



Identification of genomic variants and considering their effects in Mazandarani buffaloes

Milad, Hosseini¹, Hossein Moradi Shahrabak^{2✉}, Mohammad, Moradi Shahrabak³

1. Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: milad.hosseiny88@gmail.com
2. Corresponding Author, Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: hmoradis@ut.ac.ir
3. Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: moradim@ut.ac.ir

Article Info

Article type:

Research Article

Article history:

Received: 11 June 2023

Received in revised form: 30 July 2023

Accepted: 6 September 2023

Published online: 21 December 2024

Keywords:

Whole-Genome, Variations, Next-Generation-Sequencing, Mazandarani Buffalo.

ABSTRACT

Nowadays, with the progress obtained in molecular genetic techniques and bioinformatics, alternative methods have been invented to increase the speed and efficiency of DNA sequencing and their roles in genome level. In the whole-genome sequencing method, the whole genome sequence of organism (nuclear-genome with mitochondrial DNA) is sequenced. One of the important topics in genomics is genome differences, including single-nucleotide polymorphisms and INDELS for the relationship between genotype and phenotype. Polymorphisms are powerful tools for molecular analysis of economic traits and are important in breeding programs. For this purpose, whole-genome variations of Mazandarani buffaloes were identified and classified. In this study, the whole-genome of 4 Mazandarani buffaloes was sequenced with the Illumina platform. Data quality was measured by FastQC software. BWA-MEM was used for alignment with reference genome. Finally, the variants were obtained using freebayes and the SnpEff was used to calculate the effects of the variants. The result of aligning led us to identification of 56537534 SNPs, and 6128529 indels with an average coverage of x4 to x13. The most number of variants were observed on 1 and X chromosomes, and the least number were in 23 and mitochondrial chromosomes. The transition, transversion and rate of transition/transversion mutations were 236549743, 108015966 and 2/19 respectively. Also, the mutations was calculated. The frequency of variants in intergenic regions was estimated to be 52746727, intron 23560994, downstream 3713594, upstream 3571409 and exon 574093. Considering that this research is the only one that carried out to identify the genomic variations of Mazandarani buffalo, the identified genomic variations can be used for the development of SNP-arrays for Iranian breeds.

Cite this article: Hosseini, M., Moradi Shahrabak, H. & Moradi Shahrabak, M. (2024). Identification of genomic variants and considering their effects in Mazandarani buffaloes. *Iranian Journal of Animal Science*, 55 (4), 713-725. DOI: <https://doi.org/10.22059/ijas.2023.359835.653948>



© The Author(s).

DOI: <https://doi.org/10.22059/ijas.2023.359835.653948>

Publisher: The University of Tehran Press.

Extended Abstract

Introduction

Livestock breeding is one of the main sectors of animal production, with important role in economy, self-sufficiency and food security. Existing genetic diversity and resource in native populations with good adaptation to their environmental conditions are very valuable for breeding and production plans. Buffalo is one of the important native livestock of the country and has an important contribution in the production of milk, meat.

Objective

Research on DNA variants that directly affect the phenotype is one of the main and key field of genetic research in domestic animals. In the past few years, single nucleotide polymorphisms have played a important

role in the field of genetics studies of domestic animals. One of the important topics in genomics studies is genome differences, including single-nucleotide polymorphisms and INDELs to consider relation between genotype and phenotype. Polymorphisms are powerful tools for molecular analysis of economic traits and are important in breeding programs. Thus the aim of this research was to identify single-nucleotide polymorphisms and INDELs across the whole genome of Mazanderani buffaloes.

Research method

In this study, the whole-genome of 4 Mazandarani buffaloes was sequenced with the Illumina platform. Data quality was measured by FastQC software. This software uses 11 different tests to measure data quality. Trimmomatic software was used to edit the data. This software is a flexible tool with effective pre-processing and compatible with paired end data, and it is optimized for Illumina company's next generation sequencing data. The tasks of this software include removing adapters and removing or editing poor quality readings. Then, using the samtools software package, we converted the output file in sam format to bam. To obtain the alignment and coverage percentage, we used the flagstat and depth commands used in SamTools software. The file of genomic variants was obtained using freebayes. After receiving the reference genome file and its annotation file, it was necessary first to index the information related to the reference genome. BWA-MEM software package was used to align and index the data with the cow reference genome (UMD3.1). Compared to other algorithms, it has a higher processing speed. The SnpEff was used to calculate the effects of the variants.

Results

The result of aligning led to identification of 56537534 SNPs, 6128529 indels with an average coverage of x4 to x13. The most number of variants were observed on 1 and X chromosomes, and the least number were in 23 and mitochondrial chromosomes. The transition, transversion and rate of transition/transversion mutations were 236549743, 108015966 and 2/19 respectively. Also, the mutations were calculated as. The frequency of variants in intergenic regions was estimated to be 52746727, intron 23560994, downstream 3713594, upstream 3571409 and exon 574093.

Conclusion

Considering the important role of buffalo in providing part of the income and necessities of the rural population, special attention should be paid to these animals in order to raise the level of welfare of the rural population and also to increase the production efficiency of buffaloes in the country. Therefore, in order to improve and raise the production level of these animals, it is very important to know the genetic variations. The present study is the only research carried out to identify the genomic variations of Mazandarani buffalo, so the genomic variations identified in this study can be used for the development of high-density SNP arrays for genetic and breeding applications in Iranian breeds.



شناسایی تنوع‌های ژنومی و بررسی نقش آن‌ها در سطح ژنوم گاومیش‌های نژاد مازندرانی

میلاد، حسینی^۱ | حسین مرادی شهراباک^۲ | محمد، مرادی شهراباک^۳

۱. گروه علوم دامی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه: milad.hosseiny88@gmail.com

۲. نویسنده مسئول، گروه علوم دامی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه: hmoradis@ut.ac.ir

۳. گروه علوم دامی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه: moradim@ut.ac.ir

اطلاعات مقاله	چکیده
<p>نوع مقاله:</p> <p>مقاله پژوهشی</p> <p>تاریخ دریافت: ۱۴۰۲/۰۳/۲۱</p> <p>تاریخ بازنگری: ۱۴۰۲/۰۵/۰۸</p> <p>تاریخ پذیرش: ۱۴۰۲/۰۶/۱۵</p> <p>تاریخ انتشار: ۱۴۰۳/۱۰/۰۱</p> <p>کلیدواژه‌ها:</p> <p>تنوع‌های کل ژنوم، توالی‌یابی نسل جدید، گاومیش مازندرانی.</p>	<p>امروزه با پیشرفت فناوری‌های ژنتیک مولکولی و توسعه پایگاه‌های بیوانفورماتیکی، روش‌های جایگزینی جهت افزایش سرعت و بازدهی تعیین توالی دی‌ان‌ای و بررسی نقش آنها در سطح ژنوم توسعه یافته است. در روش توالی‌یابی کل ژنوم، ژنوم موجود زنده (ژنوم هسته‌ای به همراه دی‌ان‌ای میتوکندریایی) بطور کامل توالی‌یابی می‌گردد. یکی از مباحث مهم در حوزه ارزیابی‌های ژنومی، مطالعه تفاوت‌ها و تنوع ژنوم، شامل چندشکلی‌های تک‌نوکلئوتیدی (SNP) و حذف‌ودرج‌ها بمنظور شناخت ارتباط میان ژنوتیپ و فنوتیپ می‌باشد. در ژنوم موجودات، چندشکلی‌ها، ابزارهای نیرومندی برای واکاوی مولکولی صفات اقتصادی هستند و کاربردهای بالقوه ای در برنامه‌های اصلاحی دارند. برای دستیابی به این اهداف، تنوع‌های ژنومی گاومیش‌های مازندرانی شناسایی و طبقه‌بندی شدند. در این مطالعه ژنوم ۴راس گاومیش مازندرانی با فناوری ایلومینا توالی‌یابی شد. سنجش کیفیت داده‌ها توسط نرم‌افزار FastQC انجام شد. برای هم‌ردیفی با ژنوم مرجع از نرم‌افزار BWA-MEM استفاده شد. در نهایت واریانت‌ها با استفاده از freebayes شناسایی و بمنظور محاسبه‌ی اثرات واریانت‌ها با ذکر نوع، محل و تعداد آن‌ها از برنامه SnpEff استفاده شد. یافته‌ها: نتیجه‌ی هم‌ردیفی خوانش‌ها، با ژنوم مرجع منجر به شناسایی ۵۶۵۳۷۵۳۴ نشانگر SNP و ۶۱۲۸۵۲۹ حذف‌ودرج (ایندل) با میانگین پوشش ۴ تا ۱۳۱ x شد. بیشترین واریانت‌ها در کروموزوم‌های ۱ و جنسی (X) کم‌ترین آن در کروموزوم‌های ۲۳ و میتوکندریایی مشاهده شد. جهش‌های جابه‌جایی ۲۳۶۵۴۹۷۴۳ و معکوس ۱۰۸۰۱۵۹۶۶ بود. همچنین نرخ جهش‌های جابه‌جایی / معکوس ۲/۱۹ محاسبه شد. فراوانی واریانت‌ها در مناطق بین ژنی ۵۲۷۴۶۷۲۷، اینترون ۲۳۵۶۰۹۹۴، پایین‌دست ژنی ۳۷۱۳۵۹۴، بالادست ژنی ۳۵۷۱۴۰۹ و آگزون ۵۷۴۰۹۳ برآورد شدند. نتیجه‌گیری: باتوجه به این که مطالعه‌ی حاضر تنها تحقیق انجام شده در زمینه‌ی شناسایی تنوع‌های ژنومی گاومیش نژاد مازندرانی می‌باشد، تنوع‌های ژنومی شناسایی شده در این مطالعه می‌تواند برای توسعه‌ی آرایه‌های نشانگری با چگالی بالادر نژادهای ایرانی مورد استفاده قرار گیرد.</p>

استناد: حسینی، میلاد؛ مرادی شهراباک، حسین و مرادی شهراباک، محمد (۱۴۰۳). شناسایی تنوع‌های ژنومی و بررسی نقش آن‌ها در سطح ژنوم گاومیش‌های نژاد مازندرانی.

نشریه علوم دامی ایران، ۵۵ (۴)، ۷۱۳-۷۲۵. DOI: <https://doi.org/10.22059/ijas.2023.359835.653948>



© نویسندگان.

DOI: <https://doi.org/10.22059/ijas.2023.359835.653948>

ناشر: مؤسسه انتشارات دانشگاه تهران.

مقدمه

دامپروری به عنوان یکی از بخش‌های اصلی کشاورزی است که نقش آن در اقتصاد و خودکفایی و امنیت غذایی انکارناپذیر می‌باشد و وجود تنوع و استعدادهای ژنتیکی در نژادهای بومی که برای زیست در شرایط محیطی خاص خود تطابق یافته‌اند، سرمایه‌های ژنی با ارزشی هستند که می‌بایست به بهترین شکل ممکن مورد استفاده قرار گیرند. یکی از دام‌های مهم بومی ایران گاو میش بوده و سهمی هرچند اندک در تولید شیر و گوشت و سایر فرآورده‌های دامی دارد (Arefnejad *et al.*, 2015). جمعیت گاو میش‌های جهان بیش از ۲۰۷ میلیون راس می‌باشد که این تعداد در سال ۱۹۶۲ در حدود ۸۸ میلیون رأس بوده و در طی این پنج دهه میانگین رشد سالانه جمعیت در حدود ۶/۱ بوده است. در حالی که این تعداد برای گاو، گوسفند و بز به ترتیب حدود ۴۹/۱، ۱۷/۱ و ۱ میلیارد رأس و ۱ میلیارد رأس می‌باشد. جمعیت گاو میش دنیا در ۱۲۹ کشور جهان و عموماً در قاره آسیا پراکنده شده‌اند و فقط تعداد کمی در قاره‌های دیگر وجود دارد. در حدود ۱۹۴ میلیون رأس یا ۹۷ درصد از گاو میش‌های دنیا در آسیا پرورش داده می‌شود. گاو میش از لحاظ تولید شیر و گوشت نقش حیاتی در اقتصاد و سلامتی مردم کشورهای مختلف ایفا می‌کند (Iamartino *et al.*, 2013). گاو میش‌های ایران به دلیل سازگاری با محیط مقاومت در برابر بیماری‌ها، پایین بودن هزینه‌های نگهداری و استفاده مؤثر از ضایعات کشاورزی و مواد خشبی کم‌ارزش، یکی از ذخایر ژنتیکی با ارزش محسوب می‌شود و حفظ این ذخیره ژنتیکی ضروری می‌باشد. همچنین این حیوان نقش اساسی در اشتغال اقشار کم درآمد روستایی و تأمین بخشی از نیاز غذایی کشور به شیر و گوشت را دارد. سودآوری گاو میش تابع میزان تولید شیر در هر زایش و درصد چربی آن می‌باشد. هدف اصلی از اصلاح نژاد گاو میش، افزایش راندمان تولید در حیوانات گله از طریق ایجاد پیشرفت ژنتیکی برای صفات تولیدی است (Aminafshar *et al.*, 2008). وجود تنوع و استعدادهای ژنتیکی در توده‌های بومی که برای زیست در شرایط محیطی خاص خود حاصل شده‌اند، سرمایه‌های ژنی باارزشی هستند که چنانچه این سرمایه‌ها بدون هرگونه بهره‌وری و استفاده از دست برون امکان بازیابی مجدد آنها وجود نخواهد داشت. لذا لازم است همگام با حفاظت از توده‌های بومی، پتانسیل ژنتیکی صفات اقتصادی آنها به منظور طراحی برنامه‌های مناسب اصلاح‌نژادی، به طور دقیق شناسایی گردد. در تلاش برای اصلاح دام‌های بومی، باید جنبه‌های مختلفی مورد بررسی قرار گیرد. از جمله این موارد می‌توان به شناخت حیوان با توجه به داده‌های تولیدی، انتخاب افراد شایسته که با دارا بودن مقاومت لازم در برابر عوامل نامساعد محیطی از سطح تولید خوبی نیز برخوردار باشند و سامانه آمیزشی مناسب که باعث جمع‌آوری صفات مطلوب در یک فرد و انتقال آن به نسل بعد می‌شود، اشاره نمود. زیست‌داده‌ورزی، دانش استفاده از علوم رایانه و آمار و احتمالات در شاخه زیست‌شناسی مولکولی است. در چند دهه اخیر، پیشرفت در زیست‌شناسی مولکولی و تجهیزات مورد نیاز تحقیق در این زمینه باعث افزایش سریع تعیین توالی ژنوم بسیاری از گونه‌های موجودات شد، تا جایی که پروژه‌های تعیین توالی ژنوم از پروژه‌های بسیار رایج به حساب می‌آیند. امروزه توالی ژنوم بسیاری از موجودات ساده مانند باکتری‌ها تا موجودات بسیار پیشرفته چون یوکاریوت‌های پیچیده شناسایی شده است (Claverie & Notredame, 2006). استفاده از داده‌های کل ژنوم یا به طور کلی داده‌های DNA و به کارگیری آن می‌تواند زمینه‌ساز ورود به عصر پساژنومی در اصلاح‌نژاد حیوانات اهلی گردد. امروزه درباره چندشکلی‌های تک نوکلئوتیدی^۱ (SNP) کارهای بسیاری درباره گاو میش در سطح جهان انجام گرفته است، ولی متأسفانه در کشور ایران با توجه به کمبود امکانات و هزینه‌های توالی‌یابی و همچنین نیاز به سیستم‌های بالا برای آنالیز داده‌ها، کار آنچنانی صورت نگرفته است. با توجه به اهمیت تنوع‌ها در مطالعات اصلاح‌نژادی، مطالعه‌ی حاضر ضروری به نظر می‌آید. در این پژوهش آنالیز کل توالی ژنوم ۴ راس از گاو میش‌های مازندرانی، به منظور شناسایی واریانت‌های ژنومی از جمله حذف^۲ و اضافه‌های^۳ کوچک و

1 Bioinformatic

2 Single nucleotide polymorphism

3 Deletion

4 Insertions

چند شکلی‌های چند نوکلئوتیدی^۱ (MNP) و همچنین شناسایی چندشکلی تک نوکلئوتیدی در سرتاسر ژنوم گاومیش‌های ایران صورت گرفت. بنابراین فصل نوینی در ارزیابی‌های ژنتیکی گاومیش بر اساس واریانت‌های ژنومی، و برنامه‌های اصلاح نژادی گشوده خواهد شد.

پیشینه پژوهش

تحقیق در مورد واریانت‌های DNA که مستقیماً فنوتیپ افراد را تحت‌تاثیر قرار می‌دهد، یکی از زمینه‌های اصلی و کلیدی تحقیقات ژنتیکی در حیوانات اهلی است. در چند سال گذشته چند شکلی‌های SNP به واسطه کاربرد وسیع و گسترده آنها در آزمایش با توان عملیاتی بالا، نقش محوری را در زمینه ژنتیک حیوانات اهلی ایفا کرده است. در همین رابطه مطالعات پویش ژنومی با استفاده از ده‌ها هزار SNP مکانیسم ژنتیکی بیماری‌ها و همچنین معماری ژنتیکی برخی از صفات پیچیده در گاوها، گوسفندان و دیگر حیوانات اهلی شناسایی و ارزیابی دام‌ها بر اساس نشانگرهای SNP با موفقیت انجام شده است (Kidd *et al.*, 2008). جهت نقشه‌یابی ژنتیکی ژنوم از سه روش می‌توان استفاده کرد: الف) نقشه‌یابی بر اساس آنالیز پیوستگی^۳ (ب) نقشه‌یابی بر اساس (RH)^۴ و ج) نقشه‌یابی بر اساس تکنیک (FISH)^۵؛ دی موو^۶ و همکارانش (۲۰۰۸) اولین نقشه سیتوژنتیکی برای گاومیش‌های آبی را تنها با استفاده از تکنیک (FISH) برای ۶۸ جایگاه ارائه کردند. پس از آن ۱۳۱ ژن به همراه ۱۲۲ ریزماوه‌ها بر روی ژنوم گاومیش آبی معرفی گردید و امروزه ۳۰۹ جایگاه نقشه‌یابی شده بر روی کلیه کروموزوم‌ها در سرتاسر ژنوم شناسایی شده است. همچنین با استفاده از این تکنیک قطعات تکرار شده در سرتاسر ژنوم با طول‌های مشخص به نام رتروترانسپوزون‌ها^۶ که قسمتی از ژن‌های عملکردی در گاو و گاومیش هستند، نقشه‌یابی شده است. پانل‌های RH جهت نقشه‌یابی فیزیکی کروموزوم‌های گاومیش اولین بار توسط گلدامر و همکارانش (۲۰۰۷) مورد استفاده قرار گرفت (Di Meo *et al.*, 2008) و پس از آن تلاش‌هایی جهت تولید اولین نقشه ژنتیکی RH کل ژنوم برای تمام کروموزوم‌های گاومیش توسط آمارال و همکاران (۲۰۰۸) صورت گرفت (Amaral *et al.*, 2008). پس از آن نقشه RH کروموزوم Y گاومیش آبی (۲۰۰۹) توسط استافوزا ارائه گردید (Stafuzza *et al.*, 2007). اخیراً هم نقشه بهبود یافته کروموزوم X گاومیش نیز توسط پروکاتی و همکاران (۲۰۱۲) انتشار یافت (Perucatti *et al.*, 2012). ژنوم میتوکندریایی (mtDNA) گاومیش آبی یک مولکول DNA حلقوی است که اولین بار با همکاری سه گروه از دانشمندان چینی، هندی و ایتالیایی توالی‌یابی و توسط پیترو و همکاران (۲۰۰۴) منتشر گردید. mtDNA در گاومیش‌های آبی حاوی ۳۷ ژن، شامل ۱۳ ژن جهت کدگذاری پلی پپتید و ۲۲ ژن جهت رمزگذاری tRNA و هر کدام از آنها نیز در ارتباط با زیر واحدهای بزرگ و کوچک (rRNA) خواهند بود (Lei *et al.*, 2007). شناسایی واریانت‌های ساختاری گوناگون با به کارگیری توالی‌یابی، نخستین بار به وسیله توزون و همکاران (۲۰۰۵) و با نقشه‌یابی دو سویه ارایه شد. هر چند کارایی توالی‌یابی و وضوح این پژوهش بهینه نبود ولی با توجه به این که قطعات با طول بلند توالی‌یابی شده بودند، امکان شناسایی دقیق واریانت‌های ساختاری بزرگ فراهم شد (Tuzun *et al.*, 2005). گسترش نسل جدید تکنولوژی‌های توالی‌یابی با برونده بالا، امکان توالی‌یابی همه ژنوم‌های بزرگ (با اندازه ژنوم انسان) را در زمان یک هفته فراهم کرد. بی‌درنگ در پی آن الگوریتم‌های محاسباتی گوناگون برای شناسایی واریانت‌های ساختاری بر پایه توالی‌های کوتاه

1 Multi-nucleotide polymorphisms

2 High throughput assays

3 Linkage

4 Radiation hybrid mapping

5 Fluorescent in situ hybridization

6 Retrotransposon

گسترش پیدا کردند. این الگوریتمها دربرگیرنده روشهای بر پایه جفت‌های خوانش؛ عمق خوانش؛ خوانش‌های تکه‌ای^۳ و سرهم سازی ژنوم هستند. ماتوکومالی و همکاران (۲۰۰۹) اولین تحقیقات را در مورد یافتن مکان‌های حذف شده (deletion) روی ژنوم گاو انجام داده و بیش از ۲۰۰ مکان CNV را مورد شناسایی قرار دادند (Mardis, 2011). بیک هارت و همکاران (۲۰۱۲) با استفاده از اطلاعات تعیین توالی کل ژنوم ۶ گاو از ۴ نژاد آنگوس، هلشتاین، هرورد و نیلور هندی در حدود ۱۲۶۵ منطقه حاوی CNV را شناسایی کردند (Bickhart et al., 2012). داس و همکاران (۲۰۱۵) در یک مطالعه که به بررسی و شناسایی پروتئین‌های کم اثر بود به شناسایی SNP با استفاده از داده‌های توالی‌یابی کل ژنوم پرداختند. در این تحقیق ۴ گاو هلشتاین دانمارکی با میانگین پوشش ۲۷ بار با تکنولوژی Illumina 2000Hiseq توالی‌یابی گردید. در فراخوانی چندباره برای نمونه‌ها تعداد ۱۰ میلیون چند شکلی تک نوکلئوتیدی شناسایی شد، که از این مقدار تعداد ۴۸۲ هزار چندشکلی (۴/۵ درصد)، جدید بودند (Das et al., 2015). رمزگشایی توالی DNA، برای تمام شاخه‌های تحقیقاتی بیولوژیکی ضروری است. یامارتینو و همکاران (۲۰۱۳) در گزارشی از نتایج و آینده توالی‌یابی ژنوم گاومیش بیان داشتند که آگاهی از ژنوم گاومیش که دامی با اهمیت برای دامداران خرده پا در کشورهای در حال توسعه است می‌تواند به آشکارسازی رابطه تکاملی موجود، میان نژادهای مختلف کمک نموده و نقش مهمی در مدیریت و کنترل ژنتیکی روی تنوع فنوتیپی ایفا نماید (Iamartino et al., 2013).

روش‌شناسی پژوهش

جمع‌آوری داده‌ها

در این پروژه از اطلاعات توالی‌یابی شده ۴ راس گاومیش نژاد مازندرانی 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 توالی‌یابی^۴ (NGS) انجام شد.

سنجش کیفیت و ویرایش داده‌ها

برای سنجش کیفیت داده‌ها از نرم‌افزار FastQC (v0.11.5) تحت جاوا استفاده شد (Andrews, 2010). این نرم‌افزار برای سنجش کیفیت داده‌ها از ۱۱ آزمون مختلف استفاده می‌کند که شامل: کیفیت توالی به ازای هر باز، کیفیت توالی به ازای tile، امتیاز کیفیت هر توالی، محتوای نوکلئوتیدی توالی به ازای هر باز، محتوای GC به ازای توالی، محتوای باز خوانده نشده (N) به ازای هر باز، توزیع طول توالی‌ها، تکراری بودن توالی، محتوای آدپتوری، محتوای Kmer و توالی‌هایی که بیش از حد وجود دارند است. نرم‌افزار نتیجه هر آزمون را به طور جداگانه و به صورت گرافیکی گزارش می‌دهد. رد، قبول و هشدار برای آنها را به ترتیب با رنگ‌های قرمز، سبز و نارنجی مشخص می‌کند. وضعیت قبول، رد یا هشدار برای آزمون‌های ۱۱ گانه ملاک سنجش کیفیت داده‌ها قرار می‌گیرد. برای ویرایش داده‌ها از نرم‌افزار Trimmomatic (v0.40) استفاده شد (Bolger et al., 2014). این نرم‌افزار به عنوان ابزاری منعطف با پیش پردازش‌های مؤثر و منطبق با داده‌های paired-end است و برای داده‌های توالی‌یابی نسل بعد دستگاه شرکت ایلومینا بهینه‌سازی شده است. وظایف این نرم‌افزار شامل حذف آدپتورها و حذف یا ویرایش خوانش‌های بی کیفیت می‌باشد. بعد از ویرایش داده‌ها، کیفیت داده‌ها مجدداً ارزیابی شد.

1Read Pairs
2Read Depth
3Split Reads
4Next generation sequencing

همردیفی و مکان‌یابی خوانش‌ها بر روی ژنوم مرجع

پس از دریافت فایل ژنوم مرجع و فایل حاشیه‌نویسی آن ابتدا لازم است که اطلاعات مربوط به ژنوم مرجع را شاخص کرد برای هم‌ردیفی و شاخص کردن داده‌ها با ژنوم مرجع گاو (UMD3.1) از بسته نرم افزار BWA-MEM (v0.7.17) استفاده شد (Li, 2013). که نسبت به الگوریتم‌های دیگر دارای سرعت پردازش بالاتری می‌باشد. شاخص کردن شبیه به شاخص‌بندی و فهرست‌بندی یک کتاب می‌باشد. فهرست‌بندی ژنوم مرجع به هم‌ردیف‌کننده اجازه می‌دهد تا منشأ یک توالی را در ژنوم به آسانی پیدا کند و از این طریق در زمان و حافظه صرفه‌جویی می‌گردد. سپس با استفاده از بسته نرم‌افزاری SamTools (v1.18) فایل خروجی با فرمت sam را به bam تبدیل می‌کنیم (Li et al., 2009). برای به دست آوردن درصد هم‌ردیفی و کاوریج از دستورات flagstat و depth به کار برده شده در نرم‌افزار SamTools استفاده کردیم. فایل واریانت‌های ژنومی با استفاده از freebayes (v1.3.7) به دست آمد (Garrison & Marth, 2012).

مستندسازی نتایج

به منظور محاسبه‌ی آثار واریانت‌ها با ذکر نوع و محل آنها (اگر، بین ژنی، اینترون و غیره)، تعداد جهش‌های جابجایی و معکوس، نسبت جابجایی به معکوس برای چند شکلی‌های تک نوکلئوتیدی (SNP) و همچنین تعداد SNP‌های هموزیگوت و هتروزیگوت در ژنوم گاوهای مازندرانی از نرم‌افزار SnpEff (v5.1) استفاده شد (Cingolani et al., 2012). این برنامه از طریق مستندسازی، آثار واریانت‌ها را روی ژن‌ها پیش‌بینی می‌کند. این برنامه به زبان جاوا نوشته شده است و تحت محیط لینوکس قابل اجرا است.

یافته‌های پژوهش

در این مطالعه توالی‌یابی کل ژنوم ۴ راس گاو میش نژاد مازندرانی با استفاده از فناوری توالی‌یابی با برونده بالا یا نسل دوم توالی‌یابی تعیین گردید. توالی‌یابی کل ژنوم با میانگین ۱۷۰ میلیون خوانش با میانگین طول ۳۵ جفت باز برای هر نمونه به دست آمد که در مجموع همه‌ی نمونه‌ها ۸۰ گیگابایت اطلاعات تولید شد. میانگین محتوای GC خوانش‌ها برای همه‌ی نمونه‌ها ۴۲ درصد به دست آمد. طول همه خوانش‌ها برای تمام نمونه‌ها ۱۰۱ جفت باز بود و میانگین تعداد نوکلئوتیدهای خوانده نشده (N) نیز برای همه خوانش‌ها صفر محاسبه شد. پس از حذف خوانش‌های بی‌کیفیت (کیفیت پایین)، همردیفی خوانش‌های باکیفیت، با ژنوم مرجع گاو منجر به شناسایی ۵۶۵۳۷۵۳۴ میلیون SNP و ۶۱۲۸۵۲۹ ایندل (درج و حذف) شد. میانگین پوشش^۱ برای همه‌ی ۴ نمونه بین ۴x تا ۱۳x محاسبه شد. واریانت‌ها در سرتاسر همه‌ی کروموزوم‌های گاو میش‌های مازندرانی شناسایی شدند. بیشترین تعداد واریانت‌ها به ترتیب در کروموزوم‌های ۱ و X، و کمترین آن در کروموزوم‌های ۲۳ و DNA میتوکندریایی مشاهده شد (جدول ۱). مقایسه واریانت‌های شناسایی شده نشان داد که نمونه‌ها در حدود ۹۸ درصد از آلل‌های سطح ژنوم را با ژنوم مرجع به اشتراک گذاشتند. نتیجه‌ی همردیفی خوانش‌های با کیفیت بالا با ژنوم مرجع نشان داد درصد همردیفی برای هر ۴ نمونه بالای ۹۸/۰۳ درصد بود که نشان دهنده‌ی کیفیت بالای خوانش‌های کوتاه است. میزان شباهت هر یک از نمونه‌ها با ژنوم مرجع و جزئیات توالی‌یابی در (جدول ۲) آورده شده است. میانگین کیفیت واریانت‌ها ۲۶۱/۴۱ و میانگین طول ایندل‌ها ۱/۴۰۹ bp برآورد شد. معمولاً واریانت‌های SNP به دو گروه هموزیگوت و هتروزیگوت تقسیم می‌شوند. طبق نتایج به آمده در تمام نمونه‌های گاو میش‌های ایرانی مورد مطالعه، میزان واریانت‌های هموزیگوت بیشتر از واریانت‌های هتروزیگوت بود.

جدول ۱. تعداد تنوع‌های شناسایی شده به تفکیک هر کروموزوم

کروموزوم	طول کروموزوم (جفت باز)	تعداد تنوع‌ها
۱	۱۵۸۳۳۷۰۶۷	۴۴۰۱۲۹۶
۲	۱۳۷۰۶۰۴۲۴	۳۷۷۱۰۹۲
۳	۱۲۱۴۳۰۴۰۵	۳۳۲۷۹۰۸
۴	۱۲۸۲۹۶۹۹	۳۴۱۸۵۳۴
۵	۱۲۱۱۹۱۴۹۰	۳۳۵۱۰۱۹
۶	۱۱۹۴۵۸۷۳۶	۳۳۱۱۰۴۸
۷	۱۱۲۶۳۸۶۵۹	۳۰۴۷۶۳۳
۸	۱۱۳۳۸۴۸۳۶	۳۱۱۲۴۲۳
۹	۱۰۵۷۰۸۲۵۰	۲۹۵۲۲۵۵
۱۰	۱۰۴۳۰۵۰۱۶	۲۸۶۲۵۲۸
۱۱	۱۰۷۳۱۰۷۶۳	۳۰۰۷۰۶۰
۱۲	۹۱۱۶۳۱۲۵	۲۶۷۴۱۳۱
۱۳	۸۴۲۴۰۳۵۰	۲۳۹۱۴۱۷
۱۴	۸۴۶۴۸۳۹۰	۲۳۸۳۱۵۵
۱۵	۸۵۲۹۶۶۷۶	۲۴۳۷۱۶۵
۱۶	۸۱۷۲۴۶۷۸	۲۲۹۹۰۳۷
۱۷	۷۵۱۵۸۵۹۶	۲۱۳۵۶۵۳
۱۸	۶۶۰۰۴۰۲۳	۱۸۷۰۵۱۰
۱۹	۶۴۰۵۷۴۵۷	۱۷۹۵۹۲۸
۲۰	۷۲۰۴۲۶۵۵	۲۱۴۳۶۹۶
۲۱	۷۱۵۹۹۰۹۶	۲۰۵۵۲۶۶
۲۲	۶۱۴۳۵۸۷۴	۱۷۷۵۴۱۰
۲۳	۵۲۵۳۰۰۶۳	۱۵۵۳۱۶۷
۲۴	۶۳۷۱۴۹۳۰	۱۸۶۴۵۹۰
MT	۱۶۳۳۸	۸۹۰
X	۱۴۸۱۲۳۸۹۹	۳۱۷۷۰۵۱

کل جهش‌های جابه‌جایی کشف شده در این مطالعه ۲۳۶۵۴۹۷۴۳ و کل جهش‌های معکوس ۱۰۸۰۱۵۹۶۶ بود. همچنین نرخ جهش‌های جابه‌جایی به معکوس، ۲/۱۹ تخمین زده شد. در مجموع ۸۴۴۷۲۶۵۴ اثر بر اساس موقعیت ژنومی از ۷۴۴۳۸۸۵۶ واریانت در سطح کل ژنوم گاو میش‌های مازندرانی پیش‌بینی شد. بر اساس شدتی که SNPها میتوانند روی ژن‌ها داشته باشند اثرات واریانت‌ها به چهار دسته‌ی تأثیر اصلاح‌کننده، کم، متوسط و تأثیر زیاد طبقه‌بندی شدند. نتایج آنالیزها نشان داد که بیشترین تعداد SNPها شدت اصلاح‌کننده (۹۹/۲۶ درصد) داشتند و کوچکترین ارزش برای SNPهایی با شدت زیاد (۰/۱۲ درصد) ثبت شد (جدول ۳).

جدول ۲. خلاصه‌ی اطلاعات مربوط به توالی‌یابی کل ژنوم گاو میش‌های مازندرانی

درصد خوانش‌های جفت شده صحیح	تعداد خوانش‌های جفت شده صحیح	خوانش ۲	خوانش ۱	خوانش‌های جفت شده در توالی‌یابی	تعداد خوانش‌های نقشه‌یابی شده	درصد خوانش‌های نقشه‌یابی شده	خوانش‌های کنترل کیفیت شده
۸۸/۳۷	۲۵۶۶۶۵۵۱۸	۱۴۵۲۲۰۷۱۲	۱۴۵۲۲۰۷۱۲	۲۹۰۴۴۱۴۲۴	۲۸۷۸۵۱۲۵۳	۹۸/۲۱	۲۹۳۰۹۴۹۸۱
۸۵/۹۸	۹۵۸۵۹۳۳۰	۵۵۷۴۳۲۵۶	۵۵۷۴۳۲۵۶	۱۱۱۴۸۶۵۱۲	۱۱۰۵۲۲۶۱۸	۹۸/۰۳	۱۱۲۷۴۴۳۵۹
۸۸/۲۳	۱۵۲۳۱۸۰۳۸	۸۶۳۱۶۱۳۹	۸۶۳۱۶۱۳۹	۱۷۲۶۳۲۲۷۸	۱۷۱۱۵۷۵۰۷	۹۸/۲۱	۱۷۴۲۸۲۷۶۴
۸۸/۰۴	۱۸۰۹۲۶۳۸۸	۱۰۲۷۵۵۴۷۶	۱۰۲۷۵۵۴۷۶	۲۰۵۵۱۰۹۵۲	۲۰۳۴۷۳۴۹۳	۹۸/۱۱	۲۰۷۴۰۲۶۶۸

جدول ۳. تعداد اثرات تنوع‌ها بر اساس شدت

نوع شدت اثر	تعداد	درصد
زیاد	۹۸۷۳	۰/۰۱۲
کم	۴۰۰۵۳۸	۰/۴۷۴
متوسط	۲۱۴۱۵۸	۰/۲۵۴
اصلاح کننده	۸۳۸۴۸۰۸۵	۹۹/۲۶۱

نتایج مستندسازی واریانت‌های سطح ژنوم گاو میش‌های مازندرانی نشان داد که از تعداد کل واریانت‌ها، فراوانی واریانت‌ها به ترتیب در مناطق بین ژنی تعداد ۵۲۷۴۶۷۲۷ (۶۲/۴۴ درصد)، اینترون ۲۳۵۶۰۹۹۴ (۲۷/۸۹ درصد)، پایین دست ژنی ۳۷۱۳۵۹۴ (۴/۳۹ درصد)، بالادست ژنی ۳۵۷۱۴۰۹ (۴/۲۲ درصد) و اگزون ۵۷۴۰۹۳ (۰/۶۸ درصد) برآورد شد. همچنین مقدار کمی از واریانت‌ها در مناطق UTR_3_PRIME (۰/۲۴۲ درصد) و UTR_5_PRIME (۰/۴۷ درصد) واقع شده بودند. کمتر از ۰/۴ درصد از کل واریانت‌ها نیز در دیگر مناطق شناسایی شدند.

بحث

یکی از مهم‌ترین داده‌های متکی بر DNA که هم در پزشکی انسانی و هم در اصلاح نژاد حیوانات و گیاهان جایگاه ویژه‌ای برای خود باز کرده، نشانگرهای مبتنی بر چندشکلی‌های تک نوکلئوتیدی یا SNP می‌باشند. در این مطالعه تعداد ۵۶۵۳۷۵۳۴ میلیون چندشکلی تک نوکلئوتیدی شناسایی و معرفی شد. لی رویز و همکاران (۲۰۱۲) طی یک پژوهش روی بوفالوی افریقایی به عنوان یک گونه مقاوم در اکوسیستم ساوانا و مقاوم در برابر انگل‌های خارجی و داخلی و همچنین بیماری‌های عفونی اقدام به شناسایی SNP کردند. در این پژوهش DNA ۹ حیوان با استفاده از دستگاه Sequencer ABI SOLID4 توالی‌یابی گردید. سپس توالی‌های بدست آمده با ژنوم مرجع گونه Bos Taurus مکان‌یابی گردید. پس از آنالیز و بررسی کنترل کیفیت داده‌های بدست آمده تعداد ۲ میلیون SNP بدست آمد (Le Roex et al., 2012). عوامل زیادی همچون نرم‌افزارهای مورد استفاده، پیش‌فرض‌های مربوط به نرم‌افزارها، نوع داده‌ها و دستگاه‌های توالی‌یابی کننده، حیوانات نمونه‌برداری شده، نژاد و شرایط فیلتر

کردن SNP های خام در تعداد چندشکلی های شناسایی شده دخیل می باشند (Maghsoodi *et al.*, 2011). همانطور که در جدول ۱ گزارش شد، به طور کلی نسبت تعداد SNP های شناسایی شده به تفکیک کروموزوم در برخی از کروموزومها ارتباط مستقیمی با طول کروموزوم دارد. ولی این موضوع برای همه کروموزومها صادق نیست (کروموزوم ۳ و ۵). همچنین شمار اندکی از کل SNP های شناسایی شده بر روی توالی های غیر کروموزومی یا میتوکندریایی قرار گرفته بودند. با مقایسه تعداد SNP های شناسایی شده و تعداد SNP های موجود بر روی کروموزومها مشخص شد تقریباً ۱ الی ۲ درصد SNP های شناسایی شده غیر کروموزومی می باشند. همچنین وجود تعداد زیاد SNP موجود بر روی کروموزوم بیانگر تاثیر زیاد آن کروموزوم در رونویسی نمی باشد. مقصودی و همکاران (۲۰۱۱) در پژوهشی که بر داده های ترانسکریپتوم گاو انجام گرفت اعلام کردند طول کروموزوم رابطه مستقیمی با تعداد SNP شناسایی شده در داده های ترانسکریپتومی ندارد و رونویسی از تمام طول ژنوم با توزیع همگن و پوشش یکسان رخ نمی دهد (Maghsoodi *et al.*, 2011).

SNP، تعویضات تک نوکلئوتیدی هستند که به طور عمده به دودسته جابه جایی (مبادلات دو پورین یا دو پیریمیدین از قبیل آدنین-گوانین یا سیتوزین-تیمین) و معکوسها^۱ (تبادلات بین پورین ها و پیریمیدین ها مثلاً آدنین-تیمین یا آدنین-سیتوزین) تقسیم بندی می شوند. نسبت شمار جهشها شاخص و ابزار سودمندی برای محاسبه خطای شناسایی تعداد نشانگرهای SNP می باشد. هرچه این مقدار بالاتر باشد نشان دهنده خطای کمتر در شناسایی می باشد (Guo *et al.*, 2012). تعداد بیشتر اثرات در مقایسه با تعداد واریانتها به این دلیل است که یک واریانت خاص می تواند بر چندین ژن تاثیر بگذارد (مثلاً یک واریانت می تواند در پایین دست^۲ یک ژن و بالادست^۳ از یک ژن دیگر باشد). نتایج آنالیزها نشان داد که بیشتر از ۹۹ درصد واریانتها شدت اصلاح کننده داشتند، این نوع از واریانتها روی مناطق غیر کدکننده تاثیر دارند. تنها ۰/۶۸ درصد از اثرات واریانتها در مناطق کدکننده قرار گرفتند. در مجموع ۹۸۷۳ واریانت با تاثیر زیاد (۰/۰۱۲ درصد) مشاهده شد که کمترین مقدار بین این ۴ دسته اثر بود. این نوع از واریانتها دارای اثر مخرب بر روی مسیرهای پروتیین سازی و آنزیمها هستند و تاثیر مستقیمی بر عملکرد ژنها دارند (جدول ۳) (Shirasawa *et al.*, 2013). از متداول ترین اثرات ناشی از SNP های با تاثیر زیاد می توان به از دست دادن کدون توقف^۴ و افزایش کدون توقف^۵ اشاره کرد که ممکن است منجر به تغییر گسترده ای عملکرد ژنها شود (Bohry *et al.*, 2021). همچنین، حذف و درج های با تاثیر زیاد، عمدتاً باعث اختلال در قالب خواندن ترجمه^۶ می شود و ممکن است باعث تولید محصولات پروتئینی غیر طبیعی با یک توالی اسید آمینه نادرست شوند. در مجموع تعداد ۲۱۴۱۵۸ اثر با شدت متوسط مشاهده شد. SNP های دارای تاثیر متوسط، باعث تغییر در یک اسید آمینه به دلیل یک جایگزینی غیر مترادف^۷ می شوند. SNP های با شدت تاثیر کم عمدتاً از جانشینی های مترادف تشکیل شده اند که در آنها هیچ گونه تغییری در کدون اسیدهای آمینه ایجاد نمی شود (Bohry *et al.*, 2021). تنوع های ژنومی با بسیاری از مکان های صفت کمی در حیوانات اهلی مرتبط هستند و یک تنوع به تنهایی می تواند تاثیر زیادی بر فنوتیپ فرد داشته باشد (Pavlopoulos *et al.*, 2013). حذف و درجها هنگامی که در مناطق کدگذاری یافت می شود، به طور کلی قالب خواندن ترجمه (نوع تغییر فریم) را مختل می کند، به جز زمانی که جهش مضربی از سه نوکلئوتید باشد (Bohry *et al.*, 2021). همچنین حذف و درجها نیز نقش مهمی در تنوع فنوتیپی مشاهده شده بین افراد یک گونه ایفا می کند (Wheeler *et al.*, 2007). این نتایج نشان داد که بیشتر تنوعها در مناطق غیر

1Transition

2Transversion

3Downstream

4Upstream

5High Impact

6Stop codon lost

7Stop codon gain

8Translational reading frame

9Non-synonymous

10Synonymous

کدکننده قرار گرفته‌اند. پیش از این نیز نتایج مشابه در مطالعات دیگر در گاو، گوسفند و اسب بدست آمده است (Doan *et al.*, 2014; Zhang *et al.*, 2011; Kijas *et al.*, 2011; Fontanesi *et al.*, 2012). اگرچه اکثر تنوع‌ها در داخل ژن‌ها قرار ندارند، اما فراوانی و استحکام، آنها را به منبع مهمی برای کمک به برنامه‌های اصلاح‌نژاد گاوهای مازندرانی تبدیل می‌کند. همچنین می‌توان با افزایش اندازه‌ی نمونه از نتایج تحقیق حاضر بمنظور شناسایی SNP‌های مرتبط با صفات تولیدی در سایر نژادهای ایرانی، بررسی نشانه‌های انتخاب بین نژادهای مختلف و شناسایی جایگاه‌های تحت انتخاب استفاده کرد.

نتیجه‌گیری

در این مطالعه به منظور شناسایی واریانت‌های موجود در سطح ژنوم گاوهای مازندرانی، کل ژنوم ۴ نمونه گاوهای بومی ایران توسط روش توالی‌یابی نسل جدید، با پوشش‌دهی بالا توالی‌یابی و تنوع‌های کل ژنوم آنها شناسایی و طبقه‌بندی شد. دام‌ها طی فرایند اهلی شدن در معرض عوامل ژنتیکی مختلفی مانند انتخاب مصنوعی، رانش ژنتیکی و جهش قرار داشته‌اند. در این پژوهش علاوه بر آنالیز توالی ژنوم گاوهای مازندرانی، شناسایی تنوع‌های ژنومی، و همچنین شناسایی تنوع‌های تک‌نوکلئوتیدی در سرتاسر ژنوم گاوهای مازندرانی انجام شد. نتایج این تحقیق می‌تواند فصل نوینی در ارزیابی‌های ژنتیکی گاوهای موجود در ایران بر اساس تمامی تنوع‌های ژنوم، بررسی‌های جمعیتی و یا انتخاب حیوان برتر، بگشاید. با توجه به اینکه مطالعه‌ی حاضر تنها تحقیق انجام شده در زمینه‌ی شناسایی تنوع‌های ژنومی گاوهای مازندرانی می‌باشد، و از طرفی با توجه به تفاوت عملکرد بین نژادهای گاوهای مازندرانی به لحاظ تولید و اندازه بدن، می‌توان از نتایج این تحقیق، برای توسعه آرایه‌های SNP با چگالی بالا در نژادهای ایرانی برای کاربردهای ژنتیکی، و اصلاحی استفاده شود.

منابع

عارف نژاد، بابک؛ کهرام، حمید؛ محمد، مرادی شهربابک؛ شاکری، ملک؛ دونگ، یانگ؛ زانگ، خیائولی؛ وانگ، ون؛ حسینی‌سالکده، قاسم (۲۰۱۵). شناسایی واریانت‌های ژنی اسب کاسپین با استفاده از نسل جدید توالی‌یابی ژنوم با کارایی بالا. *بیوتکنولوژی کشاورزی*, ۶(۴), ۳۴-۴۶.

REFERENCES

- Amaral, M. E. J., Grant, J. R., Riggs, P. K., Stafuzza, N. B., Goldammer, T., Weikard, R., Jeong, J. (2008). A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC genomics*, 9(1), 1-11.
- Aminafshar, M., Amirinia, C., & Torshizi, R. V. (2008). Genetic diversity in buffalo population of guilan using microsatellite markers. *Journal of Animal Veterinary advances*, (7), 1499-1502.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online. Retrieved May, 17, 2018.
- Arefnejad, B., Kohram, H., Moradi Shahrabak, M., Shakeri, M., Dong, Y., Zhang, X., Hoseini Salekdeh, G. (2015). Genic Variant Detection of Caspian Horse Using High-throughput Sequencing Technology. *Agricultural Biotechnology Journal*, 6(4), 101-116. (in persian).
- Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., Taylor, J. F. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome research*, 22(4), 778-790.
- Bohry, D., Ramos, H. C. C., Dos Santos, P. H. D., Boechat, M. S. B., Arêdes, F. A. S., Pirovani, A. A. V., & Pereira, M. G. (2021). Discovery of SNPs and InDels in papaya genotypes and its potential for marker assisted selection of fruit quality traits. *Scientific Reports*, 11(1), 292.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.

- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *fly*, 6(2), 80-92.
- Claverie, J.-M., & Notredame, C. (2006). *Bioinformatics for dummies*. John Wiley & Sons.
- Das, A., Panitz, F., Gregersen, V. R., Bendixen, C., & Holm, L.-E. (2015). Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. *BMC genomics*, 16(1), 1-12.
- Di Meo, G., Perucatti, A., Floriot, S., Hayes, H., Schibler, L., Incarnato, D., Eggen, A. (2008). An extended river buffalo (*Bubalus bubalis*, 2n= 50) cytogenetic map: assignment of 68 autosomal loci by FISH-mapping and R-banding and comparison with human chromosomes. *Chromosome research*, 16, 827-837.
- Doan, R., Cohen, N., Harrington, J., Veazy, K., Juras, R., Cothran, G., Dindot, S. V. (2012). Identification of copy number variants in horses. *Genome research*, 22(5), 899-907.
- Fontanesi, L., Beretti, F., Martelli, P., Colombo, M., Dall'Olio, S., Occidente, M., Russo, V. (2011). A first comparative map of copy number variations in the sheep genome. *Genomics*, 97(3), 158-165.
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.
- Guo, Y., Li, J., Li, C.-I., Long, J., Samuels, D. C., & Shyr, Y. (2012). The effect of strand bias in Illumina short-read sequencing data. *BMC genomics*, 13, 1-11.
- Iamartino, D., Williams, J. L., Sonstegard, T., Reecy, J., Tassell, C. v., Nicolazzi, E. L., de Oliveira, D. A. (2013). The buffalo genome and the application of genomics in animal management and improvement. *Buffalo Bulletin*, 32(Special Issue 1), 151-158.
- Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Antonacci, F. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, 453(7191), 56-64.
- Kijas, J. W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S., & Whan, V. (2011). Analysis of copy number variants in the cattle genome. *Gene*, 482(1-2), 73-77.
- Le Roex, N., Noyes, H., Brass, A., Bradley, D. G., Kemp, S. J., Kay, S., Hoal, E. G. (2012). Novel SNP discovery in African buffalo, *Syncerus caffer*, using high-throughput sequencing. *PLoS one*, 7(11), e48792.
- Lei, C., Zhang, W., Chen, H., Lu, F., Liu, R., Yang, X., Lu, Z. (2007). Independent maternal origin of Chinese swamp buffalo (*Bubalus bubalis*). *Animal genetics*, 38(2), 97-102.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *bioinformatics*, 25(16), 2078-2079.
- Maghsoodi, S. M., MIRAEI, A. S. R., Banabazi, M. H., & MEHRABANI, Y. H. (2011). Polymorphism of prion protein gene (PRNP) in Iranian Holstein and two local cattle populations (Golpayegani and Sistani) of Iran.
- Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198-203.
- Pavlopoulos, G. A., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., Iliopoulos, I. (2013). Unraveling genomic variation from next generation sequencing data. *BioData mining*, 6, 1-25.
- Perucatti, A., Genuardo, V., Iannuzzi, A., Rebl, A., Di Bernardino, D., Goldammer, T., & Iannuzzi, L. (2012). Advanced comparative cytogenetic analysis of X chromosomes in river buffalo, cattle, sheep, and human. *Chromosome research*, 20, 413-425.
- Shirasawa, K., Fukuoka, H., Matsunaga, H., Kobayashi, Y., Kobayashi, I., Hirakawa, H., Tabata, S. (2013). Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. *DNA research*, 20(6), 593-

603.

- Stafuzza, N., Ianella, P., Miziara, M., Agarwala, R., Schäffer, A., Riggs, P., Amaral, M. (2007). Preliminary radiation hybrid map for river buffalo chromosome 6 and comparison to bovine chromosome 3. *Animal genetics*, 38(4), 406-409.
- Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., Pinkel, D. (2005). Fine-scale structural variation of the human genome. *Nature genetics*, 37(7), 727-732.
- Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Federhen, S. (2007). Database resources of the national center for biotechnology information. *Nucleic acids research*, 35(suppl_1), D5-D12.
- Zhang, L., Jia, S., Yang, M., Xu, Y., Li, C., Sun, J., Zhou, Y. (2014). Detection of copy number variations and their effects in Chinese bulls. *BMC genomics*, 15, 1-9.