

## **Comparison of the Different Clustering Methods for Population Structure of Sarabi and Nadjdi Cows by Using Dense Genetic Markers**

**Ahad Khasaliaghtaei**

PhD Student, Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol, Iran, Email: [ahad\\_48@yahoo.com](mailto:ahad_48@yahoo.com)

**Mehdi Vafaye Valleh**

\* Assistant Professor, Department of Animal Science, College of Agriculture, University of Zabol, Zabol, Iran, Tel +98 9358237550, Email: [Mehdi.Valleh@uoz.ac.ir](mailto:Mehdi.Valleh@uoz.ac.ir)

**Gholam Reza Dashab**

Associate Professor, Department of Animal Science, College of Agriculture, University of Zabol, Iran, Email: [dashab@uoz.ac.ir](mailto:dashab@uoz.ac.ir)

**Hossein Moradi Shahrabak**

Associate Professor, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, Email: [hmoradis@ut.ac.ir](mailto:hmoradis@ut.ac.ir)

### **Abstract**

#### **Objective**

So far, various methods have been used to investigate the structure of the population using the markers available in the whole genome (single-nucleotide polymorphism (SNP)), each of which has Weakness and strength. In the present study, an unsupervised network clustering (SPC), a data-mining method, was used to survey the population structure of the Sarabi and Najdi cows.

#### **Materials and methods**

The study population 424 cattle consisted of 213sarabi cattle and 211 Najdi cattle, sequenced with Illumina Bead Chip 40 K v 2 for single nucleotide markers. SORTING POINTS INTO NEIGHBORHOOD (SPIN) was used to analyze population structure. After editing data, 27859 autosomal markers were analyzed.

#### **Results**

Clustering results based on the similarities and differences between nucleotides led to the classification of two base populations and nine clusters.

#### **Conclusions**

Comparison the number of samples and other existing methods for population layering, the use of the SPIN method with high computational efficiency and the needn't for prior assumptions makes it possible to analyze the structure of populations.

**Keywords:** Single-nucleotide polymorphism, Unsupervised network clustering, Data mining, Demographic structure.

**Citation:** Khasaliaghtaei A, Vafaye Valleh M, Dashab GR, Moradi Shahrabak H (2019) Comparison of the Different Clustering Methods for Population Structure of Sarabi and Nadjdi Cows by Using Dense Genetic Markers. *Agricultural Biotechnology Journal* 11 (1), 25-54.

*Agricultural Biotechnology Journal* 11 (1), 25-54.

DOI: 10.22103/jab.2019.13193.1098

Received: January 16, 2019; Accepted: April 28, 2019

© Faculty of Agriculture, Shahid Bahonar University of Kerman-Iranian Biotechnology Society

## مقایسه روش‌های مختلف لایه‌بندی جمعیتی گاوهای سرابی و نجدی با استفاده از

### نشانه‌های ژنتیکی متراکم

احد خصالی اقطاعی

دانشجوی دکترا ژنتیک و اصلاح نژاد دام، دانشگاه زابل، زابل، ایران، ایمیل: [ahad\\_48@yahoo.com](mailto:ahad_48@yahoo.com)

مهدی وفای واله

\* استادیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل، ایران، تلفن 9358237550 +98، ایمیل:

[Mehdi.Valleh@uoz.ac.ir](mailto:Mehdi.Valleh@uoz.ac.ir)

غلامرضا داشاب

دانشیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل، ایران، ایمیل: [dashab@uoz.ac.ir](mailto:dashab@uoz.ac.ir)

حسین مرادی شهریابک

دانشیار، گروه علوم دامی، دانشکده علوم کشاورزی و مهندسی، دانشکده کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران، ایمیل:

[hmoradis@ut.ac.ir](mailto:hmoradis@ut.ac.ir)

تاریخ دریافت: 1397/10/26، تاریخ پذیرش: 1398/02/08

### چکیده

**هدف:** تاکنون از روش‌های مختلفی برای بررسی ساختار جمعیتی با استفاده از نشانه‌های موجود در کل ژنوم (چند شکلی‌های تک‌نوکلئوتیدی (SNP)) استفاده شده است که هر کدام نقاط ضعف و قوتی دارند. در پژوهش حاضر از خوشه‌بندی شبکه‌ای بدون نظارت یا SPC که روشی مبتنی بر داده‌کاوی است، برای بررسی ساختار جمعیت گاوهای سرابی و نجدی استفاده شد.

**مواد و روش‌ها:** جمعیت مورد مطالعه 424 رأس گاو شامل 213 رأس گاو نجدی و 211 رأس گاو سرابی بود که با Chip 40 K Illumina Bead v 2 برای نشانه‌های تک‌نوکلئوتیدی تعیین توالی شدند. برای تجزیه ساختار جمعیت از بسته‌ی نرم‌افزاری SORTING POINTS INTO NEIGHBORHOOD (SPIN) استفاده شد. بعد از ویرایش داده‌ها، 27859 نشانه‌ی اتوزومی تجزیه و تحلیل شدند.

**نتایج:** خوشه‌بندی بر اساس شباهت‌ها و تفاوت‌های نوکلئوتیدها، منجر به طبقه‌بندی دو جمعیت پایه و نه زیر جمعیت شد.

**نتیجه‌گیری:** در مقایسه با سایر روش‌های موجود برای لایه‌بندی جمعیت، در حال حاضر روش SPIN با توجه به عدم نیاز به فرض‌های پیشین، کارایی مناسبی جهت تجزیه و تحلیل ساختار جوامع دارد.

**واژه‌های کلیدی:** چند شکلی‌های تک‌نوکلئوتیدی، خوشه‌بندی شبکه‌ای بدون نظارت، داده‌کاوی، ساختار جمعیتی.

## مقدمه

طبق آخرین آمار رسمی وزارت جهاد کشاورزی، تعداد 18830 واحد صنعتی گاوداری با ظرفیت 2048563 راس گاو شیرده در کشور مشغول فعالیت هستند. طی سالهای 1383 تا 1387 تولید شیر دارای یک روند رو به رشد بوده است (Kharrati Koopaei et al. 2011). با وجود روند افزایشی تولید شیر در کشور اما هنوز سرانه مصرف شیر از حد استاندارد جهانی پایین‌تر است (Kharrati Koopaei et al. 2012a). سرانه مصرف شیر در کشور برای هر نفر برابر با 95 کیلوگرم می‌باشد، در حالی که سرانه مصرف شیر در جهان برابر با 169 کیلوگرم و در اروپا برابر با 350 کیلوگرم در سال است (Kharrati Koopaei, et al. 2011).

با توجه به آمار و اطلاعات موجود می‌توان دریافت که اهداف اصلاح نژادی در ایران بایستی برای افزایش تولید شیر در کشور برنامه‌ریزی شود. لذا، مطالعه و بررسی نژادهای مختلف گاو در ایران اهمیتی دو چندان می‌یابد (Kharrati Koopaei et al. 2012b). از سوی دیگر، با پیشرفت علم ژنتیک، ژن به عامل اصلی در برنامه‌ریزی عملکرد سلول و به دنبال آن کنترل ویژگی‌های موجود زنده شناخته شد. به این ترتیب تمایل برای شناخت هر چه بیشتر ژن‌ها به منظور توجه پدیده‌های زیستی افزایش یافت؛ تا حدی که در چند دهه اخیر، تجهیزات مورد نیاز در تحقیقات مولکولی به طور گسترده‌ای افزایش یافته است و امروزه تحقیقات مولکولی جزو مطالعات رایج آزمایشگاه‌های زیستی است.

اطلاعات به دست آمده از تحلیل داده‌های زیستی به وسیله علم بیوانفورماتیک، در به خط کردن توالی‌ها در بانک‌های اطلاعاتی برای یافتن شباهت‌ها و تفاوت‌های ژنی، پیش‌گویی ساختار و عملکرد محصولات ژن‌ها و یافتن ارتباط فیلوژنتیک میان ژن‌ها و توالی‌های پروتئینی کمک می‌کند (Hadizadeh et al. 2013; Hadizadeh et al. 2014).

رشد روزافزون اطلاعات حاصل از تعیین ژنوتیپ نمونه‌ها به ویژه با استفاده از توالی‌یابی سبب تحول در تجزیه و تحلیل دقیق ساختار جوامع در گونه‌های مختلف شده است. امروزه نشانگرهای ژنتیکی نقش مهمی در مطالعات مرتبط با منشأ، تاریخچه و سیر تکاملی جوامع دامی ایفا می‌کنند (Fernández et al. 2013; Lirón et al. 2002; Troy et al. 2001). پیشرفت‌های کنونی در زمینه‌ی توالی‌یابی ژنوم با تراکم بالا با استفاده از تراشه‌های نشانگری، توسعه‌ی نرم‌افزارهای مربوطه و علم بیوانفورماتیک، استفاده

از چند شکلی‌های تک‌نوکلئوتیدی<sup>1</sup> (SNP) را بسیار مرسوم کرده است؛ هر چند به دلیل دو آلی بودن نشانگرهای SNP، اطلاعات ژنتیکی آشکار شده توسط آن‌ها در مقایسه با نشانگرهای ریزماهورهای کمتر است.

برای رسیدن به دقتی معادل با دقت ریزماهورها برای تعیین هویت افراد و مطالعه‌ی روابط خویشاوندی حداقل تعداد نشانگر SNP دو تا شش برابر مورد نیاز است (Morin et al. 2004)، اما SNPها دارای مزایایی همچون فراوانی بیشتر (Heaton et al. 2005)، پایداری ژنتیکی آن در پستانداران (Markovtsova et al. 2000)، نام‌گذاری آسان و مناسب بودن برای تجزیه و تحلیل‌ها و تفسیر نتایج هستند (Wang et al. 1998).

تنوع ژنتیکی موجود در بین جوامع، حاصل عوامل بسیاری از جمله مهاجرت<sup>2</sup>، جهش<sup>3</sup>، اختلاط جمعیت‌ها<sup>4</sup> و پدیده‌ی تنگنا جمعیتی<sup>5</sup> است (Serre et al. 2008). این تنوع نشانه‌هایی از رانش تصادفی<sup>6</sup> و انتخاب را با خود به همراه دارد. آثار متقابل این عوامل می‌تواند سبب ایجاد جوامع و زیرجمعیت‌های<sup>7</sup> مرتبط با آن شود. برای به دست آوردن دیدگاهی درباره‌ی تاریخچه و سیر تکاملی جوامع، مشخص کردن زیر ساختارهای جمعیتی<sup>8</sup> بسیار با اهمیت است. در مطالعات ارتباط صفات با ژن‌ها (نشانگرها)، برای کاهش ارتباط‌های کاذب زیرساختارهای جمعیتی (Serre et al. 2008) و همچنین فراهم کردن اطلاعاتی ارزشمند در جهت حفظ ذخایر ژنتیکی، لازم است افراد در زیرجمعیت‌های مربوط به خود قرار گیرند (Bowden et al. 2012). امروزه مطالعات متنوعی برای مشخص کردن زیرساختارهای جمعیتی در دام‌ها صورت گرفته است (Decker et al. 2009; Gautier et al. 2010; Kijas et al. 2009). با گسترش تکنولوژی‌های آنالیز گسترده توالی‌های ژنوم، امروزه محققان اطلاعات نشانگری با تراکم بالا را در اختیار دارند و وجود ابزاری برای تجزیه و تحلیل ساختار جامعه با استفاده از این حجم اطلاعات بدون نیاز به پیشینه‌ی اجدادی، به منظور تعیین زیرساختارهای جمعیتی، لازم و ضروری است (Lee et al. 2009; Paschou et al. 2007).

تا به امروز روش‌های آماری بسیاری برای بررسی ارتباط بین جوامع و مشخص کردن موقعیت افراد در جوامعی که از آن نشأت گرفته‌اند، پیشنهاد شده است. دو روش مهم از میان این روش‌ها، روش‌های خوشه‌بندی مبتنی بر فاصله<sup>9</sup> و روش‌های مبتنی بر مدل<sup>10</sup> هستند. در روش اول از فنون کاهش ابعاد از قبیل تجزیه و تحلیل مؤلفه‌های اصلی<sup>11</sup> (PCA) به همراه ابزارهای خوشه‌بندی همچون K-means استفاده می‌شود (Lee et al. 2009; Paschou et al. 2000).

1. Single nucleotide polymorphisem
2. Migration
3. Mutation
4. Mixing of populations
5. Bottle neck population
6. Random drift
7. Sub population
8. Sub-populations structure
9. Clustering base on distanc
10. Clustring base on model
11. Principle component analysis

یکی از روش‌های مبتنی بر فاصله، روش خوشه‌بندی<sup>1</sup> (SPC) است، ماهیت این روش به لحاظ اصولی با داده‌کاوی<sup>2</sup> و کشف دانش<sup>3</sup> مرتبط است. این روش دارای مزایای بسیاری است که از آن جمله می‌توان قوی بودن در برابر اختلالات<sup>4</sup> و منشاء<sup>5</sup>، توانایی در تولید ساختار سلسله مراتبی خود به خودی و از همه مهم‌تر نیاز نداشتن آن به اطلاعات قبلی درباره‌ی توزیع داده‌ها یا ساختار خوشه را نام برد (Radjiman & Sugiarto 2005).

کارایی این روش در آنالیز اطلاعات حاصل از داده‌های مخابره شده‌ی ماهواره‌ای و پروفایل بیان ژن در SPC مخمر، خوشه‌بندی توالی‌های پروتئین به دست آمده از پایگاه‌های داده‌ی SwissProt و SCOP (Tetko et al. 2005)، الگوریتم داده‌های سرطان سینه، سرطان کولون روده‌ی بزرگ و سرطان خون به اثبات رسیده است (Domany 2003; Getz et al. 2003; Getz et al. 2000a; Getz et al. 2000b).

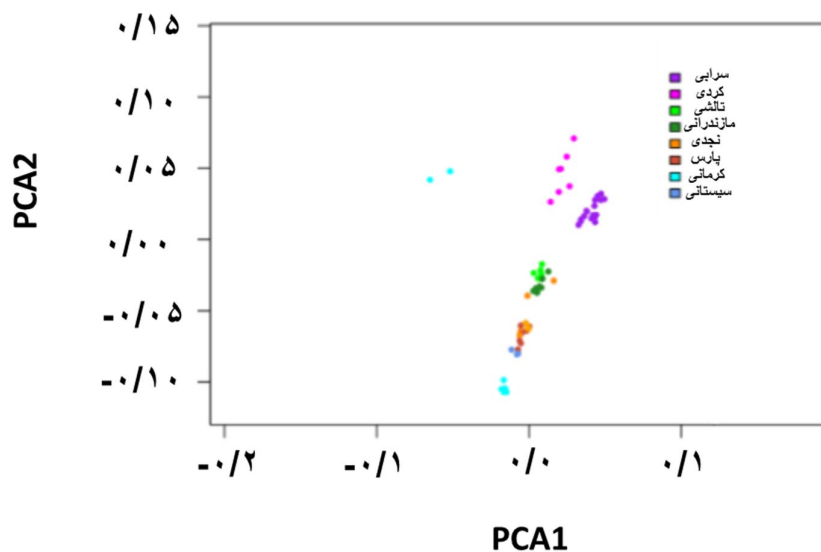
نرم‌افزار SPIN از یک تابع هزینه به نام Hamiltonian inhomogeneous ferromagnetic Potts برای ارزیابی عملکرد خوشه‌بندی توسط SPC در فضای چند بعدی استفاده می‌کند.  $H(S) = -\sum_{ij} J_{ij} \delta_{s_i, s_j}$  در این تابع، طبقه‌بندی (S) با تابعی به نام همبستگی Spin-Spin ( $G_{ij}$ ) و ثابت اتصال  $J_{ij}$  که تابعی مثبت و کاهشی از فواصل است، معین می‌شود. مدل‌های Ferro magnetic Potts در یک توالی از دماها<sup>6</sup> (T) با داشتن تغییرات دمایی ثابت ( $\Delta T$ )، شبیه سازی می‌کنند؛ بنا براین خوشه‌بندی می‌تواند در هر سطحی از T بیان شود. به طور کلی در این تابع برای مقادیر کوچک T، همه‌ی داده‌ها همبستگی ندارند.

با افزایش دما همبستگی Spin-Spin بین نقاط هم‌جوار افزایش یافته و داده‌ها در امتداد دماها ( $T \leq T_{MAX} \leq 0$ ) با استفاده از تقریب میانگین، خوشه‌بندی می‌شوند. در نهایت سطح خوشه‌های ایجاد شده براساس T یا دما، از طریق تعیین مقدار حدی Spin‌های ij با  $G_{ij}(T)$  حداکثری برای هر نقطه از داده، مشخص می‌شود؛ بنابراین محدوده‌ی دمایی  $\Delta T = T_2 - T_1$  که یک خوشه از خوشه‌های بالا دستی (والدین) جدا می‌شود، معیاری برای پایداری و خوشه‌بندی داده‌های متناظر با آن است. هر چه خوشه پایدارتر باشد، محدوده‌ی  $\Delta T$  بزرگ‌تر است. روش کار بر دو مرحله استوار است.

مرحله‌ی نخست، فرآیند مرتب‌سازی است و سپس عمل خوشه‌بندی انجام می‌گیرد. در نخستین مرحله با استفاده از دو الگوریتم<sup>7</sup> STS یا Neighborhood دوباره مرتب‌سازی ماتریس فواصل<sup>8</sup> از قبل انجام می‌گیرد و عناصری که با فاصله‌ی بیشتری از هم قرار دارند، در گوشه‌ی سمت راست بالایی و همچنین سمت چپ پایینی قرار می‌گیرند (Tsafirir et al. 2005).

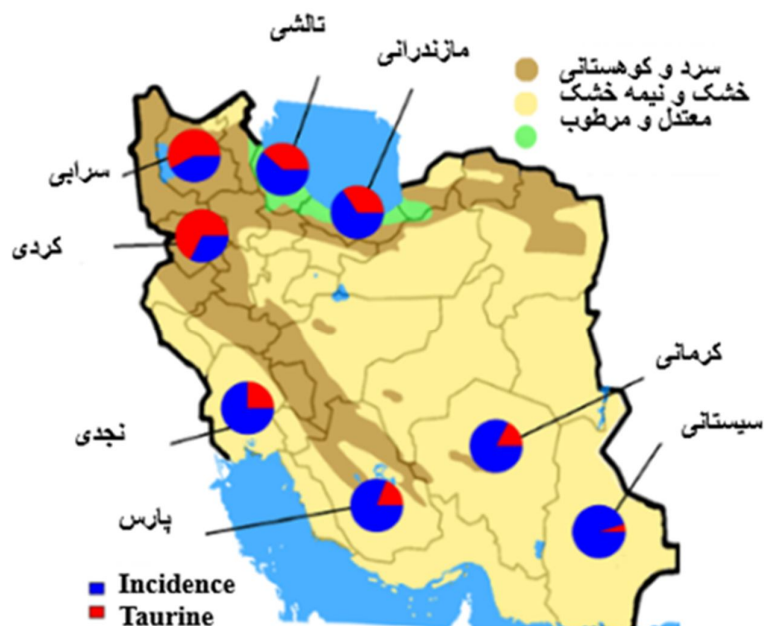
3. Super Paramagnetic
4. Data mining
5. Knowledge discovery
6. Noise
7. Initialization
8. Temperature
1. Side2Side
2. Distance Matrix

در روش دوم، از روش‌های استاندارد آماری برای برآورد فراسنجه‌های جمعیتی استفاده می‌شود از جمله معایب روش‌های مبتنی بر مدل این است که معمولاً برای هر جامعه فرض تعادل هاردی وینبرگ لحاظ می‌شود؛ بنابراین اگر نمونه‌های تهیه شده کوچک باشند، به دلیل برآورد ناصحیح فراوانی آلی، استنتاج‌های مناسبی به دست نخواهد آمد (Gao & Starmer 2007). براساس روش‌های مبتنی بر مدل، نرم‌افزارهای مختلفی توسعه یافته‌اند که می‌توان نرم‌افزارهای STRUCTURE (Pritchard et al. 2000)، ADMIXTURE (Alexander et al. 2009) و FRAPPE (Tang et al. 2005) را نام برد. این دو روش دارای محدودیت در فرض‌های پیشین و عدم توانایی پشتیبانی داده‌های گسترده‌ی ژنومی موجود و همچنین مشکل به تصویر درآوردن و تفسیر نتایج هستند. توده گاوهای بومی ایران بر اساس خوشه‌بندی به 8 گروه (شکل 1) و بر اساس شرایط آب و هوایی به سه دسته اصلی تقسیم می‌شوند: گاوهای مناطق معتدل و مرطوب (مازندرانی و تالشی)، گاوهای منطقه سرد و کوهستانی (سرایی و کردی) و گاوهای منطقه خشک و نیمه خشک (نجدی، پارس، کرمانی و سیستانی) (Karimi et al. 2016) (شکل 2).



شکل 1. خوشه‌بندی 8 توده نژاد گاو ایرانی با استفاده از تجزیه و تحلیل مؤلفه‌های اصلی (PCA)، با استفاده از چند شکلی‌های تکنوکلنوتیدی اتوزومی

Figure 1. Clustering of 8 Iranian Native Cows, using Basic Component Analysis (PCA), using Autosomal SNPs



شکل 2. توزیع جغرافیایی گونه‌های نژاد گاو ایران و نسبت indicine و taurine آنها بر اساس درصد ژنتیک گاوهای نژاد هولشتاین، جرزی و برهمن (K=2)

Figure 2. Distribution of species of Iranian Cows breeds and their ratios of indicine and taurine based on the genetic rate of Holstein, Jerzy and Brahman breeds (K = 2)

گاوهای سرابی<sup>1</sup> و نجدی<sup>2</sup> دو توده بومی دارای اهمیت اقتصادی، اما با اندازه کم جمعیت هستند. گاو سرابی یکی از مهم‌ترین گاوهای دو منظوره بومی در ایران است که سازگاری بالایی با آب و هوای سرد، بیماری‌ها و انگل‌ها دارد. این توده از قره‌باغ وارد ایران شده و در مناطق سراب، میانه و اردبیل پرورش داده می‌شود، تولید شیر این توده به میزان قابل توجهی بالاتر از سایر توده‌های گاو ایرانی است (Khansefid 2010). اکثر گاوهای سرابی، از دید پهلو، دارای بدنی دوزنقه‌ای شکل بوده و به خوبی با شرایط سخت مناطق کوهستانی شمال غربی سازگار شده و اهمیت اقتصادی آنها مربوط به سیستم‌های تولید سنتی پنیر و کره محلی می‌باشد (Tavakkilian 2000).

گاو نجدی از گاوهای بومی کم تولید است که به طور عمده در خوزستان در جنوب غربی ایران پرورش داده می‌شود، مقاومت در برابر بیماری‌ها و نقش آنها در اقتصاد خانواده‌های روستایی از خصوصیات این توده است (Karimi et al. 2009). گاو نجدی

1. Sarabi  
2. Najdi



در طول سالیان متمادی خود را با شرایط آب و هوایی گرم و خشک و منابع غذایی محدود نواحی جنوب غربی ایران که بیشتر از گیاهان مرتعی ناحیه جلگه تشکیل شده است، تطبیق داده و برای تولید گوشت و شیر نگهداری می‌شود (Tavakkilian 2000). از طرف دیگر، اطلاع از ساختار ژنتیکی جمعیت‌ها کمک بزرگی به منظور برنامه‌ریزی برای طرح‌های اصلاح نژادی و از همه مهم‌تر، حفظ ذخایر ژنتیکی می‌باشد (Mohammadabadi et al. 2017). روش‌های مولکولی و استفاده از نشانگرهای مولکولی در این زمینه یکی از بهترین گزینه‌ها به حساب می‌آید، زیرا با توجه به اطلاعات زیادی که به دست می‌دهد می‌تواند نتایجی که از تجزیه و تحلیل رکوردها با روش‌های آماری به دست آمده است را تأیید و تکمیل نموده و حتی ممکن است که آنها را رد کند (Alinaghizadeh et al. 2010).

به علاوه، استفاده از ژنتیک مولکولی فواید زیادی دارد که یکی از این فواید معنی‌دار تعیین ژنوتیپ افراد برای جایگاه خاصی است (Vajed Ebrahimi et al. 2016). همچنین استفاده از نشانگرهای ژنتیکی در انتخاب و اصلاح نژاد حیوانات ممکن است به طور مهبجی پیشرفت ژنتیکی را تسریع کند (Mohammadabadi 2017). علاوه بر این، مطالعه تنوع ژنتیکی نژادهای بومی برای حفاظت از منابع ژنتیکی ذخایر بومی لازم و ضروری است (Mohammadi et al. 2009; Vajed Ebrahimi et al. 2017). حفاظت باید بر اساس دانش عمیقی از منابع ژنتیکی نژادهای خاص باشد، لذا تلاش برای شناسایی و تعیین خصوصیات ژنتیکی نژادهای بومی و محلی بسیار اهمیت دارد (Khodabakhshzadeh et al. 2016; Zamani et al. 2015).

با توجه به نتایج پژوهش‌های مختلف انجام شده بر روی گاوهای هولشتاین ایران (Ghasemi et al. 2010; Mohammad Abadi & Mohammadi, 2010; Kharrati Koopaei et al. 2012; Ebrahimi et al. 2015a and 2016; Barazandeh et al. 2016; Pasandideh et al. 2015b) و آمار و اطلاعات موجود می‌توان دریافت که این دام‌ها باید از جنبه‌های گوناگون مورد مطالعه قرار گیرند. بنابراین، هدف از تحقیق حاضر امکان‌سنجی خوشه‌بندی شبکه‌ای بدون نظارت در تجزیه‌ی ساختار جمعیتی به کمک نشانگرهای متراکم SNP گاوهای سرابی و نجدی به منظور دسترسی به خوشه‌های همگن است.

## مواد و روش‌ها

در این مطالعه از تعداد 424 رأس گاوهای سرابی و نجدی شامل: 213 رأس گاو سرابی (93 رأس مربوط به ایستگاه اصلاح نژاد و 120 رأس مربوط به گله‌های مردمی) و 211 رأس گاو نجدی (121 رأس مربوط به ایستگاه اصلاح نژاد و 90 رأس مربوط به گله‌های مردمی) خون‌گیری به عمل آمد. محل نمونه‌گیری گاوهای سرابی، ایستگاه اصلاح نژاد شهرستان سراب و گله‌های مردمی از شهرستان سراب و محل نمونه‌گیری گاوهای نجدی، ایستگاه اصلاح نژاد شهرستان شوشتر و گله‌های مردمی از شهرستان‌های سوسنگرد، هندیجان و ماهشهر بودند.

نمونه‌های خون جهت استخراج DNA به آزمایشگاه ژنتیک مولکولی گروه علوم دامی دانشگاه علوم کشاورزی و منابع طبیعی کرج انتقال داده شدند. استخراج DNA از خون کامل به روش بهینه یافته استخراج نمکی انجام شد. نمونه‌های DNA

استخراج شده در دمای 20- درجه سانتیگراد تا زمان تعیین ژنوتیپ در آزمایشگاه نگهداری شدند و سپس به آزمایشگاه GeneSeek کشور آمریکا جهت تعیین ژنوتیپ انتقال داده شد. ژنوتیپ با استفاده از تراشه نشانگرهای تک نوکلئوتیدی Illumina Bead Chip 40 K v 2 و با بکارگیری پروتکل استاندارد ارائه شده توسط شرکت ایلومینا انجام گرفت. بعد از انجام مراحل تعیین ژنوتیپ خوانش تراشه‌ها به منظور تعیین نوع SNP های موجود روی تراشه‌های مربوط به هر حیوان با استفاده از اسکنر ارائه شده توسط شرکت ایلومینا<sup>1</sup> انجام شد.

کنترل کیفیت SNP ها یک مرحله مهم در مطالعات ژنومی برای به حداقل رساندن خطای تعیین ژنوتیپ است. برای اطمینان از این که کیفیت داده‌های مورد استفاده در آنالیزهای بعدی بالا باشد با اعمال نرخ فراخوانی<sup>2</sup> کمتر از 10 درصد، حداقل فراوانی آلی<sup>3</sup> کمتر از 95 درصد و تعادل هاردی واینبرگ، داده‌های با کیفیت پایین و با اطلاعات کم از آنالیزهای بعدی حذف شدند. پایش یا ارزیابی چندبعدی<sup>4</sup> یا MDS برای بررسی ساختار جمعیت و ارتباط میان افراد بر اساس ماتریس همبستگی همسانی وجودی در جایگاه<sup>5</sup> IBS بین دو فرد، عمل می‌کند که برای این کار از نرم‌افزار PLINK (Purcell et al. 2007) استفاده شد. سپس MDS در این ماتریس با تابع cmd scale در نرم‌افزار R اجرا شد.

کنترل ژنومیک یا GC، برای برآورد لایه‌بندی جمعیتی با استفاده از آماره لامبدا و نمودار Q-Q plot انجام گرفت. آماره لامبدا که از تقسیم میان مقادیر کای مربع مشاهده شده بر میانه مورد انتظار حاصل می‌شود، فاکتور inflation می‌باشد. اگر مقدار آماره لامبدا کمتر یا مساوی یک باشد، نشان دهنده عدم وجود اثر لایه‌بندی می‌باشد. در این مطالعه روش کنترل ژنومیک و ترسیم-Q plot در نرم‌افزار R و با پکیج SNP assoc اجرا گردید.

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

بررسی آمیختگی بین جمعیت‌ها، آنالیز اختلاط نژادی با استفاده از نرم‌افزار Admixture 1.23 در محیط لینوکس صورت گرفت و میزان اشتراک ژنتیکی این دو توده به صورت گرافیکی بدست آمد (Alexander et al. 2009) از فرمت داده‌های استاندارد و باینری PLINK به عنوان داده‌های ورودی این نرم‌افزار استفاده شد و فایل خروجی آن یک فایل متنی بود و گراف مربوط به این

1. Illumina scan reder
2. Call Rate
3. Minor AllelFerquency
4. Multiple Dimensional Scaling
5. Identical by State

فایل خروجی در محیط R رسم شد. در این نرم‌افزار فاکتور K (فاکتور مربوط به تعداد جمعیت مورد بررسی) در تشخیص تعداد جمعیت‌ها نقش دارد و مبنای تفکیک جمعیت‌ها مقدار عددی این فاکتور می‌باشد و با توجه به خطای اعتبارسنجی متقابل<sup>1</sup> کم‌ترین مقدار برای این منظور انتخاب می‌شود. اندازه‌گیری فاصله‌ی ژنتیکی این فواصل در نرم‌افزار PLINK و توسط بررسی همسانی وجودی در جایگاه (IBS) با کسر از عدد 1 محاسبه شدند.

خوشه‌بندی فوق پارامغناطیسی، خوشه‌بندی شبکه‌ای بدون نظارت در آنالیز ساختار جمعیتی با استفاده از چند شکلی‌های تک نوکلئوتیدی یا خوشه‌بندی (SPC) به منظور استنتاج ساختار جمعیتی گاوهای سرابی و نجدی استفاده شد. برای این منظور از بسته‌ی نرم‌افزاری SPIN<sup>2</sup> (Blatt et al. 1996b; Tsafirir et al. 2005) استفاده شد که در آن از مدل Potts Hamiltonian برای معین کردن ساختارهای جمعیتی در شبکه‌ها استفاده می‌شود (Neuditschko et al. 2012).

انتخاب بهترین K-Nearest-Neighbor (KNN)، برای تعیین تعداد بهینه‌ی KNN فراسنجه‌ای به نام Modularity به عنوان معیاری برای اندازه‌گیری کیفیت شبکه‌ها در فرایند پس از ارزیابی، معرفی شد. برای این منظور از نرم‌افزار MODULAR (Marquitti et al. 2014) استفاده شد. محدوده‌ی Modularity از 0 تا 1 است که 1 نشانگر تقسیمات اجتماعی بهتر است (Holmström et al. 2009).

برای یافتن رابطه‌ی خویشاوندی دقیق‌تر بین افراد، نمودار استخراجی SPC برای پاسخ‌های نهایی خوشه‌بندی به تصویر کشیده شد. برای این منظور، ابتدا با استفاده از ابزار تحلیل‌گر شبکه (Brohée et al. 2008) NeAT<sup>3</sup> داده‌ها به قالب مختصات تبدیل شدند و سپس از نرم‌افزار Cytoscape استفاده شد (Shannon et al. 2003).

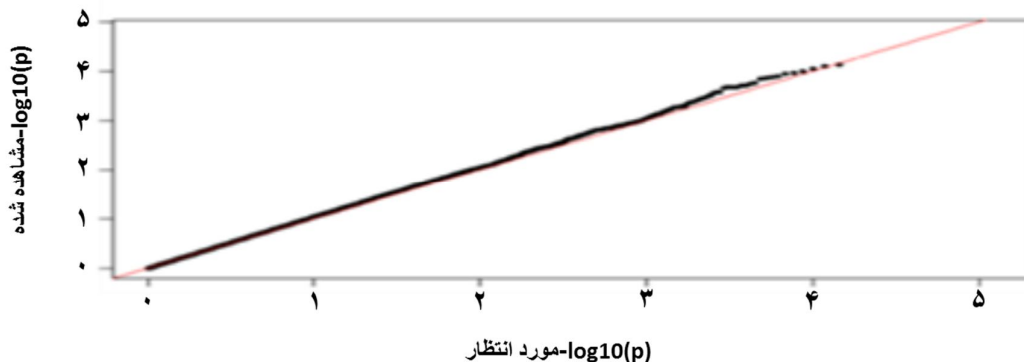
## نتایج و بحث

تعداد 113 رأس گاو شامل سرابی ایستگاهی 15 رأس، سرابی مردمی 18 رأس، نجدی ایستگاهی 43 رأس، نجدی مردمی 37 رأس، و تعداد 415 نشانگر به علت نرخ فراخوانی کمتر از 10 درصد حذف شدند، تعداد 1211 نشانگر به علت حداقل فراوانی آلی کمتر از 95 درصد و تعداد 621 نشانگر به علت تعادل هاردی واینبرگ حذف شدند. در انتها 27859 نشانگر اتوزومی برای تجزیه و تحلیل‌های نهایی در نظر گرفته شدند.

چندین روش برای تخمین کنترل تورم ژنومیک<sup>4</sup> وجود دارد که در این تحقیق روش تخمین‌گر میانه اجرا شد. پارامتر تورم ژنتیکی حدود 1/0343 ( $\lambda_{GC}=1.0343$ ) تخمین زده شده که لایه‌بندی جمعیتی ضعیفی را نشان می‌دهد. به عبارتی انحراف از یک نشان دهنده این موضوع است که گاوهای سرابی و نجدی کاملاً خالص نیستند و اختلاط ژنتیکی بین این حیوانات وجود دارد.

- 
1. Cross validation error
  2. Sorting Points Into Neighborhood
  3. Net analysis tools
  1. Genomic inflation

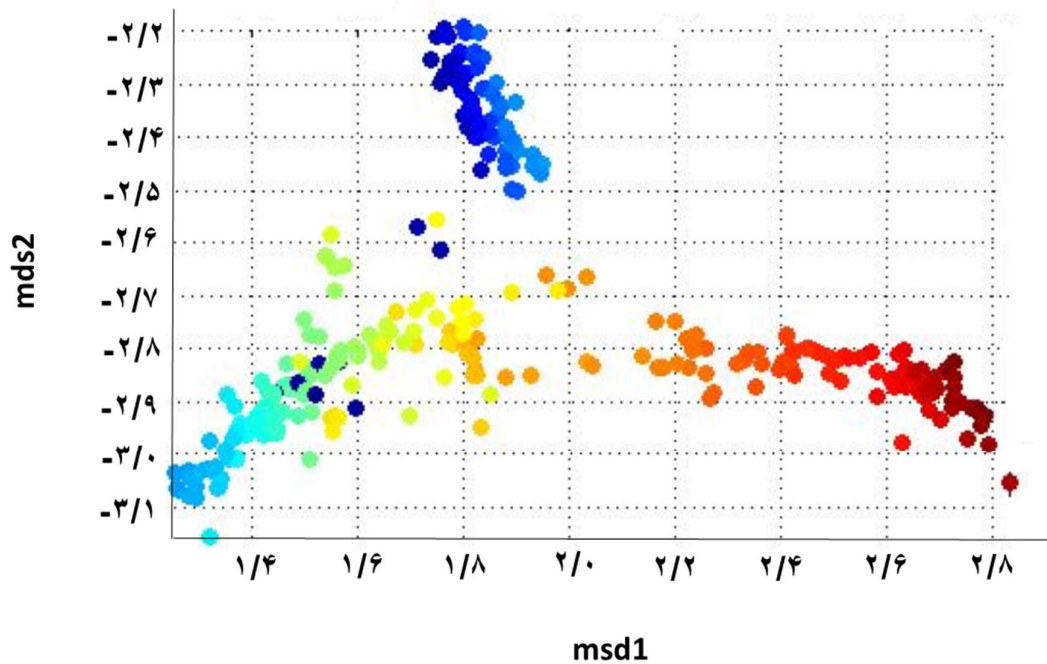
گراف مربوط به Q-Q (شکل 3) ارائه شده است. هرگونه انحراف از خط نشان دهنده وجود اثر لایه‌بندی جمعیتی است.



شکل 3. Q-Q plot برای ارزیابی لایه‌بندی جمعیت‌های مختلف گاوهای سرابی و نجدی

Figure 3. Q-Q plot to evaluate the layering of different populations of Sarabi and Najdi Cows

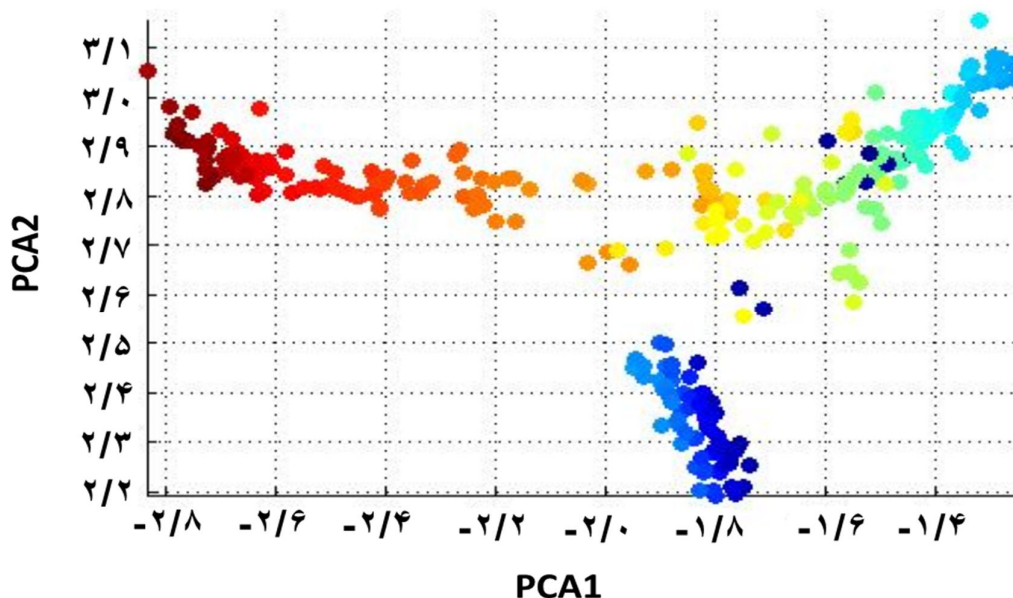
تجزیه ارزیابی چند بعدی که برای مشاهده شباهت یا تفاوت در فضای با ابعاد کم است، توزیع افراد دو جمعیت را نشان می‌دهد (شکل 4)، نتایج حاکی از این است که افرادی که دورترند خلوص بیشتری نسبت به افرادی که در میانه قرار گرفته‌اند، دارند و افرادی که در میانه هستند اختلاط داشته و هیبرید هستند. در این شکل رنگ‌های مختلف خوشه‌های موجود در دو جمعیت را نشان می‌دهد که بیشترین فاصله را گاوهای سرابی و نجدی ایستگاه دارند، ولی با توجه به اختلاطی که در میانه شکل رخ داده است تفکیک کامل گاوهای سرابی و نجدی مربوط به گله‌های مردمی امکان‌پذیر نیست.



شکل 4. لایه‌بندی جمعیتی به روش MDS گاوهای سرابی و نجدی ایران

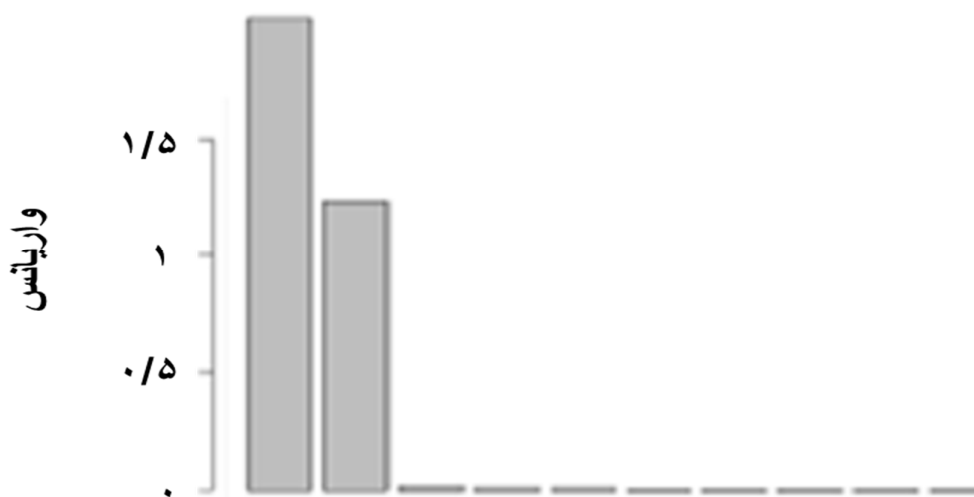
Figure 4. MDS layering in Iranian Sarabi and Najdi Cows

آنالیز به روش PCA، مؤلفه‌های اصلی که ساختار جمعیت را بر اساس همبستگی ژنتیکی میان افراد بیان می‌دارد، شناسایی می‌کند. برای ارزیابی اختلاف ژنتیکی میان جمعیت گاوهای سرابی و نجدی PCA ترسیم شد (شکل 5) که نشان دهنده نزدیکی ژنتیکی افراد در دو جمعیت مزبور است که مثل نتایج MDS در این شکل رنگ‌های مختلف مربوط به خوشه‌های موجود در دو جمعیت می‌باشند، که بیشترین فاصله را گاوهای ایستگاه اصلاح نژاد سرابی و نجدی دارند، ولی نتایج آنالیز PCA بر اساس PCA1 و PCA2 با توجه به اختلاطی که در میانه شکل رخ داده است نشان داد که تا حدودی ارتباط ژنتیکی بین گاوهای گله‌های مردمی وجود دارد، که می‌تواند مربوط به استفاده از گاوهای نر هولشتاین باشد. مقدار واریانس توجیه شده توسط PCA بیانگر تأثیر 58 درصدی مؤلفه‌ی اول (PCA1) و تأثیر 32 درصدی مؤلفه‌ی دوم (PCA2) است. که بالا بودن مقدار واریانس توجیهی نشان دهنده تمایز بیشتر جمعیت‌ها است (شکل 6).



شکل 5. لایه‌بندی جمعیتی با روش PCA برای گاوهای سرابی و نجدی ایران

Figure 5. Demographic layering with PCA method for Iranian Sarabi and Najdi Cows

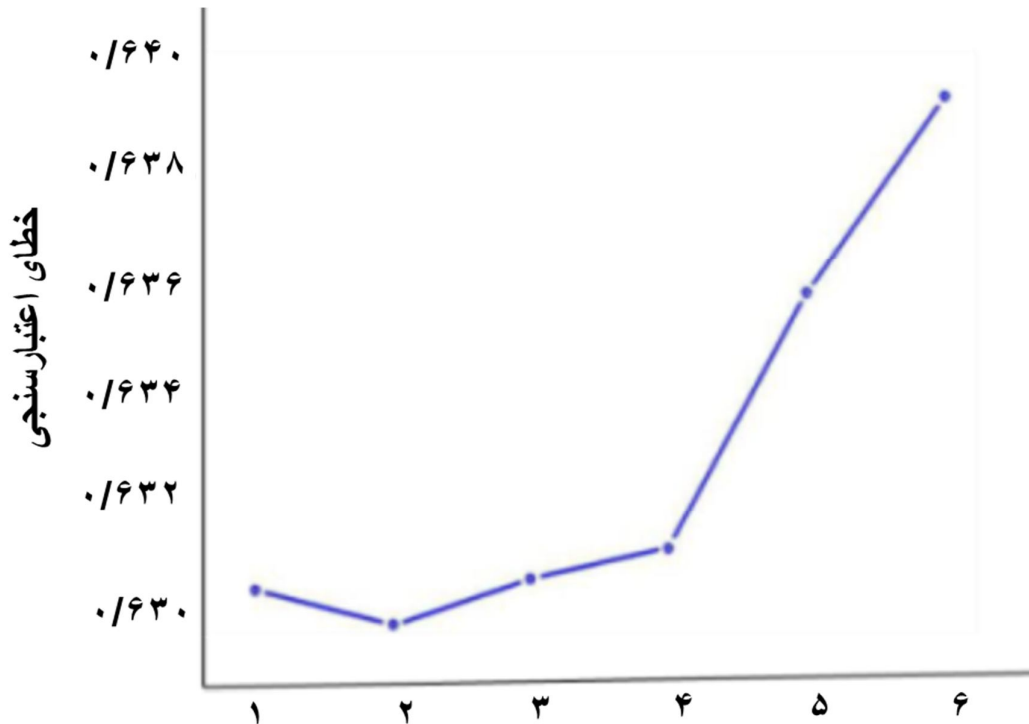


شکل 6. نمودار مقدار واریانس توجیه شده نمایانگر تأثیر 58 درصدی مؤلفه‌ی اول (PCA1) و تأثیر 32

(PCA2) درصدی مؤلفه‌ی دوم در گاوهای سرابی و نجدی ایران است

Figure 6. The magnitude of the variance value justifies the effect of 58% of the first component (PCA1) and the effect of 32 (PCA2) percentage of the second component in Iranian Sarabi and Najdi Cows

آمیختگی بین جمعیت‌ها یا Admixture، فاکتور  $k$  که بیانگر تعداد جمعیت‌ها می‌باشد در دامنه بین 1 تا 6 در نظر گرفته شد و  $k=2$  با 10 بار اعتبار سنجی متقابل کمترین خطا را نشان داد (شکل 7). طبق شکل شماره 8 که درصد اختلاط در سمت چپ قابل مشاهده است افرادی که تا بالای نمودار با یک رنگ مشخص ادامه دارند خالص هستند و افرادی که رنگ‌های مختلف را با درصدهای متفاوت دارند مخلوط بوده و هیبرید می‌باشند. نتایج حاکی از این است که اختلاط ژنتیکی بین دو جمعیت وجود دارد.



شکل 7. اعتبار سنجی متقابل برای داده‌های مربوط به گاوهای سرابی و نجدی ایران برای تعیین  $K$  مربوط به تعداد جمعیت

Figure 7. Cross validation for data related to Iranian Sarabi and Najdi Cows for determining  $K$  population



شکل 8. آنالیز Admixture دو توده نژاد گاو سرابی و نجدی ایران، تعداد جمعیت 2 می باشد

**Figure 8. Admixture analysis of two populations of Sarabi and Najdi Native Iranian Cows, the Number populations are 2**

تعدادی از پاسخ‌های خوشه‌بندی فوق پارامغناطیسی برای تعداد متفاوت KNN؛ خوشه‌ها با اعداد متفاوت از یکدیگر مشخص شده‌اند که با توجه به دمای مشخص شده (محور yها)، خوشه‌ها ابتدا به تقسیمات بزرگ‌تر و به مرور به تقسیمات خردتر تفکیک می‌شوند. هر عدد بر روی تقسیمات (nها) نشان‌دهنده‌ی گروه است. اعداد کوچک‌تر برای n، مربوط به گروه‌های بزرگ و اعداد بزرگ‌تر مربوط به دسته‌بندی‌های خردتر است. با کاهش KNN، تعداد خوشه‌ها (K) افزایش یافت (شکل 9).





مقدار بهینه KNN به وسیلهی ترسیم مقادیر Modularity برای محدودهی KNN مورد بررسی یعنی  $KNN=2$  تا  $KNN=70$  بدست آمد. مشاهده می‌شود که در  $KNN=2$ ، یک حداکثر با مقدار  $Q = 0/986$  به دست آمد (شکل 10).

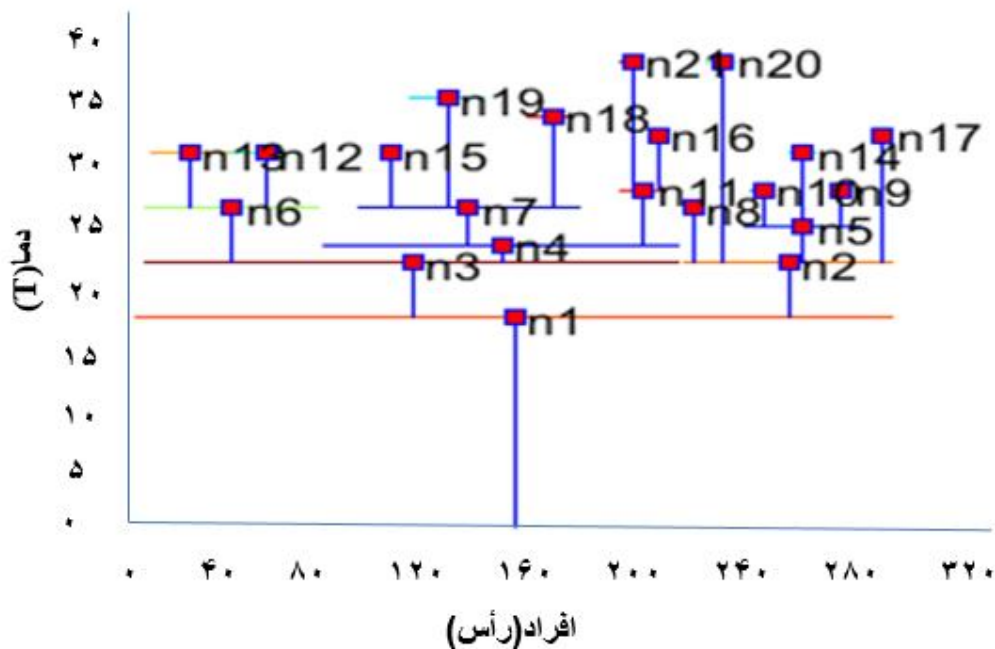


شکل 10. تعیین مقدار بهینه KNN با استفاده از مقادیر مختلف Modularity برای گاوهای سرابی و نجدی ایران

Figure 10. Determining the optimal value of KNN using different Modularity values for Iranian Sarabi and Najdi Cows

پس از انتخاب بهینه KNN نمودار تأیید شدهی خوشه‌بندی به وسیلهی SPC برای جمعیت گاوهای سرابی و نجدی مشخص شد (شکل 11). این نمودار نمایانگر سطح بهینهی خوشه‌بندی با  $KNN=2$  است. حیوانات در 9 خوشه گروه‌بندی شدند که مربوط به 2 جامعه‌ی پایه و اصلی هستند که خود به زیرجمعیت‌هایی تقسیم شده‌اند. فاصله‌ی افقی نشانه‌ی سطح هم‌جواری بین خوشه‌هاست و دما یا همان محور عمودی نشان دهنده‌ی پایداری هر خوشه است.

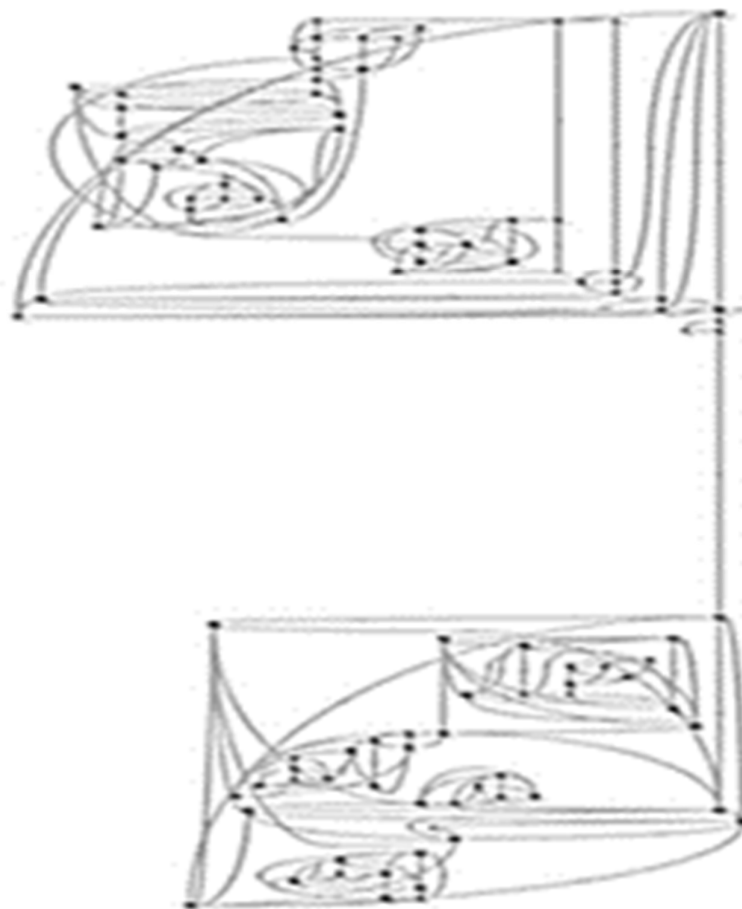
ساختار جمعیتی ارائه شده در شکل 12، جواب‌های خوشه‌بندی شکل 11 را منعکس می‌کند. در این شکل، گره‌ها نمایانگر افراد و خطوط اتصالی بین دو گره نشانگر رابطه‌ی بین حیوانات مرتبط است. هر حیوان با یک گره مشخص شده است و حیوانات قرار گرفته در هر جامعه به خوبی از هم مجزا شده‌اند و خرده جوامع هم در درون جامعه‌ها به صورت توده‌ی متراکم معین شده‌اند. در این شکل به وضوح جدایی دو جامعه‌ی اصلی گاوهای سرابی و نجدی مشاهده می‌شود. این تصویر، بیانگر تفسیر راحت این رویکرد در مقایسه با استفاده از سایر روش‌ها مانند مقیاس‌گذاری چند بعدی (شکل 13) یا کاربرد رویکرد مبتنی بر مدل با استفاده از نرم‌افزار STRUCTURE (شکل 14) است که به منظور مقایسه روی همین مجموعه داده اعمال شدند.



شکل 11. نمودار تأیید شده‌ی خوشه‌بندی به وسیله‌ی SPC برای جمعیت گاوهای سرابی و نجدی ایران

Figure 11. Confirmation chart for clustering by SPC for Iranian Sarabi and Najdi Cows

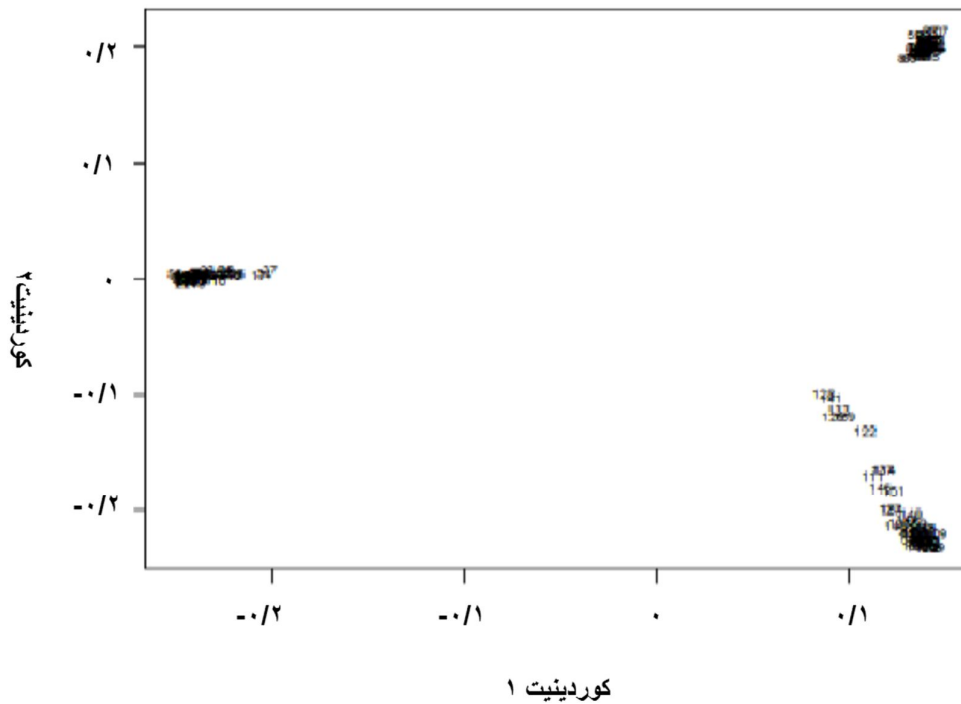
تفسیر نتایج در روش‌های سنتی با توجه به فضای موجود، تصویری کامل و دقیق نیست و با توجه به داده‌های واقعی این تحقیق از جوامع دامی، به سبب اختلاط وسیع جمعیتی و عدم وضوح جمعیت‌ها و زیرجمعیت‌ها، تفسیر نمودارهای حاصل از بررسی PCA بسیار پیچیده می‌شود (شکل 5). محدودیت دیگر PCA این است که به تصویر کشیدن تعداد زیادی مؤلفه به طور همزمان برای نمایان کردن ساختار جامعه ممکن نیست (Gao & Starmer 2007). از طرفی برخی محققان معتقدند روش PCA به سبب اینکه در آن از برخی اطلاعات ژنتیکی چشم‌پوشی می‌شود (همانند فراوانی آللی)، برای کشف ساختار جمعیتی مناسب نیست (Pritchard et al. 2000).



شکل 12. شبکه‌ی ارتباطی بین جوامع و خرده جوامع گاوهای سرابی و نجدی ایران از طریق روش SPC

**Figure 12. Communication network between populations and Sub populations of Iranian Sarabi and Najdi Cows by the SPC method**

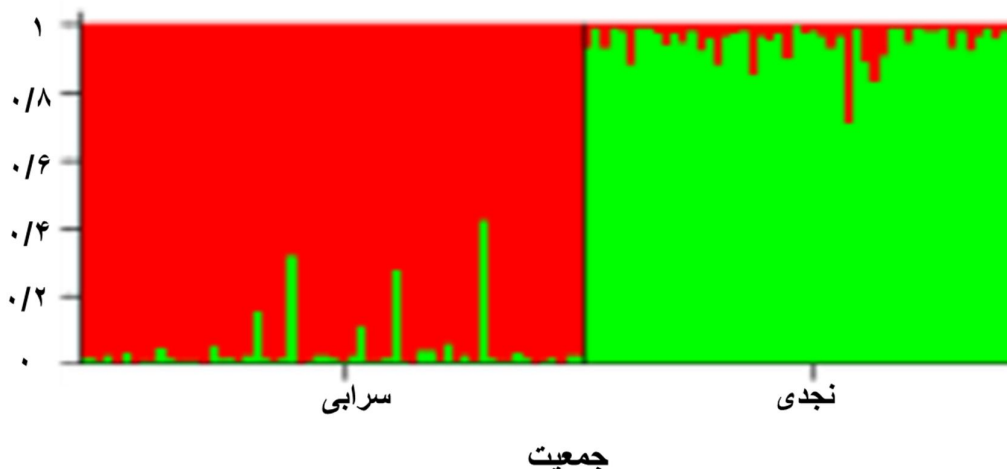
نتایج رویکرد مبتنی بر مدل هم نشان داد که کشف خرده ساختارهای جمعیتی در این روش با محدودیت مواجه است و فقط ساختارهای عمده‌ی جمعیتی را نمایان می‌کند. ساختار جمعیت گاوهای سرابی و نجدی توسط نرم‌افزار ADMIXTURE بررسی شد که نتیجه در شکل 14 نمایش داده شده است. در آنالیز ساختار جمعیت نژادهای گاو میش با روش خوشه‌بندی بیزی و PCA با استفاده از 935 چند شکلی تک‌نکلئوتیدی از Bovine SNP 50 K Bead Chip، نژادها از نظر ساختاری به هم نزدیک بودند و این در حالی بود که این نژادها پیش‌زمینه تفاوت ژنتیکی معنی‌داری هم نداشتند (Wu et al., 2013). در مطالعه‌ای روی جمعیت‌های گوسفند دنیا، از 20 مؤلفه اول در مجموع فقط 16 درصد واریانس را توجیه کردند که بزرگ‌ترین مقدار مربوط به مؤلفه اول به مقدار 1/98 درصد بود (Kijas et al. 2012).



شکل 13. اجرای مقیاس گذاری چند بعدی برای تعیین طبقه بندی جمعیت گاوهای سرابی و نجدی ایران

**Figure 13. Multi-dimensional Scaling to determine the classification of the population of Iranian Sarabi and Najdi Cows**

در مطالعه‌ای روی هفت نژاد بز ایرانی با 14 نشانگر ریزماهواره مؤلفه اول 25/39 درصد واریانس و مؤلفه دوم 19/70 درصد واریانس را توجیه نمود، آماره  $F_{ST}$  نیز محاسبه گردید و آنالیز STRUCTURE که جزء روش‌های مبتنی بر مدل است، نژادهای بز ایرانی را به سه خوشه تقسیم کرد و اختلاطی بین خوشه‌های مرکزی و شمالی مشهود بود در حالی که خوشه غربی استخر ژنی<sup>1</sup> کاملاً مجزایی تشکیل داد (Vahidi et al. 2014). در بررسی ساختار ژنتیکی گاوهای بومی ایران بر اساس نشانگرهای چند شکل تکنوکلوتیدی، روش مبتنی بر مدل به کار گرفته شد و نتایج تعداد سه خوشه را توجیه کرد (Karimi et al. 2015).



شکل 14. نمودار مربوط به خوشه‌بندی جمعیت گاوهای سرابی و نجدی توسط نرم‌افزار STRUCTURE، هر فرد با یک خط عمودی که به K قطعه‌ی رنگی تقسیم شده، نشان داده شده است که طول آن نشانگر سهم هر کدام از K خوشه‌ی استنتاجی است

Figure 14. shows the cluster diagram of the population of Sarabi and Najdi Cows by the STRUCTURE software, each person is represented by a vertical line divided into K colored segments, the length of which represents the contribution of each of the K clusters

ساختار ژنتیکی و تنوع ژنتیکی گاوهای وحشی و اهلی بنگلادش با اسنیپ چیپ 80K بررسی شد. نتایج ساختار جمعیت و آنالیز مؤلفه‌ی اصلی پیشنهاد کرد که گابال<sup>1</sup> جدا از بوس ایندیکوس<sup>2</sup> بوده و دو جمعیت زبو<sup>3</sup> ساختار ضعیفی داشتند (Uzzaman et al. 2014). این روش‌هایی که ذکر شد (PCA و MDS)، جزء روش‌های بدون نظارت هستند که اطلاعات اولیه در آنالیزها وارد نمی‌شود. همچنین، این روش‌ها توانایی اجرای سریع داده‌های با حجم زیاد (صدها هزار نشانگر و هزاران نمونه) را دارند. در مقایسه دو روش PCA و MDS برای بررسی ساختار جمعیتی، در حالی که هر دو روش ساختار جمعیتی مشابهی را شناسایی کردند، PCA کمی بهتر از MDS در تصحیح لایه‌بندی جمعیتی عمل کرد (Wang et al. 2009).

در بررسی الگوهای اختلاطی و ساختار جمعیتی، جمعیت ساکنان شمال غرب آمریکای شمالی با روش Admixture نتایج برخی از شباهت‌ها را نشان داد، ولی اختلاف در بین الگوهای اختلاط در شمال غرب اقیانوس آرام و آمریکای لاتین وجود داشت (Verdu et al. 2014). روش‌های GC برای تست آماری<sup>4</sup> GWAS با مدل‌های متنوع توارثی می‌تواند اعمال شود (Tsepilov

2. Gaial

3. Bus indicus

4. Zebu

1. Genomic wide association study

(et al. 2013). در مطالعه‌ای برای بررسی لایه‌بندی جمعیتی روش‌های مختلفی مقایسه شدند که در این بین، روش‌های مبتنی بر مدل و PCA عملکرد مشابهی داشتند اما روش کنترل ژنومیک در جمعیت‌های با لایه‌بندی جمعیتی معنی‌دار، عملکرد خوبی نداشت (Zhang et al. 2008). از روش Admixture می‌توان برای تخمین تعداد جمعیت‌های پایه از طریق اعتبار سنجی متقابل استفاده کرد و افراد از انساب شناخته شده را برای دقت بیشتر تخمین اجداد بکار برد (Alexander & Lange 2011).

روش MDS انعطاف‌پذیرتر از روش PCA بود. PCA نیازمند پیش فرض توزیع نرمال چند جمله‌ای داده‌ها است در حالی که MDS می‌تواند از ماتریس کواریانس بدست آید، در این صورت نتایج دو روش در تصحیح لایه‌بندی جمعیتی یکسان می‌شود (Li & Yu 2008). مقادیر  $\Delta K$  محاسبه شده با استفاده از روش Evanno et al. (2005) توزیع مقادیر، نشان دهنده‌ی بهترین تعداد خوشه با توجه به ساختار داده‌ها است که برای این کار تعداد 2 خوشه مناسب شناخته شد (شکل 15).



شکل 15. مقادیر  $\Delta K$  محاسبه شده با استفاده از روش Evanno et al.، برای گاوهای سرابی و نجدی ایران  
 Figure 15. Calculated values of  $\Delta K$  using Evanno et al. For Iranian Sarabi and Najdi Cows

روش STRUCTURE که از رویکردهای استاندارد آماری برای برآورد پارامترهای جمعیتی استفاده می‌کند، به شدت به فرض‌های مدل وابسته است و زمانی که اندازه‌ی نمونه محدود و کوچک باشد، کارایی مناسبی ندارد؛ چرا که فراوانی‌های آلی را بدون دقت برآورد می‌کند. از طرفی به دلیل پیچیدگی محاسباتی، سرعت کمی دارد و محاسبات با تعداد زیاد نشانگر، قابل اجرا نیست. مزیت اصلی که در این بررسی به آن اشاره شد، کارایی محاسباتی بالا و کمترین نیاز به فرض‌های پیشین در روش SPC است و بنابراین به محقق این امکان را می‌دهد که اطلاعات هزاران حیوان را بدون داشتن هر گونه اطلاعاتی از شجره و نژاد، برای بررسی

ساختار جمعیتی تجزیه و تحلیل کند. نتایج به دست آمده از این تحقیق با نتایج حاصل از خوشه‌بندی سه جمعیت دامی شبیه‌سازی شده (Rahmanienia et al. 2015a) و بررسی تنوع ژنتکی و ساختار جمعیتی توده‌های گاومیش ایران با استفاده از نشانگرهای متراکم SNP مطابقت دارد (Rahmanienia et al. 2015a).

در این مطالعه با توجه به اینکه روش‌های مختلفی برای بررسی ساختار جمعیت بکار برده شد، و همه روش‌های ذکر شده توانستند ساختار جمعیت‌های دو توده نژادی را نشان دهند و نتایج بدست آمده حاکی از این بود که با وجود افراد خالص در این جمعیت‌ها، این افراد از دو جمعیت مختلف متعلق به یک نژاد هستند و اشتراک ژنتیکی زیادی دارند. نتایج این تحقیق در صورت وجود صفات فنوتیپی و نتایج حاصل از آنالیز GWAS و با استفاده از اطلاعات نژادهای دیگر و با در نظر گرفتن زیرجمعیت‌ها و مقایسه آن با نتایج حاصل از در نظر نگرفتن زیر جمعیت‌ها، می‌تواند منجر به فهم بهتر اهمیت بررسی ساختار و لایه‌بندی جمعیت شود.

### سپاسگزاری

از آقای دکتر جواد رحمانی‌نیا به دلیل همکاری در استفاده از نرم‌افزار SPIN، تشکر و قدردانی می‌گردد.

### منابع

- بحرینی بهزادی محمدرضا، امینی علی، اسلمی نژاد علی اصغر، طهمورث‌پور مجتبی (1392) برآورد پارامترهای ژنتیکی صفات تولیدی گاوهای شیری هولشتاین ایران. تحقیقات دامداری توسعه روستایی 25(9)، 21-24.
- توکلیان جواد (1377) ذخائر ژنتیکی دام و طیور بومی ایران ص 4-27.
- خراتی‌کوپایی حامد، محمدآبادی محمدرضا و همکاران (1390) تغییرات ژنتیکی ژن DGAT1 و ارتباط آن با تولید شیر در جمعیت گاو هولشتاین ایران. مجله علمی پژوهشی ایران 3(2)، 185-192.
- خراتی‌کوپایی حامد، محمدآبادی محمدرضا، ترنگ علی‌رضا و همکاران (1391) بررسی ارتباط بین تغییرات آلل ژن DGAT1 با ورم پستان در گاوهای هولشتاین ایران. مجله ژنتیک نوین 7(1)، 104-101.
- رحمانی‌نیا جواد، میرائی‌آشتیانی سیدرضا، مرادی شهربابک حسین (1394) تجزیه و تحلیل خوشه‌ای بدون نظارت ساختار جمعیت و زیرجمعیت با استفاده از نشانگرهای SNP متراکم. مجله علوم دامی ایران 3(46)، 277-877.
- رحمانی‌نیا جواد، میرائی‌آشتیانی سیدرضا، مرادی شهربابک حسین (1394) تنوع ژنتیکی و ساختار جمعیتی گاومیش‌های ایران با استفاده از نشانگرهای متراکم SNP. رساله دکترا. ص 120-114.
- علینقی‌زاده روح‌الله، محمدآبادی محمدرضا، زکی‌زاده سونیا (1389) چند شکلی اکزون 2 ژن BMP15 در بز قرمز جبال بارز. مجله بیوتکنولوژی کشاورزی 2(1)، 69-80.



- محمدآبادی محمدرضا، محمدی اکرم (1389) بررسی ژنوتیپ‌های بتا لاکتوگلوبولین در گاوهای بومی و هولشتاین استان کرمان. مجله تولیدات دامی 12(2)، 61-67.
- محمدی اکرم، محمدآبادی محمدرضا، میرزایی حمیدرضا و همکاران (1389) بررسی ژن کاپاکازین گاوهای شیری بومی و هولشتاین در استان کرمان با استفاده از روش PCR-RFLP. مجله علوم کشاورزی و منابع طبیعی 16(2) 125-132.
- واجدابراهیمی محمدتقی، محمدآبادی محمدرضا، اسمعیلی‌زاده کشکوئیه علی (1396) تجزیه و تحلیل تنوع ژنتیکی در پنج گروه گوسفند ایرانی با استفاده از نشانگرهای ریزماهوره. مجله بیوتکنولوژی کشاورزی 7، 143-158.
- هادی‌زاده مرتضی، محمدآبادی محمدرضا، نیازی علی و همکاران (1392) استفاده از ابزارهای بیوانفورماتیک برای مطالعه اگزون 2 ژن GDF9 در بز های تالی و بی‌تال. مجله ژنتیک نوین 8 (334)، 283-288.
- هادی‌زاده مرتضی، نیازی علی، محمدآبادی محمدرضا و همکاران (1392) تجزیه و تحلیل بیوانفورماتیک از اگزون 2 ژن BMP15 در بزهای تالی و بی‌تال. ژنتیک نوین 9(1)، 117-120.

## References

- Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12, 246.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655-1664.
- Alinaghizadeh H, Mohammadabadi MR, Zakizadeh S (2010) Exon 2 of BMP15 gene polymorphism in Jabal Barez Red Goat. Agric Biotechnol J 2, 69-80 (In Persian).
- Barazandeh A, Mohammadabadi MR, Ghaderi M, Nezamabadipour H (2016) Predicting CpG Islands and Their Relationship with Genomic Feature in Cattle by Hidden Markov Model Algorithm. Iran J Appl Anim Sci 6, 571-579.
- Behzadi Behzadi MR, Amini A, Slaminejad A, Tahmoospour M (2013) Estimation of genetic parameters for production traits of Iranian Holstein dairy cattle. Livest Res Rural Develop 25, 21-24 (In Persian).
- Bowden R, MacFie TS, Myers S et al. (2012) Genomic tools for evolution and conservation in the chimpanzee: Pan troglodyteselliotti is a genetically distinct population. PLOS Genet 8, e1002504.
- Brohée S, Faust K, Lima-Mendez G et al. (2008) NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. Nucleic Acids Res 36, W444-W451.
- Decker JE, Pires JC, Conant GC et al. (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proceedings of the National Academy of Sciences 106, 18644-18649.

- Domany E (2003) Cluster analysis of gene expression data. *J Statistical Physics* 110, 1117-1139.
- Ebrahimi Z, Mohammadabadi MR, Esmailzadeh AK et al. (2015a). Association of PIT1 gene with milk fat percentage in Holstein cattle. *Iran J Appl Anim Sci* 5, 575-582.
- Ebrahimi Z, Mohammadabadi MR, Esmailzadeh AK et al. (2015b) Association of PIT1 gene and milk protein percentage in Holstein cattle. *J Livest Sci Technol* 3, 41-49.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14, 2611-2620.
- Faraji-Arough H, Aslaminejad A, Tahmoorespur M et al. (2015) Bayesian Inference of (Co) Variance Components and Genetic Parameters for Economic Traits in Iranian Holsteins via Gibbs Sampling. *Iran J Appl Anim Sci* 5, 51-60.
- Fernández ME, Goszczynski DE, Lirón JP et al. (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genet Mol Biol* 36, 185-191.
- Gao X, Starmer J (2007) Human population structure detection via multilocus genotype clustering. *BMC Genet* 8, 34.
- Gautier M, Laloë D, Moazami-Goudarzi K (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PloS One* 5, e13038.
- Getz G, Gal H, Kela I, Notterman DA et al. (2003) Coupled two-way clustering analysis of breast cancer and colon cancer gene expression data. *Bioinformatics* 19, 1079-1089.
- Getz G, Levine E, Domany E (2000a) Coupled two-way clustering analysis of gene microarray data. *Proceedings of the National Academy of Sciences* 97, 12079-12084.
- Getz G, Levine E, Domany E et al. (2000b) Super-paramagnetic clustering of yeast gene expression profiles. *Physica A: Statistic Mechanic Appl* 279, 457-464.
- Ghasemi M, Baghizadeh A, Abadi MRM (2010) Determination of genetic polymorphism in Kerman Holstein and Jersey cattle population using ISSR markers. *Australian J Basic Appl Sci* 4, 5758-5760
- Hadizadeh M, Mohammadabadi MR, Niazi A et al. (2013) Use of bioinformatics tools to study exon 2 of GDF9 gene in Tali and Beetal goats. *Modern Genet J* 8, 283-288 (In Persian).
- Hadizadeh M, Niazi A, Mohammadabadi MR et al. (2014) Bioinformatics analysis of the BMP15 exon 2 in Tali and Beetal goats. *Modern Genet J* 9, 117-120 (In Persian).
- Heaton MP, Keen JE, Clawson ML et al. (2005) Use of bovine single nucleotide polymorphism markers to verify sample tracking in beef processing. *J Am Vet Med Assoc* 226, 1311-1314.
- Holmström E, Bock N, Brännlund J (2009) Modularity density of network community divisions. *Physica D: Nonlinear Phenomena* 238, 1161-1167.

- Karimi K, Strucken Eva M, Moghaddar N et al. (2016) Local and global patterns of admixture and population structure in Iranian native cattle. *BMC Genet* 17, 108.
- Karimi K, Beigi Nassiri MT, Mirzadeh Kh et al. (2009) Polymorphism of the  $\beta$ -lactoglobulin gene and its association with milk production traits in Iranian Najdi cattle. *Iran J Biotchnol* 7, 82-85.
- Khansefid SM (2009) An introduction to Sarabi cow breed a vernacular cow breed of Iran. *Nolan Special Magazine on Dairy cows*. No3/Jan-Mar.
- Kharrati Koopaei H, Mohammadabadi MR, Ansari.Mahyari S et al. (2012a) Effect of DGAT1 variants on milk composition traits in Iranian Holstein cattle population. *Anim Sci Paper Report* 30, 231-240
- Kharrati Koopaei H, Mohammadabadi MR, Ansari Mehryari S et al. (2011) Genetic Variation of DGAT1 Gene and its Association with Milk Production in Iranian Holstein Cattle Breed Population. *Iran J Anim Sci Res* 3, 185-192 (In Persian).
- Kharrati Koopaei H, Mohammadabadi MR, Tarang A et al. (2012b) Study of the association between the allelic variations in DGAT1 gene with mastitis in Iranian Holstein cattle. *Modern Genet J* 7, 101-104 (In Persian).
- 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, 331.
- Kijas JW, Townley D, Dalrymple BP et al. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS One* 4, e4668.
- Lee C, Abdool A, Huang CH (2009) PCA-based population structure inference with generic clustering algorithms. *BMC Bioinformatics* 10, S73.
- Levy M, Feingold J (2000) Estimating prevalence in single-gene kidney diseases progressing to renal failure. *Kidney International* 58, 925-943.
- Li Q, Yu K (2008) Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* 32, 215-226.
- Lirón J, Ripoli M, De Luca J et al. (2002) Analysis of genetic diversity and population structure in Argentine and Bolivian Creole cattle using five loci related to milk production. *Genet Mol Biol* 25, 413-419.
- Markovtsova L, Marjoram P, Tavaré S (2000) The age of a unique event polymorphism. *Genet* 156, 401-409.
- Marquitti FMD, Guimarães PR, Pires MM, Bittencourt LF (2014) MODULAR: software for the autonomous computation of modularity in large network sets. *Ecography* 37, 221-224.

- Meuwissen T, Solberg TR, Shepherd R, Woolliams JA (2009) A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genet Sel Evol* 41, Doi: 10.1186/1297-9686-41-2.
- Mohammadabadi MR, Mohammadi A (2010) Study of beta-lactoglobulin genotypes in native and Holstein cattle of Kerman province. *J Anim Prod* 12, 61-67 (In Persian).
- Mohammadabadi MR (2017) Role of clostridium perfringens in pathogenicity of some domestic animals. *Journal Adv Agric* 7, 1117-1121.
- Mohammadabadi MR, Esfandyarpoor E, Mousapour A (2017) Using Inter Simple Sequence Repeat Multi-Loci Markers for Studying Genetic Diversity in Kermani Sheep. *J Res Dev* 5, 154-157.
- Mohammadi A, Mohammadabadi MR, Mirzaei H et al. (2009) Study of Kappa Casein gene of local and Holstein dairy cattle in Kerman province using PCR-RFLP method. *J Agric Sci Natural Resources* 16, 125-132 (In Persian).
- Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends Ecol Evol* 19, 208-216.
- Neuditschko M, Khatkar MS, Raadsma HW (2012) NetView: A High-Definition Network-Visualization Approach to Detect Fine-Scale Population Structures from Genome-Wide Patterns of Variation. *PloS One* 7, e48375.
- Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev* 69, 026113.
- Ojango J, Pollott G (2001) Genetics of milk yield and fertility traits in Holstein-Friesian cattle on large-scale Kenyan farms. *J Anim Sci* 79, 1742-1750.
- Pasandideh M, Mohammadabadi MR, Esmailizadeh AK, Tarang A (2015) Association of bovine PPARGC1A and OPN genes with milk production and composition in Holstein cattle. *Czech J Anim Sci* 60, 97-104.
- Paschou P, Ziv E, Burchard EG, Choudhry S et al. (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 3, e160.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genet* 155, 945-959.
- 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.
- Radjiman & Sugiarto (2005) Super Paramagnetic Clustering of DNA Sequences. Master of Science dissertation. National University of Singapore.

- Rahmanienia J, Miri Ashtiani SR, Moradi Shahrbabak H (2015a) Unsupervised clustering analysis of population and subpopulation structure using dense SNP markers. *Iran J Anim Sci* 46, 287-277 (In Persian).
- Rahmanienia J, Miriya Ashtiani R, Moradi Shahrbabak H (2015b) Genetic diversity and demographic structure of Iranian buffaloes using dense SNP markers. Ph.D. thesis p. 120-114 (In Persian).
- Sargolzaei M, Schenkel FS (2009) QMSim: a large-scale genome simulator for livestock. *Bioinform* 25, 680-681.
- Seo TK, Thorne JL, Hasegawa M, Kishino H (2002) Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genet* 160, 1283-1293.
- Serre D, Montpetit A, Paré G et al. (2008) Correction of population stratification in large multi-ethnic association studies. *PloS One* 3, e1382.
- Shannon P, Markiel A, Ozier O et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498-2504.
- Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28, 289-301.
- Tavakkolian J (2000) Genetic Resources of Native Farm Animals in Iran. P. 4-27 (In Persian).
- Tetko IV, Facius A, Ruepp A, Mewes HW (2005) Super paramagnetic clustering of protein sequences. *BMC Bioinform* 6, 82.
- Troy CS, MacHugh DE, Bailey JF et al. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088-1091.
- Tsafirir D, Tsafirir I, Ein-Dor L et al. (2005) Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinform* 21, 2301-2308.
- Tsepilov YA, Ried JS, Strauch K et al. (2013) Development and application of genomic control methods for genome-wide association studies using non-additive models. *Plos One* 8, e81431.
- Uzzaman MR, Edea Z, Bhuiyan MSA et al. (2014) Genome-wide Single Nucleotide Polymorphism Analyses Reveal Genetic Diversity and Structure of Wild and Domestic Cattle in Bangladesh. *Asian-Australas J Anim Sci* 27, 1381.
- Vahidi SM, Tarang AR, Naqvi AU et al. (2014) Investigation of the genetic diversity of domestic *Capra hircus* breeds reared within an early goat domestication area in Iran. *Genet Sel Evol* 46, 27.
- Vajed Ebrahimi MT, Mohammadabadi MR, Esmailizadeh AK (2016) Analysis of genetic diversity in five Iranian sheep population using microsatellites markers. *Agric Biotechnol J* 7, 143-158 (In Persian).

- Vajed Ebrahimi MT, Mohammadabadi MR, Esmailizadeh AK (2017) Using microsatellite markers to analyze genetic diversity in 14 sheep types in Iran. *Archiv fuer Tierzucht (Archives Anim Breed)* 60, 183-189.
- Verdu P, Pemberton TJ, Laurent R et al. (2014) Patterns of admixture and population structure in native populations of Northwest North America. *Plos Genet* 10, e1004530.
- Wang D, Sun Y, Stang P et al. (2009) Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling, *BMC proceedings, BioMed Central* 3, S109.
- Wang DG, Fan JB, Siao CJ et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Sci* 280, 1077-1082.
- Zamani P, Akhondi M, Mohammadabadi MR (2015) Associations of Inter-Simple Sequence Repeat loci with predicted breeding values of body weight in sheep. *Small Rumin Res* 132, 123-127.