

شناسایی واریانت‌های ژنتیکی در لاین آرین و بررسی عملکرد آنها با استفاده از توالی‌یابی کل ژنوم

حامد خراتی کوپایی^۱، اسماعیل ابراهیمی^{۲*}، محمد دادپسند^۳، علی نیازی^۴، علی اسمعیلی زاده^۵

^۱ دانشجوی دوره دکتری پژوهشکده زیست فناوری، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران.

^۲ دانشیار پژوهشکده زیست فناوری، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران.

^۳ دانشیار بخش علوم دامی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران.

^۴ استاد پژوهشکده زیست فناوری، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران.

^۵ استاد بخش علوم دامی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران.

تاریخ دریافت: ۱۳۹۷/۰۱/۲۶، تاریخ پذیرش: ۱۳۹۷/۰۶/۲۹

چکیده

مطالعه حاضر، اولین پژوهش برای شناسایی واریانت‌های ژنتیکی در لاین آرین با اطلاعات توالی‌یابی کل ژنوم است. مطالعه خصوصیات لاین آرین در سطح ژنوم، می‌تواند تفاوت‌های ژنتیکی آن را در ارتباط صفات مهم اقتصادی از جمله رشد شرح دهد. در این پژوهش، از اطلاعات توالی کل ژنوم سه قطعه مرغ آرین استفاده شد. نتایج نشان داد که ۶۳۸۶۶ واریانت بین نمونه‌ها مشترک است که از این تعداد، ۱۷۶۰۴ واریانت به عنوان واریانت جدید برای اولین بار در این پژوهش گزارش شدند. همچنین، ۶۰۴ واریانت در مناطق کدکننده ژنوم وجود دارند که از این تعداد، ۲۰۴ واریانت منجر به تغییر در نوع آمینواسیدی پروتئین‌ها می‌گردند. با توجه به مهم بودن کیفیت پروتئین جیره برای رشد و افزایش وزن برای طیور، نتایج ژن آنتولوژی نشان داد که مسیر کاتابولیسم پروتئین‌ها و مسیرهای تغییرات پس از ترجمه‌ی پروتئین‌های مرتبط با رشد دارای اهمیتی دوچندان هستند. همچنین، ژن‌های کاندیدای *SMURF2* و *SUMO1* برای این مسیرهای بیولوژیک پیشنهاد شد. یکی دیگر از نتایج قابل توجه ژن آنتولوژی، معنی‌دار شدن مسیرهای پاسخ به تنش بود. در لاین آرین به دلیل انتخاب شدید برای دستیابی به پیشرفت ژنتیکی بالا در صفات رشد، وجود تنش اجتناب‌ناپذیر است. بنابراین، معنی‌دار شدن مسیرهای پاسخ به تنش، می‌تواند نشان دهنده اهمیت این مسیرها در ارتباط با عملکرد مطلوب باشد.

واژه‌های کلیدی: لاین آرین، توالی‌یابی کل ژنوم، واریانت.

متفاوت باشد (Alkan *et al.*, 2011). روش‌های قدیمی‌تر برای شناسایی واریانت‌ها مانند روش‌های سیتوژنتیکی از تکنیک‌هایی مانند تهیه کاریوتایپ و (Fluorescence in situ hybridization) FISH بهره می‌بردند که به دلیل پوشش محدود ژنوم و قدرت تفکیک پایین، توانایی لازم را برای شناسایی واریانت‌ها و به‌خصوص جهش‌های نادر و جدید نداشتند (Buysse *et al.*, 2009). اما، امروزه فناوری NGS (Next generation sequencing) با مزیت‌هایی مانند پوشش گسترده‌تر ژنوم، قابلیت تفکیک بالاتر، زمان کم توالی‌یابی و 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، تصویر آشکاری از خصوصیات ژنومی به نمایش گذارد (Metzker, 2010). با توجه به وارث‌پذیر بودن و فراوانی بالای واریانت‌ها، این احتمال وجود دارد که برخی از واریانت‌ها با ژن‌های اصلی تولید و سلامتی دام‌ها در ارتباط باشند (Kharrati-Koopae and Esmailzadeh, 2014). پرورش ماکیان در ایران و انتشار آن از طریق این کشور تاریخچه‌ای بسیار کهن دارد. ایران یک امپراطوری بزرگ از قرن ۵ قبل از میلاد تا تقریباً قرن هفت میلادی بود و از هند (دهلی) تا دریا‌های سیاه و مدیترانه گسترده بود. در آن

دو فرآیند اصلی انتخاب طبیعی و مصنوعی زمینه اصلی تغییرات ژنتیکی در ژنوم حیوانات مزرعه‌ایی را بوجود آورده‌اند. انتخاب طبیعی به منظور بقای موجود با تغییرات شرایط محیطی، موجب سازگار شدن ژنوم با محیط می‌گردد و انتخاب مصنوعی به دست بشر، برای افزایش پیشرفت ژنتیکی و بهبود تولید صورت گرفته است (Zhang *et al.*, 2012)؛ یک نمونه از انتخاب مصنوعی، لاین آرین است. بدون تردید، لاین آرین یکی از ارزشمندترین خزانه ژنتیکی مرغ گوشتی مولد در ایران و خاورمیانه است؛ بنابراین، مطالعه آن در سطح ژنوم و شناسایی واریانت‌های ژنتیکی می‌تواند به نحو چشمگیری در شناسایی عوامل بهبود کمیت و کیفیت تولید، تاثیرگذار باشد. شناسایی مناطق ژنومی مرتبط با صفات اقتصادی، مانند تولید گوشت و مقاومت به بیماری، همیشه از موضوعات مورد علاقه دانش‌پژوهان در جهت بهبود تولید و اصلاح نژاد بوده است (Fleming *et al.*, 2016) به طور کلی، تنوع ژنتیکی دربردارنده واریانت‌های تک‌نوکلئوتیدی^۱، حذف و اضافه‌های کوتاه، تنوع در تعداد نسخه‌ها^۲ و تنوعات ساختاری بزرگ^۳ است و محدوده این تغییرات می‌تواند از یک جفت باز تا جایگزینی‌های بزرگ در حد کروموزوم،

¹ Single nucleotide variations

² Copy number variations

³ Structural variants

پژوهش دیگر در طیور با استفاده از اطلاعات یک میلیون SNP نشان داد که ژن‌های *NCOA1* موثر بر صفات تولیدمثلی هستند. همچنین، ژن‌های *PRKAB2* و *SPP1*, *JGF2*, *CAPN1*, *TGFb2* در ارتباط با صفات لاشه و رشد گزارش شدند (Gholami *et al.*, 2014). پژوهشی دیگر برای شناسایی نواحی ژنومی مرتبط با ذخیره چربی در ناحیه شکمی در دو لاین جوجه گوشتی تجاری که به لحاظ ذخیره چربی شکمی تفاوت فنوتیپی قابل ملاحظه‌ایی داشتند (لاین پرچرب و کم چرب) انجام گرفت. نتایج نشان دادند ۵۱ ناحیه ژنومی در لاین کم چرب و ۵۷ ناحیه در لاین پرچرب می‌توانند بر ذخیره چربی موثر باشند. علاوه بر این، ژن‌های کانید *RBI*, *BS7* و *MAOA* برای این نواحی پیشنهاد شدند (Zhang *et al.*, 2012). پژوهش دیگری روی نژادهای تجاری و غیرتجاری نشان داد که ۳۹۶ ناحیه کروموزومی مرتبط با صفات عملکردی در طیور وجود دارد که ۱۳ ناحیه برای اولین بار در این پژوهش گزارش شدند (Elferink *et al.*, 2012). علاوه بر طیور، مطالعات دیگری نیز در سایر حیوانات مزرعه‌ایی با شناسایی واریانت‌ها انجام گرفته است. برای نمونه، پژوهش‌ها نشان داده‌اند که ۳۶ ناحیه ژنومی در ارتباط با اهلی شدن سگ وجود دارند که افزایش قابلیت هضم نشاسته در سگ‌ها را نسبت به اجداد گرگ خود را در طی زمان

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 راه خشکی و هم از راه دریا قرار داشت. جنگ‌های زیادی 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 پژوهش‌های انجام شده، استخوان‌های یافت شده در ایران در سه منطقه وجود داشته‌اند: دو کشف در تپه یحیی (Tepe Yahya) (جنوب شرقی ایران) به ترتیب متعلق به ۳۸۰۰ تا ۳۹۰۰ قبل از میلاد و ۱۰۰۰ قبل از میلاد و دیگری در تخت سلیمان (شمال غربی ایران) متعلق به ۱۰۰۰ قبل از میلاد (Mohammadabadi *et al.*, 2010). تاکنون پژوهش‌های گسترده‌ایی برای شناسایی ژن‌های کانیدای صفات اقتصادی با استفاده از SNPchip ها و فناوری NGS انجام شده است. برای نمونه، در پژوهشی آنالیز ژنومیکس مقایسه‌ایی بین نژادهای گوشتی، تخم‌گذار و مرغ جنگلی قرمز با استفاده از اطلاعات هفت میلیون SNP انجام گرفت. نتایج نشان دادند که ژن گیرنده هورمون تیروئید که نقش اصلی در سوخت و ساز بدن دارد، دارای بیشترین واریانت‌ها و تغییرات در بین سه جمعیت اصلی است (Rubin *et al.*, 2010).

اطلاعات ژنوم رفرنس و تفسیر ژنوم^۴ از پایگاه Ensembl (ftp://ftp.ensembl.org/pub/release-84/fasta/gallus_gallus) دانلود گردید، که شامل ژنوم رفرنس ورژن ۸۴ و اطلاعات تفسیر ژنوم شامل Gene، CDS و واریانت‌های شناسایی شده^۵ بودند. واریانت‌های شناسایی شده، نشان‌دهنده تمام واریانت‌های شناسایی شده در طیور بودند که برای شناسایی واریانت‌های جدید در این پژوهش استفاده شدند. CDS حاوی اطلاعات نواحی کدینگ در ژنوم طیور است که برای تعیین واریانت‌هایی کدینگ استفاده شدند. اطلاعات Gene نیز برای شناسایی واریانت‌هایی که منجر به تغییرات آمینواسیدی در ساختار پروتئین می‌شدند، استفاده گردید. توالی‌یابی کل ژنوم با استفاده از پلات فرم ایلومینا Hiseq 2000 انجام شد. خوانش‌ها به شکل Pair end و به طول ۱۲۰ جفت باز تهیه شدند. کنترل کیفیت برای طول خوانش‌ها، محتوای GC، بازهای خوانده نشده، خوانش‌های تکراری و شاخص Phred انجام شد. با توجه به اینکه توالی آدپتورها توسط کمپانی ایلومینا حذف شده بودند، ویرایش تنها بر اساس پارامترهای کنترل کیفیت انجام گرفت.

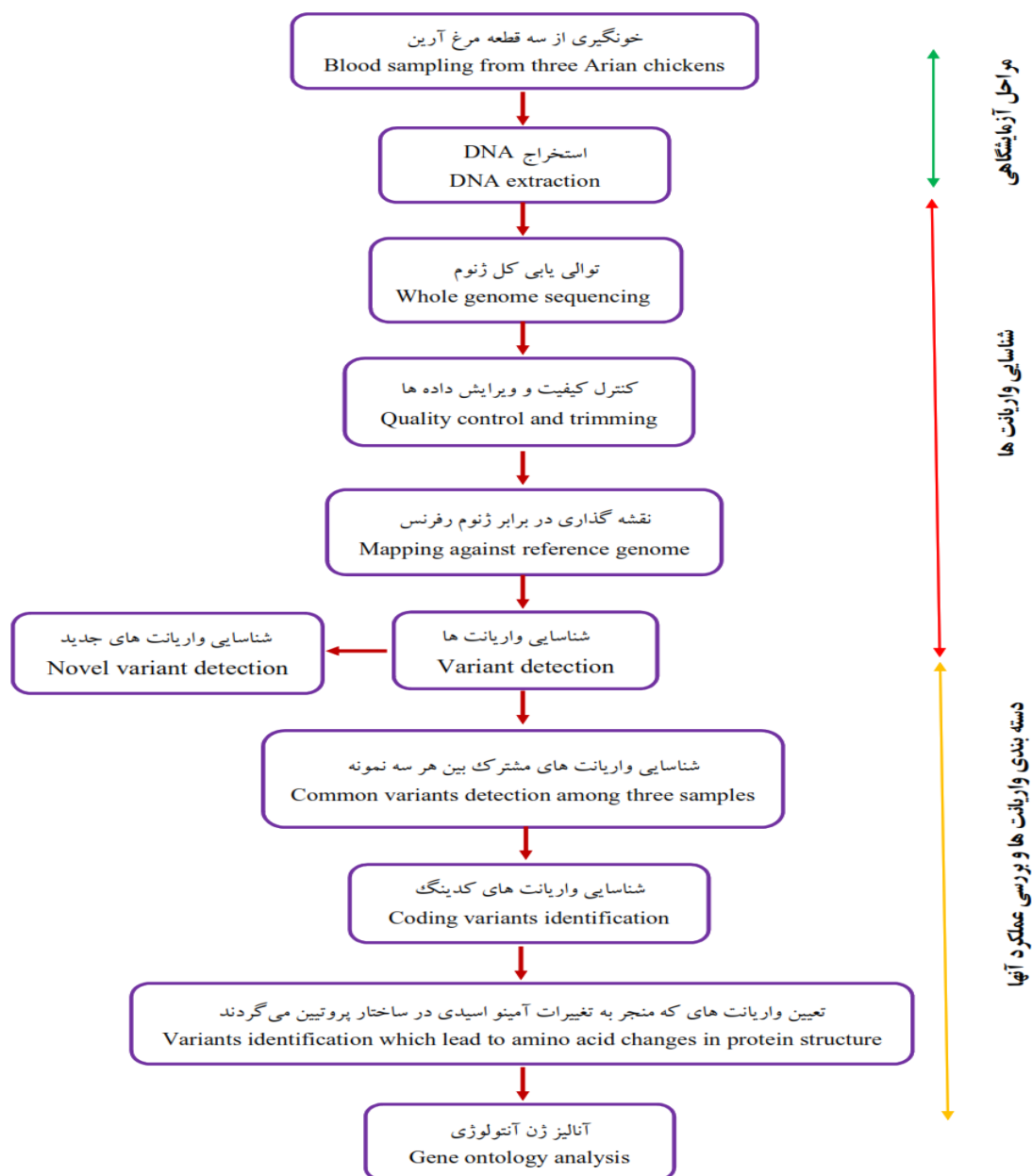
⁴ Annotations

⁵ Known variants

ایجاد کرده‌اند (Axelsson *et al.*, 2013). اگر چه مطالعات ملکولی متعددی روی طیور در ایران انجام شده است (Zandi *et al.*, 2014; Mohammadifar *et al.*, 2011; Moazeni *et al.*, 2016a; Moazeni *et al.*, 2016b; Shahdadnejad *et al.*, 2014). اما تاکنون شناسایی واریانت‌های ژنتیکی در مرغ لاین آرین به منظور بررسی عملکرد آنها در رابطه با صفات اقتصادی مورد مطالعه قرار نگرفته است؛ لذا، هدف این پژوهش، شناسایی واریانت‌های ژنتیکی در مرغ لاین آرین به منظور بررسی عملکرد آنها در رابطه با صفات اقتصادی بود.

مواد و روش‌ها

روش کلی مراحل نمونه‌برداری و آنالیز داده‌ها در شکل ۱ نشان داده شده است. قابل ذکر است که کلیه مراحل واکاوی داده‌های NGS با نرم‌افزار CLC genomic workbench(8.5) انجام شد. بنابراین، از تکرار نام نرم‌افزار در بخش‌های مختلف مقاله خودداری شده است. خون‌گیری از سه قطعه مرغ لاین تجاری آرین واقع در بابل کنار انجام گرفت. استخراج DNA با روش نمکی انجام شد (Miller, 1988) و کیفیت DNA با استفاده از نانودراپ و ژل گذاری تائید گردید. نمونه‌های DNA حداقل با غلظت ۵۰ نانوگرم در هر میکرولیتر در حجم نهایی ۱۰۰ میکرولیتر برای انجام توالی‌یابی آماده شدند.



شکل ۱- خلاصه‌ی مراحل نمونه برداری و آنالیز داده‌های NGS برای شناسایی واریانت ها در لاین آراین.

Figure 1- The Summary of sampling and NGS data analysis for variant discovery in Arian line.

شدند و واریانت‌های مشترک برای آنالیزهای بعدی و دسته‌بندی کردن ذخیره گردیدند. برای انجام این منظور میزان آستانه فراوانی واریانت ۱۰۰ درصد در نظر گرفته شد؛ به عبارت دیگر، واریانتی به عنوان واریانت مشترک انتخاب شد که در هر سه نمونه وجود داشته باشد. دسته‌بندی اول، شامل شناسایی واریانت‌هایی بود که برای نخستین بار در این پژوهش گزارش می‌شدند و در پایگاه‌های اطلاعاتی ثبت نشده بودند. این دسته‌بندی با کمک فایل واریانت‌های شناسایی شده انجام گرفت. دسته‌بندی دوم، شامل واریانت‌های بود که در مناطق کدکننده ژنوم وجود داشتند که با استفاده از اطلاعات CDS انجام شد. واریانت‌هایی که در نواحی کدکننده تعیین شدند در دسته‌بندی دیگری که منجر به تغییرات آمینواسیدی در پروتئین می‌گشت قرار گرفتند، که این آنالیز با استفاده از اطلاعات CDS، Gene و توالی ژنوم رفرنس انجام شد. برای انجام آنالیز ژن آنتولوژی، فایل ژن آنتولوژی مربوط به مرغ (*Gallus gallus*) از پایگاه Gene ontology consortium دانلود گردید و با استفاده از فایل تغییرات آمینواسیدی، آنالیز ژن آنتولوژی در سه سطح Biological process (BP)، Cellular Molecular function (MF) و component (CC) با سطح معنی‌داری پنج درصد انجام گرفت.

بعد از ویرایش، کنترل کیفیت برای حصول اطمینان از مطلوب بودن نتایج ویرایش داده‌ها انجام گرفت. برای حصول اطمینان از عمق خوانش مناسب نمونه‌ها برای شناسایی واریانت‌ها از پارامتر Coverage و رابطه لندر- واترمن به شرح زیر استفاده شد (Lander and waterman, 1988).

$$C = LN / G$$

در این رابطه، C برابر با Coverage، L طول خوانش، N نشان دهنده تعداد خوانش‌ها و G طول ژنوم رفرنس بر اساس جفت باز است. قابل ذکر است که این پارامتر برای هر سه نمونه در دو مرحله قبل و بعد از ویرایش کردن محاسبه گردید.

نقشه‌گذاری داده‌های ویرایش شده در برابر ژنوم رفرنس بر اساس پارامترهای زیر انجام شد. میزان جریمه هر باز ناجور دو، میزان جریمه حذف و یا اضافه، سه و میزان تشابه خوانش‌ها و ژنوم رفرنس ۸۰ درصد در نظر گرفته شد (Doan et al., 2012) با توجه به ثابت بودن سطح پلوئیدی در طیور ($2n=78$)، شناسایی واریانت‌ها با الگوریتم فیکس پلوئیدی (Fixed ploidy) و تحت پارامترهای زیر انجام شد. امتیاز باز واریانت برابر با ۳۰، امتیاز بازهای مجاور واریانت ۱۵ و میزان حداقل فراوانی واریانت ۳۰ درصد گرفته شد (Dias et al., 2017).

به منظور تفسیر بهتر نتایج واریانت‌های شناسایی شده در بین سه نمونه با همدیگر مقایسه

نتایج و بحث

دارای کیفیت خوانش نامطلوبی بودند، از کتابخانه‌های ژنومی حذف شدند. نتایج حاصل از هم‌ردیفی خوانش‌ها و ژنوم رفرنس نشان داد که نقشه‌گذاری به نحو مطلوبی انجام شده است؛ به طوری که درصد نقشه‌گذاری برای نمونه‌ها بین ۹۱ تا ۹۳ درصد برآورد شدند. همچنین، میزان Coverage محاسبه نشان داد که کتابخانه‌های ژنومی، عمق خوانش مناسبی برای انجام شناسایی واریانت‌ها دارند. نتایج بیشتر در جدول ۱ ارائه شده است.

نتایج کنترل کیفیت داده‌ها نشان داد که کتابخانه‌های ژنومی با کیفیتی مطلوب تهیه شده‌اند و نیاز به انجام ویرایش روی داده‌ها نیست؛ با این حال، به منظور کاهش خطای نقشه‌گذاری، ویرایش داده‌ها تنها بر اساس شاخص Phred و حذف بازه‌هایی که کیفیت خوانش پایینی داشتند، انجام شد. به این ترتیب، پنج درصد از خوانش‌هایی که دارای کمترین میزان شاخص Phred بودند به همراه ده جفت باز ناحیه ۳ که

جدول ۱- خلاصه نتایج حاصل نقشه‌گذاری خوانش‌ها در برابر ژنوم رفرنس و محاسبه Coverage.

Table 1- The summary of short reads alignments against reference genome and coverage calculation.

تعداد کل خوانش‌ها Total reads	درصد نقشه‌گذاری Mapped reads%	Coverage قبل از ویرایش Coverage before trimming	Coverage بعد از ویرایش Coverage after trimming
74894016	92.45	8.98	8.23
67495154	91.90	8.09	7.42
70314620	93.40	8.43	7.73

(nucleotide variation)، حذف (Deletion) و واریانت جایگزینی (Replacement) در نمونه وجود دارند. جزییات بیشتر از تعداد و انواع واریانت‌های شناسایی شده در جدول ۲ نشان داده شده است.

نتایج شناسایی واریانت‌ها نشان داد که بیش از چهار میلیون واریانت شامل واریانت‌های تک‌نوکلئوتیدی (SNV: single nucleotide variation)، اضافه شدن (Insertion)، واریانت چندنوکلئوتیدی (Multi-

جدول ۲- نتایج حاصل از شناسایی واریانت‌ها در سه نمونه مرغ لاین آرین

Table 2- The results of variant discovery in three Arian samples

Total variants کل واریانت‌ها	SNV% تک نوکلئوتیدی	MNV% چند نوکلئوتیدی	Insertion % اضافه شدن	Deletion% حذف شدن	Replacement% جایگزینی
1257551	88.10	3.90	4.00	3.50	0.50
1705857	87.50	3.50	4.60	3.80	0.60
1243857	87.90	3.80	4.30	3.60	0.40

SNV: Single nucleotide variation, MNV: Multi nucleotide variation

پروتئین‌ها گردد و بدین شکل بر فعالیت آنها تاثیر بگذارد (Mukhopadhyay *et al.*, 2007). با توجه به اینکه یکی از اصلی‌ترین اهداف لاین آرین، تولید گوشت و سرعت رشد است، وجود مسیرهای بیولوژیکی کارآمد برای کاتابولیسم پروتئین جیره‌های مصرفی، ضروری به نظر می‌رسد. بنابراین، می‌توان گفت که در لاین آرین به دلیل پیشرفت ژنتیکی قابل توجهی که برای صفات رشد در سال‌های متمادی بر اثر انتخاب بوجود آمده است، مسیرهای کاتابولیک پروتئین از اهمیتی دوچندان برخوردار شده‌اند. در تایید مطلب فوق، پژوهش‌های گوناگون نشان داده‌اند که کیفیت پروتئین‌های جیره به نحو مطلوبی می‌توانند با تاثیر بر بازده مسیر متابولیسم پروتئین بر رشد توده ماهیچه‌ای بدن تاثیرگذار باشد (Mudalal *et al.*, 2014). همچنین، در پژوهش حاضر، ژن *SMURF2* به عنوان ژن کاندیدا برای این مسیر بیولوژیک پیشنهاد گردید. این ژن با کد کردن آنزیم E3 ubiquitin-protein ligase در پذیرش ملکول

مقایسه واریانت‌های شناسایی شده در بین هر سه نمونه مورد مطالعه، نشان داد که ۶۳۸۶۶ واریانت بین نمونه‌ها مشترک است که از این تعداد ۱۷۶۰۴ واریانت به عنوان واریانت جدید برای اولین بار در این پژوهش گزارش شد و اطلاعاتی از آنها در پایگاه‌های اطلاعاتی موجود نیست. همچنین، نتایج نشان داد که ۶۰۴ واریانت در مناطق کدکننده ژنوم وجود دارند که از این تعداد، ۲۰۴ واریانت منجر به تغییر در آمینواسید پروتئین‌ها (Missense variants) می‌گردند. نتایج آنالیز ژن آنتولوژی در سطح BP نشان داد که کاتابولیسم پروتئین وابسته به Ubiquitin (GO: 0042787) دارای بیشترین فراوانی است. این فرآیند بیولوژیک با کمک پروتئینی کوچکی به نام Ubiquitin انجام می‌گیرد. Ubiquitin با نشان‌دار کردن سایر پروتئین‌ها به عنوان پیام‌رسان برای پروتئاز عمل می‌کند و عمل هضم پروتئین‌های ناکارآمد را در سلول تسهیل می‌کند. همچنین، Ubiquitin می‌تواند موجب افزایش یا کاهش اثر متقابل بین

نقش داشته باشند (Glaever *et al.*, 2002). یکی از اصلی‌ترین نقش SUMO ها در رشد و قدرت زنده‌مانی سلول است؛ پژوهش‌ها نشان داده‌اند که حذف پارالوگ‌های SUMO در مخمر می‌تواند موجب مرگ سلول و کاهش رشد گردد (Craig and Henley, 2012).

بدون تردید، یکی از اصلی‌ترین مشکلات در لاین‌های تجاری مانند آرین به دلیل شدت انتخاب بالا و پیشرفت ژنتیکی زیاد برای صفات گوناگون، وجود تنش است. یکی دیگر از نتایج جالب توجه این پژوهش وجود مسیر بیولوژیک Stress-activated protein (GO: 0031098) kinase signaling است. این مسیر بیولوژیک در برابر وجود توکسین‌ها و تنش‌های فیزیکی فعال می‌گردد (Tibbles and Woodgett, 1999). همچنین ژن کاندیدای *TAOK1* برای این مسیر بیولوژیک پیشنهاد گردید. این ژن با کد کردن Serine/threonine-protein kinase در پاسخ به تنش‌ها مانند تخریب DNA، پایداری اسکلت سلولی و آپوپتوز نقش‌های مهمی ایفا می‌کند (Uniprot: Q7L7X3). جزئیات بیشتر از آنالیز BP در جدول ۳ نشان داده شده است. نتایج آنالیز ژن آنتولوژی در سطح MF نشان داد که بیشترین فراوانی مربوط به ژن آنتولوژی پروتئین Serine/threonine kinase activity (GO: 0004674) است که تاییدکننده نتایج ذکر شده در آنالیز BP است. همچنین، ژن کاندیدای *PIK3C2A* برای این

Ubiquitin و انتقال آن به پروتئین هدف نقش کلیدی ایفا می‌کند (Uniprot: Q9HAU4). علاوه بر این، نتایج نشان دادند که Protein neddylation می‌تواند به عنوان یک مسیر بیولوژیک که در رشد، توسعه و حیات سلول‌ها دخیل است، مورد توجه قرار گیرد. در این فرآیند، پروتئین کوچکی بنام NEDD8 (Neural-precursor-cell-expressed developmentally down-regulated 8) اتصال پروتئین هدف و بعنوان یکی از تغییرات پس از ترجمه، باعث بهبود نقش تنظیمی پروتئین‌ها می‌گردد (Xirodimas 2008; Dharmasiri *et al.*, 2003). نتایج مشابهی نشان داده‌اند که تجمع مولکول موتانت NEDD8 به شکل نامحسوسی می‌تواند بر میزان رشد تاثیرگذار باشد که این امر خود نشان دهنده نقش کلیدی این پروتئین در رشد و حفظ هموستازی است (Mergner *et al.*, 2017). یکی دیگر از نتایج قابل توجه این پژوهش، فرآیند بیولوژیک SUMOylation (GO: 0016925) است. SUMOylation یا Small Ubiquitin-like Modifier به عنوان تغییرات پس از ترجمه می‌تواند به پروتئین‌های هدف متصل شوند و عملکرد پروتئین‌ها را بهبود ببخشند. این پروتئین می‌تواند در فرآیندهای مرگ سلولی، انتقال مواد بین هسته و سیتوپلاسم، پایداری پروتئین‌ها، پاسخ به تنش و تنظیم چرخه تقسیم سلولی

دستگاه گلژی مشارکت دارد (Uniprot: P48563). در تایید نتایج بدست آمده، پژوهش‌های گوناگونی نشان داده‌اند که شبکه انتقال دستگاه گلژی علاوه بر انتقال، در بیوستز برخی پروتئین‌ها نیز مشارکت دارند که از این طریق می‌توانند بر متابولیسم کلی پروتئین‌ها و بهره‌وری بهتر آنها تاثیرگذار باشند (Connerly, 2010; Groen, 2014). جزئیات بیشتر از نتایج آنالیز CC در جدول ۵ خلاصه شده‌اند

از وجود نتایج مشابه ژن آنتولوژی در بخش‌های BP، CC و MF می‌توان دریافت که وجود واریانت‌های ژنتیکی در مسیرهای متابولیسم پروتئین در لاین آرین بیان‌گر یک مزیت برای بهره‌وری بهتر از پروتئین جیره مصرفی و در نهایت رشد بیشتر است، که در نتیجه‌ی سال‌ها انتخاب مصنوعی و آمیزش‌های هدفمند جهت بروز صفات مطلوب اقتصادی به ثمر رسیده است. با این حال، مطالعه دقیق‌تر واریانت‌های ژنتیکی مانند تایید آزمایشگاهی و بررسی عملکرد آنها در شرایط مدل‌های درون‌تنی می‌تواند بیش از پیش اطلاعات مفیدی را در اختیار پژوهشگران قرار دهد (Zhu et al., 2017).

عملگر مولکولی پیشنهاد گردید. این ژن با کد کردن 3-Phosphatidylinositol 4-phosphate kinase C2 domain-containing subunit alpha در تنظیم فعالیت سلول نقش اساسی دارد. که از مهم‌ترین آنها می‌توان به انتقال گلوکز به پلاسما، ترشح انسولین و ایجاد پاسخ به اثرات مضر UV و در حالت کلی تنش اشاره کرد (Uniprot: O00443). یکی دیگر از نتایج مشابه این تحقیق، وجود ژن آنتولوژی Ubiquitin-protein ligase activity (GO: 0004842) است که با فراوانی ۲۴/۲ درصد در مرتبه دوم اهمیت برای آنالیز MF است و همانطور که بیان شد در تغییرات پس از ترجمه پروتئین‌ها و بهبود عملکرد آنها نقش کلیدی دارد. همچنین، پیشنهاد شد ژن *AP4SI* دارای نقش کلیدی برای انتقال پروتئین‌ها در سلول به کمک دستگاه گلژی است (Uniprot: Q9Y587). نتایج بیشتر از آنالیز MF در جدول ۴ نشان داده شده است. نتایج ژن آنتولوژی در سطح CC نشان داد که محیط داخل سلول و شبکه انتقال دستگاه گلژی دارای بیشترین فراوانی در بین سایر ژن آنتولوژی‌های یافت شده هستند. همچنین، ژن *MON2* برای این موقعیت سلولی گزارش گردید که در فرآیندهای آندوسیتوز و بسته بندی پروتئین‌ها و انتقال آنها بوسیله

جدول ۳- آنالیز ژن آنتولوژی در سطح **Biological process** برای واریانت‌های مشترک شناسایی شده در لاین آرین.

Table 3- Gene ontology enrichments analysis in biological process level for identified common variants in Arian line.

Go Num	Go description	Frequency%	Gene name	Type	Chr
شماره ژن آنتولوژی	شرح ژن آنتولوژی	درصد فراوانی	نام ژن	نوع	کروموزوم
0042787	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process	14.5	<i>SMURF2</i>	SNV	18
0045116	Protein neddylation	5.5	<i>UBA3</i>	SNV	12
0016192	Vesicle-mediated transport	4.6	<i>NCK2</i>	SNV	1
0035556	Intracellular signal transduction	13	<i>PRKCZ</i>	SNV	2
0014065	Phosphatidylinositol 3-kinase cascade	3.75	<i>GATA3</i>	Insertion	1
0007157	Heterophilic cell-cell adhesion	2.1	<i>CADM2</i>	SNV	1
0061024	Membrane organization	3.5	<i>SGCG</i>	SNV	1
0031098	Stress-activated protein kinase signaling	7.3	<i>TAOK1</i>	SNV	19
0016310	Phosphorylation	4.25	<i>EPHA7</i>	SNV	3
0023014	Signal transduction by phosphorylation	6.4	<i>STK3</i>	Deletion	2
0016925	Protein sumoylation	4.5	<i>SUMO1</i>	SNV	7
0048762	Mesenchymal cell differentiation	5.5	<i>TGFBR2</i>	SNV	2
0030705	Cytoskeleton-dependent intracellular transport	4.5	<i>KIF14</i>	Insertion	14
0051480	Cytosolic calcium ion homeostasis	5	<i>TRPC6</i>	SNV	1
0061025	Membrane fusion	3.1	<i>STX18</i>	SNV	4
0086010	Membrane depolarization involved in regulation of action potential	6.5	<i>SCN5A</i>	SNV	2
0007018	Microtubule-based movement	6	<i>KIF23</i>	SNV	10

جدول ۴- آنالیز ژن آنتولوژی در سطح Molecular function برای واریانتهای مشترک شناسایی شده در لاین آرین.

Table 4- Gene ontology enrichments analysis in molecular function level for identified common variants in Arian line.

Go term	Go description شرح ژن آنتولوژی	Frequency درصد فراوانی	Gene name نام ژن	Type نوع	Chr کروموزوم
0005086	ARF guanyl-nucleotide exchange factor activity	5.1	MON2	SNV	1
0042393	Histone binding	7	NCAPD2	Deletion	1
0016301	Kinase activity	3.8	ATM	SNV	1
0005487	Nucleocytoplasmic transporter activity	3.2	NUP133	SNV	3
0004674	Protein serine/threonine kinase activity	27.4	RPS6KA2	SNV	3
0035005	1-phosphatidylinositol-4-phosphate 3-kinase	3.8	PIK3C2A	Deletion	5
0004842	Ubiquitin-protein ligase activity	24.2	AP4S1	Insertion	5
0000175	3'-5'-exoribonuclease activity	3.2	DIS3L2	SNV	9
0015279	Store-operated calcium channel activity	3.8	TRPC7	SNV	13
0016303	1-phosphatidylinositol-3-kinase activity	3.8	TMEM201	SNV	21
0004016	Adenylate cyclase activity	5.1	GUCY2C	SNV	1
0044212	Transcription regulatory region DNA binding	9.6	TCF20	SNV	1

جدول ۵- آنالیز ژن آنتولوژی در سطح the cellular component برای واریانت های مشترک شناسایی شده در لاین آرین.

Table 5- Analysis of the ontology gene at the cellular component level for the common variant identified in Arian line.

GO	Go description	Frequency%	Gene name	Type	Chr
ژن	شرح ژن آنتولوژی	درصد فراوانی	نام ژن	نوع	کروموزوم
آنتولوژی					
0005802	Trans-Golgi network	15.10	<i>MON2</i>	SNV	1
0005622	Intracellular	27.35	<i>PIK3C2G</i>	Deletion	1
0043235	Receptor complex	5.70	<i>LRP6</i>	SNV	1
0005911	Cell-cell junction	6.60	<i>GPA33</i>	Insertion	1
0005913	Cell-cell adherens junction	3.75	<i>CADM2</i>	SNV	1
0005681	Spliceosomal complex	8.50	<i>DHX15</i>	Deletion	4
0032040	Small-subunit processome	10.40	<i>NOP14</i>	SNV	4
0015630	Microtubule cytoskeleton	5.70	<i>TACC3</i>	SNV	4
0030424	Axon	8.50	<i>HSP90AA1</i>	MNV	5
0000123	Histone acetyltransferase complex	2.80	<i>PHF20</i>	SNV	20
0000124	SAGA complex	2.80	<i>ENY2</i>	SNV	2
0016012	Sarcoglycan complex	2.80	<i>SGCZ</i>	SNV	4

منابع

- Alkan C, Coe BP, Eichler EE (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 12: 363-376.
- Axelsson E, Ratnakumar A, Arendt M, Maqbool K, Webster M, Perloski M, Liberg O, Arnemo J, Hedhammar A, Lindblad-Toh K (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495: 360-365.
- Buyse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mo111rtier G, Speleman F, Menten B (2009). Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. *European Journal of Medical Genetics* 52: 398-403.
- Connerly PL (2010). How do proteins move through the golgi apparatus? *Nature Education* 3: 60.
- Craig TJ, Henley JM (2012). Protein SUMOylation in spine structure and function. *Current Opinion in Neurobiology* 22: 480-487.
- Dharmasiri S, Dharmasiri N, Hellmann H, Estelle M (2003). The RUB/Nedd8 conjugation pathway is required for early development in Arabidopsis. *EMBO Journal* 22: 1762-1770.

- Dias, M. Cánovas A, Mantilla-Rojas C, Riley DG, Luna-Nevarez P, Coleman SJ, Speidel SE, Enns RM5, Islas-Trejo A, Medrano JF, Moore SS, Fortes MR, Venus B, Diaz IS, Souza FR, Fonseca LF, Baldi F, Albuquerque LG, Thomas MG, Oliveira HN (2017). SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. *Genetics Molecular Research* 16, DOI: 10.4238/gmr16019522.
- Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson DC, Dindot SV (2012). Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. *BMC genomics* 13, DOI: 10.1186/1471-2164-13-78.
- Elferink MG, Megens HJ, Vereijken A, Hu X, Crooijmans RPMA, Groenen MAM (2012). signatures of selection in the genomes of commercial and non-commercial chicken breeds. *Plos one* 7: e32720.
- Fleming DS, Koltjes JE, Markey AD, Schmidt CJ, Ashwell CM, Rothchild MF, Reecy MJ, Lamont SJ (2016). Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600K genotyping array. *BMC Genomics*. DOI:10.1186/s12864-016-2711-5.
- Gholami M, Erbe M, Garke C, Preisinger R, Weigend A, Weigend S, Simianer H (2014). Population Genomic Analyses Based on 1 Million SNPs in Commercial Egg Layers. *Plos one* 9: e94509.
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, *et al* (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418: 387–391.
- Groen AJ, Sancho-Andrés G, Breckels LM, Gatto L, Aniento F, Kathryn S. Lilley KS (2014). Identification of trans-golgi network proteins in *Arabidopsis thaliana* root tissue. *Journal dharof Proteome Research* 13: 763–776.
- Kharrati-Koopae H, Esmailzadeh AK (2014). SNPs Genotyping Technologies and Their Applications in Farm Animals Breeding Programs: Review. *Brazilian Archives of Biology and Technology* 57: 87-95.
- Lander ES, Waterman MS (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 2: 231-239.
- Mergner J, Kuster B, Schwechheimer C (2017). DENEDDYLASE1 Protein counters auto-modification of neddylation enzymes to maintain NEDD8 protein homeostasis in *Arabidopsis*. *Journal of Biological Chemistry* 292: 3854-3865.
- Metzker ML (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics* 11:31-46.
- Miller SA, Dykes DD, Polesky HF (1988). A simple salting-out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Researches* 16: 1215.
- Moazeni S, Mohammadabadi MR, Sadeghi M, Shahrabak H, Koshkoieh A, Bordbar F (2016a). Association between UCP Gene Polymorphisms and Growth, Breeding Value of Growth and Reproductive Traits in Mazandaran Indigenous Chicken. *Open Journal of Animal Sciences* 6: 1-8.
- Moazeni SM, Mohammadabadi MR, Sadeghi M, Moradi Shahrabak H, Esmailzadeh AK (2016b). Association of the melanocortin-3(MC3R) receptor gene with growth and reproductive traits in Mazandaran indigenous chicken. *Journal of Livestock Science and Technologies* 4: 51-56.
- 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.

- Mudalal S, Babini E, Cavani C, Petracci M (2014). Quantity and functionality of protein fractions in chicken breast fillets affected by white striping. *Poultry Science* 93: 2108–2116.
- Mukhopadhyay D, Riezman H (2007). Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315: 201–205.
- Rubin C, Zody M, Erikson J, Meadows J, Sherwood E, Webster M, Jiang I, Ingman M, Sharpe T, Ka S, Hallbook F, Besnier F, Carlborg O, Bedhom B, Jensen P (2010). Whole genome resequencing reveals loci under selection during chicken domestication. *Nature Letters* 464: 587-593.
- Shahdadnejad N, Mohammadabadi MR, Shamsadini M (2016). Typing of *Clostridium Perfringens* Isolated from Broiler Chickens Using Multiplex PCR. *Genetics in the 3rd millennium* 14: 4368-4374.
- Tibbles LA, Woodgett JR (1999). The stress-activated protein kinase pathways. *Cellular Molecular Life Science* 55: 1230-1254.
- Xirodimas DP (2008). Novel substrates and functions for the ubiquitin-like molecule NEDD8. *Biochemical Society Transactions* 36: 802-806.
- Zandi E, Mohammadabadi MR, Ezzatkah M, Esmailizadeh AK (2014). Typing of Toxigenic Isolates of *Clostridium Perfringens* by Multiplex PCR in Ostrich. *Iranian Journal of Applied Animal Science* 4: 509-514.
- Zhang H, Wang SZ, Wang ZP, Da Y, Wang N, Hu XX, Zhang YD, Wang YX, Leng L, Tang ZQ, Li h (2012). A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. *BMC Genetics* 13: 704-720.
- Zhu JY, Fu Y, Nettleton M, Richman A, Han Z (2017). High throughput in vivo functional validation of candidate congenital heart disease genes in *Drosophila*. *Developmental Biology, Human Biology and Medicine* DOI: 10.7554/eLife.22617.

Identification of genetic variants in Arian line and investigation of their performance using whole genome sequencing

Kharrati-Koopae H.¹, Ebrahimie E*.², Dadpasand M.³, Niazi A.⁴, Esmailizadeh A.⁵

¹Ph.D student, Institute of Biotechnology, School of Agriculture, Shiraz University, Shiraz, Iran.

²Associate professor, Institute of Biotechnology, School of Agriculture, Shiraz University, Shiraz, Iran.

³Associate professor, Department of Animal science, School of Agriculture, Shiraz University, Shiraz, Iran.

⁴Professor, Institute of Biotechnology, School of Agriculture, Shiraz University, Shiraz, Iran.

⁵Professor, Department of Animal science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran.

Abstract

This is the first study for discovering genetic variants in Arian line by whole genome sequencing data. The study of Arian line features at genomic level can unravel the genetic background of economic traits such as growth. In this study, the information of three whole genomes of Arian chickens were used. The results of this study indicated that, there were 63866 common variants among samples. 17604 variants Out of 63866 variants were reported as novel variants in this study. Also 604 variants are located in coding regions, which 204 out of them lead to change in the amino acid sequence of the proteins. Considering the importance of dietary protein quality for growth and weight gain for poultry, the results of gene ontology analysis indicated that the pathways of protein catabolism and post-translational modification of proteins, related to growth are considerable. Also candidate genes such as, *SMURF2* and *SUMO1* was suggested for these biological pathways. Another interesting result of gene ontology was the significance of response to stress pathways. In the Arian line due to intense selection, the existence of stress is inevitable. Therefore, the significance of stress response can indicate the importance of these pathways in relation to the desired performance.

Keywords: *Arian line, Whole-genome sequencing, Variants.*

* Corresponding Author: Ebrahimie E.

Tel: 09122952785

Email: ebrahimiet@gmail.com.