Shahid Bahonar University of Kerman
Agricultural Biotechnology Journal
Iranian Biotechnology Society
Print ISSN: 2228-6705 Online ISSN: 2228-6500

Population structure identification of Turkmen and Darehshori horses using PCA, DAPC, and SPC methods

Ghazaleh Javanmard
MSc Student, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: gh.javanmard@ut.ac.ir

Mohammad Moradi Shahrbabak
Professor, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: moradim@ut.ac.ir

Hossein Moradi Shahrbabak
*Corresponding author. Assistant Professor, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: hmoradis@ut.ac.ir

Javad Rahmaninia
Assistant Professor, Animal Sciences Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran. E-mail address: javad_rahmaninia@yahoo.com

Mahdi Abbasi Firoozjaei
MSc Student, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: mahdiabbasi@alumni.ut.ac.ir

Mohammad Bagher Zandi
Assistant Professor, Department of Animal Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran. E-mail address: mbzandi@znu.ac.ir

Abstract

Objective

Conservation of the genetic diversity of indigenous animals is very important. For the sustainable use of genetic resources, it is necessary to first study the genetic structure of populations. The main goals of this research were to identify the population structure of Turkmen and Darehshori horses using dense SNP markers and to compare the effectiveness of PCA, DAPC, and SPC methods in clustering these populations.

Materials and methods

For this purpose, 67 Turkmen and 39 Darehshori horses were genotyped using Illumina EquineSNP70 BeadChip. After applying quality control steps, five Turkmen horses and one Darehshori horse were removed. Then, the structure of populations was identified by three methods of principal component analysis (PCA), discriminant analysis of principal components

(DAPC), and superparamagnetic clustering (SPC). These methods do not depend on previous assumptions and make it possible to analyze very large genome databases without prior knowledge of individual ancestry. These methods are also very fast and efficient.

Results

This study compared the efficiency of these three clustering methods in identifying population structures. All three methods were successful in separating the two breeds, and Turkmen and Darehshori breeds were grouped into separate genetic groups. The difference is that the DAPC method only separated the two main populations, but the PCA and SPC methods could identify several subpopulations in each breed. The results of this study showed that the SPC method for studying the population structure of indigenous breeds with unknown information can be more useful than other methods. Therefore, using this method, a suitable program can be designed to conserve and use genetic resources.

Conclusions

PCA, DAPC, and SPC methods were able to successfully identify the genetic structure of Turkmen and Darehshori breeds, and in general, it can be said that the information obtained from dense SNP markers can be a powerful tool for identifying the population structure of indigenous breeds.

Keywords: Discriminant analysis of principal components, Principal component analysis, Subpopulation, Superparamagnetic clustering

Paper Type: Research Paper.

Citation: Javanmard G, Moradi Shahrbabak M, Moradi Shahrbabak H, Rahmaninia J, Abbasi Firoozjaei M, Zandi MB (2022) Population structure identification of Turkmen and Darehshori horses using PCA, DAPC, and SPC methods. *Agricultural Biotechnology Journal* 14 (4), 200-220.

Agricultural Biotechnology Journal 14 (4), 200-220. DOI: 10.22103/jab.2022.18795.1372

Received: October 23, 2022. Received in revised form: November 23, 2022.

Accepted: November 24, 2022. Published online: November 30, 2022

Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

مجله بیوتکنولوژی کشاورزی
شاپا چاپی: ۲۲۲۸-۶۷۰۵ شاپا الکترونیکی: ۲۲۲۸-۶۵۰۰

شناسایی ساختار جمعیتی اسب‌های ترکمن و دره‌شوری با استفاده از روش‌های PCA، DAPC و SPC

غزاله جوانمرد

دانشجوی کارشناسی ارشد، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: gh.javanmard@ut.ac.ir

محمد مرادی شهربابک

استاد، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: moradim@ut.ac.ir

حسین مرادی شهربابک

*نویسنده مسئول: استادیار، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: hmoradis@ut.ac.ir

جواد رحمانی‌نیا

استادیار پژوهشی، موسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران. رایانامه: javad_rahmaninia@yahoo.com

مهدی عباسی فیروزجایی

دانشجوی کارشناسی ارشد، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: mahdiabbasi@alumni.ut.ac.ir

محمدباقر زندی

استادیار، دانشگاه زنجان، دانشکده کشاورزی، گروه علوم دامی، زنجان، ایران. رایانامه: mbzandi@znu.ac.ir

تاریخ دریافت: ۱۴۰۰/۱۱/۱۹ تاریخ دریافت فایل اصلاح شده نهایی: ۱۴۰۰/۱۲/۱۶ تاریخ پذیرش: ۱۴۰۰/۱۲/۱۷

چکیده

هدف: حفاظت از تنوع ژنتیکی حیوانات بومی، امر بسیار مهمی است. برای استفاده پایدار از منابع ژنتیکی، باید ابتدا به مطالعه ساختار ژنتیکی جمعیت‌ها پرداخت. اهداف اصلی این پژوهش، شناسایی ساختار جمعیتی اسب‌های نژاد ترکمن و دره‌شوری با استفاده از نشانگرهای متراکم SNP و مقایسه کارایی روش‌های PCA، DAPC و SPC در خوشه بندی این جمعیت ها بود.

مواد و روش‌ها: برای این منظور، ۶۷ راس اسب از نژاد ترکمن و ۳۹ راس اسب از نژاد دره‌شوری با استفاده از تراشه ۷۰k شرکت ایلومینا (Illumina EquineSNP70 BeadChip) تعیین ژنوتیپ شـدند. پس از اعمال مراحل کنترل کیفیت، پنج راس اسب ترکمن و یک راس اسب دره شوری حذف شدند. سپس با استفاده از سه روش تجزیه و تحلیل مولفه‌های اصلی (PCA)، تجزیه و تحلیل تفکیکی مولفه‌های اصلی (DAPC) و خوشه‌بندی فوق پارامغناطیسی (SPC) شناسایی ساختار جمعیت‌ها انجام شد. این روش‌ها به مفروضات پیشین وابسته نیستند و در نتیجه، تجزیه و تحلیل مجموعه داده‌های ژنومی بسیار حجیم را بدون دانش قبلی درباره انساب افراد، امکان‌پذیر می‌کنند. همچنین این روش‌ها بسیار سریع و کارآمد هستند.

نتایج: در این پژوهش، کارایی سه روش خوشه‌بندی یادشده در تشخیص ساختارهای جمعیتی مقایسه شد. هر سه روش در تفکیک دو نژاد موفق بودند و نژادهای ترکمن و دره‌شوری در گروه‌های مجزای ژنتیکی قرار گرفتند؛ با این تفاوت که روش DAPC صرفا دو جمعیت اصـلی را از هم تفکیک کرد اما روش‌های PCA و SPC زیرجمعیت‌های متعددی را نیز در هر نژاد شـناسـایی کردند. نتایج این پژوهش نشان داد روش SPC برای مطالعه ساختار جمعیت نژادهای بومی با اطلاعات ناشناخته، می‌تواند مفیدتر از سایر روش‌های مورد بررسی باشد. بنابراین، با استفاده از این روش می‌توان برنامه مناسبی برای حفظ و استفاده از منابع ژنتیکی طراحی کرد.

نتیجه‌گیری: روش‌های PCA، DAPC و SPC توانستند ساختار ژنتیکی نژادهای ترکمن و دره‌شوری را با موفقیت شناسایی کنند و به‌طور کلی می‌توان گفت اطلاعات به‌دست‌آمده از نشانگرهای متراکم SNP می‌تواند ابزار قدرتمندی برای شناسایی ساختار جمعیتی نژادهای بومی باشد.

کلیدواژه‌ها: تجزیه تفکیکی مولفه‌های اصلی، تجزیه مولفه‌های اصلی، خوشه‌بندی فوق پارامغناطیسی، زیرجمعیت.

نوع مقاله: پژوهشی.

استناد: جوانمرد غزاله، مرادی شهربابک محمد، مرادی شهربابک حسین، رحمانی‌نیا جواد، عباسی فیروزجایی مهدی، زندی محمدباقر (۱۴۰۱) شناسایی ساختار جمعیتی اسب‌های ترکمن و دره‌شوری با استفاده از روش‌های PCA، DAPC و SPC. *مجله بیوتکنولوژی کشاورزی*، ۱۴(۴)، ۲۰۱-۲۲۰.

Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

مقدمه

سرزمین پهناور ایران، خاستگاه نژادهای مختلف اسب است. اسب‌های عرب در مناطق جنوب و جنوب غربی، اسب‌های کرد در مناطق غربی و مرکزی، اسب‌های قره باغ در مناطق شمال غربی، اسب‌های ترکمن در مناطق شمال شرقی و اسب‌های دره‌شوری در مناطق مرکزی و جنوب غربی کشور پرورش می‌یابند (Hedayat-Evrigh et al. 2018). در دو دهه اخیر، واردات اسب های خارجی به کشور افزایش قابل توجهی داشته است. از دلایل این امر می‌توان به تغییر رویه در واردات کشور، افزایش توجه به ورزش سوارکاری و افزایش حضور بخش خصوصی در این عرصه اشاره کرد (Moladoust et al. 2020). واردات مداوم اسب‌های خارجی و آمیزش‌های کنترل نشده بین نژادی، خلوص ژنتیکی نژادهای بومی را کاهش می‌دهد (Khadka 2010; Hassan et al. 2019). برای حفاظت از منابع ژنتیکی، به برنامه‌ی مدیریت ژنتیکی مناسب، نیاز است و لازمه تدوین این برنامه، مطالعه ساختار ژنتیکی و تنوع ژنتیکی جمعیت‌ها است (Jemaa et al. 2015; Laliotis & Avdi 2017). استنباط ساختار جمعیت از داده‌های ژنتیکی، اغلب برای درک فرایندهای تکاملی و جمعیت شناختی، مورد استفاده قرار می‌گیرد و یک جنبه مهم در بسیاری از مطالعات ژنتیکی است. چنین استنباطی عمدتا توسط خوشه‌بندی افراد در گروه‌هایی به نام زیرجمعیت انجام می‌شود. ارزیابی ساختار جمعیت، اجازه استنتاج در مورد الگوهای مهاجرت و پیامدهای ژنتیکی آن‌ها را می‌دهد (Greenbaum et al. 2016). پیشرفت در فناوری‌های تعیین توالی DNA، فرصت استفاده از تعداد زیادی چند شکلی تک نوکلئوتیدی (SNP) را برای بررسی ساختار ژنتیکی و تنوع گونه‌های دام، مانند گاو (Neuditschko 2011; Karimi et al. 2016)، گوسفند (Kijas et al. 2012)، بز (Visser et al. 2016; Colli et al. 2018) و اسب (Petersen et al. 2013; Petersen et al. 2014) فراهم کرده‌است. استفاده از تکنیک‌های مولکولی در سال‌های اخیر جهت مطالعه موجودات بومی و حفاظت شده، کاربرد گسترده‌ای یافته است (Askari et al. 2010; Mohammadabadi 2017). میزان اطلاعات به‌دست‌آمده از این تکنیک های ژنتیکی، یکی از پارامتر های قابل ارزیابی برای مطالعه جمعیت های مختلف و درک تفاوت های ژنتیکی بین جمعیت هاست (Mohammadifar et al. 2014; Mohammadifar and Mohammadabadi 2018). همچنین، مطالعه نژادهای مختلف با استفاده از تکنیک های مولکولی، بسیار مهم و برای طبقه‌بندی آن ها مفید است (Mohammadabadi et al. 2017). حفاظت باید بر اساس دانش عمیقی از منابع ژنتیکی نژادهای خاص باشد، بنابراین، تلاش برای شناسایی و تعیین خصوصیات ژنتیکی نژادهای مختلف بسیار اهمیت دارد (Ghasemi et al. 2010; Mohammadabadi et al. 2021) تنوع ژنتیکی یک عنصر اساسی برای پیشرفت ژنتیکی، حفظ جمعیت‌ها، تکامل و سازگاری با شرایط محیطی متغییر و مختلف است (Askari et al. 2008; Mohammadifar and Mohammadabadi 2011). تجزیه و تحلیل داده‌های حجیم و شناسایی ساختار جمعیت بدون اطلاع از نسب افراد می‌تواند یک چالش بزرگ باشد؛ اما می‌توان این مسئله را با به کارگیری روش‌های خوشه‌بندی بدون نظارت حل کرد (Neuditschko

2011). بدین جهت در این پژوهش، برای شناسایی ساختار ژنتیکی اسب‌های نژاد ترکمن و دره‌شوری، روش‌های تجزیه و تحلیل مولفه‌های اصلی (PCA)، تجزیه و تحلیل تفکیکی مولفه‌های اصلی (DAPC) و خوشه‌بندی فوق پارامغناطیسی (SPC) مورد استفاده قرار گرفتند. PCA یک روش مبتنی بر فاصله برای تجزیه و تحلیل مجموعه داده‌هایی با ابعاد بزرگ است. این روش، مجموعه داده‌ی متشکل از تعداد زیادی متغیر همبسته را به مجموعه کوچکی از متغیرهای غیر همبسته تبدیل می‌کند. اعضای این مجموعه جدید، مؤلفه‌های اصلی (PCs) نامیده می‌شوند (Jolliffe 2003; Reich et al. 2008). هر PC، ترکیبی خطی از متغیرهای اولیه است (Lavine & Mirjankar 2006). برای شناسایی ساختار جمعیت، معمولا PCA به همراه الگوریتم خوشه‌بندی k-means اعمال می شود (Liu & Zhao 2006). DAPC یک روش مبتنی بر فاصله برای شناسایی و توصیف خوشه‌ها در مجموعه داده‌های بزرگ است (Campoy et al. 2016) روش DAPC توسط Jombart et al. (2010) پیشنهاد شد (Jombart et al. 2010). این روش یک رویکرد چند متغیره است که ترکیبی از تجزیه و تحلیل مولفه‌های اصلی (PCA) و تجزیه و تحلیل تفکیکی (DA) برای خلاصه کردن تمایز ژنتیکی بین گروه ها است (García-Girón et al. 2019). SPC یک روش خوشه‌بندی مبتنی بر فاصله از نوع سلسله مراتبی است که از رفتار فیزیکی دانه‌های فرومغناطیس ناهمگن الهام گرفته شده‌است (Blatt et al. 1996). در مطالعات پیشین، ساختار جمعیتی اسب‌های بومی ایران با استفاده از ژنوم میتوکندریایی (Moridi et al. 2012)، نشانگرهای ریز ماهواره (Seyedsharifi et al. 2019a; Seyedsharifi et al. 2019b; Hedayat-Evrigh et al. 2020; Khamisabad et al. 2020; Abdoli et al. 2021) و نشانگرهای SNP (Maghsoodi et al. 2017; Sadeghi et al. 2019) مورد بررسی قرار گرفته‌است. در مطالعه ای، ساختار ژنتیکی پنج نژاد اسب بومی ایران (عرب، ترکمن، دره‌شوری، کرد و کاسپین) بررسی شد و نتایج حاکی از آن بود که نژادهای کاسپین و کرد دارای تشابه ژنتیکی بیشتری بوده و در یک گروه نژادی واقع شدند و نژادهای ترکمن، عرب و دره‌شوری در گروه‌های مجزای ژنتیکی قرار گرفتند (Abdoli et al. 2021). در مطالعه دیگری، تنوع ژنتیکی در اسب‌های عرب و رابطه ژنتیکی این نژاد با نژادهای ترکمن، دره‌شوری، کرد و خزر بررسی شد و نتایج نشان داد اسب‌های عرب را می‌توان به سه خوشه تقسیم کرد؛ همچنین ساختار جمعیتی متمایزی بین نژادهای عرب، ترکمن و خزر مشاهده شد، درحالی‌که بین اسب‌های عرب، کرد و دره شوری همپوشانی وجود داشت (Sadeghi et al. 2019). اهداف این پژوهش، بررسی ساختار ژنتیکی جمعیت‌های اسب ترکمن و دره‌شوری و همچنین مقایسه عملکرد روش‌های PCA، DAPC و SPC در خوشه بندی این جمعیت ها بود.

مواد و روش‌ها

نمونه‌گیری، استخراج DNA و تعیین ژنوتیپ SNPها: در این بخش، نمونه خون از ۶۷ راس اسب نژاد ترکمن (۴۸ راس مادیان و ۱۹ راس نریان) از شهرهای گنبدکاووس، بندر ترکمن، آق‌قلا، گرگان و کلاله و همچنین ۳۹ راس اسب نژاد

دره‌شوری (۲۸ راس مادیان و ۱۱ راس نریان) از شهرهای شیراز، خان‌زنیان، کازرون، کوار، استهبان، نی‌ریز، زروان و مرودشت توسط شرکت دانش‌بنیان ساینا گستر البرز تهیه شد. در این پژوهش سعی شد نمونه‌گیری از حیواناتی به عمل آید که رابطه خویشاوندی با یکدیگر نداشته و از لحاظ مورفولوژیکی و فنوتیپی دارای حداقل خصوصیات نژادی باشند. استخراج DNA از خون به روش کلرفرم در آزمایشگاه بیوتکنولوژی علوم دامی پردیس کشاورزی و منابع طبیعی دانشگاه تهران انجام شد. سپس نمونه‌ها در پلت مخصوص، جایگذاری و برای تعیین ژنوتیپ به دانشگاه کنتاکی آمریکا ارسال شدند. پس از اطمینان از کمیت و کیفیت DNA استخراج شده توسط دستگاه پیکوگرین، نمونه‌ها با استفاده از آرایه‌های شرکت ایلومینا (Illumina EquineSNP70 BeadChip) تعیین ژنوتیپ شدند.

کنترل کیفیت داده‌ها: مراحل کنترل کیفیت داده‌ها توسط نرم‌افزار PLINK(1.07) (Purcell et al. 2007) اعمال شد. در این فرایند، فقط نشانگرهایی با فراوانی آلل حداقل (Minor Allele Frequency) بیش از ۱٪ و نرخ فراخوانی (call rate) بیش از ۹۵٪ حفظ شدند. در معیار تعادل هاردی- واینبرگ نیز جایگاه هایی که P-value آزمون کای اسکور آن‌ها کمتر از 10^{-6} بود حذف شدند.

خوشه‌بندی با روش PCA به همراه الگوریتم k-means: با هدف ارزیابی ساختار جمعیت، یک مسیر دو مرحله‌ای انتخاب گردید. ابتدا از روش PCA برای کاهش ابعاد داده‌ها و سپس برای خوشه‌بندی داده‌های کاهش یافته، از الگوریتم k-means استفاده شد (Ding & He 2004). در مرحله نخست، برای یافتن تعداد PCهای معنادار، از تجزیه و تحلیل موازی Horn استفاده کردیم که توسط بسته نرم‌افزاری paran در محیط R اجرا شد (Dinno & Dinno 2018). سپس روش PCA توسط تابع prcomp روی ماتریس داده‌ها اجرا شد (Holland 2008). در مرحله دوم، از الگوریتم خوشه‌بندی k-means برای تخمین تعداد خوشه‌ها (k) و سپس اختصاص افراد به این خوشه‌ها استفاده شد. با استفاده از روش‌های آرنج (Elbow) و نیمرخ (Silhouette) شناسایی تعداد بهینه خوشه برای الگوریتم k-means انجام شد. هر دو روش توسط تابع fviz_nbclust بسته نرم‌افزاری factoextra در محیط R اجرا شدند (Kassambara & Mundt 2017). روش آرنج، الگوریتم k-means را به ازای محدوده‌ای از مقادیر k (به‌طور مثال با در نظر گرفتن مقدار k در بازه ۱ تا ۱۵) محاسبه می‌کند و سپس برای هر مقدار k، مجموع مربعات داخل خوشه (WSS) را محاسبه می‌کند و منحنی WSS را بر حسب مقادیر مختلف k رسم می‌کند. محل خمیدگی (آرنج) در نمودار، به عنوان شاخص تعداد بهینه خوشه‌ها در نظر گرفته می‌شود (Jeon et al. 2016). روش نیمرخ، الگوریتم k-means را به ازای محدوده‌ای از مقادیر k اجرا می‌کند و ضریب نیمرخ هر یک از مشاهدات را به دست می‌آورد (رابطه ۱) و سپس میانگین آن‌ها را محاسبه می‌کند (رابطه ۲) و مقدار میانگین عرض نیمرخ بر حسب مقادیر مختلف k را در یک نمودار به تصویر می کشد. نقطه ماکزیمم نمودار رسم شده، تعداد بهینه خوشه‌ها را نشان می‌دهد (Kassambara 2017). مقدار ضریب نیمرخ برای هر مشاهده طبق (رابطه ۱) به دست می آید (Reddy 2018):

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$ رابطه ۱)

s(i) ضریب نیمرخ داده i است؛ a(i) میانگین فاصله بین نقطه i و سایر نقاط خوشه‌ای است که i به آن تعلق دارد و b(i) میانگین فاصله از نقطه i تا تمام خوشه‌هایی است که i به آن‌ها تعلق ندارد (Reddy 2018). s(i) در محدوده بین ۱ تا ۱- است (Kaufman & Rousseeuw 2009). میانگین ضرایب نیمرخ تمام نقاط (میانگین عرض نیمرخ) توسط (رابطه ۲) محاسبه می‌شود. هرچه میانگین عرض نیمرخ مجموعه داده بالاتر باشد، خوشه‌بندی بهتر انجام شده است (Reddy 2018).

$$S = \frac{\sum_{i=1}^{N} \frac{b(i) - a(i)}{max\{a(i), b(i)\}}}{N}$$ رابطه ۲)

خوشه‌بندی با روش DAPC: روش DAPC با استفاده از بسته نرم‌افزاری adegenet 2.0.0 در محیط R انجام شد (Jombart et al. 2018). DAPC نیاز دارد گروه ها از پیش تعیین شده باشند، بنابراین از الگوریتم k-means برای تعیین تعداد خوشه‌ها استفاده کردیم. تعیین خوشه‌ها توسط تابع find.clusters انجام شد (Jombart et al. 2018). این تابع، الگوریتم k-means را مرتبا با افزایش مقادیر k اجرا می‌کند و حالت‌های مختلف خوشه‌بندی را با استفاده از معیار اطلاعات بیزی (BIC) مقایسه می‌کند. بهترین تعداد خوشه، اغلب با یک آرنج در منحنی مقادیر BIC به عنوان تابعی از k نمایان می‌شود (Jombart & Collins 2015). سپس برای ارزیابی تعداد بهینه مولفه‌های اصلی حفظ شده، از روش اعتبارسنجی متقابل که توسط تابع xvalDapc اجرا می شود، استفاده کردیم و در نهایت تجزیه و تحلیل DAPC توسط تابع dapc اجرا شد (Jombart et al. 2018).

خوشه‌بندی با روش SPC: الگوریتم خوشه‌بندی SPC در بسته نرم‌افزاری Sorting points into neighborhoods (SPIN) پیاده‌سازی شده است (Tsafrir et al. 2005). ورودی‌های این الگوریتم عبارتند از: ماتریس فواصل (D)، k نزدیک‌ترین همسایگان (k-NN)، یک متغیر اسپین (s) که مقادیر صحیح ۱ تا q را به خود می گیرد: (s = 1,2,….q)، یک ΔT ثابت و حداقل سایز خوشه (Neuditschko et al. 2010). در معیار k-NN، هر فرد به k نزدیک‌ترین هم سایگان خود متصل می شود. بنابراین، عملکرد خوشه‌بندی SPC به شدت به تعداد k-NN وابسته است، به عنوان مثال با کاهش k-NN تعداد خوشه‌ها افزایش می‌یابد (Neuditschko et al. 2010). با توجه به ماتریس D، هر نقطه (داده) با یک متغیر Pott spin یعنی s مرتبط می شود. نتیجه خوشه‌بندی نسبت به انتخاب q حساس نیست و ما با مقدار معمول q = 20 کار کردیم (Neuditschko 2011). پس از مرتبط شدن Pott spinها، خوشه‌بندی در یک محدوده دمایی (ΔT) انجام می شود که تعاملات را به تعداد معینی از k-NN محدود می‌کند (Tsiafouli et al. 2017). برای ارزیابی کیفیت زیرشبکه‌ها و تعیین تعداد بهینه k-NN از پارامتری به نام مدولاریته (Q) استفاده کردیم. محدوده مدولاریته از ۰ تا ۱ متغیر است و هر چه مقدار آن بیشتر

با شد، تقسیم بندی جامعه بهتر بوده ا ست (Newman 2006). بسته نرم‌افزاری SPIN برای ارزیابی عملکرد خو شه‌بندی SPC در یک فضـــای چند بعدی، از یک تابع هزینه اســـتفاده می‌کند. تابع هزینه اعمال شـــده برای SPC، مدل پاتس همیلتونی فرو مغناطیس ناهمگن است که با (رابطه ۳) محاسبه می‌شود (Neuditschko et al. 2010).

$$H[\{s\}] = \sum_{i,j} J_{ij} \left[1 - \left(\delta_{si,sj}\right)\right] \qquad \text{(رابطه ۳)}$$

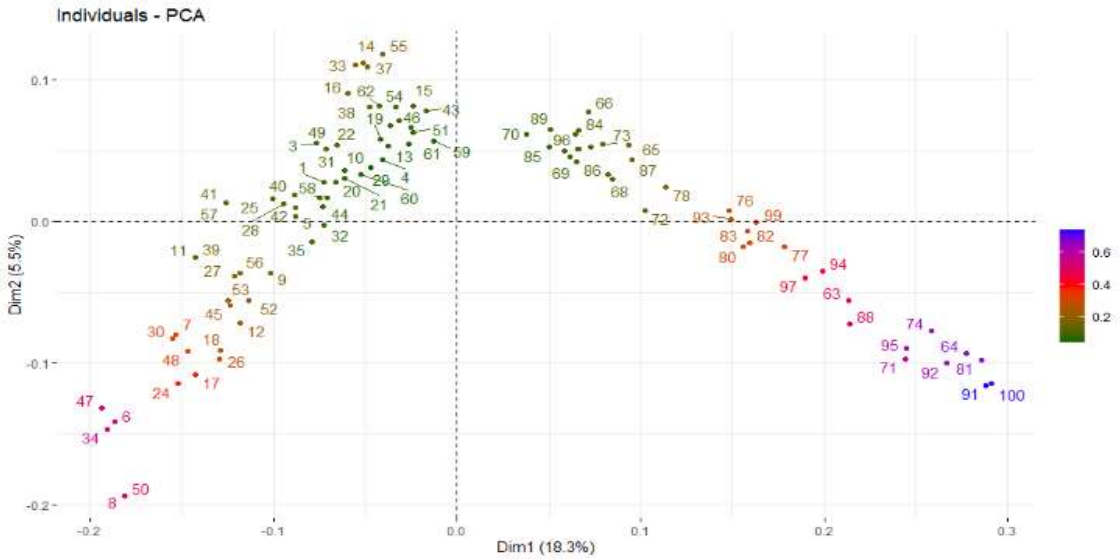
طبقه بندی $\{s\}$ تو سط تابع همبستگی ا سپین–ا سپین $\left(\delta_{si,sj}\right)$ و اثر متقابل نزدیک‌ترین هم‌سایه J_{ij} تعیین می شود، که تابعی کاهشی و مثبت از فوا صل بین نقاط هم‌سایه i و j ا ست. مدل‌های پاتس فرومغناطیسی در دنباله‌ای از دما با یک ΔT ثابت، شبیه سازی می شوند، به‌طوری‌که خو شه‌بندی می‌تواند در هر سطحی از T بیان شود (Neuditschko et al. 2010). این فرایند از $T = 0$ شروع می‌شود، یعنی زمانی که همه نقاط داده، یک خو شه را تشکیل داده‌اند. این خو شه‌بندی اولیه فرضی، پی در پی در یک شیب دمایی پیوسته، انشعاب می‌یابد؛ به عبارت دیگر، الگوریتم SPC شروع به تقسیم دورترین داده‌های ترکیب شده می‌کند و بدین ترتیب یک ساختار سل سله مراتبی از خو شه‌ها ایجاد می شود. بنابراین، محدوده دمایی $\Delta T = T2 - T1$ که یک خو شه از خو شه‌های بالاد ستی جدا می شود به عنوان معیاری برای ثبات و اهمیت خو شه ا ستفاده می شود. هرچه ثبات خو شه بی‌شتر با شد، محدوده ΔT بزرگ‌تر است (Neuditschko 2011).

نتایج و بحث

کنترل کیفیت داده‌ها: در حین مراحل کنترل کیفیت، شش حیوان از بررسی‌ها حذف شدند و ۱۰۰ حیوان باقی ماندند که از این تعداد، ۶۲ راس متعلق به نژاد ترکمن و ۳۸ راس متعلق به نژاد دره‌شـــوری بودند. تعداد ۵۶۹۰۸ جایگاه در این حیوانات حفظ شدند.

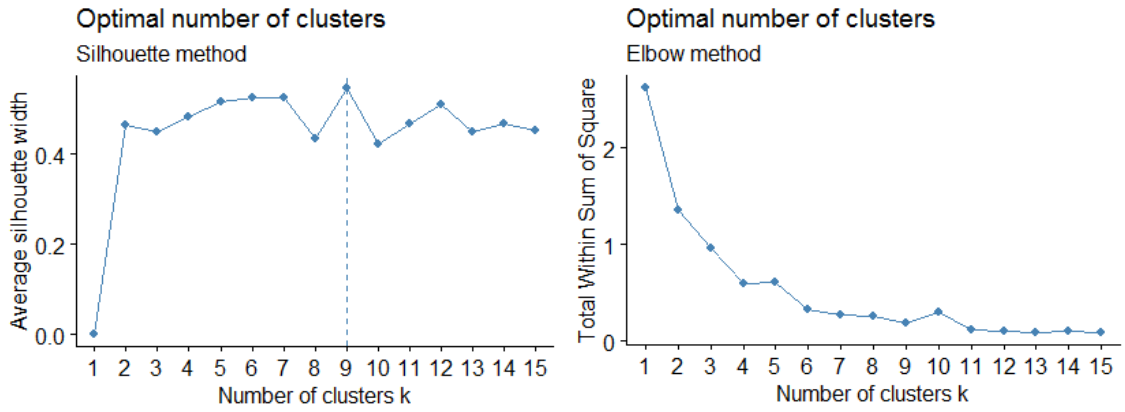
خوشـــه‌بندی با روش PCA به همراه الگوریتم k-means: با به کارگیری تجزیه و تحلیل موازی Horn، تعداد سه مولفه ا صلی حفظ شدند. در این فرایند، واریانس موجود در داده‌ها به مؤلفه‌هایی تجزیه می شود به این صورت که اولین مؤلفه علت بیشترین واریانس موجود در داده‌ها باشد، دومین مؤلفه علت بیشترین واریانس ممکن پس از مؤلفهٔ اول باشد و الی‌آخر.

ســه مولفه اصــلی ابتدایی به ترتیب ۱۸٫۳٪، ۵٫۵٪ و ۵٪ از کل واریانس را توجیه کردند، ســپس ابعاد داده‌ها توســط روش PCA کاهش یافت و پراکنش داده‌ها در یک نمودار دو بعدی به نمایش درآمد، به‌طوری‌که فواصل هندسی بین افراد، منعکس‌کننده فا صله ژنتیکی بین آن‌ها ا ست (شکل ۱). در تعیین تعداد بهینه خو شه‌ها، نتیجه روش آرنج ابهام آمیز بود اما روش نیمرخ تعداد ۹ خوشه را پیشنهاد کرد (شکل ۲).



شکل ۱. نحوه توزیع نمونه‌ها در دو بعد. اعداد ۱ تا ٦۲ متعلق به نمونه‌های نژاد ترکمن و اعداد ٦۳ تا ۱۰۰ متعلق به نمونه‌های نژاد دره‌شوری است

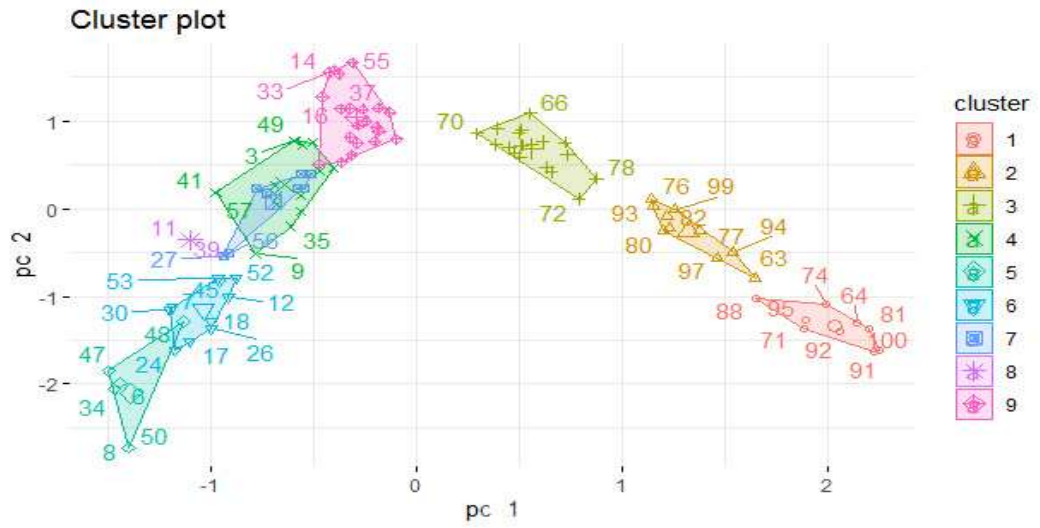
Figure 1. The distribution of the first two dimensions. The numbers 1 to 62 belong to the Turkmen breed and the numbers 63 to 100 belong to the Darehshori breed



شکل ۲. تعیین تعداد بهینه خوشه‌ها با روش آرنج (سمت راست) و نیمرخ (سمت چپ). نتیجه روش آرنج مبهم بود اما روش نیمرخ تعداد ۹ خوشه را پیشنهاد کرد

Figure 2. Determining the optimal number of clusters by Elbow (right) and Silhouette (left) methods. The result of the Elbow method was ambiguous, but the Silhouette method suggested the number of 9 clusters

بنابراین الگوریتم k-means با تعداد ۹ خوشــه اجرا شــد. در نتیجه، نمونه‌های متعلق به نژاد ترکمن (۶۲ راس) در شــش خوشــه و نمونه‌های متعلق به نژاد دره‌شــوری (۳۸ راس) در ســه خوشــه جای گرفتند. بین نمونه‌های دو نژاد هیچگونه اختلاط و همپوشانی وجود نداشت و کاملا از هم تفکیک شدند. (شکل ۳)



شکل ۳. محور x مولفه اصلی اول و محور y مولفه اصلی دوم را نشان می‌دهد. نقاط، نشانگر افراد هستند و خوشه‌ها با رنگ های مختلف نمایش داده شده است. خوشه‌های ۱، ۲ و ۳ نمونه‌های نژاد دره‌شوری هستند و خوشه‌های ٤، ٥، ٦، ۷، ۸ و ۹ نمونه‌های نژاد ترکمن هستند

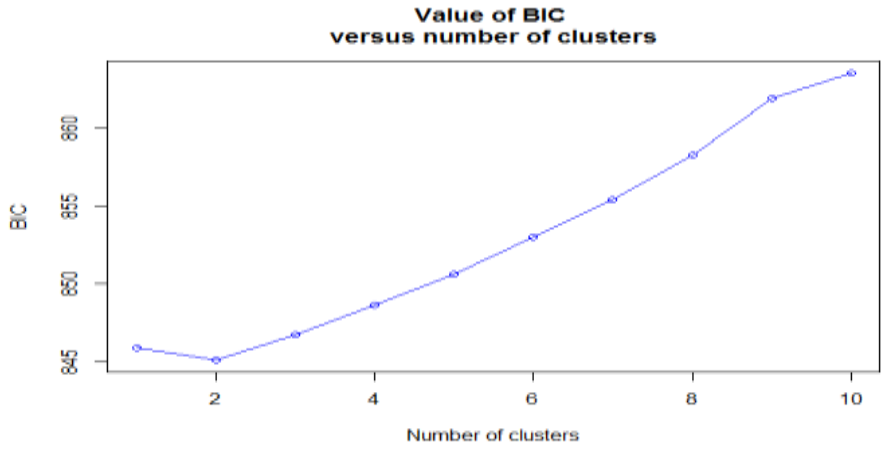
Figure 3. The x-axis represents the first principal component and the y-axis represents the second principal component. The dots represent the samples and the clusters are displayed in different colors. Clusters 1, 2, and 3 are samples of the Darehshori breed, and clusters 4, 5, 6, 7, 8, and 9 are samples of the Turkmen breed

خوشـه‌بندی با روش DAPC: با اجرای الگوریتم k-means، با اسـتفاده از معیار اطلاعات بیزی (BIC) تعداد دو خوشـه به عنوان تعداد بهینه خوشـه‌ها انتخاب شـد (شـکل ۴). برای تعیین تعداد PCهای حفظ شـده برای انجام تجزیه تحلیل تفکیکی، روش اعتبار سنجی متقابل به کار برده شد. در شکل حا صله (شکل ۵) تعداد PC با کمترین خطای جذر میانگین مربعات (RMSE) به عنوان تعداد بهینه PC حفظ شده در نظر گرفته شد (Jombart & Collins 2015) که در این‌جا ۴۰ عدد ا ست. روش DAPC با در نظر گرفتن ۴۰ مؤلفه اصـلی ابتدایی و یک تابع تفکیکی اجرا شـد و نمونه‌ها در دو خوشـه جای گرفتند. نمونه‌های نژاد ترکمن و دره‌شوری کاملا از هم تفکیک شدند و فاصله نسبتا زیادی از هم دارند (شکل ۶).

خوشـه‌بندی با روش SPC: برای تعیین مقدار بهینه k-NN، ابتدا مقداری به k-NN تعلق می‌گیرد که در آن، همه نمونه‌ها در یک خوشـه قرار می‌گیرند. این مقدار برای داده‌های ما، برابر k-NN = 33 بود. سـپس مقدار k-NN به تدریج کاهش یافت و در k-NN = ۲ تعداد ۱۶ خو شه حا صل شد. این فرایند در شکل ۷ به نمایش درآمده ا ست. با توجه به (شکل ۸) حداکثر میزان مدولاریته در k-NN = 2 حاصـل شـد (۰،۸۰۷۶۴۲) و اجرای نهایی الگوریتم SPC با این مقدار k-NN انجام پذیرفت (شکل ۹). در سطح k-NN = 2 تعداد ۱۶ خوشه حاصل شد؛ حیوانات متعلق به نژاد ترکمن از ۱۲ زیرجمعیت و حیوانات متعلق به

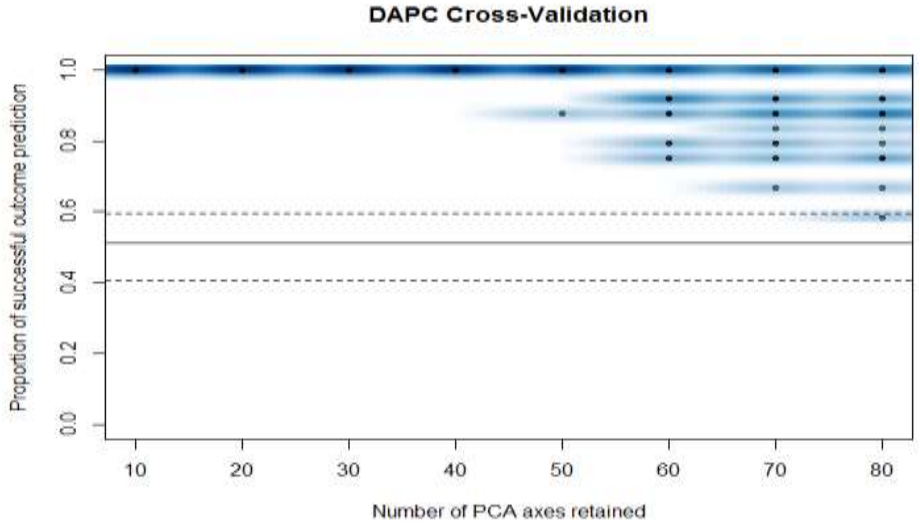
Agricultural Biotechnology Journal; Print ISSN: 2228-6705, Electronic ISSN: 2228-6500

نژاد دره شوری از ۴ زیرجمعیت تشکیل شده اند و اختلاطی بین دو نژاد مشاهده نشد. این ۱۶ خوشه از دو خوشه اصلی انشعاب یافته‌اند. در (شکل۹) تفکیک دو نژاد، به وضوح مشخص است.



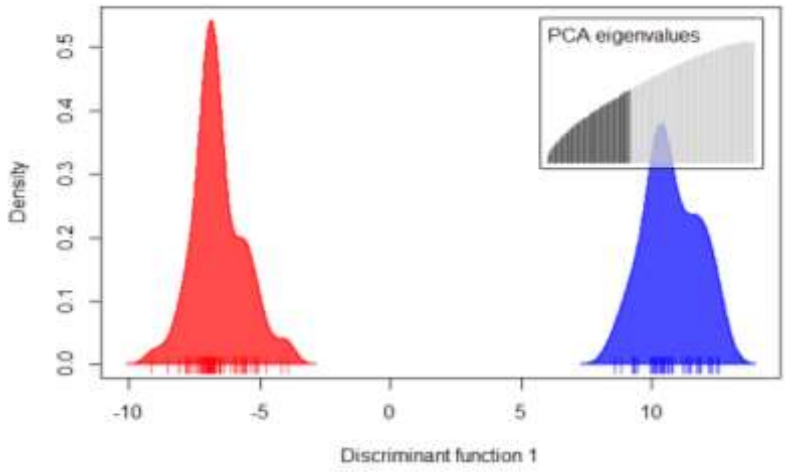
شکل ٤. استنباط تعداد خوشه بهینه با استفاده از معیار اطلاعات بیزی (BIC)

Figure 4. Inferring the number of optimal clusters using the Bayesian information criterion (BIC)



شکل ٥. اعتبارسنجی متقابل DAPC: تعداد PCهای حفظ شده در محور x است و نسبت موفقیت پیش‌بینی نتایج در محور y است. تکرارهای جداگانه به صورت نقاط، ظاهر شده و تراکم آن نقاط در مناطق مختلف شکل با رنگ آبی نشان داده شده است

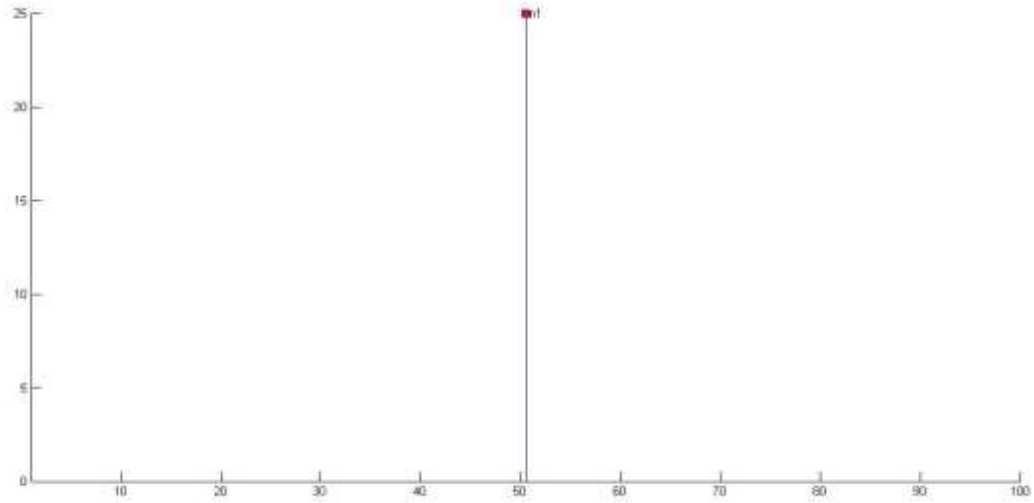
Figure 5. DAPC cross-validation. The number of PCs retained is on the x-axis, and the proportion of successful outcome prediction is on the y-axis. Single repetitions appear as dots, and the density of those dots in different areas is shown in blue



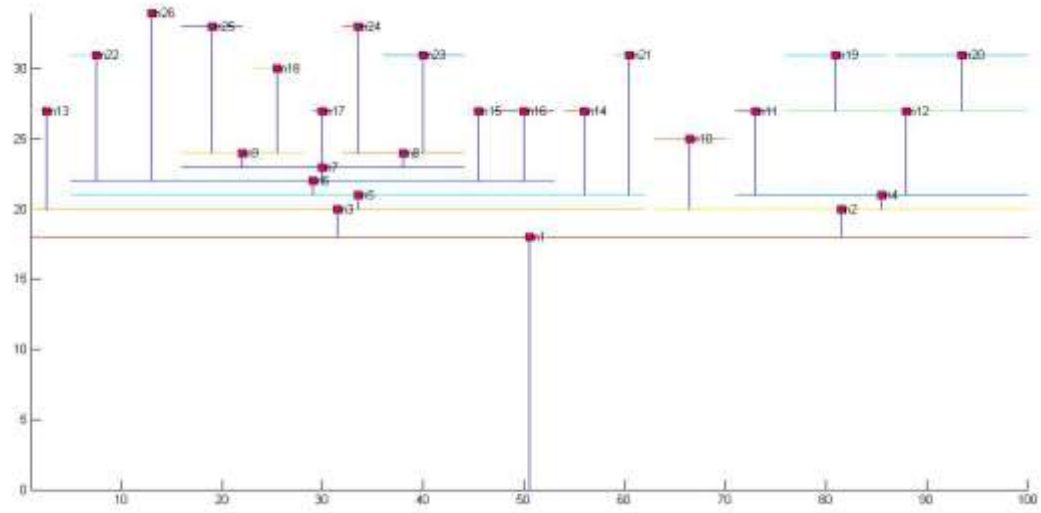
شکل ٦. نتیجه DAPC. نژاد ترکمن با رنگ قرمز و نژاد دره‌شوری با رنگ آبی نمایش داده شده است. هیستوگرام سمت راست بالا، میزان واریانس توجیه شده توسط مولفه‌های اصلی حفظ شده را نشان می‌دهد (PCs = 40)

Figure 6. DAPC result. The Turkmen breed is shown in red and the Darehshori breed is shown in blue. The upper right histogram shows the amount of variance explained by the remaining principal components (PCs = 40)

نتایج این پژوهش نشان داد هر سه روش مورد استفاده (PCA، DAPC و SPC) نمونه های دو نژاد را کاملا از هم تفکیک کردند و هیچگونه همپوشانی بین نمونه‌های دو نژاد وجود نداشت؛ در مطالعه Abdoli et al. (2021) نیز نژادهای ترکمن و دره شوری در گروه‌های مجزای ژنتیکی قرار گرفتند (Abdoli et al. 2021). نتایج این مطالعات نشاندهنده مطابقت خوشه‌بندی ژنتیکی جمعیت‌ها با فاصله جغرافیایی محل پرورش آن‌هاست. ویژگی مثبتی که بین هر سه روش، مشترک بود، این است که همه آن‌ها بسیار سریع و کارآمد هستند. همچنین به فرض‌های پیشین وابسته نیستند و تجزیه و تحلیل مجموعه داده‌های ژنوم بسیار حجیم را بدون دانش قبلی درباره شجره امکان‌پذیر می‌کنند. در این پژوهش، روش DAPC هیچ اطلاعاتی راجع به زیرجمعیت‌ها ارائه نداد و صرفا دو جمعیت اصلی را از هم تفکیک کرد، اما روش‌های PCA و SPC اطلاعات مفیدی را درباره تعداد زیرجمعیت‌ها و نحوه اختصاص افراد به هر زیرجمعیت فراهم کردند. در مطالعه Zargar et al. (2018) روش DAPC در تعیین تعداد بهینه خوشه‌ها بهتر از روش PCA عمل کرد و در انتساب افراد به خوشه خودشان موفق‌تر بود (Zargar et al. 2018). در مطالعه Jemaa et al. (2015) روش DAPC نتایجی مشابه با نتایج PCA نشان داد؛ اما DAPC تفکیک بهتری بین نژادهای با منشا یکسان ارائه داد (Jemaa et al. 2015).



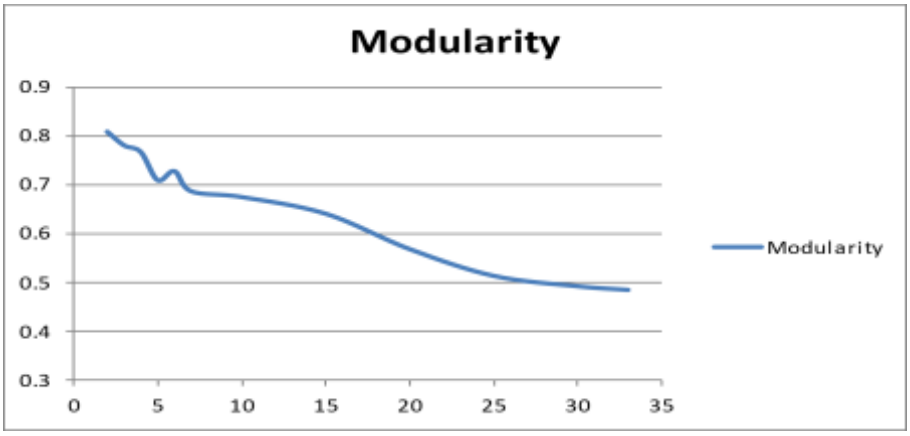
k-NN = 33 k = 1



k-NN = 2 k = 16

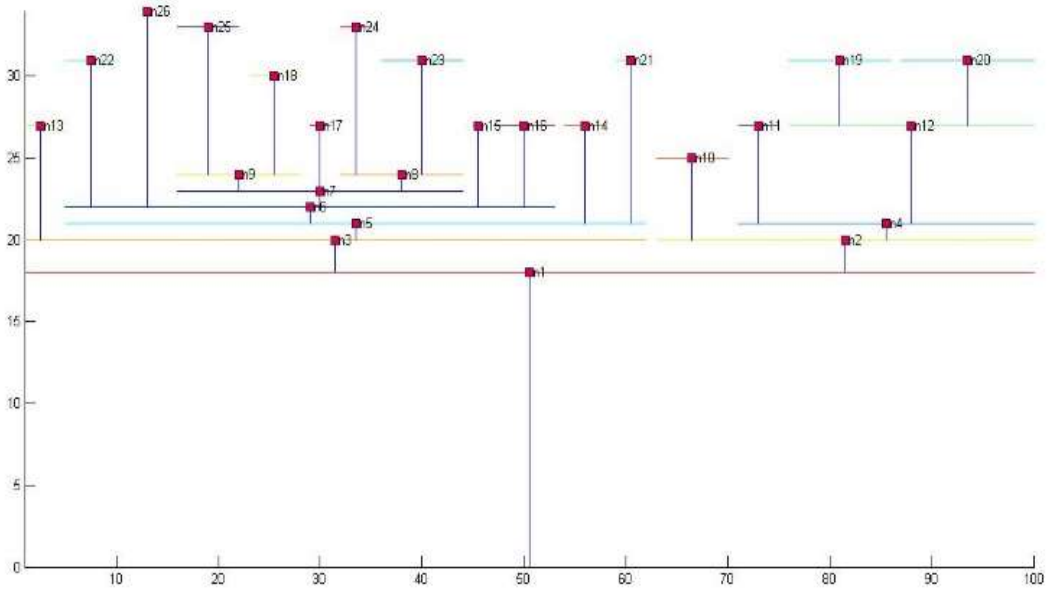
شکل ۷. پاسخ های خوشه‌بندی SPC به ازای تعداد متفاوت k-NN. محور y نشانگر دما و محور x نشانگر افراد است. فاصله افقی خوشه‌ها از یکدیگر نشانگر میزان همجواری آن‌هاست و فاصله عمودی خوشه‌ها نشانگر میزان پایداری خوشه‌هاست

Figure 7. SPC clustering responses for different k-NN values. The y-axis represents the temperature and the x-axis represents the samples. The horizontal distance of the clusters from each other indicates the degree of their neighborhood and the vertical distance of the clusters indicates the degree of stability of the clusters



شکل ۸. برای هر پیمایش SPC میزان مدولاریته را نشان می‌دهد. محور x نشانگر مقدار k-NN و محور y نشانگر میزان مدولاریته است

Figure 8. Indicates the degree of modularity for each SPC run. The x-axis represents the value of k-NN and the y-axis represents the degree of modularity



شکل ۹. دندروگرام خوشه بندی با k-NN بهینه. محور y نشانگر دما و محور x نشانگر افراد است

Figure 9. Clustering dendrogram with optimal k-NN. The y-axis represents the temperature and the x-axis represents the samples

در مطالعه Azizi et al. (2017) روش DAPC احتمال عضـویت افراد جمعیت‌ها را با صـحت ۱۰۰٪ پیش بینی کرد؛ همچنین این روش برای به دسـت آوردن تصـویری روشـن از واریانس بین جمعیت‌ها بهتر از PCA بود (Azizi et al. 2017). نتایج پژوهش حا ضر نیز ن شان می‌دهد با توجه به توانایی روش DAPC در حداکثر کردن واریانس بین خو شه ای و نادیده گرفتن

واریانس درون خوشه ای، این روش در به تصویر کشیدن واریانس بین خوشه ها بهتر از PCA عمل می کند. از نقاط ضعف روش PCA می‌توان به این نکته اشـاره کرد که این روش برای خوشـه‌بندی نمونه‌ها به یک الگوریتم خوشـه‌بندی (مانند الگوریتم K-means) نیاز دارد و این الگوریتم نمی‌تواند تعداد بهینه خو شه‌ها را به طور عینی ارائه دهد (Gao & Starmer 2007). اما روش SPC با این محدودیت مواجه نی ست و تعیین تعداد خو شه‌ها در SPC، "خود سازماندهی شده" ا ست. یکی دیگر از نقاط ضعف روش PCA این اسـت که تجسـم تعداد زیادی PC به طور همزمان امکان‌پذیر نیسـت و تفسـیر روابط به مجموعه‌ای از حالات بصـری دو بعدی محدود می‌شـود (Neuditschko 2011). همچنین الگوریتم K-means در روش PCA ممکن اسـت گرفتار بهینه‌های محلی شود و بسته به نقطه شروع، نتایج متفاوتی تولید کند (Pham et al. 2005) اما SPC دارای یک بهینه سراسری واحد است و بنابراین چندین بار اجرای SPC، نتایج یکسانی ارائه می‌دهد. در مطالعه Rahmaninia et al. (2015) روش SPC در شناسایی ساختار جمعیت بدون پیش‌آگاهی از انساب افراد، موفقیت‌آمیز عمل کرد (Rahmaninia et al. 2015). روش های PCA و DAPC می‌توانند حیوانات را بر ا ساس نژاد تفکیک کنند، اما این روش‌ها نمی‌توانند روابط فیلوژنتیک بین نژادها را نشان دهند. در مقابل، روش SPC، ساختار سلسله مراتبی بین نژادها را نشان می‌دهد (شکل ٩) که کاملا با روابط فیلوژنتیک بین نژادها مطابقت دارد و بدین ترتیب، زیرجمعیت‌های درون نژادها را نشـان می‌دهد و تاریخچه نژادها را منعکس می‌کند (Neuditschko 2011). نتایج ما نشان می‌دهد روش SPC می‌تواند برای مطالعه ساختار جمعیت نژادهای بومی با اطلاعات نا شناخته، بسیار مفید باشد. بنابراین، با کمک این روش می‌توان برنامه مناسبی برای حفظ منابع ژنتیکی و استفاده پایدار از آن‌ها طراحی کرد. سیاست‌های حفاظت از نژادهای بومی تا حد زیادی به دانش ما از روابط ژنتیکی بین نژادها بسـتگی دارد (Marwal et al. 2014). بنابراین، می‌توان از نتایج این پژوهش و پژوهش‌های مشابه، در تدوین برنامه‌های اصلاحی و تولیدمثلی استفاده کرد (Hedayat-Evrigh et al. 2018). این پژوهش‌ها اطلاعات مفیدی در مورد سـطح فعلی تنوع ژنتیکی فراهم می‌کنند. این اطلاعات می‌توانند برای پیش‌بینی چگونگی تأثیر برنامه‌های مدیریتی خاص بر تنوع ژنتیکی گله، مورد اسـتفاده قرار گیرند (Rahimi-Mianji et al. 2015). به طور کلی، برای چشم‌اندازهای طولانی مدت، ا ستفاده از این اطلاعات، راه منا سبی برای ر سیدن به تعادل تنوع ژنتیکی است ((Eusebi et al. 2020).

نتیجه‌گیری: نتایج این پژوهش ن شان داد که روش‌های PCA، DAPC و SPC توان ستند بدون هیچ‌گونه دانش قبلی از انسـاب افراد، سـاختار ژنتیکی نژادهای ترکمن و دره‌شـوری را با موفقیت شـناسـایی کنند. این روش‌ها به دلیل توانایی در کاهش پیچیدگی مجموعه داده‌های حجیم، می‌توانند در تدوین برنامه‌های حفاظت از منابع ژنتیکی نقش قابل توجهی داشـته باشـند. این روش‌ها می‌توانند یک مکمل یا جایگزین برای الگوریتم‌های خوشـه‌بندی وقت‌گیر باشـند، زیرا در عرض چند ثانیه می‌توانند ساختارهای جمعیتی را ت شخیص دهند. همچنین می‌توان گفت اطلاعات به‌د ست‌آمده از ن شانگرهای متراکم SNP می‌تواند برای شناسایی ساختار جمعیتی نژادهای بومی بسیار کاربردی باشد.

سپاسگزاری: از معاونت محترم پژوهشی پردیس کشاورزی و منابع طبیعی دانشگاه تهران و شرکت دانش‌بنیان ساینا گستر البرز بخاطر حمایت و همکاری در اجرای پژوهش حاضر سپاسگزاری می‌شود.

اختصارات: PCA: principal component analysis, PC: principal component, DAPC: discriminant analysis of principal components, DA: discriminant analysis, SPC: superparamagnetic clustering, SNP: single nucleotide polymorphism, WSS: within sum of squares, k-NN: k-nearest neighbor, RMSE: root mean squared error

منابع

رحمانی نیا جواد، میرائی آشتیانی سیدرضا، مرادی شهربابک حسین (۱۳۹۴) بررسی ساختارهای جوامع و خرده جوامع دامی به روش خوشه‌بندی شبکه‌ای بدون نظارت با استفاده از نشانگرهای ژنتیکی متراکم. علوم دامی ایران ۴۶، ۲۸۷-۲۷۷.

زرگر محمدرضا، فیاضی جمال، بیگی نصیری محمدتقی، مرادی شهربابک حسین (۱۳۹۷) بررسی ژنومی ساختار جمعیتی و ارتباط فیلوژنتیکی گاومیش نژاد خوزستانی. پژوهش‌های علوم دامی ۲۸، ۱۹۴-۱۸۱.

سیدشریفی رضا، بادبرین سجاد، خمیس آبادی حسن، هدایت ایوریق نعمت، سیف دواتی جمال (۱۳۹۸) بررسی ساختار ژنتیکی و دقت انتساب افراد به پنج جمعیت اسب با استفاده از نشانگرهای ریزماهواره. پژوهش‌های تولیدات دامی ۱۰، ۱۲۶-۱۲۰.

سیدشریفی رضا، بادبرین سجاد، هدایت ایوریق نعمت، ساورسفلی سیما، سیف دواتی جمال، خمیس آبادی حسن (۱۳۹۸) بررسی ساختار ژنتیکی و روابط فیلوژنی اسب‌های کاسپین، عرب و تالشی. پژوهش‌های علوم دامی ایران ۱۱، ۲۳۲-۲۲۳.

عبدلی محمد، زندی محمدباقر، هرکی نژاد طاهر، خلیلی مسعود (۱۴۰۰) بررسی ساختار ژنتیکی اسب‌های بومی ایران با استفاده از نشانگرهای ریزماهواره. تولیدات دامی ۲۳، ۱۶۳-۱۵۵.

عزیزی زهرا، مرادی شهربابک حسین، مرادی شهربابک محمد (۱۳۹۶) مقایسه روش‌های *PCA* و *DAPC* در تجزیه و تحلیل ساختار جمعیتی گاومیش های ایران با تراشه‌های اسنیپ ۹۰*k*. علوم دامی ایران ۴۸، ۱۶۱-۱۵۳.

عسکری ناهید، باقی زاده امین، محمدآبادی محمدرضا (۱۳۸۹) مطالعه تنوع ژنتیکی در چهار جمعیت بز کرکی راینی با استفاده از نشانگرهای ISSR. ژنتیک نوین ۵، ۵۶-۴۹.

محمدی فر آمنه، فقیه ایمانی سید علی، محمدآبادی محمدرضا، سفلایی محمد (۱۳۹۲) تأثیر ژن TGFβ۳ بر ارزش‌های فنوتیپی و ارثی صفات وزن بدن در مرغ بومی استان فارس. بیوتکنولوژی کشاورزی ۵(۴)، ۱۳۶-۱۲۵.

مریدی میثاق، مسعودی علی‌اکبر، واعظ ترشیزی رسول (۱۳۹۱) مطالعه ساختار ژنتیکی اسب‌های بومی ایران با استفاده از توالی *D-loop* ژنوم میتوکندری. علوم دامی ایران ۴۳، ۱۸۲-۱۷۲.

مقصودی صابرمحمد، مهربانی یگانه حسن، نجاتی جوارمی اردشیر، یوسفی مشعوف نوید (۱۳۹۶) بررسی ساختار جمعیت و شناسایی نواحی تحت انتخاب در ژنگان اسب های کرد و عرب ایرانی. علوم دامی ایران ۴۸، ۴۳۸-۴۲۹.

مولادوست کیومرث، حسینی سیدصفدر، مقدسی رضا (۱۳۹۸) مدیریت تقاضای واردات اسب در ایران. پژوهش در حسابداری و علوم اقتصادی (۳) ۱۷، ۱۳-۲۴.

هدایت ایوریق نعمت، آزادمرد الهام، سیدشریفی رضا، نیک بین سعید، شکوری میرداریوش، خلخالی ایوریق رضا (۱۳۹۸) بررسی تنوع ژنتیکی جمعیت اسب های شمالغرب ایران با استفاده از نشانگرهای ریزماهواره ای. بیوتکنولوژی کشاورزی ۱۱، ۵۰-۳۵.

References

Abdoli M, Zandi MB, Harkinezhad T et al. (2021) Genetic structure survey of Iranian native horse breeds by microsatellite markers. J Anim Pro 23**,** 155-163. (In Persian).

Askari N, Baghizadeh A, Mohammadabadi MR (2010) Study of genetic diversity in four populations of Raeini cashmere goat using ISSR markers. Modern Genet J 5 (2), 49-56 (In Persian).

Askari N, Baghizadeh A, Mohammadabadi MR (2008) Analysis of the genetic structure of Iranian indigenous Raeni cashmere goat populations using microsatellite markers. Biotechnol 2 (3), 1-4.

Azizi Z, Moradi Shahrbabak H, Moradi Shahrbabak M (2017) Comparison of PCA and DAPC methods for analysis of Iranian buffalo population structure using SNPchip90k data. Iran J animal Sci 48**,** 153-161. (In Persian).

Blatt M, Wiseman S, Domany E (1996) Superparamagnetic clustering of data. Physical review letters 76**,** 3251.

Campoy JA, Lerigoleur-Balsemin E, Christmann H et al. (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of Prunus avium L. landraces and bred cultivars. BMC Plant Biol 16**,** 1-15.

Colli L, Milanesi M, Talenti A et al. (2018) Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet Sel 50**,** 1-20.

Ding C, He X (2004) K-means clustering via principal component analysis. In: Proceedings of the twenty-first international conference on Machine learning. pp. 29.

Dinno A, Dinno MA (2018) Package 'paran'. CRAN.

Eusebi PG, Martinez A, Cortes O (2020) Genomic tools for effective conservation of livestock breed diversity. Diversity 12(1), 8.

Gao X, Starmer J (2007) Human population structure detection via multilocus genotype clustering. BMC Genet 8**,** 1-11.

García-Girón J, García P, Fernández-Aláez M et al. (2019) Bridging population genetics and the metacommunity perspective to unravel the biogeographic processes shaping genetic differentiation of Myriophyllum alterniflorum DC. Sci Rep 9**,** 1-10.

Ghasemi M, Baghizadeh A, Abadi MRM (2010) Determination of genetic polymorphism in Kerman Holstein and Jersey cattle population using ISSR markers. Aust J Basic Appl Sci 4 (12), 5758-5760.

Greenbaum G, Templeton AR, Bar-David S (2016) Inference and analysis of population structure using genetic data and network theory. Genetics 202**,** 1299-1312.

Hassan F-u, Khan MS, Saif-ur-Rehman M et al. (2019) Genetic diversity among some horse breeds in Pakistan. Pak J Zool 51**,** 1203-1209.

Hedayat-Evrigh N, Azadmard E, Seyed Sharifi R et al. (2020) Investigation of genetic diversity of Iran northwest horses using microsatellite markers. Agric Biotechnol J 11**,** 35-50. (In Persian).

Holland SM (2008) Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA**,** 1-12.

Jemaa SB, Boussaha M, Mehdi MB et al. (2015) Genome-wide insights into population structure and genetic history of tunisian local cattle using the illumina bovinesnp50 beadchip. BMC Genom 1, 1-12.

Jeon J-Y, Choi J-S, Byun H-G (2016) Implementation of Elbow method to improve the gases classification performance based on the RBFN-NSG algorithm. J Sens Sci Technol 25**,** 431-434.

Jolliffe I (2003) Principal component analysis. Technometrics 45**,** 276.

Jombart T, Collins C (2015) A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11**,** 1-15.

Jombart T, Kamvar ZN, Collins C et al. (2018) Package 'adegenet'. CRAN.

Karimi K, Strucken EM, Moghaddar N et al. (2016) Local and global patterns of admixture and population structure in Iranian native cattle. BMC Genet 17**,** 1-14.

Kassambara A (2017) Practical guide to cluster analysis in R: Unsupervised machine learning. Sthda.

Kassambara A, Mundt F (2017) Package 'factoextra'. Extract and visualize the results of multivariate data analyses 76.

Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.

Khadka R (2010) Global horse population with respect to breeds and risk status. In: Department of Animal Breeding and Genetics. Swedish University of Agricultural Sciences.

Khamisabad H, Badbarin S, Seyedsharifi R (2020) Genetic structure and assignment tests of Kurdish horse based on microsatellite markers. Mod Genet 14**,** 337-344.

Kijas JW, Lenstra JA, Hayes B et al. (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol 10(2)

Laliotis GP, Avdi M (2017) Genetic diversity assessment of an indigenous horse population of Greece. Biotechnol Anim Husb 33**,** 81-90.

Lavine BK, Mirjankar N (2006) Clustering and classification of analytical data. Encyclopedia of Analytical Chemistry: Applications ,Theory and Instrumentation.

Liu N, Zhao H (2006) A non-parametric approach to population structure inference using multilocus genotypes. Hum Genomics 2**,** 1-12.

Maghsoodi SM, Mehrabani Yeganeh H, Nejati Javaremi A et al. (2017) Investigating population structure and identifying signatures of selection in Iranian Kurdish and Arabian horses. Iran J animal Sci 48**,** 429-438. (In Persian).

Marwal A, Sahu AK, Gaur R (2014) Molecular markers: Tool for genetic analysis, in animal biotechnology. Elsevier 289-305.

Mohammadabadi M, Bordbar F, Jensen J et al. (2021) Key genes regulating skeletal muscle development and growth in farm animals. Animals 11 (3), e835.

Mohammadabadi MR (2017) Inter-simple sequence repeat loci associations with predicted breeding values of body weight in Kermani sheep. Genet millenn 14 (4), 4383-4390.

Mohammadabadi MR, Esfandyarpoor E, Mousapour A (2017) Using inter simple sequence repeat multi-loci markers for studying genetic diversity in Kermani sheep. J Res Develop 5 (2), e154.

Mohammadifar A, Mohammadabadi MR (2011) Application of microsatellite markers for a study of Kermani sheep genome. Iran J animal Sci 42 (4), 337-344.

Mohammadifar A, Faghih Imani SA, Mohammadabadi MR, Soflaei M (2014) The effect of TGFb3 gene on phenotypic and breeding values of body weight traits in Fars native fowls. Agric Biotechnol J 5 (4), 125-136.

Mohammadifar A, Mohammadabadi M (2018) Melanocortin-3 receptor (MC3R) gene association with growth and egg production traits in fars indigenous chicken. Malays Appl Biol 47 (3), 85-90.

Moladoust K, Hosseini SS, Moghadasi R (2020) Horse import demand management in Iran. Research in Accounting and Economic Sciences 17(3), 13-24. (In Persian).

Moridi M, Masoudi Aa, Vaez torshizi R (2012) A study of the genetic structure of Iranian native horses using mitochondrial DNA sequence. Iran J animal Sci 43**,** 172-182. (In Persian).

Neuditschko M (2011) A whole-genome population structure analysis within cattle breeds. Department of Veterinary Sciences. Ludwig-Maximilians-University München.

Neuditschko M, Maxa J, Russ I et al. (2010) Spinnet: a new tool to study the population structure with a genome-wide SNP survey. In: Proceedings of the 9 th World Congress on Genetics Applied to Livestock production Leipzig, Germany.

Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci U.S.A. 103**,** 8577-8582.

Petersen JL, Mickelson JR, Cleary KD et al. (2014) The American Quarter Horse: population structure and relationship to the Thoroughbred. J Hered 105**,** 148-162.

Petersen JL, Mickelson JR, Cothran EG et al. (2013) Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS One 8**,** e54997.

Pham DT, Dimov SS, Nguyen CD (2005) Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: J Mech Eng Sci 219**,** 103-119.

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81**,** 559-575.

Rahimi-Mianji G, Nejati-Javaremi A, Farhadi A (2015) Genetic diversity, parentage verification, and genetic bottlenecks evaluation in Iranian Turkmen horse. Russ J Genet 51(9), 916-924.

Rahmaninia J, Miraei-Ashtiani SR, Moradi Shahrbabak H (2015) Unsupervised clustering analysis of population and subpopulation structure using dense SNP markers. Iran J animal Sci 46**,** 277-287. (In Persian).

Reddy CK (2018) Data clustering: Algorithms and applications. Chapman and Hall/CRC.

Reich D, Price AL, Patterson N (2008) Principal component analysis of genetic data. Nat Genet 40**,** 491-492.

Sadeghi R, Moradi Shahrbabak M, Miraei Ashtiani SR et al. (2019) Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. J Hered 110**,** 173-182.

Seyedsharifi R, Badbarin S, Hedayat N et al. (2019a) Investigation of the genetic structure and phylogenic relationships of Caspian, Arabic and Taleshi horses. Iranian Journal of Animal Science Research 11**,** 223-232. (In Persian).

Seyedsharifi R, Badbarin S, Khamisabadi H et al. (2019b) Study of genetic structure and accuracy of assignment of individuals to five horse populations using microsatellite markers. Research on Animal Production 10**,** 120-126. (In Persian).

Tsafrir D, Tsafrir I, Ein-Dor L et al. (2005) Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. Bioinformatics 21**,** 2301-2308.

Tsiafouli MA, Drakou EG, Orgiazzi A et al. (2017) Optimizing the delivery of multiple ecosystem goods and services in agricultural systems. Frontiers Media.

Visser C, Lashmar SF, Van Marle-Köster E et al. (2016) Genetic diversity and population structure in South African, French and Argentinian Angora goats from genome-wide SNP data. PLoS One 11**,** e0154353.

Zargar M, Fayazi J, Beigi M et al. (2018) Genomic study of population structure and phylogenetic relationship of Khuzestani buffaloes. Journal of Animal Science Research 28**,** 181-194. (In Persian).