

ژنومی برای شناسایی تنوع تعداد کپی در اسب‌های کاسپین و ترکمن با استفاده از آرایه چند شکلی تک نوکلئوتیدی با تراکم بالا

سمیه منوچهر^{۱*}، قدرت رحیمی میانجی^۲، سید حسن حافظیان^۳ و محمد باقر زندی^۳

تاریخ دریافت: ۱۴۰۰/۱۲/۱۵ تاریخ پذیرش: ۱۴۰۱/۰۲/۲۸

^۱ دانشجوی دکتری، گروه علوم دامی، دانشگاه کشاورزی و منابع طبیعی ساری

^۲ استاد، گروه علوم دامی، دانشگاه کشاورزی و منابع طبیعی ساری

^۳ استادیار، گروه علوم دامی دانشکده کشاورزی، دانشگاه زنجان

*مسئول مکاتبه: s.manoochehr60@yahoo.com

چکیده

زمینه مطالعاتی: تنوع در تعداد کپی (CNV) همراه با چند شکلی‌های تک نوکلئوتیدی (SNPs)، نقش کلیدی در ارزیابی تنوع ژنتیکی در گونه‌های اهلی ایفا می‌کنند. با این وجود، در مورد تنوع تعداد کپی در اسب ایرانی اطلاعات کمی موجود است. روش کار: در این مطالعه، شناسایی CNVها و (CNVR)های موجود در ژنوم اسب کاسپین و ترکمن بر اساس داده‌های SNP آرایه اسب (Equine70k) انجام شد. در این مطالعه، از داده‌های تعیین ژنوتیپ شده ۴۸ نمونه (۲۶ راس اسب ترکمن و ۲۲ راس اسب کاسپین) از آرایه کمپانی ایلومینا (Equine 70k) که توسط شرکت Gene seek انجام گرفته بود، استفاده شد. تعداد کپی فقط در کروموزوم‌های اتوزوم با استفاده از نرم افزار PennCNV انجام شد. تجزیه و تحلیل CNVRها برای محتوای ژن با استفاده از BioMart و پایگاه داده Ensembl Genes 93 Horse EquCab3 انجام شد. در حاشیه نویسی کاربردی از پایگاه داده DAVID (<http://david.abcc.ncifcrf.gov>) و برای آنالیز Gene Ontology و تجزیه و تحلیل مسیر از پایگاه داده KEGG استفاده شد. برای تأیید CNVهای شناسایی شده تعداد ۵ جایگاه CNV به وسیله Real-time qPCR مورد آزمایش قرار گرفت. در این آزمون از ژن گلیسرآلدئید-۳-فسفات دهیدروژناز (GAPDH) به عنوان کنترل داخلی استفاده شد (متزگر و همکاران ۲۰۱۳). نتایج: در مجموع تعداد ۲۰۲ و ۱۰۵ به ترتیب CNV و CNVRs در اسب‌های مورد مطالعه شناسایی شدند که ۱/۰۸ درصد ژنوم اسب را پوشش می‌دهند. تنوع در تعداد کپی در نژاد اسب کاسپین نسبت به نژاد اسب ترکمن ۱/۶ برابر بیشتر بود. همچنین، متوسط طول CNV در نژاد کاسپین بزرگتر از نژاد ترکمن بود. در هر دو نژاد، تعداد رخداد ژنتیکی اضافه نسبت به رخداد ژنتیکی حذف بیشتر بود. در اسب‌های نژاد کاسپین کروموزوم‌های شماره ۱، ۳ و ۱۲ و در اسب‌های نژاد ترکمن کروموزوم شماره ۱، ۶ و ۱۲ به ترتیب بیشترین تنوع در تعداد کپی را نشان دادند. آنالیز عملکردی نشان داد که CNVRهای شناسایی شده با ۴۳۴ ژن همپوشانی داشتند که بیشتر این ژن‌ها در بین نمونه‌های اسب دو نژاد مشترک بودند (بیش از ۶۰ درصد). در میان این ژن‌ها، PPARG و GALR ارتباط بالقوه ای با صفات خاص نژادی داشته باشند. همچنین آنالیز KEGG چندین مسیر بیولوژیکی از جمله حس بویایی، محرک شیمیایی، پردازش آنتی ژن و مسیر سیگنال دهی پروتئین G را نشان داد. همچنین ۶۰ درصد از CNVRهای شناسایی شده با موفقیت توسط Real-time qPCR تایید شد. نتیجه‌گیری نهایی: CNVها ممکن است بخشی از تنوع فنوتیپی و شواهدی از انطباق پذیری اسب ایرانی را توضیح دهند. با توجه به اینکه ژن‌های دخیل در تعدادی از اجزای سلولی، فرآیندهای بیولوژیکی و عملکردهای مولکولی در CNVRهای شناسایی شده گزارش شد؛ اهمیت چنین CNVRها و تأثیر احتمالی آن باید مشخص شود و ممکن

است بینش جالبی را در مورد پیامد عملکردی و تطبیقی CNVs در اسب داشته باشد. در نهایت یافته‌های این تحقیق می‌توانند اطلاعات ارزشمندی را برای درک بهتر ژنوم اسب و هم چنین ارتباط صفات مهم عملکردی با CNVRها و ژن‌های همراه آنها در مطالعات آتی در نژادهای اسب فراهم نمایند.

واژگان کلیدی: آرایه SNP، CNV، ترکمن، کاسپین

مقدمه

می‌توانند، باعث تنوع و اختلالات فنوتیپی شوند. اگر چه، مطالعات متعددی در ارتباط با CNVها در انسان و حیوانات اهلی گزارش شده است (بوتو و همکاران ۲۰۱۹)؛ اما، مطالعات روی CNVها در بعضی از گونه‌های دام مانند اسب هنوز بسیار محدود است. چندین مطالعه برای بررسی تغییرات تعداد کپی در اسب‌ها با استفاده از تکنیک‌های مختلف انجام گرفته است (چن و همکاران ۲۰۱۲ و دون و همکاران ۲۰۱۲ و دوپیس و همکاران ۲۰۱۳ و گاش و همکاران ۲۰۱۴ و جون و همکاران ۲۰۱۴ و کادر و همکاران ۲۰۱۶ و شورینک و همکاران ۲۰۱۸ و وانگ و همکاران، ۲۰۱۴). بخشی از این مطالعات به منظور بررسی ارتباط بین CNVs و یک ویژگی خاص، برای مثال یک بیماری یا برای بررسی میزان بیان ژن انجام شد (گاش و همکاران ۲۰۱۴ و دون و همکاران ۲۰۱۲ و دوپیس و همکاران ۲۰۱۳ و پالینا و همکاران ۲۰۱۷) به عنوان مثال، به نظر می‌رسد یک تکرار ۶۲ کیلو جفت بازی روی کروموزوم ۱۰ با نوروپاتی حنجره عودکننده مرتبط باشد. و همچنین یک تکرار ۴/۶ کیلو جفت بازی اینترونیک در ژن STX17 روی کروموزوم ۲۵ با سفید شدن مو و ملانوم (رزنگرن و همکاران ۲۰۰۸) و حذف‌های بزرگ در یک ناحیه شبه اتوزومی با اوتیسم اسکلتی در پونی‌های شتلند^۲ (رافاتی و همکاران ۲۰۱۶) مرتبط است. اگر چه این مطالعات پایه ای، برای درک نقش CNVs در زیست‌شناسی اسب مهم هستند، اما اطلاعات فعلی برای کشف موثر این نواحی که بر سلامت و دیگر اختلالات در اسب تاثیر می‌گذارند ناکافی است. از دلایل ناکافی بودن این اطلاعات و یا نتایج متفاوت می‌تواند به سبب استفاده از پلتفرم‌های مختلف کشف CNV، پائین بودن تعداد نژادها

تنوع ژنتیکی، در سطح ژنوم حاصل چند شکلی‌های تک نوکلئوتیدی (SNPها)، درج / حذف شدگی‌ها (INDELS) و CNV^۱ است که به عنوان منبع اصلی تنوع فنوتیپ‌ها در حیوانات و انسان‌ها شناخته شده‌اند (فریمن و همکاران ۲۰۱۴). CNV، یکی از انواع تغییرات ساختاری است که با اضافه یا حذف بخش‌هایی از DNA به طول یک کیلو جفت باز تا چندین مگاجفت باز نسبت به ژنوم مرجع شناخته می‌شود (عبدلی و همکاران ۲۰۲۱). در مقایسه با چند شکلی‌های تک نوکلئوتیدی، CNVها با توجه به اندازه بزرگتر و در نتیجه تعداد بیشتری نوکلئوتیدهای درگیر، منبع مناسبی برای بررسی تنوع هستند (ردون و همکاران ۲۰۰۶). CNVها میزان بروز بیماری‌های مندلی و چند ژنی را توضیح می‌دهند و مسئول تغییر در بیان فنوتیپی صفات پیچیده هستند (دون و همکاران ۲۰۱۲). ثابت شده است که CNVها بر بیان ژن، تغییرات فنوتیپی، پیشرفت بیماری و سازگاری گونه‌ها تأثیر می‌گذارند (دیبزینسکی و همکاران ۱۹۹۸). ساختار CNVها می‌تواند ساده باشد، مانند تکرارهای پشت سر هم، یا ممکن است شامل اضافه‌شدگی یا حذف توالی‌های همولوگ در چندین مکان ژنومی باشد (ردون و همکاران ۲۰۰۶). توسعه سریع فن‌آوری‌های ژنوتایپینگ DNA همراه با ظهور پویای کل ژنوم، به تشخیص انواع تنوع ساختاری مانند CNVها کمک کرده است. امروزه، متداول‌ترین روش‌ها برای شناسایی CNV، آرایه‌های ترکیبی ژنومی مقایسه‌ای (CGH)، توالی‌یابی نسل بعدی (NGS) و آرایه‌های SNP هستند. در اسب نشان داده شده است که CNVها

² Shetland

¹ Copy Number Variation

(۲۰۰۷)، تعداد ۶۱۱۷۵ SNP اتوزومی در ادامه بررسی برای تجزیه و تحلیل CNV حفظ شدند. موقعیت ژنی نشانگرها براساس توالی تراشه جدید ژنوم اسب EquCab3.0 مکان یابی مجدد شد. خوانش CNV فقط در کروموزم‌های اتوزوم با استفاده از نرم افزار PennCNV انجام شد (وانگ و همکاران ۲۰۰۷). برای خوانش CNV توسط PennCNV ابتدا، فایل‌های شدت سیگنال برای هر ۴۸ نمونه اسب ساخته شد. فایل‌های شدت سیگنال حاوی اطلاعات شامل Log R Ratio (LRR) و فراوانی آلل (BAF) برای هر SNP تشکیل شد. فایل‌های حاوی فراوانی جمعیت آلل (PFB) B بر اساس مقدار متوسط BAF هر نشانگر در جمعیت ایجاد شدند. PennCNV شامل یک آرگومان به نام GCmodel است که با استفاده از یک مدل رگرسیون برای تنظیم محتوای GC بالا و بهبود نمونه تحت تاثیر "امواج ژنومی" بکار می‌رود (دوپیس و همکاران ۲۰۱۳). امواج ژنومی می‌تواند بر خوانش ژنوتیپ‌ها و صحت تعیین تنوع ساختار کروموزومی اثرگذار باشد. بنابراین باید برای این امر تصحیح صورت پذیرد (دوپیس و همکاران ۲۰۱۳). فایل مدل GC اسب برای تنظیم امواج ژنومی از گزینه "gcmode" ساخته شد. این فایل محتوای GC ناحیه ژنومیک مگاجفت باز اطراف هر نشانگر را مشخص می‌کند (هر ۵۰۰ کیلوگفت باز در هر طرف). در ادامه نمونه‌های با انحراف استاندارد $LRR > 0.35$ ، $BAF > 0.05$ و $waviness > 0.1$ factor حذف شدند. مناطق تغییر تعداد کپی (CNVRs) به صورت همپوشانی CNVها در حداقل یک جفت باز تعیین شد (هو و همکاران ۲۰۱۱). از نرم افزار CNVRuler (کیم و همکاران ۲۰۱۲) برای استخراج CNVRs استفاده شد. CNVRهای مشتق شده بر اساس حذف، اضافه و حذف/اضافه تعریف شدند. برای به حداقل رساندن خوانش‌های CNVRs مثبت کاذب، پارامتر آستانه تراکم منطقه‌ای CNV زیر ۱۰ درصد در طی فرآیند ادغام نرم افزار CNVRuler، با استفاده از مقدار پیش فرض (۰/۱) اصلاح شد. تجزیه و تحلیل CNVRها برای محتوای ژن با استفاده از BioMart و پایگاه داده Ensembl Genes 93 Horse

و همچنین تعداد افراد در هر یک از این نژادها باشد. علاوه بر این، اکثر CNVهای گزارش شده مختص یک مطالعه خاص هستند که توسط دو یا چند مطالعه مستقل دیگر تایید نشده‌اند. بنابراین هنوز هم تعداد زیادی از CNVها باقی مانده‌اند که تاکنون در نژادهای مختلف اسب از جمله اسب‌های ایرانی شناسایی نشده‌اند و نیازمند مطالعات بیشتر برای شناخت این CNVها و متعاقباً، بررسی نقش آنها در کنترل صفات پیچیده فنوتیپی است. نژادهای اسب عرب ایرانی، کاسپین، ترکمن، دره شوری و کرد دارای اهمیت دیرینه از نقطه نظر پرورشی، فرهنگی و تاریخی در ایران هستند (فیروز ۱۹۹۸ و فتوتی ۲۰۰۰). وضعیت تنوع ژنتیکی در این نژادها، در مطالعات مختلفی مورد بررسی قرار گرفته است و در غالب این مطالعات تنوع ژنتیکی بالایی را در اسب‌های بومی ایران گزارش کرده‌اند (عبدلی و همکاران ۲۰۲۱ و امجدی و همکاران ۲۰۱۷ و صادقی و همکاران ۲۰۱۹ و یوسفی و همکاران ۲۰۲۱). از این رو با توجه به اهمیت موضوع، در مطالعه حاضر تنوع ژنتیکی در ژنوم اسب‌های نژاد کاسپین و ترکمن از نظر CNV مورد بررسی قرار گرفت.

مواد و روش‌ها

در این مطالعه، از داده‌های تعیین ژنوتیپ شده ۴۸ نمونه (۲۶ راس اسب ترکمن و ۲۲ راس اسب کاسپین) از آراییه کمپانی ایلومینا (Equine 70k) که توسط شرکت Geene seek انجام گرفته بود، استفاده شد (زند و همکاران ۱۳۹۳). در ابتدا، کروموزوم X و سپس SNPهای بدون مکان مشخص از ادامه آنالیز حذف شدند (هو و همکاران ۲۰۱۱). در ادامه برای تعیین آستانه عبور هر یک از نمونه‌ها و SNPها از فرایند کنترل کیفی، از شاخص‌های: $MAF < 0.05$ ، $Animal\ call\ rate < 0.95$ ، $SNPs\ call\ rate < 0.95$ و شاخص تعادل هاردی-واینبرگ با سطح احتمال کمتر از 10^{-6} استفاده شدند. نمونه‌ها و SNPهایی که هر یک از معیارها را نداشته‌اند از ادامه آنالیز حذف شدند. پس از کنترل کیفیت با استفاده از بسته نرم‌افزاری PLINK (پرسل و همکاران

زیر بود: واسرشت سازی اولیه در ۹۴ درجه سانتیگراد به مدت ۳ دقیقه، به دنبال آن ۴۰ چرخه دناتوراسیون در ۹۴ درجه سانتیگراد به مدت ۱۵ ثانیه و دمای اتصال و دمای بسط در ۶۰ درجه سانتیگراد به مدت ۶۰ ثانیه. qPCR برای هر نمونه در سه تکرار انجام شد. برای برآورد CNV از روش $2^{-\Delta\Delta C_t}$ استفاده شد. که در آن $\Delta\Delta C_t$ معادل است با: $(C_t \text{ ژن هدف در هر نمونه دارای CNV مورد نظر} - C_t \text{ ژن کنترل در هر نمونه دارای CNV مورد نظر}) - (C_t \text{ ژن هدف در نمونه‌های بدون CNV مورد نظر} - C_t \text{ ژن کنترل در نمونه‌های بدون CNV مورد نظر})$. طبق گزارش لیو و همکاران (۲۰۱۳) مقادیر حدود ۲، ۳ یا بالاتر و ۱ یا پایین تر برای کروموزوم‌های اتوزومی به ترتیب وضعیت طبیعی، اضافه و حذف را نشان می‌دهد. در این مطالعه عدد کمتر از ۲ بعنوان حذف در نظر گرفته شد. در نهایت برای ارزیابی قابلیت اطمینان CNVRهای شناسایی شده، نتایج به دست آمده در این تحقیق با مطالعات منتشر شده روی CNVهای اسب مقایسه شد (دون و همکاران ۲۰۱۲ و دوپیس و همکاران ۲۰۱۳ و گاش و همکاران ۲۰۱۴ و جون و همکاران ۲۰۱۴ و متزگر و همکاران ۲۰۱۳ و وانگ و همکاران ۲۰۱۴).

EquCab3 انجام شد. در حاشیه نویسی کاربردی از پایگاه داده DAVID (<http://david.abcc.ncifcrf.gov/>) و برای آنالیز Gene Ontology و تجزیه و تحلیل مسیر از پایگاه داده KEGG استفاده شد. از آنجا که تعداد ژن‌های اسب حاشیه نویسی شده در بانک اطلاعاتی هستی شناسی ژن (GO) هنوز محدود است، لیست ژن‌های اسب بازیابی شده ENSEMBL با استفاده از نرم افزار Biomart به شناسه‌های ژن انسانی ENSEMBL متعامد آن تبدیل شد. حاشیه نویسی عملکردی در قالب پارامترهای اجزاء سلولی (cellular component)، عملکرد مولکولی (molecular function) و فرآیندهای بیولوژیکی (biological processes) انجام شد. برای تایید CNVهای شناسایی شده تعداد ۵ جایگاه CNV به وسیله Real-time qPCR مورد آزمایش قرار گرفت. در این آزمون از ژن گلیسرآلدئید-۳-فسفات دهیدروژناز (GAPDH) به عنوان کنترل داخلی استفاده شد (متزگر و همکاران ۲۰۱۳). تعداد ۶ جفت پرایمر مربوط به جایگاه‌های مورد نظر توسط برنامه‌ی Primer3 طراحی و توسط پرایمر بلست اختصاصی بودن آنها مورد بررسی قرار گرفت (جدول ۱). از معرف Power SYBR Green Power PCR در واکنش qPCR استفاده شد. شرایط qPCR به شرح

Table 1. Specifications of primers used in real-time qPCR

CNVR	Direction	Sequence	Product Size	Event*
GAPDH	F	TGAGAGGGGAATGGAGAAGTG	119	normal
	R	TCCCTCCACGATGCCAAAGT		
8: 5751419-5755950	F	ACCAACAGTGAGACTCGGAG	210	loss
	R	CAATGTGCTATGGACCTGGC		
3: 67780091-67799192	F	AGTGGGTGGAGCATTGAGAA	368	gain
	R	TGCTCTGGTATTTGGCTCAC		
3: 8600487-8669915	F	ATCCCAGGTTTGTTCGCTCG	230	loss
	R	CCTCTCATCTTCTGGTGCTG		
6: 4334586-4414882	F	GCCAGGAAGCAAAAGGGAAA	229	gain
	R	TGGTGCCTGCTGTTTCTTGA		
7: 75259939:75634194	F	TACAACGAGGAAGCAGCAGA	245	loss
	R	TGCAGCATGGACTTCAGGAA		

Normal: no change in copy number; **Loss:** Deleted in copy number; **Gain:** Added in copy number

نتایج و بحث

CNV و CNVRها

پس از کنترل کیفی CNVهای شناسایی شده، در مجموع تعداد ۲۰۲ CNV در دو نژاد اسب ایرانی شناسایی و متوسط خوانش برای هر حیوان ۴/۲ رویداد CNV برآورد شد. دامنه تعداد CNV برای هر نمونه از ۱ تا ۱۸ عدد متغیر بود. از این میان تعداد ۵۲ CNV حذف و ۱۵۰ عدد CNV اضافه مورد شناسایی قرار گرفتند. بنابراین تعداد وقایع اضافه تقریباً سه برابر وقایع حذف بود. از نظر نژادی نیز در هر دو نژاد تعداد اضافه شدگی بیشتری نسبت به حذف شدگی شناسایی شد. از تعداد ۲۰۲ CNV شناسایی شده، تعداد ۱۲۶ و ۷۶ CNV به ترتیب در اسب‌های نژاد کاسپین و ترکمن ردیابی شدند که نشان می‌دهد CNV در نژاد کاسپین نسبت به نژاد ترکمن ۱/۶ برابر بیشتر و متوسط طول CNVها ۳۰۹/۸ کیلو جفت باز بود. به تفکیک نژاد نیز متوسط طول CNV در نژاد کاسپین بزرگتر از نژاد ترکمن و در مجموع تعداد ۱۹ (۹/۱۴ درصد) در بین اسب‌های دو نژاد کاسپین و ترکمن مشترک بودند (جدول ۲).

از نظر پراکنش طولی، به طور کلی تقریباً ۳۸/۱۲ درصد از CNVها در بازه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز، ۷۳/۲۶ درصد بین ۵۰ و ۱۰۰ کیلو جفت باز و ۳/۹۶ درصد در بازه طولی ۵۰۰ تا ۱۰۰۰ کیلو جفت باز، ۱۲/۳۱ درصد در بازه طولی ۱۰ تا ۵۰ کیلو جفت باز و ۸/۲۴ درصد مربوط به رویدادهای CNV بزرگتر از ۱ مگاجفت باز بود. همچنین ۴/۱۰ درصد تنوعات تعداد کپی کمتر از ۱۰ کیلو جفت باز شناسایی شد (شکل A۱). بیشترین فراوانی (۲۶/۷۳ درصد) طول قطعاتی که دچار اضافه شدگی شده‌اند در دامنه طولی تا ۵۰۰ کیلو جفت باز و کمترین فراوانی (۶/۹۳ درصد) در دامنه طولی بیشتر از یک مگاجفت باز قرار داشتند. همچنین بیشترین فراوانی (۱۱/۳۹ درصد) طول قطعاتی که دچار حذف شدگی شده‌اند در دامنه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز و

کمترین فراوانی (۱ درصد) در دامنه طولی ۵۰۰ تا ۱۰۰۰ جفت باز قرار داشتند (شکل A۱).

ادغام CNVهای همپوشان منجر به شناسایی ۱۰۵ CNVRs اتوزومی شد که ۲۳/۹۷ مگاجفت باز (۱۰۰۸ درصد) از توالی ژنوم اسب را پوشش می‌دهد. اندازه CNVRs بین ۴۵ جفت باز تا ۲ مگاجفت باز (با میانگین ۲۳۵/۲۲۲ کیلو جفت باز و میانه ۲۱۰/۶ جفت باز) متغیر بود. در نژاد کاسپین کروموزم‌های شماره ۱، ۳ و ۱۲ و در نژاد ترکمن کروموزم‌های شماره ۱، ۶ و ۱۲ به ترتیب دچار بیشترین تغییر در تعداد کپی بودند. ۴۷/۴۴ درصد CNVRs بین ۵۰ تا ۵۰۰ کیلو جفت باز در حالیکه ۰/۱۲ درصد از CNVRs کمتر از ۱۰ کیلو جفت باز طول داشتند (شکل B۱). بیشترین فراوانی (۲۹/۳۳ درصد) طول نواحی قطعاتی که دچار اضافه شدگی شده‌اند در دامنه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز و کمترین فراوانی (۰/۰۴ درصد) در دامنه طولی کمتر از ۱۰ کیلو جفت باز قرار داشت. همچنین بیشترین فراوانی (۱۰/۶۹ درصد) طول قطعاتی که دچار حذف شدگی شده‌اند در دامنه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز و کمترین فراوانی (۰ درصد) در دامنه طولی بزرگتر از یک مگاجفت باز قرار داشت. ۱۲/۳ درصد از نواحی نیز دچار حذف/اضافه شده است (شکل B۱).

مناطق CNVRs به جز در کروموزم‌های ۱۶، ۲۱، ۲۲، ۲۷، ۳۰ در بقیه کروموزم‌های اتوزوم پراکنده بودند (شکل ۲). از نظر توزیع کروموزمی، روی کروموزم‌های ۱ و ۱۲ بیشترین تعداد CNV شناسایی شد در حالی که هیچ CNV روی کروموزم‌های ۱۶، ۲۱، ۲۲، ۲۷ و ۳۰ شناسایی نشد.

پوشش کروموزومی CNVRs در بین کروموزم‌های مختلف از ۰/۰۰۴ تا ۸/۲۷ درصد متغیر بود به طوری که بالاترین درصد پوشش در کروموزوم ۱۲ و کمترین درصد پوشش روی کروموزوم ۱۰ مشاهده شد (جدول ۲). از نظر وقایع CNVR، ۷۲ نمونه (۶۸/۵۸ درصد) به صورت اضافه در حالی که ۳۱ نمونه (۱۸/۸۸ درصد)

وقایع حذف و ۲ نمونه (۱۲/۵۴ درصد) حذف/اضافه را نشان دادند. جدول ۳ به طور خلاصه جزئیات نتایج مربوط به CNVRs را در ژنوم اسب‌های کاسپین و ترکمن نشان می‌دهد.

Table 2. Detected of CNVs in the genome of Caspian and Turkmen horses

Breed	Samples(n)	CNVs(n)	CNVs Average size (kb)	CNVs Median size (kb)	Average CNVs per sample(n)
Caspian	22	126	323.97	96.11	5.25
Turkman	26	76	286.48	105.79	3.16
Total	48	202	309.86	103.61	4.2

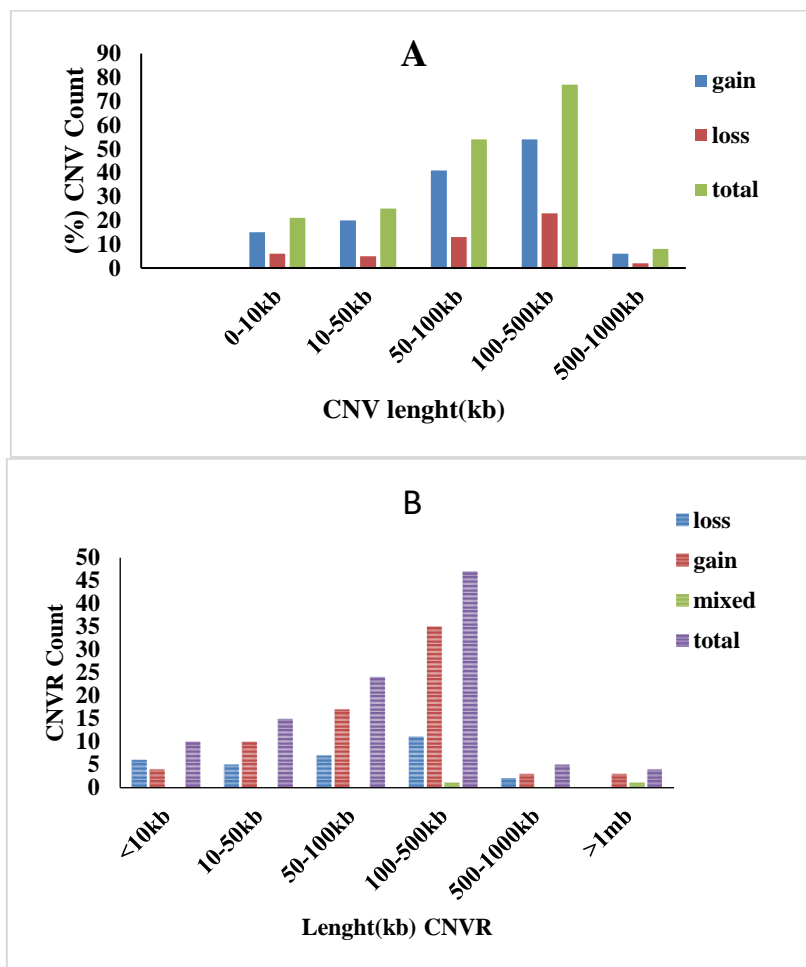


Figure 1. CNV (A) and CNVR (B) length distributions in the genomes of Caspian and Turkmen horses

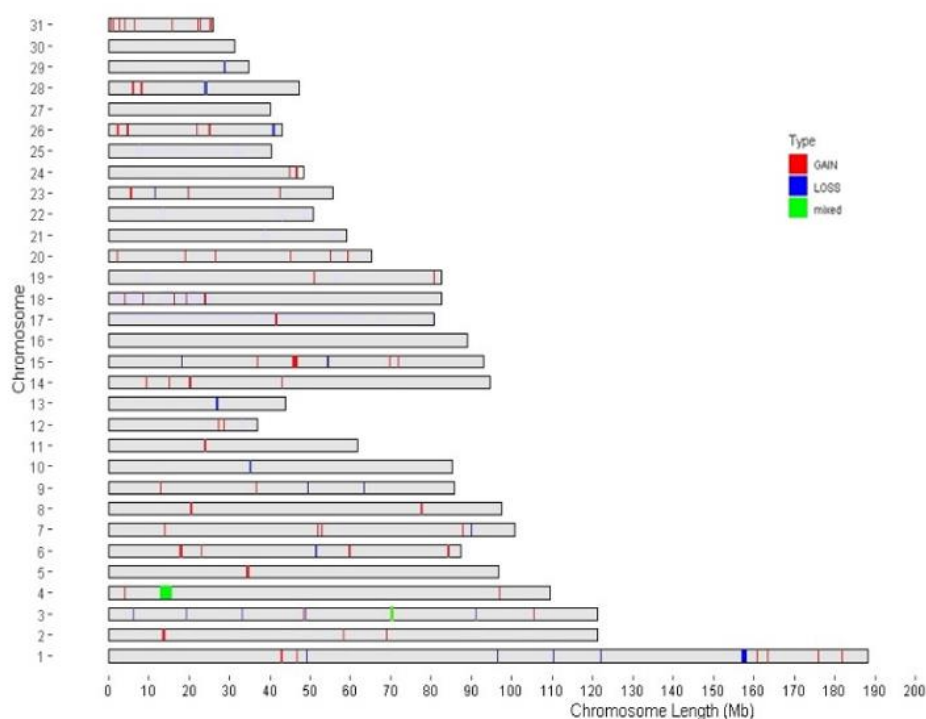


Figure 2. Chromosomal distribution of CNVRs (gain, loss and mixed) in the genomes of Caspian and Turkmen horses

Table 3. Details of CNVRs identified in the Caspian and Turkmen equine genomes

Chromosome	CNVR(n)	CNVR total length (kb)	Chromosome length(kb)	Genomic Coverage(%)	CNVR Average length (kb)
1	13	2238.709	188260	1.19	172.20
2	3	289.3	121350	0.23	96.43
3	8	800.53	121350	0.65	100.06
4	8	968.93	109460	0.88	121.11
5	1	93.17	96760	0.09	93.17
6	5	1008.46	87230	1.15	201.96
7	5	1024.59	100790	1.01	204.91
8	2	333.805	97560	0.34	166.90
9	4	443.35	85790	0.51	110.83
10	1	3.49	85160	0.004	3.49
11	1	21.27	61680	0.03	21.26
12	2	3062.69	36990	8.28	1531.35
13	1	26.03	43780	0.05	26.03
14	4	492.96	94600	0.52	123.24
15	6	803.77	92850	0.86	133.97
17	1	82.93	80720	0.1	82.93
18	5	413.1	82640	0.5	82.62
19	2	115.87	82680	0.18	57.97
20	6	728.67	65340	1.11	121.44
23	4	930.173	55560	1.67	232.54
24	2	36.65	48290	0.07	18.82
25	2	607.66	40280	1.5	303.83

26	5	1129.35	40280	2.8	22587
28	3	263.45	43150	0.61	87.82
29	1	119.06	47350	0.25	119.06
31	10	8659.32	26000	33.30	865.93
Total	105	4698.31	2280920	1.08	235.22

*Chromosomes without CNVR are not included in the table.

احتمالی عدم تطابق این نتایج در بین مطالعات مختلف را می‌توان به پارامترهای مختلفی مانند اندازه و زمینه ژنتیکی نمونه‌ها، پلت فرم‌های مختلف تشخیص و الگوریتم‌های فراخوانی CNV، تعاریف CNV و CNVRs و همچنین روش‌های برآورد تصادفی خطاها مرتبط دانست (پینتو و همکاران ۲۰۱۱). به عنوان مثال هر چه تعداد نمونه‌های مورد مطالعه و نشانگرهای تک نوکلئوتیدی استفاده شده برای شناسایی CNVها بیشتر باشد در مجموع تعداد CNVهای شناسایی شده نیز بیشتر خواهد بود. تعداد پائین CNVRهایی که توسط PennCNV در مطالعه حاضر شناسایی شده بودند، مشابه نتایج سایر مطالعات CNV در انسان و پستانداران دیگر بود (جیانگ و همکاران ۲۰۱۲).

در مطالعه حاضر، متوسط تعداد CNVها برای هر فرد ۴/۵ بود (دامنه ۱ تا ۱۸)، که به طور قابل توجهی کمتر از تعداد شناسایی شده توسط دون و همکاران (۲۰۱۲) و بیشتر از دوپیس و همکاران (۲۰۱۳) و متزگر و همکاران (۲۰۱۳) و کادر و همکاران (۲۰۱۶) (۳ = ۹۶/۲۸۷) بود.

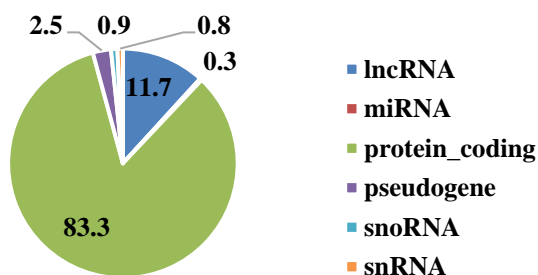


Figure 3. Percentage of genes identified in CNVR regions in the Caspian and Turkmen horse genomes

تا به امروز، نشان داده شده است که CNVها بزرگترین بخش تغییرات ژنتیکی در ژنوم بسیاری از گونه‌ها از جمله اسب را توضیح می‌دهند (۸،۱۳). اما وقوع CNVها در اسب‌ها و تأثیر بعدی آنها بر تغییرات فنوتیپی تا به امروز نسبت به سایر حیوانات اهلی به خوبی مورد بررسی قرار نگرفته است. اگرچه کارآیی شناسایی CNV با استفاده از آرایه‌های SNP در مقایسه با آرایه‌های دیگر (به عنوان مثال آرایه CGH) کمتر است، اما نتایج مطالعه حاضر که از آرایه SNPها حاصل شده‌اند را می‌توان با سایر مطالعات CNV انجام شده در اسب مقایسه کرد.

این مطالعه اولین تجزیه و تحلیل تنوع ژنوم از نظر تعداد کپی در اسب‌های ایرانی است. در مجموع از تجزیه و تحلیل ۶۱۱۷۵ داده ژنوتیپی حاصل از SNP در ۴۸ اسب ایرانی در دو نژاد کاسپین و ترکمن، ۲۰۲ رخداد CNV اتوزومی و ۱۰۵ CNVR شناسایی شد که معادل ۲۳/۹۷ مگاجفت باز از ژنوم اسب می‌باشد. این نتایج را می‌توان با آنچه قبلاً در مطالعات CNV / CNVR در اسب گزارش شده بود مقایسه کرد (دون و همکاران ۲۰۱۲ و دوپیس و همکاران ۲۰۱۳ و گاش و همکاران ۲۰۱۴ و جان و همکاران ۲۰۱۴ و کادر و همکاران ۲۰۱۶ و متزگر و همکاران ۲۰۱۳ و وانگ و همکاران ۲۰۱۴). در این مطالعه نتایج حاصل از چند متغیر با نتایج هشت مطالعه دیگر مقایسه شد (جدول ۴). به عنوان مثال، میانگین اندازه CNVRهایی که توسط آرایه‌های ۵۰K و آرایه‌های ۵۰Kv شناسایی شدند، به طور قابل توجهی بزرگتر از میانگین اندازه‌های بدست آمده توسط آرایه‌های CGH و NGS بودند. این امر ممکن است به دلیل پوشش نسبتاً کم و توزیع غیریکواخت SNP در ژنوم اسب در آرایه‌های SNP ارتباط داشته باشد. دلایل

Table 4. Summary of studies performed to identify CNV in horses during the years 2012 to 2019

	This Study	Doan et al (2012b)	Metzger et al (2013)	Dupuis et al (2013)	Wang et al (2014)	Jun et al (2014)	Gosh et al (2014)	Kader et al (2016)	Corbi-Botto et al (2019)
Platform	Equine SNP 70	WGS	Equine SNP 50	Equine SNP 50	Array CGH	WGS	Array CGH	Equine SNP 70	Equine SNP 70
Number of samples / breeds	2.48	1.1	717.17	447.4	6.6	1.1	38.16	96.3	24.1
Number of CNVRs	105	282	50	478	353	2569	258	122	87
CNVR domain	3.48-1.5	3.7-4.8	0.5-0.9	0.1-2.7	0.5-6.1	3-6.4	2.1-5	0.2-2.2	0.5-2
Number of gains	72	274	n/a	238	109	869	64	50	48
Number of losses	31	6	n/a	240	234	1710	172	69	39
Mixed number	2	2	n/a	n/a	n/a	n/a	22	3	0
Median	105.794	n/a	169	61	131	n/a	46	98	n/a
Mean	235.22	n/a	487	114	39	56	110	1552	143.3
Genomic coverage	1.08	3.5	1.7-22	2.3	0.6	n/a	1.2	0.8	0.6

دون و همکاران (۲۰۱۲) (۱۵۰۹ اضافه و ۸۵۹ حذف) و شرینک و همکاران (۲۰۱۸) (۸۵/۵ درصد اضافه و ۱۴.۵ درصد حذف) و کوربی و همکاران (۲۰۱۹) (۵۵.۲ درصد اضافه و ۴۴.۸ درصد حذف) اما متفاوت با بعضی دیگر از مطالعات بود (دوپیس و همکاران ۲۰۱۳ و گاش و همکاران ۲۰۱۴ و متزگر و همکاران ۲۰۱۳ و وانگ و همکاران ۲۰۰۷).

هستی شناسی ژنها و مسیرهای ژنتیکی موجود در CNVR

در مجموع ۴۳۴ ژن که به طور کامل در CNVRs قرار داشتند، شناسایی شدند. در میان این ژن‌های همپوشان CNVR، ۸۳/۳۱ درصد ژن‌های کد کننده پروتئین، ۱۱/۷۲ درصد lncRNA، ۲/۵ درصد ژن‌های کاذب، ۰/۹ درصد snoRNA و ۰/۲۵ درصد مرتبط به miRNA بودند (شکل ۳).

در ادامه از ژن‌های ارتولوگ انسانی برای انجام تجزیه و تحلیل GO برای تعیین اثرات بیولوژیکی ۴۳۴ ژن‌های همپوشان با CNV‌های شناسایی شده، استفاده شد.

علاوه بر تجزیه و تحلیل مسیر KEGG نشان داد مسیرهای بیولوژیکی غنی شده در CNVRهای شناسایی شده عمدتاً با ادراک حسی، متابولیسم و ایمنی انتقال بویایی ($P=10^{-3} * 3/4$) و پردازش و ارائه آنتی ژن ($P=0.0000016$) پس از تصحیح Bonferroni مرتبط شده‌اند (جدول ۵). مسیرهای بیولوژیکی شناسایی شده در این تحقیق با مطالعات قبلی CNV در اسب‌ها (شورنیک و همکاران ۲۰۱۸ و بوتو و همکاران ۲۰۱۹) و سایر پستانداران به عنوان مثال گوسفند و گاو مطابقت داشت (لیو و همکاران ۲۰۱۳ و هو و همکاران ۲۰۱۱).

درصد پوشش ژنومی CNVRs در این مطالعه (۱/۰۸ درصد) به دست آمد. به این معنا که از کل ژنوم اسب بیش از یک درصد آن دارای CNV بوده است. در مطالعات پیشین بررسی تغییرات تعداد کپی در انسان و موش نیز نشان داده است که ۵ تا ۱۰ درصد از ژنوم در این گونه‌ها دچار تغییرات ژنومی شامل تکرار یا حذف شدند (شارپ و همکاران ۲۰۰۶ و استچویک و همکاران ۲۰۱۷). همچنین درصد پوشش ژنومی در این مطالعه در مقایسه با نتایج ارائه شده توسط کادر و همکاران (۲۰۱۶)، کوربی و همکاران (۲۰۱۹) و وانگ و همکاران (۲۰۱۴) بیشتر بود، اما از نتایج گزارش دون و همکاران (۲۰۱۲)، دوپیس و همکاران (۲۰۱۳) کمتر، اما مشابه نتایج به دست آمده توسط گاش و همکاران (۲۰۱۴) و متزگر و همکاران (۲۰۱۳) بود. در مطالعه حاضر، کروموزم‌های شماره ۱ و ۱۲، دارای بالاترین تعداد CNV بودند. این نتایج مطابق با گزارشات دوپیس و همکاران (۲۰۱۳)، دون و همکاران (۲۰۱۲)، متزگر و همکاران (۲۰۱۳)، گاش و همکاران (۲۰۱۴) بود که در کروموزم‌های شماره ۳، ۴ و ۳۱ دارای بیشترین پوشش CNVRs بودند. CNV روی غالب کروموزم‌های اتوزوم به غیر از کروموزم‌های ۱۶، ۲۱، ۲۲، ۲۷، ۳۰ پراکنده شده بودند که مطابق با نتایج مطالعه کوربی و همکاران (۲۰۱۹) بوده است. همچنین متزگر و همکاران (۲۰۱۳) نیز نتایج مشابهی را به دست آورده‌اند (کروموزم‌های ۱۶، ۱۹، ۲۱، ۲۶، ۲۲، ۲۷ و ۳۱ فاقد تنوع بوده‌اند). در مطالعه حاضر، رخداد اضافه تقریباً سه برابر (۱۵۰) رخداد حذف (۵۲) بدست آمد. نتایج بدست آمده مشابه نتایج گزارش شده در اسب توسط کادر و همکاران (۲۰۱۶) (۱۴۲ حذف و ۱۴۵ اضافه در سه نژاد اسب چینی)،

Table 5. Functional analysis of overlapping genes with CNVRs identified in the Caspian and Turkmen equine genomes

Category	Term	Count	P-value	Bonferroni Correction
GOTERM_BP_DIRECT	GO:0050911~detection of chemical stimulus involved in sensory perception of smell	50	4.7E-39	3.8E-36
GOTERM_BP_DIRECT	GO:0007186~G-protein coupled receptor signaling pathway	52	9.7E-26	3.9E-23
GOTERM_BP_DIRECT	GO:0007608~sensory perception of smell	32	2.4E-26	.9E-23
GOTERM_BP_DIRECT	GO:0050907~detection of chemical stimulus involved in sensory perception	23	1.5E-24	4.1E-22
GOTERM_BP_DIRECT	GO:0002480~antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent	6	1E-08	0.0000021
GOTERM_BP_DIRECT	GO:0010529~negative regulation of transposition	5	6E-07	0.000096
GOTERM_BP_DIRECT	GO:0002474~antigen processing and presentation of peptide antigen via MHC class I	6	9.8E-06	0.0013
GOTERM_MF_DIRECT	GO:0004984~olfactory receptor activity	52	3.4E-42	8.9E-40
GOTERM_MF_DIRECT	GO:0004930~G-protein coupled receptor activity	53	1.8E-32	2.3E-30
GOTERM_MF_DIRECT	O:0004888~transmembrane signaling receptor activity	26	1.9E-20	1.6E-18
GOTERM_MF_DIRECT	GO:0005549~odorant binding	19	4E-19	2.6E-17
GOTERM_MF_DIRECT	GO:0016814~hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	5	2.4E-06	0.00012
GOTERM_CC_DIRECT	GO:0005886~plasma membrane	83	4.6E-14	8.2E-12
GOTERM_CC_DIRECT	GO:0042612~MHC class I protein complex	6	2.7E-08	0.0000024
GOTERM_CC_DIRECT	GO:0016021~integral component of membrane	79	3.2E-07	0.000019
KEGG_PATHWAY	hsa04740: Olfactory transduction	51	3.4E-35	5.1E-33
KEGG_PATHWAY	hsa04612: Antigen processing and presentation	10	1.6E-06	0.00013

کننده سازگاری موجود با محرک‌های محیطی و همچنین یک مزیت تکاملی محسوب می‌شود. به طور کلی انتخاب طبیعی باعث بیان بسیاری از ویژگی‌های سازگاری مانند طول عمر، باروری، سلامتی، مقاومت فیزیکی زیاد، تحمل درد و مقاومت هوایی و خصوصیات خاص نژادی در اسب می‌شود. به عنوان مثال شناسایی CNVR همپوشان با ژنهایی مانند PPARG و GALR که مرتبط با خصوصیات خاص نژادی هستند نشان از تطابق پذیری نژادها با شرایط محیطی دارد. PPARG نقش مهمی در حفظ عملکرد انسولین ماهیچه‌های اسکلتی دارد و حذف این ژن باعث مقاومت به انسولین می‌شود (هوینر و همکاران ۲۰۰۳). GALR ممکن است در اختلالات هیپوفیز نقش داشته باشد و نقش واسطه ای در عملکرد گالانین در محور

تجزیه و تحلیل عملکردی و مسیر KEGG نشان داد یکی دیگر از مهمترین مسیر غنی شده سیگنالینگ بویایی (olfactory receptors) در هر دو نژاد اسب کاسپین و ترکمن بود و CNVهای موجود در ژن‌های موجود در این مسیر در گزارش‌های قبلی نیز در بسیاری از گونه‌ها، از جمله گاو (هو و همکاران ۲۰۱۵)، گوسفند (لیو و همکاران ۲۰۱۳)، خوک (پادل و همکاران ۲۰۱۳) و اسب (متزگر و همکاران ۲۰۱۳) یافت شده بودند. اسب‌ها به دلیل داشتن ارگان تکامل یافته سیستم بویایی 1 و مجموعه وسیعی از ژن‌های عملکردی گیرنده بویایی دارای ظرفیت بالایی برای درک بویایی هستند (روگر و همکاران ۲۰۱۱). غنی سازی ژن‌های گیرنده‌های بویایی که به طور گسترده در مطالعات قبلی CNV نیز گزارش شده است (کوربی و همکاران ۲۰۱۹ و کیل و همکاران ۲۰۱۹ و پادل و همکاران ۲۰۱۳) منعکس

ژن‌هایی که در CNVR قرار گرفته بودند به طور معنی داری با ادراک حسی، ایمنی و متابولیسم غنی شدند.

نتیجه‌گیری کلی

تحقیقات CNV در اسب در سال‌های گذشته پیشرفت قابل توجهی داشته است. نقشه‌های CNV اسب در سطح ژنوم با استفاده از انواع پلتفرم‌ها و الگوریتم‌های تشخیص تولید شده‌اند. با این حال، همپوشانی بین مجموعه داده‌های CNV موجود بسیار کم است. این اختلافات احتمالاً ناشی از تفاوت در نژادها، اندازه نمونه، پلت فرم و الگوریتم تشخیص است. به طور کلی این مطالعه اولین بررسی CNV در اسب‌های نژاد ترکمن و کاسپین بود که نشان داد بخش قابل توجهی از ژنوم اسب دچار تغییرات ژنتیکی ناشی از حذف یا اضافه شدگی قطعات ژنی شده است. همچنین بعضی از این تغییرات در داخل یا همپوشان با ژنهای مهم قرار داشتند که ممکن است تاثیر مهم و اثر گذاری بر خصوصیات ژنتیکی و فنوتیپی داشته باشد. داده‌های حاصل از این پژوهش می‌تواند زمینه را برای مطالعات بعدی CNV در اسب در رابطه با اثرات فنوتیپی CNV‌های خاص، درک بهتر نقش تنوع ژنتیکی بر صفات مهم در اسب مانند عملکردهای ورزشی، تولید مثل و بیماری‌ها را فراهم کند.

هیپوتالاموس-هیپوفیز داشته باشد (دیپزینسکی و همکاران ۱۹۹۸).

اعتبارسنجی CNVها توسط Real time qPCR

به منظور تایید صحت CNV‌های شناسایی شده توسط PennCNV، از qPCR برای اعتبارسنجی پنج ناحیه CNVR (۲۲، ۴۴، ۴۵، ۳۴، ۴۸) به صورت تصادفی از بین تعداد ۱۰۵ جایگاه CNVR شناسایی شده در این مطالعه استفاده شد (شکل ۴).

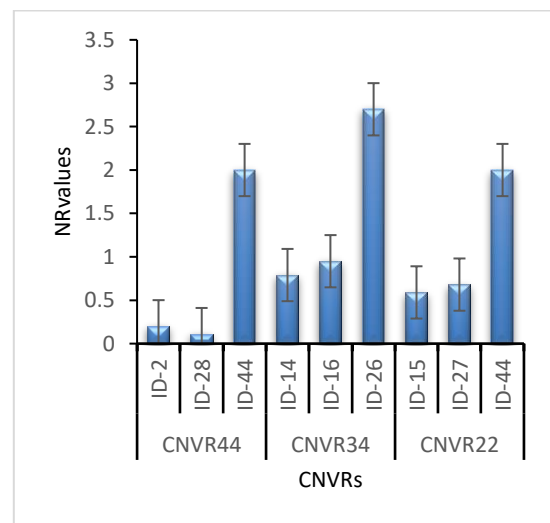


Figure 4. qPCR validation of three selected CNVR regions.

از میان پنج ناحیه مورد نظر، سه ناحیه CNVR‌های انتخاب شده (با وضعیت حذف loss) با موفقیت تایید شدند (شکل ۴). با توجه به اینکه نواحی مورد نظر ما دچار حذف شده بودند لذا وجود مقادیر پایین تر از عدد ۲ در نمودار مشهود است. به طور کلی برای هر یک از نواحی سه نمونه مورد آزمایش قرار گرفت که در میان هر کدام دو نمونه تایید و یک نمونه (مقادیر بالاتر از عدد ۲) رد شد. بقیه CNVRها احتمالاً به دلیل نرخ مثبت کاذب ناشی از نرم افزار و خارج بودن پرایمرها از مرزهای واقعی CNV برای برخی افراد تایید نشدند. در مطالعه حاضر در مجموع تعداد تنوعات تعداد کپی در نژاد کاسپین نسبت به نژاد ترکمن بیشتر و همچنین طول CNV در نژاد کاسپین بزرگتر از نژاد ترکمن بود. همچنین در هر دو نژاد

منابع مورد استفاده

- Abdoli M, Zandi MB, Harkinezhad T and Kahlili M. 2021. Genetic structure survey of Iranian native horse breeds by microsatellite markers, *Journal of Animal Productions* 23(2):155-163.
- Ala-Amjadi M, Yeganeh H.M and Sadeghi M. 2017. Study of genetic variation in Iranian Kurdish horse using microsatellite marker. *Iranian Journal of Animal Science* 48(3):335-342.
- Beckmann JS, Estivill X and Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews Genetics* 8:639–46
- Chen CY, Qiao RM, Wei RX, Guo YM, Ai HS, Ma JW, Ren J and Huang LS. 2012. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC Genomics* 13, 733:1-12
- Clop A, Vidal O and Amills M, 2012. Copy number variation in the genomes of domestic animals. *Animal Genetics*. 43: 503–17.
- Corbi-Botto CM, Morales-Durand H, Zappa ME, Sadaba SA, Peral-García P, Giovambattista G, and Díaz S. 2019. Genomic structural diversity in Criollo Argentino horses: Analysis of copy number variations. *Gene* 695: 26-31.
- Depczynski B, Nichol K, Faithi Z, Iismaa T, Shine J and Cunningham A. 1998. Distribution and characterization of the cell type sex expressing GALR2 mRNA in brain and pituitary gland. *Annals of the New York Academy of Sciences* 863:120–128.
- Doan R, Cohen N, Harrington J, Veazey K, Juras R, Cothran G, McCue ME, Skow L and Dindot SV. 2012. Identification of copy number variants in horses. *Genome Research*. 22: 899–907.
- Dupuis, MC, Zhang Z, Durkin K, Charlier C, Lekeux P and Georges M. 2013. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Animal Genetics* 44 (2), 206–208.
- Firouz L. 1998. Original ancestors of the Turkoman and Caspian horses. *First International Conference on Turkoman horses; Ashgabat, Turkmenistan*.
- Fotovati A, 2000. Persian horse breeds from ancient time to present and their rules in development of world horse breeds. *Asian-Australasian Journal of Animal Sciences*, 13(Suppl): 401.
- Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME and Carter NP. 2006. Copy number variation: new insights in genome diversity. *Genome Research*, 16(8): 949-61.
- Ghosh S, Qu Z, Das PJ., Fang E, Juras R, Cothran EG, McDonell S, Lear TL, Adelson DL, Chowdhary BP, Raudsepp T, 2014. Copy number variation in the horse genome. *PLoS Genetics*, (10): e1004712.
- Hevener AL, He W, Barak Y, Le J, Bandyopadhyay G, Olson P, Wilkes J, Evans RM. and Olefsky J, 2003. Muscle-specific Pparg deletion causes insulin resistance. *Nature Medicine*, 9: 1491–1497.
- Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM and Sonstegard TS. 2011. Genomic characteristics of cattle copy number variations. *BMC Genomics*, 12(1): 127.
- Hussain A, Saraiva LR and Korsching SI. 2009. Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. *Proceedings of the National Academy of Sciences of the United States of America* 106, 4313–8.
- Jiang L, Jiang J, Wang J, Ding X, Liu J. and Zhang Q, 2012. Genome-wide identification of copy number variations in Chinese Holstein. *Plos One*, 7(11): e48732.
- Jun J, Cho, YS, Hu H, Kim HM, Jho S, Gadhvi P, Park KM, Lim J and Paek WK, Han K. 2014. Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics* 15, S4.
- Kader A, Liu X, Dong K, Song S, Pan J, Yang M, Chen X, He X, Jiang L and Ma Y. 2016. Identification of copy number variations in three Chinese horse breeds using 70K single nucleotide polymorphism BeadChip array. *Animal Genetics* 47 (5): 560–569.
- Keel B, Nonneman D, Lindholm-Perry AK, Oliver W. and Rohrer G, 2019. A survey of copy number variation in the porcine genome detected from whole-genome sequence. *Frontiers in Genetics* 10: 737.
- Kim JH, Hu HJ, Yim SH, Bae JS, Kim SY and Chung YJ. 2012. CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics* 28(13): 1790-1792.
- Krueger K and Flauger B. 2011. Olfactory recognition of individual competitors by means of faeces in horse (*Equus caballus*). *Animal Cognition* 14, 245–57.

- Liu J, Calhoun VD, Chen J, Claus E D and Hutchison KE, 2013. Effect of homozygous deletions at 22q13. 1 on alcohol dependence severity and cue-elicited BOLD response in the precuneus. *Addiction Biology* 18(3): 548-558.
- Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F and Du L. 2013. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14(1): 229.
- Metzger J, Philipp U, Lopes MS, Machado da Camara, A, Felicetti, M, Silvestrelli M and Distl O. 2013. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics* 14 (1), 487.
- Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Bastiaansen JW, Crooijmans RP and Groenen MA, 2013. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics*, 14(1): 449.
- Pawlina-Tyszko K, Gurgul A, Szmatola T, Ropka-Molik K, Semik-GurgulEKlukowska-Rötzler J, Koch C, Mählmann K and Bugno-Poniewierska M, 2017. Genomic landscape of copy number variation and copy neutral loss of heterozygosity events in equine sarcoids reveals increased instability of the sarcoid genome. *Biochimie* 140: 122–32.
- Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R and Prasad A. 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology* 29(6): 512.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. and Sham PC, 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81(3): 559-75.
- Rafati N, Andersson LS, Mikko S, Feng C, Raudsepp T and Pettersson J, 2016. Large deletions at the SHOX locus in the pseudoautosomal region are associated with skeletal atavism in Shetland ponies. *G3 (Bethesda)*, 6(7): 2213–23.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH and Andrews TD. 2006. Global variation in copy number in the human genome. *Nature*, 444:444–454.
- Rosengren PG, Golovko A, Sundstrom E, Curik I, Lennartsson J and Druml T, 2008. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics* 40:1004–9.
- Sadeghi R, Moradi-Shahrbabak M, MiraeiAshtiani SR, Schlamp F, Cosgrove EJ and Antczak DF. 2019. Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. *Journal of Heredity*, 110(2): 173-182.
- Schurink A, da Silva VH, Velie BD, Dibbitts BW, Crooijmans RPMA and François L. 2018. Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity. *BMC Genomic* 19: 49.
- Sharp AJ, Cheng Z and Eichler EE, 2006. Structural variation of the human genome. *Annual Review of Genomics and Human Genetics* 7: 407–442.
- Stachowiak M, Szczerbal I, Nowacka-Woszek J, Jackowiak H, Sledzinski P, Iskrzak P, Stanislaw D and Switonski M. 2017. Polymorphisms in the SOX9 region and testicular disorder of sex development (38, XX; SRY-negative) in pigs. *Livestock Science* 203:48–53.
- Vila C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, Liden K, WayneRK and Ellegren H. 2001. Widespread origins of domestic horse lineages. *Science* 291: 474–477.
- Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF. and Bucan M, 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* 17(11):1665-1674.
- Wang W, Wang S, Hou C, Xing Y, Cao J, Wu K, Liu C, Zhang D, Zhang L and Zhang Y. 2014. Genome-wide detection of copy number variations among diverse horse breeds by array CGH. *Plos One*, 9, e86860.
- Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM. and Trask BJ, 2008. Extensive copy-number variation of the human olfactory receptor gene family. *The American Journal of Human Genetics* 83(2): 228-42.
- Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM and Trask BJ. 2008. Extensive copy-number variation of the human olfactory receptor gene family. *The American Journal of Human Genetics* 83(2): 228-42.

- Yousefi-Mashouf N, Mehrabani-Yeganeh H, Nejati-Javaremi A, Bailey E and Petersen JL. 2021. Genomic comparisons of Persian Kurdish, Persian Arabian and American Thoroughbred horse populations. *Plos One*, 16(2): e0247123.
- Zandi Baghcheh Maryam MB, 2014. Genomic scan for detection of selective sweeps in Turkmen horse population. PhD. Campus of Agriculture and Natural Resources, University of Tehran, Tehran, Iran, 112 pages.

Genomic scan for detection of copy number variations in Caspian and Turkman horses using high density SNP array

S Manoochehr¹, Gh Rahimi Mianji², SH Hafezian² and MB Zandi³

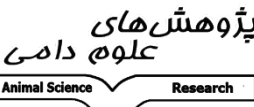

Received: March 6, 2022 Accepted: May 8, 2022

¹PhD Student, Department of Animal Science, Sari Agriculture Sciences and Natural Resources University, Sari, Iran

²Professor, Department of Animal Science, Sari Agriculture Sciences and Natural Resources University, Sari, Iran

³Assistant Professor, Department of Animal Science, University of Zanjan, Zanjan, Iran

*Corresponding Author: s.manoochehr60@yahoo.com

 <p>پژوهش‌های علوم دامی Animal Science Research</p>	<p>Journal of Animal Science/vol.33 No.4/ 2024/pp 29-45 https://animalscience.tabrizu.ac.ir</p>	
<p>© 2009 Copyright by Faculty of Agriculture, University of Tabriz, Tabriz, Iran This is an open access article under the CC BY NC license (https://creativecommons.org/licenses/by-nc/2.0/) DOI: 10.22034/AS.2022.50663.1652</p>		

Introduction: Horses have played an important role in the history of Iranians during different centuries. They kept horses for various aims such as agriculture, transportation, sport, food sources. Iran has a suitable climatic, social and economic potential for keeping and breeding horses, that is why it has been created different breeds using the selection and breeding. But due to problems such as mechanization, lack of government support, export ban, high costs of breeding and maintenance, import of foreign horses, lack of proper planning, interest in keeping horses has decreased. So, unfortunately, only a few native Iranian breeds remain. Additional investigation of the equine genomic architecture is critical for a better understanding of the equine genome, as well as for expanded comparisons across diverse mammalian species. Turkmen and Caspian horses are well-known breeds of Iranian horse breeds. These breeds were historically selected to perform distinct tasks and therefore may harbor a wealth of unique variation at the genome level. Copy number variation (CNV) along with single nucleotide polymorphisms (SNPs) play a key role in genetic diversity in livestock species. CNVs, a term that refers to a change in the number of copies of a genomic segment, are responsible for more sequence differences between individuals than SNPs and are considered to be a major source of genetic variation contributing to differences in phenotypes (Beckmann et al, 2007). Several studies identified copy number variations in horses using different techniques (Doan et al, 2012). Part of these studies tried to establish associations between CNVs and a specific trait, a disease or even gene expression (Schurink et al, 2017). Most of these studies found either no association or inconclusive associations as the number of horses with phenotypic information or with specific CNVs were limited. For example, a 62 kb duplication on Equus caballus (ECA) chromosome 10 seemed to be related to recurrent laryngeal neuropathy (Dupuis et al, 2013). However, little is known about CNV in Iranian horses.

Materials and methods: In this study, detection of CNVs and CNVRs were performed based on SNP data from Caspian and Turkman horse breeds were genotyped via Equine70k SNP beadchip. PennCNV software was only used to detect CNV on autosomes. The PennCNV algorithm was only applied to autosomes (command: -lastchr 26) to identify individual-based CNVs. To increase the confidence of the detected CNVs, quality control was performed by employing standard exclusions of the LRR (standard deviation of LRR) <0.3, a BAF drift <0.01 and a waviness factor <0.05. We classified the status of these CNV into two categories: “loss” (CNV containing a deletion) and “gain”

(CNV containing a duplication). The CNVRs were determined by aggregating the overlapping CNVs with CNVRuler. BioMart in the Ensembl database and DAVID was employed to identify genes located in CNVRs and GO terms and KEGG pathway analyses respectively. Quantitative real-time PCR (qPCR) was applied to validate the CNVRs that were detected in this study.

Results and discussion: A total of 202 and 105 CNVs and CNVRs were identified in the studied horses, respectively, which cover 1.08% of the horse genome. The number CNVs in Caspian breed were 1.6 times more than Turkmen. Also, the average size of CNVs in Caspian breed was longer than Turkmen. In both breeds, the genetic event of gain was higher than the genetic event of deletion. In Caspian breed, chromosomes 1, 3, 12, and in the Turkmen breed, chromosomes 1, 6 and 12 showed the most changes in CNVs, respectively. Functional analysis showed that the identified CNVRs overlapped with 434 genes and the most of these genes were common between the two horse breeds (more than 60%). Among these genes, PPARG and GALR have potential related with breed-specific traits. The KEGG pathway analysis also identified several pathways that are significantly enriched in olfactory sensory perception, chemical stimulus sensory perception, antigen processing, and G protein signaling pathway. Also, 60% of successfully detected CNVRs were confirmed by Real-time qPCR. The results of this study were compared with the results of eight other studies. For example, we concluded that the average size of the CNVRs detected by the 70k arrays and the 50K arrays were significantly larger than obtained by the CGH and NGS arrays. This may be due to the relatively low coverage and non-uniform distribution of SNP in the equine genome in SNP arrays. Possible reasons for the differences between our results and some CNV studies can be related to different parameters such as sample size and genetic background, different detection platforms and CNV retrieval algorithms, CNV definitions and CNVRs, as well as random error estimation methods (Pinto et al. 2011).

Conclusions: CNVs can describe part of the phenotypic diversity and adaptation evidence in Iranian horses. With regard to Genes identified in a number of cellular components, biological processes and molecular functions within CNVRs, the importance of such CNVRs and the possible effect needs to be studied and may interest insight into the functional and adaptive consequence of CNVs in horse. In total, the number of CNVs in the Caspian breed was greater than in the Turkmen breed, and also the CNV length in the number of copies in the Caspian breed was greater than in the Turkmen breed. In both breeds, there were overlapping genes with CNVRs that were significantly enriched in biological pathways, including sensory perception, immunity, and metabolism. This is the first CNV report on Turkmen and Caspian horses and the findings of this study could provide valuable information for better understanding of the horse genome and also the important performance traits with CNVRs and associated genes for the future studies in horse breeds.

Keywords: SNP BeadChip, Copy Number Variation, Turkmen, Caspian