تحقیقات تولیدات دامی

سال نهم/شماره اول/بهار ۱۳۹۹ (۴۴-۲۹)

Animal Production Research

دانشگاه گیلان

مقاله پژوهشی

بررسی تنوع تعداد کپی در ژنوم گوسفندان بلوچی با استفاده از تجزیه مقایسه‌ای الگوریتم‌های PennCNV و QuantiSNP

کبری تقی‌زاده[۱]، محسن قلی‌زاده[۲*]، محمد حسین مرادی[۳]، قدرت‌الله رحیمی میانجی[۴]

۱- دانشجوی دکتری، گروه علوم دامی، دانشکده علوم دامی و شیلات، دانشگاه کشاورزی و منابع طبیعی ساری

۲- استادیار، گروه علوم دامی، دانشکده علوم دامی و شیلات، دانشگاه کشاورزی و منابع طبیعی ساری

۳- استادیار، گروه علوم دامی، دانشکده علوم دامی، دانشگاه اراک

۴- استاد، گروه علوم دامی، دانشکده علوم دامی و شیلات، دانشگاه کشاورزی و منابع طبیعی ساری

(تاریخ دریافت: ۹۸/۰۳/۰۷ - تاریخ پذیرش: ۹۸/۰۹/۲۱)

چکیده

تنوع تعداد کپی (CNV)، از مهمترین تغییرات ساختاری ژنوم، به عنوان منبع مهم تنوع ژنتیکی و فنوتیپی شناخته شده است. هدف از این مطالعه بررسی مقایسه‌ای CNV در گوسفندان نژاد بلوچی با استفاده از الگوریتم‌های PennCNV و QuantiSNP بود. تجزیه داده‌ها با استفاده از آرایه تعیین ژنوتیپ SNP50K گوسفندی روی ۹۶ گوسفند بلوچی انجام شد. پس از تشخیص CNVها، مناطق تنوع تعداد کپی (CNVRs) با استفاده از برنامه CNVRuler تعیین شدند. در مجموع تعداد ۲۰۱ و ۹۱۶ CNV به ترتیب با الگوریتم‌های PennCNV و QuantiSNP شناسایی شد. همچنین ۹۱ CNVR (به طول ۱۸/۷۵ تا ۵۱۱/۷ کیلو جفت باز) با الگوریتم PennCNV و ۳۱۶ CNVR (به طول ۵/۷ تا ۱۲۸۰ کیلو جفت باز) با الگوریتم QuantiSNP شناسایی شد که به ترتیب در برگیرنده ۰/۴۶ و ۱/۳۳ درصد از کل ژنوم گوسفند بود. تعداد CNVهای نوع حذف در الگوریتم QuantiSNP حدود پنج برابر و در الگوریتم PennCNV حدود سه برابر بیشتر از تعداد اضافه‌ها بود. همچنین تعداد CNVهای شناسایی شده با الگوریتم QuantiSNP حدود چهار برابر بیشتر بود. میزان ۸۶/۶ درصد (۱۷۴ CNV با متوسط طول ۱۲۲/۶۷ کیلو جفت باز) از CNVهای شناسایی شده به وسیله الگوریتم PennCNV با CNVهای شناسایی شده در الگوریتم QuantiSNP مطابقت داشتند. در مجموع، نتایج این تحقیق نشان داد که استفاده از چندین الگوریتم می‌تواند تغییرات ساختاری ژنوم را با دقت بیشتری تشخیص دهد و منجر به درک بهتری از ژنوم گوسفند شود.

واژه‌های کلیدی: آرایه تعیین ژنوتیپ، الگوریتم‌های PennCNV و QuantiSNP، تغییرات ساختاری، تنوع تعداد کپی

* نویسندهٔ مسئول: mh_gholizadeh@yahoo.com

doi: 10.22124/ar.2020.13356.1418

مقدمه

در دسترس بودن اطلاعات توالی ژنوم افراد درون یک گونه امکان بررسی تنوع ژنتیکی را با وضوح بالا فراهم کرده است. منابع اصلی تنوع ژنتیکی در ژنوم افراد شامل چندشکلی‌های تک‌نوکلئوتیدی (SNP)، درج یا حذف‌های کوچک (Indels) و تغییرات بزرگ مقیاس هستند. تغییرات بزرگ مقیاس ممکن است به صورت تفاوت در تعداد کپی (حذف یا اضافه شدن قطعات کروموزومی) یا تغییرات خنثی تعداد کپی (مانند وارونگی‌ها یا جابه‌جایی‌های متعادل کروموزومی) رخ دهد (Hastings *et al*., 2009; Park *et al*., 2015). تنوع تعداد کپی (CNV) نیز به شکل اضافه یا حذف شدن بخش‌های ژنومی یا قطعاتی از DNA با طول یک کیلو جفت باز تا چند مگا جفت باز تعریف می‌شوند که در تعداد کپی در مقایسه با ژنوم مرجع متفاوت هستند (Freeman *et al*., 2006; Xi *et al*., 2011). CNVها به دلیل تغییرات در سطوح بیان ژن می‌توانند یا به طور مستقیم با تکرار یا حذف یک ژن، یا به طور غیر مستقیم از راه آثار موضعی در پایین دست و شبکه‌های تنظیمی بر فنوتیپ تأثیر بگذارند (Dermitzakis and Stranger 2006; Reymond *et al*., 2007). CNVها به عنوان یک شکل اصلی ژنتیکی تغییرات ساختاری به طور گسترده‌ای در ژنوم انسان توزیع شده‌اند و بر بیان ژن، تنوع فنوتیپی و انطباق‌پذیری به وسیله اختلال ژن‌ها و تغییر دادن دُز ژن مؤثر باشند (Feuk *et al*., 2006; Zhang *et al*., 2009). همچنین در مطالعات انسانی گزارش شده است که SNPها و CNVها، به ترتیب تقریباً ۸۲ درصد و ۱۸ درصد از کل تغییرات ژنتیکی تشریح شده در بیان ژن را به خود اختصاص می‌دهند، که این امر نشان‌دهنده اهمیت مشارکت CNVها در حساسیت به بیماری و تنوع فنوتیپی است (Stranger *et al*., 2007). مطالعات تنوع تعداد کپی برای درک ساز و کار تکاملی در اهلی کردن دام و سازگاری آن‌ها با شرایط محیطی، شناخت زیر بنای جایگاه صفات کمّی (QTL) و شناخت CNVهای مرتبط با صفات مهم اقتصادی برای سرعت بخشیدن به پیشرفت ژنتیکی انجام می‌شود (Bhanuprakash *et al*., 2018). تأثیر CNVها بر صفات متفاوت فنوتیپی در حیوانات مختلف شامل گاو، گوسفند، بز، خوک، اسب، سگ، مرغ، بوقلمون و اردک گزارش شده است (de Lemos *et al*., 2018). آثار قابل توجه CNV مشخص شده در حیوانات عبارتند از رنگ سفید و سیاه پوشش در گوسفندان به علت تکرار ژن پروتئین سیگنالی آگوتی (ASIP) (Norris and Whan, 2008) و در بز (Fontanesi *et al*., 2009)؛ رنگ پوشش سفید در خوک‌ها به علت تغییر تعداد کپی در ژن KIT (Giuffra *et al*., 2002) و فنوتیپ تاج نخودی در جوجه تحت تأثیر CNV در اینترون اول ژن SOX5 ایجاد می‌شود (Wrigh *et al*., 2009). CNVها با روش‌های مختلفی از جمله روش هیبریداسیون درجا فلورسنت (FISH)، آرایه هیبریداسیون ژنومی مقایسه‌ای (aCGH)، آرایه SNP و نسل‌های جدید توالی‌یابی (NGS) شناسایی می‌شوند (Pinkel *et al*., 1998; Carter, 2007; Wain *et al*., 2009). در مقایسه با روش‌های اشاره شده، مزایای استفاده از آرایه SNP شامل هزینه‌های کم آن، پوشش متراکم و توان عملیاتی بالا است. داده‌های توالی‌یابی تولید شده در مطالعات ارتباط ژنومی (GWAS) می‌تواند به طور مستقیم برای تجزیه CNV مورد استفاده قرار گیرد (Zhao *et al*., 2013). اخیراً تعدادی از الگوریتم‌های خوانش CNV توسعه داده شده‌اند که GWAS مبتنی بر CNV را تسهیل می‌کنند (Lai *et al*., 2005; Kim *et al*., 2012). با این حال، با توجه به محدودیت اساسی داده‌های SNP برای اندازه‌گیری شدت سیگنال، در مورد احتمال پاسخ‌های کاذب یا حساسیت اندک برای تشخیص CNVها نگرانی وجود دارد (Kim *et al*., 2012). در واقع خوانش CNV وابسته به انواع پلت‌فرم‌های آرایه و ابزارهای تحلیلی است. هر پلت‌فرم و الگوریتم خوانش دارای مزایا و معایب خاص خود است؛ بنابراین، یک الگوریتم یکپارچه یا پلت‌فرم آرایه همیشه برای تعیین CNVها کافی نیست (Baumbusch *et al*., 2008; Hester *et al*., 2009). اخیراً نشان داده شده است ابزارهای تحلیلی مختلفی که برای داده‌های خام مشابه اعمال می‌شوند؛ معمولاً خوانش‌های CNV را با مطابقت کمتر از ۵۰ درصد مشخص می‌کنند و با استفاده از الگوریتم‌های متعدد، تعداد پاسخ‌های مثبت کاذب به حداقل می‌رسند (Pinto *et al*., 2011). طیف گسترده‌ای از ابزارهای کشف CNV بر اساس داده‌های حاصل از آرایه‌های SNP، مانند CNVpartion (Illumina) (Korn *et al*., 2008)، Birdsuite (Wang *et al*., 2007)، PennCNV

و QuantiSNP (Collea *et al*., 2007) توسعه داده شده‌اند. PennCNV و QuantiSNP دو الگوریتم هستند که بر اساس مدل‌های مخفی مارکوف (HMMs) طراحی شده‌اند. هر دو این الگوریتم‌ها می‌توانند داده‌های Illumina و SNP Affymetrix را پردازش کنند. PennCNV چندین منبع اطلاعات شامل نسبت لگاریتم R (LRR) و فراوانی آلل B (BAF) در هر SNP، فاصله بین SNP مجاور و فراوانی آلل SNP را استفاده می‌کند. علاوه بر این، PennCNV قادر به در نظر گرفتن اطلاعات شجره‌نامه (والدین- فرزند) برای بهبود نرخ خوانش و دقت پیش‌بینی نقاط انفصال و همچنین استنباط SNP مخصوص در CNVها است. در مقابل، الگوریتم QuantiSNP از یک رویکرد Bayes Objective برای بدست آوردن تعداد حالات کپی بر اساس نسبت LRR و BAF برای هر SNP استفاده می‌کند. الگوریتم PennCNV با استفاده از یک ماتریس انتقال، حالات واقعی کپی‌ها را بین پروب‌های SNP مدل می‌کند، در حالی که QuantiSNP احتمال‌های بیزی را برای هر یک از جفت SNPها محاسبه می‌کند و سپس با استفاده از SNPها برای خوانش CNVها از HMM استفاده می‌کند (Colella *et al*., 2007; Wang *et al*., 2007). در گوسفند مطالعات متعددی درباره CNV منتشر شده است؛ به عنوان مثال، Fontanesi *et al*. (2011) و Hou *et al*.(2011) بر اساس aCGH؛ Liu *et al*. (2013) با استفاده از ارایه SNP 50k نقشه‌های CNV را بدست آوردند، در حالی که Jenkins *et al*. (2016) از رویکرد چندگانه (مقایسه حد واسطی بین آرایه SNP و توالی‌یابی کل ژنوم) استفاده کردند. همچنین، مطالعات مختلفی در بررسی CNV با الگوریتم‌های مختلف انجام شده است؛ به عنوان مثال Yan *et al*. (2017) با استفاده از الگوریتم‌های SVS، PennCNV و CNVpartion و داده‌های ژنوتیپی آرایه SNP50k، CNV را در نژادهای مختلف گوسفند تجزیه کردند و بیان داشتند که تفاوت-های زیادی در تعداد و اندازه CNVR بین الگوریتم‌ها وجود دارد. در تجزیه تغییرات تعداد کپی در ژنوم گوسفند با استفاده از آرایه SNP BeadChip 50k به وسیله Liu *et al*. (2013) از دو الگوریتم PennCNV و CNVpartion استفاده و گزارش شد که تفاوت دو ابزار تشخیص و ادغام الگوریتم‌های متعدد می‌تواند تشخیص تغییرات ساختاری ژنوم گوسفند را ارتقاء بخشد. بنابراین این پژوهش با هدف استفاده از دو الگوریتم خوانش CNV شامل PennCNV و QuantiSNP برای تشخیص CNV در ژنوم گوسفند بلوچی بر اساس آرایه Ovine SNP50k و مقایسه آن دو با هم و سایر مطالعات مشابه انجام شد.

مواد و روش‌ها

نمونه‌های مورد مطالعه: در این تحقیق از داده‌های ژنومی مربوط به ۹۶ گوسفند نژاد بلوچی واقع در ایستگاه پرورش گوسفند عباس آباد مشهد و تعیین ژنوتیپ شده با استفاده از آرایه‌های Illumina OvineSNP50K Beadchip (دارای بیش از ۵۴۲۴۱ SNP) استفاده شد.

کنترل کیفیت: مطابق با سایر مطالعه‌ها در فراخوانی CNVها (Hou *et al*., 2011; Yang *et al*., 2018; Di Gerlando *et al*., 2019)، در ابتدا SNPهای مرتبط با کروموزوم X و بدون مکان کروموزومی مشخص از ادامه تجزیه حذف شدند. الگوریتم‌هایی که برای شناسایی CNVها استفاده می‌شوند به طور نرمال تعداد دو کپی را برای هر SNP فرض می‌کنند که این فرض برای نواحی اتوزمومی کاذب و دو برابر شدن‌های قطعه‌ای (Segmental Duplication) کروموزوم X محتمل نیست. همچنین توالی کروموزوم X و منتسب کردن آن بین دو اسمبلی ژنوم گوسفند به صورت چشمگیر متفاوت است که سبب غیرقابل اعتماد شدن CNVهای فراخوانی شده روی کروموزوم X شده و از ادامه تجزیه و تحلیل‌ها حذف می-شوند (Hou *et al*., 2011). برای بدست آوردن نتایج قابل اعتماد و دقیق برای تشخیص CNV، کنترل کیفیت در دو مرحله شامل SNP تعیین ژنوتیپ شده و CNVهای خوانش شده انجام شد. در مرحله اول، کنترل کیفیت با استفاده از نرم‌افزار Plink v1.07 (Purcell *et al*., 2007) اجرا شد. کنترل کیفیت SNPها به وسیله آستانه‌هایی اعمال می‌شود تا SNPهای که در محدوده متغیرهای کنترل کیفیت از جمله تعادل هاردی-وینبرگ، نسبت خوانش SNP و فراوانی آللی جزئی (MAF) قرار دارند، انتخاب شوند. دلیل این امر این است که انحراف شدید از تعادل هاردی-وینبرگ برای شناسایی خطای بزرگ تعیین ژنوتیپ مورد استفاده قرار می‌گیرد (Teo *et al*., 2007). وجود مقادیر بالای نسبت SNP از دست رفته نشان‌دهنده عملکرد ضعیف پروب و دقت پایین تعیین ژنوتیپ است. SNPهایی که دارای MAF پایین هستند بیشتر مستعد

خطا هستند، زیرا نمونه‌های کمتری در یک خوشه ژنوتیپی قرار می‌گیرند و بیشتر الگوریتم‌های فراخوانی مبتنی بر خوشه‌ای با آلل‌های نادر عملکرد خوبی ندارند (Pongpanich *et al*., 2010). به طور کلی برای مطالعات ژنومی، آستانه‌هایی تعیین شده است (Craddock and Jones, 2007) که در سایر مطالعات به میزان اندکی تغییر می‌کنند. در مطالعه حاضر، آستانه حذف هر یک از نمونه‌ها و SNPها، به صورت زیر تعیین شد: MAF<0.01، Animal call rate <0.99، SNPs call rate<0.95 و تعادل هاردی-واینبرگ <0.000001.

خوانش CNVها با PennCNV: خوانش CNVها با استفاده از داده‌های شدت سیگنال SNP انجام شد. همچنین نسبت شدت سیگنال طبیعی مشاهده شده نمونه آزمایشی به شدت مورد انتظار هر جایگاه (LRR) و فراوانی آلل B - مقدار نرمال شده نسبت شدت سیگنال نسبی آلل‌های B و A- (BAF) هر SNP برای هر نمونه با استفاده از GenomeStudio گزارش شد. با توجه به نحوه تعیین ژنوتیپ SNPها، از شدت سیگنال ساطع شده آلل A و B (مقادیر X وY) برای تهیه فایل‌های LRR و BAF استفاده شد. به این ترتیب که LRR کل شدت سیگنال نرمال شده برای هر دو آلل SNP است که در زمان حذف کاهش و در زمان اضافه شدن، افزایش می‌یابد. نحوه محاسبه آن به این صورت است: R = Y + X، $LRR = Log_2 (R_{observed}/ R_{expected})$ که در آن $R_{expected}$ از راه درون‌یابی خطی دسته‌های ژنتیکی SNP نمونه‌ها بدست می‌آید. همچنین BAF نرمال شده، نسبت شدت سیگنال نسبی آلل A و B است که از راه محاسبه $\theta=arc\text{-}\tan(Y/X)/(\pi/2)$ بدست آمد (Peiffer *et al*., 2006). در نهایت، فایل داده شدت سیگنال حاوی نام SNP، شماره کروموزوم، موقعیت SNP، BAF و LRR آماده‌سازی شد. فراوانی جمعیت آلل B (PFB) بر اساس BAF هر SNP در جمعیت برای نمونه‌ها با برهان compile_pfb.pl در الگوریتم PennCNV تولید شد. محل و موقعیت SNPها روی کروموزم از فایل PFB اخذ می‌شود، لذا در صورت تغییر در اسمبلی ژنوم، تعیین محل SNP به راحتی امکان‌پذیر است. تجزیه CNV با نرم‌افزارهای (Wang *et al*., 2007) PennCNV 1.0.4 و (Wang *et al*., 2007) QuantiSNP 2.0 انجام شد. برای تجزیه موازی، فایل‌های شدت به فایل‌های انفرادی با استفاده از گزینه "split-" در بسته PennCNV تبدیل شد. تجزیه مناطق اتوزومال برای CNV در دو الگوریتم مختلف انجام شد. الگوریتم خوانش انفرادی با استفاده از گزینه 'test-' انجام شد. PennCNV شامل یک برهان 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 (Diskin *et al*., 2008). امواج ژنومی ممکن است به دلایل متعددی از جمله کیفیت پایین نمونه‌های DNA، هیبریداسیون ضعیف DNA به آرایه، نوسانات در غلظت DNA، محتوای GC متفاوت، عملکرد ضعیف پروب‌های تعیین ژنوتیپ و همچنین مخلوط نمونه‌ها با آلودگی ایجاد می‌شود. این پدیده می‌تواند بر خوانش ژنوتیپ‌ها و در آینده بر صحت تعیین تنوع ساختار کروموزومی نیز اثرگذار باشد. بنابراین باید برای این امر تصحیح انجام شود (Diskin *et al*., 2008). در این مطالعه، برای تنظیم امواج ژنومی از گزینه "-gcmodel" و فایل GC (oviAri1gc5Base.txt) استفاده شد. سپس فایل مدل GC به وسیله محاسبه محتوای GC از یک مگا جفت باز منطقه ژنومی اطراف هر SNP (۵۰۰ کیلو جفت باز از هر طرف) بدست آمد. فایل مدل پنهان مارکوف (HMM)، مدل HMM را برای نرم‌افزار PennCNV فراهم می‌کند و مقدار شدت سیگنال مورد انتظار برای تنوع‌های مختلف را به الگوریتم القاء می‌کند. با توجه به اینکه مدل HMM برای همه ژنوم‌ها تقریباً یکسان است از فایل hh550.hmm، برگرفته از چیپ انسانی ۵۵۰k، استفاده شد. به این ترتیب CNV قطعه ژنومی با استفاده از مدل زنجیره مارکوف در نرم‌افزار PennCNV محاسبه شد (Wang *et al*., 2007). نمونه‌های با انحراف استاندارد LRR کمتر از ۰/۳، BAF برابر با حداقل ۰/۰۱ و مقدار فاکتور امواج بین ۰/۰۵ تا ۰/۰۵- انتخاب شدند.

خوانش CNVها با QuantiSNP. برای تجزیه CNV از QuantiSNP 2.0 - یک الگوریتم مبتنی بر مدل(HMM) Object Bayes Hidden-Markov- استفاده شد. برای هر تغییر تعداد کپی، نمره مشخصی به آن اختصاص داده می‌شود. این معیار، فاکتور Log Bayes نام دارد و مقداری است که نشان‌دهنده پشتیبانی از وجود تغییر تعداد کپی در محل مشخص شده به وسیله SNP است (Colella *et al*., 2007). آستانه دقیق برای فاکتور Log Bayes، حداقل ۳۰، برای بدست آوردن نرخ مثبت کاذب پایین (کمتر از

یک درصد) توصیه می‌شود. آستانه بین ۱۰ تا ۳۰ سبب افزایش قدرت تشخیص تغییرات کوچکتر تعداد کپی، اما افزایش نرخ مثبت‌های کاذب (تا ۱۰ درصد) می‌شود. آستانه ۱۰ برای فاکتور Log Bayes تعیین شد و تغییرات تعداد کپی شناسایی شده با فاکتورLog Bayes کمتر از ۱۰ به طور کلی ناچیز بوده و توصیه می‌شود که حذف شوند (Colella *et al*., 2007; Lin *et al*., 2011).

تعیین نواحی CNVها (CNVRs): پس از تشخیص CNV، مناطق تغییرات تعداد کپی (CNVRs) با ادغام CNVهای همپوشان انفرادی با استفاده از برنامه CNVRuler (Kim *et al*., 2012) تعیین شد. این فرایند ساده ممکن است زمانی که طول CNVهای همپوشان بسیار بزرگ باشد؛ اندازه CNVRها را بیش از حد تعیین کند. به منظور به حداقل رساندن این احتمال، CNVRuler گزینه‌ای را برای ارزیابی باز با باز تراکم منطقه‌ای CNV شرکت‌کننده و مرتب کردن مناطق کم تراکم فراهم می‌کند. لذا مناطقی از ژنوم با تراکم کمتر از ۱۰ درصد حذف می‌شوند (recurrence trim 0.1). این معیار، CNVR را بر اساس فراوانی آنها تشکیل می‌دهد تا از پیش‌بینی‌های مثبت کاذب جلوگیری و محدوده‌های قوی‌تر از مناطق شروع و پایان را تعریف کند. این برنامه، CNVR را با ادغام CNVهایی که حداقل یک جفت باز همپوشانی دارند؛ تولید می‌کند. در نهایت CNVR تحت عنوان «loss» (CNVR حاوی حذف)، «gain» (CNVR حاوی اضافه-شدگی) و «mixed» (CNVR شامل حذف و اضافه) طبقه‌بندی می‌شود. CNVR مشترک بین هر دو الگوریتم با همپوشانی متقابل تعیین شد.

نتایج و بحث

بعد از کنترل کیفیت دقیق SNPها از مجموع ۹۶ گوسفند نژاد بلوچی، ۸۷ نمونه برای آنالیز CNVها باقی ماند. شناسایی CNVها بر اساس داده‌های Illumina Ovine SNP50 beadchip و با استفاده از نرم‌افزارهای PennCNV و QuantiSNP انجام شد. جدول ۱ آمار کلی خصوصیات CNVهای شناسایی شده را نشان می‌دهد. با استفاده از الگوریتم QuantiSNP تعداد کل CNV ۲۹۷۶ شناسایی شد که از این تعداد، ۹۶/۲۳ درصد تنوع‌ها یک جفت بازی بوده و از ادامه تجزیه حذف شدند و در نهایت تعداد ۹۱۶ CNV (۱۰/۵۲ تنوع به ازای هر نمونه) با میانگین و میانه به ترتیب برابر با ۹۸/۱۳ و ۶۶/۰۵ کیلو جفت باز برای تجزیه‌های بعدی باقی ماند. همچنین تعداد ۷۸۹ CNV حذف و ۱۲۷ اضافه شناسایی شد. نتایج QuantiSNP بیشترین تعداد CNV را روی کروموزم‌های یک، شش و سه با فراوانی‌های به ترتیب ۹/۲۵، ۱۷/۳ و ۷/۴۳ درصد شناسایی کرد (شکل ۱).

طول CNVها از ۵۷۰۱ جفت باز تا ۱۷۷۵۸۶۸ جفت باز متغیر بود. برای بررسی توزیع اندازه طول CNVR، آن‌ها به شش گروه تقسیم شدند: کمتر از ۱۰ کیلو جفت باز، ۱۰ تا ۵۰ کیلو جفت باز، ۵۰ تا ۱۰۰ کیلو جفت باز، ۱۰۰ تا ۵۰۰ کیلو جفت باز، ۵۰۰ تا ۱۰۰۰ کیلو جفت باز و بیشتر از یک مگا جفت باز. توزیع فراوانی گروه‌های مختلف اندازه طول CNVR تفاوت بین الگوریتم‌ها را نشان می‌دهد. جدول ۲ مشخصات CNVRها در دو الگوریتم PennCNV و QuantiSNP و جداول ۳ و ۴ میزان پوشش کروموزومی CNVRها را در کروموزوم‌های اوتوزومی با استفاده از این دو الگوریتم نشان می‌دهند. با تقسیم طول کل CNVR شناسایی شده در هر کروموزم بر طول کروموزوم، این میزان پوشش بدست می‌آید. بیشترین میزان پوشش با استفاده از الگوریتم QuantiSNP در کروموزم‌های ۱۷، ۲۵ و ۶ شناسایی شد.

بیشتر CNVها برای الگوریتم QuantiSNP در دامنه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز (۵۹/۰۳ درصد) و ۵۰ تا ۱۰۰ کیلو جفت باز (۲۴/۸۹ درصد) قرار داشتند (جدول ۱). در مجموع ۳۱۶ CNVR اتوزومی به طول ۳۸/۷۶ مگا جفت باز شناسایی شد که در برگیرنده ۱/۴۵ درصد از توالی ژنوم اتوزومی گوسفند و ۱/۳۳ درصد از کل ژنوم گوسفند است. دامنه طولی CNVRها از ۵/۷ کیلو جفت باز تا ۱/۲۸ مگا جفت باز با میانگین ۱۲۲/۶۷ کیلو جفت باز و میانه ۸۰/۳۵ کیلو جفت باز مشخص شدند. تعداد حذف و اضافه‌ها به ترتیب ۲۶۷ و ۴۹ بود. بنابراین تعداد حذف‌ها ۵/۴۴ برابر اضافه‌ها بدست آمد (جدول ۲). بیشتر CNVRها (۸۲ درصد) در دامنه ۵۰ تا ۵۰۰ کیلو جفت باز قرار داشتند (جدول ۱). با استفاده از الگوریتم PennCNV در مجموع تعداد ۲۰۱ CNV با میانگین طول ۱۱۹/۸۴ کیلو جفت باز و میانه ۹۵/۶۴ کیلو جفت باز شناسایی شد (جدول ۱).

جدول ۱- خصوصیات CNVهای شناسایی شده به وسیله دو الگوریتم QuantiSNP و PennCNV در ژنوم گوسفند بلوچی

Table 1. Characteristics of CNVs detected by two QuantiSNP and PennCNV algorithms in Baluchi sheep genome

Measurement	QuantiSNP	PennCNV
Total CNVs	916	201
Total CNVs length (Mb)	89.89	24.08
Average CNVs length (Kb)	98.13	119.84
Median CNVs length (Kb)	66.05	98.64
Loss	789	160
Gain	127	41
CNVs length range		
<10 kb	0.39	-
10-50 kb	8.79	1.33
50-100 kb	24.89	29.63
100-500 kb	59.03	66.92
500-100 kb	0.55	2.12
>1Mb	3.4	-

جدول ۲- مشخصات CNVRها در دو الگوریتم PennCNV و QuantiSN.

Table 2. Characteristics of CNVRs in two algorithms of PennCNV and QuantiSNP

Measurement	PennCNV	QuantiSNP
Total CNVRs	91	316
Total length (Mb)	12.21	38.77
Average length (Kb)	134.18	122.67
Median	110.82	80.35
Percentage of genome covered by CNVRs (%)	0.46	1.45
Loss	66	267
Gain	23	49
Mixed	2	-
CNVR ranges (Kb)	18.75-517.7	5.7- 1280

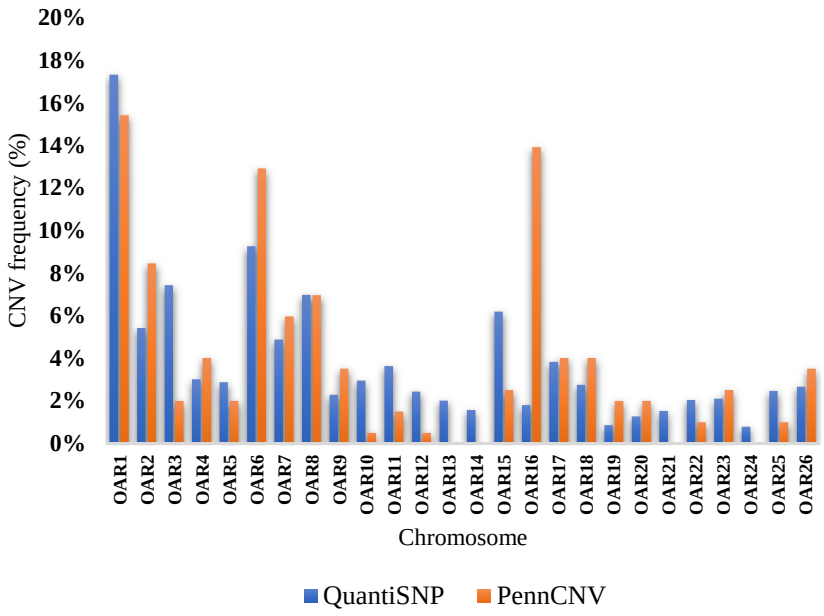

Fig. 1. Chromosomal distribution of detected CNVs using two algorithms of PennCNV and QuantiSNP

شکل ۱- توزیع کروموزمی CNVهای شناسایی شده به وسیله دو الگوریتم PennCNV و QuantiSNP

جدول ۳- مشخصات کروموزم‌های اُتوزوم گوسفند به لحاظ پوشش CNVR قطعه ژنومی با استفاده از الگوریتم PennCNV

Table 3. The sheep autosomal chromosome characterization for CNVR coverage of genome segment using PennCNV algorithm

Chromosome	*Length (Mb)	CNVR length (Mb)	CNVR count	Chromosome coverage (%)	Chromosome	Length (Mb)	CNVR length (Mb)	CNVR count	Chromosome coverage (%)
OAR1	301.31	1.7	13	0.56	OAR12	84.52	0.07	1	0.08
OAR2	265.69	1.38	12	0.52	OAR15	90.32	0.43	4	0.47
OAR3	241.14	0.39	3	0.16	OAR16	78.35	0.37	4	0.47
OAR4	130.07	0.50	5	0.38	OAR17	82.58	0.44	5	0.54
OAR5	117.63	0.27	2	0.23	OAR18	70.86	0.57	3	0.76
OAR6	129.79	1.69	11	1.30	OAR19	62.71	0.14	2	0.23
OAR7	107.7	1.13	5	1.05	OAR20	55.94	0.15	2	0.27
OAR8	98.77	0.78	5	0.79	OAR22	55.67	0.32	2	0.58
OAR9	104.71	0.31	2	0.30	OAR25	47.66	0.36	2	0.76
OAR10	97.21	0.14	1	0.15	OAR26	49.39	0.88	4	1.79
OAR11	60.98	0.20	3	0.34					

* The chromosomes length was obtained from NCBI site.

جدول ۴- مشخصات کروموزم‌های اُتوزوم گوسفند به لحاظ پوشش CNVR قطعه ژنومی با استفاده از الگوریتم QuantiSNP

Table 4. The sheep autosomal chromosome characterization for CNVR coverage of genome using QuantiSNP algorithm

Chromosome	*Length (Mb)	CNVR length (Mb)	CNVR count	Chromosome coverage (%)	Chromosome	Length (Mb)	CNVR length (Mb)	CNVR count	Chromosome coverage (%)
OAR1	301.31	5.76	40	1.91	OAR13	87.26	1.04	10	1.19
OAR2	265.69	2.98	27	1.12	OAR14	71.11	0.85	6	1.20
OAR3	241.14	2.93	25	1.21	OAR15	90.32	1.73	14	1.91
OAR4	130.07	1.16	14	0.89	OAR16	78.35	0.73	7	0.93
OAR5	117.63	1.18	15	1.00	OAR17	82.58	2.44	12	2.95
OAR6	129.79	2.67	19	2.06	OAR18	70.86	1.15	12	1.62
OAR7	107.7	1.17	13	1.09	OAR19	62.71	0.54	5	0.86
OAR8	98.77	1.44	14	1.46	OAR20	55.94	0.79	8	1.41
OAR9	104.71	1.70	13	1.63	OAR21	52.95	0.43	7	0.81
OAR10	97.21	2.13	13	2.19	OAR22	55.67	1.37	8	2.46
OAR11	60.98	1.00	7	1.64	OAR25	47.66	1.14	6	2.39
OAR12	84.52	1.65	14	1.95	OAR26	49.39	0.80	7	1.62

* The chromosomes length was obtained from NCBI site.

با توجه به توزیع CNVها روی کروموزوم، نتایج الگوریتم PennCNV بیشترین CNV را روی کروموزم های ۱، ۱۶ و ۶ به ترتیب با درصد توزیع ۱۵/۴۲، ۱۳/۹۳ و ۱۲/۹۴ مشخص کرد. روی کروموزوم‌های ۱۳، ۱۴، ۲۱ و ۲۴، هیچ CNV شناسایی نشد (شکل ۱). متوسط تعداد CNV در هر نمونه ۲/۳۱ و در هر کروموزوم ۸/۷۳ بدست آمد. ۶۶/۹۲ درصد از CNV در بازه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز قرار دارد. همچنین تنوع‌های کمتر از ۱۰ کیلو جفت باز و بیشتر از یک مگا جفت باز شناسایی نشد (جدول ۱). بعد از تجمیع نواحی همپوشان با نرم‌افزار CNVRuler، تعداد ۹۱ CNVR ژنومی در کروموزوم‌های اتوزوم شناسایی شد (جدول ۲). این تعداد تنوع، ۱۲/۲۱ مگا جفت باز از ژنوم معادل ۰/۴۶ درصد از کل کروموزم‌های اتوزوم گوسفند را با میانگین ۱۳۴/۱۸ و میانه ۱۱۰/۸۲ کیلو جفت باز به خود اختصاص داد. طول CNVR قطعه ژنومی از ۱۸/۷۵ تا ۵۱۱/۷ کیلو جفت باز متغیر بود. ۷۰/۵ درصد از CNVR ژنومی در دامنه طولی ۱۰۰ تا ۵۰۰ کیلو جفت باز قرار داشت (جدول ۱). مقایسه بین دو الگوریتم PennCNV و QuantiSNP نشان داد که تعداد CNVهای بیشتری با QuantiSNP شناسایی شد، به طوری‌که تعداد CNVهای شناسایی شده حدود چهار برابر بیشتر بود. همچنین هر دو الگوریتم، CNVهای مشابهی روی کروموزوم‌ها شناسایی کردند، به طوری‌که ۸۶/۶ درصد از CNVهای شناسایی شده به وسیله الگوریتم PennCNV با CNVهای شناسایی شده به وسیله الگوریتم QuantiSNP مطابقت داشت و تنها ۱۳/۳درصد (۲۷ عدد) از CNVهای شناسایی شده با الگوریتم PennCNV منحصر به فرد بودند. همچنین تعداد ۷۴۲ CNV معادل ۸۱ درصد از کل CNVهای شناسایی شده به وسیله الگوریتم QuantiSNP منحصر به فرد بودند که با الگوریتم PennCNV خوانده نشدند. در کل، تعداد ۷۶۹ CNV منحصر به فرد و تعداد ۱۷۴ CNV مشترک به وسیله دو الگوریتم شناسایی شد. متوسط طول CNVهای مشترک برابر با ۱۲۲/۶۷ کیلو جفت باز و در دامنه ۲۵۷۱۰ تا ۵۱۱۱۷۰۵ جفت باز و با میانه برابر با ۹۷/۵۲ کیلو جفت باز بود. در این مطالعه، نتایج اولیه بدست آمده از خوانش CNVها در تعداد و طول CNVهای بدست آمده با دو الگوریتم متفاوت بود که نشان می‌دهد توانایی دو الگوریتم در این موارد متفاوت است. این نتیجه در مطالعات دیگر که از چندین الگوریتم برای خوانش CNVها استفاده کرده‌اند، مشاهده شده است (Liu *et al*., 2013; Metzger *et al*., 2013; Xu *et al*., 2013; Salomón *et al*., 2015; Yan *et al*., 2017). در بررسی CNV در ژنوم اسب با استفاده از سه الگوریتم CNVpartion، PennCNV و QuantiSNP، به ترتیب ۱۶۶، ۸۶۰ و ۱۰۹۰ با میانگین ۳۵۰/۷، ۳۱۲/۰۵ و ۸۶۹/۴۷ کیلو جفت باز شناسایی شد (Metzger *et al*., 2013). همچنین نویسندگان دیگری در بررسی CNV در ژنوم گوسفندان چینی با الگوریتم‌های PennCNV، CNVpartion و SVS به ترتیب ۴۱۹، ۱۰۴ و ۷۲۵ CNVR را شناسایی کردند (Yan *et al*., 2017). همچنین یک مطالعه انسانی با داده‌های CNV مبتنی بر CGH مربوط به پروژه HapMap به وسیله چهار الگوریتم خوانش CNV شامل Birdsuite، dChip، GTC و PennCNV به ترتیب تعداد ۹۵۱، ۶۳۹، ۲۰۵ و ۵۶۴ شناسایی شد. میانگین طول CNVها به ترتیب ۱۲۷/۲۳۵، ۱۷۸/۹۲۰، ۳۱۶/۹۳۲ و ۱۵۲/۶۳۴ بود (Zhang *et al*., 2014). این امر نشان‌دهنده فقدان یک استاندارد کلی برای برآورد دقت الگوریتم‌های تشخیص CNV است. از این‌رو نویسندگان این تحقیقات بیان داشته‌اند که نمی‌توان بیان کرد که کدام الگوریتم قابل اعتمادتر است. برای درک بهتر از مقایسه الگوریتم‌ها، Zhu *et al*. (2013) در مقاله منتشر شده خود، حداقل ۱۰ مقایسه از نقاط قوت و ضعف ابزارهای خوانش CNV با استفاده از داده CNV انسانی را بررسی کردند. در این مقایسه‌ها، یک موضوع کلی مورد توافق است. اولین مورد، عدم وجود یک روش استاندارد برای جمع‌آوری داده‌ها و فقدان نمونه‌های مرجع استاندارد است. این امر مقایسه نتایج CNV را در مطالعات مختلف دشوار می‌کند. دوم اینکه، نتایج CNV نیز بسته به روش‌های تشخیص CNV بسیار متفاوت است. به عنوان مثال، به عنوان جامع‌ترین مطالعه در مورد این موضوع، Pinto *et al*. (2011) به طور سیستماتیک تشخیص CNV در ۱۱ پلت‌فرم ریزآرایه را برای ارزیابی کیفیت داده‌ها و خوانش CNV، تکرارپذیری، تطابق‌پذیری پلت‌فرم‌ها و تغییرپذیری ابزار تجزیه و تحلیل مقایسه کردند. نکته عجیب آن بود که تکرارپذیری در آزمایش‌های تکرار شده برای بیشتر پلت‌فرم‌ها کمتر از هفتاد درصد بود و ابزارهای مختلف تحلیلی که برای داده‌های خام یکسان استفاده

می‌شوند، معمولاً خوانش‌های CNV را با مطابقت کمتر از پنجاه درصد نشان دادند. نویسندگان این مشاهدات، ضریب تکرارپذیری را به این حقایق نسبت داده‌اند: (۱) CNVهای بزرگ در بیشتر موارد با دو برابر شدن قطعه‌ها در مناطق ژنومی پیچیده همپوشانی دارند، (۲) CNVهای بزرگ همچنین منجر به خوانش‌های تکه‌ای می‌شوند (یک CNV منفرد به عنوان چندین نوع کوچک‌تر شناسایی می‌شود). این امر نویسندگان را به این نتیجه رسانده است که تفاوت‌های چشمگیر بین خوانش‌های CNV از پلت-فرم‌های مختلف و ابزارهای تحلیلی، اهمیت ارزیابی دقیق طراحی آزمایش را در مطالعات کشف و ارتباط دقیق داده‌ها و فیلتر کردن در تشخیص برجسته می‌کند (Xu *et al*., 2013). لذا، بهترین انتخاب ممکن، استفاده از الگوریتم‌های متعدد برای تشخیص CNV همپوشان، به جای استفاده از یک الگوریتم واحد است. در این مطالعه، میانگین و میانه CNVهای بدست آمده از الگوریتم PennCNV بزرگتر از الگوریتم QuantiSNP شناسایی شد که مشابه نتایج Kim *et al*. (2012) بود که گزارش کردند در یک مطالعه انسانی میانگین و میانه CNVهای شناسایی شده برای الگوریتم PennCNV؛ ۴۴/۶۸ و ۱۲/۹۸ کیلو جفت باز و برای الگوریتم QuantiSNP؛ ۳۹/۹۱ و ۷/۶ کیلو جفت باز است. همچنین نتیجه‌گیری شد که اندازه CNVهایی که با دو یا تعداد بیشتر الگوریتم تشخیص داده می‌شود عموماً بزرگتر از CNVهایی هستند که با یک روش شناسایی شده‌اند (Kim *et al*., 2012). تعداد CNVRهای گزارش شده به وسیله الگوریتم‌های PennCNV و QuantiSNP به ترتیب ۹۱ و ۳۱۶ با میانگین طول ۱۳۴/۱۸ و ۱۲۲/۶۷ کیلو جفت باز و معادل حدود ۰/۴۶ درصد و ۱/۳۳ درصد از کل ژنوم گوسفند بدست آمد. در مطالعات دیگر که روی گوسفند انجام شده است میانگین طول CNVRها از ۱۹ کیلو جفت باز تا ۲۵۴ کیلو جفت باز با پوشش ژنومی ۰/۰۶ تا ۴/۸ درصد متغیر بود (Hou *et al*., 2015; Ma *et al*., 2015; Jenkins *et al*., 2016; Zhu *et al*., 2016). اختلاف در این مقادیر می‌تواند به نوع پلت‌فرم و الگوریتم مورد استفاده، زمینه‌های ژنتیکی متفاوت و تعداد نمونه مرتبط باشد، زیرا هر چه تعداد نمونه و SNPهای مورد استفاده برای شناسایی CNVها بیشتر باشد در مجموع تعداد CNVهای شناسایی شده نیز بیشتر خواهد بود.

تعداد کمتر CNVRهایی که با الگوریتم PennCNV شناسایی شده بودند، مشابه نتایج سایر مطالعات CNV در انسان و پستانداران دیگر بود (Jiang *et al*., 2012; Zhu *et al*., 2016). در این مطالعه، تعداد CNV بدست آمده از الگوریتم‌های PennCNV و QuantiSNP به ازای هر رأس گوسفند به ترتیب ۲/۳۱ و ۱۰/۵۲ بود. در مطالعات دیگر؛ تعداد CNV به ازای هر رأس بـز (۱۷/۹) (Fontanesi *et al*., 2010)، گوسـفند (۱۶/۹) (Fontanesi *et al*., 2010; Wang *et al*., 2010; Fontanesi *et al*., 2011; Liu *et al*., 2013; Metzger *et al*., 2013; Karimi *et al*., 2018)، جوجـه گوشـــتی (۹/۶) (Wang *et al*., 2010)، اسب (۰/۱) (Metzger *et al*., 2013) و گاو بومی ایران (۱۹/۱۶) (Esmailizadeh *et al*., 2018) شناسایی شد. اخـتلاف در نتـایج مطالعـات مختلـف می‌تواند دلایل متعددی داشته باشد که از آن جمله می‌توان به تعـــداد نمونه، روش شناسایی، نرم‌افزار مورد استفاده، گونـه دام و تـراکم SNP مـورد اسـتفاده بـرای شناسـایی و اسمبلی ژنوم اشاره کرد (Metzger *et al*., 2013; Salomón *et al*., 2015). همچنین با افزایش تعداد نمونه، تعداد CNVهای شناسایی شده در سطح ژنوم افزایش می‌یابد (Bae *et al*., 2010). بیشتر CNV و CNVRهای مشاهده شده در هر دو نرم‌افزار از نوع حذف (loss) بودند. در حالی‌که وقایع اضافه (gain) به تعداد کمتری مشاهده شد، به طوری‌که تعداد CNVهای نوع حذف در نرم‌افزار QuantiSNP حدود پنج برابر و در نرم‌افزار PennCNV حدود سه برابر بیشتر از تعداد اضافه‌ها بود. بالاتر بودن تعداد حذف‌ها نسبت به اضافه‌ها در مطالعات دیگر نیز مشاهده شده است (Liu *et al*., 2013; Xu *et al*., 2013; Ma *et al*., 2015; Zhu *et al*., 2016; Yan *et al*., 2017). از دلایل این امر می‌توان به موارد زیر اشاره کرد: ۱) انحراف آرایه‌های SNP (Yan *et al*., 2017)؛ از آن‌جا که تراشه‌های SNP برای تعیین ژنوتیپ نشانگرهای SNP طراحی شده‌اند، برخی از نویزهای پس‌زمینه که روی خوانش SNP تاثیر نمی‌گذارند، ممکن است مشکلی برای الگوریتم‌های خوانش CNV به وجود آورد. برای این منظور، داده‌های SNP به طور معمول در برابر یک جمعیت مرجع به منظور کاهش تغییرات بین آرایه و آثار هیبریداسیون خاص پروب، نرمال می‌شوند. این فرضیه که بیشتر نمونه‌های رفرنس دارای دو کپی مشابه هستند برای نواحی CNV

مشترک برقرار نیست. بنابراین نرمال‌سازی در این مناطق برای بدست آوردن پارامترهای صحیح، بیشتر انجام می‌شود (Xu *et al.*, 2013)، ۲) عوامل زیستی؛ نوترکیبی همولوگوس غیر آللی بیشتر تمایل به ایجاد حذف دارد تا اضافه (Clop *et al.*, 2012). حذف قطعات CNV در یک فرد می‌تواند منجر به از دست رفتن تعدادی ژن شود، اما اگر یک ژن یک کپی را از دست بدهد، کپی دیگر ممکن است تولید پروتئین کد شده به وسیله ژن را جبران کند. همچنین بسیاری از مطالعات تأیید کرده‌اند که آثار زیان‌آور حذف‌ها بیشتر از آثار زیان‌آور اضافه‌ها است (Yeo *et al.*, 2011; Liu *et al.*, 2012; Liu *et al.*, 2013). با این حال، افزایش CNV ممکن است مقدار محصول ژن را به طور چشمگیری افزایش دهد که به طور بالقوه منجر به ناهنجاری‌های رشد و یا تغییر عملکرد می‌شود. به عنوان مثال، در انسان، نشان داده شده است که اضافه‌شدگی یک عنصر نظارتی موثر بر بیان پروتئین ۲ مورفوژنیک استخوان، منجر به ایجاد ناهنجاری هیپوپلاستی انگشت‌های دوم و پنجم می‌شود (Dathe *et al.*, 2009). با این حال، چنین اضافه‌شدگی‌های CNV ممکن است با انتخاب طبیعی حذف شودند، ۳) فراخوانی ژنوتیپ‌ها از ریزآرایه؛ به طوری که یک حذف نشان‌دهنده ۵۰ درصد کاهش شدت سیگنال است، در عوض یک اضافه نشان-دهنده افزایش ۳۳ درصدی شدت سیگنال است (Lin *et al.*, 2011). علاوه بر این، فراوانی آلل B- مقدار گزارش شده از ریزآرایه- SNPهای یک منطقه حذف خاص، معمولاً مقدار بین صفر تا ۱۰۰ را در بر می‌گیرد که سبب ایجاد یک الگوی مشخص می‌شود که به راحتی قابل تشخیص است (Lin *et al.*, 2011)، ۴) این نتیجه ممکن است ناشی از این واقعیت باشد که حذف‌ها نسبت به اضافه‌ها به راحتی قابل شناسایی است، زیرا داده‌های شدت با تعداد کپی به صورت خطی ارتباط دارند (Kim *et al.*, 2012). مقایسه نتایج دو الگوریتم در این مطالعه نشان داد که الگوریتم QuantiSNP قابلیت شناسایی CNV نسبتاً کوچک‌تر (کمتر از ۱۰ کیلو جفت باز) و نسبتاً بزرگ‌تر (بیشتر از یک مگا جفت باز) را نسبت به الگوریتم PennCNV افزایش می‌دهد و این نشان می‌دهد که چرا CNVR بیشتری به وسیله الگوریتم QuantiSNP نسبت به الگوریتم PennCNV تشخیص داده شده است؛ البته در الگوریتم QuantiSNP، به منظور افزایش اعتماد کشف CNV، فیلتر کردن دقیق (LBF <30) برای تعریف CNVها در این مورد استفاده شد، اما میزان شناسایی CNVهای کوچکتر در هر دو نسبت به مطالعات دیگر پایین بود. به طور کلی، تراکم پایین آرایه SNP، تشخیص CNVهای کوتاه را دشوارتر از روش توالی‌یابی کل ژنوم می‌کند. بررسی هر دو الگوریتم نشان داد بیشتر CNVها در دامنه طولی ۱۰۰ تا ۵۰۰ جفت باز قرار داشتند. Yan *et al.* (2017) در بررسی خود با سه الگوریتم، بیشترین فراوانیCNVها را با نرم‌افزار PennCNV بین ۱۰۰ تا ۴۰۰ جفت باز، با SVS بین ۱۰۰ تا ۲۰۰ جفت باز و با CNVpartion بین ۱۰۰ تا ۲۰۰ و بیشتر از یک مگا جفت باز گزارش کردند. در مطالعه حاضر، تعداد CNVRهای بدست آمده با استفاده از الگوریتم PennCNV نسبت به سه مطالعه دیگر که از الگوریتم مشابه استفاده کرده بودند، کمتر بود (Liu *et al.*, 2013; Ma *et al.*, 2015; Yan *et al.*, 2017). این نشان می‌دهد که نتایج خوانش CNVها با استفاده از یک الگوریتم در یک نژاد خاص می‌تواند تحت تاثیر اندازه جمعیت و نسخه اسمبلی ژنوم قرار گیرد. در این مطالعه، با استفاده از الگوریتم PennCNV روی کروموزم‌های ۱۳، ۱۴، ۲۱ و ۲۴ تنوع CNV شناسایی نشد. Metzger *et al.* (2013) با استفاده از سه الگوریتم PennCNV، QuantiSNP و CNVpartion روی کروموزم‌های ۱۶، ۱۹، ۲۱، ۲۲، ۲۶، ۲۷ و ۳۱ اسب تنوعی با الگوریتم CNVpartion شناسایی نکردند. Salomón *et al.* (2015) با استفاده از دو الگوریتم PennCNV و QuantiSNP، با توجه به نتایج PennCNV روی کروموزوم ۲۷ گاو هیچ CNV مشاهده نکردند. بررسی دقیق‌تر هر الگوریتم نشان داد که الگوریتم‌های PennCNV و QuantiSNP تقریباً توزیع مشابهی از CNVها را روی کروموزوم‌ها ارائه کرده و تعداد قابل توجهی از رویدادهای تشخیص را در مقایسه با CNVpartion نشان می‌دهد. تجزیه مقایسه‌ای روش‌های تشخیص CNV برای آرایه‌های SNP تایید کرد که الگوریتم‌های PennCNV و QuantiSNP دارای همپوشانی زیادی با وقایع تشخیص هستند (Winchester *et al.*, 2009; Yuan *et al.*, 2017). مطالعه سرطان مثانه در انسان که از طراحی مطالعه مشابه با این تحقیق استفاده کرد، الگوریتم‌های PennCNV و QuantiSNP را به عنوان یک روش قابل اعتماد برای تشخیص CNVs نسبت به

CNVpartion پیشنهاد کرد و حتی الگوریتم QuantiSNP را به عنوان الگوریتم بهتر از این دو روش اعلام نمود. در این مطالعه، ۷۳/۶۳ درصد از CNVهای شناسایی شده به وسیله الگوریتم PennCNV با CNVهای یافته شده به وسیله QuantiSNP همپوشانی داشت. Salomón (2015) *et al.* نیز در تجزیه CNVها در گاو با SNP chip 648 Kb، ۴۹/۶۷ درصد از CNVهای یافته شده به وسیله الگوریتم PennCNV را همپوشان با الگوریتم QuantiSNP گزارش کردند. همچنین تنها ۲۷ CNVR (۶/۶ درصد از کل CNVR) مشترک بین دو الگوریتم یافت شد. در مطالعه Yan *et al.* (2017)، میزان مشترک بودن CNVRها بین سه الگورریتم PennCNV، SVS و CNVpartion، ۶۹ CNVR و معادل با ۵/۶ درصد بود. همچنین در مطالعه Liu *et al.* (2013)، تعداد ۴۷ CNVR بین دو الگوریتم PennCNV و CNVpartion یعنی ۱۹/۵ درصد مشترک بود. منحصر به فرد بودن بیشتر CNVRها برای هر یک از الگوریتم‌ها نشان می‌دهد یک استاندارد برای ارزیابی الگوریتم‌های تشخیصCNV لازم است. یک مسئله که ممکن است تا حدی به تفاوت‌های مشاهده شده بین الگوریتم‌ها کمک کند، مدل CG است. مدل CG معمولاً در مورد الگوریتم‌های PennCNV و QuantiSNP برای تنظیم تفاوت در محتوای GC (مانند waviness) بر اساس ژنوم انسان مورد استفاده قرار می‌گیرد و هنوز برای گونه‌های غیر انسانی بهینه‌سازی نشده است (Diskin *et al.*, 2008). در بررسی محتوای GC در CNVRها مشخص شد که این مقدار از ۳۵/۰۵ به ۶۸/۳۸ درصد در الگوریتم PennCNV و از ۲۴/۳۶ تا ۷۳/۲۹ درصد در الگوریتم QuantiSNP متغیر است که موافق با نتایج Salomón *et al.* (2015) در گاو است. این محققین این مقدار را بین ۳۳/۹ تا ۵۷/۹۵ درصد یافتند که نشان می‌دهد CNVها مناطقی با محتوای بالا GC هستند (Fadista *et al.*, 2010). مقایسه CNVR تشخیص داده شده در این مطالعه با مطالعات دیگر در جدول ۵ ارائه شده است.

بر اساس اطلاعات جدول ۵، مقایسه CNVR تشخیص داده شده در تحقیقات مختلف نشان می‌دهد که تعداد CNVRها (با توجه به اندازه نمونه) با استفاده از aCGH در مطالعه Fontanesi *et al.* (2011) و Hou *et al.* (2015) مشابه هستند. با این حال، Jenkins *et al.* (2016) در مطالعه خود تعداد بسیار زیادی از CNVR (۳۴۸۸ عدد) را فقط در ۳۶ گوسفند تشخیص دادند که می‌تواند به علت تراکم بالای پروب aCGH (۲/۱ میلیون) و روش‌های متعدد تشخیص CNV باشد. روش aCGH به علت توزیع یکنواخت پروب‌ها در سراسر ژنوم، قابلیت تشخیص تنوع‌های ساختاری ژنوم را با قدرت و دقت بیشتری دارد، لذا نواحی تنوع را دقیق‌تر تشخیص می‌دهد. از این‌رو تعداد نواحی تنوع شناسایی شده بیشتر و طول آن کوتاه‌تر است (Liu *et al.*, 2010). علاوه بر این، در مقایسه با روش aCGH، روش آرایه حاوی پروب‌های کمتری برای نواحی هم‌شکل است، لذا فقط تنوع‌های بزرگ را شناسایی می‌کند (Hou *et al.*, 2012). بنابراین با توجه به اینکه نتایج تحقیق حاضر با نتایج CGH سازگاری کمتر دارد می‌توان بیان کرد نتایج آرایه‌های CGH و SNP به راحتی قابل مقایسه نیستند.

نتیجه‌گیری کلی

هدف از مطالعه حاضر، بررسی دو الگوریتم متفاوت برای تشخیص CNVها در گوسفندان نژاد بلوچی و مقایسه آن-ها با یکدیگر بود. مانند سایر مقایسه‌های منتشر شده از روش‌های خوانش CNV، شاهد تغییرات زیادی در خوانش‌های انجام شده به وسیله الگوریتم‌های مختلف بودیم. نتایج این تحقیق همانند بسیاری از نویسندگانی که از چندین الگوریتم فراخوانی CNV استفاده کردند نشان داد که مقایسه ترکیبی از الگوریتم‌های PennCNV و QuantiSNP به جای استفاده از تنها یک الگوریتم برای شناسایی دقیق CNVها موثر است و الگوریتم‌های خوانش فعلی باید برای تجزیه و تحلیل CNV با کارآیی بالا در پویش‌های گسترده ژنومی بهبود یابد. همچنین از نرم-افزارهای استفاده شده برای هر پلت فرم استفاده شود. این نتایج نشان می‌دهد که تجزیه CNV مقایسه‌ای با الگوریتم‌های متعدد، یک رویکرد مفید برای برطرف کردن محدودیت‌های الگوریتم‌های فردی است و به افزایش قابلیت اطمینان نتایج تشخیصی کمک می‌کند. در کل نتایج این مطالعه نشان داد تفاوت‌های زیادی در بخش‌های CNV و همچنین CNVR در بین دو الگوریتم مورد مقایسه وجود دارد و تجزیه تطبیقی الگوریتم‌های تشخیص CNV برای افزایش قدرت تشخیص CNV مفید است، اما دارای محدودیت‌های خاصی است که به ابزار تشخیص وابسته است. لذا برای شناسایی بهتر CNVها

جدول ۵- مقایسه تعداد و اندازه مناطق مختلف تنوع تعدادکپی (CNVR) تشخیص داده شده در این مطالعه با مطالعات قبلی در گوسفند

Table 5. Comparison of the number and size detected (CNVRs) in this study with previous studies in sheep

References	Algorithm	Platform	Assembly	Sample size	CNVR number	Mean size (Kb)	Median (Kb)	Size range (Kb)
Fontanesi *et al*. (2011)		Bovine 385 k aCGH	Btau_v4.0	11	135	77.6	55.9	24.6-505
Liu *et al*. (2013)	PennCNV	OvineSNP50 K	OaiAri1	329	238	253.57	186.92	13.66-1300
Ma *et al*. (2015)	PennCNV	OvineSNP50 K	Oar_v 3.1	160	111	123.84	100.53	-
Zhu *et al*. (2016)	PennCNV	OvineSNP600 K	Oar_v 3.1	11	490	225	-	-
Hou *et al*. (2011)	-	1.4 M aCGH	OaiAri1	5	51	304.86	-	52–2000
Jenkins *et al*. (2016)	-	2.1 M aCGH	UMD3_OA	36	3488	19	-	1–3600
Yan *et al*. (2017)	SVS	OvineSNP50 K	Oar_v3.1	385	749	189	118	15.3-6600
	PennCNV				464	305.5	218.1	11.4-2108.8
	CNVPartion				104	1521.3	395.4	87-12093.7
This study	PennCNV	OvineSNP50 K	OaiAri3.1	87	91	134.18	110.817	18.75-511.7
	QuantiSNP				316	122.67	80.35	5.7-1280

استفاده از آرایه‌های با چگالی بالا و استاندارد طلایی برای اعتبارسنجی CNVها می‌تواند نقش مهم داشته باشد. ناسازگاری‌های زیاد در نتایج حاصل از ابزارهای مختلف فراخوانی CNV، نیاز به استانداردسازی جمع‌آوری داده‌های آرایه، ارزیابی کیفیت و اعتبار سنجی تجربی را برجسته می‌کند. این در مورد گونه‌هایی مانند گوسفند بسیار صادق است، که برای آن‌ها استاندارد طلایی خوانش‌های CNV برای مقایسه داده‌ها وجود ندارد. بنابراین پس از طراحی دقیق و تجربی و فیلتر کردن دقیق داده‌ها، می‌توان آثار CNV بر تغییرات طبیعی فنوتیپی و حساسیت بیماری را به طور کامل نشان داد.

فهرست منابع

Bae J. S., Cheong H. S., Kim L. H., NamGung S., Park T. J., Chun J. Y. and Shin H. D. 2010. Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genomics, 11(1): 232.

Baumbusch L. O., Aarøe J., Johansen F. E., Hicks J., Sun H., Bruhn L. and Børresen-Dale A. L. 2008. Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. BMC Genomics, 9(1): 379.

Bhanuprakash V., Chhotaray S., Pruthviraj D. R., Rawat C., Karthikeyan A. and Panigrahi M. 2018. Copy number variation in livestock: A mini review. Veterinary World, 11(4): 535-541.

Carter N. P. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. Nature Genetics, 39(7s), S16.

Clop A., Vidal O. and Amills M .2012. Copy number variation in the genomes of domestic animals. Animal Genetics, 43(5): 503-517.

Colella S., Yau C., Taylor J. M., Mirza G., Butler H., Clouston P. and Ragoussis J. 2007. QuantiSNP: An Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Research, 35(6): 2013-2025.

Craddock N. J. and Jones I. R. J. N. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447: 661-678.

Dathe K., Kjaer K. W., Brehm A., Meinecke P., Nürnberg P., Neto J. C. and Seemann P. 2009. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. The American Journal of Human Genetics, 84(4): 483-492.

de Lemos M. V. A., Berton M. P., de Camargo G. M. F., Peripolli E., de Oliveira Silva R. M., Olivieri B. F. and Tonhati H. 2018. Copy number variation regions in Nellore cattle: evidences of environment adaptation. Livestock Science, 207: 51-58.

Dermitzakis E. T. and Stranger B. E. 2006. Genetic variation in human gene expression. Mammalian Genome, 17(6): 503-508.

Diskin S. J., Li M., Hou C., Yang S., Glessner J., Hakonarson H. and Wang K. 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Research, 36(19): e126-e126.

Fadista J., Thomsen B., Holm L. E. and Bendixen C. 2010. Copy number variation in the bovine genome. BMC Genomics, 11(1): 284.

Feuk L., Carson, A. R. and Scherer S. W. 2006. Structural variation in the human genome. Nature Reviews Genetics, 7(2): 85.

Fontanesi L., Beretti F., Riggio V., González E. G., Dall'Olio S., Davoli R. and Portolano B. 2009. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. Cytogenetic and Genome Research, 126(4): 333-347.

Fontanesi L., Martelli P. L., Beretti F., Riggio V., Dall'Olio S., Colombo M. and Portolano B. 2010. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. BMC Genomics, 11(1): 639.

Fontanesi L., Beretti F., Martelli P. L., Colombo M., Dall'Olio S., Occidente M. and Russo V. 2011. A first comparative map of copy number variations in the sheep genome. Genomics, 97(3): 158-165.

Freeman J. L., Perry G. H., Feuk L., Redon R., McCarroll S. A., Altshuler D. M. and Carter N. P. 2006. Copy number variation: new insights in genome diversity. Genome Research, 16(8): 949-961.

Giuffra E., Törnsten A., Marklund S., Bongcam-Rudloff E., Chardon P., Kijas J. M. and Andersson L. 2002. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. Mammalian Genome, 13(10): 569-577.

Hastings P. J., Lupski J. R., Rosenberg S. M. and Ira G. 2009. Mechanisms of change in gene copy number. Nature Reviews Genetics, 10(8): 551.

Hester S. D., Reid L., Nowak N., Jones W. D., Parker J. S., Knudtson K. and Denslow N. D. 2009. Comparison of comparative genomic hybridization technologies across microarray platforms. Journal of Biomolecular Techniques, 20(2): 135.

Hou C. L., Meng F. H., Wang W., Wang S. Y., Xing Y. P., Cao J. W. and Zhou H. M. 2015. Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization. Small Ruminant Research, 128: 19-26.

Hou Y., Liu G. E., Bickhart D. M., Cardone M. F., Wang K., Kim E. S. and Sonstegard T. S. 2011. Genomic characteristics of cattle copy number variations. BMC Genomics, 12(1): 127.

Hou Y., Bickhart D. M., Hvinden M. L., Li C., Song J., Boichard D. A. and Sonstegard T. S. 2012. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics, 13(1): 376.

Jenkins G. M., Goddard M. E., Black M. A., Brauning R., Auvray B., Dodds K. G. and McEwan J. C. 2016. Copy number variants in the sheep genome detected using multiple approaches. BMC Genomics, 17(1): 441.

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.

Karimi K., Esmailizadeh A., Wu D. D. and Gondro C. 2018. Mapping of genome-wide copy number variations in the Iranian indigenous cattle using a dense SNP data set. Animal Production Science, 58(7): 1192-1200.

Kim J. H., Hu H. J., Yim S. H., Bae J. S., Kim S. Y. and Chung Y. J. 2012. CNVRuler: a copy number variation-based case–control association analysis tool. Bioinformatics, 28(13): 1790-1792.

Kim S. Y., Kim J. H. and Chung Y. J. 2012. Effect of combining multiple CNV defining algorithms on the reliability of CNV calls from SNP genotyping data. Genomics & Informatics, 10(3): 194.

Korn J. M., Kuruvilla F. G., McCarroll S. A., Wysoker A., Nemesh J., Cawley S. and Lee C. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nature Genetics, 40(10): 1253.

Lai W. R., Johnson M. D., Kucherlapati R. and Park P. J. 2005. Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics, 21(19): 3763-3770.

Lin P., Hartz S. M., Wang J. C., Krueger R. F., Foroud T. M., Edenberg H. J. and Webb B. T. 2011. Copy number variation accuracy in genome-wide association studies. Human Heredity, 71(3): 141-147.

Liu G. E., Hou Y., Zhu B., Cardone M. F., Jiang L., Cellamare A. and Gasbarre L. C. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Research, 20(5): 693-703.

Liu J., Ulloa A., Perrone-Bizzozero N., Yeo R., Chen J. and Calhoun V. D. 2012. A pilot study on collective effects of 22q13. 31 deletions on gray matter concentration in schizophrenia. PloS ONE, 7(12): 52865.

Liu J., Calhoun V. D., Chen J., Claus E. D. and Hutchison K. E. 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. and Du L. 2013. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genomics, 14(1): 229.

Ma Y., Zhang Q., Lu Z., Zhao X. and Zhang Y. 2015. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. Genomics, 106(5): 295-300.

Marioni J. C., Thorne N. P. Valsesia A., Fitzgerald T., Redon R., Fiegler H. and Carter N. P. 2007. Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. Genome Biology, 8(10): R228.

Metzger J., Philipp U., Lopes M. S., da Camara Machado 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.

Norris B. J. and Whan V. A. 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Research, 18(8): 1282-1293.

Park R. W., Kim T. M., Kasif S. and Park P. J. 2015. Identification of rare germline copy number variations over-represented in five human cancer types. Molecular Cancer, 14(1): 25.

Peiffer D. A., Le J. M., Steemers F. J., Chang W., Jenniges T., Garcia F. and Cheung S. W. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Research, 16(9): 1136-1148.

Pinkel D., Segraves R., Sudar D., Clark S., Poole I., Kowbel D. and Dairkee S. H. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nature Genetics, 20(2): 207.

Pinto D., Darvishi K., Shi X., Rajan D., Rigler D., Fitzgerald T. and Prasad A. 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nature Biotechnology, 29(6): 512.

Pongpanich M., Sullivan P. F. and Tzeng J.-Y. J. B. 2010. A quality control algorithm for filtering SNPs in genome-wide association studies. Bioinformatics, 26(14): 1731-1737.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D. and Sham P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3): 559-575.

Reymond A., Henrichsen C. N., Harewood L. and Merla G. 2007. Side effects of genome structural changes. Current Opinion in Genetics & Development, 17(5): 381-386.

Salomón-Torres R., González-Vizcarra V. M., Medina-Basulto G. E., Montaño-Gómez M. F., Mahadevan P., Yaurima-Basaldúa V. H. and Villa-Angulo R. 2015. Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. Genetics and Molecular Research, 14(4): 11848-11859.

Stranger B. E., Forrest M. S., Dunning M., Ingle C. E., Beazley C., Thorne N. and Tyler-Smith C. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science, 315(5813): 848-853.

Teo Y. Y., Fry A. E., Clark T. G., Tai E. S. and Seielstad M. 2007. On the usage of HWE for identifying genotyping errors. Annals of Human Genetics, 71(5): 701-703.

Wain L. V., Armour J. A. and Tobin M. D. 2009. Genomic copy number variation, human health, and disease. The Lancet, 374(9686): 340-350.

Wang K., Li M., Hadley D., Liu R., Glessner J., Grant S. F. 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 X., Nahashon S., Feaster T. K., Bohannon-Stewart A. and Adefope N. 2010. An initial map of chromosomal segmental copy number variations in the chicken. BMC Genomics, 11(1): 351.

Winchester L., Yau C. and Ragoussis J. 2009. Comparing CNV detection methods for SNP arrays. Briefings in Functional Genomics and Proteomics, 8(5): 353-366.

Wright D., Boije H., Meadows J. R., Bed'Hom B., Gourichon D., Vieaud, A. and Andersson L. 2009. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genetics, 5(6): 1000512.

Xi R., Hadjipanayis A. G., Luquette L. J., Kim T. M., Lee E., Zhang J. and Kucherlapati R. 2011. Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. Proceedings of the National Academy of Sciences, 108(46): 1128-1136.

Xu L., Hou Y., Bickhart D., Song J. and Liu G. 2013. Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. Microarrays, 2(3): 171-185.

Yan J., Blair H. T., Liu M., Li W., He S., Chen L. and Dukkipati V. S. 2017. Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips. Small Ruminant Research, 155: 24-32.

Yang L., Xu L., Zhou Y., Liu M., Wang L., Kijas J. W. and Liu G. E. 2018. Diversity of copy number variation in a worldwide population of sheep. Genomics, 110(3): 143-148.

Yeo R. A., Gangestad S. W., Liu J., Calhoun V. D. and Hutchison K. E. 2011. Rare copy number deletions predict individual variation in intelligence. PloS ONE, 6(1): 16339.

Yuan Z., Liu E., Liu Z., Kijas J. W., Zhu C., Hu S. and Wei C. 2017. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. Animal Genetics, 48(1): 55-66.

Zhang F., GU W., Hurles M. E. and Lupski J. R. 2009. Copy number variation in human health, disease, and evolution Annual Review of Genomics and Human Genetics, 10: 451-481.

Zhang X., Du R., Li S., Zhang F., Jin L. and Wang H. 2014. Evaluation of copy number variation detection for a SNP array platform. BMC Bioinformatics, 15(1): 50.

Zhao M., Wang Q., Wang Q., Jia P. and Zhao Z. 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics, 14(11): S1.

Zhu C., Fan H., Yuan Z., Hu S., Ma X., Xuan J. and Zhao F. 2016. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. Scientific Reports, 6: 27822.

Animal Production Research

Vol. 9, No. 1, 2020 (29-44)

University of Guilan

Research paper

Investigation of copy number variation in Baluchi sheep genome using comparative analysis of PennCNV and QuantiSNP algorithms

K. Taghizadeh[1], M. Gholizadeh[2*], M. H. Moradi[3], GH. Rahimi Mianji[4]

1. Ph.D Student, Department of Animal Science, Sari Agricultural Sciences and Natural Resources University, Sari, Iran
2. Assistant Professor, Department of Animal Science, Sari Agricultural Sciences and Natural Resources University, Sari, Iran
3. Assistant Professor, Department of Animal Science, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran
4. Professor, Department of Animal Science, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

(Received: 28-05-2019 – Accepted: 12-12-2019)

Abstract

Copy number variation (CNV), one of the most important structural changes in the genome, has been known as an important source of genetic and phenotypic variations. The purpose of this study was to compare the two different algorithms in CNV detection consisting PennCNV and QuantiSNP in Baluchi sheep. Data analysis was performed using the Illumina OvineSNP50k BeadChip on 96 Baluchi sheep. After CNV calling, the copy number variation regions (CNVRs) were determined using the CNVRuler program. 91 CNVRs with a length range of 18.75 up to 511.7 kbp were identified by the PennCNV algorithm, covering 0.46% of whole sheep genome. Also, 316 CNVRs with the length range of 7.5 up to 1280 kbp were obtained using QuantiSNP algorithm, covering 1.33% of whole sheep genome. The number of loss events was about five and three times more than the number of gain events for QuantiSNP and PennCNV algorithms, respectively. Also, the number of CNVs detected by QuantiSNP was about four times higher than PennCNV. Also, 86.6% of total CNVs (174 CNVs with average length of 12.62 kb) identified by PennCNV were common with CNVs detected by QuantiSNP. In general, the results showed that the use of several algorithms could improve the accuracy for detecting the structural variation in the genome and led to a better understanding of the sheep genome.

Keywords: Genotyping array, PennCNV and QuantiSNP algorithms, Structural variation, Copy number variation

*Corresponding author: mh_gholizadeh@yahoo.com

doi: 10.22124/ar.2020.13356.1418