Shahid Bahonar University of Kerman
Agricultural Biotechnology Journal
Iranian Biotechnology Society
Print ISSN: 2228-6705 Online ISSN: 2228-6500

Identification of genomic variations of Azeri buffaloes using whole genome sequencing

Milad Hosseini

Ph.D. Student, Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: milad.hosseiny88@gmail.com

Hossein Moradi shahrbabak

*Corresponding author. Assistant Professor, Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: hmoradis@ut.ac.ir

Mohammad Moradi shahrbabak

Professor, Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail address: moradim@ut.ac.ir

Abstract

Objective

Advances in next generation sequencing technologies have created the ability to efficiently and economically sequence the whole genome more than ever and provide the opportunity to discover and introduce multiple polymorphisms throughout the genome of organisms. Azeri buffalo is one of the most important breeds of Iran and it is scattered in the north to the northwest of the country and is completely adapted to the environmental conditions of this geographical region. The main purpose of this study is to introduce the genomic diversity of Iranian buffaloes and categorize them. Also, introducing the effects of these variations on different genomic regions of Azeri buffaloes have potential applications in breeding programs.

Materials and methods

In this study, whole genome sequencing of 5 heads of Azeri buffaloes native to Iran was done by Illumina sequencing platform. Data quality was measured by FastQC software. BWA-MEM software was used for alignment with the reference genome. Finally, the variants were identified using freebayes and the SnpEff program was used to calculate the effects of the variants by mentioning their type, location and number. The alignment result of high quality reads with the reference genome showed that the alignment percentage for all 5 samples was over 97.5%, which

indicates the high quality of short reads. The coverage in the sequenced samples was determined between 4x and 12.8x.

Results

Finally, 76,298,858 million variants were identified, including 57,921,822 SNPs, 6,162,328 indels, 10,534,042 MNPs, and 1,680,666 MIXEDs. Small deletions and insertions with a minimum length of 1 bp, a maximum length of 28 bp and an average of 1.39 bp were identified. From the total number of variants, the highest frequency of variants was observed in intergenic regions, 53,789,879 (62.022 percent), intron 24,003,682 (27.677 percent), respectively, and the variants detected in exons had the lowest number.

Conclusions

Identification of genome-level variations such as snps, small insertions and deletions, and multi-nucleotide polymorphisms in different populations are a valuable resource in genetic research and can be useful in locating genomic segments responsible for important economic traits. Also, the large volume of SNPs enables genome-wide association studies in animals. In addition, the genomic variations identified in the present study can be used to develop high-density SNP arrays in Iranian breeds for genetic and breeding applications.

Keywords: Azari, sequencing, genome, buffalo, variant.

Paper Type: Research Paper.

Citation: Hosseini M, Moradi-Shahrbabak H, Moradi-Shahrbabak M (2023) Identification of genomic variations of Azeri buffaloes using whole genome sequencing. *Agricultural Biotechnology Journal* 15 (4), 227-238.

Agricultural Biotechnology Journal 15 (4), 227-238. DOI: 10.22103/jab.2023.21310.1474

Received: October 08, 2023. Received in revised form: November 25, 2023.

Accepted: November 26, 2023. Published online: December 30, 2023.

Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

© the authors

مجله بیوتکنولوژی کشاورزی
شاپا چاپی: ۶۷۰۵-۲۲۲۸ شاپا الکترونیکی: ۶۵۰۰-۲۲۲۸

شناسایی تنوع‌های ژنومی در گاومیش‌های آذری با استفاده از توالی‌یابی کل ژنوم

سیدمیلاد حسینی

دانشجوی دکتری، گروه علوم دامی، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران. رایانامه: milad.hosseiny88@gmail.com

حسین مرادی‌شهربابک

*نویسنده مسئول: دانشیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران. رایانامه: hmoradis@ut.ac.ir

محمد مرادی‌شهربابک

استاد، دانشیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران. رایانامه: moradim@ut.ac.ir

تاریخ دریافت: ۱۴۰۲/۰۷/۱۶ تاریخ دریافت فایل اصلاح شده نهایی: ۱۴۰۲/۰۹/۰۴ تاریخ پذیرش: ۱۴۰۲/۰۹/۰۵

چکیده

هدف: پیشرفت در فن‌آوری‌های توالی‌یابی نسل جدید توانایی توالی‌یابی کارآمد و اقتصادی کل ژنوم را بیشتر از همیشه ایجاد کرده و فرصت کشف و معرفی چندشکلی‌های متعدد در سرتاسر ژنوم موجودات را فراهم می‌کند. گاومیش نژاد آذری یکی از مهمترین نژادهای ایران است که در شمال تا شمال غرب کشور پراکنده شده و کاملا با شرایط محیطی این منطقه‌ی جغرافیایی سازگار شده است. هدف اصلی این مطالعه معرفی تنوع‌های ژنومی گاومیش‌های ایرانی و دسته‌بندی آن‌ها می‌باشد. همچنین معرفی اثرات این تنوع‌ها روی مناطق مختلف ژنومی گاومیش‌های آذری، کاربردهای بالقوه‌ای در برنامه‌های اصلاحی دارند.

مواد و روش‌ها: در این مطالعه توالی یابی کل ژنوم ۵ راس گاومیش آذری بومی ایران به وسیله ی پلتفرم توالی یابی ایلومینا انجام شد. سنجش کیفیت داده‌ها توسط نرم‌افزار FastQC انجام شد. برای هم‌ردیفی با ژنوم مرجع از نرم‌افزار BWA-MEM استفاده شد. در نهایت واریانت‌ها بااستفاده از freebayes شناسایی و به‌منظور محاسبه‌ی اثرات واریانت‌ها با ذکر نوع، محل و تعداد آن‌ها از برنامه SnpEff استفاده شد. نتیجه‌ی همردیفی خوانش‌های با کیفیت بالا با ژنوم مرجع نشان داد درصد همردیفی برای هر ۵ نمونه بالای ۹۷/۵ درصد بود که نشان دهنده‌ی کیفیت بالای خوانش‌های کوتاه است. میزان هم‌پوشانی در نمونه‌های توالی‌یابی شده بین ۴x تا ۱۲/۸x تعیین شد.

نتایج: در این پژوهش تعداد ۷۶،۲۹۸،۸۵۸ میلیون واریانت شناسایی شد که تعداد ۵۷،۹۲۱،۸۲۲ SNP، تعداد ۶،۱۶۲،۳۲۸ Indels، تعداد ۱۰،۵۳۴،۰۴۲ MNP و تعداد ۱،۶۸۰،۶۶۶ MIXED بودند. حذف و اضافه‌های کوچک با حداقل طول ۱، حداکثر

طول ۲۸ جفت باز و میانگین ۱/۳۹ جفت باز شناسایی شدند. از تعداد کل واریانت‌ها، بیشترین فراوانی واریانت ها به ترتیب در مناطق اینترژنیک تعداد ۵۳،۷۸۹،۸۷۹ (۶۲/۰۲ درصد)، اینترون ۲۴،۰۰۳،۶۸۲ (۲۷/۶۷ درصد) مشاهده شد و واریانت های شناسایی شده در اگزون ها کمترین تعداد را داشتند.

نتیجه‌گیری: شنا سایی تنوع‌های سطح ژنوم مانند ا سنیپ‌ها، حذف و ا ضافه‌های کوچک و چند شکلی‌های چند نوکلئوتیدی در جمعیت‌های مختلف منبعی ارزشمند در تحقیقات ژنتیکی هستند و می‌توانند در مکان‌یابی بخش‌های ژنومی مسئول ویژگی‌های مهم اقتصادی مفید با شند. همچنین حجم بسیار زیاد SNP ها، مطالعات ارتباط ژنومی را در حیوانات امکان پذیر می کند. علاوه بر این، تنوع‌های ژنومی شنا سایی شده در مطالعه حا ضر می‌تواند برای تو سعه آرایه‌های SNP با چگالی بالا در نژادهای ایرانی برای کاربردهای ژنتیکی و اصلاحی استفاده شود.

کلیدواژه‌ها: آذری، توالی‌یابی، ژنوم، گاومیش، واریانت.

نوع مقاله: پژوهشی.

استناد: حسینی میلاد، مرادی‌شهربابک حسین، مرادی‌شهربابک محمد (۱۴۰۲) شناسایی تنوع‌های ژنومی در گاومیش‌های آذری با استفاده از توالی‌یابی کل ژنوم. *مجله بیوتکنولوژی کشاورزی*، ۱۵(۴)، ۲۲۷-۲۳۸.

Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

© the authors

مقدمه

گاومیش یکی از حیوانات اهلی مهم و سودمند برای بشر است و به دلیل قابلیت‌های ویژه ، از جمله مقاومت در برابر بیماری‌ها و انگل‌ها، طول عمر مفید بالا و قابلیت استفاده از علوفه کم ارزش، مورد توجه بسیاری از محققان قرار گرفته است (Yindee et al. 2010). جمعیت گاومیش های آبی جهان حدود ۲۰۲ میلیون (http://www.fao.org/faostat/) راس ا ست که از این تعداد ۱۷۴ میلیون در آ سیا، ۵ میلیون در آفریقا (فقط در م صر)، ۳/۵ میلیون در آمریکا (به‌طور عمده در برزیل)، و برخی از جمعیت های کوچک در اروپا و استرالیا وجود دارند (Borghese 2013; Williams et al. 2017; Zhang et al 2020; Rafiepour et al. 2021). دو زیرگونه از گاومیش آبی اهلی شده وجود دارد که تصور می شود فرآیندهای اهلی سازی مستقل را پشت سر گذاشته‌اند. این دو گونه شامل گاومیش مردابی (*Bubalus bubalis carabanesis*، 2N=48) و گاومیش رودخانه‌ای (*Bubalus bubalis bubalis*، 2N=50) ه ستند (Tanaka et al. 1996; Kumar et al. 2007). گاومیش های مردابی در کشورهای آسیای شرقی و جنوب شرقی (مانند چین، ویتنام، تایلند و غیره) گسترش پیدا کرده اند. گاومیش‌های رودخانه‌ای عمدتاً

در هند، پاکستان، خاورمیانه و ایتالیا گسترش یافته اند. این حیوانات به دلیل شیر متمایز و باکیفیت شان که مشخصه آن چربی و ماده خشک بالا است، اهمیت ویژه ای در اقتصاد روستایی دارند. شیر گاومیش نسبت به شیر گاو دارای کلسترول کمتر، اما کالری و چربی بیشتری می‌باشد و پنیرهای با ارزش، غلیظ و خامه‌ای از آن تولید می شود (Borghese 2013). جمعیت بوفالوهای آبی در سراسر جهان به طور مداوم با نرخ ۱/۶۵٪ در سال در طول پنج دهه گذشته افزایش یافته است. با این حال، از قابلیت‌های گاومیش به طور کامل استفاده نشده است. پرورش دهندگان و کشاورزان گاومیش آبی با چالش های زیادی مانند کارایی تولید مثل ضعیف، تولید نامناسب و نرخ پایین بقای گوساله مواجه هستند (Michelizzi et al. 2010; Mokhber et al. 2018). توجه و بهبود این صفات می‌تواند باعث افزایش کارایی تولید گاومیش‌ها و همچنین منجر به حمایت بیشتر از اقتصاد روستایی شود. گاومیش‌های ایرانی از شبه قاره‌ی هند منشا گرفته‌اند و در سرتاسر شمال، شمال غرب، جنوب و جنوب غرب بر اساس ویژگی‌های زیست بومی منطقه مانند دما، رطوبت و ارتفاع گسترش پیدا کرده‌اند. به نظر می‌رسد گاومیش‌های ایرانی از نوع رودخانه‌ای هستند یکی از مهمترین نژادهای ایران، گاومیش آذری است (Safari 2018). گاومیش آذری یکی از دو نژاد غالب در ایران است و در شمال تا شمال غرب کشور عمدتا در استان‌های گیلان، اردبیل و آذربایجان پراکنده شده‌اند. این نژاد دارای جثه کوچک و رنگ بدن خاکستری با لکه‌های سفید است (Safari 2018; Rafiepour et al. 2021). گاومیش‌های آذری از نظر فنوتیپی شبیه به گاومیش‌های رودخانه‌ای مدیترانه‌ای هستند که تصور می‌شود از اجداد یکسانی هستند (Borghese 2013; Rafiepour et al. 2021). دوره‌ی شیردهی این نژاد ۲۰۰ تا ۲۲۰ روز با تولید متوسط شیر ۱۳۰۰ کیلوگرم و محتوای چربی ۶/۸ درصد است. وزن بدن برای نرهای بالغ ۴۰۰ تا ۶۰۰ کیلوگرم و برای ماده‌ها ۳۵۰ تا ۵۰۰ کیلوگرم است. اولین زایش در حدود سن ۳۴ ماهگی انجام می شود. همچنین وزن تولد برای گو ساله‌های نر ۳۷ کیلوگرم و برای گو ساله‌های ماده ۳۰ کیلوگرم است. گاومیش‌های آذری به خوبی با دمای متوسط ۳۵ درجه سانتیگراد در تابستان و ۵ درجه سانتیگراد در زمستان سازگار هستند (Borghese 2013; Mokhber et al. 2018; Safari et al. 2018). تعیین توالی کل ژنوم و معرفی تنوع‌های سطح ژنوم، به شناسایی تغییراتی که منجر به تنوع فنوتیپی می‌شوند کمک می‌کند. فناوری‌های توالی‌یابی در حال حاضر به سرعت دانش موجود در مورد ژنوم موجودات مختلف را افزایش می‌دهند (Michelizzi et al. 2010). استفاده از رویکردهای ژنتیکی مولکولی بهره‌وری را در برنامه‌های اصلاحی افزایش می‌دهد و می‌تواند برای توصیف، درک و حفاظت از تنوع ژنتیکی جهانی گاومیش مورد استفاده قرار گیرد (Zimin et al. 2013). به‌علاوه، مطالعات و بررسی‌های به عمل آمده در اواخر دهه ۸۰ میلادی روشن نمود که مکانیسم‌های مولکولی در زمره مهم‌ترین فرایندهای ژنتیکی (مشتمل بر همانندسازی DNA، رونویسی، ترجمه و حتی نحوه تنظیم ژن ها) هستند (Barazandeh et al. 2016a; Safaei et al. 2022). ماده ژنتیکی یک سلول از تعداد زیادی ژن تشکیل شده است که هیچ‌گاه همه به طور همزمان بیان نمی‌شوند و در هر لحظه فقط تعداد کمی از آن‌ها بیان شده و پروتئین یا آنزیم مورد نیاز سلول را تولید می‌نمایند (Barazandeh et al. 2016b). محیطی که موجود در آن رشد می‌کند مشخص می‌کند که آیا ژن بیان شود و یا این که نیازی به فرآورده آن نیست و باید غیرفعال و یا خاموش شود (Bordbar et al. 2022; Jafari

Ahmadabadi et al. 2023). ساز و کار بیان ژن اولین بار در باکتری E.coli کشف شد (Mohamadipoor et al. 2021; Masoudzadeh et al. 2020). ژن‌های یوکاریوتی بیان شان تحت کنترل موقت و چندبعدی ا ست. در هر یک از انواع بافت‌ها تنها یک مجموعه نسبتاً کوچک از تمام ژنوم بیان می شود و نیز بیان ژن‌ها به مرحله نمو بستگی دارد (Masoudzadeh et al. 2020). بنابراین، بیان ژن در یوکاریوت‌ها برای هر بافت اختصا صی ا ست. همچنین مقدار محصولات ژن که در همان بافت و نیز در سایر بافت‌هایی که آن محصول را می سازند، ساخته شده سبب تنظیم بیان آن ژن می شود (Mohammadabadi et al. 2021; Shokri et al. 2023). یکی از اقدامات اساسی در به‌نژادی ملکولی مطالعه ژن‌ها و پروتئین‌های مرتبط با صفات مهم و مطالعه آن‌ها در سطح سلولی یا کروموزومی است (Mohammadinejad et al. 2022; Shahsavari et al. 2022). لذا، در این پژوهش آنالیز توالی ژنوم ۵ راس از گاومیش‌های آذری، به منظور شناسایی واریانت‌های ژنومی از جمله حذف و اضافه‌های کوچک و چند شکلی‌های چندنوکلئوتیدی (MNP) و همچنین شناسایی چند شکلی‌های تک‌نوکلئوتیدی (SNP) در سرتاسر ژنوم گاومیش‌های ایران صورت گرفت.

مواد و روش‌ها

جمع آوری داده‌ها: در این پروژه از اطلاعات توالی‌یابی شده ۵ راس گاومیش نژاد آذری استفاده شد. توالی‌یابی کل ژنوم به صورت paired-end یا دو سویه توسط دستگاه ایلومینا Hiseq 2500 انجام شد. در این روش، هر کدام از دو رشته رفت (Reverse) و برگشت (Forward) مربوط به هر قطعه توالی کوتاه توالی‌یابی می‌شوند. توالی‌یابی ژنوم در موسسه PTP ایتالیا با حمایت شرکت دانشگاهی و دانش‌بنیان توسعه کشت و دام نو اندیش البرز و با استفاده از تکنیک توالی‌یابی با برونده بالا یا نسل دوم توالی یابی (NGS)[1] انجام شد.

کنترل کیفیت داده‌ها و ویرایش داده‌ها: برای سنجش کیفیت داده‌ها از نرم افزار FastQC (version 0.11.5) تحت جاوا استفاده شد (Andrews 2010). برای ویرایش داده‌ها از نرم افزار Trimmomatic استفاده شد (Bolger et al. 2014). این نرم افزار به عنوان ابزاری منعطف با پیش پردازش‌های مؤثر و منطبق با داده‌های pairedend است و برای داده‌های توالی‌یابی نسل بعد دستگاه شرکت ایلومینا بهینه سازی شده است. وظایف این نرم افزار شامل حذف آداپتورها و حذف یا ویرایش خوانش‌های بی‌کیفیت می باشد.

همردیفی و مکان‌یابی خوانش‌ها بر روی ژنوم مرجع: برای هم‌ردیفی داده‌ها با ژنوم مرجع گاو (UMD3.1) از بسته نرم افزاری BWA-MEM استفاده شد که نسبت به الگوریتم‌های دیگر دارای سرعت پردازش بالاتری می‌باشد. سپس با

[1] Next generation sequencing

استفاده از بسته نرم افزاری samtools فایل خروجی با فرمت sam به bam تبدیل شد (Li et al. 2009). برای به دست آوردن درصد همردیفی و هم‌پوشانی از دستورات flagstat و depth موجود در نرم افزار SamTools استفاده شد. فایل واریانت‌های ژنومیک با استفاده از (https://github.com/freebayes/freebayes) freebayes به دست آمد (Garrison & Marth 2012).

مستند سازی واریانت‌های ژنومی: به‌نظور مستند سازی یا انوتیشن، محاسبه شمار هموزیگوس‌ها وهتروزیگوس‌ها درچندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک، و همچنین شمار جهش‌های جابه‌جایی و معکوس در چندریختی‌های تک‌نوکلئوتیدی، از برنامه SnpEff software version 5.1 (http://pcingola.github.io/SnpEff/) استفاده شد. این برنامه از طریق مستندسازی، آثار واریانت‌ها را روی ژن‌ها پیش بینی می‌کند.

نتایج و بحث

راهکارهای توالی‌یابی کل ژنوم روش‌های قدرتمندی برای کشف تنوع‌های موجود در سطح ژنوم هستند. تغییرات گسترده ژنوم برای نقشه‌برداری صفات و بهبود تولید بسیار مهم هستند و در سال‌های اخیر بسیار اهمیت پیدا کرده‌اند. در این مطالعه، ما شناسایی واریانت‌های ژنوم گاومیش‌های نژاد آذری را با استفاده از توالی کل ژنوم انجام دادیم. توالی‌یابی کل ژنوم ۵ راس گاومیش بومی ایران به وسیله‌ی پلتفرم توالی‌یابی ایلیومینا انجام شد. نتایج کنترل کیفیت داده‌ها نشان داد که خوانش‌های کوتاه[2] کیفیت بالا و بدون آلودگی اولیه(پرایمری) داشتند. طول خوانش‌های کوتاه برای همهٔ داده‌ها ۱۰۱ بود. درصد همردیفی برای هر ۵ نمونه بالای ۹۷/۵ درصد بود که نشان دهنده‌ی کیفیت بالای خوانش‌های کوتاه است (جدول ۱). میزان کاوریج در نمونه‌های توالی‌یابی شده بین x ۴ تا x ۱۲/۸ بود.

واریانت‌های ژنومی بعد از همردیفی توالی‌های کوتاه کنترل کیفیت شده با ژنوم مرجع استخراج شدند تعداد چندریختی‌های تکنوکلئوتیدی در گاومیش‌های نژاد آذری محاسبه گردید. در این پژوهش تعداد ۷۶،۲۹۸،۸۵۸ میلیون واریانت شناسایی شد. که از این‌ها، تعداد ۵۷،۹۲۱،۸۲۲ SNP، تعداد ۶،۱۶۲،۳۲۸ Indels، تعداد ۱۰،۵۳۴،۰۴۲ MNP و تعداد ۱،۶۸۰،۶۶۶ MIXED بودند. در همه‌ی نمونه‌ها میزان جهش‌های هموزایگوت بیشتر از دو برابر هتروزایگوت‌ها بود. جایگزینی نوکلئوتیدها می‌تواند به صورت جهش جابه‌جایی که جایگزینی یک پورین به جای پورین یا یک پیریمیدین به جای پیریمیدین است یا به صورت جهش معکوس که جایگزینی یک پورین با پیریمیدین یا برعکس است رخ دهد. کل جهش‌های جابه‌جایی[3] و معکوس[4] کشف شده در این تحقیق به ترتیب ۲۸۵،۲۲۶،۹۹۱ و ۱۲۹،۷۱۱،۹۵۵ بود. نتایج مستند سازی واریانت‌های سطح ژنوم گاومیش‌های آذری نشان داد که از

[2] Short reads
[3] Transition
[4] Transversion

تعداد کل واریانت‌ها، فراوانی واریانت‌ها به ترتیب در مناطق اینترژنیک تعداد ۵۳،۷۸۹،۸۷۹ (۶۲/۰۲ درصـد)، اینترون ۲۴،۰۰۳،۶۸۲ (۲۷/۶۷ در صد)، پایین د ست ژنی ۳،۸۹۱،۸۸۳ (۴/۴۸ در صد)، بالاد ست ژنی ۳،۶۸۷،۸۴۴ (۴/۲۵در صد) و اگزون ۹۹۵،۷۸۷ (۱/۱۴ درصـد) برآورد شـد. کمتر از ۰/۵ درصـد از کل واریانت‌ها نیز در دیگر مناطق شـناسـایی شـدند. این نتایج نشـان داد واریانت‌های شنا سایی شده در اگزون‌ها کمترین تعداد و واریانت‌ها در مناطق اینترژنیک و اینترون بی‌شترین تعداد را دا شتند. (شکل۱). تغییرات تک نوکلئوتیدی غیر مترادف[5] با پتانسـیل تغییر عملکردهای ژن مرتبط، در مطالعات اصـلاح مولکولی به دلیل نقش عمیق آن‌ها در روابط صفات و انتخاب به کمک نشانگر بسیار حائز اهمیت است (Guajardo et al. 2020).

جدول ۱. خلاصه‌ی اطلاعات مربوط به توالی‌یابی کل ژنوم گاومیش‌های آذری

Table 1. Summary of the Azari buffaloes whole-genome sequencing data

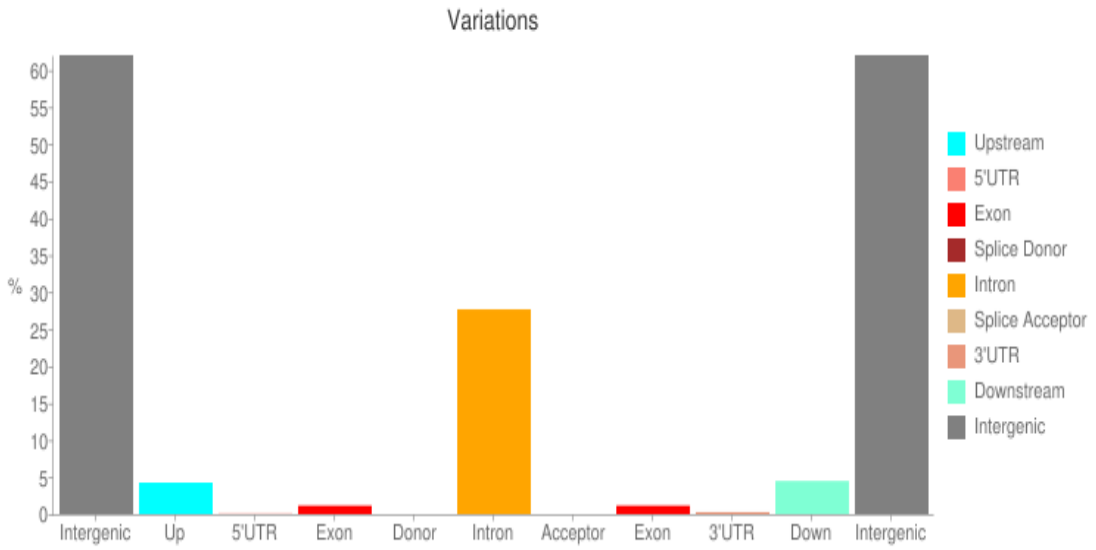
	خوانش‌های کنترل کیفیت شده QC-passed reads	درصد خوانش-های نقشه‌یابی شده Mapped (%)	تعداد خوانش-های نقشه-یابی شده Mapped	خوانش‌های جفت شده در توالی یابی Paired in sequencing	خوانش۱ Read 1	خوانش۲ Read 2	تعداد خوانش-های جفت شده صحیح Properly paired	درصد خوانش-های جفت شده صحیح Properly paired (%)
نمونه ۱ Sample 1	125008525	97.8	122264309	123861824	61930912	61930912	108986230	87.99
نمونه ۲ Sample 2	118969441	98.2	116827070	117813322	58906661	58906661	104272440	88.51
نمونه ۳ Sample 3	261447560	97.92	256016516	259052172	129526086	129526086	222207332	85.78
نمونه ٤ Sample 4	197673159	98.23	194171863	195875376	97937688	97937688	173159588	88.40
آذری ۵ Sample 5	117311351	97.5	114381481	116189586	58094793	58094793	104157176	89.64

همچنین تعداد کل جهش‌های خاموش[6] (هم معنی) ۷۲/۴۷ در صد، بی معنی[7] ۲۷/۳۴ در صد و بد معنی[8] ۰/۱۸ در صد بد ست آمد. به علت مکانیسـم‌های مولکولی ایجاد کننده جهش در سـطح مولکول DNA مانند دی آمیناسـیون اکسـیدتیو تعداد چندریختی‌های تک‌نوکلئوتیدی که توسـط جهش‌های جابه‌جایی تولید می‌شـوند، تقریباً دو برابر تعداد چندریختی‌های تک

[5] Non-synonymous
[6] Silent
[7] Missense
[8] Nonsense

Agricultural Biotechnology Journal; Print ISSN: 2228-6705, Electronic ISSN: 2228-6500

نوکلئوتیدی ایجاد شـده توسـط جهش‌های معکوس هسـتند (Collins & Jukes 1994). همانطور که در دیگر مطالعات نیز گزارش شده است، نرخ بالاتر شاخص ts/tv نشان دهنده‌ی کیفیت بهتر فراخوانی سنیپ‌ها است (Bakhtiarizadeh & Alamouti 2020). در این مطالعه تعداد ۶,۱۶۲,۳۲۸ حذف و اضـافه‌ی کوچک با حداقل طول ۱، حداکثر طول ۲۸ جفت باز و میانگین ۱/۳۹ جفت باز شناسایی شد. حذف و اضافه‌های کوچک نقش مهمی در تنوع فنوتیپی مشاهده شده بین افراد یک گونه ایفا می‌کند. برای مثال در انبه، یک جهش درج دی نوکلئوتید در ژن کد کننده آنزیم لیکوپن β-سـیکلاز (CpCYC-b) باعث تنوع فنوتیپی قرمز و زرد در میوه می شود (Blas et al. 2010). نتایج آنالیزها نشان داد که بیشتر از ۹۸ درصد واریانت‌ها شدت ملایم کننده[9] دا شتند، این نوع از واریانت‌ها تاثیر روی مناطق غیرکدکننده دارند. همچنین کمترین مقدار برای واریانت‌های با شدت زیاد[10] ثبت شـد (۰/۰۱۳درصـد) که این نوع از وارینت‌ها دارای اثر مخربی بر روی مسـیرهای پروتیین سـازی و آنزیم‌ها دارند (جدول ۲) (Shirasawa et al. 2013).



شکل ۱. نمودار درصد پراکنش واریانت‌ها در نواحی مختلف ژنوم گاومیش‌های آذری

Figure 1. Distribution percentage of variants in different regions of Azeri buffalo genome

نتیجه‌گیری: در این پژوهش به منظور شناسایی تنوع‌های کل ژنوم گاومیش‌های آذری بومی ایران، کل ژنوم پنج نمونه گاومیش بومی ایران توسط روش توالی‌یابی نسل جدید، به صورت Paired-End با پوشش دهی بالا توالییابی و بررسی شد. تجزیه وتحلیل داده‌های توالی‌یابی نسل جدید تنوع‌های کل ژنوم گاومیش‌ها را به خوبی نشان داد و تنوع‌ها تقسیم‌بندی و معرفی شدند. همانطور که می‌دانیم تنوع‌های موجود در جمعیت‌ها از مهمترین منابع برای دست‌یابی به اهداف مختلف در اصلاح نژاد است.

[9] Modifier Impact
[10] High Impact

با توجه به اینکه انجام برنامه‌های محافظت از ذخایر ژنتیکی پرهزینه بوده و امکان ذخیره تمام تنوع در محیط‌های آزمایشـــگاهی و ایستگاه‌های تحقیقاتی وجود ندارد، بنابراین حفظ تنوع ژنتیکی در محل زندگی حیوان بهترین راهکار جهت محافظت از ویژگی‌های نژادهای بومی تحت سیستم اصلاح نژاد می‌باشد. حذف و اضافه‌های کوچک و چند شکلی‌های چندنوکلئوتیدی در جمعیت‌های مختلف منبعی ارزشمند در تحقیقات ژنتیکی هستند و می‌توانند در مکان‌یابی بخش‌های ژنومی مسئول ویژگی‌های مهم اقتصادی مفید باشند. همچنین حجم بسیار زیاد SNP ها، مطالعات ارتباط ژنومی را در حیوانات امکان پذیر می‌کند. علاوه بر این، تنوع‌های ژنومی شـناسـایی‌شـده در مطالعه حاضـر می‌تواند برای توسـعه آرایه‌های SNP با چگالی بالا در نژادهای ایرانی برای کاربردهای ژنتیکی و اصلاحی استفاده شود.

جدول ۲. تعداد اثرات تنوع‌ها بر اساس شدت

Table 2. Number of variant effects by impact

نوع شدت اثر Type	تعداد Count	درصد Percent
زیاد HIGH	11,525	0.013%
کم LOW	720,545	0.831%
متوسط MODERATE	310,311	0.358%
اصلاح کننده MODIFIER	85,684,379	98.798%

سپاسگزاری: نگارندگان بر خود لازم می‌دانند از آقای دکتر مهدی مخبر و داوران محترم مجله‌ی بیوتکنولوژی کشاورزی به خاطر ارائه نظرها و راهنمایی‌های ارزشمند سپاسگزاری نمایند.

منابع

جعفری احمدآبادی سید علی اصغر، عسکری‌همت حشمت‌اله، محمدآبادی محمدرضا (۱۴۰۲) تاثیر شاهدانه بر بیان ژن DLK1 در بافت قلب بره‌های کرمانی. مجله بیوتکنولوژی کشاورزی، ۱۵(۱)، ۲۱۷-۲۳۴.

شکری سمیرا، خضری امین، محمدآبادی محمدرضا، خیرالدین حمید (۱۴۰۲) بررسی بیان ژن MYH7 در بافت‌های ران، دست و راسته بره‌های پرواری نژاد کرمانی. مجله بیوتکنولوژی کشاورزی، ۱۵(۲)، ۲۱۷-۲۳۶.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bakhtiarizadeh MR, Alamouti AA (2020) RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. Sci Rep 10**,** 13525.

Barazandeh A, Mohammadabadi MR, Ghaderi-Zefrehei M, Nezamabadipour H (2016a) Predicting CpG Islands and Their Relationship with Genomic Feature in Cattle by Hidden Markov Model Algorithm. Iran J Appl Anim Sci 6 (3), 571-579 .

Barazandeh A, Mohammadabadi MR, Ghaderi-Zefrehei M, Nezamabadipour H (2016b) Genome-wide analysis of CpG islands in some livestock genomes and their relationship with genomic features. Czech J Anim Sci 61, 487.

Blas AL, Ming R, Liu Z et al. (2010) Cloning of the papaya chromoplast-specific lycopene β-cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. Plnt Physio 152**,** 2013-2022.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform 30**,** 2114-2120.

Bordbar F, Mohammadabadi M, Jensen J, et al. (2022) Identification of candidate genes regulating carcass depth and hind leg circumference in simmental beef cattle using Illumina Bovine Beadchip and next-generation sequencing. J Anim 12 (9), e1103.

Borghese A (2013) Buffalo livestock and products in Europe. Buffalo Bulletin 32**,** 50-74.

Collins DW, Jukes TH (1994) Rates of transition and transversion in coding sequences since the human-rodent divergence. Genomics 20**,** 386-396.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907.

Guajardo V, Solís S, Almada R et al. (2020) Genome-wide SNP identification in Prunus rootstocks germplasm collections using genotyping-by-sequencing: phylogenetic analysis, distribution of SNPs and prediction of their effect on gene function. Sci Rep 10**,** 1467.

Jafari Ahmadabadi SAA, Askari-Hemmat H, Mohammadabadi M, et al. (2023) The effect of Cannabis seed on DLK1 gene expression in heart tissue of Kermani lambs. Agric Biotechnol J 15 (1), 217-234 (In Persian).

Kumar S, Nagarajan M, Sandhu J et al. (2007) Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. Anim Genet 38**,** 227-232.

Li H, Handsaker B, Wysoker A et al. (2009) The sequence alignment/map format and SAMtools. Bioinform 25**,** 2078-2079.

Masoudzadeh, S.H., Mohammadabadi, M.R., Khezri, A., et al. (2020) Dlk1 gene expression in different Tissues of lamb. Iran J Appl Anim Sci 10, 669-677.

Michelizzi VN, Dodson MV, Pan Z et al. (2010) Water buffalo genome science comes of age. Int J Biol Sci 6**,** 333.

Mohamadipoor L, Mohammadabadi M, Amiri Z, et al. (2021) Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. BMC Vet Res 17 (1), 1-9.

Mohammadabadi M, Masoudzadeh SH, Khezri A, et al. (2021) Fennel (Foeniculum vulgare) seed powder increases Delta-Like Non-Canonical Notch Ligand 1 gene expression in testis, liver, and humeral muscle tissues of growing lambs. Heliyon 7 (12), e08542 .

Mohammadinejad F, Mohammadabadi M, Roudbari Z, Sadkowski T (2022) Identification of Key Genes and Biological Pathways Associated with Skeletal Muscle Maturation and Hypertrophy in Bos taurus, Ovis aries, and Sus scrofa. J Anim 12 (24), 3471.

Mokhber M, Moradi-Shahrbabak M, Sadeghi M et al. (2018) A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. BMC Genom 19**,** 1-9.

Rafiepour M, Ebrahimie E, Vahidi MF et al. (2021) Whole-genome resequencing reveals adaptation prior to the divergence of buffalo subspecies. Genome Biol Evol 13**,** evaa231.

Safaei SMH, Dadpasand M, Mohammadabadi M, et al. (2022) An Origanum majorana Leaf Diet Influences Myogenin Gene Expression, Performance, and Carcass Characteristics in Lambs. J Anim 13 (1), e14.

Safari A Ghavi Hossein-Zadeh N, Shadparvar AA, Abdollahi AR. (2018) A review on breeding and genetic strategies in Iranian buffaloes. Trop Anim Health Prod 50 (4)**,** 707-714.

Shahsavari M, Mohammadabadi M, Khezri A, et al. (2022) Effect of Fennel (Foeniculum Vulgare) Seed Powder Consumption on Insulin-like Growth Factor 1 Gene Expression in the Liver Tissue of Growing Lambs. Gene Expr 21 (2), 21-26.

Shirasawa K, Fukuoka H, Matsunaga H et al. (2013) Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. DNA Res 20**,** 593-603.

Shokri S, Khezri A, Mohammadabadi M, Kheyrodin H (2023). The expression of MYH7 gene in femur, humeral muscle and back muscle tissues of fattening lambs of the Kermani breed. Agric Biotechnol J 15 (2), 217-236 (In Persian).

Tanaka K, Solis CD, Masangkay JS et al. (1996) Phylogenetic relationship among all living species of the genus Bubalus based on DNA sequences of the cytochrome b gene. Biochem Genet 34**,** 443-452.

Williams JL, Iamartino D, Pruitt KD et al. (2017) Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2 n= 50). Gigascience 6**,** gix088.

Yindee M, Vlamings B, Wajjwalku W et al. (2010) Y-chromosomal variation confirms independent domestications of swamp and river buffalo. Anim Genet 41**,** 433-435.

Zhang Y, Colli L, Barker J (2020) Asian water buffalo: domestication, history and genetics. Anim genet 51**,** 177-191.

Zimin AV, Marçais G, Puiu D et al. (2013) The MaSuRCA genome assembler. Bioinform 29**,** 2669-2677.