

## بررسی ساختار بلوک‌های هاپلوتیپی و پویش ژنومی صفت افزایش وزن بدن در کروموزوم‌های آتوزومال گوسفند نژاد زندی با استفاده از مدل هاپلوتایپ

حسین محمدی<sup>۱</sup>، سید عباس رأفت<sup>۲\*</sup>، حسین مرادی شهر بابک<sup>۳</sup>، جلیل شجاع<sup>۴</sup> و محمد حسین مرادی<sup>۲</sup>

تاریخ دریافت: ۹۶/۸/۱ تاریخ پذیرش: ۹۶/۹/۱۹

<sup>۱</sup> دانش‌آموخته دکتری گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

<sup>۲</sup> استاد گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

<sup>۳</sup> استادیار گروه علوم دامی، پردیس کشاورزی و منابع طبیعی کرج، دانشگاه تهران

<sup>۴</sup> استادیار گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک

\*مسئول مکاتبه: Email: abbasrafat@hotmail.com

### چکیده

**زمینه مطالعاتی:** افزایش وزن روزانه از مهمترین سنج‌های تعیین‌کننده سود اقتصادی در پرورش گوسفند است. آگاهی از ویژگی‌های عدم تعادل پیوستگی (LD) و ساختار بلوک‌های هاپلوتیپی در مطالعات پویش ژنوم و انتخاب ژنومی معیارهای کلیدی می‌باشند. هدف: این تحقیق به منظور مطالعه گستره LD، ساختار بلوک هاپلوتیپی و ارتباط ژنومی هاپلوتیپی گوسفند برای شناسایی مناطق ژنومی مؤثر بر صفات افزایش وزن روزانه قبل (AGW) و بعد از شیرگیری (PWG) در گوسفند زندی اجرا شد. روش کار: از ۹۶ رأس گوسفند زندی نمونه خون تهیه شد و با استفاده از آرایه‌های SNPChip 50 K شرکت ایلومینا تعیین ژنوتیپ شدند. پس از مراحل کنترل کیفی، در نهایت ۴۰۸۷۹ نشانگر SNP مربوط به ۹۴ حیوان آنالیز شدند. مقدار LD با محاسبه آماره  $r^2$  بین تمام جفت جایگاه‌ها از طریق نرم افزار PLINK و بلوک‌های هاپلوتیپی بوسیله نرم افزار Haploview برای هر کروموزوم محاسبه شدند. پس از شناسایی اثرات ثابت معنی‌دار (سال تولد و تیپ تولد)، مطالعه پویش ژنومی در نرم افزار PLINK ارزیابی و برای کنترل نرخ اشتباه از تصحیح بنفرونی استفاده شد. نتایج: در این مطالعه گستره مفید عدم تعادل پیوستگی در 40K برابر با  $r^2 = 0/2$  برآورد شد. ۷/۵۸ درصد از کل SNPها درون بلوک‌های هاپلوتیپی و ۱/۴۵ درصد از ژنوم اتوزومی توسط بلوک‌ها پوشش داده شد. با انجام آنالیزهای پویش ژنومی، در مجموع چهار جایگاه هاپلوتیپی روی کروموزوم‌های ۳، ۵، ۶ و ۷ شناسایی شد، به طوری که، ارتباط معنی‌داری بین هاپلوتیپ‌های کروموزوم ۵، ۶ و ۷ با صفت AWG و توجیه ۳/۴۳ درصد از واریانس صفت و هاپلوتیپ کروموزوم ۳ با PWG و توجیه ۱/۵۲ درصد از واریانس صفت به دست آمد. نتیجه‌گیری نهایی: ژن‌های کاندیدای شناسایی شده حاصل از آنالیزهای هاپلوتیپی عملکرد مولکولی مرتبط با صفات رشد داشتند که قابل استفاده بودن این یافته‌ها در ارزیابی‌ها، سبب تسریع در پیشرفت ژنتیکی افزایش وزن خواهد شد.

**واژگان کلیدی:** عدم تعادل پیوستگی، بلوک هاپلوتیپی، مطالعات پویش ژنومی، ژن کاندیدا، افزایش وزن روزانه

## مقدمه

میزان که طول بلوک‌های هاپلوتیپی بلندتر باشد نیاز به نشانگرهای ژنتیکی کمتری برای مطالعات ژنومیکی نیاز است. همچنین بلوک‌های هاپلوتیپی می‌توانند در شناسایی مناطق ژنومی تحت انتخاب طی تکامل استفاده شوند (ثابتی و همکاران ۲۰۰۲ و محمدی و همکاران ۲۰۱۸). اولین بار ساختار بلوک‌های هاپلوتیپی در گوسفند چورا اسپانیایی توسط (گارسیا-گامز و همکاران ۲۰۱۲) گزارش شد و توانستند در حدود ۲۰۹۹ بلوک با میانگین ۲۷/۰۳ کیلوباز روی کروموزوم‌های اتوزومی شناسایی کنند که ۲/۳۲ درصد از طول کروموزوم را پوشش می‌داد.

از سویی دیگر در مطالعه GWAS در گوسفند با استفاده از SNPChip 50K توسط جانستن و همکاران (۲۰۱۱) در صفت تیپ شاخ ارائه شد و جایگاه‌های ژنی مرتبط با این صفت به طور موفقیت آمیزی شناسایی شد. ریجیو و همکاران (۲۰۱۳) در مطالعه ارتباط ژنومی برای شناسایی نواحی ژنومی مؤثر بر صفات رشد شامل وزن بدن در سنین مختلف در گوسفند، ۷۵۲ رأس گوسفند نژاد بلک فیس اسکاتلندی را با استفاده از SNPChip 50K مورد ارتباط پویش ژنومی قرار دادند. در این مطالعه ارتباط معنی‌داری با صفات رشد شامل وزن تولد، وزن ۱۶ هفتگی و ۲۴ هفتگی نشان داده شد. تا به امروز مطالعات پویش ژنومی متعددی در ارتباط با صفات رشد گزارش شده است (آل-ممن و همکاران ۲۰۱۵؛ ماتیکا و همکاران ۲۰۱۶ و بولورما و همکاران ۲۰۱۶). اما طبق بررسی‌های انجام شده، تا به امروز در گوسفندان نژاد دنبه‌دار تحقیقاتی مبنی بر ارتباط ژنومی هاپلوتیپی با استفاده از نشانگرهای 50K با صفات رشد افزایش وزن روزانه در سنین مختلف گزارش نشده است. هدف این تحقیق تعیین میزان عدم تعادل پیوستگی و شناسایی بلوک‌های هاپلوتیپی با استفاده از داده‌های حاصل از SNPChip 50K

افزایش وزن روزانه یکی از مهمترین سنجه‌های تعیین کننده سود اقتصادی در پرورش گوسفند در ایران است. در راستای دستیابی به بیشترین بازده تولید در گوسفند، صفات رشد به عنوان ملاک انتخاب در پرورش گوسفند پیشنهاد شده است (محمدی و همکاران ۲۰۱۳ و بهلولی و همکاران ۲۰۱۳). یکی از راه‌های کمکی برای افزایش بهبود این صفات، علاوه بر انتخاب کلاسیک استفاده از اطلاعات ژنومیکی می‌باشد که امروزه از طریق فن آوری-های مختلف مولکولی مبتنی بر DNA، دستیابی به این اهداف امکان پذیر شده است. به طوری که در ژانویه ۲۰۰۹ طراحی اولین آرایه SNPChip 50K در گوسفند و استفاده موفقیت‌آمیز از آن در قالب پروژه Sheep HapMap گزارش شد (کیجاس و همکاران ۲۰۱۲).

عدم تعادل پیوستگی (LD) یا عدم تعادل فاز گامتی به ارتباط غیر تصادفی آلل‌های دو جایگاه بر روی یک کروموزوم در یک گامت گفته می‌شود (وال و پریچارد ۲۰۰۳). معیار تعیین کننده موفقیت استفاده از مطالعات پویش ژنومی (GWAS)<sup>۲</sup> و انتخاب ژنومیک (GS)<sup>۳</sup> به مقدار LD بین نشانگرها و جایگاه‌های ژنومی مسبب صفات کمی در طول کل ژنوم بستگی دارد (ژائو و همکاران ۲۰۱۴). اساس مطالعات ژنومی بر مبنای مقدار LD در جمعیت مورد مطالعه و پایداری فاز عدم تعادل پیوستگی بین نشانگر و جایگاه ژنومی صفات کمی (QTL)<sup>۴</sup> در طی نسل‌ها است (ژائو و همکاران ۲۰۱۴). بلوک‌های هاپلوتیپی گروهی از آلل‌ها در بخشی از توالی DNA می‌باشند که دارای پیوستگی بسیار بالایی بوده و به احتمال زیاد با همدیگر به ارث می‌رسند (گابریل و همکاران ۲۰۰۲). علاوه بر این که آگاهی از ساختار بلوک-های هاپلوتیپی می‌تواند اطلاعات مهمی از مطالعات پویش ژنومی و انتخاب ژنومی ارائه دهد به عبارت دیگر به هر

۳. Genomic selection

۴. Quantitative trait loci

۱. Linkage disequilibrium

۲. Genome wide association study

های Illumina OvineSNP50K BeadChip انجام شدند. این آرایه‌ها امکان تعیین ژنوتیپ همزمان ۵۴۲۴۱ جایگاه نشانگری با میانگین فاصله بین نشانگرها ۵۰/۹ کیلوباز را فراهم می‌کنند (Illumina OvineSNP50 BeadChip) (Datasheet).

### مراحل کنترل کیفیت

برای تصحیح نمودن داده‌های تعیین ژنوتیپ شده از معیارهای فراوانی نرخ ژنوتیپ شده در نمونه‌ها، نرخ ژنوتیپ شده در نشانگرها در هر نمونه و حداقل فراوانی آلی (MAF) استفاده شد. در ابتدا نمونه‌هایی که فراوانی نرخ تعیین ژنوتیپ آنها کمتر از ۹۰٪ بود، شناسایی و حذف شد. این نمونه‌ها احتمال دارد که با داده‌های گمشده همراه بوده و خطای تعیین ژنوتیپ در آنها بالا باشد. در مرحله بعد نشانگرهایی که حداقل فراوانی آلی در آنها کمتر از ۵٪ بود حذف شدند. زمانیکه فراوانی آلی نشانگرها کمتر از ۵٪ باشد سبب کمتر برآورد شدن<sup>۲</sup> آماره  $r^2$  برای میزان LD بین جفت نشانگرها خواهد شد (قنبری و همکاران ۲۰۱۰؛ اسپیگلان و همکاران ۲۰۱۳). سپس نشانگرهایی که نرخ ژنوتیپ شده آن‌ها در نمونه‌ها کمتر از ۹۵٪ بود شناسایی و حذف شدند. در نهایت برای SNP‌های باقیمانده، آن‌هایی که در تعادل هاردی-واینبرگ نبودند به عنوان معیاری از خطای تعیین ژنوتیپ (تئو و همکاران ۲۰۰۷) کنار گذاشته شدند. سطح احتمال برابر با  $10^{-6}$  در نظر گرفته شد که با استفاده از تصحیح بونفرونی به دست آمد. مراحل مختلف کنترل کیفیت با استفاده از نرم افزار PLINK (پورسل و همکاران ۲۰۰۷) انجام شد.

### اندازه‌گیری عدم تعادل پیوستگی

برای محاسبه مقدار LD در بین نشانگرها از آماره  $r^2$  استفاده شد (ویلیام ۱۹۷۴):

و همچنین مطالعه ارتباط ژنومی هاپلویتی برای شناسایی مناطق ژنومی مؤثر روی صفات رشد شامل افزایش وزن روزانه قبل و بعد از شیرگیری گوسفند نژاد زندی بود.

### مواد و روش‌ها

از ۲۰۰ رأس گوسفند نژاد زندی حاضر در مرکز اصلاح نژاد خجیر وابسته به سازمان جهاد کشاورزی استان تهران خون گیری شد. از میان نمونه‌های اخذ شده پس از آنالیز شجره، حیواناتی که کمترین رابطه خویشاوندی را با هم داشته و حداکثر میزان تنوع در صفات رشد را داشتند به تعداد ۹۶ رأس انتخاب شدند. بدین منظور ابتدا مؤلفه‌های واریانس و کوواریانس و پارامترهای ژنتیکی به وسیله برنامه ASREML (گیلمور و همکاران ۱۹۹۹) برآورد گردید. بعد از تجزیه و تحلیل داده‌ها و پیش‌بینی ارزش‌های اصلاحی حیوانات با استفاده از مدل حیوانی تک متغیره، دام‌هایی که در دو انتهای توزیع ارزش‌های اصلاحی بالا و پایین قرار داشتند انتخاب شدند. آمار توصیفی صفات مورد بررسی در جدول ۱ نشان داده شده است. افزایش وزن روزانه قبل و بعد از شیرگیری به ترتیب برای ۱۱۰ و ۳۶۵ روزگی تصحیح شدند. مقدار ۵-۷ سی سی خون از رگ وداجی همراه با ۰/۵ میلی لیتر EDTA در لوله‌های خلاء با pH ۵-۷/۸ گرفته شد. ۹۶ نمونه خون گرفته شده بلافاصله بعد از شماره گذاری و با حفظ شرایط سرد به آزمایشگاه منتقل و تا زمان استخراج DNA در دمای ۴ درجه سانتی گراد نگهداری شد.

استخراج DNA با استفاده از روش بهینه یافته استخراج نمکی از خون انجام شد. پس از اطمینان از کمیت و کیفیت نمونه‌ها از طریق الکتروفورز بر روی ژل آگارز و دستگاه نانودراپ، غلظت آنها تا ۵۰ ng/μl جهت تعیین ژنوتیپ رقیق شد. سپس ژنوتیپ نمونه‌ها در کمپانی Neogen (<http://genomics.neogen.com>) با استفاده از آرایه

یک جفت نوترکیب در نظر گرفته می‌شود و حالت‌هایی جز این دو حالت، نامعلوم در نظر گرفته می‌شوند. در هر منطقه از ژنوم، بزرگترین ناحیه‌ای که در آن بیش از ۹۵٪ از جفت نشانگر تک نوکلئوتیدی با همبستگی زیاد باشند (حداقل ۳ نشانگر) یک بلوک هاپلوتیپی را معین می‌کند (گابریل و همکاران ۲۰۰۲).

### بررسی ارتباط هاپلوتیپی با صفات افزایش وزن روزانه

برای بررسی وجود اثرات ثابت معنی‌دار با استفاده از رویه GLM در نرم افزار R 3.4.0 عوامل ثابت سال تولد، سن مادر و تیپ تولد حیوان بررسی شدند. این اثرات ثابت در صورت معنی‌دار بودن به عنوان اثرات ثابت در فایلی با پسوند کووریت در تجزیه و تحلیل‌های پویش ژنومی در نظر گرفته شدند. در این تحقیق برای بررسی ارتباط هاپلوتیپ‌ها با صفات افزایش وزن روزانه قبل و بعد از شیرگیری از مدل ژنتیکی زیر استفاده شد:

$$y \sim \text{Haps} + \text{year} + \text{birth type}$$

با بررسی اثرات ثابت و با در نظر گرفتن عوامل سال تولد (year) و تیپ تولد به عنوان عوامل ثابت معنی‌دار (birth type) و اثر هاپلوتیپ (Haps)، مدل رگرسیون و با استفاده از نرم افزار PLINK (پورسل و همکاران ۲۰۰۷) انجام گرفت. برای کنترل نرخ خطای نوع اول از آزمون تصحیح بنفرونی استفاده شد که به صورت تصحیح خطای نوع اول و با تقسیم کردن احتمال خطای نوع اول (۰/۰۵) بر تعداد آزمون‌های مستقل (تعداد نشانگرهای بعد از کنترل کیفیت) صورت می‌گیرد. بنابراین آستانه معنی‌داری تعیین شده در آنالیزهای پویش کل ژنومی در این تحقیق برابر با  $10^{-6} \times 1/2 = 5 \times 10^{-7}$  (۰/۰۵) بود. همچنین در این تحقیق برای بررسی وجود یا عدم وجود لایه‌بندی جمعیتی با استفاده از فاکتور تورم کنترل جمعیتی ( $\lambda$ ) و GC در نرم افزار PLINK و ترسیم پلات-های Q-Q در نرم افزار SNPEVWin\_3.2 (وانگ و همکاران ۲۰۱۲) بررسی شد.

$$r^2 = \frac{D^2}{P_{A1}P_{A2}P_{B1}P_{B2}}$$

که در آن  $D = P_{11}P_{22} - P_{12}P_{21}$  و  $P_{11}$ ،  $P_{12}$ ،  $P_{21}$  و  $P_{22}$  فراوانی هاپلوتیپ‌ها به ترتیب برای  $A_1B_1$ ،  $A_1B_2$ ،  $A_2B_1$  و  $A_2B_2$  می‌باشد. همچنین  $P_{A1}$ ،  $P_{A2}$ ،  $P_{B1}$  و  $P_{B2}$  فراوانی‌های هر یک از آلل‌ها در دو جایگاه A و B است. آماره D در معادله فوق به تنهایی یک روش محاسبه میزان LD است که به شدت تحت تأثیر فراوانی تک تک آلل‌ها می‌باشد (گارسیا-گامز و همکاران ۲۰۱۲)، بنابراین میزان LD برای هر کروموزوم اتوزومی به صورت جداگانه و در فواصل 0-10 kb، 10-25 kb، 25-50 kb، 50-100 kb، 100-500 kb، 0.5-1 Mb، 1-5 Mb و 5-10 Mb با استفاده از نرم افزار PLINK (پورسل و همکاران ۲۰۰۷) بر اساس آماره  $r^2$  محاسبه شد.

### ساختار بلوک هاپلوتیپی

برای تخمین هاپلوتیپ‌ها از نرم افزار fastPHASE نسخه ۱/۲ (اسچیت و استیفن ۲۰۰۶) استفاده شد. در این برنامه از روش بیزی براساس مدل مخفی مارکوف (HMM) برای فازبندی هاپلوتیپ‌ها استفاده می‌شود. برای اجرا، فایل ورودی شامل داده ژنوتیپی SNP‌ها برای هر کروموزوم در تمام حیوانات است. بلوک‌های هاپلوتیپی با استفاده از الگوریتم ارائه شده توسط گابریل و همکاران (۲۰۰۲) به وسیله برنامه Haploview نسخه ۴/۲ (بارت و همکاران ۲۰۰۵) برای هر یک از کروموزوم‌های اتوزومی شناسایی شدند. مقدار LD بین جفت نشانگر تک نوکلئوتیدی در روش گابریل، به صورت یک شاخص سه وضعیتی خلاصه سازی می‌شود به طوری که اگر کران‌های بالا و پائین برآورد فاصله‌ای  $D'$   $\left(\frac{|D|}{D_{Max}}\right)$  به ترتیب بیشتر از ۰/۹۸ و ۰/۷ باشند جفت نشانگر تک نوکلئوتیدی متناظر به عنوان جفت نشانگر تک نوکلئوتیدی با همبستگی زیاد و اگر کران بالای همین برآورد کمتر از ۰/۹ باشد جفت نشانگر تک نوکلئوتیدی متناظر به عنوان

پایگاه‌های اطلاعاتی آنلاین GeneCards  
 UniProtKB و (<http://www.genecards.org>)  
 برای بررسی (<http://www.uniprot.org>) استفاده شد. برای بررسی همپوشانی مناطق مورد نظر با QTL‌های شناسایی شده مرکز اطلاعاتی آنلاین QTL (<http://i.animalgenome.org/jbrowse/>) مورد استفاده قرار گرفت.

برای تعیین موقعیت ژنومی هاپلوتیپ‌ها در سطح ژنوم گوسفند از آخرین نسخه ژنومی گوسفند (assembly NCBI Oar\_v4.0) پایگاه اطلاعاتی NCBI (<https://www.ncbi.nlm.nih.gov/genome/83>) و همچنین به منظور شناسایی ژن‌های مرتبط با بخش‌های معنی‌دار هاپلوتیپی ژنوم، از اطلاعات پایگاه اطلاعاتی NCBI ([https://www.ncbi.nlm.nih.gov/genome?term=ovis%20\(aries\)](https://www.ncbi.nlm.nih.gov/genome?term=ovis%20(aries))) و برای تفسیر عملکرد ژن‌های به دست آمده از

### جدول ۱- آمار توصیفی صفات رشد مورد بررسی در گوسفند نژاد زندی

Table 1- Descriptive statistics of studied growth traits in Zandi sheep breed

صفات Traits	حداکثر Maximum	حداقل Minimum	انحراف استاندارد Standard deviation	میانگین Mean
افزایش وزن روزانه قبل از شیرگیری (kg) Pre-weaning daily gain	0.34	0.08	0.04	0.197
افزایش وزن روزانه پس از شیرگیری (kg) Post-weaning daily gain	0.21	0.07	0.03	0.126

تعیین ژنوتیپ اطمینان حاصل شد (۹۹/۹٪). از مجموع ۵۴۲۴۱ نشانگر به کار رفته در این تحقیق، ۴۰۸۷۹ نشانگر توانستند مراحل مختلف کنترل کیفیت را بگذرانند. به طور کلی ۷۷۱۷ نشانگر به دلیل حداقل فراوانی آلی کمتر از ۵٪، ۱۰۷۰ نشانگر به دلیل نرخ تعیین ژنوتیپ کمتر از ۹۵٪ در هر نمونه، ۱۱۷ نشانگر به دلیل عدم تعادل هاردی واینبرگ، ۱۵۴۲ نشانگر روی کروموزوم X و ۲۹۱۶ نشانگر با موقعیت ناشناخته و همچنین ۲ نمونه به دلیل فراوانی نرخ تعیین ژنوتیپ کمتر از ۹۰٪ حذف شدند. بیشترین تعداد SNP‌ها بر روی کروموزوم شماره یک و کمترین روی کروموزوم ۲۴ قرار داشتند. میانگین فاصله بین دو SNP، ۵۹ کیلوباز و حداقل و حداکثر فاصله بین دو SNP به ترتیب ۵/۱۴ کیلوباز در داخل کروموزوم ۶ و ۲/۸۹ مگاباز در داخل کروموزوم ۲۲ بود.

#### عدم تعادل پیوستگی و ساختار بلوک هاپلوتیپی

نتایج آنالیز LD در بین جفت نشانگرها به تفکیک هر کروموزوم تا فاصله ۱۰ مگاباز بین جفت نشانگرها در

#### نتایج و بحث

##### بررسی کنترل کیفیت داده‌های حاصل از تعیین ژنوتیپ

ابتدا جنس بره‌های ثبت شده در مقابل جنس برآورد شده مورد آزمون قرار گرفت که نتایج این بخش صحت کامل تعیین ژنوتیپ نمونه‌ها و جنس بره‌های ثبت شده را نشان داد. در این مرحله جنسیت هر حیوان با توجه به میزان هتروزیگوسیتی SNP‌های موجود در کروموزوم X برآورد می‌شود و نتایج به دست آمده با جنسیت ثبت شده برای آنها مقایسه می‌شود. همچنین میزان تکرار پذیری در ۳ نمونه مورد بررسی قرار گرفت. این سه نمونه شامل حیواناتی بودند که در طی تحقیقاتی در گذشته نیز تعیین ژنوتیپ شده بودند و نتایج ژنوتیپ آنها در دسترس بود. در این مرحله تکرارپذیری نتایج حاصل از تعیین ژنوتیپ این نمونه‌ها در طی دو آزمایش مستقل مورد بررسی قرار می‌گیرد. نتایج این نمونه‌ها با نتایج قبلی همخوانی کامل داشته و به این ترتیب از صحت بالای

طور متوسط مقدار LD برابر با ۰/۳۴ بوده و بیشترین و کمترین میزان LD به ترتیب روی کروموزوم‌های ۲۵ و ۱۷ مشاهده شد.

جدول ۲ نشان داده شده است. با افزایش فاصله جفت نشانگرها مقدار LD کاهش و مقدار LD حدود ۰/۱۸ در کمتر از 50 Kb به دست آمد. در فاصله کمتر از 10 Kb به

### جدول ۲- آمار توصیفی میانگین $r^2$ بین جفت نشانگرها تا فاصله ۱۰ مگاباز در ژنوم

Table 2- Statistical information for average  $r^2$  as distance between pairs of SNP up to 10 Mb in the genome

کروموزوم Chromosome	کروموزوم							
	5-10 Mb	1-5 Mb	0.5-1 Mb	100-500 Kb	50-100 Kb	25-50 Kb	10-25 Kb	0-10 Kb
1	0.031	0.061	0.079	0.089	0.105	0.186	0.212	0.347
2	0.021	0.061	0.070	0.083	0.114	0.175	0.283	0.354
3	0.031	0.061	0.071	0.093	0.114	0.184	0.256	0.338
4	0.020	0.051	0.079	0.083	0.112	0.193	0.269	0.352
5	0.030	0.051	0.050	0.083	0.124	0.178	0.245	0.353
6	0.031	0.051	0.065	0.093	0.122	0.192	0.284	0.395
7	0.031	0.051	0.060	0.072	0.108	0.163	0.263	0.339
8	0.030	0.061	0.079	0.092	0.157	0.197	0.290	0.401
9	0.010	0.051	0.068	0.082	0.104	0.214	0.282	0.355
10	0.022	0.047	0.072	0.073	0.111	0.203	0.282	0.343
11	0.029	0.044	0.059	0.072	0.132	0.156	0.197	0.371
12	0.020	0.045	0.078	0.092	0.105	0.162	0.170	0.332
13	0.030	0.051	0.080	0.093	0.117	0.176	0.185	0.367
14	0.030	0.051	0.079	0.092	0.109	0.167	0.202	0.352
15	0.020	0.049	0.079	0.092	0.104	0.158	0.181	0.350
16	0.020	0.049	0.070	0.082	0.106	0.191	0.254	0.371
17	0.020	0.050	0.072	0.089	0.159	0.186	0.195	0.256
18	0.020	0.049	0.077	0.082	0.093	0.132	0.144	0.291
19	0.030	0.049	0.079	0.082	0.123	0.164	0.172	0.383
20	0.030	0.050	0.070	0.087	0.190	0.192	0.242	0.319
21	0.030	0.050	0.072	0.087	0.140	0.161	0.199	0.348
22	0.032	0.049	0.072	0.087	0.130	0.159	0.164	0.287
23	0.032	0.050	0.078	0.082	0.155	0.192	0.211	0.303
24	0.032	0.043	0.077	0.083	0.133	0.156	0.191	0.284
25	0.042	0.068	0.088	0.099	0.162	0.242	0.257	0.417
26	0.011	0.056	0.078	0.083	0.119	0.139	0.155	0.276

(۲۰۱۴) و لیو و همکاران (۲۰۱۷) گزارش شده است. این تفاوت‌ها می‌تواند به نرخ متفاوت نوترکیبی درون و بین کروموزوم‌ها، هتروزیگوسیتی، دریافت ژنتیکی و اثر انتخاب صفات مهم اقتصادی مربوط شود (لیو و همکاران ۲۰۱۷). بعلاوه در داخل یک کروموزوم نیز نرخ نوترکیبی از ناحیه سانترومیک به ناحیه تلومریک افزایش می‌یابد. این موضوع می‌تواند منجر به تنوع در مقدار و وسعت عدم تعادل پیوستگی در نواحی مختلف ژنوم شود (آریاس و همکاران ۲۰۰۹).

در مطالعه‌ی حاضر، برای تمامی کروموزوم‌های اتوزومی، ساختار بلوک‌های هاپلوتیپی تا حداکثر فاصله

مطالعات نشان می‌دهد که بازدهی نقشه‌یابی LD به میزان LD در جمعیت مورد مطالعه، عدم تجانس LD در طول ژنوم، تراکم نشانگر و عدم تجانس آلی QTL بستگی خواهد داشت. جمعیتی با مقدار بالای LD به تراکم نشانگری کمتر نیازمند می‌باشد و بالعکس (میدوز و همکاران ۲۰۰۸). نتایج این تحقیق نشان داد که مقدار  $r^2$  حدود ۰/۲۰ در گوسفند نژاد زندگی کمتر از 40 Kb گسترش یافته است که با نتیجه به دست آمده در تحقیق قلی‌زاده و همکاران (۲۰۱۴) در گوسفند بلوچی ایرانی مطابقت دارد. مقدار LD در بین کروموزوم‌های مختلف در این تحقیق متفاوت بود. تفاوت در مقدار LD مابین کروموزوم‌ها در گوسفند توسط قلی‌زاده و همکاران

متوسط طول بلوک‌های هاپلوتیپی ۲۳/۰۸ کیلو باز بود. در مطالعه‌ای در گوسفندان چورا، تعداد ۲۰۹۹ بلوک هاپلوتیپی در کروموزوم‌های اتوزومی، که ۲/۳۲ درصد از ژنوم را پوشش می‌داد، شناسایی نمودند (گارسیا-گامس و همکاران ۲۰۱۲). در مطالعه‌ی این محققین، میانگین طول بلوک‌ها ۲۷/۰۳ کیلو باز با ۴۷۸۰ SNP که ۱۰ درصد کل SNP‌ها را شامل می‌شد، بود.

۵ مگاباز مورد مطالعه قرار گرفت. آنالیز ساختار بلوک-های هاپلوتیپی برای جمعیت گوسفند زندی در جدول ۳ ارائه شده است. تعداد ۱۴۷۲ بلوک با پوشش ۱/۴۵ درصد از ژنوم مشاهده شد. علاوه براین، تعداد ۳۴۶۴ SNP درون بلوک‌های هاپلوتیپی با متوسط ۳ SNP درون هر بلوک قرار گرفت که ۷/۵۳ درصد از کل SNP‌ها را تشکیل می‌داد. بیشترین و کمترین تعداد SNP درون بلوک، به ترتیب روی کروموزوم ۱ و ۲۲ با ۴۹۶ و ۲۲ مشاهده شد.

جدول ۳- ساختار بلوک هاپلوتیپی به ازای هر کروموزوم برای جمعیت گوسفند زندی  
Table 3- Block haplotype structure per each chromosome for Zandi sheep population

کروموزوم Chr	درصد در SNP‌ها در بلوک % of SNPs in blocks	تعداد SNP‌ها در بلوک Number of SNPs in Blocks	درصد طول کروموزوم در بلوک % of Chr length in blocks	میانگین طول بلوک Mean Block length (kb)	تعداد بلوک Number of blocks	تعداد SNP‌ها Number of SNPs	طول کروموزوم Chr length (Mb)
	9.60	496	2.75	36.5	187	5168	275.41
2	10.60	428	1.79	35.2	175	4036	248.97
3	12.28	392	2.29	39.8	168	3191	224.00
4	10.53	221	1.89	37.9	102	2098	119.22
5	10.17	181	1.39	35.4	95	1778	107.84
6	8.09	176	1.31	32.1	86	1957	116.89
7	8.34	154	1.37	39.5	85	1845	100.01
8	8.95	148	1.38	32.6	79	1653	90.62
9	8.92	138	1.37	32.3	79	1546	94.58
10	8.87	130	1.40	30.5	68	1461	86.38
11	9.70	98	1.46	27.2	51	991	62.17
12	6.68	86	1.34	22.1	41	1301	79.03
13	7.60	79	1.35	20.2	34	1270	82.95
14	7.46	76	1.42	22.7	32	991	62.57
15	5.47	70	1.31	20.4	29	1278	80.78
16	5.72	67	1.34	18.3	25	1170	71.69
17	5.23	63	1.32	19.8	22	1204	72.25
18	4.74	57	1.28	17.1	20	1201	68.49
19	5.55	53	1.36	13.6	18	954	60.45
20	6.15	52	1.38	12.3	15	845	51.05
21	5.11	43	1.33	11.5	13	841	49.99
22	7.16	67	1.26	10.8	12	935	50.78
23	6.09	60	1.20	8.3	11	985	62.28
24	9.23	58	1.22	7.4	10	628	41.98
25	5.09	22	1.14	8.7	8	824	45.22
26	3.98	29	1.13	6.4	7	728	44.05
<b>کل</b>	<b>7.58</b>	<b>3464</b>	<b>1.45</b>	<b>23.08</b>	<b>1472</b>	<b>40879</b>	<b>2449.65</b>

معنی‌دار و استفاده همزمان هر دو عامل به عنوان عوامل غیر ژنتیکی با عنوان فایل کواریت در آنالیز محاسباتی نرم افزار Plink استفاده شد. در این حالت، ارتباط معنی-دار هاپلوتیپ‌ها با صفت افزایش وزن روزانه قبل از شیرگیری روی کروموزوم‌های ۵، ۶ و ۷ و برای صفت افزایش وزن روزانه بعد از شیرگیری روی کروموزوم ۳ شناسایی شد (جدول ۴). اولین ناحیه هاپلوتیپی معنی‌دار روی کروموزوم شماره ۳ در ناحیه ۱۵۴ مگابازی شامل ۲۰ نشانگر SNP و حاوی ژن‌های MSR3، LEMD3 و WIF1 می‌باشد. ژن MSR3 در ناحیه‌ای از کروموزوم ۳ واقع شده که در بررسی نژادهای مختلف گوسفند به عنوان ناحیه‌ای که تحت تأثیر فشار بالای انتخاب بوده است، گزارش شده است (کیجاس و همکاران ۲۰۱۲). علاوه بر این، ژن کاندیدای WIF1 کد کننده پروتئین‌های بازدارنده مسیره‌های سیگنالی Wnt می‌باشد که تکثیر سلولی و تعیین سرنوشت سلولی را در طی تکوین جنینی و هموستازی بافتی هدف قرار می‌دهد. همچنین در فرآیند بیولوژیکی تنظیم مثبت تفرق سلول‌های چربی در ارتباط هستند (<http://www.genecards.org>).

شکل ۱ لایه بندی جمعیتی خیلی کمی را نشان می‌دهد که به ترتیب برابر با ۱/۰۳۹ و ۱/۰۷۳ برای صفات مورد بررسی بودند. به عبارت دیگر، هر گونه انحراف از خط نشان دهنده وجود ارتباط ژنتیکی بین افراد در صفت مربوطه می‌باشد. علاوه بر این، در این تحقیق جهت بررسی نحوه اثر لحاظ نمودن اثرات ثابت معنی‌دار در کنترل نتایج، مدل فاقد هر گونه اثرات ثابت منجر به شناسایی ۷ منطقه هاپلوتیپی روی ۵ کروموزوم شد. در حالی که با در نظر گرفتن نتایج اثر سال تولد و تیپ تولد (اثرات ثابت معنی‌دار) به عنوان عواملی که می‌توانند نتایج مثبت کاذب (خطای نوع اول) ایجاد کنند، این تعداد جایگاه به ۴ منطقه هاپلوتیپی روی ۴ کروموزوم کاهش یافت. استفاده از آزمون هاپلوتیپ‌ها می‌تواند سبب بهبود قدرت آزمون‌های ارتباط ژنومی شود، زیرا هاپلوتیپ‌ها نسبت به نشانگر انفرادی می‌توانند در LD بالاتری با QTL یا ژن‌های کاندیدا باشند. بنابراین، استفاده از آزمون هاپلوتیپ‌ها ممکن است سبب شناسایی نواحی ژنومی معنی‌دار مرتبط با صفت مورد بررسی باشد که در آزمون ارتباط تک نشانگری شناسایی نمی‌شوند. با استفاده از نتایج رویه GLM و شناسایی دو عامل ثابت

جدول ۴- ارتباط هاپلوتیپ‌ها با صفات رشد در گوسفند نژاد زندی

Table 4- Association between haplotypes with growth traits in Zandi sheep breed

کروموزوم- مگاباز Chr- Mb	ژن (های) کاندیدا Candidate gene(s)	صفت (های) مرتبط Related- trait(s)	اندازه پنجره (تعداد SNP) Window size (no.SNP)	۲-SNP SNP2	۱-SNP SNP1
3-154	MSRB3, LEMD3, WIF1	POWG	20	OAR3_166103396.1	OAR3_165009241.1
5-88	ARRDC3	PRWG	15	OAR5_97280023.1	OAR5_96279928.1
6-37	LCORL, NCAPG, DCAF16, FAM184B, TMC05A, SNORA70, SPRED1, FAM98B	PRWG	15	OAR6_42317192.1	OAR6_41424992.1
			20	OAR7_35045583.1	OAR7_33975956.1



ژن‌های NCAPG، LCORL و DACAF16 که روی کروموزوم ۶ واقع شده‌اند بر روی تقسیم سلولی، پایداری کروموزوم، تنظیم رونویسی و اندازه اسکلت بدنی نقش کلیدی دارند (<http://www.genecards.org>).

چهارمین ناحیه هاپلوتیپی معنی‌دار بر روی کروموزوم ۷ در ناحیه ۳۰ مگا بازی که شامل ۲۰ نشانگر بود، تشخیص داده شد. از مهمترین ژن در این ناحیه، ژن SPRED1 است که جزء خانواده ژنی پروتئین‌های SPRED بوده و در پاسخ به فاکتورهای مختلف رشد از مسیرهای سیگنالینگ گیرنده‌های فیبروبلاستی فاکتورهای رشد و سیتوکین‌ها نقش دارند (<http://www.genecards.org>).

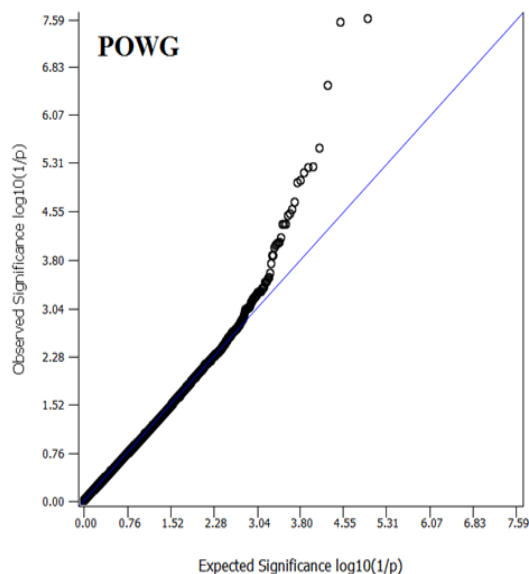
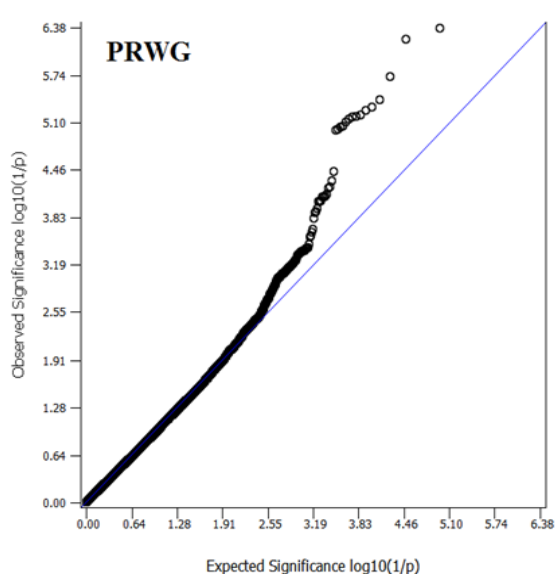
دومین ناحیه هاپلوتیپی روی کروموزوم ۵ در ناحیه ۸۸ مگابازی شامل ۱۵ نشانگر و حاوی ژن ARRDC3 می‌باشد. ژن ARRDC3 جزو خانواده بزرگ ژنی آرستین‌ها می‌باشد که دارای نقش سیگنالینگ و کنترل متابولیسم می‌باشند (<http://www.genecards.org>). ارتباط قوی و معنی‌داری بین چند شکلی در ژن کاندیدای ARRDC3 با صفات رشد، وزن لاشه و کیفیت گوشت در نژادهای مختلف گاو گوشتی گزارش شده است (بولورما و همکاران ۲۰۱۴).

سومین ناحیه هاپلوتیپی روی کروموزوم ۶ در ناحیه ۳۷ مگابازی که شامل ۱۵ نشانگر بود، تشخیص داده شد.

جدول ۵- برآورد اثرات هاپلوتیپی‌های معنی‌دار بر صفات مرتبط با رشد

Table 5- Estimate effects of significant haplotypes for the traits related to growth

هاپلوتیپ Haplotype	درصد واریانس توجیه شده به وسیله هاپلوتیپ percentage of variance explained by Haplotype	P-value	صفت مرتبط Related-trait(s)
AGCCGCGGCATAAAGCCGG	1.52	5.923E-07	POWG
TTTACCCACATATT	1.68	4.21E-07	PRWG
AAAGGGGCCAACGGA	1.31	3.896E-07	PRWG
TAATATAATAGCCTTATAAA	0.65	2.924E-08	PRWG



شکل ۱- پلات‌های Q-Q صفات مرتبط با افزایش وزن روزانه قبل (PRWG) و پس (POWG) از شیرگیری در نژاد گوسفند زندی برای ارزیابی لایه‌بندی جمعیتی

**Figure 1- Plots Q-Q for evaluation of stratification for PRWG (Pre-weaning daily gain) and POWG (Post-weaning daily gain) in Zandi sheep breed**

بررسی قرار گرفت و سطح مفید عدم تعادل پیوستگی در 40 Kb برابر با  $r^2 = 0.2$  مشاهده شد. نتایج مطالعات عدم تعادل پیوستگی اولین قدم در تعیین تراکم نشانگر و اندازه نمونه مناسب در مطالعات ژنومی می‌باشد. لذا تهیه نقشه LD امری اجتناب ناپذیر برای افزایش صحت می‌باشد. همچنین تعداد ۱۴۷۲ بلوک هاپلوتیپی که  $7/58$  درصد از کل SNPها را پوشش می‌داد شناسایی شد. ۴ ناحیه هاپلوتیپی جدید و معنی‌دار روی کروموزوم‌های ۳، ۵، ۶ و ۷ گوسفندی مرتبط با افزایش وزن روزانه قبل و بعد از شیرگیری شناسایی شد.

#### سپاسگزاری

از مرکز اصلاح نژاد و بهبود تولیدات دامی کشور به خاطر حمایت مالی بخشی از پروژه و ایستگاه پرورش و اصلاح نژاد گوسفند زندی استان تهران (ایستگاه خجیر) برای در اختیار گذاشتن حیوانات در اجرای این پروژه و همچنین از شرکت دانش بنیان سایننا گستر البرز به خاطر حمایت مالی عمده این طرح کمال تشکر را داریم.

پس از شناسایی هاپلوتیپ‌های معنی‌دار مرتبط با صفت افزایش وزن روزانه قبل و بعد از شیرگیری و تشکیل ماتریس خویشاوندی تمام جایگاه‌ها به طور همزمان در داخل معادله قرار گرفتند و میزان واریانس توجیه شده توسط هر یک از هاپلوتیپ‌های معنی‌دار محاسبه شد (جدول ۵). برای صفت افزایش وزن روزانه قبل از شیرگیری سه هاپلوتیپ معنی‌دار شناسایی شده بود که در مجموع  $3/43$  درصد از تنوع واریانس صفت مذکور را توجیه می‌نمود. همچنین در ارتباط با صفت افزایش وزن روزانه پس از شیرگیری تنها یک جایگاه هاپلوتیپی معنی‌دار شناسایی شده بود که  $1/52$  درصد از تنوع صفت را توجیه می‌نمود. با این حال در تفسیر این نتایج بایستی محتاط بود، چرا که تعداد ۹۴ حیوان برای آنالیزها استفاده شد، برای این چنین آنالیزهایی کم است. پیشنهاد می‌شود در تحقیقات آینده با تعداد افراد بیشتری میزان توجیه واریانس توسط هاپلوتیپ‌ها محاسبه شود.

#### نتیجه‌گیری نهایی

در این تحقیق، میزان و گستره عدم تعادل پیوستگی بر روی کروموزوم‌های مختلف اتوزومی گوسفندی مورد

#### منابع مورد استفاده

- Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R and Gondro C, 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genetics Selection Evaluation* 47:66.
- Arias JA, Keehan M, Fisher P, Coppieters W and Spelman R, 2009. A high density linkage map of the bovine genome. *BMC Genetics* 10: 18-30.
- Barrett JC, Fry B, Maller J and Daly MJ, 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.
- Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ and Goddard ME, 2014. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genetics* 10(3): e1004198.
- Bolormaa S, Hayes BJ, van der Werf JH, Pethick D, Goddard ME and Daetwyler HD, 2016. Detailed phenotyping identifies genes with pleiotropic effects on body composition. *BMC Genomics* 17:224.
- Bohlouli M, Mohammadi H and Alijani S, 2013. Genetic evaluation and genetic trend of growth traits of Zandi sheep in semi-arid Iran using random regression models. *Small Ruminant Research* 114, 195-201.

- Espigolan R, Baldi F, Boligon AA, Souza FRP, Gordo DGM, Tonussi RL, Cardoso DF, Oliveira HN, Tonhati H, Sargolzaei M, Schenkel FS, Carneiro R, Ferro JA and Albuquerque LG, 2013. Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics* 14:305.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J and Blumenstiel B, 2002. The structure of haplotype blocks in the human genome. *Science* 296: 2225-2229.
- Garcia-Gamez E, Sahana G, Gutierrez-Gil B and Arranz J, 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* 13:43.
- Gilmour A, Gogel B, Cullis and Thompson R, 1999. ASReml User Guide Release 2.0. VSN International Ltd, Hemphstead, UK.
- Gholizadeh M, Rahimi Mianji G and Nejati Javaremi A, 2014. Linkage disequilibrium estimation and haplotype based genome-wide association to detect QTLs affecting twinning rate in Baluchi sheep. *Research on Animal Production* 10: 166-178.
- Johnston SE, McEwan JC, Pickering NK, Kijas JW, Beraldi D, Pilkington JG, Pemberton JM and Slate J, 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* 20(12): 2555–2566.
- Kijas JW, Johannes A, Ben Hayes L, Boitard S, PortoNeto LR, Cristobal MS, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J and Dalrymple B, 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* 10: 1-14.
- Liu S, He S, Chen L, Li W, Di J and Liu M, 2017. Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (Xinjiang type) sheep by genome-wide SNPs. *Genes Genome* 17: 1-13.
- Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL and Bishop SC, 2016. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genetics Selection Evaluation* 48:11.
- Meadows JR, Chan EK and Kijas JW, 2008. Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* 30; 9:61.
- Mohammadi H, Moradi Shahrehabak M and Sadeghi M, 2013. Association between single nucleotide polymorphism in the ovine DGAT1 gene and carcass traits in two Iranian sheep breeds. *Animal Biotechnology* 24: 159–167.
- Mohammadi H, Rafat SA, Moradi shahrehabak H, Shodja J and Moradi MH, 2018. An assessment of population stratification and haplotype based Genome-wide association for wool quality traits in Zandi sheep breed. *Journal of Animal Science Researches (Agricultural science)* 28(2); 193-204.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC, 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *The American Journal of Human Genetics* 81: 559-575.
- Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR and Simianer H, 2010. The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 41: 346–356.
- Riggio V, Matika O, Pong-Wong R, Stear MJ and Bishop SC, 2013. Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity* 110:4 20–9.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF and Ackerman HC, 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419(6909): 832-837.
- Scheet P and Stephens M, 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal Human Genetics* 78(4): 629–644.
- Teo YY, Fry AE, Clark TG, Tai ES and Seielstad M, 2007. On the usage of HWE for identifying genotyping errors. *Annals of Human Genetics* 71: 701-703.
- Wall JD, Pritchard JK, 2003. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* 4: 587–597.

- Wang S., Dvorkin D and Da Y, 2012. SNPEVG: a graphical tool for GWAS graphing with mouse clicks. *BMC Bioinformatics* 13: 319-326.
- William GH, 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33: 229–239.
- Zhao FP, Wang GK, Zeng T, Wei CH, Zhang L, Wang HH, Zhang SZ, Liu RZ, Liu Z and Du LX, 2014. Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations. *Livestock Science* 170: 22–29.

## Study of haplotype blocks structure and haplotype based genome scan in autosomal chromosomes to body weight gain trait in Zandi sheep breed

H Mohammadi<sup>1</sup>, SA Rafat<sup>2\*</sup>, H Moradi Shahrebabak<sup>3</sup>, J Shodja<sup>3</sup> and MH Moradi<sup>4</sup>

Received: October 23, 2017 Accepted: December 10, 2017

<sup>1</sup>PhD Graduated, Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz, Tabriz, Iran

<sup>2</sup> Professors, Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz, Tabriz, Iran

<sup>3</sup>Assistant Professor, Department of Animal Sciences, University College of Agriculture and Natural Resources, University of Tehran, Tehran, Iran

<sup>4</sup>Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, University of Arak, Arak, Iran

\*Corresponding author: abbasrafat@hotmail.com

**Introduction:** Average daily gain (ADG) is the most economically important trait in sheep industry. In genome wide association study and genomic selection, determining of extent and level of linkage disequilibrium (LD) and haplotype block structure are critical in sample size and marker density. Haplotype blocks are defined as long stretches of SNPs along a chromosome that have low recombination rates, which characterized by relatively few haplotypes. Understanding haplotype structure in genome can greatly facilitate LD analysis. Haplotype-based association analysis can offer a powerful approach for mapping functional genes (Gabriel et al., 2002). Therefore, the objective of this research were to study LD pattern, determine haplotype block structure and genome wide haplotype association study in Zandi sheep for identifying the genomic region associated with pre-weaning (AGW) and post-weaning daily gain (PWG).

**Materials and method:** A total of 96 Iranian Zandi sheep were used in the study. The following two traits were analyzed: pre-weaning daily gain and post-weaning daily gain. Animals were genotyped using 50 K SNPChip panel. Quality control of the genotype data consisted in removing SNPs with a call rate less than 95 %, SNPs with a minor allele frequency (MAF) less than 5 %, individual with more than 10 % missing genotypes, and SNPs that deviated strongly from Hardy–Weinberg equilibrium ( $P < 10^{-6}$ ). LD between all pairs were calculated with  $r^2$  by PLINK v1.07. Haplotype blocks were identified on base algorithm Gabriel et al. (2002) for all autosomes, using Haploview software (Barrett et al. 2005), based on estimates of  $D'$  for all pairwise combinations of SNPs within each chromosome. Following Gabriel et al. (2002), a pair of SNPs is defined to be in “strong LD” if the upper 95% confidence bound of  $D'$  is  $> 0.98$  (consistent with no historical recombination) and the lower bound is  $> 0.7$ . Using the Haploview default values for blocks (Gabriel et al. 2002), a haplotype block is defined as a region over which 95 % of informative SNP pairs show “strong LD”. The PLINK was used to generate the matrix using the GLM algorithm. In this analysis, because of the previous selection history of the flock, it was important to identify and correct for population stratification. To evaluate whether estimates were overinflated, we used the genomic inflation factor  $\lambda$  using the PLINK software (Purcell et al., 2007). We also assessed their deviation from the expected distribution of no SNPs being associated with the trait of interest using a quantile-quantile (Q-Q) plot, which is commonly used to analyze population stratification in GWAS. We use SNPEVG tool to show the (Q-Q) plot. The Bonferroni method was used to adjust for multiple testing from the number of SNP loci detected. We declared a significant SNP at the genome-wide significance level if the raw P-value was,  $0.05/N$ , here N is the number of SNP loci tested in the analyses. Therefore, for each trait, the threshold P-value for declaring genome-wide significance was  $(0.05)/40,879=1.2 \times 10^{-6}$ . The exact positions of the annotated genes were extracted from the latest sheep genome *Oar\_v4.0* assembly along with the NCBI annotation release 102 of the sheep genome. To investigate whether the

significant SNPs detected in this study were within the range of previously identified QTL for relevant traits, we searched for meat or production QTL in the Animal QTLdb within a 1-Mb region on both sides of each significant haplotype.

**Results and discussion:** After quality control, 2 individuals were excluded, leaving 94 sheep for the association analysis. Additionally, we removed 1070 SNPs with call rates less than 95% and 7717 SNPs with MAF less than 0.05. A total of 40,879 SNPs passed these quality-control filters and were retained in the dataset. These SNPs were distributed across 26 autosomes, with the number of SNPs per chromosome ranging from 747 to 5694, and with a mean distance between adjacent SNPs ranging from 50.4 to 68.7 kb. Also, in this study, the extent of LD was 40 kb with  $r^2=0.2$ . Overall, 1472 blocks were observed in the 7.58% of all SNPs were classified into haplotype blocks, covering 1.45% of the total autosomal genome size. The results showed a reduction in LD level with the increase in distance between markers. The average pre-weaning daily gain was  $0.197\pm 0.04$  kg with an individual sheep range of 0.08-0.34 kg. Average post-weaning daily gain was  $0.126\pm 0.07$  kg with an individual sheep range of 0.03-0.41 kg. The result from genomic control showed weak population stratification for AGW and PGW in between population of Zandi sheep. The genomic inflation factors ( $\lambda_{gc}$ ) for the two traits were equal to 1.039 and 1.073, for AGW and PGW, respectively. However, as the Q-Q plots clearly show, there is no evidence of any systematic bias ( $\lambda_{gc}<1.1$ ) due to population structure or analytical approach in our case. Considering the significant fixed effects in the genomic wide association analysis, four haplotypes on chromosomes 3, 5, 6 and 7 identified to affect significantly AGW and PWG traits.

**Conclusion:** The results of this study could provide a suite of novel SNP markers and candidate genes associated with growth traits and hence, may play an important role for understanding the biology of average daily gain in sheep.

**Keywords:** Average daily gain, Candidate gene, Genome wide association study, Linkage disequilibrium, haplotype block.