مجله بیوتکنولوژی کشاورزی

علمی-پژوهشی و ISC

بررسی ساختار و لایه بندی جمعیت گاومیش‌های اکوتیپ آذری و شمالی با نشانگرهای متراکم چند شکل تک نوکلئوتیدی با استفاده از روش‌های Admixture، GC، PCA و MDS

زهرا عزیزی[۱]، عباس رأفت[۲]، جلیل شجاع[۳]، حسین مرادی شهربابک*[٤]، محمد مرادی شهربابک[٥]

[۱]دانشجوی دکتری،گروه علوم دامی،دانشکده کشاورزی، دانشگاه تبریز

[۲]دانشیار، گروه علوم دامی،دانشکده کشاورزی، دانشگاه تبریز

[۳]استاد، گروه علوم دامی،دانشکده کشاورزی، دانشگاه تبریز

[٤]استادیار، گروه علوم دامی،پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

[٥]استاد،، گروه علوم دامی،پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

تاریخ دریافت: ۱۳۹٤/۰۹/۳۰، تاریخ پذیرش: ۱۳۹۵/۰۲/۰٤

چکیده

در کاربردهای ژنتیک جمعیت، اختصاص افراد به جمعیت‌های مربوط به خود اهمیت دارد. با توسعه تکنولوژی تعیین ژنوتیپ در مقیاس وسیع بسیاری از نشانگرها از جمله اسنیپ ها برای این مطالعات قابل دسترس شده‌اند که این اسنیپ‌ها در مطالعه تنوع ژنتیکی دام‌های اهلی و ساختار جمعیت سودمند هستند. هدف این تحقیق بررسی ساختار و لایه‌بندی گاومیش‌های مناطق مختلف دو اکوتیپ آذری و شمالی با استفاده از داده‌های SNPChip 90 با روش‌های معمول بررسی ساختار جمعیت بود که برای این منظور تعداد ۲۵۸ گاومیش از استان‌های آذربایجان شرقی، آذربایجان غربی و اردبیل مربوط به اکوتیپ آذری و از استان گیلان مربوط به اکوتیپ شمالی نمونه‌گیری و تعیین ژنوتیپ شدند. نتایج حاصل از کنترل ژنومیک لایه بندی ضعیفی با $\lambda=1.056$ نشان داد که حاکی از وجود اختلاط (ساختار ضعیفی) در بین دو اکوتیپ است. پلات-های حاصل از تجزیه مولفه‌های اصلی و مقیاس بندی چند بعدی، تفکیک این دو اکوتیپ و استان‌های مختلف دو اکوتیپ را براساس فواصل انجام داد. روش Admixture نیز نزدیکی فاصله ژنتیکی افراد استان-های مختلف دو اکوتیپ را نشان داد که البته افراد خالصی هم در این بین وجود داشتند و k=3 خطای اعتبارسنجی پایینی داشت. این روش‌ها قادر به جداسازی کلی حیوانات به توده‌های مربوطه بودند و نتایج این تحقیق گویای ارتباط ژنتیکی نزدیک افراد چهار استان مختلف از دو اکوتیپ آذری و شمالی بود.

واژه‌های کلیدی: *لایه بندی جمعیت، تراشه اسنیپ 90K، گاومیش، MDS، PCA*

* نویسنده مسئول: حسین مرادی شهربابک تلفن: ۰۹۱۳۳۹۱۵۳۰٦ Email: hmoradis@ut.ac.ir

مقدمه

حیوانات و گیاهان بومی به عنوان سرمایه ملی و ذخایر استراتژیک هر کشور محسوب می شوند و حفظ و تکثیرآنها از ارزش و اهمیت بسیاری برخوردار است. این موجودات پس از هزاران سال انتخاب طبیعی و مصنوعی و نیز گذر از موانع بسیار و با غلبه بر تمامی شرایط نامساعد محیطی همچنان به حیات خویش ادامه داده و به تکثیر و ازدیاد نسل پرداخته‌اند همچنین نسبت به بسیاری از محدودیت‌های محیطی سازگاری پیدا کرده اند. این مسئله، بخصوص با افزایش تولید محصولات دامی و تولید محصولات پیش‌بینی‌نشده در آینده، لزوم حفظ تنوع ژنتیکی در دام‌های بومی را الزامی ساخته است چرا که یک گونه بدون تنوع ژنتیکی کافی قادر به سازگاری با تغییرات محیطی و مبارزه با انگل‌ها نیست (Askari *et al*., 2011). همچنین مطالعه تنوع ژنتیکی نژادهای بومی برای حفاظت از منابع ژنتیکی ذخایر بومی لازم و ضروری است (Mohammadi *et al*., 2009).

حفاظت باید بر اساس دانش عمیقی از منابع ژنتیکی نژادهای خاص باشد، لذا تلاش برای شناسایی و تعیین خصوصیات ژنتیکی نژادهای بومی و محلی بسیار اهمیت دارد (Zamani *et al*., 2013; Shojaei *et al*., 2011). استفاده از نشانگرهای مولکولی در سال‌های اخیر جهت تعیین تنوع ژنتیکی بین جمعیت‌ها و حیوانات حفاظت‌شده، کاربرد گسترده‌ای یافته است. میزان چندشکلی به‌دست‌آمده از این نشانگرهای ژنتیکی، یکی از پارامترهای قابل ارزیابی برای مطالعه جمعیت‌های مختلف و درک تفاوت‌های ژنتیکی بین جمعیت‌هاست. تکنیک-های ملکولی جدید، به همراه پیشرفت‌های ژنتیک آماری افق جالب و جذابی را برای انجام تحقیقات در زمینه نقشه‌یابی QTL و تنوع صفات مورد نظر گشوده‌اند (Abadi *et al*., 2009). گاومیش‌های ایران به دلیل سازگاری با محیط، مقاومت در برابر بیماری‌ها، هزینه‌های نگهداری پایین و استفاده از ضایعات کشاورزی و مواد خشبی کم ارزش، یکی از ذخایر ژنتیکی با ارزش محسوب می‌شوند. گاومیش ایران براساس شرایط آب و هوایی به سه دسته اصلی تقسیم بندی می-شود: اکوتیپ آذری (آذربایجان غربی و شرقی)، اکوتیپ شمالی (گیلان و مازندران) و اکوتیپ خوزستانی (خوزستان). در حوزه ژنتیک و اصلاح نژاد دام، اطلاع از ساختار ژنتیکی جمعیت در راستای اجرای بهتر برنامه‌های اصلاح نژادی و از همه مهم‌تر، حفظ ذخیره ژنتیکی بسیار ارزشمند است. از سویی دیگر ساختارهای زیر جمعیتی درون جمعیت های مورد مطالعه باعث ایجاد اریب در مطالعات GWAS می شود (Thomas Wacholder *et al*., 2002; and Witte, 2002) و لایه‌بندی جمعیتی (مخلوطی از افراد از پس زمینه‌های ژنتیکی متفاوت) به دلیل ایجاد اشتباه نوع اول، چالشی برای مطالعات GWAS است، چرا که در مطالعات GWAS فرض بر همگنی جامعه است که این فرض می‌تواند به آسانی

نقض شود (Marchini *et al.*, 2004). لایه بندی جمعیتی ناشی از تفاوت در فراوانی اللی زیر جمعیت ها به دلیل تفاوت ژنتیکی جد مشترک است بخصوص در مطالعات ارتباطی که باعث ایجاد نتایج اریب مرتبط با صفت مورد نظر می-گردد (Price *et al.*, 2010). لایه بندی جمعیتی در این زمینه اشاره به ساختار جمعیت دارد. نسل-های جدید فن‌آوری توالی یابی، باعث ایجاد مقادیر بی سابقه‌ای از داده ها برای جوامع دامی در حوزه ژنتیک شده است و داده‌های ژنومی فرصتی برای حل پیچیدگی تاریخچه تکاملی جمعیت‌ها و بازسازی حتی وقایع تاریخی نادر، فراهم می‌آورند. نتیجه منشا تاریخی پیچیده در ارتباط با انتخاب طبیعی و مصنوعی منجر به اختلاف متعدد نژادهای مختلف که تنوع فنوتیپی گسترده در طی یک دوره کوتاه زمانی نشان می-دهند، شده است (Epps *et al.*, 2013). استنتاج ساختار جمعیت از نشانگرهای ژنتیکی، در شرایط گوناگون مثل مطالعات ارتباطی و تکاملی، دسته-بندی زیرگونه‌ها و تعیین موانع ژنتیکی مفید می باشد. روش‌های متعددی برای تعیین ساختار ژنتیکی و لایه‌بندی جمعیت وجود دارد. آنالیز مولفه‌های اصلی (Price Patterson *et al.*, 2006 *et al.*, 2006;) و مقیاس بندی چند بعدی (Purcell *et al.*, 2007) روش‌هایی هستند که قادر به تعیین ساختار جمعیت می‌باشند. هدف از آنالیز MDS، کشف ساختار در داده‌ها و ابعاد معنی دار مرتبط است که شمایی تصویری از عدم تشابه (تشابه) در بین عناصر را می‌دهد. این روش ابعادی که فاصله ژنتیکی مشاهده شده را توضیح می‌دهد براساس روش Identity by state شناسایی می‌نماید. تجزیه مؤلفه‌های اصلی جزئی از تحلیل‌های عاملی است که به عنوان یک روش بسیار مفید برای تصحیح لایه بندی جمعیتی در مطالعات GWAS کاربرد دارد (Liu *et al.*, 2013). این روش علاوه بر نشانگرهای تک نوکلئوتیدی و ریز ماهواره، بر فراوانی‌های هاپلوتایپی نیز اعمال می‌شود. آنالیز مولفه اصلی ابزار استانداری در ژنتیک جمعیت است که برای کشف ساختار جمعیت کاربرد دارد و می‌تواند برای داده‌های با حجم زیاد استفاده شود، برخلاف STRUCTURE که برای داده‌های با حجم زیاد غیر عملی است. روش PCA آزمونی را برای وجود ساختار جمعیتی در داده های ژنتیکی فراهم می‌آورد (Patterson *et al.*, 2006) و در مطالعه‌ی جمعیت‌های اروپائی و هندی استفاده شده است (Lao *et al.*, 2008). همچنین در کنترل کیفیت در مطالعات ژنتیکی استفاده می‌شود. روش کنترل ژنومیک برای پیمایش ارتباطات نشانگرها تحت فرضیه صفر و برآورد لایه بندی جمعیتی با استفاده از آماره لامبدا و نمودار Q-Q plot استفاده می‌گردد. کنترل ژنومیک روش ناپارامتری برای کنترل لایه‌بندی جمعیتی در مطالعات case-control می‌باشد (Devlin and Roeder, 1999). این روش از نظر محاسباتی آسان و سریع است (Hinrichs *et al.*, 2009). علاوه براین می‌تواند برای تصحیح برای ساختار خانواده و ارتباطات نهان استفاده شود (Thornton

and McPeek, 2010). در مطالعات ارتباطی کنترل ژنومیک نشانگرهای کل ژنوم را برای تصحیح هر گونه تورم در تست آماری به دلیل وجود زیر ساختار، استفاده می‌کند (Bacanu *et al.*, 2002). روش‌های مبتنی بر مدل نیز برای استباط ساختار جمعیت و انتساب افراد به جمعیت‌ها به کار برده شده است (Pritchard *et al.*, 2000). از انواع الگوریتم‌های مبتنی بر مدل، مدل Admixture برای دستیابی به ساختار جمعیت با استفاده از کل ژنوم ارائه شده است که این به لحاظ محاسباتی کارآمد بوده و برای داده های بزرگ کاربرد دارد و نیازمند توزیع پیشین برای پارامترهای مدل و متکی به اطلاعات انساب می‌باشد (Alexander *et al.*, 2009). الگوهای اختلاط و ساختار جمعیتی در جمعیت‌های شمال آمریکای شمالی با روش Admixture بررسی شده است (Verdu *et al.*, 2014). لذا، در این مطالعه روش‌های GC[1] (کنترل ژنومیک)، PCA[2] (آنالیز مولفه اصلی)، MDS[3] (مقیاس بندی چند بعدی) و Admixture برای بررسی ساختار جمعیتی گاومیش‌های اکوتیپ آذری و شمالی اجرا شد.

مواد و روش ها

نمونه‌های حیوانی و تعیین ژنوتیپ

نمونه‌ها از گله‌های مردمی و گله‌های تحت سیستم ثبت شجره و رکوردهای مرکز بهبود تولیدات دامی و اصلاح نژاد جمع آوری شد. فاکتورهای مورد توجه در گزینش حیوانات، انتخاب حیوانات غیر خویشاوند و حیواناتی بودند که تا حد ممکن پراکنش متفاوتی داشته و بیانگر تنوع موجود در جمعیت ها بودند. نمونه برداری از استان های آذربایجان غربی (از سه شهر خوی، ارومیه و مهاباد)، آذربایجان شرقی (از ۵ شهر شامل تبریز، سراب، بستان آباد، اسکو و ایلخچی)، اردبیل (از دو شهر نمین و مشکین شهر) و گیلان (از ۷ شهر ماسال، تالش، صومعه سرا، بندر انزلی، طاهر گوراب، رضوانشهر و اسالم) انجام گرفت. در کل ۲۶۲ نمونه به ترتیب ۶۸، ۶۵،۵۶ و ۷۳ نمونه از استان‌های آذربایجان غربی، اردبیل، آذربایجان شرقی و گیلان جمع-آوری شد. استخراج DNA ژنومیک از ریشه مو و خون با روش بهینه نمکی انجام شد. نمونه‌ها جهت انجام مراحل بعدی توالی‌یابی به آزمایشگاه ژنومیک مرکز تحقیقات پادانو (Parco Tecnologico Padano) کشور ایتالیا منتقل شدند سپس نمونه‌ها با استفاده از تراشه‌های Array Axiom® Buffalo Genotyping 90K مربوط به شرکت افی‌متریکس کشور ایتالیا تعیین ژنوتیپ شدند. این آرایه‌ها امکان تعیین ژنوتیپ بیش از۸۵ هزار جایگاه نشانگریا اسنیپ را فراهم می‌آورند.

مراحل فیلتراسیون داده های حاصل از تعیین ژنوتیپ جهت انجام آنالیزهای نهایی

برای اطمینان از کیفیت داده‌های حاصل از تعیین ژنوتیپ، در آنالیزهای نهایی مراحل

[1] Genomic Control
[2] Principle Component Analysis
[3] Multiple Dimensional Scaling

مختلف فیلتراسیون بر روی داده های اولیه با استفاده از نرم افزار Plink، اعمال شد بدین ترتیب که در ابتدا حیوانات دارای بیش از ۵٪ ژنوتیپ از دست رفته از آنالیزهای بعدی کنار گذاشته شد چون نمونه‌های با کیفیت پایین با احتمال بیشتری با داده های گمشده همراه هستند و منجر به افزایش خطای ژنوتایپ می‌شود (Barendse *et al.*, 2009). دو فاکتور حداقل فراوانی آللی(MAF) و درصد نمونه‌هایی که برای آن نشانگر ژنوتایپ شده‌اند (Call rate) برای هر اسنیپ محاسبه شدند واسنیپ‌هایی که در مجموع دارای Call rate و MAF به ترتیب کمتر از ۹۵٪ و ۱ ٪ بودند، حذف شدند. برای اسنیپ‌های باقی مانده در صورت عدم تعادل هاردی–وینبرگ به عنوان معیاری از خطای ژنوتایپینگ کنار گذاشته شدند (Teo *et al.*, 2007). برای تعیین سطح معنی داری مطلوب در این آزمون از تصحیح بنفرونی($\beta=\alpha/n$) استفاده شد. به عبارتی پس از تعیین ژنوتیپ، با عمل غربالگری توضیح داده شده، اسنیپ‌های منتخب وارد مرحله دیگر آنالیز شدند. در این مطالعه کنترل کیفیت اولیه روی داده‌ها توسط شرکت پادانو انجام گرفت که بعد از کنترل کیفیت اولیه، ٤ نمونه در جریان تعیین ژنوتیپ(دو نمونه از استان اردبیل و دو نمونه از استان گیلان) با بیش از ۵ درصد ژنوتیپ گم شده حذف شدند. در مجموع تعداد ۸۸۵۵ اسنیپ به دلیل MAF کمتر از ۱٪ ، ۳۳٦ اسنیپ به دلیل انحراف از تعادل هاردی–وینبرگ در سطح ۵٪ و ۱۹ اسنیپ بخاطر موقعیت ناشناخته حذف شدند. در نهایت۲۵۸ حیوان با ٦٤٧۵۰ اسنیپ ، مراحل کنترل کیفیت را با MAF>0.01 و call rate>0.99 گذراندند و همه اسنیپ‌های باقی مانده در سطح ۵٪ در تعادل هاردی وینبرگ بودند.

آنالیزهای آماری

مقیاس چند بعدی یا MDS[1]

این روش برای بررسی ساختار جمعیت و ارتباط میان افراد، براساس ماتریس همبستگی IBS بین دو فرد، عمل می کند که برای این کار از نرم افزار PLINK (Purcell *et al.*, 2007) استفاده شد. سپس MDS در این ماتریس با تابع cmdscale در نرم افزار R اجرا شد.

کنترل ژنومیک یا GC

روش کنترل ژنومیک برای برآورد لایه بندی جمعیتی با استفاده از آماره لامبدا و نمودار Q-Q plotاستفاده می‌گردد. آماره لامبدا که از تقسیم میانه مقادیر کای مربع مشاهده شده بر میانه مورد انتظار (۰/٤۵٦) حاصل می شود، فاکتور inflation می باشد. اگر مقدار آماره لامبدا کمتر یا مساوی یک باشد، نشان دهنده عدم وجود اثر لایه بندی می باشد. در این مطالعه روش کنترل ژنومیک و ترسیم Q-Q plot در نرم افزار R و با پکیج SNPassoc اجرا شد.

آنالیز به مولفه های اصلی یا PCA

هدف از تجزیه به مؤلفه‌های اصلی آن است که واریانس موجود در داده‌های چندمتغیره را به مؤلفه‌هایی تجزیه کند که اولین مؤلفه تا آنجا که ممکن است علت بیشترین واریانس موجود در داده‌ها باشد. دومین مؤلفه علت بیشترین واریانس ممکن بعد از مؤلفه اول و الی آخر باشد. بعلاوه، در این روش هر مؤلفه مستقل از مؤلفه‌های دیگر است، یعنی بین هر مؤلفه و مؤلفه‌های دیگر همبستگی وجود ندارد. یعنی در فضا هر مؤلفه از نظر جهت در زاویه طرف راست مؤلفه‌های دیگر قرار دارد. آنالیز PCA با تابع prcomp در نرم افزار R انجام شد.

بررسی آمیختگی بین جمعیت‌ها

برای بررسی آمیختگی بین جمعیت‌ها، آنالیز اختلاط نژادی با استفاده از نرم افزار Admixture 1.23 در محیط لینوکس صورت گرفت و میزان اشتراک ژنتیکی این نژادها به‌صورت گراف بدست آمد (Alexander *et al.*, 2009). از فرمت داده‌های استاندارد و باینری PLINK به‌عنوان داده‌های ورودی این نرم افزار استفاده شد و فایل خروجی آن یک فایل متنی بود و گراف مربوط به این فایل خروجی در محیط R رسم شد. فاکتور K که عدد آن در این نرم‌افزار در تشخیص تعداد جمعیت‌ها نقش دارد و مبنای تفکیک جمعیت‌ها مقدار عددی این فاکتور می‌باشد. با توجه به خطای اعتبارسنجی متقابل پایین انتخاب می‌شود.

نتایج و بحث

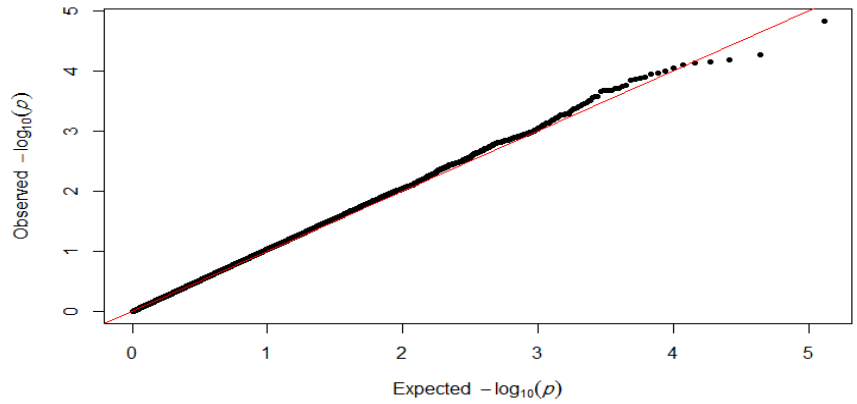
نتایج کنترل کیفیت

کنترل کیفیت روی ٦٤٧٥٠ اسنیپ بدست آمده از کنترل کیفیت اولیه اجرا شد که در ابتدا ۱۹ اسنیپ به دلیل موقعیت ناشناخته حذف شدند و در مراحل مختلف کنترل کیفیت روی اسنیپ های باقیمانده ۷ اسنیپ با MAF کمتر از ۱ درصد حذف شدند و ۵ اسنیپ هم به دلیل انحراف از تعادل هاردی –وینبرگ از آنالیزهای نهایی حذف شدند و در مجموع ۲۵۸ حیوان از ٤ استان مختلف از دو اکوتیپ با ٦٤٧١٩ اسنیپ وارد مرحله آنالیز نهایی شدند.

آنالیز آماری

کنترل ژنومیک

چندین روش برای تخمین کنترل تورم ژنومیکی وجود دارد که در این تحقیق روش تخمین‌گر میانه اجرا شد. پارامتر تورم ژنتیکی تخمین زده شده حدود ۱/۰٥٦۹ (λ_{GC}=1.0569) بود که لایه بندی جمعیتی ضعیفی را نشان می‌دهد. به عبارتی انحراف از یک نشان دهنده این موضوع است که حیوانات ٤ استان کاملا خالص نیستند و اختلاط و ارتباط ژنتیکی بین این حیوانات وجود دارد. گراف مربوط به Q-Q (شکل ۱) ارائه شده است. هر گونه انحراف از خط نشان دهنده وجود اثر لایه بندی جمعیتی می باشد.



شکل ۱- Q-Q plot برای ارزیابی لایه بندی.

Figure 1- Plot Q-Q for evaluation of stratification.

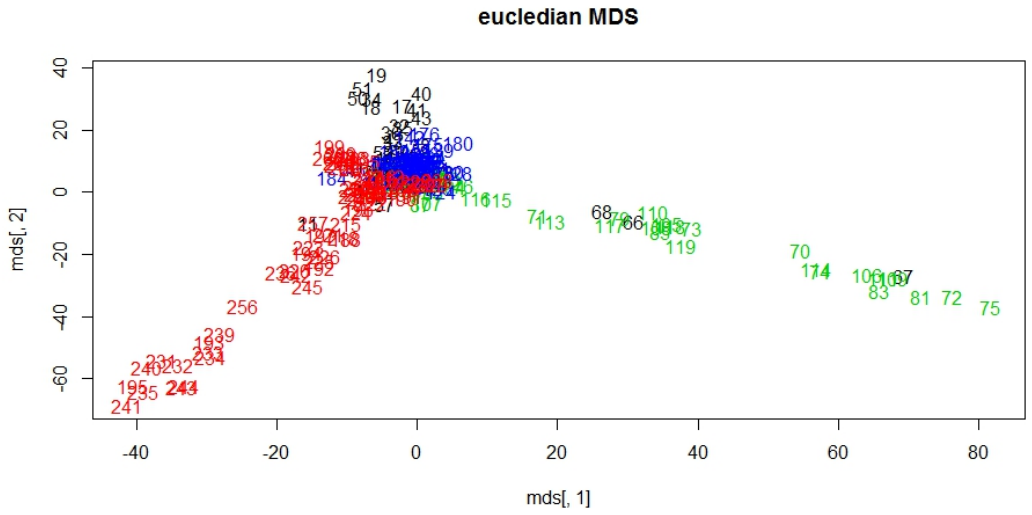
مقیاس بندی چند بعدی

آنالیز مقیاس بندی چند بعدی که برای مشاهده شباهت یا تفاوت در فضای با ابعاد کم است توزیع افراد استان‌های مختلف را نشان می-دهد (شکل ۲) که نتایج حاکی از این است که افرادی که دورترند خلوص بیشتری نسبت به افرادی که در میانه قرار گرفته اند دارند و افرادی که در میانه هستند اختلاط داشته و هیبرید هستند. در این شکل رنگ های سبز، قرمز، آبی و سیاه به ترتیب استان‌های اردبیل، گیلان، تبریز و ارومیه می‌باشد که بیشترین فاصله را حیوانات استان اردبیل با استان گیلان دارند ولی با توجه به اختلاطی که در میانه شکل رخ داده است تفکیک کامل امکان پذیر نیست و نمی‌توان گفت که حیوانات دو استان از هم جدا هستند.

آنالیز مولفه های اصلی

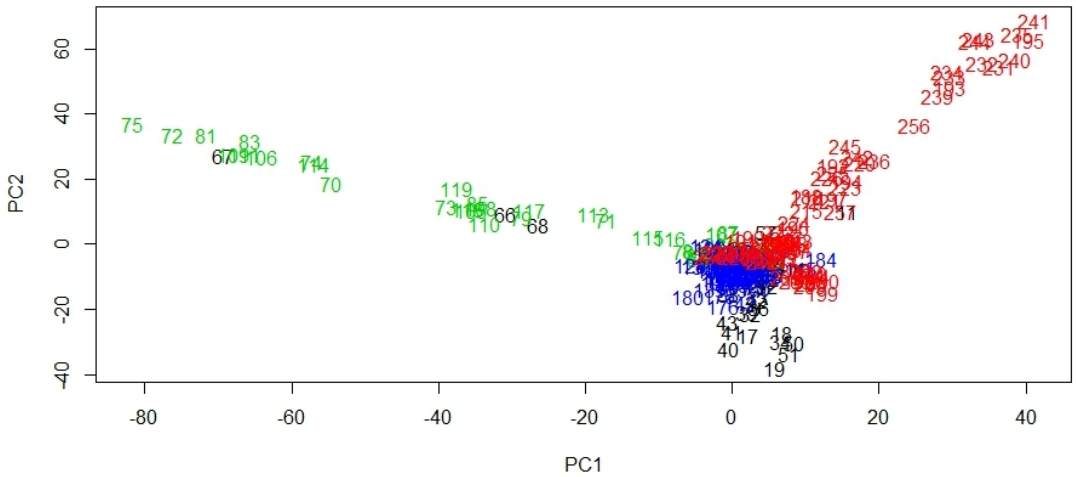
روش PCA، مولفه‌های اصلی را که ساختار جمعیت را براساس همبستگی ژنتیکی میان افراد بیان می‌کند، شناسایی می‌کند. برای ارزیابی اختلاف ژنتیکی میان جمعیت‌ها (٤ استان) در دو اکوتیپ شکلPCA (شکل ۳) ترسیم شد که نشان دهنده نزدیکی ژنتیکی افراد استان های مختلف دو اکوتیپ است که مثل نتایج MDS، در این شکل رنگ های قرمز، سبز، آبی و سیاه به ترتیب استان های اردبیل، گیلان، تبریز و ارومیه می‌باشد که بیشترین فاصله را حیوانات استان اردبیل با استان گیلان دارند ولی با توجه به اختلاطی که در میانه شکل رخ داده است ارتباط ژنتیکی بین این استان ها وجود دارد. نتایج آنالیز PCAبر اساس PC1 و PC2 نشان داد که این ٤ استان در میانه هم پوشانی دارند. این دو PCA در ۲/٤ درصد واریانس را توجیه می کنند. PCA۲٦

اول ۲۰ درصد واریانس را در این جمعیت توجیه می کنند که پایین بودن مقدار واریانس توجیهی نشان دهنده این است که این جمعیت ها تمایز کمتری دارند.



شکل ۲– MDS و دسته بندی افراد استان های مختلف (رنگ های سبز، قرمز، سیاه و آبی به ترتیب متعلق به استان های اردبیل، گیلان، آذربایجان غربی و آذربایجان شرقی می باشد).

Figure 2- MDS and categorization of the different provinces (green, red, blue and black color shows Ardebil, Guilan, West Azerbaijan and East Azerbaijan, respectively).



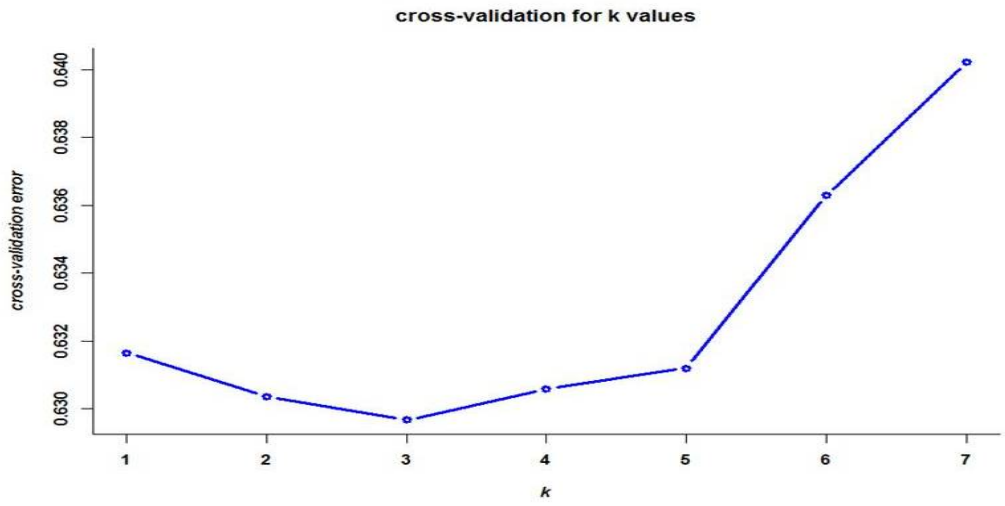
شکل ۳– آنالیز PCA مربوط به گاومیش های استان های مختلف دو اکوتیپ (رنگ‌های قرمز، سبز، سیاه و آبی به ترتیب استان های اردبیل، گیلان، آذربایجان غربی و آذربایجان شرقی را نشان می‌دهد).

Figure 3- PCA analysis of the buffalo from the provinces of two ecotypes (red, yellow, green and blue color shows Ardebil, Guilan, West Azerbaijan and East Azerbaijan, respectively).

آمیختگی بین جمعیت‌ها یا Admixture

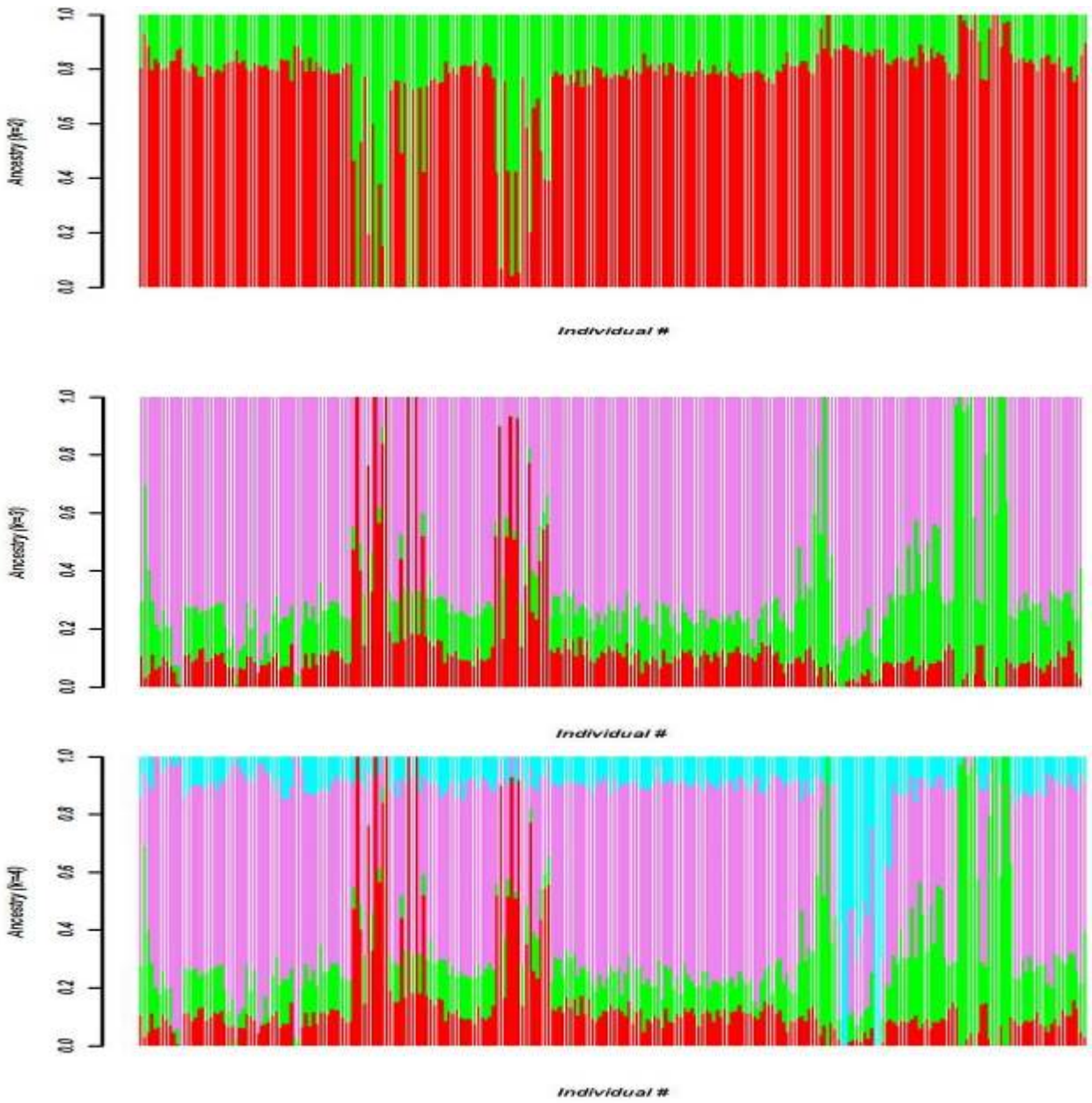
فاکتور k که تعداد جمعیت ها را نشان می‌دهد از ۲ تا ۷ در نظر گرفته شد و k=3 با ۱۰ بار اعتبار سنجی متقابل کمترین خطا را نشان داد (شکل ٤). گراف‌های مربوط به k از ۲ تا ٤ برای مثال ارائه شده است (شکل ٥). طبق شکل‌های بالا که درصد اختلاط در سمت چپ قابل مشاهده است افرادی که تا بالای نمودار با یک رنگ مشخص ادامه دارند خالص هستند و افرادی که رنگ‌های مختلف را با درصدهای مختلف دارند مخلوط بوده و هیبرید هستند. نتایج حاکی از این است که اختلاط و ارتباط ژنتیکی افراد استان‌های مختلف دو اکوتیپ بالاست و طبق نتایج بدست آمده، این حیوانات متعلق به یک جمعیت می‌باشند.

روش MDS انعطاف پذیرتر از روش PCA بود. PCA نیازمند پیش فرض توزیع نرمال چند جمله‌ای داده‌ها است در حای که MDS این محدودیت را نداشته و می‌تواند برای هر نوع از تشابهات و فواصل نیز به کار رود. در حالتی نیز که MDS می‌تواند از ماتریس کواریانس بدست آید در این صورت نتایج دو روش در تصحیح لایه‌بندی جمعیتی یکسان می‌شود (Li and Yu, 2008).



شکل ٤- اعتبارسنجی متقابل برای داده های گاومیش برای تعیین K.

Figure 4- Cross validation for buffalo data to determine K.



شکل ۵- ساختار Admixture. در هر شکل، هر خوشه یا کلاس با رنگ ها مختلف بیان می شود و هر فرد توسط خطوط عمودی به k بخش رنگی با ارتفاع متناسب با سهم ژنوتیپ در خوشه ها تقسیم شده است.

Figure 5- Structure Admixture. In each plot, each cluster classes with different colors and each one is represented by vertical lines to k-color section height divided according to the proportion of genotype in clusters.

محدودیت دیگر PCA، به تصویر کشیدن تعداد زیادی PC به طور همزمان برای نمایان کردن ساختار جامعه ممکن نیست (Gao and Starmer, 2007). در آنالیز ساختار جمعیت نژادهای گاومیش با روش خوشه بندی بیزی و PCA، با استفاده از ۹۳۵ اسنیپ پلی مورفیسم

ازBovine SNP50K Bead Chip، نژادها از نظر ساختاری به هم نزدیک بودند و این درحالی بود که این نژادها پیش زمینه تفاوت ژنتیکی معنی داری هم نداشتند (Wu *et al.*, 2013). در مطالعه‌ای روی جمعیت های گوسفند دنیا، ۲۰ تا PCاول در مجموع فقط ۱۶ درصد واریانس را توجیه کردند که بزرگ ترین مقدار مربوط به PC1 به مقدار ۱/۹۸ درصد بود (Kijas *et al.*, 2012). در مطالعه‌ای روی هفت نژاد بز ایرانی با ۱۴ نشانگر ریزماهواره PC اول ۲۵/۳۹ درصد واریانس و PC دوم ۱۹/۷۰ درصد واریانس را توجیه کرد، آماره F_{ST} نیز محاسبه گردید و آنالیزSTRUCTURE که جزو روش‌های مبتنی بر مدل است، نژادهای بز ایرانی را به سه خوشه تقسیم کرد و اختلاطی بین خوشه‌های مرکزی و شمالی مشهود بود در حالی که خوشه غربی استخر ژنی کاملاً مجزایی تشکیل داد (Vahidi *et al.*, 2014). در بررسی ساختار ژنتیکی گاوهای بومی ایران براساس نشانگرهای چند شکل تک نوکلئوتیدی، روش مبتنی بر مدل به کار گرفته شد و نتایج تعداد سه خوشه را توجیه کرد (Karimi *et al*, 2015). ساختار ژنتیکی و تنوع ژنتیکی گاوهای وحشی و اهلی بنگلادش با اسنیپ چیپ 80k بررسی شد. نتایج ساختار جمعیت و آنالیز مولفه اصلی پیشنهاد کرد که گایال جدا از بوس ایندیکوس بوده و دو جمعیت زبو ساختار ضعیفی داشتند (Uzzaman *et al.*, 2014). این روش‌هایی که ذکر شد جزء روش های بدون نظارت هستند که اطلاعات اولیه در آنالیزها وارد نمی‌شود همچنین این روش‌ها توانایی اجرای سریع داده-های با حجم زیاد (صدها هزار نشانگر و هزاران نمونه)را دارند. در مقایسه‌ی دو روش PCA و MDS برای بررسی ساختار جمعیتی، در حالی که هر دو روش ساختار جمعیتی مشابهی را شناسایی کردند، PCA کمی بهتر از MDS در تصحیح لایه بندی جمعیتی عمل کرد (Wang *et al.*, 2009). در بررسی الگوهای اختلاطی و ساختار جمعیتی جمعیت شمال غرب آمریکای شمالی با روش Admixture نتایج برخی از شباهت ها را نشان داد ولی اختلاف در بین الگوهای اختلاط در شمال غرب اقیانوس آرام و آمریکای لاتین جود داشت (Verdu *et al.*, 2014). روش‌های GC برای تست آماری GWAS با مدل‌های متنوع توارثی می‌تواند اعمال شود (Tsepilov *et al.*, 2013). در مطالعه‌ای برای بررسی لایه بندی جمعیتی روش‌های مختلفی مقایسه شدند که در این بین روش‌های مبتنی بر مدل و PCA عملکرد مشابهی داشتند و این در حالی بود که روش کنترل ژنومیک در جمعیت‌هایی که لایه‌بندی جمعیتی معنی داری وجود داشت، عملکرد خوبی نداشت (Zhang *et al.*, 2008). روش Admixture می‌تواند برای تخمین تعداد جمعیت‌های پایه از طریق اعتبارسنجی متقابل استفاده شود و افراد از انساب شناخته شده می-توانند در یادگیری با نظارت برای تولید دقت بیشتر تخمین اجداد بکار برده شوند (Alexander and Lange, 2011). در این مطالعه با توجه به اینکه روش‌های مختلف برای بررسی ساختار

جمعیت به کار برده شد، و همه روش‌های ذکر شده توانستند ساختار جمعیت‌های دو اکوتیپ را نشان دهند و نتایج بدست آمده حاکی از این بود که با وجود افراد خالص در این جمعیت ها، این افراد از دو اکوتیپ مختلف متعلق به یک نژاد هستند و اشتراک ژنتیکی زیادی دارند. نتایج این تحقیق می‌تواند در صورت وجود صفات فنوتیپی اجرا شود و نتایج حاصل از آنالیز GWAS با استفاده از اطلاعات نژادهای دیگر و با در نظر گرفتن زیر جمعیت ها و مقایسه آن با نتایج حاصل از بدون درنظر گرفتن زیر جمعیت ها منجر به فهم بهتر اهمیت بررسی ساختار و لایه بندی جمعیت شود.

منابع

Mohammad Abadi MR, Askari N, Baghizadeh A, Esmailizadeh AK (2009). A directed search around caprine candidate loci provided evidence for microsatellites linkage to growth and cashmere yield in Rayini goats. Small Ruminant Research 81:146-151.

Alexander DH, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655-1664.

Askari N, Abadi MM, Baghizadeh A (2011). ISSR markers for assessing DNA polymorphism and genetic characterization of cattle, goat and sheep populations. Iranian Journal of Biotechnology 9:222-9.

Bacanu SA, Devlin B, Roeder K (2002). Association studies for quantitative traits in structured populations. Genetic Epidemiology 22:78-93.

Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009). Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:178.

Devlin B, Roeder K (1999). Genomic control for association studies. Biometrics 55:997-1004.

Epps CW, Castillo JA, Schmidt-Küntzel A, du Preez P, Stuart-Hill G, Jago M, Naidoo R (2013). Contrasting historical and recent gene flow among African buffalo herds in the Caprivi Strip of Namibia. Journal of Heredity ess142.

Gao X, Starmer J (2007). Human population structure detection via multilocus genotype clustering. BMC Genetics 8:34.

Hinrichs AL, Larkin EK, Suarez BK (2009). Population stratification and patterns of linkage disequilibrium. Genetic Epidemiology 33:S88-S92.

Karimi K, Esmailizadeh Koshkoiyeh A, Asadi Fuzi M (2015). Analysis of genetic structure of Iranian indigenous cattle populations using dense single nucleotide polymorphism markers. Animal Production Research 4:93-104.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Plos-Biology 10:331.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D (2008). Correlation between genetic and geographic structure in Europe. Current Biology 18:1241-1248.

Li Q, Yu K (2008). Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genetic Epidemiology 32:215-226.

Liu L, Zhang D, Liu H, Arendt C (2013). Robust methods for population stratification in genome wide association studies. BMC Bioinformatics 14:1.

Marchini J, Cardon LR, Phillips MS, Donnelly P (2004). The effects of human population structure on large genetic association studies. Nature Genetics 36:512-517.

Mohammadi A, Nassiry M, Mosafer J, Mohammadabadi M, Sulimova G (2009). Distribution of BoLA-DRB3 allelic frequencies and identification of a new allele in the Iranian cattle breed Sistani (Bos indicus). Russian Journal of Genetics 45:198-202.

Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. PLoS Genetetics 2:e190.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909.

Price AL, Zaitlen NA, Reich D, Patterson N (2010). New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11:459-463.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.

Shojaei M, Mohammad Abadi M, Asadi Fozi M, Dayani O, Khezri A, Akhondi M (2011). Association of growth trait and Leptin gene polymorphism in Kermani sheep. Journal of Cell and Molecular Research 2:67-73.

Teo YY, Fry AE, Clark TG, Tai E, Seielstad M (2007). On the usage of HWE for identifying genotyping errors. Annals of Human Genetics 71:701-703.

Thomas DC, Witte JS (2002). Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiology Biomarkers & Prevention 11:505-512.

Thornton T, McPeek MS (2010). ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. The American Journal of Human Genetics 86:172-184.

Tsepilov YA, Ried JS, Strauch K, Grallert H, van Duijn CM, Axenovich TI, Aulchenko YS (2013). Development and application of genomic control methods for genome-wide association studies using non-additive models. Plos One **8**, e81431.

Uzzaman MR, Edea Z, Bhuiyan MSA, Walker J, Bhuiyan A, Kim KS (2014). Genome-wide Single Nucleotide Polymorphism Analyses Reveal Genetic Diversity and Structure of Wild and Domestic Cattle in Bangladesh. Asian-Australasian journal of Animal Sciences 27:1381.

Vahidi SM, Tarang AR, Naqvi AU, Falahati Anbaran M, Boettcher P, Joost S, Colli L, Garcia JF, Ajmone-Marsan P (2014). Investigation of the genetic diversity of domestic Capra hircus breeds reared within an early goat domestication area in Iran. Genetic Selection Evolution **46**, 27.

Verdu P, Pemberton TJ, Laurent R, Kemp BM, Gonzalez-Oliver A, Gorodezky C, Hughes CE, Shattuck MR, Petzelt B, Mitchell J (2014). Patterns of admixture and population

structure in native populations of Northwest North America. Plos Genetics **10**, e1004530.

Wacholder S, Rothman N, Caporaso N (2002). Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiology Biomarkers & Prevention 11:513-520.

Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q (2009). Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling, BMC proceedings, BioMed Central 3:S109

Wu JJ, Song LJ, Wu FJ, Liang XW, Yang BZ, Wathes DC, Pollott GE, Cheng Z, Shi DS, Liu QY (2013). Investigation of transferability of BovineSNP50 BeadChip from cattle to water buffalo for genome wide association study. Molecular Biology Reports 40:743-750.

Zamani P, Akhondi M, Mohammadabadi MR, Saki AA, Ershadi A, Banabazi MH, Abdolmohammadi AR (2013). Genetic variation of Mehraban sheep using two intersimple sequence repeat (ISSR) markers. African Journal of Biotechnology 10:1812-1817.

Zhang F, Wang Y, Deng H.-W (2008). Comparison of population-based association study methods correcting for population stratification. Plos One 3:e3392.

Study of population structure and stratification two ecotypes buffalo with dense single nucleotide polymorphism markers using Admixture, MDS, PCA and GC methods

Azizi Z.[1], Rafat A.[2], Shoja J.[3], Moradi Shahrbabak H.[*4], Moradi Shahrbabak M.[5]

[1] PhD student, Department of Animal Sciences, Faculty of Agricultural Sciences, University of Tabriz.
[2] Associate Professor, Department of Animal Sciences, Faculty of Agricultural Sciences, University of Tabriz.
[3] Professor, Department of Animal Sciences, University College of Agriculture and Natural Resources, University of Tehran.
[4] Assistant Professor, Department of Animal Sciences, University College of Agriculture and Natural Resources, University of Tehran.
[5] Assistant Professor, Department of Animal Sciences, University College of Agriculture and Natural Resources, University of Tehran.

Abstract

In applications of population genetics, classification of individuals in a sample into populations is important. With the development of high throughput genotyping technologies many markers such as SNPs are available which useful in the study of genetic diversity and structure population. The purpose of this research was to study of population structure and stratification buffaloes from different areas of the two ecotypes (Azari and North) using data SNPChip 90K. A total of 258 buffalo from Ardabil, West Azarbaijan, East Azarbaijan and Guilan provinces were sampled and genotyped. The result showed weak population stratification with λ =1.056 for GC method. Also the plots obtained from PCA and MDS showed separation of different provinces based on genetic distance and these animals have closed genetic relationship. Admixture method represents same results and admixture between individual from different provinces of two ecotypes and k=3 have low error cross validation. These methods are generally able to separate the animals. The results showed the close genetic relationship between two ecotypes from 4 different provinces.

Keywords: *Population Stratification, Buffalo, SNPChip 90K, MDS, PCA*.

[*] Corresponding Author: Moradi Shahrbabak H. Tel: 09133915306 Email: hmoradis@ut.ac.ir