نشریه علوم دامی
(پژوهش و سازندگی)

شماره ۱۲۱، زمستان ۱۳۹۷
ص‌ص: ۶۷~۷۶

بررسی همگنی ساختار ژنتیکی در جمعیت گوسفند نژاد زندی با استفاده از داده‌های ژنومی

- **حمیدرضا کورش‌نیا**
  دانش آموخته کارشناسی ارشد ژنتیک و اصلاح نژاد دام گروه مهندسی علوم دامی –دانشکده علوم و مهندسی کشاورزی-پردیس کشاورزی و منابع طبیعی دانشگاه تهران
- **حسین مرادی شهربابک** (نویسنده مسئول)
  استادیار بخش ژنتیک و اصلاح نژاد دام گروه مهندسی علوم دامی –دانشکده علوم و مهندسی کشاورزی-پردیس کشاورزی و منابع طبیعی دانشگاه تهران.
- **مصطفی صادقی**
  دانشیار بخش ژنتیک و اصلاح نژاد دام گروه مهندسی علوم دامی –دانشکده علوم و مهندسی کشاورزی-پردیس کشاورزی و منابع طبیعی دانشگاه تهران

تاریخ دریافت : شهریور ۱۳۹۶ تاریخ پذیرش : بهمن ۱۳۹۶

شماره تماس نویسنده مسئول: ۰۹۱۳۳۹۱۵۳۰۶

Email: hmoradis@ut.ac.ir

شناسه دیجیتال (DOI): 10.22092/asj.2018.115547.1539

چکیده

داده‌های ژنومی می‌تواند ما را به چگونگی شکل‌گیری نژادها و جمعیت‌ها و روند تأثیرگذاری رخدادهای ژنتیکی هرچند کمیاب در گذر زمان رهنمون سازد. از موارد بسیار ارزشمند جهت حفظ ذخایر ژنتیکی که خود موضوعی پر اهمیت تلقی می‌گردد و همچنین بهبود برنامه‌های اصلاح نژادی، پی‌بردن به ساختار ژنتیکی جوامع مورد مطالعه است. جهت مطالعه ساختار ژنتیکی جمعیت گوسفند زندی واقع در ایستگاه پرورش و اصلاح نژاد گوسفند زندی تهران از روش آنالیز تفکیکی مؤلفه‌های اصلی (DAPC) و همچنین از آنالیز مؤلفه‌های اصلی (PCA) استفاده شد .پس از۹۹ رأس گوسفند نژاد زندی خونگیری و تعیین ژنوتیپ با تراشه‌های اسنیپ ۵۰K شرکت ایلومینا صورت گرفت. روش تجزیه‌ی تفکیکی مؤلفه‌های اصلی، به وضوح ساختار ژنتیکی جمعیت مورد مطالعه و تفکیک حیوانات را در دو گروه نشان داد که می‌تواند ناشی از حساسیت روش DAPC باشد که قادر به بررسی همگنی واریانس در جوامع حیوانی است. در روش DAPC، برای ارزیابی تعداد بهینه‌ی خوشه با معیار BIC، k=۲ بهترین نتیجه را نشان داد. بررسی نتایج حاصل جهت حفظ تعداد مؤلفه‌ی اصلی برای آنالیز تفکیکی، ۳۱ مؤلفه‌ی اول را تعداد بهینه‌ی مؤلفه برای مراحل بعدی آنالیز در نظر گرفت. با توجه به اهمیت در نظر گرفتن واریانس درون گروهی و همچنین ساختار ژنتیکی جوامع جهت آنالیزهای مهم ژنومی مشخص شد که روش DAPC در مطالعه‌ی ساختار ژنتیکی گوسفند زندی به دلیل در نظر گرفتن تعداد مؤلفهی بیشتر و متعاقباً افزایش واریانس در نظر گرفته شده نسبت به روش PCA کاراتر می باشد.

واژه‌های کلیدی: گوسفند زندی، ساختار ژنتیکی جمعیت، PCA، DAPC، تراشه‌های اسنیپ

Animal Science Journal (Pajouhesh & Sazandegi) No 121 pp: 67-76

The effect of *Prosopis farcta* on the performance, some blood parameters, immune and antioxidant system of broiler chickens under heat stress conditions.

By: Hamidreza Kouroshnia[1], Hossein MoradiShahrbabak[2*], Mostafa Sadeghi[3]

[1]MSc in Genetics and Animal Breeding, [2]Assistant Professor and [3]Associate Professor of Genetics and Animal Breeding Province, Department of Animal Science Engineering, Faculty of Agricultural Science and Engineering, College of Agriculture and Natural Resources, University of Tehran, [*]responsible author's Email: hmoradis@ut.ac.ir

Received: September 2017 **Accepted: February 2018**

Genomic data can lead us to how of breeds and populations forming and process of genetic occurrences effecting even infrequent cases of them at time passage. A very worthy instance to genetic resources preservation that itself is counted an important subject, and improvement of animal breeding plans, is genetic structure discovery of studying populations. For investigation about genetic structure homogeneity in Zandi breed sheep population located in Tehran Zandi sheep breeding station, discriminant analysis of principal components (DAPC) and in inside of this analysis, principal components analysis (PCA) were performed. For this purpose, 99 heads of sheep were bled and genotyped using Illumina SNP chip 50K. Discriminant analysis of principal component obviously showed the genetic structure of studying population and its members separation to two groups that it may be arising from DAPC high sensitiveness that is able to investigate variance homogeneity in animal populations. In DAPC, to evaluate the clusters optimized number using BIC, k=2 was shown as best result. Investigation of results to maintain the number of principal components for discriminant analysis, takes into account that the first 31 components is the optimized number of component for analysis next steps. Re the importance of within group variance contemplating and also populations genetic structure for important genomic analyses, became clear that DAPC in study of Zandi population genetic structure, because that contemplates of more principal component number and subsequently increasing of contemplated variance, is more effective than PCA.

Key words: Zandi sheep, population genetic structure, PCA, DAPC, SNP chips.

مقدمه

گوسفند زندی یکی از ذخایر ژنتیکی با ارزش ایران و بالطبع دارای سازگاری بالایی با محیط و مقاومت در برابر بیماری‌ها و پیرو آن هزینه‌های نگهداری پایین است و استفاده از مواد خشبی کم ارزش و ضایعات کشاورزی و و داشتن آمار دوقلوزایی، این نژاد را مورد توجه برنامه‌های اصلاح نژاد قرار داده است.

نشان‌گرهای مولکولی جهت بررسی تفاوت ژنتیکی بین جمعیت‌ها و حیوانات حفاظت‌شده، بسیار مورد توجه قرار گرفته‌اند. پلی‌مورفیسم نشانگرهای ژنتیکی، در مطالعه‌ی جمعیت‌های مختلف و تفاوت‌های ژنتیکی بین آن‌ها مورد استفاده است. تکنیک‌های مولکولی جدید، به همراه پیشرفت‌های ژنتیک آماری، چشم‌اندازی مطبوع برای تحقیقات در زمینه‌ی نقشه‌یابی QTL و تنوع صفات است (محمدآبادی و همکاران، ۲۰۰۹). استنتاج ساختار جمعیت از مارکرهای ژنتیکی، در شرایط گوناگون مثل مطالعات ارتباطی و تکاملی، دسته‌بندی زیرگونه‌ها و تعیین موانع ژنتیکی مفید است (لیو و ژائو، ۲۰۰۶). بررسی منشأهای تاریخی پیچیده نشان می‌دهد انتخاب طبیعی و مصنوعی، منجر به اختلاف

متعدد نژادها با تنوع فنوتیپی گسترده در یک دوره‌ی کوتاه زمانی شده است (اِپس و همکاران، ۲۰۱۳). پیشرفت‌های فناوری توالی‌یابی، باعث ایجاد انبوهی از داده‌های ژنتیک و در نتیجه مطالعه‌ی دقیق‌تر جوامع دامی شده است و داده‌های ژنومی می‌تواند ما را به چگونگی شکل‌گیری نژادها و جمعیت‌ها و روند تأثیرگذاری رخدادهای ژنتیکی هرچند کمیاب در گذر زمان رهنمون سازد.

از موارد بسیار ارزشمند جهت حفظ ذخایر ژنتیکی که خود موضوعی پر اهمیت تلقی می‌گردد و همچنین بهبود برنامه‌های اصلاح نژادی پی‌بردن به ساختار ژنتیکی جوامع مورد مطالعه است. این مورد از آن لحاظ حائز اهمیت است که زیر جمعیت‌های دارای ساختار متفاوت درون جوامع مورد مطالعه باعث ایجاد اریب در مطالعات پویش کل ژنوم (GWAS)[1] میشود (واکُلدِر و همکاران، ۲۰۰۲؛ توماس و ویت، ۲۰۰۲) و افرادی با پس زمینه های ژنتیک متفاوت که در لایه‌های مختلف جمعیتی قرار می‌گیرند درون یک جمعیت به دلیل ایجاد اشتباه نوع اول، چالشی برای GWAS محسوب می‌شود. چرا که در GWAS فرض بر همگنی جامعه است که این فرض می‌تواند به آسانی نقض شود (مارچینی و همکاران، ۲۰۰۴).

برای بررسی و تحلیل ساختار جمعیتی با استفاده از حجم بالای داده‌های ژنومی ضروری است که ابزاری مناسب بر اساس روشی کارآمد از میان روش‌های گوناگون استفاده شود. شناخت ساختار ژنتیکی و شباهت‌ها و تفاوت‌های موجود میان نژادها و همچنین افراد درون یک نژاد ضمن آشکار ساختن ارزش حفاظتی آنها می‌تواند اولویت هر یک از برنامه‌های حفاظت ژنتیکی را نیز پیشنهاد دهد. روش‌های متعددی برای تعیین ساختار ژنتیکی و تعیین گروه‌های نژادی وجود دارد.

تجزیه‌ی تفکیکی مؤلفه‌های اصلی (DAPC)[2] با تمرکز بر روی تنوع بین گروه‌ها و نادیده گرفتن تنوع داخل گروهی برای ارزیابی ارتباط بین گروه‌ها، روش مناسبی است. تجزیه‌ی تفکیکی مؤلفه-های اصلی، روش چند متغییره است که برای شناسایی و توصیف خوشه‌های افرادی که به طور ژنتیکی با هم ارتباط دارند، طراحی شده است. در این روش ابتدا داده‌ها به منظور اطمینان از اینکه متغییرها برای ورود به مرحله‌ی آنالیز تفکیکی (DA)[3] مستقل از هم هستند، بدون از دست رفتن اطلاعات، با استفاده از تجزیه‌ی مؤلفه‌های اصلی (PCA)[4] (پَترسُن و همکاران، ۲۰۰۶؛ پُرایس و همکاران، ۲۰۰۶) تبدیل می‌شوند و سپس خوشه‌ها با استفاده از آنالیز تفکیکی شناسایی می‌شوند (جُمبارت و همکاران، ۲۰۱۰).

آنالیز مؤلفه‌ی اصلی ابزار استانداری در ژنتیک جمعیت است که در مطالعه‌ی جمعیت‌های اروپائی و هندی استفاده شده است (لائو و همکاران، ۲۰۰۸). این روش، روشی غیرپارامتری است که از گذشته به عنوان یک تکنیک کاهش خطی ابعاد جهت نمایان سازی ساختارهای جمعیتی استفاده می شود. در این روش تمام مولفه‌های اصلی مطابق با واریانسی که توجیه می‌کنند مرتب می-شوند (لاُلئی و همکاران، ۲۰۰۷). همچنین در کنترل کیفیت در مطالعات ژنتیکی استفاده می‌شود. در عمل PCA به عنوان جایگزینی برای الگوریتم‌های خوشه‌بندی بیزی پیشنهاد شده است (لیو و ژائو، ۲۰۰۶؛ پُرایس و همکاران، ۲۰۰۶؛ لی و همکاران، ۲۰۰۹). مزیت اصلی PCA توانایی آن برای شناسایی ساختارهای ژنتیکی مجموعه‌داده‌های بزرگ با زمان محاسباتی قابل اغماض و عدم وجود فرضیه‌هایی درباره‌ی مدل ژنتیک جمعیتی می باشد. با این حال، PCA فاقد برخی از ویژگی‌های ضروری بررسی ساختار جمعیت‌های بیولوژیکی است. PCA یک ارزیابی گروهی را فراهم نمی‌سازد و نیازمند تعریف پیشینی از خوشه‌ها برای مطالعه‌ی ساختار جمعیتی است و حتی برای بدست آوردن تصویر واضح از تنوع بین جمعیت‌ها مناسب نیست (جُمبارت و همکاران، ۲۰۱۰). تجزیه‌ی مؤلفه های اصلی در این تلاش است که تنوع کلی میان افراد که شامل هر دو تنوع بین گروه‌ها و تنوع داخل گروه‌ها می‌شود را خلاصه نماید.

اُلانو-مارین و همکاران در سال ۲۰۱۴ با استفاده از روش DAPC، نشان دادند که تفاوت ژنتیکی چندانی بر اساس نشانگرهای میکروستلایت میان سه جمعیت گوزن کوچک

[1] Genome Wide Association Studies

[2] Discriminant Analysis of Principal Components

[3] Discriminant Analysis

[4] Principal Components Analysis

اروپایی در سه منطقه‌ی متفاوت با توجه به تنوع ژنتیکی بالا وجود ندارد.

روش DAPC ، جمعیت سمور اروپایی را از نظر ساختار ژنتیکی حاصل از نتایج ژنوتایپ ۲۳ نشان‌گر میکروستلایت و یک توالی ۶۲۱ جفت‌بازی mtDNA، به بیشترین تعداد گروه نسبت به سایر روش‌ها تفکیک نمود هرچند یک گروه نسبت به دیگر گروه‌ها بیشتر متمایز بود (وِرگارا و همکاران، ۲۰۱۵).

در بررسی ساختار ژنتیکی گله‌های گاو شیری تایلند بر اساس تراشه‌های اسنیپ 9K، لائودیم و همکاران (۲۰۱۷) به این نتیجه رسیدند که روش DAPC تفاوت ژنتیکی میان نمونه‌ها را بر اساس فراوانی آللی بدون در نظرگرفتن نژاد اصلی آنها تعیین می‌کند.

روش تجزیه‌ی تفکیکی مؤلفه‌های اصلی هم می‌تواند به عنوان روش بدون نظارت و هم با نظارت عمل کند. این روش زمانی که گروه‌های پیشین وجود ندارند، از k-means پی‌درپی و انتخاب مدل را برای تفکیک ژنتیکی خوشه‌ها استفاده می‌کند. k-means تعداد گروه‌ها(k)یی را که تنوع بین گروه‌ها را حداکثر می‌کند پیدا می‌کند. DAPC تفکیک افراد را به گروه‌های از پیش تعیین شده امکان‌پذیر می‌سازد و همچنین این روش احتمالاتی را برای انتساب افراد به هر گروه فراهم می‌آورد. این روش را میتوان به داده‌های با حجم زیاد، هزاران مارکر از هزاران فرد اعمال کرد تجزیه‌ی تفکیکی مؤلفه‌های اصلی، ارزیابی بصری از اختلاف بین جمعیت‌ها را فراهم می‌کند. این رویکرد سریع‌تر از روش‌های خوشه بندی بیزی بوده و با طیف وسیعی از داده ها قابل اجرا است (جُمبارت و همکاران، ۲۰۱۰). یکی از مزیت های DAPC ، تطبیق پذیری بالای آن است در واقع DAPC به مدل ژنتیک جمعیت خاصی وابسته نیست و در نتیجه فرض تعادل هاردی واینبرگ و تعادل لینکاژی ندارد. به این ترتیب برای انواع موجودات صرف نظر از پلوئیدی بودن و نرخ نوترکیبی آن‌ها، کاربرد دارد. همچنین این روش محدود به داده‌ی ژنومیکی نبوده و برای داده های کمی دیگر مثل داده‌های مورفولوژی نیز قابل اجرا است (جُمبارت و کُلینز، ۲۰۱۵). هدف این تحقیق بررسی همگنی ساختار جمعیت گوسفند زندی ایستگاه پرورش و اصلاح نژاد گوسفند زندی تهران با استفاده از داده‌های SNP-Chip 50K و همچنین تفکیک حیوانات در صورت وجود تفاوت‌های ساختاری ژنتیکی در گروه‌های مختلف از راه مسئله‌ی یادگیری با نظارت به روش DAPC است.

مواد و روش‌ها

نمونه‌گیری و تعیین ژنوتیپ دام‌ها

داده‌های این تحقیق از طرح ملی ژنومیک گوسفند زندی با نظارت مرکز اصلاح نژاد دام و بهبود تولیدات دامی کشور و سازمان جهاد کشاورزی استان تهران که توسط شرکت دانش‌بنیان "ساینا‌گستر البرز" اجرا شده، فراهم گردیده است. نمونه‌ها از گله‌های تحت سیستم ثبت شجره و رکورد ایستگاه پرورش و اصلاح نژاد گوسفند زندی تهران که از سال ۱۳۷۰ با گردآوری قوچ و میش برتر این نژاد از گله‌های مردمی در تهران، قم و مرکزی آغاز به کار کرده‌است واقع در طول جغرافیایی '۴۳و۵۱° شرقی و عرض جغرافیایی '۱۸و۳۵° شمالی می باشد. در کل از ۹۹ رأس میش شامل ۵۲ رأس میش دوقلوزا و ۴۷ رأس میش تک‌قلوزا به طور تصادفی خون‌گیری انجام شد. و نحوه‌ی تعیین چندقلوزایی بر اساس معیار نسبت تعداد بره به تعداد زایش بود اگر این نسبت بزرگتر از ۱/۳ باشد به عنوان میش چندقلوزا در نظر گرفته شده‌است و کمتر یا مساوی این نسبت، به عنوان حیوان تک‌قلوزا در نظر گرفته شد. استخراج DNA ژنومیک از خون با روش بهینه‌ی نمکی (گِریمبِرگ و همکاران، ۱۹۸۹) انجام شد و پس از استخراج DNA در آزمایشگاه علوم دامی دانشگاه تهران و بررسی کیفیت نمونه‌های مورد نظر، نمونه‌ها جهت انجام مراحل بعدی توالی‌یابی به آزمایشگاه ژنومیک دانشگاه کنتاکی[5] کشور آمریکا منتقل شدند. سپس نمونه‌ها با استفاده از تراشه‌های ۵۰K شرکت ایلومینا[6] مخصوص گوسفند که تنها نمونه‌ی تجاری قابل دسترسی در ایران است و بیشترین پوشش ژنوم را نسبت به دیگر نشانگرها دارد تعیین ژنوتیپ شدند.

[1] University of Kentucky

[2] Illumina

مراحل فیلتراسیون داده‌های حاصل از تعیین ژنوتیپ جهت انجام آنالیزهای نهایی

برای اطمینان از کیفیت داده‌های حاصل از تعیین ژنوتیپ، در آنالیزهای نهایی مراحل مختلف فیلتراسیون بر روی داده های حاصل از کنترل کیفیت اولیه با استفاده از نرم افزار Plink، اعمال شد. به این دلیل که نمونه‌های با کیفیت پایین با احتمال بیشتری با داده های گمشده همراه هستند و منجر به افزایش خطای ژنوتیپ می‌شود (بَرِندز و همکاران، ۲۰۰۹) در ابتدا حیوانات دارای بیش از ٪۵ ژنوتیپ از دست رفته از آنالیزهای بعدی کنار گذاشته شد. دو فاکتور حداقل فراوانی آللی(MAF)[۷] و درصد نمونه‌هایی که برای آن نشانگر ژنوتایپ شده‌اند (نرخ خوانش)[۸] برای هر اسنیپ محاسبه شدند و اسنیپ‌هایی که دارای نرخ خوانش و MAF به ترتیب کمتر از ٪۹۵ و ٪۱ بودند، حذف شدند. برای اسنیپ‌های باقی‌مانده در صورت عدم تعادل هاردی–واینبرگ به عنوان معیاری از خطای تعیین ژنوتیپ کنار گذاشته شدند (تِئو و همکاران، ۲۰۰۷). تعادل هاردی–واینبرگ در کلیه جایگاه‌ها مورد بررسی قرار گرفت(). برای تعیین سطح معنی =>P× ۱ 10^{-6}) استفاده $\beta=\alpha/n$داری مطلوب در این آزمون از تصحیح بنفرونی(پس از تعیین ژنوتیپ، با عمل فیلتراسیون توضیح داده شده، .شد اسنیپ‌های منتخب وارد مرحله دیگر آنالیز شدند. در نهایت داده‌ها ژنومی جهت آنالیز از فرمت داده‌های آللی به فرمت داده‌های تبدیل شد.Plinkژنوتیپی(۰،۱،۲) با نرم افزار

آماده سازی داده‌ها برای DAPC با نظارت

یک روش آماده‌سازی مجموعه‌داده‌ها، تقسیم کل نمونه‌ها به دو دسته‌ی آموزش و آزمایش به صورت تصادفی است. راهکار دیگری که برای آماده‌سازی داده‌ها وجود دارد استفاده از روش اعتبار سنجی k مرتبه (۱ تا k مرتبه) می‌باشد. در این روش کل داده‌ها به k دسته‌ی تقریباً هم‌اندازه تقسیم می‌شوند. ۱-k دسته برای آموزش مدل و دسته باقیمانده برای آزمایش مدل به کار می‌رود .به این ترتیب، به تعداد k مرتبه، مدل، آموزش و آزمایش می‌شود. مزیت این روش این است که سرانجام همه‌ی نمونه‌های موجود، در هر دو فرآیند آموزش و آزمایش شرکت می‌کنند. برای گسترش یک مدل ماشین یادگیری با نظارت، داده‌ها به دو دسته‌ی آموزش و آزمایش تقسیم بندی می‌شوند تا مدل مورد نظر توسط داده‌های مجموعه‌ی آموزش، آموزش داده شود و کارایی مدل در پیش بینی کمیت مورد نظر به کمک داده‌هایی که در طول آموزش مدل، توسط مدل تجربه نشده‌اند (مجموعه‌داده‌های آزمایش)، بررسی گردد.

آنالیزهای آماری

آنالیز DAPC و در بطن آن PCA با پکیج adegenet اجرا شد. برای شناسایی تعداد بهینه‌ی خوشه‌ها، k-means به صورت متوالی با افزایش مقادیر k، اجرا می‌شود و جواب‌های خوشه‌بندی مختلف با استفاده از معیار اطلاعات بیزی (BIC)[۹] مقایسه شد. در حالت ایده‌آل، خوشه‌ی بهینه بایستی BIC پایین تری داشته باشد در عمل بهترین BIC اغلب با خمیدگی در منحنی مقادیر BIC به عنوان تابعی از k نشان داده شد. در این روش خوشه‌ها با تابع find.cluster بدست می‌آیند که این تابع ابتدا داده‌ها را با PCA تبدیل می‌کند، سپس الگوریتم k-means با افزایش مقادیر k اجرا می‌شود. برای انتخاب تعداد بهینه PCهای نگه داشته شده جهت آنالیز تفکیکی از پیام‌های متقابل آن مبنی بر ناپایداری نتایج، استفاده شد.

[۱] Bayesian Information Criterion

[۳] Minor Allele Frequency

[۴] Call Rate

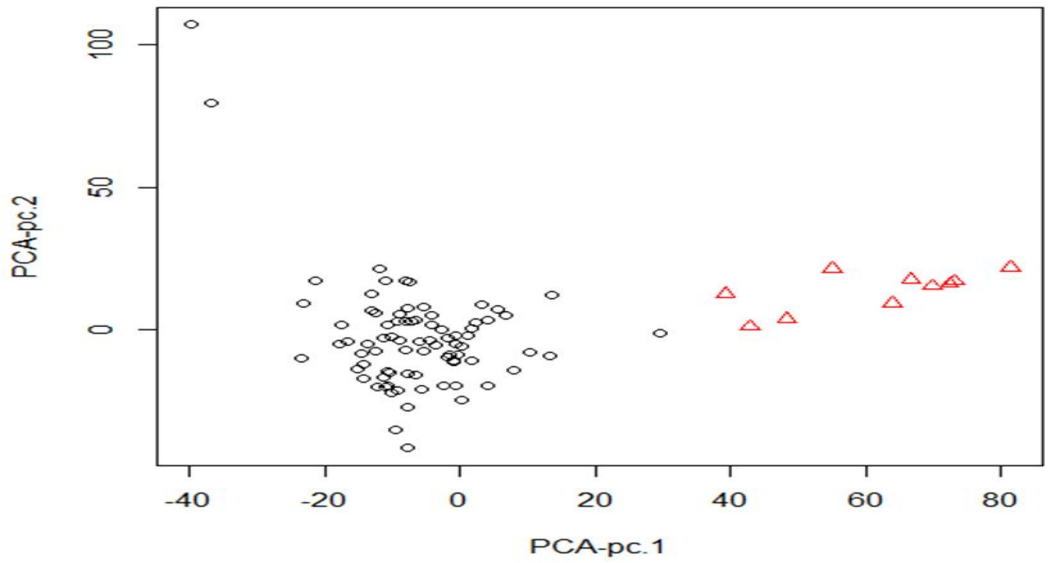
نتایج و بحث

کنترل کیفیت و فیلتراسیون داده‌های ژنومی

کنترل کیفیت روی ۵۴۲۴۱ اسنیپ بدست آمده از کنترل کیفیت اولیه اجرا شد که اسنیپی به دلیل موقعیت ناشناخته حذف نشد و در مجموع تعداد ۳۰۲۷ اسنیپ به دلیل MAF کمتر از ٪۱ ، ۴۶ اسنیپ به دلیل انحراف از تعادل هاردی-وینبرگ در سطح ٪۵ و ۱۳۸۹ اسنیپ به دلیل نرخ خوانش SNP کمتر از ٪۹۵ و ۴ حیوان به دلیل نرخ خوانش حیوان کمتر از ٪۹۵ حذف شدند. در نهایت ۹۵ حیوان با ۵۰۶۴۹ اسنیپ، مراحل کنترل کیفیت را با MAF>0.01 و call rate>0.95 گذراندند و همه‌ی اسنیپ-های باقی مانده در تعادل هاردی-واینبرگ بودند.

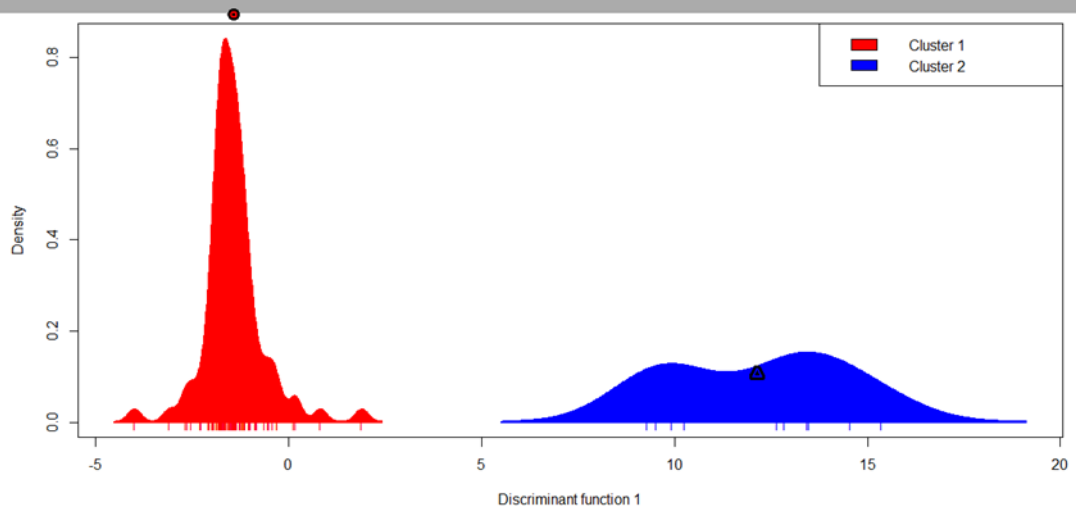
آنالیز آماری

روش PCA، مؤلفه‌های اصلی را که ساختار جمعیت را براساس همبستگی ژنتیکی میان افراد بیان می‌کند، شناسایی می‌کند. برای ارزیابی اختلاف ژنتیکی میان جمعیت نمودار حاصل از PCA(شکل ۱) ترسیم شد که نشان دهنده‌ی میزان تمایز افراد از همدیگر است. نتایج آنالیز PCA بر اساس PC1 و PC2 نشان داد که برای این جمعیت، دو PCA اول ۵/۶۳ درصد واریانس را توجیه می‌کنند و برای توجیه ۹۰ درصد از واریانس تقریبا ۷۴ مؤلفه‌ی اول نیاز است.



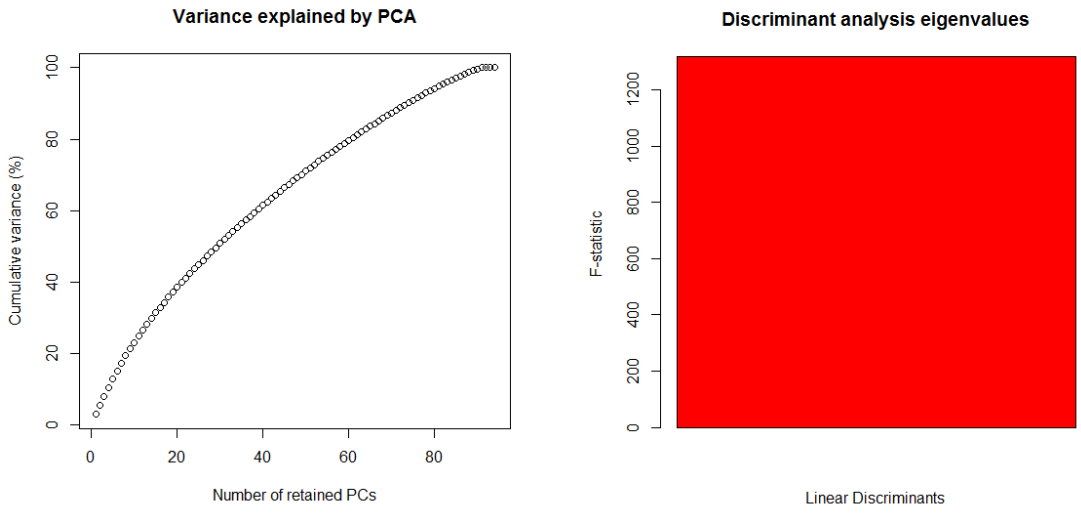
شکل ۱: آنالیز PCA گوسفند زندی

نتایج حاصل از آنالیز تفکیکی مؤلفه‌های اصلی، با دو PC و یک تابع تفکیکی ۵/۶۳۱ درصد واریانس را توجیه کرد. نمودار مربوط به خوشه‌بندی با این روش ارئه شده است (شکل ۲). همانطور که مشاهده می شود در روش PCA نمودار تنها بر اساس دو مؤلفه‌ی اصلی ترسیم می شود و نقاط روی نمودار مربوط به هر دام می باشد و این نمودار حاکی از وجود ناهمگنی واریانس به دلیل تمایز زیرگروه‌ها از یکدیگر است که این تفاوت دو گروه حاصل از آنالیز DAPC نسبت به مؤلفه‌ی اول در شکل ۱ کاملاً واضح است.



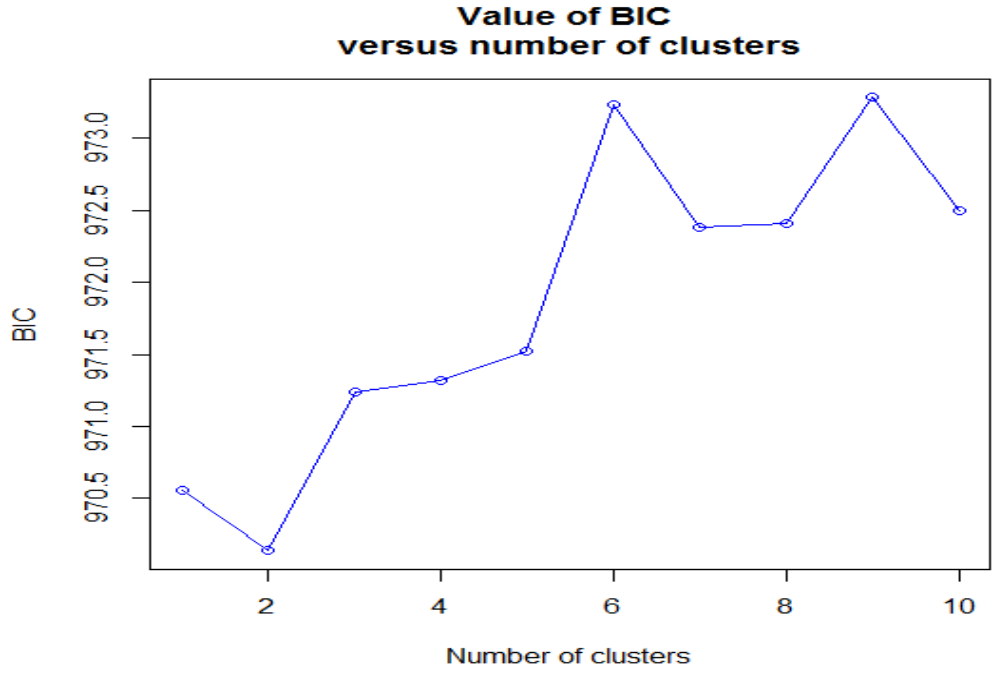
شکل ۲: این شکل یک تابع تفکیکی آنالیز تفکیکی مؤلفه‌های اصلی را نشان می‌دهد. خوشه‌ها و چگالی هر خوشه با رنگ‌های مختلف مشخص شده‌اند.

روش DAPC واریانس بین گروهی را حداکثر و واریانس داخل گروهی را حداقل می‌کند. همچنین با توجه به خوشه‌هایی که تعیین و مرکز هر خوشه را مشخص می‌کند تصویر واضحی از گروه‌های افراد نمایان می‌سازد. پیامهای متقابل آنالیز مبنی بر ناپایداری نتایج برای بدست آوردن تعداد PC (مؤلفه‌ی اصلی)هایی که باید برای آنالیز نگه داشته شوند، نشان داد که ۳۱ مؤلفه‌ی اول دارای پایداری نتایج هستند و تابع تفکیکی حاصل، افراد را در گروه‌های مشخصی گروه‌بندی می‌نماید، این ۳۱ مؤلفه اول با یک تابع تفکیکی، ۵۱/۹۷ درصد واریانس را توجیه کرد. نتایج نشان داد که ۸۹ مؤلفه‌ی اول، حدود ۹۹ درصد واریانس را توجیه می‌کنند (شکل ۳).



شکل ۳: واریانس تجمعی توصیفی توسط مقادیر ویژه‌ی مؤلفه‌های اصلی و ۱ تابع تفکیکی ۳۱ مؤلفه‌ی اول

برای تعیین تعداد بهینه‌ی خوشه بندی کمترین مقدار BIC معیار ارزیابی است که در این مطالعه k=۲ کمترین مقدار را نشان داد (شکل ۴).



شکل ۴: : مقادیر BIC به ازای تعداد خوشه‌ها

طبق شکل ۴، کمترین مقدار BIC، k=۲ است که بعد از آن، نمودار صعودی است و به طور واضح نشان می دهد که این تعداد خوشه بایستی در نظر گرفته شود. با توجه به اینکه صعود از k=۲ اتفاق افتاده است، دو خوشه‌ای بودن این جمعیت را تأیید می‌کند و k=۲ بیانگر وجود دو زیر گروه (دو زیر جمعیت) در این نژاد است و می‌تواند نتیجه‌ی حساسیت بالای روش DAPC باشد که افراد را بدون توجه به نژاد و بر اساس ساختار ژنتیکی جمعیت گروه‌بندی می‌نماید که این امر نیز می‌تواند ناشی از تلاقی‌های انتخابی قوچ و میش در جمعیت مورد مطالعه باشد که بایستی حتماً در آنالیز های GWAS و سایر آنالیز های ژنومی مد نظر قرار گیرد.

نتایج حاصل از این تحقیق در نژاد گوسفند زندی نشان داد که PCA در تفکیک گروه‌ها خیلی حساس نبوده در حالی که روش DAPC اختلاف گروهی را در این نژاد به طور مناسب نشان داد. و این نتایج با نتایج آنالیز شبیه سازی (جُمبارت و همکاران، ۲۰۱۰) و بررسی ساختار جمعیتی (پُمِتّی و همکاران، ۲۰۱۴) مطابقت دارد و آنها نشان دادند که روش DAPC در مشخص کردن زیر بخش‌های جمعیتی و استنباط مقدار k جمعیت بهتر از نرم‌افزار STRUCTURE عمل می‌کند. این روش همچنین سریع‌تر از روش‌های خوشه‌بندی بیزی بوده و برای داده‌های با ابعاد زیاد کارآمد است. این روش مزیتی که نسبت به PCA دارد، در تجسم اختلاف بین گروهی نسبت به PCA است و مشخصه اصلی PCA توانایی آن برای شناسایی ساختارهای ژنتیکی در مجموعه‌داده‌های بزرگ در زمان محاسباتی ناچیز و بدون هیچ فرضی درباره‌ی زمینه‌ی مدل ژنتیکی جمعیت است (جُمبارت و همکاران، ۲۰۱۰). همچنین قابل کر است که نتایج حاصل با مطالعات زیر مطابقت داشت و نشان دادند که بررسی ساختار جمعیتی با روش‌های مبتنی بر مدل و روش اکتشافی DAPC تداوم بالایی در تخمین ساختار جمعیت و استنباط احتمالات عضویت افراد به هر گروه (پُمِتّی و همکاران، ۲۰۱۴) داشت و

همچنین روش PCA به عنوان جایگزینی برای روش‌های خوشه بندی بیزی پیشنهاد شده است (لیو و ژائو، ۲۰۰۶؛ لی و همکاران، ۲۰۰۹). با این حال PCA فاقد برخی از ویژگی‌های ضروری برای بررسی ساختار جمعیت بیولوژیکی از جمله ناتوانی در ارزیابی گروهی و نیازمند تعریف پیش فرض خوشه‌ها برای مطالعه‌ی ساختار جمعیت است. همچنین در تصحیح لایه‌بندی جمعیتی، روش DAPC به علت اینکه واریانس بین گروهی را افزایش و واریانس داخل گروهی را کاهش می‌دهد بهتر از روش PCA است (جُمبارت و همکاران، ۲۰۱۰).علاوه براین DAPC، احتمالات عضویت هر فرد را برای گروه‌های مختلف براساس توابع تفکیکی جایگشتی فراهم می‌آورد. این در حالی است که این ضرایب می‌توانند نزدیکی افراد به کلاسترهای مختلف را تفسیر کنند. احتمالات عضویت افراد، خوشه‌های ژنتیکی صریح و واضحی را فراهم می‌آورد. روش تفکیکی مؤلفه‌های اصلی به عنوان روش با نظارت نیز عمل می‌کند. برای مثال در DAPC ممکن است گروه‌های بیشتر افراد شناخته شده باشند ولی برخی افراد گروه نامشخص یا ناشناخته داشته باشند. در این صورت افرادی که گروه مشخص دارند برای آموزش مدل استفاده می-شوند و داده‌های جدید که گروه‌های نامشخص دارند طبق مدل آموزشی، پیش بینی می‌شوند. نرخ انتساب صحیح یا صحت روش DAPC در انتساب افراد به گروه ها از ۸۰ درصد تا ۹۷ درصد بسته به تعداد تکرار متفاوت بود (جُمبارت و همکاران، ۲۰۱۰). جان کلام این است که در این مطالعه، روش DAPC به علت تعیین تعداد بهینه‌ی k جهت تعیین همگنی و ناهمگنی واریانس داخل نژادی بهتر از روش PCA عمل کرد و تصویر بهتری از ارتباط بین افراد نسبت به PCA ارائه داد. همچنین در انتساب افراد به گروه‌های خودشان صحت بسیار خوبی نشان داد. با توجه به این مزیت‌ها و توانایی انتساب افراد، این روش برای بررسی ساختار جمعیتی و مطالعه همگنی و نا همگنی واریانس نژادی جهت جلوگیری از اریب در مطالعات پویش ژنوم و نشانه های انتخاب و همچنین انتخاب براساس ژنوم روش قابل اطمینانی است.

سپاسگزاری

در اینجا بر خود لازم می‌دانیم که از حمایت‌های مرکز اصلاح نژاد و بهبود تولیدات دامی کشور و جهاد کشاورزی استان تهران و ایستگاه پرورش و اصلاح نژاد گوسفند زندی کمال تشکر و تقدیر را داشته باشیم.

منابع

Barendse, W., Harrison, B. E., Bunch, R. J., Thomas, M.B. & Turner, L.B. (2009). Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC genomics, 10: 178.

Epps, C. W., Castillo, J. A., Schmidt-Küntzel, A., du Preez, P., … & Stuart-Hill, G. (2013). Contrasting historical and recent gene flow among African buffalo herds in the Caprivi Strip of Namibia. Journal of Heredity, ess142.

Grimberg, J., Nawoschik, S., Belluscio, L., McKee, R., … & Turck, A. (1989). A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. Nucleic acids research, 17: 8390-8390.

Jombart, T. & Collins, C. (2015). A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0.

Jombart, T., Devillard, S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics, 11: 94.

Laloë, D., Jombart, T., Dufour, A. B. & Moazami-Goudarzi, K. (2007). Consensus genetic structuring and typological value of markers using multiple co-inertia analysis. Genetics Selection Evolution, 39: 1-23.

Lao, O., Lu T. T., Nothnagel, M., Junge, O., … & Freitag-Wolf, S. (2008). Correlation between genetic and geographic structure in Europe. Current Biology, 18: 1241-1248 .

Laodim, T., Elzo, M. A., Koonawootrittriron, S., & Suwanasopee, T. (2017). Population structure in a Thai multibreed dairy cattle population. *Journal of Animal Science*, *95*, 79.

Lee, C., Abdool, A., Huang, C. H. (2009). PCA-based population structure inference with generic clustering algorithms. BMC bioinformatics, 10: S73.

Liu, N. & Zhao, H. (2006). A non-parametric approach to population structure inference using multilocus genotypes. Human genomics, 2: 1.

Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. Nature genetics, 36: 512-517.

Mohammadabadi, M., Askari, N., Baghizadeh, A. & Esmailizadeh, A. (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.

Olano-Marin, J., Plis, K., Sönnichsen, L., Borowik, T., Niedziałkowska, M., & Jędrzejewska, B. (2014). Weak population structure in European roe deer (Capreolus capreolus) and evidence of introgressive hybridization with Siberian roe deer (C. pygargus) in northeastern Poland. *PloS one*, *9*(10), e109147.

Patterson, N. J., Price, A. L. & Reich, D. (2006). Population structure and eigenanalysis.

Pometti, C. L., Bessega, C. F., Saidman, B. O. & Vilardi, J. C. (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.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., … & Shadick, N. A. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics, 38: 904-909.

Teo, Y. Y., Fry, A.E., Clark, T. G., Tai, E. & Seielstad, M. (2007). On the usage of HWE for identifying genotyping errors. Annals of Human genetics, 71: 701-703.

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

Vergara, M., Basto, M. P., Madeira, M. J., Gómez-Moliner, B. J., Santos-Reis, M., Fernandes, C., & Ruiz-González, A. (2015). Inferring population genetic structure in widely and continuously distributed carnivores: the stone marten (Martes foina) as a case study. *PloS one*, *10*(7), e0134257.

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.