



## "مقاله پژوهشی"

# تجزیه داده‌های ژنومی حاصل از 70 SNP chip جهت شناسایی جایگاه‌های مرتبط با تمایز دو نژاد اسب کرد و کاسپین با استفاده از جاروب انتخاب

محدثه نصیرپور<sup>۱</sup>، محمد مرادی شهر بابک<sup>۲</sup>، حسین مرادی شهر بابک<sup>۳</sup>، حسن مهربانی یگانه<sup>۴</sup> و یونس دوستی<sup>۵</sup>

۱- دانش‌آموخته کارشناسی ارشد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

۲- استاد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

۳- استادیار، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

۴- استاد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

۵- دانشجوی کارشناسی ارشد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران، (نویسنده مسوول: y.devisty@ut.ac.ir)

تاریخ دریافت: ۱۴۰۱/۹/۱۳ تاریخ پذیرش: ۱۴۰۱/۱۱/۵

صفحه: ۱۵۴ تا ۱۶۲

### چکیده مبسوط

**مقدمه و هدف:** انتخاب طبیعی و مصنوعی در جهت افزایش فراوانی آلل‌های مطلوب در جمعیت‌های حیوانی، سبب برجای ماندن برخی نشانه‌ها در ژنوم حیوانات می‌شود که معمولاً با صفات مهمی در ارتباط هستند. امروزه با پیشرفت توالی‌یابی نسل جدید و دسترسی آسان‌تر به اطلاعات ژنومی حیوانات، مدل‌هایی برای شناسایی نشانه‌های انتخاب بر پایه فراوانی آللی و طول هاپلوئیدی ارائه شده است. در این پژوهش نشانه‌های انتخاب متمایزکننده دو جمعیت اسب نژاد کرد و کاسپین با استفاده از تراشه 70k مورد بررسی قرار گرفت.

**مواد و روش‌ها:** در این تحقیق از ۳۵ راس اسب نژاد کاسپین و ۳۱ راس اسب نژاد کردی نمونه خون جمع‌آوری شد. پس از استخراج DNA و تعیین توالی (توسط شرکت ایلومینا) کنترل کیفی داده‌ها برای فراوانی آلل حداقل  $(P_{MAF} < 0.01)$  و نرخ تعیین ژنوتیپ  $(P_{GENO} < 0.05)$  و انحراف از تعادل هاردی-واینبرگ  $(P_{H-W} < 1 \times 10^{-6})$  انجام شد. در ادامه کنترل کیفی با استفاده از آزمون تناسلی (0) انجام گرفت و با استفاده از سایت Ensemble به شناسایی جایگاه‌های SNPها پرداخته و در نهایت ژن‌های مرتبط با این جایگاه‌ها مشخص شدند.

**یافته‌ها:** پس از انجام کنترل کیفیت داده‌های ژنومی و ادغام اطلاعات دو جمعیت بر اساس بسته نرم‌افزاری R 6.1 اسب با 52650 SNP برای ادامه آنالیزها باقی ماندند. در این پژوهش با استفاده از آزمون‌های آماری Fst به روش برادرگر ناریب تناسلی در نهایت تعداد 31 نشانگر متمایزکننده دو نژاد اسب مورد مطالعه شناسایی شدند.

**نتیجه‌گیری:** نتایج نشان داد که دو نژاد اسب کاسپین و اسب کرد در دو دسته جداگانه قرار دارند. نژاد اسب کردی دارای تنوع و پراکندگی بیشتری نسبت به اسب کاسپین بود. از مهم‌ترین ژن‌های شناسایی شده مرتبط با جایگاه‌های معنی‌دار، ژن‌های *GM6B*, *ITGB6*, *DOCK3*, *R3HCC1*, *HPSE2*, *INPP5F*، *THOC2*, *GRIA3*, *LRCH2*, *WDR13*, *PHOX* بودند. بیشتر ژن‌های شناسایی شده در این پژوهش با تبادلات بین سلولی، انقباضات ماهیچه‌ای، ایمنی، پشتیبانی غشا پایه، پایداری DNA، سیستم عصبی در ارتباط بودند.

**واژه‌های کلیدی:** اسب کاسپین، اسب کردی، نشانه‌های انتخاب، Fst، PCA

### مقدمه

بالا نگه می‌دارد)، یال پرپشت، دم در موقع راه رفتن به صورت فواره‌ای و قوس‌دار پایین آمده، کپل پر حجم و قوی، قسمت جلویی سینه عریض و عضلانی و شکم به صورت استوانه‌ای مدور در امتداد سینه قرار دارد. اسب کردی از نظر رنگ متنوع است و به رنگ‌های سیاه، سفید، گریگ و کهر دیده می‌شود (۲۶). قدمت اسب کاسپین به بیش از سه هزار سال می‌رسد. تصویر این اسب‌ها در نگاره‌های پلکان آپادانای تخت جمشید موجود است (۲۵). خصوصیات رفتاری این نژاد حاکی از هوش سرشار آن بوده و مهم‌ترین ویژگی آن فراگیری سریع و آسان رفتارهای آموزشی است. تلفیق این هوش با خلق‌و‌خوی مهربان در این نژاد، آن را جزو بهترین نژادها برای سوارکاری و آموزش سوارکاری کودکان در جهان قرار داده است. آموزش و تربیت این نژاد در باشگاه‌های سوارکاری نشان داد که از نظر پرش از استعداد خاصی برخوردار بوده و در بسیاری موارد قابل رقابت با اسب‌های بزرگ جثه می‌باشند (۱). اسب کاسپین با وجود قدی کوتاه یعنی ۹۰ تا ۱۲۰ سانتی‌متر به یک اسب با تناسب بدنی کامل شباهت دارد و دارای پرش زیاد و رشد سریع‌تری از اسب معمولی است (۲۷). قد اسب‌های کاسپین ۱۰۰ تا ۱۳۰ سانتی‌متر گزارش شده است البته نمونه‌های کمتر از ۱۰۰ سانتی‌متر هم به‌ندرت دیده می‌شود. عوامل متعددی بر

اسب حیوانی است که در سیر تکاملی و تاریخی انسان را همراهی کرده و در طول تاریخ بسیاری از پیشرفت‌ها و موفقیت‌های بشر در امور نظامی، کشاورزی، حمل‌ونقل و ورزشی تحت تأثیر اسب قرار گرفته است. سرزمین پهناور ایران مهد پرورش اسب و مبدأ نژادهای گوناگون است. در ایران ثروتی بی‌همتا با سابقه‌ای به طول تاریخ تمدن جهان، اسب بومی وجود دارد که ویژگی‌های منحصر به فرد داشته و به گواهی تاریخ جد تمام اسب‌های اصیل و ممتاز امروزی است (۲۲). در تاریخ آمده است که از ۵۰۰ سال قبل از میلاد مسیح (ع) در دامنه‌های سلسله جبال زاگرس در منطقه‌ای به نام نسا، اسب کردی وجود داشته است. با توجه به خصوصیات ظاهری و توانایی‌ها، این اسب به تیره‌های متفاوتی مانند جاف، افشاری، سنجایی و غیره تقسیم می‌شود. این نژاد اسب، اسبی قدرتمند، عضلانی و با استقامت بوده و دارای سر نسبتاً بزرگ است. توانایی این نژاد برای استفاده در مسافت‌های طولانی و مسیرهای کوهستانی و همچنین شرکت در مسابقات چوگان بی‌نظیر است (۸). اسب کردی دارای سم‌های سخت و محکم، چشم‌های درشت و گرد با حدقه برجسته، اتصال سر و گردن قوی، گردن قوی و حجیم (معمولاً سر را

می‌شود. در صورتی که جهش به وجود آمده در سطح ژنوم باعث افزایش شایستگی افراد در جامعه شود، انتخاب طبیعی باعث می‌شود تا افرادی که دارای شایستگی بیشتری هستند، در تشکیل نسل بعد مشارکت بیشتری داشته باشند (۲). به دنبال افزایش فراوانی آلل‌های مفید، فراوانی آلل‌های موجود در جایگاه‌های خنثی یا نسبتاً خنثی که به این جایگاه‌ها متصل هستند، نیز افزایش می‌یابد. به این پدیده انتقال همراه می‌گویند. به کاهش یا حذف تنوع در جایگاه‌های مجاور یک جهش (آلل جهش‌یافته) سودمند در طی انتخاب طبیعی، جاروب انتخاب<sup>۲</sup> گفته می‌شود. چنانچه یک جهش جدید باعث افزایش شایستگی افراد حامل آن نسبت به سایر افراد جامعه گردد، انتخاب باعث می‌شود افرادی که دارای شایستگی بیشتری هستند در تشکیل نسل بعد مشارکت بیشتری داشته باشند. بدین ترتیب فراوانی واریانت جهش یافته بسته به سهم آن در افزایش شایستگی به سرعت شروع به افزایش خواهد کرد. به عبارتی، هنگامی که انتخاب مثبت به سمت افزایش فراوانی آلل مطلوب پیش می‌رود، جایگاه‌های نزدیک به این جایگاه اگر هم اثرشان خنثی باشد به علت پیوسته بودن با آلل مطلوب فراوانی‌شان زیاد شده که در نتیجه این پدیده، الگوی تنوع ژنتیکی و عدم تعادل لینکاژی در جایگاه‌های اطراف این جهش انتخابی تغییر خواهد کرد (۲). تاکنون چندین روش برای شناسایی نشان‌های انتخاب با استفاده از داده‌های مقایسه ژنومی گزارش گردیده است (۲۴). این روش‌ها که بر اساس عدم تعادل پیوستگی ژنی عمل می‌کنند شامل آزمون هاپلو تیپ با طول بلند (LRH)، هموزیگوسیتی هاپلو تیپ گسترش‌یافته (EHH)، انتگرال گیری هموزیگوسیتی هاپلو تیپی (iHS) می‌باشند و همچنین بررسی آماره FST که مبتنی بر فراوانی آلل می‌باشد.

استفاده از این روش‌ها برای شناسایی نشان‌های انتخاب می‌تواند ما را در تشخیص صفات تحت انتخاب، ژن‌های مؤثر بر صفات کمی و کیفی و یا جمعیتی که QTL در آن در حال تفرق است، یاری دهند. به طور کلی، مناطقی از ژنوم که تحت انتخاب هستند بایستی اهمیت شایستگی و عملکردی داشته باشند، در غیر این صورت انتخاب مثبت آن‌ها را تحت تأثیر قرار ندهد (۲۴).

هدف از این پژوهش، شناسایی مناطقی ژنومی هدف انتخاب در طی سال‌های طولانی در دو نژاد اسب ایرانی کاسپین و اسب کرد و همچنین مکان‌یابی این مناطق، شناسایی ژن‌ها و QTL‌های موجود در این مناطق، به عنوان، مناطقی متمایز کننده و یا صفات متمایز کننده در این دو نژاد با استفاده از آرایه‌های ژنومی تک نوکلئوتیدی ۷۰k بود.

### مواد و روش‌ها

جمع‌آوری نمونه‌های خون برای استخراج DNA قبل از نمونه‌گیری از مالکان در مورد هر اسب اطلاعاتی در مورد سن و سال و نژاد اسب و امکان دورگه بودن آن‌ها جمع‌آوری شد. در صورت وجود ناخالصی طبق اطلاعات مالکان، از نمونه‌گیری خودداری به عمل آمد. همچنین برای توصیف هرچه بهتر جامعه ژنوتیپی هر نژاد، از خون‌گیری اسب‌هایی

شاخص قد اسب‌های کاسپین مؤثر هستند که وراثت مهم‌ترین آن‌ها است. با تحقیقات انجام‌شده ثابت شد که اسب کاسپین جد تمامی اسبان موجود در دنیا بوده و حتی اسب‌های عرب که از قدیمی‌ترین اسبان دنیا می‌باشند از اسبچه خزر به وجود آمده‌اند. لذا به آن اسب مینیاتوری دریای کاسپین اطلاق می‌شود. این اسب دارای سم بسیار قوی، کوچک و بیضی‌شکل است، به طوری که حتی در مواقعی که روی خشن‌ترین زمین کار می‌کنند، نیازی به نعل‌بندی ندارند (۱۰).

از اسب‌ها بیشتر جهت حمل‌ونقل و کار استفاده می‌شود. اسب‌های عرب جمعیت زیادی از اسب‌های کشور را شامل می‌شوند و در سراسر کشور پراکنده هستند در حالی که محل پراکنش اسب‌های کاسپین و تالشی در حاشیه‌های دریای خزر است و جمعیت کمی دارند و در معرض خطر انقراض هستند. لازم است توجه ویژه‌ای برای نگهداری و حفظ تمام نژادهای اسب ایران از لحاظ اصلاح نژادی و مخزن ژنی مبذول گردد (۱۶). کاهش در اندازه جمعیت برخی گونه‌های بومی منجر به نگرانی‌هایی در مورد حفظ نژادهای کمیاب در بسیاری از نواحی جهان و بخصوص ایران شده است (۸).

شرح کامل تفاوت ژنتیکی بین دو نژاد امکان‌پذیر نیست ولی فاصله ژنتیکی بهترین معیار برای بیان تفاوت ژنتیکی می‌باشد. برای شناسایی فاصله ژنتیکی روش‌های متعددی پیشنهاد شده است که مبتنی بر فراوانی آللی است (۱۸). زمانی که یک آلل سودمند در طی زمان‌های مختلف، هدف انتخاب مثبت قرار می‌گیرد باعث ایجاد نشان‌هایی در سطح ژنوم می‌گردد که از طریق بررسی طیف فراوانی آللی و محاسبه عدم تعادل پیوستگی<sup>۱</sup> (LD) قابل‌شناسایی هستند (۲۴). استفاده از این روش، که به انتقال همراه معروف است در طی چند سال اخیر به یکی از کارآمدترین تکنیک‌ها در شناسایی مناطقی ژنومی مؤثر بر صفات کمی در حوزه‌های مختلف ژنتیک تبدیل شده است (۲۸). یکی از مزایای اصلی این روش امکان اجرای آن تنها با استفاده از اطلاعات مولکولی در غیاب رکوردهای فنوتیپی می‌باشد (۱۲) که این روش را در شناسایی جایگاه‌های ژنومی کاندیدا برای بسیاری از صفات از طریق مقایسه نشان‌های انتخاب در جمعیت‌هایی با صفات متمایز بخصوص تحت شرایط نداشتن شجره و رکورد برای سایر صفات، مصداق اکثر نژادهای موجود در کشور، متمایز می‌سازد. شناسایی این مناطق ژنومی، با توجه به این که عمدتاً با ژن‌های عمده اثر و صفات مهم اقتصادی همراه هستند دارای اهمیت زیادی است و می‌توانند منابع اطلاعاتی مناسبی برای بهبود راندمان، تفکیک نژادی و دیگر مؤلفه‌های مورد نیاز برای بهبود ژنتیکی دام‌ها فراهم آورند (۱۲).

آگاهی از منشأ، نژاد و تاریخچه تکامل جهت پیش‌بینی ترکیب ژنی حمل شده توسط هر نژاد ممکن است در آینده برای توصیف صفات مختلف مثل مقاومت به بیماری، استرس و سازگار شدن به محیط‌های مختلف مهم باشد (۴). سطوح بالای تمایز جمعیت می‌تواند بیانگر وقوع انتخاب مثبت روی یک آلل مفید در یک یا چندین جمعیت باشد. در مقابل، سطوح پایین‌تر نشان‌دهنده وجود انتخاب متعادل‌کننده است که منجر به حفظ نسبت خاصی از آلل‌ها در تمام جمعیت

دارای رابطه خویشاوندی که بر اساس اطلاعات مالکان بود، امتناع گردید. در این تحقیق، ۳۶ نمونه خون از اسبچه خزر در شهر گیلان و ۳۱ نمونه خون از اسب‌های نژاد کردی در شهرهای سقز، کرمانشاه، بیجار، اسدآباد و کنگاور به میزان ۵ سی‌سی گرفته شد و پس از انتقال به آزمایشگاه استخراج DNA صورت گرفت.

#### تعیین توالی نمونه‌های DNA

تمام نمونه‌ها به آزمایشگاه GeneSeek کشور آمریکا جهت تعیین ژنوتیپ انتقال داده شدند. بعد از انجام مراحل تعیین ژنوتیپ، خوانش تراشه‌ها به منظور تعیین نوع SNP‌های موجود روی تراشه‌های مربوط به هر حیوان، با استفاده از اسکنر ارائه شده توسط شرکت ایلومینا انجام شد. برای این کار از نرم‌افزار ژنوم اسکن برای اندازه‌گیری شدت سیگنال‌های قرمز و سبز رنگ به دست آمده از دوربین استفاده شد. پس از اندازه‌گیری شدت و نوع سیگنال‌های نوری مختلف و به دست آوردن نقشه فیزیکی SNP‌های موجود، تعیین ژنوتیپ انجام شد. ژنوتیپ SNP‌ها با استفاده از تراشه‌های طراحی شده برای اسب که دارای ۷۰۰۰۰ نشانگر تک نوکلئوتیدی است و با بکارگیری پروتکل استاندارد ارائه شده توسط شرکت ایلومینا انجام شد.

#### آنالیز داده‌های اولیه و کنترل کیفیت داده‌ها

کنترل کیفیت داده‌ها با نرم‌افزار Plink (نسخه ۱/۰۶) و برنامه R (نسخه ۹/۱) انجام شد. داده‌ها با فراوانی آلل حداقل بیشتر از ۰/۰۱ و نرخ تعیین ژنوتیپ بیش از ۰/۹۵ برای آنالیزهای بعدی انتخاب و حیوانات با بیش از ۰/۰۵ ژنوتیپ گم شده از آنالیز حذف شدند. انحراف از تعادل هاردی-واینبرگ در سطح  $P < 1 \times 10^{-6}$  در نظر گرفته شد و نشانگرهایی با p-value پایین‌تر از این مقدار استفاده شدند. پس از انجام مراحل کنترل کیفیت، آنالیز تجزیه مؤلفه‌های اصلی انجام شد. برای انجام این کار از بسته نرم‌افزاری geno-pca در محیط برنامه R استفاده شد. برای ترسیم پلات مربوطه نیز از برنامه R استفاده شد.

#### آنالیزهای آماری برای شناسایی نشانه‌های انتخاب

برای محاسبه Fst، به فرکانس‌های آللی در هر جمعیت مورد نیاز است. بسته به قالب ژنوتیپ‌ها، می‌توان رویکردهای مختلفی را برای تبدیل ژنوتیپ‌ها به فرکانس‌های مختلف در نظر گرفت. به‌طورکلی، هرچه اندازه نمونه برای جمعیت‌های مورد بررسی بزرگ‌تر باشد، تخمین فرکانس آلل از اطمینان بیشتری برخوردار خواهد بود.

از Fst می‌توان برای کشف تفاوت‌های ژنتیکی بین جمعیت‌های بسیار متضاد و با فاصله دور از هم استفاده نمود (۴). پیش‌بینی می‌شود که با کاهش فاصله بین جمعیت‌ها، این آماره نمی‌تواند به خوبی تفاوت‌های ژنتیکی را شناسایی نماید لذا نیاز به دقت کافی دارد. جهت بررسی نشانه‌های انتخاب ژنومی در هر دو نژاد از روش شاخص تثبیت تمایز جمعیتی Fst برای هر نشانگر SNP به روش برآوردگر ناریب تتا در محیط R و با استفاده از پکیج Fst انجام شد. این آماره برای هر SNP و برای کلیه کروموزوم‌ها محاسبه شد لازم به ذکر است، سطح آستانه معنی‌داری برای این آزمون و برای هر SNP مقدار ۰/۰۵ در نظر گرفته شد. این آماره در واقع یک معیاری است که به طور وسیعی برای مشخص کردن تمایز ژنتیکی به وجود آمده در طول زمان در ساختار جمعیتی بین جمعیت‌ها استفاده می‌شود (۵). محققین مقدار شاخص تثبیت تمایز جمعیتی کمتر از ۰/۰۵ را به‌عنوان معیار تمایز پایین و مقدار بالاتر از ۰/۶۴ را به‌عنوان یک معیار بالای تمایز جمعیتی عنوان کردند (۲۹).

آزمون مؤلفه‌های اصلی (PCA) رویکردی است برای طبقه‌بندی افراد بر اساس داده‌های ژنتیکی به گروه‌هایی بر اساس نژاد و اولاً از این روش برای بررسی و اطمینان انطباق گزارش نژادی افراد با گروه‌بندی افراد بر اساس اطلاعات ژنتیکی و ثانیاً در صورت ظاهر شدن یک فرد یا افرادی در گروه نژادی دیگر که می‌تواند به دلیل خطای نمونه‌گیری باشد. در این مطالعه پس از ادغام داده‌های هر دو نژاد این آزمون انجام شد (۱۴).

پس از شناسایی SNP‌های معنی‌دار در هر آزمون، برای شناسایی ژن‌های مرتبط با این SNP‌ها به وب‌سایت ENSEMBLE مراجعه و برای یافتن ژن‌های مرتبط با QTL‌ها و عملکرد آن‌ها به وب‌سایت‌های Medlineplus و Genecard مراجعه شد.

#### نتایج و بحث

پس از استخراج DNA و تعیین ژنوتیپ نمونه‌های ارسال شده، در ابتدا برای هر جمعیت کنترل کیفیت انجام شد. یک راس اسب از ۳۱ راس اسب نژاد کردی و چهار راس اسب از ۳۵ راس اسب نژاد کاسپین به دلیل تعداد بالای نشانگرهای گمشده از ادامه آنالیزها حذف شد. اطلاعات مرتبط با کنترل کیفیت هر نژاد در جداول (۱) نمایش داده شده است.

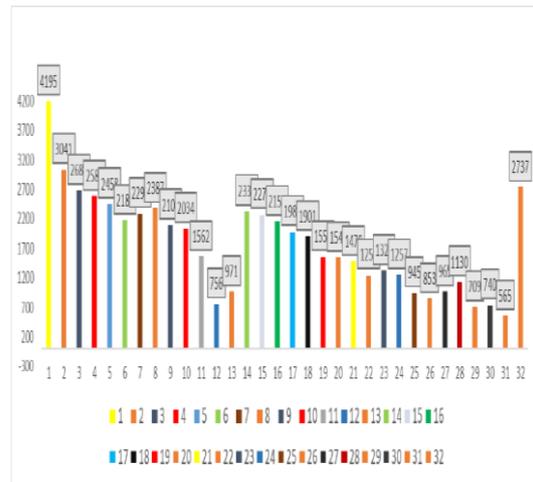
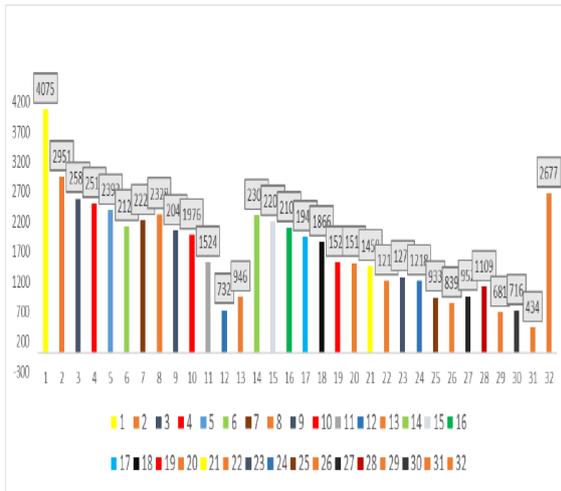
جدول ۱- کنترل کیفیت داده‌های ژنومی برای نژاد اسب کردی و کاسپین

Table1. Quality control of genomic data for Kurdish and Caspian horse breeds

ادغام اسبها (Merging horses)	اسب نژاد کاسپین (Caspian horse breed)	اسب نژاد کردی (Kurdish horse breed)	مولفه (Component)
66	35	31	تعداد کل اسب (Total number of horses)
61	31	30	تعداد کل اسب پس از کنترل کیفیت (Total number of horses after quality control)
59639	65157	65157	تعداد کل اسنپ‌ها قبل از کنترل کیفیت (Total number of snps before quality control)
0	4354	4005	تعداد نشانگرهای حذف شده با MAF کمتر از ۰/۱ (The number of markers removed with MAF less than 0/01)
6935	5411	4188	تعداد نشانگرهای حذف شده با GENO کمتر از ۰/۰۵ (The number of markers removed with GENO less than 0/05)
44	15	8	تعداد نشانگرهای حذف شده به خاطر عدم تعادل هاردی واینبرگ (Number of markers removed by Hardy-Weinberg disequilibrium)
52650	55377	56956	تعداد کل اسنپ‌ها پس از کنترل کیفیت (Total number of snps after quality control)

داده‌ها پراکنش نشانگرهای ژنتیکی روی کروموزوم‌های دو نژاد به صورت پلات ترسیم شد (شکل ۱). در اسب نژاد کردی، کروموزوم شماره یک با ۴۰۰۰ SNP دارای بالاترین تعداد نشانگر و کروموزوم‌های ۲۹، ۱۲، ۳۱، ۲۹ دارای کمترین تعداد نشانگر ژنتیکی بودند. کروموزوم ۳۲ نیز پس از کروموزوم شماره یک دارای بالاترین تعداد نشانگر ژنتیکی بود.

پس از طی مراحل کنترل کیفیت داده‌ها برای هر نژاد، دو جمعیت با یکدیگر ادغام شده و کنترل کیفیت بار دیگر انجام شد. هیچ حیوانی در کنترل کیفیت داده‌های ادغامی حذف نشد و از تعداد ۵۹۶۳۹ نشانگر در نهایت ۵۲۶۵۰ عدد برای ادامه آنالیزها باقی ماندند. در این مرحله هیچ حیوانی به دلیل MAF حذف نشد و تعداد ۴۴ نشانگر به دلیل عدم تعادل هاردی واینبرگ حذف شدند. نرخ کلی تعیین ژنوتیپ نیز عدد ۰/۹۴ به دست آمد (جدول ۱). پس از مراحل کنترل کیفیت



شکل ۱- نحوه پراکنش نشانگرهای ژنتیکی روی کروموزوم‌های اسب نژاد کردی (راست) و اسب نژاد کاسپین (چپ)

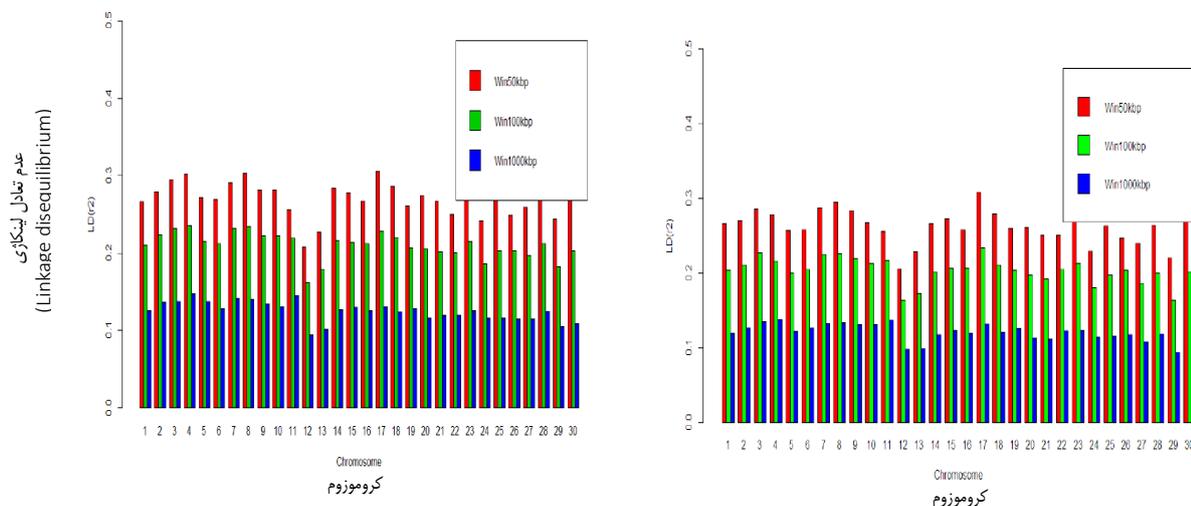
Figure1. Distribution of genetic markers on the chromosomes of Kurdish horses (right) and Caspian horses (left) breed

است. یکی از این روش‌ها، روش NextSNP و روش معمول‌تر آن روش فاصله ژنتیکی است. در این مطالعه پس از انجام مراحل کنترل کیفیت مقدار عدم تعادل لینکاژی برای سه پنجره با طول ۵۰، ۱۰۰ و ۱۰۰۰ کیلو بازی و برای هر نشانگر محاسبه شد و میانگین آن‌ها برای هر نژاد در شکل (۲) نمایش داده شده است.

پراکنش نشانگرها روی کروموزوم‌های نژاد اسب کاسپین تقریباً مشابه با اسب کردی بود و کروموزوم شماره یک و ۳۲ به ترتیب با ۴۰۷۵ و ۲۶۷۷ نشانگر دارای بالاترین تعداد SNP بودند.

**بررسی عدم تعادل لینکاژی در دو جمعیت اسب‌های کرد و کاسپین با روش فاصله فیزیکی:**

عدم تعادل لینکاژی پایه و اساس بسیاری از آنالیزهای مرتبط با جمعیت‌ها است. این مؤلفه با چند روش قابل محاسبه



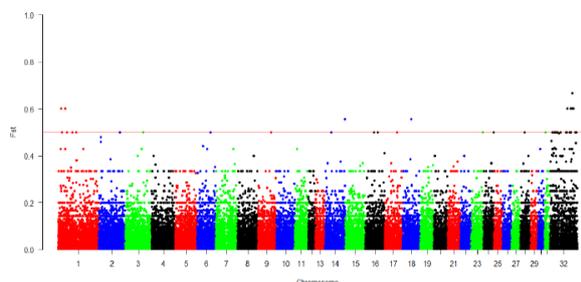
شکل ۲- عدم تعادل لینکاژی در کروموزوم‌های اسب‌های نژاد کردی (راست) و کاسپین (چپ)  
Figure 2. Linkage disequilibrium in the chromosomes of Kurdish (right) and Caspian (left) horse breed

ساختار جمعیت سایر نژادهای اسب بومی ایران و ارزیابی روابط ژنتیکی در بین این نژادها مورد مطالعه قرار دادند. هر دو تجزیه و تحلیل PCA و STRUCTURE در نژادهای مختلف اسب ایرانی نشان داد که اسب‌های عرب به‌طور جداگانه از نژادهای ترکمن و اسب‌های کاسپین خوشه می‌گیرند، اما با دره شوری و اسب‌های کرد همپوشانی دارند. آن‌ها در مجموع تعداد ۲۹ منطقه از ژنوم اسب را شناسایی کردند. در مطالعه دیگر بابائی و همکاران (۳) تحت عنوان مقایسه روش‌های تجزیه مولفه‌های اصلی (PCA) در بررسی ساختار جمعیتی نژادهای اسب آخال تکه، کاسپین و عرب با استفاده از اطلاعات ژنومی نشان داده شد که این سه نژاد به صورت مجزا گروه‌بندی شده‌اند و اسب کاسپین در گروه جدایی نسبت به اسب آخال تکه و عرب قرار دارد. به‌طور کلی بر اساس شکل (۳) و خوشه‌بندی انجام‌شده بر اساس pc1 و pc2 دو نژاد اسب کردی و کاسپین در دو گروه دور از هم و مجزا قرار گرفتند. نژاد اسب کردی دارای پراکندگی بیشتری نسبت به نژاد اسب کاسپین بود. طبق تحقیقات گذشته و تحقیق پیش رو مشخص شد که اسب کاسپین و کرد در دو گروه جداگانه قرار دارند. این نتیجه از روش آزمون مولفه‌های اصلی هم سو با نتایجی بود که محققانی همچون ریچ و همکاران (۲۳)، پاچو و همکاران (۲۰) و تومیر و همکاران (۱۹) داشتند. این محققین گزارش کردند که آنالیز مولفه‌های اصلی توانایی شناسایی مهاجرت‌ها در جمعیت‌های ایزوله شده را دارا می‌باشد. این روش همچنین به عنوان جایگزینی برای روش‌های خوشه‌بندی نیز پیشنهاد شده است (۱۴).

به‌طور کلی مقدار عدم تعادل لینکاژی در فاصله کوتاه‌تر (از نظر موقعیت فیزیکی SNPها) دارای میانگین بالاتری است و با افزایش فاصله این مقدار کاهش می‌یابد. میانگین عدم تعادل لینکاژی در پنجره ۱۰۰۰ کیلو بازی برابر با ۰/۱۲ و در پنجره ۱۰۰ کیلو بازی برابر با ۰/۲ و در پنجره ۵۰ کیلو بازی برابر با ۰/۳۰ بود. مقدار پراکندگی عدم تعادل لینکاژی در پنجره‌های مختلف پائین است. معمولاً زمانی که میانگین عدم تعادل لینکاژی کروموزوم‌ها واریانس کمی داشته باشند، جمعیت به‌عنوان یک جمعیت بسته و بدون انتخاب شدید برای یک ناحیه خاص ژنومی در نظر گرفته می‌شود و در دو نژاد مورد مطالعه نیز واریانس بین کروموزوم‌ها چندان شدید نبود. در نژاد اسب کردی کروموزوم‌های ۳، ۲۸، ۲۵، ۳۰ و ۸ و در نژاد اسب کاسپین کروموزوم‌های شماره ۳، ۴، ۳۰، ۲۳، ۱۷، ۸ و ۲۸ دارای بالاترین مقادیر عدم تعادل لینکاژی بودند.

#### نتایج بررسی آزمون مؤلفه‌های اصلی PCA

در این مطالعه پس از ادغام داده‌های هر دو نژاد و طی مراحل کنترل کیفیت این آزمون انجام شد. نتایج این آنالیز نشان داد که یک حیوان از نژاد اسب کاسپین در گروه نژاد اسب کردی قرار گرفت و همچنین سه اسب نژاد کردی در گروه اسب نژاد کاسپین قرار گرفتند. به همین دلیل این اسب‌ها از ادامه آنالیزها حذف شدند شکل (۳). طبق گزارشات همایونی و همکاران (۹) با استفاده از مجموعه داده SNP۶۷۰K میزان تنوع جایگاه‌های مورد نظر را در ۷۱ رأس اسب عرب مشخص کردند. در این تحقیق ۳۰ نمونه از ترکمن‌ها، کاسپین، دره شوری و اسب‌های کرد را برای تعیین

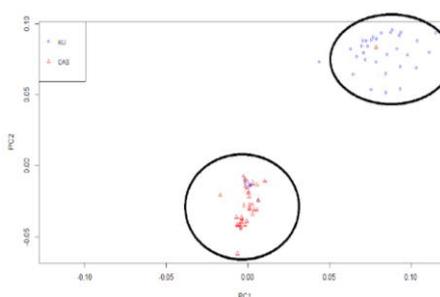


شکل ۴- مقادیر ارزش‌های  $F_{st}$  به روش برآوردگر ناریب تتا  
Figure 4.  $F_{st}$  values using unbiased theta( $\theta$ ) estimator method

هستی‌شناسی ژن مربوط به این ژن شامل اتصال کربوهیدرات است (۳۰). ژن DOCK3 کدکننده پروتئینی است که، فاکتور بالقوه تبادل نوکلئوتید گوانین (GEF) است. همچنین همراه با برخی پروتئین‌ها مانند GEF، GTPase های کوچک را فعال می‌کنند. DOCK ۳ را به‌عنوان تعدیل‌کننده جدید سلامت ماهیچه شناسایی می‌کنند و ممکن است اهداف درمانی برای درمان علائم عضله دیستروفیک داشته باشد. ژن rbms3 پروتئین متصل به RNA را رمزگذاری می‌کند که از خانواده پروتئین‌های تک‌رشته‌ای متصل به ژن c-myc است. ژن itgb6 پروتئینی را رمزگذاری می‌کند که عضوی از خانواده فوق‌العاده اینتگرین است. اعضای این خانواده گیرنده‌های چسبنده‌ای هستند که در سیگنالینگ از ماتریکس خارج سلول به داخل سلول عمل می‌کنند (۱۷). ژن GPM6B گلیکوپروتئین غشایی را کد می‌کند که متعلق به خانواده پروتئین‌های پروتئولیبید است. اعضای خانواده پروتئین لیپید در بیشتر مناطق مغز بیان می‌شوند و تصور می‌شود در عملکردهای سلولی مانند انتقال غشایی و ارتباطات سلول به سلول دخیل هستند (۷).

ژن PHEX دستورالعمل‌هایی برای ساخت آنزیمی را دربردارد که در درجه اول در استخوان‌ها و دندان‌ها فعال است. هنوز مشخص نیست که آنزیم PHEX چگونه به کنترل جذب مجدد فسفات کمک می‌کند و چه نقشی در شکل‌گیری و رشد استخوان‌ها دارد (۲۱).

ژن THOC2 ژن مهمی است که به تکامل و عملکرد سلول‌های عصبی (سلول‌های مغزی) کمک می‌کند. اختلالات مرتبط با THOC2 طیف وسیعی از اختلالات عصبی شناختی است که وقتی ژن THOC2 قادر به عملکرد صحیح نباشد، ایجاد می‌شود. اختلالات مرتبط با THOC2 با تاخیر در رشد و ناتوانی ذهنی مشخص می‌شود (۱۳).



شکل ۳- آزمون مؤلفه‌های اصلی (pca) برای اسب‌های نژاد کردی و کاسپین  
Figure 3. Principal components analysis for Kordi and Caspian horse breed

### آماره $F_{st}$

بعد از اجرای آزمون مؤلفه‌های اصلی و حذف حیوانات برهم زنده خوشه‌بندی، بقیه حیوانات جهت بررسی الگوی ژنومی انتخاب مثبت در دو جمعیت اسب نژاد کرد و اسب نژاد کاسپین مقادیر ارزش‌های  $F_{st}$  به روش برآوردگر ناریب تتا برای کلیه نشانگرهای تک نوکلئوتیدی در محیط R محاسبه شد (شکل ۴).

حد آستانه معنی‌داری SNPها در آزمون  $f_{st}$  برابر با ۰/۰۵ در نظر گرفته شد (SNPهایی که سطح معنی‌داری (p-value) کمتر از ۰/۰۵ داشتند انتخاب شدند) این حدود آستانه بسیار سخت‌گیرانه بوده و فقط مناطقی از ژنوم که دارای بیشترین تمایز جمعیتی بین دو نژاد را نشان دادند انتخاب شدند. لازم به ذکر است متوسط ارزش  $F_{st}$  در این پژوهش برابر با مقدار ۰/۰۵ به دست آمد. مقصودی و همکاران (۱۵) گزارش کردند تنوع ژنتیکی اسب عرب ایرانی نسبت به اسب‌های کرد بیش‌تر بوده و نشانه‌های انتخاب در ۶ جایگاه ژنومی بررسی و آماره  $F_{st}$  در کروموزوم شماره ۸ بیشترین میزان را دارا بوده است. ولی در تحقیق پیش رو بیشترین میزان  $F_{st}$  در روی کروموزوم ۳۲ است. جدول (۲) برخی از ژنهای شناسایی شده توسط آزمون  $F_{st}$  را نشان می‌دهد. عملکرد برخی ژن‌های شناسایی شده متمایزکننده دو نژاد:

ژن INPP5F یکی از اعضای خانواده اینوزیتول پلی فسفات-۵-فسفاتاز را رمزگذاری می‌کند در واقع این ژن کدکننده پروتئینی است. آنزیم ترشح‌شده با غیرفعال‌سازی فسفات‌های اینوزیتول در تنظیم سیگنالینگ کلسیم عمل می‌کند (۱۱). ژن HPSE۲ دستورالعمل‌هایی را برای ساخت پروتئینی به نام هیپاراناز ۲ ارائه می‌دهد (۶). ژن CNTN1 یک ژن کدکننده پروتئین است. بیماری‌های مرتبط با CNTN شامل میوپاتی مادرزادی، کامپتون-نورث، حاشیه‌نویسی

جدول ۲- ژن‌های مرتبط با جایگاه‌های معنی‌دار با استفاده از آزمون Fst

Table 2. Genes related to the significant loci using Fst test

شماره کروموزوم (Chromosome number)	نام نشانگر در فایل مپ (Marker name in the map file)	فاصله فیزیکی (Physical distance)	Fst	شروع ژن (Gene start)	پایان ژن (Gene end)	نام ژن (Gene name)
1	CUHSNP00036505	12935713	0.6	12915713	12955713	INPP5F
1	BIEC2_13989	31117195	0.6	31107195	31127159	HPSE2
6	BIEC2_1012122	60856989	0.5	60806989	60906989	CNTN1
16	BIEC2_341420	37940018	0.5	37900018	37980018	DOCK3
16	BIEC2_351603	55493263	0.5	55443263	55543263	RBMS3
18	BIEC2_411519	40888725	0.5	40838752	40938725	ITGB6
23	UKUL4181	7192727	0.4	7172727	7212727	WWC3
23	BIEC2_1109922	10444967	0.5	10404967	10484967	GPM6B
24	BIEC2_1172931	17603993	0.5	17553993	17653993	PHEX
31	BIEC2_1116780	27192136	0.5	27142136	27242136	DMD
32	BIEC2_1120232	34906699	0.5	34901699	34911699	DDX3X
32	TBIEC2_1182673	40596004	0.5	40586004	40606004	WDR13
32	BIEC2_1145451	94705145	0.6	94605145	94805145	HTR2C
32	BIEC2_1145629	95144119	0.5	95084119	95204119	IL13RA2
32	BIEC2_1206531	95266391	0.5	95246319	95286319	LRCH2
32	BIEC2_1208800	101923090	0.6	101823090	102023090	GRIA3
32	BIEC2_1208939	102230981	0.6	102130981	102330981	THOC2
32	TBIEC2_1209976	105327838	0.6	104863838	105828838	ACTRT1

### نتیجه‌گیری کلی

لینکاژی در پنجره ۱۰۰۰ کیلو بازی برابر با ۰/۱۲ و در پنجره ۱۰۰ کیلو بازی برابر با ۰/۲ و در پنجره ۵۰ کیلو بازی برابر با ۰/۳۰ بود. نتایج نشان داد بیشتر ژن‌های متمایزکننده دو نژاد با اندازه بدن، سرعت در انجام حرکات، اسکلت‌بندی بدن، انتقال دهنده‌های عصبی، ایمنی و .... در ارتباط هستند. از مهمترین ژن‌های شناسایی شده متمایزکننده دو نژاد می‌توان به THOC2، INPP5F، HPSE2، Dock3، itgb6، PHEX اشاره نمود. بسیاری از SNP‌های معنی‌دار در این پژوهش با مناطق ناشناخته ژنوم در ارتباط بودند و هنوز برای آنها عملکردی مشخص نشده است.

در این پژوهش نشانه‌های انتخاب متمایزکننده بین دو نژاد اسب کرد و کاسپین با استفاده از مولفه آماری Fst مورد بررسی قرار گرفت. نتایج تجزیه مولفه‌های اصلی نشان داد که دو نژاد اسب کاسپین و اسب کرد در دو دسته جداگانه قرار دارند. نژاد اسب کردی دارای تنوع و پراکندگی بیشتری نسبت به اسب کاسپین است. نتایج نشان داد که بیشترین نشانگرها در هر دو نژاد بر روی کروموزوم‌های ۱ و ۳۲ و کمترین آن‌ها بر روی کروموزوم‌های ۱۲ و ۱۳ قرار دارد. به طور کلی عدم تعادل لینکاژی در فاصله‌های ژنتیکی کمتر دارای میانگین بالاتری نسبت به فاصله‌های بیشتر بود و میانگین عدم تعادل

### منابع

- Amirinia, C., H. Seyedabadi, M.H. Banabazi and M.A. Kamali. 2007. Bottleneck study and genetic structure of Iranian Caspian horse population using microsatellites. *Pakistan Journal of Biological Sciences*, 10(9): 1540-1543. <https://doi.org/10.3923/pjbs.2007.1540.1543>
- Andolfatto, P. 2001. Adaptive hitchhiking effects on genome variability. *Current opinion in genetics development*, 11(6): 635-641.
- Babaei, N., A. Raft, A. Moradi and F. Derakhshi. 2021. Comparison of principal component analysis (PCA) and diagnostic analysis of principal components (DAPC) methods in investigating the population structure of Akhal-Teke, Arabian and Caspian horse breeds using genomic information. *Iran Animal Science Research*, 13(3): 453-462 (In Persian).
- Behrouzina, S., S.Z. Mirhosseini, F. Afraz, A. Sohrabi, S.A. Mohammadi, S. Shahbazi and S.B. Delirasfat. 2013. Genetic description of two populations of Iranian Turkmen horses in Turkmen Sahara and Turkmen Jorglan regions using microsatellite markers. *Iranian Animal Science Research* 1: 0-63 (In Persian).
- Bovo, S., A. Ribani, M. Munoz, E. Alves, J.P. Araujo, R. Bozzi and L. Fontanesi. 2020. Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. *Genetics Selection Evolution*, 52(1): 1-19.
- Daly, S.B., J.E. Urquhart, E. Hilton, E.A. McKenzie, R.A. Kammerer, M. Lewis and W.G. Newman. 2010. Mutations in HPSE2 cause urofacial syndrome. *The American Journal of Human Genetics*, 86(6): 963-969.
- Drabek, K., J. van-de-Poppel, M. Eijken and J.P. van-Leeuwen. 2011. GPM6B regulates osteoblast function and induction of mineralization by controlling cytoskeleton and matrix vesicle release. *Journal of Bone and Mineral Research*, 26(9): 2045-2051.
- Hedayat-Evrigh, N., E. Azadmard, R. Seyed Sharifi, S. Nikbin, M.D. Shakouri and R. Khalkhali-Evrigh. 2019. Investigation of genetic diversity of Iran northwest horses using microsatellite markers. *Agricultural Biotechnology Journal*, 11(4): 35-50.

9. Moon, S., J.W. Lee, D. Shin, K.Y. Shin, J. Kim, I.Y. Choi and H. Kim. 2015. A genome-wide scan for selective sweeps in racing horses. *Asian-Australasian journal of animal sciences*, 28(11): 1525-1531.
10. Khalkhali-Evrigh, R., N. Hedayat-Evrigh, H. Hafezian, A. Farhadi and M.R. Bakhtiarizadeh. 2020. Identification the copy number variation and its impacts on the genes of Iranian dromedary camels using whole genome sequencing data. *Iranian Journal of animal Science*, 51(2): 113-119.
11. Kim, H.S., A. Li, S. Ahn, H. Song and W. Zhang. 2014. Inositol Polyphosphate-5-Phosphatase F (INPP5F) inhibits STAT3 activity and suppresses gliomas tumorigenicity. *Scientific reports*, 4(1): 1-10.
12. Kim, Y. and W. Stephan. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2): 765-777.
13. Kumar, R., M.A. Corbett, B.W. Van-Bon, J.A. Woenig, L. Weir, E. Douglas and J. Gecz. 2015. THOC2 mutations implicate mRNA-export pathway in X-linked intellectual disability. *The American Journal of Human Genetics*, 97(2): 302-310.
14. Liu, N. and H. Zhao. 2006. A non-parametric approach to population structure inference using multilocus genotypes. *Human genomics*, 2(6): 1-12.
15. Mohammad-Maghsoudi, S., H. Mehrabani-Yeganeh and A. Nejati Javarami. 2017. Identifying regions under positive selection in the genes of Kurdish and Iranian Arabian horses using the method based on genetic linkage disequilibrium. *Animal Science of Iran*, 321-333 (In Persian).
16. Moshkelani, S., S. Rabiee and M. Javaheri-Koupaei. 2011. DNA fingerprinting of Iranian Arab horse using fourteen microsatellites marker. *Research Journal of Biological Sciences*, 6(8): 402-5.
17. Namekata, K., C. Harada, X. Guo, A. Kimura, D. Kittaka, H. Watanabe and T. Harada. 2012. Dock3 stimulates axonal outgrowth via GSK-3 $\beta$ -mediated microtubule assembly. *Journal of Neuroscience*, 32(1): 264-274.
18. Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics*, 39(1): 197-218.
19. Novembre, J. and M. Stephens. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics*, 40(5): 646-649.
20. Paschou, P., E. Ziv, E.G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M.W. Mahoney and P. Drineas. 2007. PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics*, 3(9): e160.
21. Quarles, L.D. 2003. FGF23, PHEX and MEPE regulation of phosphate homeostasis and skeletal mineralization. *American Journal of Physiology-Endocrinology And Metabolism*, 285(1): E1-E9.
22. Rafeie, F., C. Amirinia, A.N. Javaremi, S.Z. Mirhoseini and N. Amirmozafari. 2011. A study of patrilineal genetic diversity in Iranian indigenous horse breeds. *African Journal of Biotechnology*, 10(75): 17347-17352.
23. Reich, D., A.L. Price and N. Patterson. 2008. Principal component analysis of genetic data. *Nature genetics*, 40(5): 491-492.
24. Saravanan, K.A., M. Panigrahi, H. Kumar, B. Bhushan, T. Dutt and B.P. Mishra. 2020. Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science*, 241: 104257.
25. Seyedabadi, H.R. and S. Savarsofla. 2017. Microsatellite analysis for parentage verification and genetic characterization of the Turkmen horse population. *Kafkas Universitesi Veteriner Fakultesi Dergisi*, 23(3): 467-471.
26. Seyedsharifi, R., S. Badbarin, H. khamisabadi, N. hedayat-evrigh and J.S. Davati. 2019. Study of genetic structure and accuracy of assignment of individuals to five horse populations using microsatellite markers. *Research on Animal Production (Scientific and Research)* 10, no. (24): 120-126 (In Persian).
27. Shasavarani, H. and G. Rahimi-Mianji. 2010. Analysis of genetic diversity and estimation of inbreeding coefficient within Caspian horse population using microsatellite markers. *African Journal of Biotechnology*, 9(3): 293-299. <https://doi.org/10.5897/AJB09.1104>
28. Smadja, C.M., E. Loire, P. Caminade, M. Thoma, Y. Latour, C. Roux and P. Boursot. 2015. Seeking signatures of reinforcement at the genetic level: a hitchhiking mapping and candidate gene approach in the house mouse. *Molecular Ecology*, 24(16): 4222-4237.
29. Xu, L., D.M. Bickhart, J.B. Cole, S.G. Schroeder, J. Song, C.P.V. Tassell and G.E. Liu. 2015. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Molecular Biology and Evolution*, 32(3): 711-725.
30. Yan, J., D. Ojo, A. Kapoor, X. Lin, J.H. Pinthus, T. Aziz and D. Tang. 2016. Neural cell adhesion protein CNTN1 promotes the metastatic progression of prostate cancer. *Cancer research*, 76(6): 1603-1614.

## Analysis of SNP Chip 70k Genomic Data to Identify Loci Related to Differentiation of Caspian and Kurdish Horses using Selection Sweep

Mohadese Nasirpour<sup>1</sup>, Mohamad Moradi Shahr Babak<sup>2</sup>, Hossein Moradi Shahr Babak<sup>3</sup>,  
Hasan Mehrabani Yeganeh<sup>4</sup> and Younes Doosti<sup>5</sup>

1- Graduated M.Sc. Student, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran

2- Professor, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran

3- Assistant Professor, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran,

4- Professor, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran

5- M.Sc. Student, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran,

(Corresponding author: y.devisty@ut.ac.ir)

Received: 4 December, 2022 Accepted: 25 January, 2023

### Extended Abstract

**Introduction and Objective:** Selection in animal populations to increase the frequency of ideal alleles causes of remaining some signs in the animal genome, which are usually associated with important traits. Today, with developments in next-generation sequencing and easier access to animal genomic information, some models have been proposed to identify selective signatures based on allelic frequency and haplotype blocks. This research aimed to identify the selective signatures in two horse breeds of Caspian and Kurdish using a k70 SNP chip.

**Material and Methods:** In this research blood samples from 35 Caspian and 31 Kurdish horses were collected. After DNA extraction and sequencing (by Illumina company), Quality control of the data was performed for minimum allele frequency (PMAF<0.05), genotyping rate (PGENO < 0.5), and deviation from Hardy-Weinberg equilibrium (PH-W<1 × 10<sup>-6</sup>). Then the theta (θ) test and ensemble site were used for identifying selective signatures and the positions of snps respectively, and finally, the genes related to these positions were identified.

**Results:** After quality control of data and integration of genomic data of two populations based on R package protocols, 61 horses with 52,650 SNPs remained for the rest analysis. Finally, Fst statistical test based on unbiased estimator theta method 31 differentiating markers of the two studied horse breed were identified

**Conclusion:** The results showed that Caspian and Kurdish horse breeds belong to two separate groups. The Kurdish horse breed has more diversity and variety than the Caspian. The most important genes associated with significant SNPs were INPP5F, HPSE2, R3HCC1, DOCK3, ITGB6, GM6B, PHEX, WDR13, LRCH2, GRIA3 and THOC2. Most of the identified genes were associated with intercellular exchanges, muscle contractions, immunity, membrane support, DNA stability, and the nervous system.

**Keywords:** Caspian horse, Fst, Kurdish horse, PCA, Selective signatures