نشریه پژوهش‌های علوم دامی/ جلد ۲۸ شماره ۱/ سال۱۳۹۷/ صفحات ۱۸۱ تا ۱۹۴

بررسی ژنومی ساختار جمعیتی و ارتباط فیلوژنتیکی گاومیش نژاد خوزستانی

محمدرضا زرگر[۱]، جمال فیاضی*[۲]، محمدتقی بیگی نصیری[۳] و حسین مرادی شهربابک[۴]

تاریخ دریافت:۱۳۹۶/۷/۲۲ تاریخ پذیرش: ۱۳۹۶/۱۲/۲

[۱و۲و۳] به ترتیب دانش آموخته کارشناسی ارشد، دانشیار و استاد گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان

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

* مسئول مکاتبه: Email: j_fayazi@ramin.ac.ir

چکیده

زمینه مطالعاتی: باتوجه به اهمیت گاومیش در جهت سازگاری با محیط‌های خشک، مقاومت در برابر بیماری‌ها، پایین بودن هزینه‌های نگهداری و استفاده موثر از مواد خشبی کم ارزش در جهت تولید مواد پروتئینی با ارزش، ارتقا ژنتیکی آن حائز اهمیت فراوانی است. **هدف**: این تحقیق به منظور یافتن فواصل ژنتیکی گاومیش‌های مناطق مختلف خوزستان و ارتباط ژنتیکی آنها به کمک آرایه چندشکلی تک نوکلئوتیدی اجرا شد. **روش کار**: در این مطالعه از تعداد ۱۲۱ راس گاومیش از گله‌هایی که تحت پوشش سیستم ثبت مشخصات و رکوردگیری شیر در شهرستان‌های اهواز، دزفول، شادگان، شوش، شوشتر، دشت آزادگان، کرمانشاهی با منشاء خوزستانی و سایر مناطق بودند، نمونه‌گیری شد. نمونه-ها با استفاده از تراشه ژنومی اختصاصی گاومیش با تعداد ۹۰ هزار نشانگر SNP تعیین ژنوتیپ شدند. سپس جایگاه-های با حداقل فراوانی آللی کمتر از ۰/۰۱، تعادل هاردی واینبرگ کمتر از 10^{-6} و نرخ خوانش کمتر از ۰/۰۵ کنار گذاشته شدند. داده‌ها بکمک روش‌های آماری چند متغییره همچون آنالیز مولفه‌های اصلی مورد کنکاش قرار گرفتند. **نتایج**: نتایج آنالیز روابط فیلوژنتیکی نشان داد گاومیش‌های اهواز بدلیل مرکزیت، متاثر از گاومیش‌های سایر نقاط استان می‌باشند، ولی گاومیش‌های شوشتر و دزفول و دشت‌آزادگان، نسبت به اهواز اختلاط کمتری دارند. با استفاده از آنالیز مولفه‌های اصلی نتایج فوق تایید شد. جهت بررسی الگو و ساختار ژنتیکی برای جمعیت‌های گاومیش استان خوزستان ارزش‌های F_{ST} برای هر SNP به روش نااریب تتا محاسبه گردید. بیشترین میانگین ارزش F_{ST} بین جمعیت مربوط به گاومیش‌های شهرستان‌های دشت‌آزادگان و شادگان به ترتیب ۰/۰۱۵۹ و ۰/۰۱٤۷ و کمترین مربوط به شهرستان‌های اهواز و شوش به ترتیب ۰/۰۰۸۷ و ۰/۰۰۸٦ بود. آزمون‌های آماری ژنتیکی جمعیتی و تنوع نژادی نشان داد که در درون جمعیت گاومیش‌های خوزستانی تنوع ژنتیکی بالایی وجود دارد. **نتیجه گیری نهایی**: نتایج گرافیکی آزمون‌های خوشه‌بندی و انتساب حاکی از وجود حداقل ۳ زیر جمعیت نسبتاً مجزا در گاومیش‌های خوزستان می‌باشد. نتایج این تحقیق در مورد تنوع ژنتیکی مشاهده شده در سطح نوکلئوتیدها و فاصله ژنتیکی بین گروه‌های جمعیتی گاومیش استان، در مطالعات بعدی همچون تشکیل جمعیت پایه گاومیش خوزستان و انتخاب ژنومیکی شایسته توجه هستند.

واژگان کلیدی : آنالیز مولفه‌های اصلی، پویش کل ژنومی، تنوع تک نوکلئوتیدی، گاومیش خوزستانی

مقدمه

دام‌های اهلی از طریق فرایند انتخاب، تکامل و اهلی شدن، حاصل شده و دستخوش تغییرات ژنتیکی مانند رانش ژنتیکی، جهش و انتخاب مصنوعی شده‌اند. بخاطر اثرات منفی ایجاد شده بر اثر این تغییرات و استفاده وسیع از تعداد کمی از نژادهای حیوانی متعلق به گونه-های خاص به دلیل نیاز به تولید بالاتر، تنوع ژنتیکی دام‌های اهلی کاهش یافته و نگرانی‌های شدیدی را برای سیستم تولید حیوانات سراسر جهان بوجود آورده است (برومند جزی و فرازمند ۲۰۰۵ و فائو ۲۰۱۵) . گاومیش‌های ایران به دلیل سازگاری با محیط، مقاومت در برابر بیماری‌ها، پایین بودن هزینه‌های نگهداری و استفاده مؤثر از مواد خشبی کم ارزش، یکی از ذخایر ژنتیکی با ارزش محسوب می‌شوند. همچنین این حیوان نقش اساسی در اشتغال اقشار کم درآمد روستایی و تأمین بخشی از نیاز غذایی کشور به شیر و گوشت دارد (برومند جزی و فرازمند ۲۰۰۵).

یکی از مهمترین گاومیش‌های ایران، گاومیش‌های خوزستانی هستند که گستردگی زیادی دارند. این گاومیش‌ها با توجه به اینکه ممکن است به چند اکوتیپ و زیر جمعیت تقسیم شوند، پویش ژنومی را با مشکلاتی همراه خواهند کرد. یکی از چالش‌هایی که همواره درست بودن مطالعات پیوستگی ژنومی[1] و تنوع ژنتیکی را تهدید می‌کند، وجود زیر جمعیت‌ها، درون جمعیت‌های مورد بررسی است. به عنوان مثال اگر در یک پژوهش بیمار-شاهد یکی از دو زیر جمعیت مورد بررسی متعلق به یک نژاد و زیر جمعیت دیگر از نژادی دیگر باشد، همواره این امکان وجود دارد که جایگاه ژنتیکی معرفی شده با اثر معنی‌دار، منشاء گرفته از تفاوت در صفت مورد بررسی نباشد، بلکه ناشی از تفاوت‌های نژادی دو زیر جمعیت یاد شده باشند. بنابراین باید سعی شود که زیر جمعیت‌های انتخاب شده در بررسی‌ها تا حد امکان برای دیگر صفات همگن باشند و تفاوت‌های آن‌ها تنها در صفت مورد بررسی باشد (توماس و ویت ۲۰۰۲).

سودآوری گاومیش تابع میزان تولید شیر در هر زایش و درصد چربی آن می‌باشد. هدف اصلی از اصلاح‌نژاد گاومیش افزایش راندمان تولید در حیوانات گله از طریق ایجاد پیشرفت ژنتیکی برای صفات تولیدی است. در حوزه ژنتیک و اصلاح نژاد دام، اطلاع از ساختار ژنتیکی جمعیت می‌تواند کمک بزرگی به برنامه‌ریزی برای طرح-های اصلاح‌نژادی و از همه مهمتر، حفظ این ذخیره ژنتیکی کند (برومند جزی و فرازمند ۲۰۰۵).

استفاده از نشانگرهای ژنتیکی با اطلاعات زیاد جهت مطالعه منشاء، تاریخچه و تکامل جوامع دامی و تعیین گونه‌های تحت حفاظت در مبحث زیست‌شناسی حفاظتی بسیار ضروری است. استفاده از نشانگرهای ریزماهواره برای بیش از یک دهه است که مرسوم گردیده ولی امروزه دامنه کاربرد استفاده از چند شکلی-های تک‌نوکلئوتیدی[2] سبب برتری این نشانگرها بر دیگر نشانگرهای ژنتیکی مانند ریزماهواره‌ها گردیده است (کریمی و همکاران ۲۰۱۷، عارف و همکاران ۲۰۱۱، لی و همکاران، ۲۰۰۹، مدگوراک و همکاران ۲۰۰۹).

عبارت ژنومیک جمعیت[3] اولین بار در سال ۱۹۹۸ توسط استفان سون[4] و گولچس[5] به کار رفت و برای آنالیز تنوع جمعیتی درون و بین جمعیت‌ها بر اساس داده‌های کل ژنوم استفاده شد و خصوصیات خاص جمعیت‌ها از قبیل اندازه موثر جمعیت و تمایز جمعیت‌ها با استفاده از روش‌های آنالیز ژنومیک جمعیت استنباط گردید (سیمیانر۲۰۱٤).

استفاده از نشانگرهای متراکم اگر چه در دنیا به طور گسترده مورد استفاده قرار گرفته است (کریمی و همکاران ۲۰۱۵)، اما در مورد دام‌های بومی خصوصاً گاومیش در مراحل مقدماتی به سر می‌برد. اگرچه در

[1] Genome Wide Association Study

[2] SNP

[3] population genomics

[4] Stefan son

[5] Gulches

دنیا رفرنس ژنوم گاومیش در حال تکمیل است (ویلیامز و همکاران ۲۰۱۷) اما در ایران مطالعات ژنومیکی کمی در مورد گاومیش وجود دارد. در مورد پژوهشهای ژنومیکی گاومیش خوزستان شرایط بهتر نیست. مطالعات محدودی در مورد بررسی بخشی از ژنوم گاومیش خوزستان انجام شده است. در سه پژوهش مجزا در جمعیت گاومیش خوزستان، با استفاده از تکنیک PCR-RFLP چند شکلی ژن پرولاکتین، کاپاکازئین و PIT1 مورد بررسی قرار گرفت و حالت تک شکلی یا مونومورف در آن جایگاه‌های ژنی گزارش داده شد (شجاعی و همکاران ۲۰۰۹، عباسی مشایی و همکاران ۲۰۰۹، روزگار و همکاران ۲۰۰۹).

در طی تحقیقی پویش ژنومی جهت شناسایی نشانه‌های انتخاب در نژادهای گاومیش خورستانی و مازندرانی مورد مطالعه قرار گرفت و نشانه‌های انتخاب در بخش-هایی از ژنوم شناسایی شد (مخبر و همکاران ۲۰۱۵) که در تحقیقات قبلی در انسان و گاو و دام‌های دیگر نیز تایید شده بودند (ویقت و همکاران ۲۰۰۶، اوتسانومیا و همکاران ۲۰۱۳).

رحمانی نیا و همکاران(۲۰۱۵) در بررسی ساختارهای جوامع و خرده جوامع دامی به روش خوشه بندی شبکه‌ای بدون نظارت[1]با استفاده از نشانگرهای ژنتیکی متراکم نشان دادند که حیوانات براساس شباهت‌ها و تفاوت‌ها به خوبی در جوامع مربوطه قرار گرفتند و خرده جوامع موجود نیز درون جمعیت نمایان شدند. مزیت اصلی این روش، کارایی محاسباتی بالا و نیاز نبودن به فرض‌های پیشین در آن است، بنابراین به محقق این امکان را می‌دهد که ساختار جوامع متشکل از هزاران حیوان را بدون داشتن هرگونه اطلاعاتی از شجره و نژاد، تجزیه وتحلیل کند.

در بررسی ساختار و لایه‌بندی جمعیت گاومیش‌های اکوتیپ آذری و شمالی با نشانگرهای متراکم با استفاده از آنالیز مولفه‌های اصلی مشخص شد بین گاومیش‌های آذری با شمالی ارتباط ژنتیکی وجود دارد (عزیزی و همکاران ۲۰۱۶). این محققین در پژوهشی دیگر نشان دادند باروش یادگیری عصبی (SVM) افراد با ژنوتیپ مشخص، میتوانند با درستی بالایی به نژاد، منطقه یا گله ای که به آن تعلق دارند، اختصاص یابند به طوری که اگر شماری نمونه با هویت مجهول وجود داشته باشد میتوان با این روش به نژاد یا جمعیتی که به آن تعلق دارند اختصاص داده شود (عزیزی و همکاران ۲۰۱۶ب). همانگونه که اشاره شد، بررسی کل ژنوم گاومیش‌های خوزستانی و یافتن ارتباط گروه‌های احتمالی ژنتیکی در بین گاومیش خوزستانی انجام نشده است اما درکاوش ژنومیکی نشانه‌های انتخاب در نژادهای مختلف گاومیش ایرانی با استفاده از تعداد زیادی SNP، کل گاومیش‌های ایران از نظر قرابت و تنوع ژنتیکی مورد مطالعه قرار گرفتند (مخبر, ۲۰۱۵). لذا با توجه به اینکه مطالعه‌ای در خصوص قرابت و تنوع ژنتیکی گاومیش-های زیست بوم مناطق مختلف خوزستان انجام نشده بود، هدف این مطالعه تکمیل نتایج مطالعات قبلی (رحمانی‌نیا و همکاران ۲۰۱۵، مخبر و همکاران ۲۰۱۵، و عزیزی و همکاران ۲۰۱۶) به کمک بررسی کل ژنوم با استفاده از تراشه Axiom @ Buffalo Genotyping 90 K مربوط به شرکت افی‌متریکس[2] جهت بررسی تنوع داخل نژادی گاومیش خوزستانی بود.

مواد و روش‌ها

در این مطالعه از تعداد ۱۲۱ راس از گاومیش‌های شهرستان‌های گاومیش‌خیز استان خوزستان شامل ۲۸ نمونه از اهواز، دزفول ۱۹ نمونه، شادگان ٦ نمونه، شوش ٤ نمونه، شوشتر ۹ نمونه، دشت آزادگان ۲٤ نمونه، کرمانشاهی با منشاء خوزستانی ۸ نمونه و سایر مناطق ۱۷ نمونه استفاده شد. نمونه‌ها از گله‌هایی که تحت سیستم ثبت شجره و رکوردگیری مرکز اصلاح نژاد کشور قرار گرفته بودند گردآوری شد. در انتخاب

[1] Unsupervised approaches

[2] Affymetrix

حیوانات تاکید بر انتخاب حیوانات غیرخویشاوند و از شهرستان‌های مختلف استان خوزستان و اکوتیپ‌های مشابه در کشور بود. استخراج DNA ژنوم از ریشه مو و خون با روش بهینه نمکی انجام شد. باتوجه به اینکه این پروژه از محل اعتبارات ملی استان خوزستان توسط شرکت دانشگاهی نواندیش البرز انجام شده بود نمونه برای انجام مراحل بعدی تعیین ژنوتیپ توسط شرکت مذکور به آزمایشگاه ژنومیک مرکز تحقیقات پادانو[1]کشور ایتالیا منتقل شدند. نمونه‌ها با استفاده از تراشه‌های Array Axiom @ Buffalo Genotyping 90K مربوط به شرکت افی متریکس تعیین ژنوتیپ شدند. این آرایه‌ها امکان تعیین ژنوتیپ بیش از ۸۵ هزار جایگاه نشانگری SNP را فراهم می‌آورند (رحمانی نیا و همکاران ۲۰۱۵). هفتاد درصد SNPهای این آرایه برای مطالعه انواع جمعیت‌های گاومیش رودخانه ای بسیار مناسب تشخیص داده شده اند اما تنها ۳۰ درصد آنها در گاومیش باتلاقی قابل استفاده هستند (ایامارتینو و همکاران ۲۰۱۷).

برای اطمینان از بالا بودن کیفیت داده‌های مورد استفاده در آنالیزهای بعدی، با اعمال ویرایش و محدودیت‌هایی ژنوتیپ‌های با کیفیت پایین و داده‌های دارای اطلاعات کم حذف شدند. این بخش با نرم افزار PLINK انجام گرفت (پورسل و همکاران 2007). داده‌های با اطلاعات کم شامل ژنوتیپ‌های دارای حداقل فراوانی آللی[2] (MAF) کمتر از ۰/۰۱ و SNPهایی که p-value تعادل هاردی واینبرگ آنها کمتر از 10^{-6} و نرخ خوانش[3] آنها کمتر از ۰/۰۵ بود، کنار گذاشته شدند(گودارد و هیز ۲۰۰۹).

بنابراین پس از تعیین ژنوتیپ، عمل غربالگری روی ٦٤٧٥٠ SNP انجام و SNPهای دارای اطلاعات وارد مرحله تجزیه و تحلیل شدند. پس از کنترل کیفیت اولیه ۸۸۵۵ SNP به دلیل MAF کمتر از ۱ درصد و ۳۲٦ SNP به دلیل نبود تعادل هاردی واینبرگ و ۱۹ SNP به خاطر موقعیت ناشناخته حذف شدند. در نهایت ۱۲۱ حیوان با ٦٤٧٠٩ SNP مراحل کنترل کیفیت را گذراندند.

در این مطالعه از روش‌های مختلف برای ارزیابی ساختار ژنتیکی جمعیت استفاده شد. روش درخت اتصال همسایگی[4] روش خوشه‌بندی بر پایه فاصله ژنتیکی است که توسط سایتو و نی ۱۹۸۷ ارائه شد. الگوریتم NJ با یک ساختار شبیه تنه درخت آغاز میشود و مجاورین درست (شاخه‌هایی که از لحاظ فاصله نزدیک‌ترند) را با کمینه کردن مجموع طول همه شاخه‌ها می‌یابد. از این الگوریتم برای رسم درختواره تبارزایی (فیلوژنیک) بر مبنای فاصله ژنتیکی برای هر جفت از افراد با استفاده از پکیج ape نرم افزار R استفاده شد (پارادایس و همکاران ۲۰۰٤). این نرم افزار فاصله بین دو نفر از افراد را بدست آورده و برای آنها جایگاه متفاوتی را تعریف میکند. نمودار فیلوژنتیکی بر اساس ماتریس فاصله‌ها و با استفاده از پکیج Phyclust نرم افزار R بدست آمده است (چن و دورمان ۲۰۱۰). جهت داشتن یک دیدگاه کلی در خصوص ساختار جمعیتی حیوانات و نژادهای مورد مطالعه و شناسایی حیواناتی که خارج از گروه نژادی خود قرار گرفته‌اند، آنالیز مولفه‌های اصلی[5]بوسیله همه ٦٤٧٠٩ SNP در دسترس، و آنالیزDAPC[6] با استفاده از پکیجِ adegent نرم افزار R انجام شد (جومبرت و همکاران ۲۰۱۰) . برای شناسایی تعداد بهینه خوشه‌ها، K-means به صورت متوالی با افزایش مقادیر K، اجرا می‌شود و جواب‌های خوشه‌بندی مختلف با استفاده از معیار بیزی (BIC) مقایسه شد. سپس با بررسی رابطه خویشاوندی درون نژادی نمودارهای Plotnj و PCA و DAPC توسط پکیج‌های

[1] Parco Tecnologico Padano
[2] Minor Allele Ferequency
[3] Call rate
[4] Neighbor-joining
[5] Principal Component Analysis(PCA)
[6] Discriminant Analysis of Principal Components

نرم افزار R ترسیم شد. ارزش‌های F_{ST} (تتا) برای کروموزوم‌های اتوزومی و شهرستان‌ها به روش نااریب محاسبه شد. سپس میانگین فاصله ژنتیکی هر شهرستان با سایر شهرستان‌ها محاسبه گردید.

نتایج و بحث

با اعمال ویرایش و کنترل کیفیت روی اطلاعات خام تعیین ژنوتیپ شده در مجموع ۱۲۱ حیوان از شش شهرستان مختلف با ٦٤٧٠٩ نشانگر SNP وارد آنالیز نهایی شدند.

در این مطالعه ابتدا میانگین F_{ST} در جمعیت‌های گاومیش خوزستانی محاسبه گردید که نتایج این محاسبات در جدول شماره ۱ ارائه شده است. این معیار نشان دهنده توزیع تنوع بین گروه و درون گروه است. بالا بودن آن نشان دهنده اینست که بخش عمده هتروزیگوسیتی ناشی از تفاوت در بین گروه‌ها است و بلعکس پایین بودن آن حاکی از اینست که تنوع موجود ناشی از تنوع افراد درون گروه است و خود جمعیتها با هم قرابت دارند. در این جدول، بیشترین ارزش Fst بدست آمده مربوط به گاومیش‌های شهرستان‌های دشت‌آزادگان و شادگان و کمترین مربوط به اهواز و شوش بود. این امر نشان دهنده این است که گاومیش-های شهرستان‌های دشت‌آزادگان و شادگان به دلیل شرایط خاص جغرافیای آن شهرستان‌ها و دور بودن از مرکز و در حاشیه بودن، و مهم تر از همه عدم انتخاب ژنتیکی، بیشترین هتروزیگوسیتی را در بین جمعیت‌های گاومیش خوزستان نشان می‌دهند. بر عکس این مطلب برای اهواز قابل تصور است. یعنی کمترین Fst را نشان داد. با توجه به اینکه شهرستان اهواز در مرکز استان قرار دارد و گاومیش داران از سطح سواد بهتری بخاطر نزدیکی با مراکز آموزشی دارند لذا احتمالاً انتخاب دامها بیشتر به وقوع پیوسته و سطح هتروزیگوسیتی را پایین آورده است. مخبر و همکاران (۲۰۱۵) در تحقیق خود در مورد پویش ژنومی نشانه‌های انتخاب در گاومیش‌های خوزستانی و مازندرانی میانگین این شاخص را برای تک تک اسنیپ‌ها محاسبه نمودند. میانگین آن در کروموزومهای غیرجنسی و در دو جمعیت مازندرانی و خوزستانی ۰/۰۳۲ محاسبه شد که از مقادیر محاسبه شده این تحقیق بالاتر بود. دلیل پایین بودن این شاخص می‌تواند بخاطر قرابت ژنتیکی بیشتر زیرجمعیت‌های خوزستانی با هم نسبت به ادغام دو جمعیت خوزستانی و مازندرانی با هم دانست. رحمانی نیا و همکاران (۲۰۱۵) در مطالعه ای دیگر، ارزش F_{ST} میان جمعیت گاومیش‌های استان‌های مختلف گاومیش خیز کشور را محاسبه نمودند، در آن مطالعه ارزش F_{ST} گاومیش‌های خوزستان با اردبیل ۰/۱۱۹ ، با آذربایجان غربی ۰/۱۱۳ ، با آذربایجان شرقی ۰/۱۱۲ ، با گیلان ۰/۱۱٤ و با مازندران ۰/۱٦۱ محاسبه شد که نشان دهنده این موضوع است که ارزش F_{ST} بین گاومیش‌های استان‌های مختلف نسبت به گاومیش‌های شهرستان‌های مختلف یک استان که در این مطالعه بدست آمده است بزرگتر میباشد. البته این موضوع طبیعی و منطقی بوده و نشان دهنده فاصله ژنتیکی بیشتر دامهای استان‌های مختلف میباشد.

F_{ST} نشان دهنده هتروزیگوسیتی بین جمعیت‌ها میباشد و هرچه F_{ST} بین دوجمعیت بیشتر باشد هتروزیگوسیتی بین آن دو جمعیت افزایش و هتروزیگوسیتی درون آن جمعیت‌ها کاهش پیدا میکند. F_{ST}های جدول ۱ به روش نااریب تتا محاسبه شده‌اند. مزیت این روش نسبت به روش پایه‌ای F_{ST} که توسط رایت ارایه شده، این است که در روش تتا اندازه نمونه‌ها در فرمول لحاظ گردیده و در واقع خطای نمونه‌گیری در نظر گرفته شده است (ویر و کوکرهام ۱۹۸٤). با توجه به اینکه این روش نااریب است حتی احتمال به دست آمدن ارزش‌های منفی نیز وجود دارد (اکی ۲۰۰۹).

جدول ۱ - میانگین هتروزیگوسیتی SNPها (FST) در جمعیت‌ها ی گاومیش خوزستانی

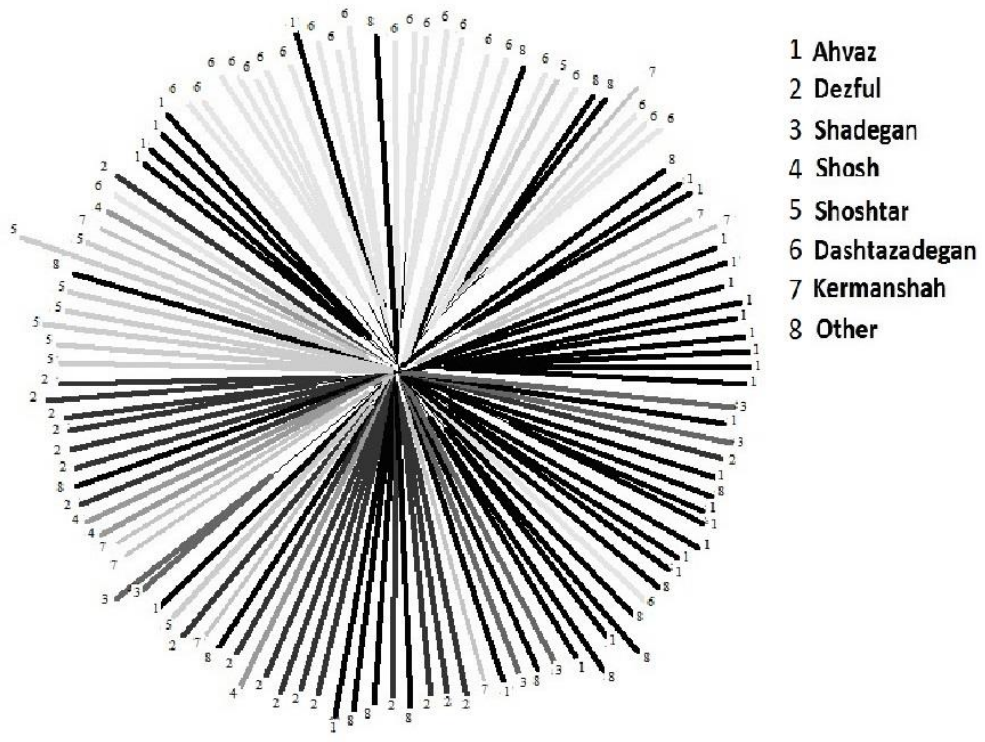
Table 1. SNPs Average Heterozigosity (FST) among Khuzestani buffalo populations

شهرستان City	میانگین هتروزیگوسیتی اسنیپ‌ها SNPs Average Heterozigosity (F_{ST})
دشت‌آزادگان Dashtazadegan	0.0160
شادگان Shadegan	0.0147
دزفول Dezfoul	0.0094
شوشتر Shoshtar	0.0091
اهواز Ahvaz	0.0087
شوش Shosh	0.0086
سایر Other	0.0054

قدم بعد یافتن میزان تعلق تک‌تک گاومیش‌های مورد مطالعه نسبت به هم به صورت گرافیکی است. در واقع اگر بخواهیم نوعی خوشه‌بندی و قرابت تک‌تک افراد را بررسی کنیم روش گرافیکی مبتنی بر اتصال همسایه‌ها مناسب است. این روش توسط نمودار درخت همسایگی[1] با استفاده از اطلاعات ژنوتیپ‌های استخراج شده که به صورت فرمت نوکلئوتیدی ذخیره شده بودند، توسط نرم‌افزار R برای جمعیت خوزستانی رسم گردید، که در شکل ۱ نشان داده شده است.

درشکل ۱ هر شعاع نشان دهنده یکی از گاومیش‌های مناطق مختلف استان خوزستان می‌باشد. گاومیش‌های اهواز که در شکل ۱ با شماره ۱ میباشند، یکجا تجمع کاملی نداشته بلکه در بین آن شعاع‌های مربوط به سایر شهرستان‌ها قرار گرفته‌اند. این امر نشان از اختلاط گاومیش‌های اهواز دارد. یکی از دلایل احتمالی آن می‌تواند بخاطر مرکزیت اهواز در استان و تبادلات زیاد دام بین گاومیش‌داران اهوازی با گاومیش‌داران سایر نقاط استان باشد. دلیل دیگر می‌تواند بخاطر عدم تمایز یا انشقاق کافی افراد در شهرستان اهواز باشد. ولی گاومیش‌های دزفول، شوشتر و دشت‌آزادگان که به ترتیب با شماره‌های ۲، ۵ و۶ مشخص شده‌اند نسبت به اهواز اختلاط کمتری داشته، یا توانسته اند مراحل بیشتری از تمایز را پشت سر بگذارند. ضمناً گله کرمانشاهی به دلیل خرید گاومیش‌های خود از استان خوزستان و قرابت ژنتیکی آن نمونه‌ها با استان در بین شهرستان‌های دشت‌آزادگان، اهواز و دزفول قرار گرفته‌اند.

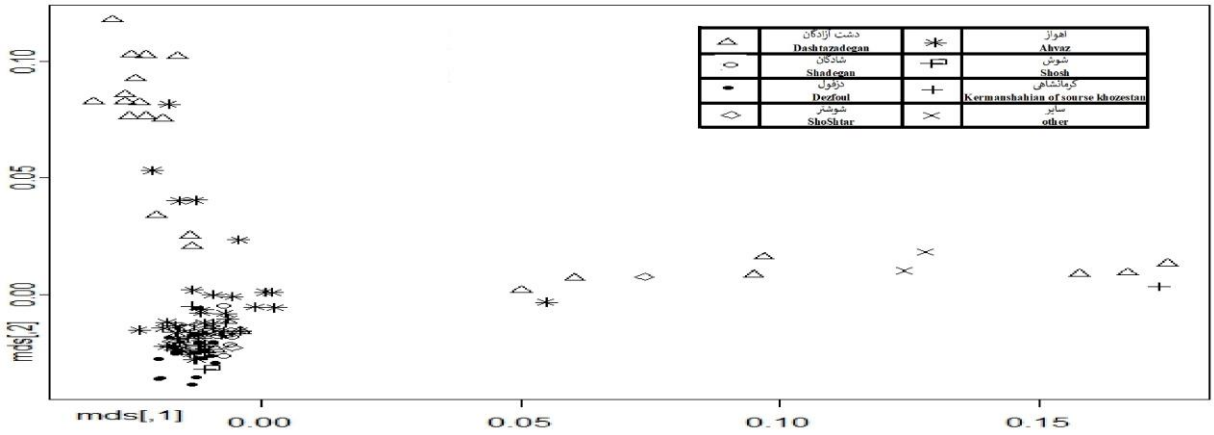
[1] Plotnj



شکل ۱- نمودار درخت همسایگی (Plotnj) بین زیرجمعیت‌های گاومیش خوزستانی

Figure 1-Unrooted neighbor-joining tree (Plotnj) between Khuzestani buffalo sub-populations

برای تایید ارتباط واریانس ژنتیکی جمعیت‌ها با پراکندگی جغرافیایی گاومیش‌ها روش تجزیه مولفه اصلی بکار رفت که نتیجه آن در شکل ۲ آورده شده است. روش PCA، مولفه‌های اصلی که ساختار جمعیت را براساس همبستگی ژنتیکی میان افراد بیان می‌کند، را شناسایی می‌کند. با ارزیابی



شکل ۲- نمودار آنالیز PCA بین زیرجمعیت‌های گاومیش خوزستانی

Figure 2- PCA plot analysis between Khuzestani buffalo sub-populations

اختلاف ژنتیکی میان جمعیت‌ها (شهرستان‌های استان خوزستان) شکل PCA (شکل ۲) ترسیم شد که نشان دهنده تمایز نسبی جمعیت شهرستان‌ها از همدیگر است. نتایج این بخش مشابه نتایج حاصل از نمودار شعاعی بود و همانطور که در این شکل مشاهده می‌شود جمعیت دشت آزادگان و شادگان از بقیه شهرستان‌های

استان جدا شده‌اند. آنالیز مولفه‌های اصلی در شکل ۲ نشان می‌دهد که علاوه بر تنوع فردی بین گاومیش‌های خوزستان تنوع بین جمعیتی نیز بین آنها مشاهده می‌شود. با توجه به پراکندگی مثلث‌های کوچک در طول محور X و Y در شکل۲، گاومیش‌های دشت‌آزادگان نسبت به سایر شهرستان‌ها دارای بیشترین تنوع از نظر ژنتیکی می‌باشند. گاومیش‌های شهرستان اهواز که با ستاره هشت‌پر نشان داده شده‌اند دارای تنوع مناسبی می‌باشند در حالیکه جمعیت گاومیش‌های دزفول، شوشتر و شوش که به ترتیب با دایره توپر، لوزی و مستطیل مشخص شده‌اند، قرابت ژنتیکی بیشتری دارند. در خصوص میزان اعتبار نتایج بدست آمده برای جمعیت‌های گاومیش در شهرستان‌ها، بررسی اعتبار روش تجزیه مولفه اصلی برای داده‌های ژنومیکی بعنوان سندی برای اعتبار نتایج مطرح می‌شود. مشخصه اصلی روش PCA، توانایی آن برای شناسایی ساختارهای ژنتیکی در مجموعه داده‌های بزرگ در زمان محاسباتی ناچیز و بدون هیچ فرضی درباره زمینه مدل ژنتیکی جمعیت است (جومبرت ۲۰۰۸). در مطالعه‌ای برای بررسی ساختار جمعیتی با روش‌های مبتنی بر مدل و روش اکتشافی DAPC، این روش‌ها تداوم بالایی در تخمین ساختار جمعیت و استنباط احتمالات عضویت افراد به هر گروه نشان دادند (پومیتی و همکاران ۲۰۱٤). روش PCA به عنوان جایگزین برای روش‌های خوشه‌بندی بیزی پیشنهاد شده است (لی و همکاران ۲۰۰۹ و لیو و ژاوو ۲۰۰٦). با این حال PCA فاقد برخی از ویژگی‌های ضروری برای بررسی ساختار جمعیت بیولوژیکی از جمله ناتوانی در ارزیابی گروهی و نیازمند تعریف پیش‌فرض خوشه‌ها برای مطالعه ساختار جمعیت است. لذا برای بررسی زیر جمعیت‌ها جدا از تاثیر مکان جغرافیایی از روش آنالیز تفکیک کننده مولفه‌های اصلی (DAPC)[1] استفاده می‌شود. در اینجا محقق نظری در خصوص تعداد زیرجمعیت‌ها نمی‌دهد بلکه این روش آماری است که تعداد زیر جمعیت‌ها را پیشنهاد می‌دهد. اغلب برای پی بردن به ساختار جمعیت با تعیین تعداد خوشه‌ها (گروه) بدون دانش قبلی اقدام می‌شود. معمولا چند روش را می‌توان استفاده کرد. بعنوان مثال برای پی‌بردن به تعداد زیر جمعیت‌ها، می‌توان روش خوشه بندی K-means، خوشه‌بندی بیزی با استفاده از ساختار و روش چند متغیره مانند آنالیز تفکیک کننده مؤلفه‌های اصلی را اشاره نمود. رویکرد روش‌های ساختاری بر این فرض استوارند که نشانگرها با هم مرتبط نبوده و جمعیت‌ها دارای آمیزش تصادفی هستند (گروارد و گروس ۲۰۱۱، جومبرت و همکاران ۲۰۱۰، ریچارد و همکاران ۲۰۰۰). لذا استفاده از روش مستقل از مدل همانند روش خوشه‌بندی K-means میانگین بر اساس فاصله ژنتیکی و یا DAPC روش راحت‌تری برای جمعیت‌های متمرکز محسوب می‌شوند. همچنین در تصحیح لایه‌بندی جمعیتی، روش DAPC به علت اینکه واریانس بین گروهی را افزایش و واریانس داخل گروهی را کاهش می‌دهد بهتر از روش PCA است (جومبرت و همکاران ۲۰۱۰). نرخ انتساب صحیح یا صحت روش DAPC در انتساب افراد به گروه‌ها از ۸۰ درصد تا ۹۷ درصد بسته به تعداد تکرار متفاوت بود است (جومبرت و همکاران ۲۰۱۰). با توجه به این موضوع، نمودار DAPC برای گاومیش‌های خوزستانی تهیه، و نتایج آن در شکل ۳ نشان داده شد. در شکل ۳، اعداد درون هر بیضی نشان دهنده شماره زیر گروه‌های تفکیک شده بر اساس اطلاعات ژنومیکی در گاومیش‌ها است. در آنالیز، سعی بر این است که واریانس درون هر کلاستر حداقل شود و واریانس بین کلاسترها بر اساس توابع تفکیک کننده[2] که عملا یک سری متغییرهای ساختگی هستند، حداکثر شود. نمودار ستونی ویژه اعداد[3] نشان دهنده تعداد توابع تفکیک

[1] Discriminant Analysis of Principal Components

[2] discriminant functions

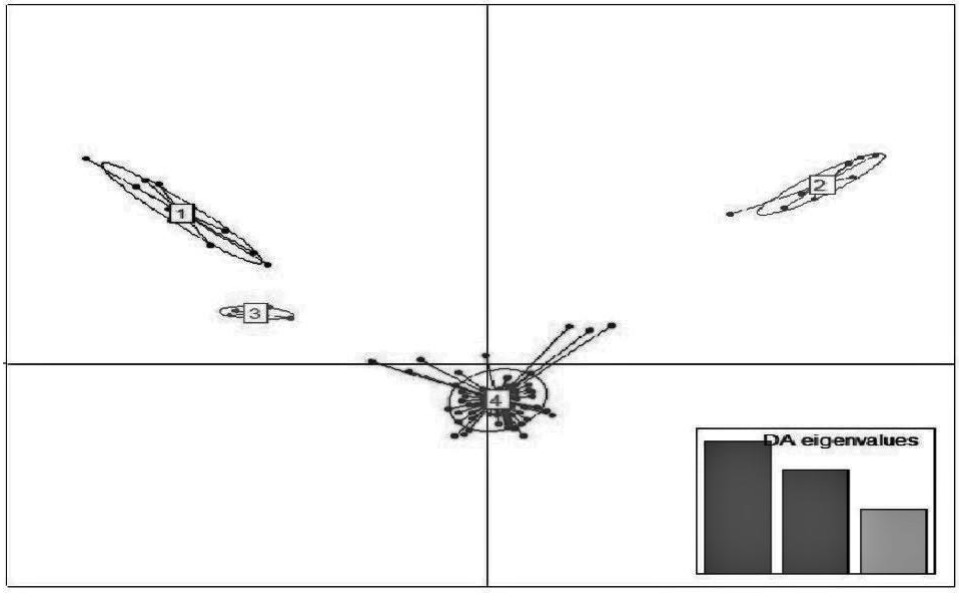
[3] eigenvalues

کننده ای که باید نگه داشته شوند، بدون اینکه مقدار زیادی از اطلاعات از دست رود. برای تعداد کم کلاسترها، تمامی ویژه مقدارها را می‌توان نگه داشت، زیرا تمامی توابع تفکیک کننده را بدون هیچ مشکلی می-توان آزمود. با بالا رفتن تعداد کلاسترها، سعی بر این خواهد بود تا ابعاد کمتری که بیشترین اطلاعات را دارند نگه داشته شوند. در شکل ۳ رنگ تیره تر بیانگر اهمیت آن ویژه مقدار است لذا می‌توان عمده واریانس را با دو تا از توابع تفکیک کننده توجیه نمود.

در این مطالعه روش DAPC در تعیین تعداد بهینه K بهتر از روش PCA عمل کرد و تصویر بهتری از ارتباط بین افراد نسبت به PCA ارائه داد. همچنین در انتساب افراد به گروه‌های خودشان صحت بسیار خوبی ارائه داد. با توجه به این مزیت‌ها و توانایی انتساب افراد، این روش برای بررسی ساختار جمعیتی و تعداد خوشه و برای کنترل کیفیت و تصحیح لایه‌بندی جمعیتی در مطالعات GWAS گزینه بهتری می‌باشد. شکل ۳ نشان می‌دهد که این گاومیش‌ها در حداقل چهار گروه مختلف خوشه‌بندی می‌شوند، که با هم فاصله مناسبی از نظر تنوع ژنتیکی دارند. اگرچه تجمع شماره ۳ بین دو زیرجمعیت ۱ و ٤ قرار گرفته و تعداد کمی گاومیش را در خود جای داده است. این تجمع می‌تواند با دو زیر جمعیت مذکور رابطه نزدیکتری داشته باشد. حتی قرابت آن به زیر جمعیت ۱ بیشتر است. با ادغام یا صرف‌نظر کردن از آن می‌توان از وجود حداقل ۳ زیر جمعیت نسبتاً مجزا در گاومیش خوزستانی خبر داد. در نظر نگرفتن این زیر جمعیت‌ها در مطالعات آتی GWAS سبب یافتن ارتباط غیرواقعی[1] می‌شود.

در تحقیقی که با عنوان اندازه گیری فواصل ژنتیکی مختلف برای تشکیل شبکه فیلوژنتیکی گاومیش‌های ایرانی با استفاده از نشانگرهای متراکم SNP انجام گرفت، با استفاده از دو سناریوی بررسی استانی و همچنین بررسی بر اساس توده‌های شناخته شده اصلی جهت تشکیل شبکه فیلوژنتیکی با روشهای مختلف اندازه‌گیری فواصل ژنتیکی و استفاده از روش پیوند همجواری بهره‌برده شد. نتایج نشان دهنده قرابت بین گاومیش‌های استان‌های آذربایجان شرقی و غربی و گیلان باهم و همچنین نزدیکی گاومیش‌های خوزستانی و کرمانشاه با یکدیگر بود (رحمانی نیا و همکاران ۲۰۱۵).

[1] False positive



شکل ۳- نمودار پراکنش DAPC برای زیرجمعیت‌های گاومیش خوزستانی

Figure 3- DAPC Scatter plot for Khuzestani buffalo sub-populations

نتیجه‌گیری کلی

یافته کلی این پژوهش حکایت از وجود تنوع زیاد درون و بین زیر جمعیت‌های گاومیش خوزستانی دارد. ولی همانگونه که انتظار می‌رفت میزان تنوع بین جمعیتی و فواصل ژنتیکی کمتر از این پارامترهای محاسبه شده توسط سایر محققین در جمعیت‌های بین استانی بود. نتایج آنالیز چند متغییره نشان داد که حداقل ۳ شبه گروه ژنتیکی نسبتاً متمایز در جمعیت گاومیش‌های خوزستان قابل تصور است. لذا باید یافته این تحقیق در سایر تحقیقات با مضمون انتخاب ژنومیکی، بررسی گسترده پیوستگی ژنوم-صفت و تشکیل جمعیت پایه در نظر گرفته شود تا اعتبار نتایج انتخاب ژنومیکی افزایش یابد.

سپاسگزاری

از مرکز اصلاح نژاد دام کشور و معاونت محترم بهبود تولیدات دامی استان خوزستان به دلیل همکاری صمیمانه و در اختیار قرار دادن اطلاعات لازم و دانشگاه علوم کشاورزی و منابع طبیعی خوزستان به دلیل حمایت‌های مالی و معنوی که در این مطالعه داشته اند و آقای دکتر محمد باقر زندی بخاطر توصیه‌های علمی، کمال تشکر را داریم.

منابع مورد استفاده

Abasi-mashee B, Fayazi J, Roshanfekr H, Nasiri-Beigi MT and Mirzade K, 2009. Molecular marker assisted study of kappa casein gene in buffalo population in khouzestan province. National Symposium of Buffalo in Iran. 2:16-21 (In Persian).

Akey J, 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome research 19: 711-722.

Arif I, Khan Haseeb A, Bahkali Ali H, Al Homaidan Ali A, Al Farhan Ahmad H, Al Sadoon M and Shobrak M, 2011. DNA marker technology for wildlife conservation. Saudi journal of biological sciences 18: 219-225.

Azizi Z, Moradi Shahrbabak H, Moradi Shahrbabak M, Rafat A and Shodja J, 2016a. Genetic classification of Azari and North ecotype Buffalo population using SVM method. Iranian Journal of Animal Science. 2:279-291 (In Persian).

Azizi Z, Rafat A, Shoja J, Moradi Shahrbabak H and Moradi Shahrbabak M, 2016b. Study of population structure and stratification two ecotypes buffalo with dense single nucleotide polymorphism markers using Admixture, MDS, PCA and GC methods. Journal of Agricultural Biotechnology. 8(2):2-8 (In Persian).

Bromad-jezi M, 2005. Buffalo Breeding .The Institute of Applied Agriculture. (In Persian).

Chen WC and Dorman K, 2010. phyclust: phylogenetic clustering (Phyloclustering) R package, http://cran.r-project.org/package=phyclust.

FAO. 2015. The second report on the state of the world's animal genetic resources for food and agriculture. Rome.

Goddard Michael E and Hayes Ben J, 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391.

Grünwald Niklaus J and Goss Erica M, 2011. Evolution and population genetics of exotic and re-emerging pathogens: novel tools and approaches. Annual review of phytopathology 49: 249-267.

Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, et al. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). PLoS ONE 12(10): e0185220.

Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405.

Jombart T, Devillard S and Balloux F, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics 1: 94.

Karimi K, Esmailizadeh KA, and Asadi Fuzi M, 2015. Analysis of genetic structure of Iranian indigenous cattle populations using dense single nucleotide polymorphism markers. Animal Production Research 3: 93-104.

Karimi K, Esmailizadeh KA, and Asadi Fuzi M, 2017. Linkage disequilibrium levels in Fars province native cattle population using high-density SNP data. Animal Science Researches 27(1): 17-27.

Lee C, Abdool A and Huang CH, 2009. PCA-based population structure inference with generic clustering algorithms. BMC bioinformatics 10: 1471-2105.

Liu N and Zhao H, 2006. A non-parametric approach to population structure inference using multilocus genotypes. Human genomics 6: 353-364.

Medugorac I, Medugorac A, Russ I, Veit-Kensch C E, Taberlet P, Luntz B, Mix Henry M and Foerster M, 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. Molecular ecology 16: 3394-3410.

Mokhber M, Moradi-Shahrbabak M, Sadegh M, Moradi-Shahrbabak H and Williams J, 2015. Genome-wide survey of signature of positive selection in Khuzestani and Mazandrani buffalo breeds. Iranian Journal of Animal Science. 46:119-31 (In Persian).

Mokhber M, 2015. A genome-wide scan for Selective signatures in Iranian buffalo breeds. PhD thesis, College of Agriculture and Natural Resources, Tehran University.

Paradis E, Claude J and Strimmer K, 2004. APE: analyses of phylogenetics and eEvolution in R language. Bioinformatics 20: 289-290.

Pometti CL, Bessega CF, Saidman BO and Vilardi JC, 2014. Analysis of genetic population structure in Acacia caven (Leguminosae, Mimosoideae), comparing one exploratory and two Bayesian-model-based methods. Genetics and molecular biology 37: 64-72.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC, 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81: 945-959.

Rahmaninia J, Miraei-Ashtiani SR and Moradi Shahrbabak H, 2015. Unsupervised clustering analysis of population and subpopulation structure using dense SNP markers. Iranian Journal of Animal Science 46(3):277-287 (In Persian).

Rozgar S, Nasiri-Beigi MT, Roshanfekr H, Fayazi J, Mirzade K and Sadr-sadat A, 2009. The study of PIT1 Gene Polymorphism in buffalo population in khouzestan province using PCR-RFLP Method. National Symposium of Buffalo in Iran. 2:78-81 (In Persian).

Shojaee Y, Fayazi J, Roshanfekr H and Mirzade K, 2009. Investigation of prolactin polymorphism in buffalo population of Khuzestan province by PCR-RFLP. National Symposium of Buffalo in Iran. 2:82-85 (In Persian).

Simianer H, Ma Y and Qanbari S, 2014. Statistical problems in livestock population genomics. In: Proceedings, 10th World Congress of genetics applied to livestock production. Vancouver; https://asas.org/docs/default-source/wcgalp-proceedings-oral/202_paper_10373_manuscript_1346_0.pdf?sfvrsn=2.

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

Utsunomiya YT, Perez OBrien AM, Sonstegard TS, Van Tassell CP and do Carmo AS, 2013. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome wide scan methods. PLoS ONE 8(5), e64280.

doi:10.1371/journal.pone.0064280.

Voight BF, Kudaravalli S, Wen X and Pritchard JK, 2006. A map of recent positive selection in the human genome. PLoS Biology 4(3), e72.

Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Williams JL, Iamartino D, Pruitt KD, Sonstegard T, Smith TPL, Low WY, Biagini T, Bomba L, et al., 2017. Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2n = 50). GigaScience 6(10): 1-6.

Genomic study of population structure and phylogenetic relationship of Khuzestani buffaloes

M R Zargar[1], J Fayazi*[2], M T Bigi Nassiri[3] and H Moradi shahr babak[4]

Received: 14 October 2017 Accepted: 21 February 2018
[1, 2 and 3] MSc Student, Associate Professor and Professor, respectively, Department of Animal Science, Khuzestan Agricultural Sciences and Natural Resources University, Iran
[4]Assistant Professor, Department of Animal Science, Faculty of Agricultural Sciences, University of Tehran, Iran
*Corresponding author: j_fayazi@ramin.ac.ir

Introduction: The science of phylogenetics, and specially the subfield of molecular systematics, has grown exponentially by amount of publications, general interest and amount of available genetic data. Modern phylogenomic studies use large genomic and transcriptomic resources. however a comprehensive molecular phylogeny of animals, including the newest types of data for all phyla, remains elusive.
Genome sequences and SNP chips are now accessible for many of species across the animal phylogeny, conveying key features of animal genome evolution into sharper emphasis. The field of animal evolutionary genomics has focused on recognizing the diversity genomic features, rebuilding the history of evolutionary variations in animal genomes and testing hypotheses about the evolutionary relationships of animals. One of the most important buffaloes in Iran are Khuzestani buffaloes, which have a wide distribution. Given the fact that they may be divided into several ecotypes and sub-populations, these may cause problems in genomic wide association study. One of the challenges that always threatens the accuracy of genomic association and genetic variation studies is the existence of sub-populations within the populations. The probable relationship between genetic groups among Khuzestani buffaloes by whole genome markers has not been studied. This study is an effort to understand and provide a detailed insight into the population structure and genomic phylogeny of Khuzestani buffaloes.
Material and methods: Blood or hair samples from 121 Khuzestani buffaloes from different regions of Khuzestan province including Ahvaz (n=28), Dezfoul (n=19), Shadegan (n=6), Shosh (n=4), Shushtar (n=9), Dashtazadegan (n=24) and Kermanshah (n=8) samples were collected. The samples were chosen from herds that were registered by Iranian Animal Breeding Center (IABC) and were under milk recording system. DNA of the samples were isolated and sent to Parco Tecnologico Padano in Itally. All samples were genotyped by Axiom@Buffalo 90K bead array of Affymetrix Company. PLINK software was used for pre-processing of raw genotype data based on minor allele frequency (MAF) ≤0.01, Hardy Weinberg disequilibrium (p-value ≤ 10-6) and call rate ≤ 0.05. Thereafter, a total of 64,709 SNPs remaind for the analysis. The genetic distance for each pair of individuals was calculated using R software (ape package). The Neighbor-joining algorithm was used to plot trees based on those distances. Also, the phylogenetic diagram is derived based on the distance matrix and using R software (Phyclust package). To identify the optimal number of clusters, K-means were sequentially increased with K values, and different clustering responses were compared using Bayesian Information Criteria (BIC). FST values (theta) for autosomal chromosomes and counties were calculated by unbiased method. Then, the average genetic distance of each area was calculated with other sampling areas. Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC) were performed for genetic clustering of individuals. Adegenet R package was used for PCA and DAPC analysis.
Results and discussion: In our study Unrooted Neighbor-Joining Tree Plot grouped the individuals mainly in one cluster depending on the origin of group or subpopulation. Dezfoul, Shushtar and Dashtazadegan individuals formed their individual clusters and due to genetic relatedness, Ahvaz

and Kermanshah clusters were formed almost overlapping with other. Along with the closeness of Dezfoul among themselves, they also showed closeness to Shosh. Kermanshah group lied between Ahvaz, Dashtazadegan and Dezfoul buffaloes.
Results obtained from DAPC analysis confirmed with those obtained from PCA analysis. Individuals were correctly assigned to their respective clusters. Based on phylogenic result, it is evident that buffaloes of Ahwaz are affected by buffaloes of other province's regions, because they are in the center. Buffaloes of Shushtar, Dezful and Dasht-E-azadegan have lower emigration rate in comparison with Ahwaz buffaloes. This result also confirmed by principal component analysis. To survey the genomic structure and patterns of Khuzestani buffaloe populations, the Fst values (unbiased Theta) for every SNPs were calculated. This criterion reflects the distribution of diversity between groups and within the group. Its high level indicates that major part of heterozygosity is due to differences in the groups. In other hand, the low level of it, suggests that the diversity is due to the individuals within the group, and the populations or groups are associated. The highest value of Fst was obtained for buffaloes of Dashte-E-azadegan (0.0159) and Shadegan (0.0147) because of its special geographical condition, and the lowest value of Fst was observed for the buffaloes of Ahwaz (0.0087) and Shosh (0.0086). The results of statistical tests of population genetic showed a high genetic diversity in buffaloes of Khuzestan. Graphical view of clustering and assignment test suggested at least three subpopulations for Khuzestani buffaloes. According to the overall results and existence of great SNP diversity, it is possible to move toward construction of base population based on this results.
Conclusion: In this study a genome-wide insight into the genomic phylogeny of Khuzestani buffaloes was provided. The results of nucleotide genetic diversity and genetic distance between buffaloes are worthy of attention in subsequent studies such as genomic wide association study, genomic selection and formation of Base populations of Khuzestani buffalo.

Keywords: Genomic wide study, Khuzestani buffalo, Principal component analysis, SNP diversity