دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشهای تولیدات دامی

بررسی ساختار جمعیتی گاوهای بومی ایران با استفاده از تحلیل افتراقی مؤلفه‌های اصلی

کریم کریمی

دانش‌آموخته دکترای تخصصی ژنتیک و اصلاح نژاد دام، دانشگاه شهید باهنر کرمان و
عضو انجمن پژوهشگران جوان، دانشگاه شهید باهنر کرمان، (نویسنده مسوول: karim.karimi81@gmail.com)
تاریخ دریافت: ۹۵/۲/۵ تاریخ پذیرش: ۹۵/۱۱/۲۴

چکیده

مدیریت مؤثر منابع ژنتیکی در دام‌های اهلی مبتنی بر شناخت ساختار و تنوع ژنتیکی جمعیت‌ها است. به کارگیری داده‌های حجیم ژنومی جهت شناسایی ارتباط ژنتیکی میان جمعیت‌ها نیازمند استفاده از روش‌هایی است که ابعاد و پیچیدگی این داده‌ها را کاهش دهند. هدف از تحقیق حاضر استفاده از روش تحلیل افتراقی مؤلفه‌های اصلی جهت بررسی ساختار ژنتیکی جمعیت گاوهای بومی ایران بر اساس داده‌های حاصل از نشانگرهای متراکم چندشکل تک نوکلئوتیدی بوده است. تعداد ۹۰ رأس گاو از هشت جمعیت مختلف از گاوهای بومی کشور شامل نژادهای سرابی، سیستانی، کردی، کرمانی، مازندرانی، تالشی، نجدی و پارس مورد نمونه‌برداری قرار گرفتند. تعیین ژنوتیپ نمونه‌ها با استفاده از Illumina BovineHD SNP chip شامل ۷۷۷۹۶۲ جایگاه SNP انجام گرفت. فواصل ژنتیکی میان نژادهای مختلف بر اساس تفاوت در فراوانی‌های آللی جمعیت‌ها تعیین شد. تحلیل افتراقی مؤلفه‌های اصلی به کمک بسته adegenet در نرم‌افزار R بر روی داده‌ها اجرا شد. دو نژاد پارس و کرمانی کمترین فاصله ژنتیکی را در بین نژادهای مورد بررسی داشتند و بیشترین فاصله ژنتیکی میان دو نژاد کردی و سیستانی مشاهده شد. تعداد سه خوشه ژنتیکی بر اساس روش K-means به عنوان مناسب‌ترین تعداد گروه استنتاج شده جهت اجرای تحلیل افتراقی مؤلفه‌های اصلی مورد انتخاب قرار گرفت. نتایج این تحقیق وجود سه گروه ژنتیکی عمده را در بین گاوهای بومی ایران ثابت کرد. جمعیت‌های مازندرانی و تالشی در انطباق با توزیع جغرافیایی این نژادها، در خوشه یکسانی قرار گرفتند. همچنین، نژادهای پراکنده شده در مناطق کوهستانی شمال غرب کشور (سرابی و کردی) و نژادهای مناطق جنوبی کشور (سیستانی، کرمانی، پارس و نجدی) دو گروه ژنتیکی متمایز دیگر را تشکیل دادند. تحلیل افتراقی مؤلفه‌های اصلی به خوبی توانست موجب تفکیک گروه‌های ژنتیکی مرتبط با افراد نژادهای مختلف شود. ساختار ژنتیکی شناسایی شده در پژوهش حاضر می‌تواند در طراحی برنامه‌های حفاظت ژنتیکی برای گاوهای بومی ایران به کار گرفته شود.

واژه‌های کلیدی: تحلیل افتراقی مؤلفه‌های اصلی، کاهش ابعاد داده‌ها، چندشکلی تک نوکلئوتیدی، گاوهای بومی ایران

مقدمه

شناسایی ساختار ژنتیکی جمعیت‌ها امکان بررسی منشأ شکل‌گیری و ارتباطات میان جمعیت‌ها را فراهم می‌آورد. مطالعه ساختار ژنتیکی جمعیت‌ها همگام با دسترسی روزافزون به داده‌های ژنومی تسهیل شده است (۳). گسترش فناوری‌های نوین در عرصه توالی‌یابی ژنوم امکان دستیابی به حجم گسترده‌ای از اطلاعات ژنتیکی را در گونه‌های مختلف دام‌های اهلی فراهم کرده است. با افزایش ابعاد و پیچیدگی داده‌های ژنتیکی، نمایش ارتباطات ژنتیکی میان افراد نیازمند استفاده از روش‌هایی است که قادر به کاهش ابعاد داده‌ها[1] باشند (۱). تجزیه و تحلیل‌های چندمتغیره از جمله روش‌هایی هستند که جهت کاهش ابعاد داده‌ها و استخراج انواع مختلفی از اطلاعات ژنتیکی به کار گرفته شده‌اند. این روش‌ها مبتنی بر یافتن متغیرهایی هستند که ترکیب خطی از آلل‌ها بوده و واریانس بین افراد را تا حد امکان منعکس می‌کنند (۳۶). تحلیل مؤلفه‌های اصلی[2] (PCA) از جمله روش‌های چند متغیره است که به فراوانی در تجزیه و تحلیل‌های ژنتیکی جمعیت‌ها به کار برده شد. این روش بدون نیاز به در نظر گرفتن فرض اولیه در مورد مدل ژنتیکی جمعیت‌ها، قادر به شناسایی ساختار ژنتیکی جمعیت‌ها است (۲۸،۱۲). با این حال، ارزیابی گروه‌های ژنتیکی شناسایی شده توسط این روش نیازمند وجود یک تعریف اولیه از خوشه‌های موجود در جمعیت است. همچنین، با آنکه واریانس بین افراد در جمعیت‌ها از دو جزء واریانس بین گروهی و واریانس داخل گروهی تشکیل شده است، این روش تنها بر واریانس کل افراد تمرکز یافته است. با این حال، روش‌هایی جهت ارزیابی ارتباط میان خوشه‌های ژنتیکی مناسب‌تر هستند که بر واریانس بین گروه‌ها تمرکز یافته باشند و واریانس داخل گروه‌ها را نادیده می‌گیرند (۲۳). روش تحلیل افتراقی مؤلفه‌های اصلی (DAPC)[3] موجب بهینه‌سازی تفاوت‌های بین گروه‌ها شده و واریانس داخل گروهی را حداقل می‌سازد. به عبارت دیگر در این روش، متغیرهای جدید به گونه‌ای تعریف می‌شوند که اختلافات بین گروه‌ها تا حد ممکن نشان داده شده و واریانس داخل خوشه‌ها حداقل شود. در مقایسه با روش‌های مبتنی بر مدل، این روش نیاز به در نظر گرفتن مدل خاصی در ارتباط با تفرق ژنتیکی جمعیت‌ها ندارد و وجود فرض‌هایی نظیر تعادل هاردی-وینبرگ و تعادل پیوستگی در جمعیت‌ها جهت اجرای این روش لازم نیست. از سوی دیگر، این روش موجب کاهش زمان محاسباتی مورد نیاز جهت استخراج اطلاعات از داده‌های حجیم می‌شود. به علاوه، این روش همانند روش‌های بیزی مبتنی بر مدل، احتمال انتساب افراد را به هر یک از گروه‌های ژنتیکی فراهم می‌آورد (۱۲).

1- Dimension reduction
2- Principal components analysis (PCA)
3- Discriminant analysis of principal components (DAPC)

[DOI: 10.29252/rap.8.17.184]

روش تحلیل افتراقی مؤلفه‌های اصلی در انواع مختلفی از مطلعات ژنومی نظیر ChIP-Seq (۱۰)، RNA-Seq (۷) و آنالیز بیان ژن‌ها (۹) به منظور کاهش ابعاد داده‌ها به کار برده شده است. با این حال، مطالعه ساختار ژنتیکی و روابط میان جمعیت‌ها به عنوان مهم‌ترین کاربرد این روش شناخته شده است. (۲۵). این روش در بررسی ساختار ژنتیکی برخی از نژادهای گاو نیز مورد استفاده قرار گرفت که از آن جمله می‌توان به مطالعه تاریخچه ژنتیکی گاوهای بومی تونس (۲)، سازگاری ژنومی گاوهای شرق آفریقا (۱۶) و ساختار جمعیتی گاو Nguni در شمال آفریقا (۳۳) اشاره کرد. در مطالعات انجام گرفته بر روی لاین‌های مختلف نژادهای خوک، DAPC علاوه بر اینکه موجب تفکیک جمعیت‌های مختلف شد، امکان شناسایی لاین‌های تحت انتخاب در هر یک از جمعیت‌ها را نیز فراهم آورد (۴،۱۷). علاوه بر این، از این روش در بررسی خصوصیات ژنتیکی برخی از جمعیت‌های گیاهی مانند ذرت (۸) و آفتابگردان (۵) نیز بهره گرفته شده است.

نسل‌های بعدی بشر بر چگونگی عملکرد نسل کنونی در حفاظت از منابع ژنتیکی اتکا خواهند داشت. برآورده کردن نیازهای آینده بشر، عدم حساسیت نسبت به تغییرات محیطی و ارزش‌های اکولوژیکی و اجتماعی به عنوان بخشی از دلایل حفاظت از تنوع ژنتیکی مورد بحث قرار گرفته‌اند (۲۹). سرزمین ایران از کهن‌ترین مراکز اهلی کردن و پرورش گاو در دنیا محسوب می‌شود (۳۰). گاوهای بومی طی سالیان متمادی در ایران زیست نموده و با شرایط محیطی مناطق مختلف کشور تطابق پیدا کرده‌اند. از جمله ویژگی‌های بارز گاوهای بومی ایران می‌توان به سازگاری با شرایط چرا، استفاده از خوراک‌های با کیفیت پایین و مقاومت نسبت به انگل‌ها و بیماری‌های منطقه اشاره نمود (۳۲). با این حال، جمعیت گاوهای بومی کشور در سال‌های اخیر دچار کاهش شدیدی شده است. تغییر شیوه زندگی روستائیان و تمایل به نگهداری گاوهای پرتولید از یک سو و انجام تلاقی‌های کنترل نشده با نژادهای خارجی از سوی دیگر، گاوهای بومی ایران را به سوی نابودی پیش برده‌اند. شناخت ساختار و روابط موجود میان این جمعیت‌ها گامی مؤثر در حفاظت از این منابع ژنتیکی خواهد بود. هدف از پژوهش حاضر استفاده از روش تحلیل افتراقی مؤلفه‌های اصلی (DAPC) جهت بررسی ساختار ژنتیکی جمعیت گاوهای بومی بر اساس داده‌های حاصل از نشانگرهای متراکم چندشکل تک نوکلئوتیدی بوده است.

مواد و روش‌ها

نمونه‌گیری، تعیین ژنوتیپ و کنترل کیفی داده‌ها

تعداد ۹۰ رأس گاو از هشت جمعیت مختلف از گاوهای بومی کشور شامل نژادهای سرابی (۲۰ رأس)، سیستانی (۱۰ رأس)، کردی (۱۰ رأس)، کرمانی (۱۰ رأس)، مازندرانی (۱۰ رأس)، تالشی (۱۰ رأس)، نجدی (۱۰ رأس) و پارس (۱۰ رأس) مورد نمونه‌برداری قرار گرفتند. نمونه‌های مربوط به نژادهای سرابی، سیستانی، تالشی و نجدی از مراکز اصلاح نژادی کشور جمع‌آوری شدند و نمونه‌های مربوط به سایر نژادها به طور تصادفی از مناطق روستایی نمونه‌برداری شدند. بر اساس اطلاعات شجره و یا با پرسش از دامداران تا حد ممکن افراد غیرخویشاوند و یا با خویشاوندی کم جهت نمونه‌گیری انتخاب شدند. نمونه‌های مو از ناحیه انتهایی دم گاوها بیرون کشیده شدند و در بسته‌های مخصوص جهت استخراج DNA به آزمایشگاه انتقال داده شد. استخراج DNA از ریشه مو انجام گرفت و تعیین ژنوتیپ نمونه‌ها با استفاده از BovineHD SNP chip (Illumina, Inc, San Diego, CA, USA) شامل ۷۷۷۹۶۲ جایگاه SNP انجام گرفت. کنترل کیفی داده‌ها با استفاده از نرم‌افزار PLINK 1.07 (۲۶) انجام شد. کلیه جایگاه‌های SNP با نرخ فراخوانی تعیین ژنوتیپ کمتر از ۰/۹ و SNPهای دارای فراوانی آلل نادر[1] کمتر از ۰/۰۱ از مجموعه داده‌ها حذف شدند. تعادل هاردی–وینبرگ در کلیه جایگاه‌ها مورد بررسی قرار گرفت و جایگاه‌های با مقدار P-value کمتر از 10^{-7} از داده‌ها کنار گذاشته شدند. همچنین، تنها جایگاه‌های واقع بر کروموزوم‌های اتوزوم مورد انتخاب قرار داده شدند. پس از کنترل کیفی مجموعه داده‌ها، تعداد ۲۸۱۳۳۳ جایگاه SNP جهت استفاده در آنالیزهای بعدی باقی ماندند.

فاصله ژنتیکی استاندارد نئی (۲۱) به کمک بسته StAMPP در نرم‌افزار R (۲۴) و فاصله ژنتیکی رینالدز (۲۷) به کمک نرم‌افزار Arlequin به ازای هر جفت از جمعیت‌ها بدست آورده شدند و به صورت ماتریسی از فواصل جفتی ارائه شدند. درخت فیلوژنتیک جمعیت‌ها با استفاده از الگوریتم جفت گروه‌های وزن نیافته با میانگین حسابی[2] (UPGMA) و بر اساس ماتریس ژنتیکی استاندارد نئی در بسته APE از نرم‌افزار R (۲۲) رسم شد. همچنین، ۱۰۰۰۰ تکرار بوت استراپ جهت ارزیابی درخت‌های تشکیل شده به کار گرفته شد.

تحلیل افتراقی مؤلفه‌های اصلی

تشریح مبانی مدل مورد استفاده

آنالیز افتراقی، DAPC و خوشه‌بندی K-means همگی بر مدل آماری مشابهی جهت اندازه‌گیری اختلافات بین گروه‌ها متکی هستند که در واقع همان مدل تجزیه واریانس (ANOVA) است. فرض کنید y بردار مربوط به n مشاهده از یک متغیر باشد که در g گروه تقسیم شده‌اند و D ماتریس قطری باشد که کلیه عناصر قطری آن برابر با 1/n هستند. همچنین، ماتریس H با ابعاد n×g حاوی بردارهایی باشد که میزان عضویت هر فرد را در گروه‌ها مشخص می‌کنند، به طوری که اگر مشاهده iام متعلق به گروه jام باشد، ۱= h_{ij} و در غیر این صورت ۰= h_{ij} باشد. بنابراین می‌توان ماتریس P را به صورت زیر تعریف کرد:

$$P = H(H^T D H)^{-1} H^T D$$

این ماتریس می‌تواند جهت تعریف مشاهدات بر اساس دو جزء تغییرات بین گروهی و درون گروهی به کار برده شود. اگر X ماتریس داده‌های ژنتیکی باشد که ابعاد آن شامل n فرد در ردیف و g فراوانی نسبی آللی در ستون‌ها باشد، آنالیز افتراقی خطی به دنبال ترکیب خطی از آل‌ها خواهد بود که:

1- Minor allele frequency (MAF)

2- Unweight pair group method with arithmetic mean (UPGMA)

[DOI: 10.29252/rap.8.17.184]

$$f(v) = \sum_{j=1}^{p} X^j v_j = Xv$$

در این معادله بردار v حاوی مقادیر وزنی داده شده به هر یک از p آلل است که در اصطلاح به آنها ضرایب افتراقی گفته می‌شود. هدف از اجرای آنالیز افتراقی انتخاب بردار v به گونه‌ای است که آماره-F مربوط به Xv حداکثر شود. ترکیب‌های خطی از آلل‌ها که موجب بهینه کردن این معیار می‌شوند در اصطلاح معادلات افتراقی نامیده می‌شوند. معادلات افتراقی به کمک آنالیز ویژه بردارهای ماتریس متقارن D قابل محاسبه هستند:

$$PX(W)^{-1}X^TP^TD$$

در این معادله W ماتریس کوواریانس داخل گروه‌ها است که به شکل زیر محاسبه می‌شود:

$$W = X^T\ 1 - P\ ^TD\ 1 - P\ X$$

وقتی که تعداد آلل‌ها بزرگتر از تعداد افراد باشند، ماتریس W معکوس‌پذیر نخواهد بود و از سوی دیگر در صورت وجود همبستگی بین متغیرها معکوس W به لحاظ عددی پایدار نخواهد بود. به همین جهت، در روش DAPC ابتدا داده‌ها به کمک روش PCA تبدیل به مؤلفه‌های اصلی می‌شوند:

$$X^TDXU = U\Lambda$$

که U ماتریس ویژه بردارهای X^TDX و Λ ماتریس قطری ویژه مقدارهای مربوطه است. در مرحله بعد آنالیز افتراقی بر روی ماتریس مؤلفه‌های اصلی اجرا می‌شود. به عبارت دیگر، آنالیز افتراقی با جایگزینی XU به جای X انجام می‌شود:

$$PXU(U^TWU)^{-1}U^TX^TP^TD$$

ویژه بردارهای vحاصل از این معادله موجب حداکثر کردن واریانس بین گروه‌ها شده و واریانس داخل گروه‌ها را به مقدار خاصی محدود می‌کنند. به عبارت دیگر، ضرایب موجود در بردار v می‌توانند جهت محاسبه ترکیبات خطی از مؤلفه‌های اصلی حاصل از PCA (XU) مورد استفاده قرار گیرند به طوری که آماره−F حداکثر مقدار خود را داشته باشد (۱۲).

وقتی که تعداد گروه‌های ژنتیکی در مراحل اولیه مشخص نباشند، می‌توان از الگوریتم خوشه‌بندی K-means جهت تعریف آنها استفاده کرد. در این حالت، با به کارگیری ماتریس مؤلفه‌های اصلی حاصل از PCA (XU)، می‌توان مدل زیر را جهت تفکیک واریانس بین و داخل جمعیت‌ها تعریف کرد:

$$VAR\,(XU) = B(XU) + W(XU)$$

در این معادله $VAR\,(XU) = tr(\Lambda)$، $X^TP^TDPXU)$ و $B(XU) = tr(U^T$ $W(XU) = tr(U^TWT)$می‌باشند.

این الگوریتم با مقادیر مختلف K اجرا شده و از معیار اطلاعات بیزی[1] (BIC) برای انتخاب بهترین مدل خوشه‌بندی به صورت زیر استفاده می‌شود (۱۲):

$$BIC = n\log(W(X)) + g\log(n)$$

آنالیز داده‌ها

تحلیل افتراقی مؤلفه‌های اصلی به کمک بسته adegenet از نرم‌افزار R انجام شد (۱۱). در مرحله نخست، داده‌های حاصل از نرم‌افزار PLINK 1.07 به فرمت داده‌های SNP در بسته adegenet (genlight object) تبدیل شدند. در روش DAPC لازم است که در ابتدا تعداد گروه‌های اولیه موجود در داده‌ها مشخص شوند، در حالی که در اغلب اوقات تعداد این گروه‌ها شناخته شده نیست. به همین جهت، در ابتدا از الگوریتم خوشه‌بندی K-means به منظور شناسایی تعداد مناسب خوشه‌ها استفاده شد. این الگوریتم برای یک تعداد خاص از خوشه‌ها (K)، واریانس بین گروه‌ها را حداکثر می‌کند. به منظور کاهش تعداد متغیرها، پیش از اجرای الگوریتم K-means، داده‌ها به کمک روش PCA به گونه‌ای مورد تبدیل قرار داده شدند که حداقل ۹۰٪ واریانس توسط تعداد مؤلفه‌های اصلی (PC) باقی مانده توجیه شود. با هدف شناسایی مناسب‌ترین تعداد خوشه‌ها، K-means به صورت متوالی همراه با افزایش مقدار K از یک تا ۱۵ اجرا گردید و پاسخ‌های حاصل از هر یک از این مقادیر با استفاده از معیار اطلاعات بیزی (BIC) مورد مقایسه قرار داده شدند. مناسب‌ترین تعداد خوشه‌ها بر اساس کمترین مقدار آماره BIC انتخاب شد. در مرحله نهایی، آنالیز DAPC بر اساس تعداد خوشه انتخاب شده و با در نظر گرفتن ۲۵ مؤلفه اصلی اول اجرا شد و دو تابع افتراقی نخست جهت ادامه آنالیزها مورد انتخاب قرار گرفتند.

نتایج و بحث

تفاوت میان فراوانی‌های آللی در بین جمعیت‌ها جهت محاسبه فواصل ژنتیکی استاندارد نئی (۱۹۷۲) و رینالدز (۱۹۸۳) مورد استفاده قرار گرفتند (جدول ۱). اگرچه مقادیر عددی بدست آمده از این دو روش با یکدیگر تفاوت دارند، اما تصویر واحدی را از فواصل ژنتیکی نسبی بین نژادها ارائه داده‌اند. به طوری که، بر اساس هر دو معیار مورد استفاده، بیشترین فاصله ژنتیکی بین نژادهای سیستانی و کردی مشاهده شد، در حالی که کمترین فاصله ژنتیکی میان دو نژاد کرمانی و پارس یافت شد. شکل ۱ ارتباطات میان جمعیت‌های مختلف را به کمک رسم درخت فیلوژنتیک حاصل از الگوریتم جفت گروه‌های وزن نیافته با میانگین حسابی (UPGMA) بر اساس فاصله ژنتیکی استاندارد نئی نمایش داده است. همان گونه که در این شکل مشخص است دو نژاد سرابی و کردی در یک سوی نمودار گره مشترکی را تشکیل داده‌اند. در سوی دیگر درخت، نژادهای پارس، کرمانی، سیستانی و نجدی روابط ژنتیکی نزدیکی را به نمایش گذاشته‌اند و نژادهای تالشی و مازندرانی نیز گره مشترکی دارند که در قسمت میانی نمودار جای گرفته است.

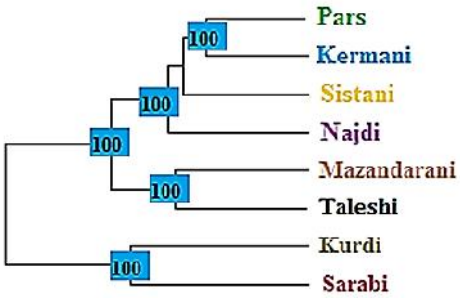
1- Bayesian information criterion (BIC)

[DOI: 10.29252/rap.8.17.184]

جدول ۱- ماتریس جفتی فاصله ژنتیکی استاندارد نئی (۱۹۷۲) (بالای قطر) و فاصله ژنتیکی رینالدز (۱۹۸۳) (پایین قطر) در بین کلیه جفت جمعیت‌های گاو بومی ایران

Table 1. Pairwise Nei's standard genetic distance (1972) (above diagonal) and Reynold's genetic distance (1983) (below diagonal) between all pairs of Iranian native cattle populations

جمعیت	سرابی	کرمانی	سیستانی	نجدی	کردی	تالشی	پارس	مازندرانی
سرابی	۰	۰/۰۹۲	۰/۱۳۴	۰/۰۸۷	۰/۰۵۷	۰/۰۶۷	۰/۰۹۱	۰/۰۶۸
کرمانی	۰/۰۸۹	۰	۰/۰۳۵	۰/۰۳۸	۰/۱۰۸	۰/۰۶۲	۰/۰۳۲	۰/۰۵۱
سیستانی	۰/۱۴۷	۰/۰۳۶	۰	۰/۰۵۴	۰/۱۵۸	۰/۰۹	۰/۰۴۴	۰/۰۷۴
نجدی	۰/۰۹۱	۰/۰۲۹	۰/۰۶۷	۰	۰/۱۰۱	۰/۰۶۲	۰/۰۴۱	۰/۰۵۳
کردی	۰/۰۶۶	۰/۰۸۸	۰/۱۶۲	۰/۰۸۹	۰	۰/۰۷۵	۰/۱۰۷	۰/۰۷۶
تالشی	۰/۰۷۷	۰/۰۵۷	۰/۱۱۱	۰/۰۶۴	۰/۰۶۹	۰	۰/۰۶۲	۰/۰۴۳
پارس	۰/۰۸	۰/۰۱	۰/۰۵۲	۰/۰۲۸	۰/۰۸۱	۰/۰۵۳	۰	۰/۰۵۲
مازندرانی	۰/۰۷۴	۰/۰۴۳	۰/۰۸۹	۰/۰۵	۰/۰۶۵	۰/۰۳۴	۰/۰۳۸	۰

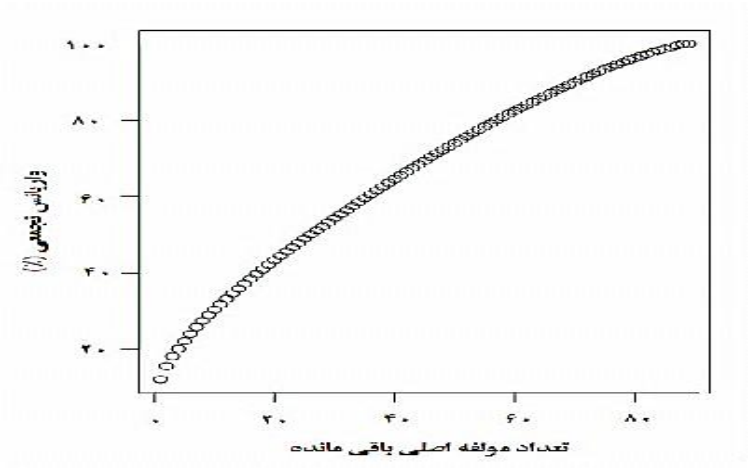


شکل ۱- درخت فیلوژنتیک حاصل از الگوریتم جفت گروه‌های وزن نیافته با میانگین حسابی (UPGMA) بر اساس فاصله ژنتیکی استاندارد نئی جهت نمایش ارتباطات ژنتیکی میان جمعیت‌های گاو بومی ایران

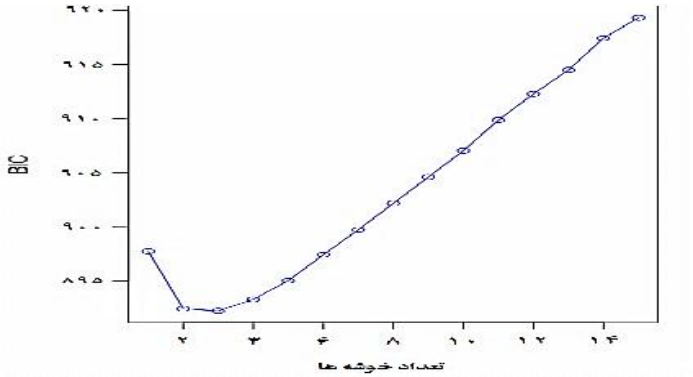
Figure 1. Phylogenetic tree from unweighted pair group method with arithmetic mean (UPGMA) algorithm representing genetic relationships among Iranian native cattle populations based on Nei's standard genetic distance

واریانس ژنتیکی میان افراد یک جمعیت را می‌توان به دو جزء واریانس بین گروهی و واریانس داخل گروهی تقسیم بندی کرد. روش تحلیل افتراقی مؤلفه‌های اصلی (DAPC) به واسطه تعریف متغیرهایی جدید (معادلات افتراقی)، اختلافات بین گروه‌ها را به بهترین شکل ممکن انعکاس می‌دهد (۱۳). در پژوهش حاضر، ساختار ژنتیکی هشت نژاد از گاوهای بومی ایران به کمک روش تحلیل افتراقی مؤلفه‌های اصلی مورد بررسی قرار گرفت. شکل ۲ میزان واریانس تجمعی توجیه شده توسط مؤلفه‌های اصلی را بر حسب تعداد مؤلفه‌های به کار گرفته شده در مدل نشان داده است. همان گونه که در شکل مشخص است همراه با افزایش تعداد مؤلفه‌ها، درصد بالاتری از واریانس نیز توسط آنها توجیه شده است، با این حال سهم چندین مؤلفه اصلی اول در توجیه واریانس داده‌ها بیشتر بوده است. مؤلفه‌های اصلی باقی مانده جهت اجرای الگوریتم K-means به گونه‌ای انتخاب شدند که قادر به توجیه حداقل ۹۰٪ تغییرات داده‌ها باشند. الگوریتم K-means به طور متوالی با در نظر گرفتن یک تا ۱۵ خوشه فرضی در داده‌ها اجرا شد. شکل ۳ نمودار تغییرات مقادیر BIC را به ازای تعداد خوشه‌های فرض شده در الگوریتم K-means نمایش داده است. این نمودار نشان‌دهنده کاهش مقادیر BIC تا K=۳ و افزایش مجدد آن پس از این مقدار می‌باشد. بنابراین، می‌توان نتیجه گرفت که تعداد ۳ خوشه در جمعیت به بهترین شکل موجب توجیه واریانس جمعیت‌ها شده است.

[DOI: 10.29252/rap.8.17.184]

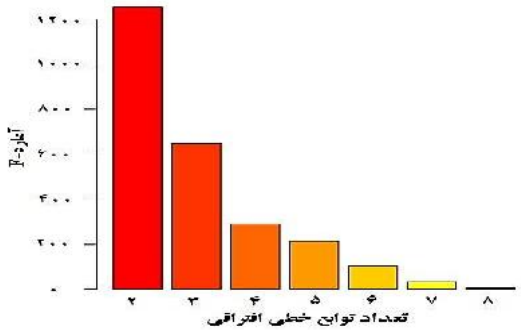


شکل ۲- میزان واریانس تجمعی توجیه شده بر حسب تعداد مؤلفه‌های اصلی باقی مانده
Figure 2. The cumulative variance explained by residual principle components remained in model



شکل ۳- تغییرات مقادیر معیار BIC به ازای تعداد خوشه‌های فرض شده در روش K-means
Figure 3. Variations of BIC criterion per number of clusters assumed in K-means method

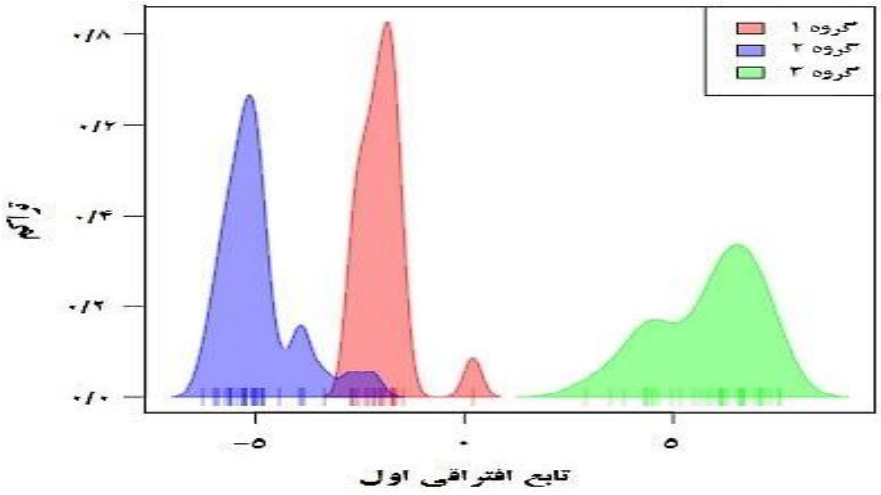
با استفاده از تعداد گروه‌های استنتاج شده از روش K-means (۳ خوشه)، تابع DAPC جهت انجام تحلیل افتراقی مؤلفه‌های اصلی بر روی داده‌ها اجرا گردید. در اینجا نیز داده‌ها ابتدا با انجام PCA تبدیل شدند و آنالیز افتراقی تنها با استفاده از ۲۵ مؤلفه اصلی اول اجرا گردید. بیشترین مقدار آماره F مربوط به دو تابع افتراقی اول بود (شکل ۴) و این توابع جهت اجرای انجام آنالیزهای بعدی انتخاب شدند.



شکل۴- سطوح معنی‌داری ویژه مقدارهای حاصل از تحلیل افتراقی مؤلفه‌های اصلی
Figure 4. Significance levels of eigenvalues obtained from discriminant analysis of principal components

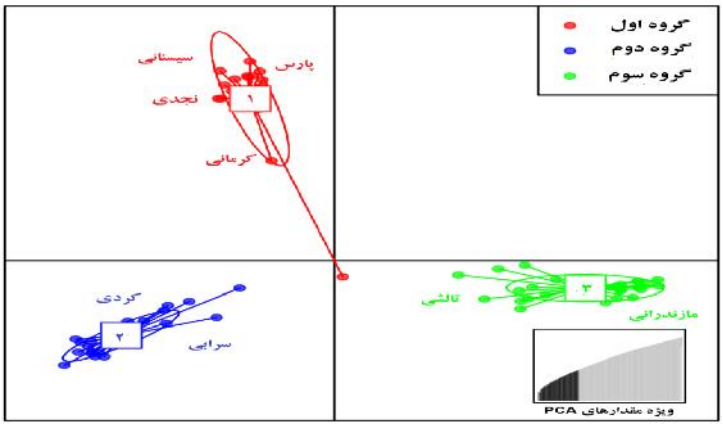
[DOI: 10.29252/rap.8.17.184]

نمودار پراکندگی افراد بر اساس دو تابع افتراقی نخست حاصل از روش DAPC در شکل ۵ نشان داده شده است. همان گونه که در این شکل مشاهده می‌شود سه گروه ژنتیکی اصلی در مجموعه داده‌ها شناسایی شدند. نژادهای مازندرانی و تالشی که در مناطق شمالی حاشیه دریای خزر پراکنده شده‌اند، در یک گروه ژنتیکی یکسان قرار گرفتند. همچنین نژادهای پراکنده شده در نواحی کوهستانی شمال غرب کشور (سرابی و کردی) در دسته یکسانی خوشه‌بندی شدند. از سوی دیگر، نژادهای سیستانی، کرمانی، پارس و نجدی در انطباق با پراکنش جغرافیایی آنها در مناطق جنوبی کشور، در گروه یکسانی دسته‌بندی شدند. علاوه بر این، نمودار تراکم افراد بر حسب تابع افتراقی اول نیز به خوبی وجود سه گروه ژنتیکی را در داده‌ها انعکاس داد (شکل ۶).



شکل ۵- پراکنش افراد بر اساس دو تابع افتراقی اول حاصل از DAPC: افراد به صورت نقاط و گروه‌ها به کمک بیضی‌های کشیده به دور افراد نمایش داده شده‌اند. ویژه مقدارهای حاصل از PCA در شکل فرعی سمت راست نمودار نشان داده‌اند.

Figure 5. Distribution of individuals based on first two discriminant functions: individuals were represented by dots and groups were shown using ovals depicted around individuals. Right hand side sub-shape represented the eigenvalues obtained from PCA



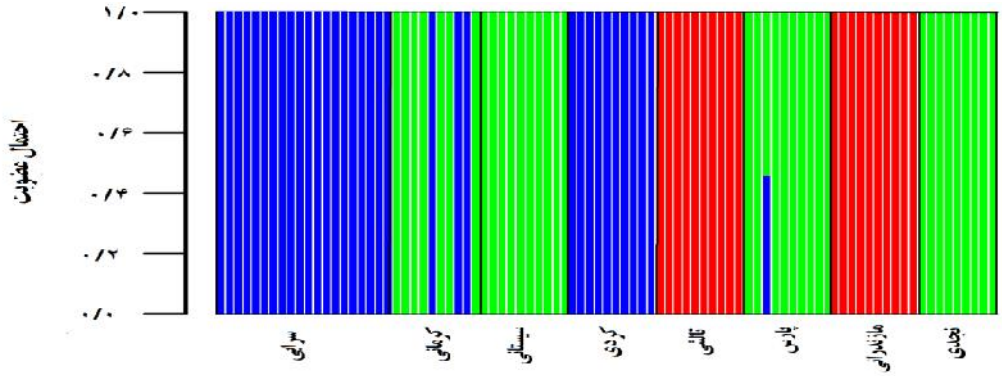
شکل ۶- تراکم افراد بر اساس تابع افتراقی اول: گروه‌های ژنتیکی مختلف با سه رنگ متفاوت نمایش داده شده‌اند.

Figure 6. Density plot of individuals based on first discriminant function: various genetic groups were presented by three different colors

شکل ۷ میزان احتمال عضویت هر یک از فرد را در گروه‌های ژنتیکی شناسایی شده نشان داده است. این شکل ثابت می‌کند که گروه‌های ژنتیکی به خوبی توانستند موجب تفکیک افراد نژادهای مختلف شوند. با این حال، در میان نژادهای کرمانی و پارس افرادی مشاهده شدند که بر خلاف سایر افراد نژاد، عمدتا در گروه مربوط به نژادهای سرابی و کردی (آبی رنگ) عضویت داشتند. از آنجا که نژادهای کردی و سرابی به عنوان نژادهایی با زمینه عمدتا تایورین (بوس تایروس تایروس) شناخته شده‌اند و از سوی دیگر، تلاقی‌های کنترل نشده نژادهای خارجی با نژادهای کرمانی و پارس به گونه‌ای بوده است که تشخیص و یافتن گاوهای خالص را بسیار مشکل کرده است، به نظر می‌رسد این افراد خصوصیات اصلی نژادی را نداشته و تحت تاثیر تلاقی‌های کنترل نشده، عمدتا زمینه تایورین پیدا کرده‌اند (نمونه‌های برون هشته[1]).

1- Outlier samples

[DOI: 10.29252/rap.8.17.184]



شکل ۷- احتمال عضویت هر یک از فرد در گروه‌های مختلف: هر یک از ستون‌ها نماینگر یک فرد است و هر گروه توسط رنگ جداگانه‌ای نشان داده شده است.

Figure 7. Probability of membership for each individual in various groups: each column is representative of one individual and each group is shown by a separate color

روش DAPC پیش از این نیز به منظور بررسی الگوهای آمیختگی موجود در برخی از نژادهای گاو دنیا به کار گرفته شده است. کیم و رتسچیلد (۱۶) از این روش در بررسی الگوهای آمیختگی گاوهای شرق آفریقا بهره گرفتند. نتایج این تحقیق نشان داد که گاوهای نمونه‌گیری شده از مزارع کوچک بیشتر تحت تاثیر اثرات آمیخته‌گری با نژادهای فریزین- هلشتاین، آیرشایر و گرنزی قرار گرفته‌اند. در مطالعه‌ای دیگر، ساختار ژنومیک جمعیت گاوهای Nguni با استفاده از روش DAPC مورد بررسی قرار گرفت و وجود پنج گروه ژنتیکی مختلف در میان ۲۶۲ حیوان مورد بررسی اثبات شد. نتایج این تحقیق اطلاعات جدیدی را در ارتباط با تلاقی‌های انجام شده میان اکوتیپ‌های مختلف نژاد Nguni فراهم آورد (۳۳). علاوه بر این، منشا آمیختگی سه جمعیت اصلی گاوهای بومی تونس نیز به کمک روش DAPC مورد مطالعه قرار گرفته است. در این مطالعه DAPC تفکیک مناسبی را میان نژادهای با منشأ یکسان به وجود آورد. نتایج این تحقیق وجود جریان ژنی از گاوهای جنوب اروپا به گاوهای شمال آفریقا را مورد تایید قرار داد (۲). خوشه‌بندی سلسه مراتبی، خوشه‌بندی مبتنی بر مدل و کاهش ابعاد داده‌ها از جمله روش‌هایی هستند که به طور متداول جهت شناسایی ساختار جمعیت‌ها بر اساس داده‌های حاصل از نشانگرهای مولکولی مورد استفاده قرار گرفته‌اند (۱۸). در این تحقیق از دو روش خوشه‌بندی سلسه مراتبی (K-means) و کاهش ابعاد داده‌ها (DAPC) به صورت متوالی جهت آنالیز تغییرات ژنتیکی جمعیت گاوهای بومی ایران بهره گرفته شد. هرچند پیش از این، استفاده از روش تحلیل مؤلفه‌های اصلی در داده‌های گاوهای بومی ایران موجب تفکیک هر یک از نژادها در خوشه‌های متمایزی شد و میزان شباهت‌ها و تفاوت‌های ژنتیکی بین نژادها را به خوبی انعکاس داد، اما این روش نتوانست تفسیر خاصی از تعداد گروه‌های ژنتیکی موجود در داده‌ها ارائه دهد (۱۴). استنتاج ساختار ژنتیکی جمعیت‌ها در روش‌های مبتنی بر مدل بر اساس در نظر گرفتن یک مدل برای تفرق ژنتیکی جمعیت‌ها صورت می‌پذیرد. کریمی و همکاران (۱۵) از روش بیزی مبتنی بر مدل در نرم‌افزار STRUCRURE جهت بررسی ساختار ژنتیکی جمعیت گاوهای بومی ایران بهره گرفتند. نتایج این تحقیق نشان داد که وجود سه گروه ژنتیکی به بهترین شکل تغییرات موجود در جمعیت‌ها را توجیه می‌کند. در این سطح از خوشه‌بندی (K=۳)، نژادهای سرابی، کردی و سیستانی خوشه‌های متمایزی را به خود اختصاص دادند و سایر نژادها آمیخته‌ای از این سه خوشه بودند.

هرچند در تحقیق حاضر نیز وجود سه گروه ژنتیکی بیشترین تطابق را با تغییرات ژنتیکی جمعیت‌ها داشت، اما نژادهای مازندرانی و تالشی به عنوان یک گروه ژنتیکی متمایز شناخته شدند و نژاد کردی به همراه نژاد سرابی در یک گروه یکسان خوشه‌بندی شد. بیشترین زمینه تایورین در بین نژادهای ایرانی متعلق به دو نژاد سرابی و کردی است و همین امر نیز موجب شد که در تحقیق حاضر این دو نژاد در یک خوشه یکسان طبقه‌بندی شوند. از سوی دیگر، افراد نژادهای سیستانی، کرمانی، پارس و نجدی (نژادهای جنوبی کشور) روابط ژنتیکی نزدیکی را با یکدیگر نمایان ساختند. به نظر می‌رسد نژاد سیستانی که عمدتا دارای زمینه ژنتیکی ایندیسین (بوس تایروس ایندیکوس)[1] است، نژادهای جنوبی کشور را تحت تاثیر خود قرار داده است و همین امر نیز کم شدن فواصل ژنتیکی این نژادها را در پی داشته است. تفکیک ژنتیکی نژادهای با زمینه تایورین و ایندیسین، به عنوان یک تقسیم بندی اولیه و برجسته در بین نژادهای گاو، در سایر مطالعات نیز به خوبی دیده شده است (۲،۶،۱۹). از سوی دیگر، قرار گرفتن دو نژاد مازندرانی و تالشی در یک گروه ژنتیکی یکسان را نیز می‌توان به فاصله ژنتیکی کم و پراکنش جغرافیایی مشابه این دو نژاد نسبت داد. هر چند در پژوهش حاضر، عدم خلوص و آمیختگی‌های موجود در برخی از افراد دو نژاد پارس و کرمانی به کمک خوشه‌بندی ژنتیکی استنتاج شده توسط روش DAPC قابل شناسایی بود (شکل ۷) اما، در مقایسه با روش بیزی مبتنی بر مدل، این روش قادر به ارائه اطلاعات چندانی از سطوح آمیختگی احتمالی در بین سایر نژادها نبود. به همین جهت، به نظر می‌رسد نتایج حاصل از روش بیزی مبتنی بر مدل قادر به ارائه تصویر مناسب‌تری از

[DOI: 10.29252/rap.8.17.184]

سطوح آمیختگی بین نژادها بوده است (۱۵). پومتی و همکاران (۲۵) میزان کارایی روش DAPC را در مقایسه با روش‌های خوشه‌بندی بیزی دو نرم‌افزار STRUCTURE و GEENLAND بررسی کردند. کارایی هر سه روش در استنتاج ساختار ژنتیکی جمعیت‌ها مورد تایید قرار گرفت و همبستگی بالایی میان نتایج حاصل از سه روش مشاهد شد. با این حال، DAPC در بین روش‌های مورد بررسی کمترین زمان محاسباتی را نیاز داشت و استفاده از این روش به عنوان نقطه شروع کار با داده‌های حجیم ژنتیکی پیشنهاد شد.

تطابق میان شباهت‌های ژنتیکی و توزیع جغرافیایی نژادها امری قابل پیش بینی است. با این حال، باید توجه نمود که در بسیاری از موارد تعداد گروه‌های ژنتیکی متمایز در میان جمعیت‌ها مشخص نیست. از سوی دیگر، در بسیاری از موارد شباهت‌های فنوتیپی می‌توانند موجب گمراهی در طبقه‌بندی نژادها شوند. در این مطالعه، بر اساس ویژگی ظاهری داشتن یا نداشتن کوهان، فرض اولیه این بود که نژادهای سرابی، کردی، پارس و کرمانی شباهت بیشتری به گروه تایورین دارند و نژادهای سیستانی، نجدی، مازندرانی و تالشی در گروه ایندیسین قابل طبقه‌بندی هستند. با این حال، نتایج آنالیز ژنتیکی نشان داد که دو نژاد مازندرانی و تالشی در مقایسه با نژادهای پارس و کرمانی زمینه تایورین بیشتری دارند. بنابراین، استفاده از داده‌های ژنتیکی مناسب‌ترین راه حل برای دسته‌بندی نژادها است.

هرچند تعداد نمونه‌های به کار گرفته شده در این مطالعه به دلیل بالا بودن هزینه‌های تعیین ژنوتیپ نسبتا کم است، اما این تعداد نمونه برای انجام آنالیزهای مرتبط با ساختار ژنتیکی جمعیت‌ها کافی می‌باشد. اسمیت و وانگ (۳۱) ثابت کردند که اندازه کوچک نمونه تنها می‌تواند موجب ایجاد اریب کوچکی در برآورد معیارهایی نظیر هتروزیگوسیتی مورد انتظار، F_{ST} جفتی و ساختار جمعیت‌ها شود، اما میزان این اریب در مورد معیاری مانند تنوع آللی بسیار بیشتر است. ویلینگ و همکاران (۳۴) گزارش کرده‌اند که به هنگام برآورد آماره F_{ST} در صورت استفاده از یک برآوردگر مناسب و در حضور تعداد زیادی از نشانگرهای ژنتیکی دو آللی، اندازه نمونه می‌تواند به طور معنی‌داری کاهش داده شود (۴ تا ۶ نمونه).

سازگاری با محیط زندگی در طی سالیان دراز موجب شکل‌گیری گروه‌های ژنتیکی متمایزی در میان گاوهای بومی ایران شده است و این امر لزوم توجه به حفاظت از این منابع ژنتیکی را روشن‌تر می‌سازد. اندازه جمعیت گاوهای بومی کشور در سال‌های اخیر روندی رو به کاهش داشته است و در حال حاضر بسیاری از نژادهای گاو بومی ایران در معرض خطر انقراض قرار گرفته‌اند (۱۴). شناخت ساختار ژنتیکی و روابط بین این نژادها می‌تواند در طراحی برنامه‌های حفاظت ژنتیکی نقش مؤثری ایفا کند. نژادهای دارای شباهت ژنتیکی بیشتر، می‌توانند در برنامه‌های حفاظت ژنتیکی یکدیگر مشارکت داشته باشند. از سوی دیگر، می‌بایست اولویت حفاظتی بیشتری به نژادهایی داده شود که دارای ترکیب ژنتیکی متمایزتری هستند. به همین جهت، نتایج پژوهش حاضر می‌تواند در طراحی برنامه‌های حفاظت ژنتیکی در گاوهای بومی ایران به کار گرفته شود.

تشکر و قدردانی

بدین وسیله از کلیه دامداران و کارکنان ایستگاه‌های اصلاح‌نژاد گاوهای بومی کشور به پاس همکاری ایشان در تهیه نمونه‌های مورد مطالعه تشکر می‌گردد.

منابع

1. Bartenhagen, C., H.U. Klein, C. Ruckert, X. Jiang and M. Dugas. 2010. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. BMC Bioinformatics, 11: 1-11.
2. Ben Jemaa, S., M. Boussaha, M. Ben Mehdi, J.H. Lee and S.H. Lee. 2015. Genome-wide insights into population structure and genetic history of tunisian local cattle using the illumina bovinesnp50 beadchip. BMC Genomics, 16: 677.
3. Boettcher, P.J., M. Tixier-Boichard, M.A. Toro, H. Simianer, H. Eding, G. Gandini, S. Joost, D. Garcia, L. Colli, P. Ajmone-Marsan and G. Consortium. 2010. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. Animal Genetics, 41: 64-77.
4. Burgos-Paz, W., C.A. Souza, A. Castello, A. Mercade, N. Okumura, I.N. Sheremet'eva and M. Perez-Enciso. 2013. Worldwide genetic relationships of pigs as inferred from X chromosome SNPs. Animal Genetics, 44(2): 130-138.
5. Filippi, C.V., A. Natalia, J.G. Rivas, J. Zubrzycki, A. Puebla, D. Cordes, M.V. Moreno, C.M. Fusari, D. Alvarez, R.A. Heinz, H.E. Hopp, N.B. Paniego and V.V. Lia. 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers .BMC Plant Biology, 15: 52.
6. Decker, J.E., S.D. McKay, M.M. Rolf, J. Kim, A. Molina Alcala, T.S. Sonstegard, O. Hanotte, A. Gotherstrom, C.M. Seabury, L. Praharani, M. Saif-Ur-Rehman, R.D. Schnabel and J.F. Taylor. 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genetics, 10: e1004254.
7. Degner, J.F., J.C. Marioni, A.A. Pai, J.K. Pickrell, E. Nkadori, Y. Gilad and J.K. Pritchard. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics, 25(24): 3207-3212.
8. Dell'Acqua, M., D.M. Gatti, G. Pea, F. Cattonaro, F. Coppens, G. Magris and M.E. Pè. 2015. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biology, 16(1): 1-23.

1- Taurine (Bos taurus taurus)

[DOI: 10.29252/rap.8.17.184]

9. Heitlinger, E., H. Taraschewski, U. Weclawski, K. Gharbi and M. Blaxter. 2014. Transcriptome analyses of Anguillicola crassus from native and novel hosts. PeerJ, 2: e684.
10. Ji, H., X. Li, Q.f. Wang and Y. Ning. 2013. Differential principal component analysis of ChIP-seq. Proceedings of the National Academy of Sciences, 110(17): 6789-6794.
11. Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403-1405.
12. Jombart, T., S. Devillard and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11: 1-15.
13. Jombart, T. and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27: 3070-3071.
14. Karimi, K., A. Esmailizadeh Koshkoiyeh, M. Asadi Fuzi, L.R. Porto-Neto and C. Gondro. 2015. Prioritization for conservation of Iranian native cattle breeds based on genome-wide SNP data. Conservation Genetics, 17: 77-89.
15. Karimi, K., A. Esmailizadeh Koshkoiyeh and M. AsadiFuzi. 2015. Analysis of genetic structure of Iranian indigenous cattle populations using dense single nucleotide polymorphism markers. Animal Production Research, 4(3): 93-104.
16. Kim, E-S. and M.F. Rothschild. 2014. Genomic adaptation of admixed dairy cattle in East Africa. Frontiers in Genetics, 5
17. Kim, E-S., R. Ros-Freixedes, R.N. Pena, T.J. Baas, J. Estany and M.F. Rothschild. 2015. Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations. Journal of Animal Science, 93: 3292-3302.
18. Lawson, D.J. and D. Falush. 2012. Population identification using genetic data. Annual Review of Genomics and Human Genetics, 13: 337-361.
19. Lin, B.Z., S. Sasazaki and H. Mannen. 2010. Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. Animal Science Journal, 81: 281-289.
20. McVean, G. 2009. A genealogical interpretation of principal components analysis. PLoS Genetics, 5:e1000686.
21. Nei, M. 1972. Genetic distance between populations. American Naturalist, 106: 283-292.
22. Paradis, E., J. Claude and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics, 20: 289-290.
23. Patterson, N., A.L. Price and D. Reich. 2006. Population Structure and Eigenanalysis. PLoS Genetics, 2:e190.
24. Pembleton, L.W., N.O. Cogan and J.W. Forster. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13: 946-952.
25. Pometti, C.L., C.F. Bessega, B.O. Saidman and J.C. Vilardi. 2014. Analysis of genetic population structure in Acacia caven (Leguminosae, Mimosoideae), comparing one exploratory and two Bayesian-model-based methods. Genetics and Molecular Biology, 37: 64-72.
26. Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81: 559-575.
27. Reynolds, J., B.S. Weir and C.C. Cockerham. 1983. Estimation of the coancestry coefficient: Basis for a short term genetic distance. Genetics, 105: 767-779.
28. Ringner, M. 2008. What is principal component analysis?. Nature Biotechnology, 26: 303-304.
29. Taberlet, P., E. Coissac, J. Pansu and F. Pompanon. 2011. Conservation genetics of cattle, sheep, and goats. Comptes Rendus Biologies, 334: 247-254.
30. 30. Scheu, A., A. Powell, R. Bollongino, J.D. Vigne, A. Tresset, C. Çakırlar and J. Burger. 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genetics, 16(1): 1-11.
31. Smith, O. and J. Wang. 2014. When can noninvasive samples provide sufficient information in conservation genetics studies?. Molecular Ecology Resources, 14: 1011-1023.
32. Tavakolian, j. 1999. Introduction on genetic resources of indigenous animals and poultry in Iran. Iranian Animal Research Institute, Karaj, Iran, 3-45 pp.
33. Wang, M.D., K. Dzama, C.A. Hefer and F.C. Muchadeyi. 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics, 16(1): 1-16. doi: 10.1186/s12864-015-2122-z.
34. Willing, E., C. Dreyer and C. Van Oosterhout. 2012. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many snp markers. PLoS One, 7(8): e42649.
35. Xie, J., R. Li, S. Li, X. Ran, J. Wang, J. Jiang, and P. Zhao. 2016. Identification of Copy Number Variations in Xiang and Kele Pigs. PLoS One, 11(2): e0148565.
36. Xu, H-M., X-W. Sun, T. Qi, W.Y. Lin, N. Liu and X.Y. Lou. 2014. Multivariate dimensionality reduction approaches to identify gene-gene and gene-environment interactions underlying multiple complex traits. PLoS One, 9: e108103.

[DOI: 10.29252/rap.8.17.184]

Investigation of the Population Structure in Iranian Native Cattle using Discriminant Analysis of Principal Components

Karim Karimi

Ph.D. gaduate of Animal Breeding and Genetics, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Member of Young Researchers, Shahid Bahonar University of Kerman, Iran
(Corresponding Author: Karim.Karimi81@gmail.com)
Received: April 24, 2016 Accepted: February 12, 2017

Abstract

Effective management of genetic resources in the domestic animals is based on characterization of genetic structure and diversity among populations. Strategies reducing complexity and dimensions of data are required to analyze the genetic relationships between populations based on dense genomic data. The objective of this study was to use the discriminant analysis of principal components (DAPC) for investigating the genetic structure of the Iranian native cattle based on a high-density single nucleotide polymorphism data. Samples were genotyped using Illumina BovineHD SNP chip containing 777962 SNPs. Genetic distances among different breeds were determined based on allele frequencies of populations. Discriminant analysis of principal components was performed using "adegenet" package in the R software. Pars and Kermani breeds had the lowest genetic distances among studied breeds and the highest genetic distances were observed between Kurdi and Sistani breeds. Three numbers of genetic clusters were selected as the best number of inferred groups to perform discriminant analysis of principal components. Results from this study confirmed that there were three main genetic groups among Iranian native cattle. In accordance with geographical distribution, Mazandarani and Taleshi breeds were classified as the same cluster. In addition, other two distinct genetic groups included breeds distributed on the North West mountainous area of the country (Sarabi and Kurdi) and those can be found in the Southern area (Sistani, Kermani, Najdi and Pars). Discriminant analysis of principal components could well distinguish individuals from different genetic groups. Genetic structure identified in the current study can be applied to design the genetic conservation programs for Iranian native cattle.

Keywords: Discriminant analysis of principal components, Data dimensionality reduction, Iranian native cattle, Single nucleotide polymorphism

[DOI: 10.29252/rap.8.17.184]