورد انتظار (۰/۴۵۶) حاصل می‌شود، فاکتور تورم می‌باشد (۳). فراوانی آلی هر کدام از SNP‌ها در اولین نسل جامعه تاریخی برابر در نظر گرفته شد. قسمتی از پردازش فایل داده‌های ژنومی توسط نرم افزار اکسل انجام شد انحراف ژنتیکی با مدل رگرسیون لجستیک که متغیرهای مستقل آن اسنیپ‌ها می‌باشد، برآورد گردید. از مدل رگرسیونی خطی تعمیم یافته (۱) برای مطالعه ارتباطی هر SNP استفاده شد:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad \text{مدل (۱)}$$

Y_i: متغیر وابسته گسسته (۲: case، ۱: control)

X_i: اثر ژنوتیپ i امین چند شکلی تک نوکلئوتیدی در مدل‌های مختلف ژنتیکی: (در مدل ژنتیکی افزایشی تعداد i امین آلل مینور، در سایر مدل‌ها اثر ژنوتیپ لحاظ شد)

نسبت بخت تمامی اسنیپ‌ها برای ۵ مدل ژنتیکی شامل افزایشی، غلبه کامل، فوق غلبه، عدم غلبه و مغلوب برآورد شد (۱). در مدل ژنتیکی افزایشی فراوانی آلل خطر در دو گروه case و control برای هر SNP مقایسه می‌شود و نسبت بخت آللی محاسبه می‌گردد. در مدل ژنتیکی غالبیت، نسبت بخت ژنوتیپ‌های حداقل دارای یک آلل خطر با ژنوتیپ بدون آلل خطر برای هر SNP محاسبه و مورد آزمون آماری قرار می‌گیرد.

^۲-Historical population

^۳-High level

^۴-Low level

^۱- Genomic Control

جدول ۱- ترکیبات مختلف case/control در مطالعه بین دو لاین با ارزش ژنتیکی بالا و پایین

Table 1- different ratios of case /control between two lines by high and low breeding value in GWAS

ترکیب composition	طرح شماره یک ^۱		طرح شماره دو ^۲		طرح شماره سه ^۳		طرح شماره چهار ^۴	
	HL	LL	LL	HL	LL	HL	LL	HL
Cases	400	400	400	400	450	350	600	200
Control	400	400	400	400	350	450	200	600
Total	800	800	800	800	800	800	800	800

^۱ طرح شماره یک تعداد افراد گروه case و control در لاین های HL و LL برابر و شامل ۴۰۰ case و ۴۰۰ control در هر دو لاین است.

^۲ طرح شماره دو: مشابه طرح شماره یک با این تفاوت که ۴۰۰ گاو اول در لاین HL به عنوان case و در لاین LL به عنوان control در نظر گرفته شده است.

^۳ طرح شماره سه شامل ۳۵۰ case و ۴۵۰ control در لاین HL و ۴۵۰ case و ۳۵۰ control در لاین LL است.

^۴ طرح شماره چهار شامل ۲۰۰ case و ۶۰۰ control در لاین HL و ۶۰۰ case و ۲۰۰ control در لاین LL است.

لامبدا در تمامی مدل های ژنتیکی در طرح شماره یک، اثر لایه بندی جمعیتی وجود نداشت، اما مقدار این آماره در طرح های شماره دو، سه و چهار بیشتر از عدد یک برآورد گردید، که نشان می دهد نتایج مطالعات ارتباطی ژنومی در طرح های مذکور به دلیل وجود اثر لایه بندی جمعیتی، اریب خواهد بود. انحراف بیشتر نسبت تعداد / case control از عدد یک، باعث افزایش مقدار آماره لامبدا و اثر لایه بندی جمعیتی می شود. زمانی که منشاء گروه های case و control از جوامع مختلف با درجات متفاوت خویشاوندی باشد، شناسایی SNP های اثر گذار و مرتبط با بیماری یا صفات چند عاملی دچار اریبی خواهد شد (۱ و ۳). لذا فرضیه استقلال مشاهدات رد می شود و اغلب باعث برآورد بیش از حد ارتباط های ژنتیکی معنی دار می شود که به خاطر روش نمونه گیری است (۶). برای کنترل اثر لایه بندی جمعیتی در مطالعات ارتباطی ژنومی، نمونه گیری از جوامع با درجات متفاوت خویشاوندی پیشنهاد نمی شود. البته در صورت توزیع کاملا یکنواخت گروه های case و control در بین این جوامع که در شرایط واقعی دور از ذهن است، اثر لایه بندی جمعیتی رخ نخواهد داد. از توزیع Q-Q plot برای بررسی تطابق توزیع اثرات چند شکلی های تک نوکلئوتیدی با توزیع مورد انتظار استفاده می شود (شکل ۱). در این نمودار آماره های اندازه گیری شده برای تک تک SNP ها منفی لگاریتم P-value در محور x از کوچک به بزرگ مرتب می شود و با مقادیر مورد انتظار آن در محور y بررسی می شود. هر گونه انحراف از خط یا به منزله وجود اثر لایه بندی جمعیتی است و یا اسنیپ ها مرتبط با بیماری یا صفات مد نظر می باشد. نمودارهای A، B، C و D به ترتیب نتایج ارزیابی اثر لایه بندی جمعیتی در طرح های شماره یک، دو، سه و چهار می باشد که این عامل در طرح شماره یک وجود نداشت، ولی در طرح های شماره دو تا چهار شناسایی شد (شکل ۱).

به همین ترتیب برای مدل ژنتیکی غلبه ماورایی نسبت بخت ژنوتیپ هتروزگوت نسبت به هموزیگوتها و در مدل عدم غلبه، نسبت بخت هر کدام از ژنوتیپ های دارای یک و دو آلل خطر نسبت به ژنوتیپ بدون آلل خطر و در مدل مغلوب، نسبت بخت ژنوتیپ بدون آلل خطر نسبت به ژنوتیپ های دارای یک یا هر دو آلل خطر محاسبه و مورد آزمون آماری قرار می گیرد (۱). برای محاسبه نسبت بخت ژنوتیپی و آللی برای هر کدام از چند شکلی های تک نوکلئوتیدی از $\exp(\beta)$ استفاده گردید. برای بررسی اثر هر کدام از SNP ها با مدل های مختلف ژنتیکی از آزمون فیشر استفاده گردید. سطح آماری معنی داری به روش بنفرونی تصحیح شد. این روش باعث رفع مشکل مقایسات چند گانه می شود. شبیه سازی داده های ژنومی با نرم افزار QMSim^۱ نسخه ۱/۰۰ صورت گرفت (۱۰). آنالیز داده ها با استفاده از نرم افزار R نسخه ۳/۰/۱ (بسته نرم افزاری SNPassoc) انجام شد. برای بررسی اثر لایه بندی جمعیتی از کروموزوم شماره ۱ استفاده شد، ولی برای بررسی انحراف ژنتیکی بین لاین های شبیه سازی شده از مطالعه ارتباطی کل ژنومی استفاده گردید.

نتایج و بحث

مطالعات ارتباطی ژنومی در بین گله های خویشاوند

لایه بندی جمعیتی با استفاده از روش GWAS بررسی شد که نتایج آن در بین گله های همخون در جدول ۳ و شکل ۱ ارائه شده است.

اگر مقدار آماره لامبدا کمتر یا مساوی یک باشد، نشان دهنده عدم وجود اثر لایه بندی جمعیتی است. بنابراین، با توجه به مقادیر

۱ - <http://www.aps.uoguelph.ca/~msargol/qmsim>

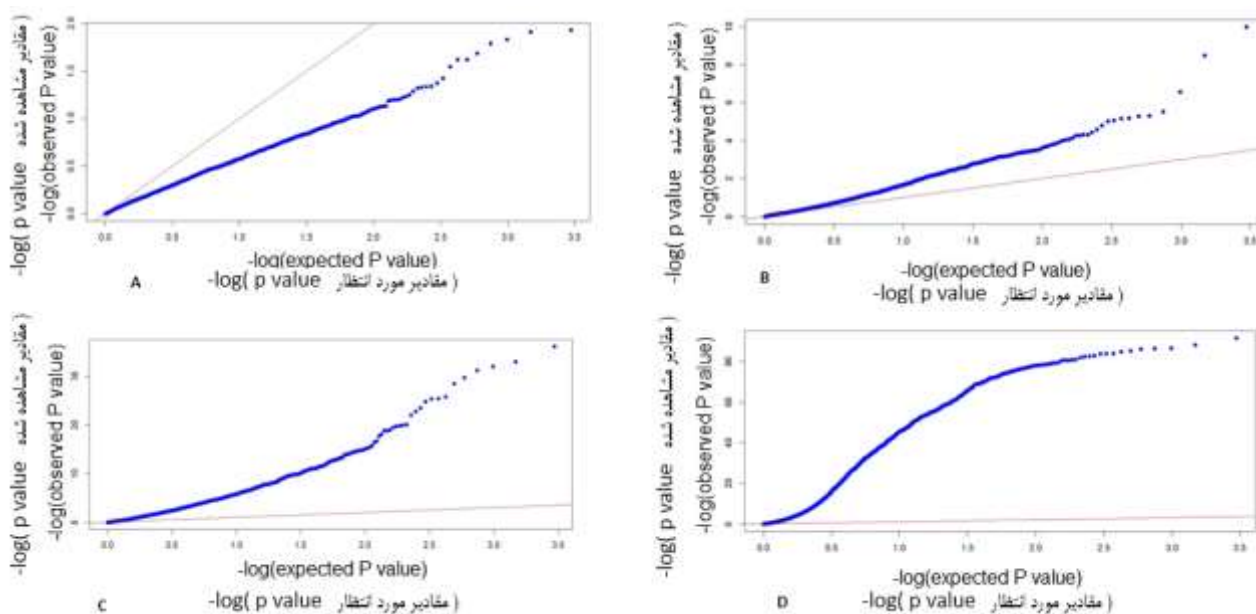
جدول ۲- تعداد اسنپ‌های شبیه‌سازی شده روی کروموزوم‌های ۲ گاو
Table 2- number of simulated SNPs on chromosome

Chromosome	Number of SNPs	chromosome	Number of SNPs
1	2970	16	1541
2	2575	17	1409
3	2274	18	1240
4	2274	19	1203
5	2274	20	1353
6	2274	21	1353
7	2236	22	1146
8	2124	23	996
9	1992	24	1184
10	1954	25	808
11	2011	26	977
12	1710	27	845
13	1578	28	864
14	1597	29	977
15	1597	X	2800

۲ تعداد حدوداً ۵۰۰۰۰ اسنپ بسته به اندازه کروموزوم‌ها روی آنها شبیه‌سازی شد.

جدول ۳- آماره لامبدا برآورد شده در بین گله‌های خویشاوند
Table 3- Estimated lambda statistics of inbred cattle

لامبدا برآورد شده Estimated lambda	مدل ژنتیکی روابط بین آلی بین SNPها Genetic model				
	عدم غلبه co dominant	غالبیت dominant	مغلوبیت Recessive	غلبه ماورایی over dominant	افزایشی additive
Design 1	0.4721	0.4105	0.5244	0.4178	0.4334
Design 2	3.6257	3.6245	5.9454	1.5319	5.3579
Design 3	8.2157	10.2572	8.8208	7.5111	11.3154
Design 4	57.4028	98.8208	33.9363	5.0125	97.7761



شکل ۱- نمودارهای Q-Q plot در مطالعات ارتباطی ژنومی بین گله‌های خویشاوند برای ارزیابی لایه بندی جمعیتی

Figure 1- quantile-quantile plots in GWAS of inbred cattle for evaluation population stratification

گله آمیخته شامل ۸۰۰ راس گاو شبیه سازی شد. از بین گاوهای آمیخته، ۴۰۰ راس به عنوان case و ۴۰۰ راس به عنوان control انتخاب شدند و مطالعه ارتباطی ژنومی بین این دو گروه انجام شد. متوسط فراوانی آلی SNPها روی هر کروموزوم در جامعه گاوهای آمیخته، نشان دهنده افزایش هتروزیگوسیتی و برابری فراوانی هر دو آلل هر SNP می باشد (جدول ۴).

میانگین مقادیر نسبت بخت و p value برای اسنیپها به ترتیب ۱/۰۱ و ۰/۵۸ برآورد شد که نشان می دهد هیچ اختلاف ژنتیکی معنی داری بین زیرگروه های جامعه گاوهای آمیخته وجود ندارد (شکل ۲ و ۳).

با افزایش انحراف نسبت تعداد case / control از عدد یک، میزان انحراف مقادیر مشاهده شده از مورد انتظار بیشتر می شود، که حاکی از افزایش اثر لایه بندی جمعیتی است. انحراف ژنتیکی بین دو لاین با ارزش اصلاحی بالا (HL) و پایین (LL) که با روش GWAS بررسی شد، نشان می دهد که ۸۸/۲ درصد از SNPها با نسبت بخت بیشتر از ۵ بین دو لاین، اختلاف معنی دار دارند ($p < 1e-06$). انتخاب طی ۲۰ نسل بر اساس ارزش ژنتیکی واقعی در دو جهت متفاوت باعث شده است تا فراوانی های آلی و ژنوتیپی اکثر اسنیپ ها بین دو لاین مورد نظر در سطح آلفای بنفرونی متفاوت باشد.

مطالعات ارتباطی ژنومی در گله های آمیخته

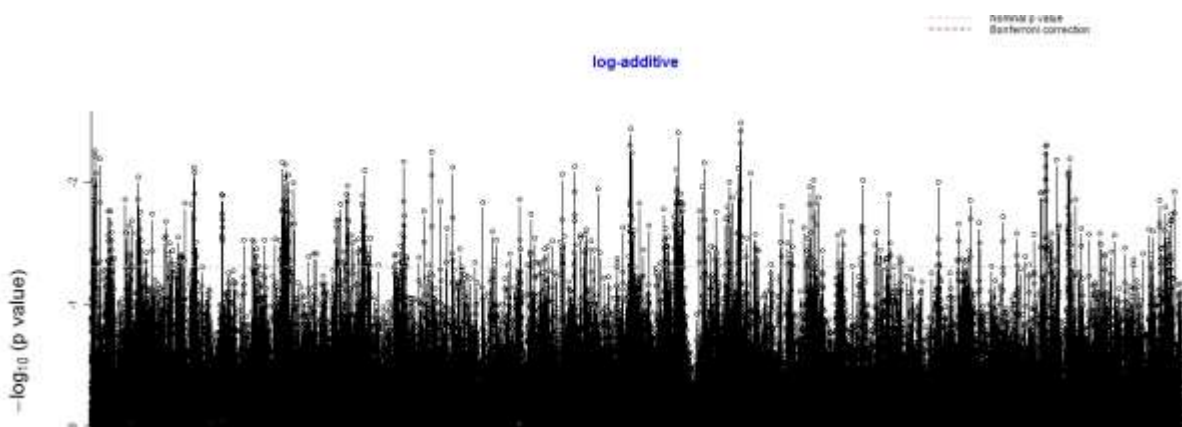
از تلاقی لاین های با ارزش اصلاحی بالا (HL) و پایین (LL)،

جدول ۴ - میانگین فراوانی آلی اسنیپ ها در جامعه گاوهای آمیخته

Table 4- mean of SNPs allele frequency in crossbred population

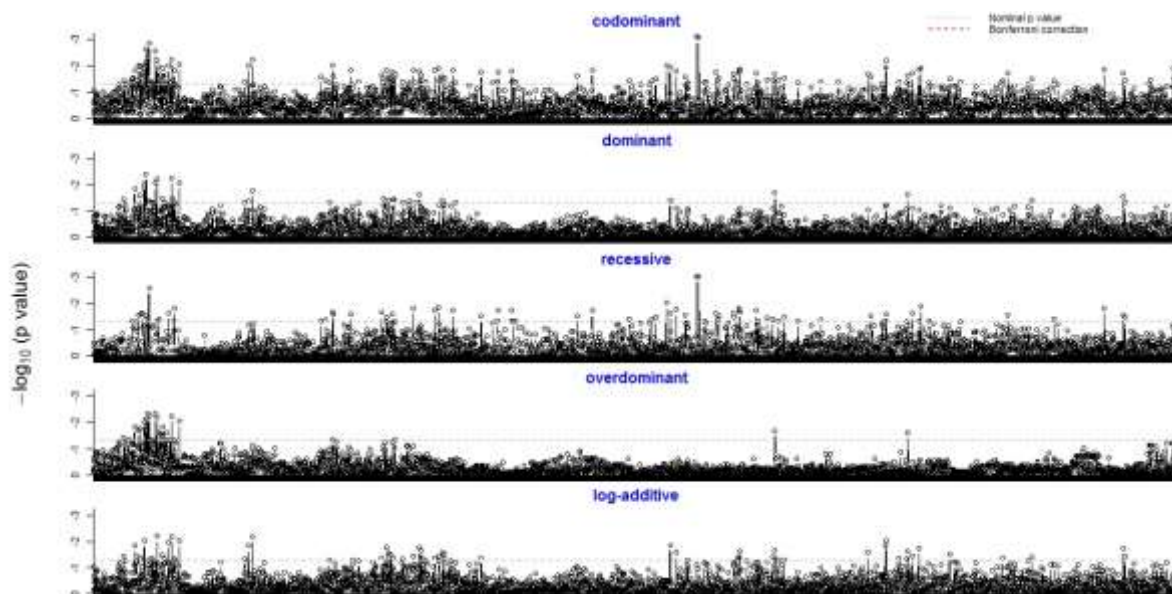
میانگین فراوانی الل	کروموزوم	میانگین فراوانی الل ^۱	کروموزوم
Average of allele frequency	Chr	Average of allele frequency	Chr
0.500	16	0.510	1
0.492	17	0.500	2
0.505	18	0.499	3
0.499	19	0.494	4
0.488	20	0.496	5
0.507	21	0.503	6
0.505	22	0.496	7
0.503	23	0.490	8
0.508	24	0.503	9
0.505	25	0.495	10
0.503	26	0.510	11
0.494	27	0.498	12
0.506	28	0.500	13
0.508	29	0.500	14
0.503	X	0.502	15

افرم های آلی هر اسنیپ، به صورت ۰ و ۱ نشان داده می شود که در این جدول فراوانی آلی فرم ۰ لحاظ شده است.



شکل ۲ - نمودار نتایج مطالعه ارتباطی ژنومی ۵۰۰۰۰ اسنیپ بین دو گروه تصادفی گاوهای آمیخته در مدل ژنتیکی افزایشی

Figure 2- Manhattan plot for GWAS of ~50000 SNP of whole genome between 2 random subpopulation of crossbred cattle for additive genetic models



شکل ۳ - نمودار نتایج مطالعات ارتباطی ژنومی در مدل های مختلف ژنتیکی روی کروموزوم شماره ۱ بین دو گروه تصادفی جامعه گاوهای آمیخته
Figure 3- Manhattan plot for genome wide association studies of 1 chromosome between 2 random subpopulation of crossbred cattle for all genetic models

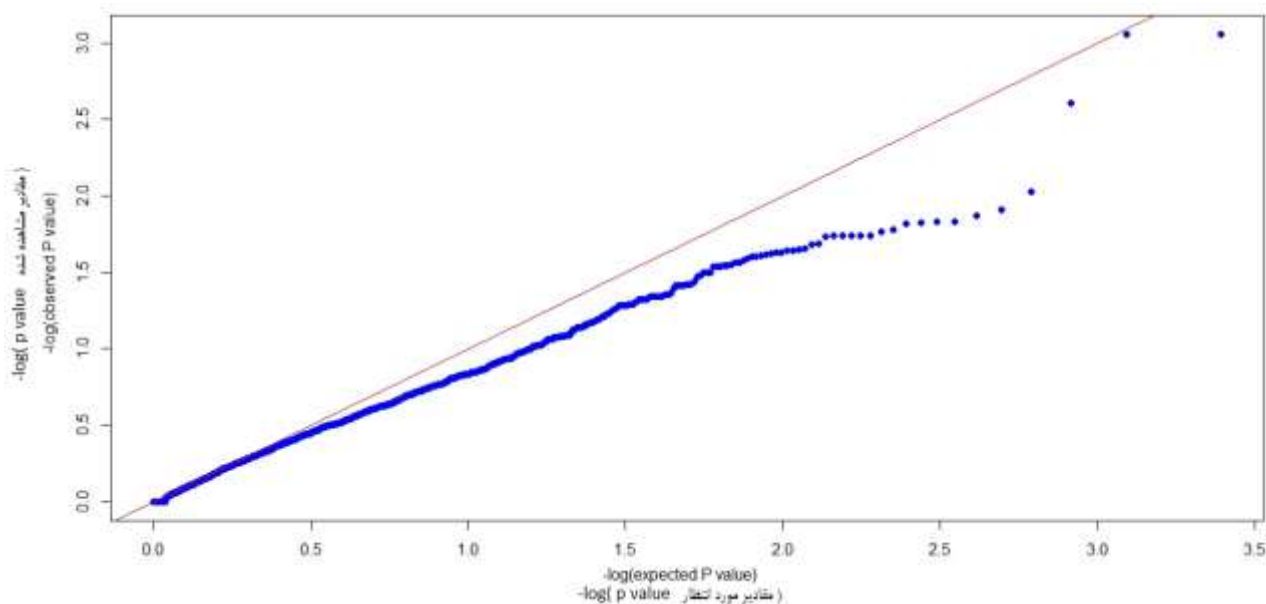
لايه‌بندی جمعیتی به خاطر تفاوت‌های سیستماتیک در فراوانی آلی بین زیرجوامع می باشد که عامل ایجاد آریبی در مطالعات ارتباطی ژنومی می شود (۹ و ۷). تلقیح مصنوعی و انتخاب ژنومی عوامل اصلی در ایجاد لایه بندی جمعیتی در گاوهای شیری هستند (۸). استفاده گسترده از تلقیح مصنوعی و انتخاب ژنومی، باعث شتاب بیشتر استفاده های مکرر از ژنوم گاوهای نر گردیده است. رانش ژنتیکی ناشی از نمونه گیری، به ناچار باعث لایه بندی جمعیتی می-شود، چرا که نسبت کمی از جامعه در ایجاد نسل بعد مشارکت خواهند داشت. رانش ژنتیکی ناشی از انتخاب که باعث تغییرات فراوانی آلی می شود، به عدم تعادل ناشی از پیوستگی ژنها بین لوکوس‌های تحت انتخاب بستگی دارد و باعث لایه بندی جمعیتی و اثرات جدی ارتباطی خواهد شد (۸). ۴۰ سال انتخاب مصنوعی در گاو شیری باعث تغییرات گسترده ژنومی شده است، به طوری که تفاوت فراوانی‌های آلی بین گروه‌های منتخبین و غیر منتخبین کاملاً مشهود است (۱۴).

مطالعات ارتباطی ژنومی، کلیه SNPها را در سراسر ژنوم برای بررسی تنوع ژنتیکی بین دو گروه case و control بررسی می کند (۴). مقادیر منفی لگاریتم P-value برای تمامی SNPها در مقایسه دو گروه مذکور برای مدل ژنتیکی افزایشی در شکل ۲ نشان داده شده است. خط چین پر رنگ، آستانه معنی‌داری تصحیح شده bonferroni و خط چین کم رنگ، آستانه معنی‌داری ۰/۰۵ را نشان می‌دهد. فراوانی آلی هیچ کدام از SNPها در بین زیرگروه‌های جامعه گاوهای آمیخته، اختلاف معنی‌داری نداشت (شکل ۲). فراوانی آلی و ژنوتیپی هیچ کدام از SNPهای کروموزوم شماره ۱ بین زیرگروه‌های جامعه گاوهای آمیخته اختلاف معنی‌داری نداشت (شکل ۳). مقادیر آماره لامبدا در همه مدل‌های ژنتیکی در مطالعات جامعه گاوهای آمیخته، کمتر از عدد یک برآورد شد (جدول ۵). بنابراین، اثر لایه بندی جمعیتی در بین مطالعات زیرگروه های جامعه گاوهای آمیخته وجود ندارد (شکل ۴).

جدول ۵ - مقادیر آماره لامبدا در مطالعات ارتباطی ژنومی جامعه گاوهای آمیخته

Table 5- estimated lambda statistics of crossbred cattle

آماره statistic	مدل ژنتیکی روابط بین آلی اسنپ ها Genetic model of SNPs alleles				
	فوق غلبه over dominant	افزایشی additive	مغلوبیت Recessive	غلبه کامل dominant	عدم غلبه co dominant
لامبدا lambda	0.41783	0.76190	0.89424	0.66333	0.55106



شکل ۴- Q-Q plot در مطالعات ارتباطی ژنومی برای ارزیابی لایه بندی جمعیتی در جامعه گاوهای آمیخته
Figure 4 - quantile-quantile plots in GWAS for evaluation population stratification

ژنومی در گاوهای هلشتاین، اثرات مرتبط با لایه بندی حاصل از انتخاب ژنومی را حذف خواهد کرد (۸). روشهای تصحیح لایه بندی جمعیتی از جمله PCA و MDS در مقایسه با روش کنترل ژنومیک حساس هستند و باعث حذف برخی از مارکرهای مرتبط واقعی می شود (۷). به منظور کنترل اثر لایه بندی جمعیتی، مطالعات ارتباطی ژنومی در جامعه گاوهای آمیخته پیشنهاد می گردد. البته برای بررسی اثرات متقابل ژنتیک و محیط در مطالعات ارتباطی ژنومی، نمی توان از QQ plot و روش کنترل ژنومی برای برآورد و تصحیح اثر لایه بندی جمعیتی استفاده کرد زیرا باعث اربیی خواهد شد (۱۵).

مهمترین مشکل مطالعات ارتباطی کل ژنومی، اثر لایه بندی جمعیتی است که در مطالعات بین جوامع گاو با درجات مختلف خویشاوندی رخ می دهد. در روش کنترل ژنومی، آماره لامبدا که نشان دهنده میزان اثر لایه بندی جمعیتی است، به نسبت $case / control$ در بین زیرگروه های مختلف جوامع همخون، بستگی دارد بطوری که هر چه این نسبت بیشتر از عدد یک منحرف شود، اربیی نتایج مطالعات ارتباطی بیشتر می شود. لایه بندی جمعیتی در مطالعات ارتباطی کل ژنومی با افزایش اندازه نمونه و تعداد مارکرها، زیادتر می شود (۹ و ۵). تصحیح لایه بندی جمعیتی در مطالعات ارتباطی کل

منابع

- Alkes, L., A. N. Price, D. Z. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11: 459–463.
- Anthony H. L., K. E. Larkin and B. K. Suarez. 2009. Population Stratification and Patterns of Linkage Disequilibrium. *Genetic Epidemiology*, 33: S88–S92.
- Devlin, B., and K. Roeder. 1999. Genomic control for association studies. *Biometrics*, 55:997–1004.
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D. 2004. Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, 36(4):388–93.
- Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. *Science*. 265:2037–2048.
- Li l, Donghui Z, Hong L and Christopher A.2013. Robust methods for population stratification in genome wide association studies. *BMC Bioinformatics*, 14:132.
- Li, M., G. R. Wiggans, Sh. Wang, T. S. Sonstegard, J. Yang, B. A. Crooker, J. B. Cole, C. P. Van Tasse, T. J. Lawlor, and Y. Da1. 2012. Effect of sample stratification on dairy GWAS results. *BMC Genomics*. 13:536.
- Marchini J, Cardon LR, Phillips MS, Donnelly P .2004.The effects of human population structure on large genetic association studies. *Nature Genetics* May; 36(5): 512-7. Epub 2004 Mar 28.
- Marchini, J., L. R. Cardon, M. S. Phillips, and P. Donnelly. 2004. The effects of human population structure on large genetic association studies. *Nature Genetics*, 36(5): 512–517.

10. Norrgard, K. 2008. Genetic variation and disease: GWAS. Nature Education: <http://www.nature.com/scitable/topicpage/genetic-variation-and-disease-gwas-682>.
11. Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association. *Nature Genetics*. 38:904-909.
12. Sargolzaei, M., and F. S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25: 680-681.
13. Sonstegard, T. S., M. Li, C. P. V. Tasse, E. S. Kim, J. B. Cole, G. R. Wiggans, B. A. Crooker, B. D. Mariani, L. K. Matukumalli, J. R. Garbe, S. C. Fahrenkrug, G. Liu, and Y. Da. 2010. Forty Years of Artificial Selection In U.S. Holstein Cattle Had Genome-wide Signatures. *Proc. Of 9th world congress on genetics applied to livestock production*.
14. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42:355-360.
15. Zhang, H., Z. Wang, S. Wang, and H. Li. 2012. Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology*, 3:26.
16. kushyar M M, nassiri M, Bitaraf sani M, Aslaminejad A A. 2013. Feasibility Study of the Detection of SNPs Associated with Breast Cancer by Genome-Wide Association Virtual Studies. *Genetics In The 3rd Millennium*. vol:3. 3190-3199. (In Persian).