Archive of SID

مجله بیوتکنولوژی کشاورزی

علمی-پژوهشی و ISC

# شناسایی نشانگرهای تک نوکلئوتیدی در مرغ بومی فارس با استفاده از روش توالی یابی کل ژنوم

**طاهره اسکندری[1]، علی اسمعیلی زاده کشکوئیه*[2]، محمدرضا محمدآبادی[2]، سعید سهرابی[3]**

[1] دانش‌آموخته کارشناسی ارشد اصلاح نژاد، بخش مهندسی علوم دامی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران.

[2] استاد بخش علوم دامی دانشگاه شهید باهنر کرمان، کرمان، ایران.

[3] دانشجوی دکتری بخش علوم دامی، دانشکده کشاورزی و عضو انجمن پژوهشگران جوان، دانشگاه شهید باهنر کرمان، کرمان، ایران.

**تاریخ دریافت: 1397/01/26، تاریخ پذیرش: 1397/04/04**

## چکیده

شناسایی تنوعات ساختاری مسئول تفاوتهای فنوتیپی در حیوانات اهلی از نظر اقتصادی اهمیت ویژه ای دارد. یکی از دقیق ترین و جدیدترین رویکردهای مورد استفاده در شناسایی این تنوعات، استفاده از تکنیک های توالی یابی نسل جدید است. در پژوهش حاضر، ژنوم مرغ بومی فارس به منظور شناسایی چندشکلی های تک نوکلئوتیدی و حذف و اضافه های کوتاه، با استفاده از داده های توالی‌یابی کل ژنوم مورد بررسی قرار گرفت. نمونه های مورد بررسی توسط سیستم توالی یابی شرکت ایلومینا (Hiseq 2000) مورد تعیین توالی قرار گرفت. پس از بررسی کیفیت داده ها توسط نرم افزار FastQC، همردیفی توالی های کوتاه با ژنوم مرجع با برنامه BWA انجام و سپس چند ریختی‌های تک نوکلئوتیدی و حذف و اضافه های کوچک به کمک برنامه GATK مشخص شدند. درصد همردیفی توالی‌های کوتاه با ژنوم مرجع بالای 99 درصد بود. در این مطالعه 8858153 چند ریختی تک نوکلئوتیدی و 895378 حذف و اضافه کوچک بدست آمد که کمترین مقدار آن در ناحیه اگزونی مشاهده شد. نتایج نشان داد چندریختی‌های تک‌نوکلئوتیدی هتروزیگوس فراوانی بیشتری نسبت به چند ریختی های هموزیگوس دارند که این نشان‌دهنده تنوع زیاد جمعیت مورد بررسی است. همچنین باتوجه به پیشینه چند هزارساله اهلی شدن مرغ در ایران، می‌توان انتظار داشت که تطابق پذیری با محیط در جمعیت مورد مطالعه اتفاق افتاده باشد و در نتیجه، تغییرات ژنومی در جهت سازگاری با محیط باعث ایجاد تنوع در جمعیت شده است.

**واژه‌های کلیدی:** *توالی‌یابی ژنوم، چندریختی‌های تک نوکلئوتیدی، حذف و اضافه کوچک، تنوع جمعیت.*

* نویسنده مسئول: علی اسمعیلی زاده تلفن: 09133958041 Email: aliesmaili@uk.ac.ir

www.SID.ir

Archive of SID


**مقدمه**

کشف و تعیین خصوصیات تنوعات بین و درون افراد می تواند منجر به شناسایی تفاوت های ژنتیکی عامل تنوع فنوتیپی شود. تا امروز روش ها و فناوری های مختلفی برای شناسایی این تنوعات ژنتیکی مورد استفاده قرار گرفته که روش های مبتنی بر فناوری های نسل جدید توالی یابی (NGS) به علت قدرت تفکیک و دقت بیشتری که نسبت به سایر روش ها در شناسایی تنوعات ژنتیکی دارند (Futschik and Schlötterer, 2010)، در چند سال گذشته به طور گسترده ای مورد استفاده قرار گرفته اند. تاکنون چندین نوع از تنوعات ژنتیکی در ژنوم موجودات مختلف شناسایی شده است که فراوانی دو گروه از این تنوعات یعنی چندشکلی های تک نوکلئوتیدی (SNP) و حذف و اضافه های کوچک (Indel) در ژنوم از بقیه بیشتر است (Mullaney *et al.*, 2010). با توجه به توزیع گسترده چندشکلی های تک نوکلئوتیدی در سراسر ژنوم، این نوع از تنوعات ژنومی کاربرد بسیار زیادی در مطالعات مختلف ژنومی ازجمله پویش های ژنومی برای شناسایی جایگاه های موثر بر خصوصیات مختلف ژنتیکی دارند و پس از آن، حذف و اضافه های کوچک از نظر تراکم در ژنوم در مرتبه دوم قرار دارند (Mills *et al.*, 2011). از آنجاکه حذف و اضافه های کوچک در مقایسه با چندشکلی های تک نوکلئوتیدی تعداد بیشتری از نوکلئوتیدهای ژنوم را در بر می گیرند، می توانند تاثیر بیشتری در عملکرد و بیان برخی ژن ها داشته باشند. بعنوان مثال، حذف و اضافه های موجود در ژن *PMEL17* مرغ مسئول یکی از جهش های موثر در رنگ پرها (Kerje *et al.*, 2004) و همچنین حذف و اضافه-های موجود در ژن گیرنده هورمون رشد مرغ مرتبط با کوتولگی هستند (Agarwal *et al.*, 1994). مرغ اهلی در طی سال ها به دلیل ارزشی که در تولید غذای مورد نیاز انسان دارد، مورد اهمیت و توجه قرار گرفته است اما در طی دهه های اخیر از این پرنده سودمند به عنوان یک مدل بیولوژیک عالی نیز استفاده های زیادی شده است (Siegel *et al.*, 2006). بیش از 8000 سال قبل، مرغ در جنوب شرقی آسیا اهلی شد (Kanginakudru *et al.*, 2008) و امروزه بعنوان یکی از منابع مهم تامین پروتئین مصرفی انسان شناخته می شود. پرورش ماکیان در ایران و انتشار آن از طریق این کشور تاریخچه‌ای بسیار کهن دارد. ایران (پرشیا) یک امپراطوری بزرگ از قرن 5 قبل از میلاد تا تقریبا قرن 7 میلادی بود و از هند (دهلی) تا دریاهای سیاه و مدیترانه گسترده بود. در آن زمان و بعد از آن، در قرون وسطی ایران در محل تقاطع راه‌ها برای حمل و نقل محصولات، از قبیل ماکیان از شرق به غرب، هم از طریق خشکی و هم از طریق دریا قرار داشت. جنگ های زیادی


www.SID.ir

Archive of SID

در حوالی ایران و کشورهای همسایه در طی این دوره‌ها نیز توسعه و گسترش جمعیت‌های ماکیان را تسهیل کرد. حفاری‌های باستان شناسی حضور ماکیان را در ایران در زمان‌های باستان تأیید کرده است (Mohammadabadi et al., 2010). بر اساس تحقیقات West and Zhou استخوان‌های یافت شده در ایران در سه منطقه وجود داشته‌اند: دو کشف در تپه یحیی (Tepe Yahya) (جنوب شرقی ایران) به ترتیب متعلق به 3800 تا 3900 قبل از میلاد و 1000 قبل از میلاد و دیگری در تخت سلیمان (شمال غربی ایران) متعلق به 1000 قبل از میلاد (Mohammadabadi et al., 2010). شناسایی عوامل تعیین کننده ژنتیکی موثر در خصوصیات مهم اقتصادی یا بیماری‌ها یکی از نقاط توجه و تمرکز مطالعات ژنتیکی مرغ است که نیازمند داشتن اطلاعاتی از تنوعات موجود در توالی DNA و همچنین توسعه تعداد زیادی نشانگر ژنتیکی دارای اطلاعات مفید می‌باشد (Yan *et al*., 2014). مرغ قرمز جنگلی (نیای وحشی مرغ اهلی امروزی) اولین گونه از حیوانات اهلی بود که توالی یابی ژنوم آن کامل شد (Bai *et al*. 2012; Burt 2005; Hillier *et al*. 2004) و از آن زمان به بعد مطالعات زیادی برای شناسایی انواع تنوعات ساختاری ژنوم این پرنده انجام گرفت. در مطالعه ای که توسط Fan *et al*. (2013) انجام شد، بیش از 5/3 میلیون چندشکلی تک نوکلئوتیدی و 635000 حذف و اضافه کوچک در نژاد مرغ سیلکی (Silkie) و همچنین 5/1 میلیون چندشکلی تک نوکلئوتیدی و 633000 حذف و اضافه کوچک در یک نژاد مرغ بومی تایوان شناسایی شد. در مطالعه ای دیگر که توسط کنسرسیوم بین المللی نقشه یابی چندشکلی های ژنتیکی مرغ (International Chicken Polymorphism Map Consortium 2004) انجام گرفت، با مقایسه توالی ژنوم 3 مرغ اهلی (گوشتی، تخمگذار و سیلکی) با نیای وحشی (مرغ قرمز جنگلی)، در حدود 2/8 میلیون چندشکلی تک نوکلئوتیدی شناسایی و نقشه ای از تنوعات ژنتیکی مرغ تهیه شد. به منظور بررسی تنوعات ساختاری ژنوم مرغ، Yan *et al.* (2014) توالی ژنوم 12 نژاد مختلف مرغ را با عمق موثر 8/6 مورد مقایسه قرار دادند. در این مطالعه درحدود 1/7 میلیون حذف و اضافه کوچک و در حدود 16 میلیون چندشکلی تک نوکلئوتیدی شناسایی شد. در پژوهشی که روی 5 نژاد مرغ بومی کره ای انجام گرفت، Seo *et al.* (2015) در حدود 4 میلیون چند شکلی تک نوکلئوتیدی را شناسایی کردند. به دلیل اهمیتی که چندشکلی‌های تک نوکلئوتیدی و حذف و اضافه های کوتاه در مطالعات ژنتیکی و شناسایی عوامل ژنتیکی موثر در خصوصیات مهم و اقتصادی دارند، و این که مطالعات انگشت شماری روی تنوع ژنتیکی مرغ بومی فارس انجام شده است (Mohammadifar *et al*., 2014; Mohammadifar and


www.SID.ir

Archive of SID

Mohammadabadi, 2017)، در پژوهش حاضر این تنوعات ساختاری در ژنوم مرغ بومی استان فارس و با استفاده از توالی کل ژنوم شناسایی و مورد بررسی قرار گرفت.

## مواد و روش‌ها

نمونه خون چهار قطعه مرغ بومی از مرکز اصلاح نژاد استان فارس گرفته شد. استخراج DNA از نمونه‌های خون تهیه شده به روش استخراج نمکی انجام شد. 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. توالی یابی کل ژنوم به صورت Paired-End یا دو سویه توسط شرکت ایلومینا Hiseq 2500 در کشور چین انجام شد. در این روش، هر کدام از دو رشته رفت (Forward) و برگشت (Reverse) مربوط به هر قطعه توالی کوتاه توالی‌یابی می شوند. در نخستین مرحله از آنالیز داده‌ها، کیفیت داده‌ها توسط برنامه FastQC بررسی شد. فایل داده‌های توالی یابی کل ژنوم با پسوند fastqc. یا fq. می باشد. برای هم‌ردیفی توالی‌های کوتاه با ژنوم مرجع، در ابتدا آخرین نسخه ژنوم مرجع مرغ (https://www.ensembl.org/Gallus-gallus/Info/Index) از سایت Ensembl دانلود شد. ژنوم مرجع حاوی 39 جفت کروموزوم است که 38 جفت آن اتوزوم و یک جفت آن کروموزوم جنسی است. در این تحقیق برای هم‌ردیفی (Alignment) خروجی‌های توالی‌یابی با ژنوم مرجع، از الگوریتم MEM در برنامه BWA استفاده شد. به دلیل اینکه داده‌های توالی یابی شده به صورت دوسویه (Paired-end) هستند، در مرحله هم‌ردیفی، فایل حاوی توالی‌های رفت و برگشت به صورت همزمان در یک مرحله با ژنوم مرجع هم‌ردیف شده و خروجی آن یک فایل با پسوند sam. است. به منظور کاهش حجم فایل خروجی، در مرحله بعد با استفاده از نرم افزار Samtools، کلیه فایل‌های با پسوند sam. به فایل‌های با پسوند bam. تبدیل شدند. جهت مرتب کردن توالی‌های کوتاه موجود در هر کدام از فایل‌های با پسوند bam. بر اساس موقعیت ژنومیک، از دستور Sort در برنامه Samtools استفاده شد. برای حذف توالی‌های کوتاه تکراری (مضاعف شده) حاصل از واکنش زنجیره‌ای پلیمراز در مرحله آماده سازی کتابخانه ژنومی قبل از توالی‌یابی، از بسته MarkDuplicates در برنامه Picard استفاده شد. برای کاهش خطاهای حاصل از مرحله هم‌ردیفی، یک مرحله هم‌ردیفی مجدد (Realignment) با استفاده از برنامه GATK انجام شد به این ترتیب که از دو ابزار RealignerTargetCreator و IndelRealigner در برنامه GATK استفاده شد. به منظور تصحیح خطاهای توالی‌یابی و دیگر خطاهای مربوط به آزمایش، با استفاده نرم‌افزار GATK نمره کیفیت بازها مجدداً کالیبره شد. کیفیت نقشه‌یابی با دو پارامتر درصد همردیفی توالی‌های کوتاه با ژنوم مرجع و عمق پوشش بررسی شد. به منظور محاسبه درصد همردیفی با


www.SID.ir

Archive of SID

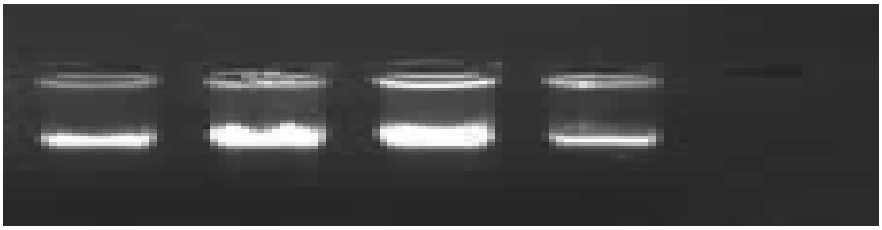
ژنوم و عمق پوشش، از دستورات flagtat و depth در نرم افزار Samtools استفاده شد. چندریختی‌های تک نوکلئوتیدی و حذف و اضافه-های کوچک با استفاده از ابزار SelectVariants در برنامه GATK از هم جدا شدند. به منظور فیلتر کردن چندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک، از ابزار Variant Filtration در برنامه GATK استفاده شد. به منظور مستندسازی یا انوتیشن، محاسبه شمار هموزیگوس‌ها و هتروزیگوس‌ها درچندریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک، و همچنین شمار جهش‌های جابه جایی و معکوس در چندریختی های تک نوکلئوتیدی، از برنامه SnpEff استفاده شد. به منظور بررسی ژن‌های موجود در نواحی شناسایی شده و بررسی تاثیرات عملکردی آنها، آنالیز ژن آنتولوژی[1] انجام گرفت. برای این کار ابتدا مختصات[2] جایگاه‌های مورد نظر شامل نام کروموزوم، نقطه شروع، نقطه پایان (برای چندریختی های تک نوکلئوتیدی، باید نقطه شروع و پایان یکسان باشد) با استفاده از ابزار VEP[3] در پایگاه Ensembl به آدرس http://www.ensembl.org/info/docs/tools/vep/index.html جستجو شد تا شناسه[4] ژن‌های دارای هم‌پوشانی با این نواحی مشخص شود. این شناسه‌ها در پایگاه DAVID[5] به آدرس https://david.ncifcrf.gov/ بارگذاری شدند تا اثرات عملکردی[6] آنها تعیین و آنالیز مسیرهای زیستی[7] مرتبط با آن انجام شود.

## نتایج و بحث

در این تحقیق کل ژنوم چهار قطعه مرغ بومی فارس به منظور مطالعه و شناسایی چندریختی‌های تک نوکلوتیدی وحذف واضافه های کوچک توالی یابی گردید. نتایج کیفیت و کمیت DNA استخراجی برای هر چهارنوع مرغ بومی فارس نشان داد که DNA مورد نیاز برای توالی‌یابی فاقد شکستگی و هرگونه مواد اضافی دیگر است (تصویر 1). از طرفی نتایج کنترل کیفیت داده‌ها بوسیله نرم افزار FastQC نشان داد که توالی‌ها (خوانش‌ها)ی کوتاه دارای کیفیت بالا و فاقد هرگونه آلودگی بودند. طول توالی‌های کوتاه برای همه داده‌ها 125 نوکلئوتید بود. درصد همردیفی برای هر چهار نمونه بالای 99 درصد بود که این میزان نشان دهنده کیفیت بالای همردیفی داده‌های توالی‌یابی شده نسبت به ژنوم مرجع بود (جدول 1). چند ریختی‌های تک‌نوکلئوتیدی بعد از همردیفی توالی‌های کوتاه با ژنوم مرجع استخراج شدند. تعداد چندریختی‌های تک‌نوکلئوتیدی برای هر چهار قطعه مرغ بومی فارس محاسبه گردید.

[1] Gene Ontology
[2] Coordinates
[3] Variant Effect Predictor
[4] Ensembl gene IDs
[5] Database for Annotation, Visualization and Integrated Discovery (DAVID )
[6] Functional enrichment
[7] Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis


www.SID.ir



**شکل 1- تعیین کیفیت نمونه های DNA استخراج شده با استفاده از ژل آگارز.**

**Figure 1- DNA quality control by agarose gel electrophoresis.**

**جدول 1- خلاصه هم ردیفی نمونه ها با ژنوم مرجع.**

**Table 1- Summary of short read alignment with reference genome.**

| نام نمونه<br>Sample name | تعداد توالی های کوتاه<br>Number of short reads | طول توالی های کوتاه<br>Short reads length | درصد همردیفی<br>Alignment percent | میانگین عمق<br>Coverage |
|---|---|---|---|---|
| Fars127s | 35774097 | 125 | 99.85% | 8.02-8.5 |
| Fars128s | 35233730 | 125 | 99.77% | 7.8-8.5 |
| Fars129s | 34264229 | 125 | 99.82% | 7.6-8.3 |
| Fars130s | 40240508 | 125 | 99.37% | 8.9-9.8 |

در این تحقیق تعداد 8858153 چندریختی تک-نوکلئوتیدی بدست آمد (جدول2) که تعداد هتروزیگوس‌ها بیشتر از هموزیگوس‌ها بود (جدول4). تعداد کل جهش‌های جابه‌جایی 24059405 و تعداد کل جهش‌های معکوس 10113770 و نسبت این دو نوع جهش به یکدیگر 2/379 بود (جدول4). نتایج مستندسازی چندریختی‌های تک‌نوکلئوتیدی نشان داد که از تعداد کل چندریختی‌های تک‌نوکلئوتیدی به ترتیب 9/677%، 1/794%، 29/158%، 48/727% و 9/945% در نواحی پایین دست، اگزون، بین‌ژنی، اینترون و بالادست قرار داشتند. این نتایج مستندسازی نشان داده‌اند که بیشتر چندریختی‌های تک‌نوکلئوتیدی در نواحی بین‌ژنی و اینترون و کمترین درصد در ناحیه اگزون قرار دارند (جدول5). همچنین تعداد کل چندریختی‌های تک‌نوکلئوتیدی خاموش(هم معنی) 25/502%، بد معنی 0/164% و بی معنی 74/335% بدست آمد (جدول 6). کالینز و جاکز (1994) در پژوهشی نشان دادند تعداد چندریختی‌های تک-نوکلئوتیدی که توسط جهش‌های جابه‌جایی تولید می‌شوند، تقریباً دو برابر تعداد چندریختی‌های تک-نوکلئوتیدی ایجاد شده توسط جهش‌های معکوس هستند، که این به علت مکانیسم‌های مولکولی ایجاد کننده جهش‌ها در مولکول DNA ازجمله دی

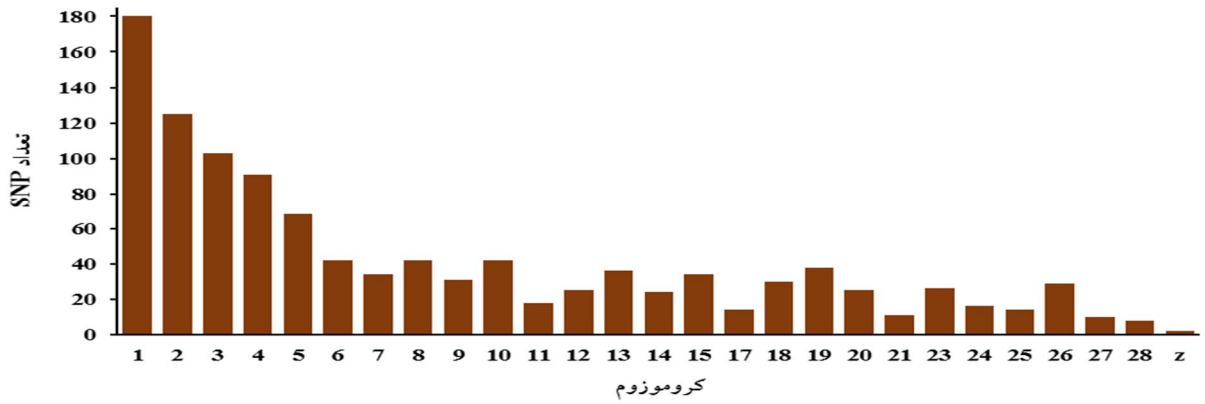

www.SID.ir

آمیناسیون اکسیداتیو است (Collins and Jukes, 1994). به طور کلی نسبت جهش‌های جابه‌جایی به معکوس یک شاخص مفید برای نشان دادن میزان خطای مثبت شناسایی چندریختی‌های تک-نوکلئوتیدی است (McKenna *et al*, 2010). شمار کل حذف و اضافه کوچک در ژنوم این چهار نمونه تعداد 895378 بود (جدول2) که در هر چهار نمونه مورد بررسی، تعداد حذف و اضافه‌های کوچک هموزیگوس از تعداد حذف و اضافه‌های کوچک هتروزیگوس بیشتر بود (جدول3). با مقایسه نتایج حاصل از پژوهش حاضر با سایر پژوهش های مشابه ، تفاوت در تعداد چند شکلی های تک نوکلئوتیدی و حذف و اضافه های کوتاه شناسایی شده را می توان دید. این تفاوت را می توان به عواملی همچون تعداد نمونه مورد بررسی، نوع سیستم توالی یابی، عمق توالی یابی و روش مورد استفاده برای آنالیز داده نسبت داد (Abel and Duncavage 2013). نتایج مستندسازی حذف و اضافه‌های کوچک نشان داد که بیشتر حذف و اضافه‌های کوچک در ناحیه اینترون ( 50%/428) و بین ژنی 29/188 ٪ و کم‌ترین مقدار مربوط به اگزون 0/984% بود (جدول 5). همچنین نتایج مشابه‌ای در سایر مطالعات شناسایی حذف و اضافه‌های ژنوم با استفاده از تکنیک توالی-یابی کل ژنوم در مرغ (Yan *et al*, 2014; Fleming *et al*, 2016) بدست آمد. در مجموع 1023 ژن با جایگاه های شناسایی شده در پژوهش حاضر مرتبط بود (فایل ضمیمه شماره 1) که توزیع این ژن ها در کروموزوم های مختلف در تصویر 2 قابل مشاهده است. همچنین پس از بررسی آنتولوژی مشخص شد که این ژن ها رابطه معنی داری ($P < 0.01$) با اصطلاحات ژنی مختلف ازجمله تنظیم فرآیندهای متابولیک سلولی، تقارن، اتصال یون ها و... داشتند. نتایج این آنالیز در فایل ضمیمه شماره 2 ارائه شده است. انجام آنالیز مسیرهای زیستی هیچ نتیجه معنی داری در پی نداشت.

**نتیجه‌گیری**

نتایج حاصل از شناسایی واریانت‌ها نشان داد که بیشترین فراوانی واریانت مربوط به چندریختی‌های تک نوکلئوتیدی است. همچنین، چندریختی‌های تک‌نوکلئوتیدی هتروزیگوس نسبت به هموزیگوس بیشتر بود که این می‌تواند نشان‌دهنده تنوع بالای جمعیت مورد نظر باشد.


www.SID.ir




**شکل 2- توزیع ژن های مرتبط با جایگاه های شناسایی شده در کروموزوم های مختلف.**

**Figure 2- Frequency of genes related to detected variants in different chromosomes.**

**جدول 2- تعداد چندریختی‌های تک‌نوکلئوتیدی و حذف و اضافه‌های کوچک برای چهار نمونه قبل و بعد از فیلترکردن.**

**Table 2- Number of raw and filtered SNPs and Indels.**

| تعداد INDELS<br>Number of Indels | تعدادSNP بعد از فیلتر<br>Number of filtered SNPs | تعداد SNP قبل از فیلتر<br>Number of raw SNPs | نام نمونه<br>Sample name |
|---|---|---|---|
| 661028 | 5646337 | 6526386 | 127s |
| 674423 | 5919961 | 6864563 | 128s |
| 659753 | 5760609 | 6673783 | 129s |
| 675141 | 6282455 | 7345131 | 130s |
| 895378 | 8858153 | 10717349 | Total |

**جدول 3- تعداد چند ریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک هموزیگوس و هتروزیگوس در نمونه ها.**

**Table 3- Number of homozygous and heterozygous SNPs and Indels.**

| تعدادIndel هموزیگوس<br>Number of homozygous Indels | تعداد Indel هتروزیگوس<br>Number of heterozygous Indels | تعداد SNP هموزیگوس<br>Number of homozygous SNPs | تعداد SNP هتروزیگوس<br>Number of heterozygous SNPs | نام نمونه<br>Sample name |
|---|---|---|---|---|
| 371365 | 289663 | 2826390 | 2819947 | 127s |
| 370909 | 303514 | 2798864 | 3121097 | 128s |
| 371547 | 288206 | 2841948 | 2918661 | 129s |
| 312214 | 362927 | 2248136 | 4034319 | 130s |


www.SID.ir

**جدول4- شمار جهش‌های جابه‌جایی و معکوس در چند ریختی‌های تک نوکلئوتیدی.**

**Table 4- Number of transition and transversion SNPs.**

| نسبت جهش های جابه-جایی به معکوس Transition to transversion ratio | تعداد جهش‌های معکوس Number of transversion mutations | تعداد جهش‌های جابه جایی Number of transition mutations | نام نمونه Sample name |
|---|---|---|---|
| 2.372 | 2501355 | 5933722 | 127s |
| 2.377 | 2570362 | 6110071 | 128s |
| 2.378 | 2535152 | 6029337 | 129s |
| 2.388 | 2506901 | 5986275 | 130s |
| 2.379 | 10113770 | 24059405 | کل Total |

**جدول 5- تعداد چند ریختی‌های تک نوکلئوتیدی و حذف و اضافه‌های کوچک در نواحی مختلف ژنوم.**

**Table 5- Distribution of SNPs and Indels in genome.**

| حذف و اضافه‌های کوچک Indels | چند ریختی‌های تک نوکلئوتیدی SNPs | نوع Type |
|---|---|---|
| (9.523%) 155709 | 1529128 (9.677%) | پایین دست Downstream |
| (0.984%) 16097 | (1.794%) 283475 | اگزون Exonic |
| (29.188%) 477258 | (29.158%) 4607664 | بین ژنی Intergenic |
| (50.428%) 824537 | 7699931 (48.727%) | اینترون Intronic |
| (0.011%) 187 | (0.003%) 520 | SPLICE_SITE_ACCEPTOR |
| (0.012%) 203 | (0.004%) 648 | SPLICE_SITE_DONOR |
| (0.221%) 3607 | (0.171%) 27055 | SPLICE_SITE_REGION |
| (9.006%) 147250 | (9.945%) 1571534 | بالا دست Upstream |
| (0.563%) 9200 | (0.427%) 67405 | UTR_3_PRIME |
| (0.062%) 1015 | (0.094%) 14931 | UTR_5_PRIME |

**جدول 6- شمار چندریختی‌های تک‌نوکلئوتیدی با اثرات مختلف در ژنوم.**

**Table 6- Number of SNPs with different effects in genome.**

| خاموش (درصد) Silent (%) | بی‌معنی(درصد) Nonsense (%) | بدمعنی(درصد) Missense (%) |
|---|---|---|
| 128189 (74.335%) | 282 (0.164%) | 43977 (25.502%) |


www.SID.ir

با توجه به پیشینه چند هزارساله اهلی شدن مرغ در ایران، می‌توان انتظار داشت که تطابق پذیری با محیط در جمعیت مورد مطالعه اتفاق افتاده باشد. در نتیجه تغییرات ژنومی در جهت سازگاری با محیط باعث ایجاد تنوع در جمعیت شده است.

باتوجه به نتایج بدست آمده این تحقیق در خصوص وجود تنوع بسیار زیاد ژنومیکی در توده مرغ بومی فارس، می‌توان اقدام به انتخاب افراد برتر این توده از لحاظ تولید و مقاومت به شرایط محیطی کرد. با توجه به اطلاعات حاصل از این مطالعه می-توان راهکارهایی برای حفاظت این نژاد در آینده ارائه داد. اطلاعات بدست آمده از آنالیز مقدماتی کل ژنوم مرغ‌های بومی فارس، زمینه مناسبی را برای ترکیب سایر داده های موجود برای تخمین مکان و تاریخ اهلی سازی مرغ فراهم نموده است.

**تشکر و قدردانی**

این پروژه با حمایت مالی مرکز مطالعات و همکاری های علمی بین المللی انجام شده است.

**منابع**

Abdolbaghi J (1990). Principles of poultry breeding. Bulletin of Animal Science Research Institute 46:23-34 (In Persian).

Abel, H. J., & Duncavage, E. J. (2013). Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer genetics 206: 432-440.

Agarwal SK, Cogburn LA, Burnside J (1994). Dysfunctional growth hormone receptor in a strain of sex-linked dwarf chicken: evidence for a mutation in the intracellular domain. Journal of Endocrinology 142: 427-434.

Amiri Ghanatsaman Z, Esmailizadeh A, Asadi Fozi M (2017). Study of structural diversity of genome Iranian native dog and wolf with the method whole genome sequencing. Iranian Journal of Animal Science 47: 271-277 (In Persian).

Arranz JJ, Bayón Y, Primitivo FS (1998). Genetic relationships among Spanish sheep using microsatellites. Animal Genetics 29: 435-440.

Azor PJ, Monteagudo LV, Luque M, Tejedor MT, Rodero E, Sierra I, Arruga MV (2005). Phylogenetic relationships among Spanish goat breeds. Animal genetics 36: 423-425.

Bai Y, Sartor M, Cavalcoli J (2012). Current status and future perspectives for sequencing livestock genomes. Journal of animal science and biotechnology 3: 8-13.

Burt DW (2005). Chicken genome: current status and future opportunities. Genome research 15: 1692-1698.

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

Dehghanzadeh H (2005). Study of genetic variation of Mazandaran, Yazd and West Azerbaijan native chicken using RAPD markers. Proceeding of the first congress on animal science and aquaculture, Iran (In Persian).



www.SID.ir

Fan WL, Ng CS, Chen CF, Lu MYJ, Chen YH, Liu CJ, Lai YT (2013). Genome-wide patterns of genetic variation in two domestic chickens. Genome biology and evolution 5: 1376-1392.

Fleming DS, Koltes JE, Fritz-Waters ER, Rothschild MF, Schmidt CJ, Ashwell CM, Lamont SJ (2016). Single nucleotide variant discovery of highly inbred Leghorn and Fayoumi chicken breeds using pooled whole genome resequencing data reveals insights into phenotype differences. BMC genomics 17: 812-817.

Futschik A, Schlötterer C (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics 186: 207-218.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-699.

International Chicken Polymorphism Map Consortium (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432: 717-721.

Kanginakudru S, Metta M, Jakati RD, Nagaraju J (2008). Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. BMC evolutionary biology 8: 174-178.

Kerje S, Sharma P, Gunnarsson U, Kim H, Bagchi S, Fredriksson R, Andersson L (2004). The Dominant white, Dun and Smoky color variants in chicken are associated with insertion/deletion polymorphisms in the PMEL17 gene. Genetics 168: 1507-1518.

Kim HM, Cho YS, Kim H, Jho S, Son B, Choi JY, Jang G (2013). Whole genome comparison of donor and cloned dogs. Scientific reports 3: 2998-3002.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, DePristo MA (2010). The genome analysis toolkit: a mapreduce frame work for analyzing next-generation dna sequencing data. Genome research 20: 1297-1303.

Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, Devine SE (2011). Natural genetic variation caused by small insertions and deletions in the human genome. Genome Research 115907.

Mirhosseini Z (1999). Study of genetic variation of silkworm using protein markers. PhD dissertation in animal breeding, Tarbiad Modares University (In Persian).

Moghaddaszadeh Ahrabi S (2003). Investigation of genetic potential of a Holstein herd using test day and random regression models. MSc dissertation in animal breeding, Zanjan University (In Persian).

Mohammadabadi MR, Nikbakhti M, Mirzaee HR, Shandi A, Saghi DA, Romanov MN, Moiseyeva IG (2010). Genetic variability in three native Iranian chicken populations of the Khorasan province based on microsatellite markers. Russian journal of genetics 46: 505-509.

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. Journal of Agricultural Biotechnology 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. Iranian Journal of Applied Animal Science 7: 679-685.

Mullaney JM, Mills RE, Pittard WS, Devine SE (2010). Small insertions and deletions (INDELs) in human genomes. Human molecular genetics 19: R131-R136.



www.SID.ir

Seo DW, Oh JD, Jin S, Song KD, Park HB, Heo KN, Lee HK (2015). Single nucleotide polymorphism analysis of Korean native chickens using next generation sequencing data. Molecular biology reports 42: 471-477.

Shamsaie AH (1992). Native chicken breeding. Animal Science Research Institute, Karaj, Iran (In Persian).

Siegel PB, Dodgson JB, Andersson L (2006). Progress from chicken genetics to the chicken genome. Poultry science 85: 2050-2060.

Snyder M, fraser AR, Laroche J, Gartner-kepkay KE, Zouros E (1987). Atypical mitochondrial DNA from the deep-sea scallop placopecten magellanicus. Proceeding of the National Academy of Science of the USA. 84:7595-7599.

Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, Moore SS (2011). Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC genomics 12: 559.

Tavakolian J (1999) A look at livestock and chicken genetic pool of Iran. Animal Research Institute of Iran (In Persian).

Wong GKS, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, Ni P (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432 (7018): 717-722.

Yan Y, Yi G, Sun C, Qu L, Yang N (2014). Genome-wide characterization of insertion and deletion variation in chicken using next generation sequencing. PloS one 9(8): e104652.



www.SID.ir

Archive of SID


# Identification of single nucleotide polymorphisms in Fars native chicken using whole genome sequencing data

Eskandari T.[1], Esmailizadeh A.K.[*2], Mohammadabadi M.R.[2], Sohrabi S.[3]

[1] MSc Student, Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran.
[2] Professor, Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran.
[3] PhD Student, Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran.

## Abstract

One of the most important approaches in animal breeding is identification of structural variations responsible for economical important traits. Next generation sequencing technologies are provided accurate tools for this purpose. In the present study, Fars province native poultry genome was analyzed using genome sequencing approach. Paired-end sequencing of the whole genome was conducted by Illumina Company in China. Data quality was determined by FastQC software. Short reads of sequences were mapped with the reference genome with the BWA program. Single nucleotide polymorphisms and small deletions and insertions were identified with the GATK program. The short sequences were compared with the reference genome of over 99% and with the mean depth of the X8 coverage. In this study, 8858153 single nucleotide polymorphisms and 895378 small deletions and insertions were identified with the lowest counts in the exon region. Since the results of identifying the variants showed that the most frequent variant is related to single nucleotide polynomials، in this study, heterozygous loci were higher than homozygous loci indicating the high diversity of the population under study. Given the history of several thousand years of chicken domestication in Iran, one can expect that adaptation to the environment has occurred in the population studied. As a result of genomic changes in order to adapt to the environment, it has created a diversity in the population.

**Keywords:** *Next Generation Sequencing, SNP, InDels, Population variation.*

[*] Corresponding Author: Esmailizadeh A.K. Tel: 09133958041 Email: aliesmaili@uk.ac.ir



www.SID.ir