ژنتیک نوین
دوره چهاردهم، شماره ۱، بهار ۱۳۹۸
صفحه ۸۸ -۸۵

شناسایی ایندل‌ها در ژنوم سگ و گرگ بومی ایران با روش توالی‌یابی کل ژنوم

Detection of deletions and insertions in genome of Iranian dogs and wolves with the method whole genome sequencing

زینب امیری قنات‌سامان[۱]، علی اسمعیلی‌زاده کشکوئیه*[۲]، مسعود اسدی‌فوزی[۲]

۱- دانش‌آموخته دکتری اصلاح نژاد دام، بخش مهندسی علوم دامی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران و عضو انجمن پژوهشگران جوان، دانشگاه شهید باهنر کرمان

۲- به‌ترتیب استاد، دانشیار، بخش مهندسی علوم دامی، دانشکدهٔ کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران

Amiri Ghanatsaman Z[1], Esmailizadeh Koshkoiyeh A*[2], Asadi Fozi M[2]

1- PhD of 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
2- Professor, Associate Professor, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

* نویسنده مسئول مکاتبات، پست الکترونیکی: aliesmaili@uk.ac.ir

(تاریخ دریافت: ۹۷/۰۸/۱٤ - تاریخ پذیرش: ۹۸/۰۳/۲٦)

چکیده

حذف و درج (ایندل) یکی از اجزای اصلی تنوع ژنتیکی و فنوتیپی است که نسبت به چند ریختی‌های تک نوکلئوتیدی و تنوع‌های ساختاری کمتر مورد توجه قرار گرفته است. به‌منظور شناسایی ایندل‌های موجود در ژنوم سگ و گرگ ایرانی و ارزیابی گروه‌های عملکردی مرتبط با آن‌ها، ۲۸۷/۵ گیگابایت داده حاصل از توالی‌یابی کل ژنوم شش قلاده سگ (دو سگ تازی از استان کردستان و یک سگ قهدریجانی از استان اصفهان) و گرگ (از استان‌های تهران، همدان و کرمان) با میانگین عمق پوشش X ١٦ و درصد همپوشانی ۹۹/٤۷ با ژنوم مرجع سگ آنالیز شد. در این تحقیق برای افزایش دقت و صحت ایندل‌های شناسایی شده، از الگوریتم قدرتمند شناسایی واریانت (HaplotypeCaller) استفاده شد. بیش از سه میلیون ایندل (باز ۷۳-۱) حاصل شد. اکثر ایندل‌ها (۹۵/۱۸ درصد) کوچک‌تر از ۱۰ باز بودند. مقایسه نتایج مستندسازی ایندل‌های ژنوم در سگ و گرگ نشان داد که درصد ایندل‌های ژنوم در نواحی اگزون و ناحیه غیر قابل ترجمه (٣′RNA utr) در سگ بیش‌تر از گرگ است. بنابراین فرآیند اهلی‌سازی احتمالاً سبب افزایش تنوع ژنتیکی در نواحی اگزون و ناحیه غیر قابل ترجمه RNA (٣′utr) شده‌است.

واژه‌های کلیدی
اهلی‌سازی سگ
ایندل
ژن آنتولوژی
گروه‌های عملکردی

[Downloaded from mail.mg.genetics.ir on 2026-11-20]

[DOR: 20.1001.1.20084439.1398.14.1.10.4]

به‌تازگی، شناسایی تغییرات ژنتیکی زمینه اهلی‌سازی سگ و مطالعه اثر تغییرات ژنتیکی در طیف گسترده‌ای از فنوتیپ‌ها از جمله صفات مختص هر نژاد مورد علاقه پژوهشگران هستند. تاکنون چند نوع از تنوع ژنتیکی در سراسر ژنوم شناخته شده‌اند که شامل چندریختی‌های تک نوکلئوتیدی، ایندل‌ها و واریانت‌های ساختاری هستند. پژوهش‌ها در انسان نشان دادند که ایندل‌ها یکی از اشکال اصلی تنوع ژنتیکی هستند که از نظر وقوع در ژنوم در رتبه دوم بعد از چند ریختی‌های تک نوکلئوتیدی قرار دارند (Mullaney et al. 2010). در حیوانات اهلی ایندل‌ها مسئول تعدادی از صفات و بیماری‌ها از قبیل صفت ماهیچه مضاعف (Grobet et al. 1997) و کمبود فاکتور XI (Kunieda et al. 2005) در گاو، توده ماهیچه در سگ (Mosher et al. 2007) و نقص اسپرم بی تحرک- دم کوتاه در خوک (Sirinen et al. 2006) هستند. ایندل‌های ۱۵-۹ بازی در ژن PMEL17 باعث رنگ پر در پرندگان (Kerje et al. 2004) و جهش ایندل در ژن گیرنده هورمون رشد باعث کوتولگی وابسته به جنس در مرغ می‌شوند (Agarwa et al. 1994). این مطالعه به‌منظور شناسایی ایندل‌ها، توزیع آن‌ها در ژنوم سگ و گرگ ایران و اثر احتمالی آن‌ها روی عملکرد ژن مورد انجام شد. نتایج حاصل از این تحقیق در درک عمیق پایه ژنتیکی اهلی‌سازی سگ و ارتباط بین تنوع‌های ژنتیکی و فنوتیپی می‌تواند مفید باشد.

برای انجام این پژوهش از سه قلاده سگ (دو سگ تازی از استان کردستان و یک سگ قهدریجانی از استان اصفهان) و سه قلاده گرگ (از استان‌های تهران، همدان و کرمان) بومی ایران نمونه به صورت بافت و خون تهیه شد. استخراج DNA از نمونه‌های خون با روش استخراج نمکی و از نمونه‌های بافت با روش کلروفورم ایزوآمیل انجام شد. توالی‌یابی کل ژنوم به‌صورت Paired-End در خوانش‌هایی با اندازه ۱۲۵ باز توسط شرکت ایلومینا Hiseq 2500 در کشور چین انجام شد. کیفیت داده‌ها توسط برنامه fastqc بررسی شد. برای هم‌ردیفی Short reads با ژنوم مرجع (*canfam* 3.1) از الگوریتم MEM در برنامه BWA[1] (Li and Durbin 2009) استفاده شد. کلیه SAM[2] فایل‌های تولید شده توسط برنامه Samtools (Li et al. 2009) به BAM[3] فایل تبدیل شدند. حذف خوانش‌های دو نسخه شده (Duplicated reads) با برنامه‌ی picard صورت گرفت. به‌منظور محاسبه درصد هم‌ردیفی با ژنوم مرجع و عمق پوشش از برنامه‌های flagstat و depth در نرم‌افزار Samtools استفاده شد. با استفاده از الگوریتم HaplotypeCaller به‌کار برده شده در برنامه GATK (Mckenna et al. 2010) ایندل‌ها شناسایی شدند و به‌وسیله برنامه Variant Filtration به‌کار برده شده در این برنامه فیلتر شدند. هیستوگرام‌های توزیع طول ایندل‌های ژنوم با استفاده از دستور hist-indel-len– به‌کار برده شده در برنامه VCFtools (Danecek et al. 2011) ترسیم شدند. مستندسازی یا انوتیشن ایندل‌ها با برنامه SnpEff 4.0e (Cingolani et al. 2012) انجام شد. ID ژن‌هایی که ایندل‌های ژنوم سگ و گرگ را شامل شدند، به IDs ژن‌های انسان تبدیل شدند. فایل رابط ژن ارتولوگوس بین انسان و سگ از پایگاه داده Ensemble دانلود شد. سپس ژن آنتولوژی و آنالیز مسیر KEGG[4] با استفاده از برنامه DAVID (Huang et al. 2009) انجام شد. مقادیر P-value کمتر از ۰/۰۵ به‌عنوان مقادیر معنی‌دار برای نتایج حاصل از ژن آنتولوژی در نظر گرفته شد.

حدود ۲۸۷/۵ گیگابایت داده حاصل از توالی‌یابی کل ژنوم شش قلاده سگ و گرگ به‌دست آمد. میانگین عمق پوشش و درصد همپوشانی با ژنوم مرجع به‌ترتیب X ۱۶ و ۹۹/۴۷ درصد بود. تعداد کل ایندل‌ها در این مطالعه ۳/۴۸ میلیون بود که ۲/۲۴ میلیون و ۳/۱۱ میلیون به‌ترتیب مربوط به نمونه‌های سگ و گرگ بودند (جدول ۱). حداقل و حداکثر طول ایندل‌ها ژنوم به‌ترتیب برابر با یک و ۷۳ بوده است (جدول ۱). در مطالعه Stafuzza et al. (2017) به‌منظور شناسایی ایندل‌ها ژنوم در گاو، اندازه ایندل‌ها ۴۳-۱ باز گزارش شد. در مطالعه‌ی که Yan et al. (2014) بر روی ژنوم مرغ با استفاده از تکنیک توالی‌یابی کل ژنوم انجام دادند، اندازه ایندل‌های ژنوم بین ۴۹-۱ باز گزارش شد. بیش‌تر ایندل‌ها دارای طولی کمتر از ۱۰ باز بودند (۹۵/۱۷۹ درصد در ژنوم شش قلاده سگ و گرگ) (شکل ۱). نتایج مشابهی در مطالعه آنالیز ژنوم در گاو (Stafuzza et al. 2017) و مرغ (Yan et al. 2014) گزارش

[1] Burrows-Wheeler Aligner
[2] Sequence Alignment Map
[3] Binary Alingnment MAP
[4] KEyoto Encyclopedia of Genes and Genomes

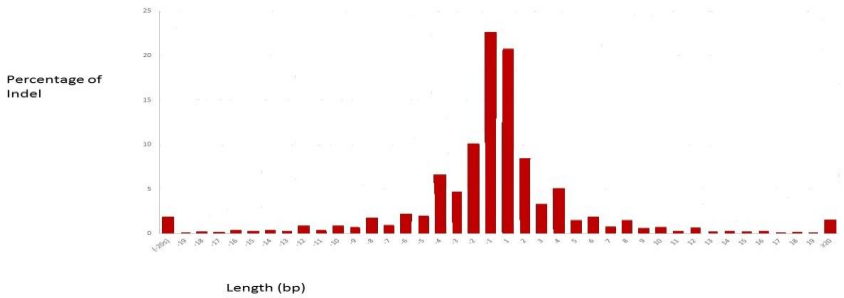
[Downloaded from mail.mg.genetics.ir on 2026-11-20]
[DOR: 20.1001.1.20084439.1398.14.1.10.4]

شد. درصد ایندل‌ها در ناحیه اگزون در سگ و گرگ به‌ترتیب ۰/۲۷ و ۰/۲٤ درصد و درصد ایندل‌ها در ناحیه ٣′utr در سگ و گرگ به‌ترتیب ۰/٤۸ و ۰/٤٦ درصد بوده است. نتایج نشان می‌دهند که درصد ایندل‌ها در نواحی اگزون و ٣′utr در سگ بیش‌تر از گرگ است. ناحیه ٣′utr اغلب نواحی تنظیمی را در بر می‌گیرد که بعد از رونویسی بیان ژن را تنظیم می‌کنند (Matoulkova et al. 2012). به‌نظر می‌رسد که اهلی‌سازی و فرآیندهای مرتبط با آن از قبیل انتخاب Relaxation باعث افزایش درصد تنوع‌های ژنومیک در نواحی کدکننده ژن و نواحی تنظیمی بیان ژن در ملکول mRNA از قبیل ٣′utr در سگ شده‌است (Björnerfeldt et al. 2006; Cruz et al. 2008).

در این مطالعه ۲۱۱۰٤ ژن از طریق مستندسازی ۳/٤۸ میلیون ایندل در سر تا سر ژنوم شش قلاده سگ و گرگ به‌دست آمد. ژن آنتولوژی و آنالیز مسیر KEGG عملکرد این ژن‌ها را در سه گروه عملکردهای ملکولی، فرآیندهای زیستی و اجزای سلولی قرار داد. اصطلاح‌های حس درک صدا (Go:۰۰۰۷٦۰۵) و درک دید (Go:۰۰۰۷٦۰۱) در میان اصطلاح‌های دسته فرآیندهای زیستی هم در سگ و هم در گرگ معنی‌دار بودند که حاکی از اهمیت ژن‌های شنوایی و بینایی در حیات سگ‌سانان است. همچنین بسیاری از این اصطلاح‌های معنی‌دار مرتبط با تولید ماده ژنتیکی DNA (تعمیر DNA (Go:۰۰۰٦۲۸۱) و تکثیر DNA (Go:۰۰۰٦۲٦۰)) یا RNA (تنظیم مثبت رونویسی از روی DNA الگو (Go:۰۰٤۵۸۹۳)، رونویسی از روی آنزیم RNA پلیمراز II (Go:۰۰۰٦۳٦٦)، (کمپلکس فاکتور رونویسی (Go:۰۰۰۵٦٦۷)، اتصال فاکتور رونویسی (Go:۰۰۰۸۱۳٤)، برش و دوخت mRNA از طریق اسپلایسوم (Go:۰۰۰۰۰۳۹۸)) هستند. همچنین بسیاری از اصطلاحات مرتبط با تنظیمات اپی‌ژنتیکی (بازآرائی کروماتین (Go:۰۰۰٦۳۳۸) و تغییر کروماتین (Go:۰۰۱٦۵٦۸)) در سگ و گرگ معنی‌دار بودند. نتایج این تحقیق همانند نتایج مطالعات قبلی (Khalkhali Ivriq et al. 2018) نشان می‌دهند که ژن‌های دخیل در همانندسازی، تعمیر و رونویسی DNA از جمله ژن‌هایی هستند که توسط ایندل‌ها دست خوش تغییر شدند. اصطلاح‌های ژن آنتولوژی از قبیل "طرح نورون" و" کمپلکس گیرنده" فقط در گرگ دیده شدند. اصطلاح طرح نورون فرآیندی است که از یک سلول عصبی مانند اکسون یا دندریت توسعه می‌یابد. نتایج این تحقیق همانند مطالعات قبلی بر کاهش رفتار تهاجمی سگ در طی فرآیند اهلی‌شدن دلالت دارند (Axelsson et al. 2013). نتایج این تحقیق نشان داد بسیاری از ژن‌های مهم و حیاتی در ژنوم سگ‌سانان توسط ایندل‌ها دستخوش تغییرات شده‌اند. با این وجود برای شناسایی تغییرات ژنتیکی زمینه اهلی‌سازی سگ، مطالعه اجزای بیشتری از منابع تنوع ژنتیکی در مطالعات پیش رو پیشنهاد می‌شود.

جدول ۱- خلاصه‌ای از آماره‌های ایندل‌ها در ژنوم سگ و گرگ

نمونه	نوع واریانت	تعداد	حداقل طول	حداکثر طول
سگ	تعداد کل	۲۲٤۳۳۵۳	۱	٦۵
گرگ	تعداد کل	۳۱۰۸۹۹۰	۱	۷۳
سگ-گرگ (کل داده‌ها)	تعداد کل	۳٤۸٤۸٤۵	۱	۷۳



شکل ۱- منحنی توزیع طول ایندل‌ها در سراسر ژنوم شش قلاده سگ و گرگ، مقادیر مثبت و منفی به‌ترتیب نشان‌دهنده اضافه و حذف می‌باشند.

[Downloaded from mail.mg.genetics.ir on 2026-11-20]
[DOR: 20.1001.1.20084439.1398.14.1.10.4]

منابع

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 Endocrinol 142:427-434.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, LindbladToh K (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360-364.

Björnerfeldt S, Webster MT, Vilà C. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research 16:990-994.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Ruden DM (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.

Cruz F, Vilà C, Webste MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. Molecular Biology and Evolution 25:2331-2336.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Durbin R (2011) The variant call format and vcftools. Bioinformatics 27:2156-2158.

Grobet L, Martin LJR, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Ménissier F, Massabanda J, Fries R (1997) A deletion in the bovine myostatin gene causes the double–muscled phenotype in cattle. Nature Genetics 17:71.

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4:44-57.

Kerje S, Sharma P, Gunnarsson U, Kim H, Bagchi S, Fredriksson R, Schütz K, Jensen P, Von Heijne G, Okimoto 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.

Khalkhali Ivriq R, Hafezian SH, Hedayat Ivriq N, Farhadi A, Bakhtiarizadeh MR (2018) Detection and analysis of deletions and insertions in genome of Iranian dromedary camels using whole genome sequencing data. Journal of Animal Production 20:109-120. (In Farsi).

Kunieda M, Tsuji T, Abbasi AR, Khalaj M, Ikeda M, Miyadera K, Ogawa H, Kunieda T (2005) An insertion mutation of the bovine F11 gene is responsible for factor XI deficiency in Japanese black cattle. Mammalian Genome 16:383-389.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R (2009) The sequence alignment/map format and sam tools. Bioinformatics 25:2078-2079.

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

Matoulkova E, Michalova E, Vojtesek B, Hrstka R (2012) The role of the 3'untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biology 9:563-576.

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.

Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, Ostrander EA (2007) A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. PLoS Genetics 3: e79.

Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. Human Molecular Genetics 19:131-136.

Sironen A, Thomsen B, Andersson M, Ahola V, Vilkki J (2006) An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig. Proceedings of the National Academy of Sciences of the United States of America 103:5006-5011.

Stafuzza NB, Zerlotini A, Lobo FP, Yamagishi MEB, Chud TCS, Caetano AR, Carvalho MR (2017) Single nucleotide variants and InDels identified from whole-genome resequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. PLoS One 12: e0173954.

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 9:104652.

[Downloaded from mail.mg.genetics.ir on 2026-11-20]

[DOR: 20.1001.1.20084439.1398.14.1.10.4]