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

Identification of genome diversity in Lari chicken using whole genome sequencing method

Hamideh Bazgir
MSc. of Genetics and Animal Breeding, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran and Yang Researchers Society, Shahid Bahonar University of Kerman, Kerman, Iran. Email: hamideh.bazgir@yahoo.com

Ali Esmailizadeh
*Corresponding author. Professor, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran. E-mail: aliesmaili@uk.ac.ir

Zeinab Amiri Ghanatsaman
Research Assistant Professor from Animal Science Research Department, Fars Agricultural and Natural Resources research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shiraz, Iran. E-mail: zeynabamiri1237@gmail.com

Masoud Asadi Fozi
Professor, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran. E-mail: masadi@uk.ac.ir

Abstract

Objective

Evaluation and conservation of native chickens as valuable genomic resources is essential. This is the first study for discovering variants in Lari chicken by whole genome sequencing data. The study of genetic diversity of Lari chicken at genomic level can provide useful information for its preservation and breeding. In this study, genomic diversity of five Lari individuals was investigated using whole genome sequencing technique.

Materials and methods

Blood samples were taken from five Lari chickens from cites of Shiraz and Zabol, Iran. Whole genome sequencing (paired end sequencing) was done by Illumina Company (Hiseq 2500). Data quality was determined by FastQC program. Whole genome sequencing data were aligned with chicken genome reference (Gallus_gallus-5.0/galGal5) using MEM algorithm applied in burrows wheeler aligner program (BWA). Processing of bam files was done in several steps. PCR

duplicates were removed using Picard program. The Percentage of alignment with the reference genome and coverage or depth were calculated using the flagstat and depth commands in samtools software. Single nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELs) were identified by the genomic analysis toolkit (GATK) program. Annotation of SNPs and Indels was done using SnpEff program. Genetic diversity of five chicken genomes was calculated with VCFtools.

Results

The mean percentage mapping of short sequences with the reference genome was 99.85% and the mean coverage depth was 7.65 X. In this study, 9.8 million SNPs and 10 million Indels were identified with the most counts of them in the intron and intergenic regions. The mean of observed and expected heterozygosity percentages for SNPs in five chicken genomes were 0.30 and 0.35, respectively.

Conclusions

Results from annotation showed that percentage of silent SNPs (74.38%) is higher than that nonsynomous SNPs (missense and nonsense, 25.62%) in Lari chicken genome. The lower observed genetic diversity than the expected genetic diversity, can be due to the forces such as inbreeding in the population of Lari chicken. The information provided herein can be useful for breed conservation and breeding programs and population structure survey.

Keywords: Indels, Lari chicken, SNPs, whole genome sequencing.

Paper Type: Research Paper.

Citation: Bazgir H, Esmailizadeh A, Amiri Ghanatsaman Z, Asadi Fozi M (2021) Identification of genome diversity in Lari chicken using whole genome sequencing method. *Agricultural Biotechnology Journal* 13 (2), 189-204.

Agricultural Biotechnology Journal 13 (2), 189-204. DOI: 10.22103/jab.2021.17102.1287

Received: April 20, 2021. Accepted: June 4, 2021.

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 تنوع ژنوم در مرغ لاری با استفاده از روش توالی‌یابی کل ژنوم

حمیده بازگیر

دانش آموخته کارشناسی ارشد ژنتیک و اصلاح نژاد دام، بخش مهندسی علوم دامی، دانشکدهٔ کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران و عضو انجمن پژوهشگران جوان، دانشگاه شهید باهنر کرمان، رایانامه: hamideh.bazgir@yahoo.com

علی اسمعیلی زاده کشکوئیه

*نویسنده مسئول: استاد بخش مهندسی علوم دامی، دانشکدهٔ کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران. رایانامه: aliesmaili@uk.ac.ir

زینب امیری قنات سامان

استادیار پژوهشی بخش تحقیقات علوم دامی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس، سازمان تحقیقات، آموزش و ترویج کشاورزی، شیراز، ایران، رایانامه: zeynabamiri1237@gmail.com

مسعود اسدی فوزی

استاد بخش مهندسی علوم دامی، دانشکدهٔ کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران. رایانامه:masadi@uk.ac.ir

تاریخ دریافت: ۱۴۰۰/۰۱/۳۱ تاریخ پذیرش: ۱۴۰۰/۰۳/۱۴

چکیده

هدف: ارزیابی و حفاظت از مرغان بومی به عنوان ذخایر ژنتیکی ارزشمند ضروری است. مطالعه حاضر اولین پژوهش برای شناسایی واریانت‌های ژنتیکی در مرغ لاری با اطلاعات توالی‌یابی کل ژنوم است. مطالعه تنوع ژنتیکی مرغ لاری در سطح ژنوم، می‌توانند اطلاعات مفیدی را در جهت حفظ و اصلاح نژاد آن فراهم سازد. در این تحقیق تنوع ژنومی پنج قطعه مرغ از نژادلاری با استفاده از تکنیک توالی‌یابی کل ژنوم بررسی شد.

مواد و روش‌ها: نمونه خون پنج قطعه مرغ بومی لاری از شهرهای شیراز و زابل گرفته شد. توالی‌یابی کل ژنوم به صورت رفت و برگشتی توسط شرکت ایلومینا Hiseq 2500 در کشور چین انجام شد. کیفیت داده‌ها توسط برنامه FastQC بررسی شدند. داده‌ها به وسیله الگوریتم MEM به کار برده شده در برنامهٔ BWA با ژنوم مرجع (Gallus_gallus-5.0/galGal) همردیف شدند. پردارش bam فایل‌ها در چندین مرحله انجام شد. PCR duplicates توسط برنامه Picard حذف شدند. درصد همردیفی با ژنوم مرجع و کاوریج یا عمق پوشش با استفاده از دستورات flagstat و depth به کار برده شده در نرم افزار samtools

محاسبه شدند. چندریختی‌های تک‌نوکلئوتیدی (SNPs) و حذف و اضافه‌های کوچک ژنوم با برنامهٔ GATK شناسایی شدند. مستندسازی چندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک ژنوم با برنامهٔ SnpEff انجام شد. تنوع ژنتیکی ژنوم پنج مرغ با برنامه VCFtools محاسبه شد.

نتایج: میانگین درصد همردیفی توالی‌های کوتاه با ژنوم مرجع ۸۵/ ۹۹ درصد بود و میانگین عمق پوشش ۷/۶۵ X بود. در این پژوهش ۹۸۵۱۷۳۱ چندریختی تک نوکلئوتیدی و ۱۰۲۴۱۳۹ حذف و اضافه کوچک بدست آمد که بیشترین مقدار آن در نواحی اینترون و بین ژنی مشاهده شد. میانگین درصد هتروزیگوسیتی مشاهده شده و مورد انتظار برای جایگاه‌های چندریختی‌های تک‌نوکلئوتیدی شناسایی شده برای ژنوم پنج مرغ به ترتیب ۰/۳۰ و ۰/۳۵ بود.

نتیجه‌گیری: نتایج مستندسازی نشان داد که درصد چندریختی‌های تک نوکلئوتیدی خاموش (۷۴٪/۳۸) بیشتر از درصد چندریختی‌های تک نوکلئوتیدی غیرمترادف (بد معنی و بی معنی، ۲۵/۶۲٪) در ژنوم مرغ است. کمتر بودن تنوع ژنتیکی مشاهده شده از تنوع ژنتیکی مورد انتظار، به وجود نیروهایی مثل همخونی در جمعیت مرغ لاری می‌توان اشاره کرد. اطلاعات بدست آمده از این پژوهش، می‌تواند برای برنامه‌های حفاظت و اصلاح نژادی و نیز بررسی ساختار جمعیتی سودمند واقع شوند.

کلیدواژه‌ها: توالی‌یابی کل ژنوم، چند ریختی‌های تک‌نوکلئوتیدی، حذف و اضافه‌های کوچک، مرغ لاری.

نوع مقاله: پژوهشی.

استناد: بازگیر حمیده، اسمعیلی زاده کشکوئیه علی، امیری قنات سامان زینب، اسدی فوزی مسعود (۱۴۰۰) شناسایی تنوع ژنوم در مرغ لاری با استفاده از روش توالی‌یابی کل ژنوم. *مجله بیوتکنولوژی کشاورزی*، ۱۳(۲)، ۲۰۴-۱۸۹.

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

© the authors

مقدمه

تنوع زیستی بستر توسعه امنیت غذایی و توسعه پایدار است و اهمیت آن در کشاورزی، سلامت انسان، تجارت، صنعت و فرهنگ کاملاً شناخته شده است. بروز تغییرات سریع در شرایط محیطی زمین و ناتوانی انسان در پیش بینی شرایط اقتصادی، کشاورزی و زیستی حاکم در آینده، ارزش تنوع زیستی را غیرقابل برآورد می‌سازد (Groeneveld et al. 2010). کشور ایران از دیرباز محل نگهداری گونه‌های متنوعی از دام و طیور بوده است. در این میان، اکوتیپ‌های متفاوتی از مرغان بومی نیز در مناطق جغرافیایی مختلف کشور دیده می‌شوند که به لحاظ ویژگی‌های ظاهری، تولیدی و اجتماعی تنوع زیادی دارند. این اکوتیپ‌ها نقش مهمی را در

اقتصاد خانوارهای روستایی به عهده داشته‌اند و تامین‌کننده بخشی از نیازهای کشور به تولیدات دامی بوده‌اند. در سال‌های اخیر عواملی نظیر از بین رفتن عرصه‌های طبیعی نگهداری دام و طیور، خشکسالی، عدم اقتصادی بودن نگهداری اکوتیپ‌های بومی، تغییر الگوی زندگی در روستاها و آلودگی‌های زیست محیطی موجب کاهش نگران کننده جمعیت این اکوتیپ‌ها در مناطق مختلف کشور شده است. تداوم کاهش جمعیت در برخی از این اکوتیپ‌ها به شکلی بوده است که آنها را با خطر جدی انقراض مواجه ساخته است. از این رو به نظر می‌رسد شناسایی ویژگی‌های این ذخایر ژنتیکی و توسعه برنامه‌هایی جهت بهره‌برداری مناسب و حفاظت از آنها ضروری باشد (Bazgir, 2017). متاسفانه آمار و اطلاعات چندانی در ارتباط با وضعیت خطر در این اکوتیپ‌ها در دسترس نیست. از جمله مشکلاتی که در این زمینه وجود دارد می‌توان به فقدان سیستم رکوردبرداری مناسب، عدم وجود شجره و آمیختگی‌های کنترل‌نشده اشاره کرد. سطح تمایز این جمعیت‌ها با یکدیگر چندان مشخص نیست و تنوع ژنتیکی داخل و بین این اکوتیپ‌ها نیز تا حدود زیادی نامشخص است. استفاده از فناوری‌های نوین ژنتیک مولکولی، به ویژه در شرایطی که اطلاعاتی نظیر شجره حیوانات، سطوح تفاوت‌های ژنتیکی داخل و بین جمعیت‌ها و میزان خلوص و یا آمیختگی آنها در دسترس نیست، می‌تواند به شناسایی خصوصیات ژنتیکی جمعیت‌ها و برنامه‌ریزی جهت حفاظت از آنها کمک شایان توجهی کند (Engelsma et al. 2011). از سوی دیگر، از آنجا که وجود واریانس ژنتیکی پیش زمینه اصلی انجام انتخاب در دام‌های اهلی محسوب می‌شود، شناسایی ساختارهای ژنومی در این اکوتیپ‌ها گام موثری در شناسایی پتانسیل‌های تولیدی این اکوتیپ‌ها خواهد بود (Zhang et al. 2012).

مرغ اهلی (*Gallus gallus domesticus*) اولین پرنده اهلی شده توسط انسان است و در بین گونه‌های پرندگان بیشترین جمعیت را به خود اختصاص داده است. مرغ علاوه بر آنکه از گونه‌های اهلی دارای اهمیت اقتصادی است، در مطالعات انسانی نیز به فراوانی به عنوان یک مدل بیولوژیک ارزشمند به کار گرفته شده است (Dohner, 2001). واریانت‌های تک نوکلئوتیدی[۱] (SNVs)، حذف و اضافه‌های کوتاه، تنوع در تعداد نسخه‌ها[۲] (CNVs) و تنوعات ساختاری بزرگ[۳] (SVs) اشکال مختلفی از تنوع ژنومی را تشکیل داده‌اند و محدوده تغییرات آنها از یک جفت باز منفرد تا جایگزینی‌های بزرگ در حد کروموزوم است (Alkan et al, 2011). SNPs فراوان‌ترین منبع تنوع ژنتیکی داخل ژنوم هستند و با تفاوت‌های ارثی بین افراد مرتبط هستند و نسبت به نشانگرهای دیگر از ثبات بیشتری (به دلیل میزان جهش کمتر) برخوردارند (Gray et al. 2000; Suh and Vijg 2005). مطالعات اخیر از چندریختی‌های تک نوکلئوتیدی در مطالعات بررسی ساختار جمعیتی (Karimi et al. 2016; Ruiz-Larrañagaet al. 2020; Deng et al. 2020)، تنوع ژنتیکی (Amiri Ghanatsaman et al. 2016., Li et al. 2019; Weldenegodguad et al. 2019) و بررسی نواحی تحت انتخاب (Asadollahpour et al. 2020., Nosrati et al. 2019., Moradian

[۱] Single nucleotide variants
[۲] Copy number variants
[۳] Structural variants

al. 2020) استفاده کرده‌اند. حذف و درج (ایندل) یکی از شکل‌های اصلی تنوع ژنتیکی هستند که از نظر وقوع در ژنوم در رتبه دوم بعد از چند ریختی‌های تک نوکلئوتیدی قرار دارند. در طیور تعدادی از صفات از قبیل رشد پر و کوتولگی وابسته[۴] به جنس با ایندل‌ها مرتبط هستند (Amiri Ghanatsaman et al. 2019). روش‌های قدیمی تر برای شناسایی واریانت‌ها مانند روش‌های سیتوژنتیکی از تکنیک‌هایی مانند تهیه کاریوتیپ و FISH (Fluorescence in situ hybridization) بهره می‌بردند که به دلیل پوشش محدود ژنوم و قدرت تفکیک پایین، توانایی لازم را برای شناسایی واریانت‌ها و به خصوص جهش‌های نادر و جدید را نداشتند (Buysse et al. 2009). اما امروزه فناوری NGS (Next genertion sequencing) با مزیت‌هایی مانند پوشش گسترده تر ژنوم، قابلیت تفکیک بالاتر، زمان کم توالی‌یابی، قیمت مناسب، انقلابی را در زمینه تحقیقات بیولوژیک بوجود آورده است، به طوری که فناوری NGS می‌تواند در مدت زمان محدودی با ارائه صدها میلیون خوانش از یک ژنوم، تصویری آشکاری از خصوصیات ژنومی به نمایش گزارد (Metzker, 2010). وجود ژنوم مرجع با کیفیت بالا، امکان همردیفی توالی‌های ژنومیک افراد با این منبع ژنتیکی و شناسایی تنوع‌های نوکلئوتیدی را فراهم ساخته است (Kerstens et al. 2009; Li and Durbin 2009). با توجه به وراثت پذیر بودن و فراونی بالای واریانت‌ها، این احتمال وجود دارد که برخی از واریانت‌ها با ژن‌های اصلی تولید و سلامتی دام‌ها در ارتباط باشند (Kharrati Koopaee and, Esmailizadeh 2014). مطالعات مختلفی با استفاده از پلت‌فرمهای NGS برای شناسایی چندریختی‌های تک‌نوکلئوتیدی و ایندل‌ها در ژنوم حیوانات و گیاهان انجام شده است (Kilian et al. 2012; Ramos et al. 2009; Stothard et al. 2011; You et al. 2011; Zeng et al. 2019) اگر چه مطالعات ملکولی متعددی روی طیور در ایران انجام شده است (Zandi et al. 2014; Mohammadifar et al. 2011; Mohammadifar et al. 2014; Mohammadifar and Mohammadabadi, 2017; Sohrabi et al. 2018; Moazeni et al. 2016a; Moazeni et al. 2016b; Nassiri and Roudbari, 2014)، اما تا کنون تنوع ژنومی در مرغ لاری با استفاده از روش توالی‌یابی کل ژنوم بررسی نشده است، لذا، نظر به موارد ذکر شده هدف از این مطالعه بررسی تنوع ژنوم و فراهم کردن منبعی برای کار ژنتیکی بعدی از قبیل انتخاب ژنومیک و مطالعات پویش ژنومی در مرغ لاری بود.

مواد و روش‌ها

در این پژوهش، از ۵ قطعه مرغ لاری شامل سه نر و دو ماده، از شهرهای شیراز و زابل نمونه خون تهیه شد. استخراج DNA از نمونه‌های خون با روش استخراج نمکی انجام شد (Abadi et al. 2009). جهت بررسی کیفیت DNA استخراج شده، از دو روش الکتروفورز با استفاده از ژل آگارز یک درصد به مدت نیم ساعت و همچنین دستگاه اسپکتروفتومتر استفاده شد. توالی‌یابی کل ژنوم به صورت رفت و برگشتی توسط شرکت ایلومینا و دستگاه توالی‌یاب Hiseq2500 در کشور چین انجام شد

[۴]. Dwarf Chicken

(www.berrygenomics.com). کیفیت داده‌ها توسط برنامه (http:// bbsrc /www.bioinformatics..ac.uk/projects/fastqc/) Fastqc بررسی شد. برای هم ردیفی توالی‌های کوتاه با ژنوم مرجع، در ابتدا آخرین نسخه ژنوم مرجع مرغ (http://ftp.ensembl.org/pub/release-103/fasta/gallus_gallus/dna/) از سایت Ensembl دانلود شد. ژنوم مرجع حاوی ۳۹ جفت کروموزوم است که ۳۸ جفت آن اتوزوم و یک جفت آن کروموزوم جنسی است. برای همردیفی[۵] Short Reads با ژنوم مرجع[۶] (Gallus_gallus-5.0/galGal5) از الگوریتم MEM در برنامه BWA-[۷] 0.7.17 (Li and Durbin, 2009) استفاده شد. SAM[۸] فایل‌های تولید شده با برنامه SAMtools1.9 (Li et al. 2009) به BAM فایل تبدیل شدند. سپس BAM[۹] فایل‌ها بر پایهٔ موقعیت ژنگانی با استفاده از برنامه SAMtools1.9 مرتب و ایندکس[۱۰] شدند. کیفیت BAM فایل‌ها توسط دو پارامتر درصد همردیفی Short read باژنوم مرجع و پوشش‌دهی[۱۱] بررسی شدند. به منظور محاسبه درصد همردیفی با ژنوم مرجع و کاوریج یا عمق پوشش از دستورات flagstat و depth به کار برده شده در نرم افزار Samtools استفاده شد. خطاهای PCR[۱۲] (توالی‌های کوتاه تکراری (مضاعف شده) حاصل از واکنش زنجیره‌ای پلیمراز در مرحله آماده‌سازی کتابخانه ژنومی قبل از توالی‌یابی) با برنامه Picard-1.105 (https://github.com/broadinstitute/picard) حذف شدند. برای کاهش خطاهای حاصل از مرحله هم ردیفی، یک مرحله همردیفی مجدد (Realignment) با استفاده از برنامه GATK (Mckenna et al. 2010) انجام شد به این ترتیب از دو ابزار RealignerTargetCreator و IndelRealigner در برنامه GATK استفاده شد. به منظور تصحیح خطاهای توالی‌یابی و دیگر خطاهای مربوط به آزمایش، با استفاده از ابزارهای BaseRecalibrator و PrintReads در برنامه GATK نمره کیفیت بازها مجدداً کالیبره شد. چند ریختی‌های تک نوکلئوتیدی (SNPs) و حذف و اضافه‌های کوچک (Indels Small) توسط ابزار SelectVariant در برنامه GATK3.4-46-[۱۳] gbc02625 از هم جدا شدند. به منظور فیلتر کردن چندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک از ابزار Variant Filtration در برنامه GATK استفاده شد. از برنامه snpEff (version 1.9.6) (Cingolani et al. 2012) برای مستندسازی[۱۴] چندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک ژنوم استفاده شد. تعداد هموزیگوس‌ها و هتروزیگوس‌ها در چند ریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک، و همچنین تعداد جهش‌های جا به جایی و معکوس در چند

5. Alignment
6. Genome Reference
7. Burrows-Wheeler Aligner
8. Sequence Alingnment MA
9. Binary Alingnment MA
10. Index
11. Coverage
12. Mark duplicates
13. Genome Analysis Toolkit
14 Annotation

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

ریختی‌های تک نوکلئوتیدی با برنامه snpEff محاسبه شدند. مقادیر تنوع ژنتیکی با استفاده از دستور het– به کار برده شده در برنامه vcftools-0.1.11 (Danecek et al. 2011) محاسبه شدند.

نتایج و بحث

در این مطالعه کل ژنوم پنج قطعه مرغ لاری به منظور مطالعه و شناسایی چندریختی‌های تک نوکلوتیدی وحذف واضافه‌های کوچک توالی‌یابی گردید. نتایج کنترل کیفیت داده‌ها نشـــان داد که Short Reads دارایی کیفیت بالایی و فاقد آلودگی پرایمر بودند. طول Short Reads برای همه داده‌ها ۱۲۵ بود. میانگین در صد همردیفی برای داده‌ها ۹۹/۸۵ درصد بود که نشان دهنده کـ یـ فیـت بـالای هـ مرد یـ فی داده‌هـای توا لی‌یـا بی شـــده بـا ژ نوم مر جع (Gallus_gallus-5.0/galGal5, https://www.ensembl.org/Gallus-gallus/Info/Index) بود. میزان پوشـش‌دهی توالی‌های کوتاه با ژنوم مرجع از X ۵/۵ X تا ۹/۲۵ است (جدول ۱).

جدول ۱. خروجی توالی‌یابی برای پنج نمونه

Table 1. Sequencing output for five samples

نام نمونه Sample name	محل نمونه‌گیری Location of sampling	تعداد توالی‌های کوتاه Number of short reads	طول توالی‌های کوتاه Short reads length	درصد همردیفی Alignment Percent	میانگین عمق Mean depth
Lari 16S	Shiraz	73805418	125	99/85	9/03X
Lari 17S	Shiraz	32279408	125	99/85	5/5X
Lari 18S	Shiraz	35030270	125	99/86	5/5X
Lari 19S	Zabol	41373546	125	99/85	9/24X
Lari 20S	Zabol	36806860	125	99/86	8/97 X

در این مطالعه تعداد ۹/۸ میلیون چندریختی تک‌نوکلئوتیدی از پنج ژنوم مرغ لاری بدست آمد (جدول ۲). تعداد چند ریختی‌های شناسایی شده در کل ژنوم مرغ لاری بیشتر از تعداد چند ریختی‌های شناسایی شده در کل ژنوم مرغ مرندی (Akbary et al. 2020) و مرغ بومی فارس (Eskandari et al. 2018) با استفاده از روش توالی‌یابی کل ژنوم در مطالعات قبلی است.

جدول ۲. تعداد چند ریختی‌های تک‌نوکلئوتیدی و حذف و اضافه‌های کوچک برای پنج نمونه پیش و پس از پالایش کردن

Table 2. Number of single nucleotide polymorphisms and small insertions and deletions for 5 samples before and after filtering

نام نمونه Sample name	تعداد SNPs قبل از فیلترکردن Number of raw SNPs	تعداد SNPsبعد از فیلترکردن Number of filtered SNP	تعداد INDELs Number of Indels
Lari 16S	7027395	6041282	721471
Lari 17S	6447551	5513158	648157
Lari 18S	6844185	5897963	713235
Lari 19S	7039644	6048492	725968
Lari 20S	6800284	5873138	713544
Total	12005711	9851731	1024139

تعداد چندریختی‌های تک‌نوکلئوتیدی هتروزیگوس بیشتر از تعداد چندریختی‌های تک نوکلئوتیدی هموزیگوس درنمونه‌ها بود (جدول ۳). نتایج مشابهی در آنالیز ژنوم آکوتیپ‌های مرغان بومی فارس و مرندی بدست آمد (Eskandari et al. 2018; Akbari et al. 2020).

جدول ۳. تعداد خالصی و ناخالصی

Table 3. Number of homozygous and heterozygous

نام نمونه Sample name	تعداد چند ریختی‌های هتروزیگوس Number of heterozygous SNPs	تعداد چند ریختی‌های هموزیگوس Number of homozygous SNPs	تعداد ایندل‌های هتروزیگوس Number of heterozygous Indels	تعداد ایندل‌های هموزیگوس Number of homozygous Indels
Lari 16S	3431018	2610264	355500	365971
Lari 17S	2803594	2709564	286497	361660
Lari 18S	3225432	2672531	342627	370608
Lari 19S	3402677	2645815	355558	370410
Lari 20S	3151464	2721674	337596	375948

نسبت جهش‌های انتقالی[15] به متقاطع[16] در چندریختی‌های تک نوکلئوتیدی از ۲/۳۷۹ تا۳۸۲/ ۲ متغیر است (جدول ۴). به علت مکانیسم‌های مولکولی ایجاد کننده جهش‌ها در مولکول DNA، تعداد چند ریختی‌های تک‌نوکلئوتیدی که توسط جهش‌های انتقالی (جهش‌هایی هستند که در آنها یک پورین به پورین دیگر یا پیریمیدین به پیریمیدین دیگر تبدیل می‌شود) تولید می‌شوند، در مقایسه با چند ریختی‌های تک نوکلئوتیدی که توسط جهش‌های متقاطع (پیریمیدین به پورین یا بالعکس تبدیل می‌شود) ایجاد می‌شوند، تقریباً دو برابر است (Collins and Jukes 1994). به طور کلی نسبت جهش‌های انتقالی به متقاطع یک شاخص مفیدی برای

15. Transition
16. Transversio

نشان دادن میزان خطای مثبت (FPR[۱۷]) شناسایی چندریختی‌های تک نوکلئوتیدی است (McKenna et al. 2010). مطالعات اخیر نشان داده است که نسبت ti/tv برای داده‌های کل ژنوم حدود ۲/۰ ~ ۲/۰۱ است. هر چه این نسبت بیشتر باشد نشانگر دقت بالای شناسایی چندریختی‌های تک نوکلئوتیدی در ژنوم است (Bainbridge et al. 2011; Guo et al. 2012). نسبت‌های بدست آمده در نتایج ما نشان‌دهنده کیفیت مطلوب شناسایی چندریختی‌های تک‌نوکلئوتیدی است. نتایج مستندسازی چند ریختی‌های تک نوکلئوتیدی نشان دادند که بیشتر چند ریختی‌های تک نوکلئوتیدی در نواحی بین ژنی و اینترون واقع شده‌اند. از تعداد کل چندریختی‌های تک نوکلئوتیدی ۲۹/۴۲۸، ۴۹/۱۰۲، ۱/۴۷۸، ۹/۶۳۴، ۹/۴۰۵، ۰/۴۱۷، ۰/۰۸۲ درصد از آنها به ترتیب درنواحی بین ژنی، اینترون، اگزون، بالادست، پایین دست، 3' UTR و 5' UTR قرارگرفتند (جدول۵). همچنین تعداد کل جهش‌های مترادف[۱۸] (چندریختی‌های تک نوکلئوتیدی خاموش یا Silent) بیش تر ازتعداد کل جهش‌های غیر مترادف[۱۹] (چند ریختی‌های تک نوکلئوتیدی بدمعنی وبی‌معنی) است (جدول ۶). تعداد کل ۱۰۲۴۱۳۹ حذف و اضافه کوچک در سراسر ژنوم نمونه‌ها شناسایی شد (جدول ۲). تعداد ایندل‌های شناسایی شده در کل ژنوم مرغ لاری بیشتر از تعداد ایندل‌های شناسایی شده در کل ژنوم مرغ مرندی (Akbary et al. 2020) و مرغ بومی فارس (Eskandari et al. 2018) با استفاده از روش توالی‌یابی کل ژنوم در مطالعات قبلی است.

نتایج مستندسازی حذف واضافه‌های کوچک نشان داد که بیشترحذف و اضافه‌های کوچک در نواحی بین ژنی و اینترون واقع شده‌اند و از تعداد کل حذف و اضافه‌های کوچک ۲۹/۷۸۳، ۵۱/۱۳۹، ۰/۸۹، ۸/۳۸، ۸/۹۹۳، ۰/۵۲۸، ۰/۰۴۴ درصد از آنها به ترتیب در نواحی بین‌ژنی، اینترون، اگزون، بالادست، پایین دست، 3' UTR و 5' UTR قرار گرفتند (جدول ۵). نتایج مشابهی با نتایج مستندسازی چند ریختی‌های تک نوکلئوتیدی و ایندل‌ها در گاو (Da Silva et al. 2015)، سگ (Amiri Ghanatsaman et al. 2016) و مرغ (Eskandari et al. 2018; Akbary et al. 2020) گزارش شده است. میانگین تنوع ژنتیکی مشاهده و مورد انتظار محاسبه شده برای چند ریختی‌های تک نوکلئوتیدی در ژنوم چهار قطعه مرغ مرندی ۰/۳۰ و ۰/۳۵ بود (جدول ۷). کمتر بودن تنوع ژنتیکی مشاهده شده از تنوع ژنتیکی مورد انتظار، به وجود نیروهایی مثل همخونی در جمعیت می‌توان اشاره کرد. همخونی در هر گله‌ای که بعد از چند نسل جوجه کشی می‌شود اتفاق می‌افتد همخونی در گله‌های مرغ لاری بیشتر اتفاق می‌افتد چون تعداد مرغ و خروس لاری محدود می‌باشد برای جلوگیری از این مشکل ژنتیکی باید حتی الامکان در هر دوره جوجه کشی خروس‌ها تعویض شوند و خروس‌های جدید از نسل جدید وارد گله شوند تا از همخونی جلوگیری شود.

17. False positive ratio
18. synonymous
19 nonsynonymous

جدول ٤. تعداد جهش‌های انتقالی و متقاطع در چند ریختی‌های تک نوکلئوتیدی

Table 4. Number of transition and transversion mutations in SNPs

نام نمونه Sample name	تعداد جهش‌های جا به جایی Number of transition mutations	تعداد جهش‌های معکوس Number of transversion mutations	نسبت جهش‌های جا به جایی به معکوس Transition to transversion ratio
Lari 16S	6060175	2543712	2.382
Lari 17S	5758826	2418403	2.381
Lari 18S	6000617	2522562	2.379
Lari 19S	6087726	2558902	2.379
Lari 20S	6018880	2528519	2.380

جدول ٥. تعداد اثر گذاری‌های چندریختی‌های تک نوکلئوتیدی وحذف و اضافه‌های کوچک در نواحی مختلف ژنوم

Table 5. Number of single nucleotide polymorphisms and small insertions and deletions effects in different regions of genome

نوع Type	چند ریختی‌های تک نوکلئوتیدی SNPs	حذف و اضافه‌های کوچک Indels
پایین دست Downstream	1642019 (9.405%	167882 (8.993%)
اگزون Exonic	305201 (1.748%)	16622 (0.89%)
بین ژنی Intergenic	5137777 (29.428%)	555976 (29.783%)
اینترون Intronic	8573921 (49.109%)	954640 (51.139%)
Splice_site_acceptor	579 (0.003%)	211 (0.011%)
Splice_site_donor	706 (0.004%)	168 (0.009)
Splice_site_region	29457 (0.169%)	4097 (0.219%)
بالادست Upstream	1681983 (9.634%)	156428 (8.38%)
utr '3	72864 (0.417%)	9861 (0.528%)
utr '5	14324 (0.082%)	818 (0.044%)

جدول ٦. تعداد اثر گذاری‌های چندریختی تک نوکلئوتیدی با دسته عملکردی

Table 6. Number of SNP effects by functional class

بد معنی (درصد) Missense (%)	بی معنی (درصد) Nonsense (%)	خاموش (درصد) Silent (%)
46605 (25.478 %)	259 (0.142 %)	136055 (74.38 %)

جدول ۷. مقادیر تنوع ژنتیکی محاسبه شده برای چندریختی‌های تک نوکلئوتیدی

Table 7. Calculated genetic diversity values for SNPs

نام نمونه Sample name	ضریب همخونی inbreeding coefficient	درصد هترزیگوسیتی مشاهده شده Proportion of the Observed Heterozygous Sites	درصد هتروزیگوسیتی مورد انتظار Proportion of the Expected Heterozygous Sites
Lari 16S	0.07359	0.328478077	0.354573
Lari 17S	0.10269	0.318111994	0.354518
Lari 18S	0.19698	0.284599836	0.354412
Lari 19S	0.16581	0.295799724	0.354456
Lari 20S	0.12269	0.300111994	0.354500
Mean	0.132352	0.30542	0.354492

نتیجه‌گیری: ژنوم پنج قطعه مرغ از نژاد لاری با اســتفاده از تکنیک توالی‌یابی کل ژنوم با میانگین عمق پوشــش **X**۷/۶۵، برای اولین بار توالی‌یابی شد. تعداد ۹/۸ میلیون چندریختی تک نوکلئوتیدی و ۱۰/۲ میلیون حذف و اضافه کوچک از ژنوم ۵ قطعه مرغ لاری استخراج شد. بسیاری از این واریانت‌های ژنومی با صفات اقتصادی مرتبط هستند که می‌توانند در پروژه‌های بعدی شــناســایی، و در برنامه‌های اصــلاح نژادی طیور بکار روند. کمتر بودن تنوع ژنتیکی مشــاهده شــده از تنوع ژنتیکی مورد انتظار، نشان‌دهنده وجود نیروهایی مثل همخونی در جمعیت مرغ لاری است.

سپاسگزاری: از معاونت محترم پژوهشی دانشگاه شهید باهنر کرمان بخاطر تامین بخشی از هزینه مالی این تحقیق در قالب گرانت پژوهشــی به نگارنده مســئول مقاله تشــکر و قدردانی می‌گردد. با تشــکر فراوان از فعالان بخش خصــوصــی و پرورش‌دهندگان مرغ لاری که مشورت و کمک زیادی در تهیه نمونه مناسب نمودند. همچنین نگارندگان مقاله از آقای دکتر حامد خراتی کوپایی که در تهیه نمونه و اســتخراج دی ان. ای و توالی‌یابی داده‌ها کمک زیادی نمودند قدردانی می‌نمایند. از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

منابع

اسکندری طاهره، اسمعیلی زاده کشکوئیه علی، اسدی فوزی مسعود (۱۳۹۷) شناسایی نشانگرهای تک نوکلئوتیدی در مرغ بومی فارس با استفاده از روش توالی‌یابی کل ژنوم. مجله بیو تکنولوژی کشاورزی ۱۰ (۱)، ۱۵۱-۱۳۹.

اکبری رسول، اسمعیلی زاده کشکوئیه علی، امیری قنات سامان زینب، آیت اللهی مهجردی احمد (۱۳۹۹) شناسایی تنوع ژنوم در مرغ مرندی با استفاده از روش توالی‌یابی کل ژنوم. مجله بیو تکنولوژی کشاورزی ۱۲ (۱)، ۱۶۱-۱۷۶.

امیری قنات سامان زینب، اسمعیلی زاده کشکوئیه علی، اسدی فوزی مسعود (۱۳۹۵) بررسی تنوع ساختاری ژنگان سگ و گرگ بومی ایران با روش توالی‌یابی کل ژنوم. مجله علوم دامی ایران ۴۷ (۲)، ۲۷۱-۲۷۷.

امیری قنات سامان زینب، اسمعیلی زاده کشکوئیه علی، اسدی فوزی مسعود (۱۳۹۸) شناسایی ایندل‌ها در ژنوم سگ و گرگ بومی ایران با روش توالی‌یابی کل ژنوم. مجله ژنتیک نوین ۱۴ (۱)، ۸۸-۸۵.

بازگیر حمیده (۱۳۹۶) نقشه یابی نشانگرهای تک نوکلئوتیدی در مرغ نژاد لاری با استفاده از توالی یابی ژنوم. پایان نامه کارشناسی ارشد، دانشگاه شهید باهنر کرمان. ۲۰-۱۵.

نصیری محمد رضا، رودباری زهرا (۱۳۹۳) تجزیه و تحلیل ناحیه سیتوکروم b در مرغ بومی خراسان. مجله بیوتکنولوژی کشاورزی ۶(۲)، ۱۹۸-۱۸۹.

محمدی فر آمنه، فقیه ایمانی سید علی، محمد آبادی محمد رضا، سفلایی محمد (۱۳۹۲) تاثیر ژن TGFB3 بر ارزش‌های فنوتیپی و ارثی صفات وزن بدن در مرغ بومی فارس. مجله بیوتکنولوژی کشاورزی ۵ (۴)، ۱۲۵-۱۳۶.

References

Abadi MRM, Askari N, Baghizadeh A, Esmailizadeh AK (2009) A directed search around caprine candidate loci provided evidence for microsatellites linkage to growth and cashmere yield in Rayini goats. Small Rumin Res 81, 146-151.

Akbary R, Esmaeelizadeh A, Amiri Ghanatsaman Z, Ayetollahi Mehrjerdi A (2020) Identification of genome diversity in marandi chicken using whole genome sequencing method. Agric Biotech J 12,161-176 (In Persian).

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping.. Nat Rev Genet 12, 363-376.

Amiri Ghanatsaman Z, Esmailizadeh Koshkoiyeh A, Asadi Fozi M (2016) Study of structural diversity of genome Iranian native dog and wolf with the method whole genome sequencing. Iran J Anim Sci 47, 271-277 (In Persian).

Amiri Ghanatsaman Z, Esmailizadeh Koshkoiyeh A, Asadi Fozi M (2019) Detection of deletions and insertions in genome of Iranian dogs and wolves with the method whole genome sequencing. Mod Genet 14, 85-88 (In Persian).

Asadollahpour Nanaei H, Dehghani Qanatqestani M, Esmailizadeh A (2020) Whole-genome resequencing reveals selection signatures associated with milk production traits in African Kenana dairy zebu cattle. Genomics 112, 880-5.

Bainbridge MN, Wang M, Wu Y et al. (2011) Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol 12, 68.

Bazgir H (2017) Mapping of single nucleotide markers in Lari chickens using genome sequencing. Master of Science Thesis, Shahid Bahonar University of Kerman. Kerman, Iran. Pp. 15-20 (In Persian).

Buysse K, Delle Chiaie B, Van Coster R et al. (2009) Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. Eur J Med Genet 52, 398-403.

Cingolani P, Platts A, Wang LL et al. (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, 80-92.

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

Da Silva JM, Giachetto PF, da Silva LOC et al. (2015) Genomic variants revealed by invariably missing genotypes in Nelore cattle. Plos One 10, 0136035.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and vcftools. Bioinformatic 27, 2156-2158.

Deng J, Xie XL, Wang DF et al. (2020) Paternal origins and migratory episodes of domestic sheep. Curr Biol 30, 4085-4095.

Dohner, J.V. (2001) The encyclopedia of endangered livestock and poultrybreeds. New Haven (CT): Yale University Press.

Engelsma KA, Veerkamp RF, Calus MPL, Windig JJ (2011) Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. J Anim Breed Genet 128, 473–481.

Eskandari T, Esmailizadeh AK, Mohammadabadi MR, Sohrabi S (2018) Identification of single nucleotide polymorphisms in Fars native chicken using whole genome sequencing data. Agric Biotech J 10, 139-151 (In Persian).

Gray IC, Campbell DA, Spurr NK (2000) Single nucleotide polymorphisms as tools in human genetics. Hum Mol Genet 9, 2403-2408.

Groeneveld LF, Lenstra JA, Eding H et al. (2010) Genetic diversity in farm animals- a review. Anim Genet 41, 6-31.

Guo Y, Li J, Li CI et al. (2012) The effect of strand bias in Illumina short-read sequencing data. BMC Genomics 13(1), 666.

Karimi K, Strucken EM, Moghaddar N et al.(2016) Local and global patterns of admixture and population structure in Iranian native cattle. BMC Genet, 17:108.

Kerstens HHD, Crooijmans RPMA, Veenendaal A et al. (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. BMC Genomics, 10: 479-10.

Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. Brief Funct Genom. 11, 38-50.

Li D, Li Y, Li M et al. (2019) Population genomics identifies patterns of genetic diversity and selection in chicken. BMC genomics 20, 263.

Li G, Ma L, Song C et al. (2009) The YH database: the first Asian diploid genome database. Nucleic Acids Res. 37, 1025-1028.

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25, 1754-1760.

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

McKenna A, Hanna M, Banks E et al. (2010) The genome analysis toolkit, a mapreduce frame work for analyzing next-generation dna sequencing data. Genome Res 20, 1297-1303.

Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11, 31-46.

Moazeni S, Mohammadabadi MR, Sadeghi M et al. (2016a) Association between UCP Gene Polymorphisms and Growth, Breeding Value of Growth and Reproductive Traits in Mazandaran Indigenous Chicken. Open J Anim Sci 6, 1-8.

Moazeni SM, Mohammadabadi MR, Sadeghi M et al. (2016b) Association of the melanocortin-3(MC3R) receptor gene with growth and reproductive traits in Mazandaran indigenous chicken. J Livest Sci Technol 4, 51-56.

Mohammadifar A, Faghih Imani SA, Mohammadabadi MR, Soflaei M (2014) The effect of TGFb3 gene on phenotypic and breeding values of body weight traits in Fars native fowls. Agric Biotech J 5, 125-136 (In Persian).

Mohammadifar A, Mohammadabadi MR **(**2017**)** The Effect of Uncoupling Protein Polymorphisms on Growth, Breeding Value of Growth and Reproductive Traits in the Fars Indigenous Chicken. Iran. J Appl Anim Sci 7, 679-685.

Moradian H, Koshkoiyeh AE, Mohammadabadi M, Fozi MA (2020) Whole genome detection of recent selection signatures in Sarabi cattle: a unique Iranian taurine breed. Genes Genom 42(2), 203-215.

Nassiri MR, Roudbari Z (2014) Genetic analysis of cytochrome b region in native chicken of Khorasan. Agric Biotech J 6,189-198 (In Persian).

Nosrati M, Asadollahpour Nanaei H, Amiri Ghanatsaman Z, Esmailizadeh A (2019) Whole genome sequence analysis to detect signatures of positive selection for high fecundity in sheep. Reprod Domest Anim 54, 358-64.

Ramos AM, Crooijmans RPMA, Affara NA et al. (2009) Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. PLoS One 4, e6524.

Ruiz-Larrañaga O, Nanaei HA, Montes I et al. (2020) Genetic structure of Iranian indigenous sheep breeds: insights for conservation. Trop Anim Health Prod 52, 2283-2290.

Sohrabi SS, Mohammadabadi MR, Wu DD, Esmailizadeh A (2018) Detection of breed-specific copy number variations in domestic chicken genome. Genome 61, 7-14.

Stothard P, Choi J-W, Basu U et al. (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12, 559-10.

Suh Y, Vijg J (2005) SNP discovery in associating genetic variation with human disease phenotypes. Mutat Res 573, 41-53.

Weldenegodguad M, Popov R, Pokharel K et al. (2019) Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. Front Genet 9, 728.

You F, Huo N, Deal K et al. (2011) Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. BMC Genomics 12, 59.

Zandi E, Mohammadabadi MR, Ezzatkhah M, Esmailizadeh AK (2014) Typing of Toxigenic Isolates of Clostridium Perfringens by Multiplex PCR in Ostrich. Iran J Appl Anim Sci 4, 509-514.

Zeng L, Tu XL, Dai H et al. (2019) Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. Genome Biol 20, 1-13.

Zhang H, Wang SZ, Wang ZP et al. (2012) A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genet 13, 704.