Iranian Journal of Field Crop Science
Vol 51, No 3, Autum 2020 (103-114)
DOI: 10.22059/ijfcs.2019277138.654586

علوم گیاهان زراعی ایران
دورهٔ ۵۱، شمارهٔ ۳، پائیز ۱۳۹۹ (ص۱۱۴-۱۰۳)

کشف، تعیین و ژنوتیپ‌سنجی نشانگرهای SNP و گروه بندی لاین‌های پیشرفته گندم نان به روش ddRAD-Seq

محمد آرمیون[۱]، محمد رضا بی‌همتا*[۲]، رضا معالی‌امیری[۳]، منوچهر خدارحمی[٤]، ساچیکو ایزوبه[٤]

۱و۲و۳- به ترتیب دانشجوی دکتری، استاد و دانشیار، گروه زراعت و اصلاح نباتات، دانشکده علوم و مهندس کشاورزی دانشگاه تهران،
٤- استادیار موسسه تحقیقات اصلاح و تهیه نهال و بذر، بخش تحقیقات غلات، سازمان تحقیقات، آموزش و ترویج کشاورزی کرج، ۵- رئیس آزمایشگاه ژنومیکس و ژنتیک کاربردی گیاهی، موسسه تحقیقات **DNA** کازوسا ژاپن

(تاریخ دریافت: ۱۳۹۷/۱۲/۱۱ - تاریخ پذیرش: ۱۳۹۸/٦/۲۷)

چکیده

به‌منظور کشف، تعیین و ژنوتیپ‌سنجی نشانگرهای SNP و گروه بندی لاین‌های پیشرفته گندم نان، از روش ddRAD-Seq استفاده شد. DNA از برگ گیاهچه استخراج شد و پس از تهیه کتابخانه، از پلاتفرم NextSeq ™ 500 Illumina ® برای توالی‌یابی استفاده شد. متوسط نمره کیفیت فرد (Phred) برای تمام افراد برابر ۳۰ بود. از مجموع ۱۷۸۸۱۱۸٤٦ خوانش توالی قرائت شده، ۱۵۰۱۰۸٦۷۸ خوانش صحیح و به‌طور متوسط به ازای هر لاین، ۲۲۰۷٤۸۰ خوانش تولید شد. بالاترین نرخ هم ردیفی، مربوط به ژنوم B و کمترین، مربوط به ژنوم D بود. با توجه به شرایط فیلتر، (DP≥5)، (Quality score=≥999)، (MAF>5%) و (Het<10%)، تعداد کل SNP های صحیح فراخونی شده برای ۵۰ درصد داده گم‌شده برابر با ۳۳٤۲ عدد، که تعداد ۱۳۲۲ روی ژنوم B، ۱۲۵۳ روی ژنوم A و ۷٦۷ روی ژنوم D و بیشترین SNP روی کروموزوم دوم و سوم از ژنوم B و کمترین آن روی کروموزوم چهارم از ژنوم D شناسائی شد. بیشترین چگالی نشانگری، روی کروموزوم‌های دوم و پنجم ژنوم B و کمترین، روی کروموزوم چهارم از ژنوم D مشاهده شد. بین چگالی نشانگری و اندازه کروموزوم در سه ژنوم، رابطه خطی بسیار معنی‌داری مشاهده شد. تجزیه به مولفه‌های اصلی و ماتریس تشابه و گروه‌بندی همزمان با استفاده از اطلاعات نشانگر SNP، قادر به شناسائی و تفکیک زیر جمعیت‌ها از یک جمعیت اصلی شد.

واژه‌های کلیدی: والی‌یابی NextSeq ™ 500، روش ddRAD-Seq، گندم نان، نشانگر SNP، DNA.

Discovery and genotyping of SNP markers and grouping of advanced bread wheat lines by ddRAD-Seq

Mohammad Armion[1], Mohammad Reza Bihamta*[1], Reza Moali Amiri[1], Manuchehr Khodarahmi[2], Sachiko Isobe[3]

1. Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Tehran, 2. Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Tehran, 3. Department of Applied Plant GenomicsKazusa DNA Research Institute, Kisarazu, Chiba, Japan.

(Received: March 2, 2019 - Accepted: September 18, 2019)

ABSTRACT

The aim of this study was to discover, genotyping and determining the genotype, the number, distribution and density of SNP markers and grouping of an advanced breeding population using the ddRAD-Seq method. DNA was extracted from 14-old-day seedlings and the NextSeq ™ 500 Illumina ® platform was used for sequencing. The average quality score for all individual was Phred's 30. The correct reads were 150108678 out of 178811846 and the average of 2207480 reads produced by individual. The highest and the lowest alignment rate were related to B and D genomes, respectively. Based on the filter conditions, (DP≥5), quality score=≥999, MAF>5% and Het<10%, the total number of SNP calling for 50% missing data were 3342 which identified 1322, 1253, and 767 on B, A and D genomes, respectively. The highest SNP markers were identified on 2B and 3B and the lowest on 4D chromosomes. A significant linear regression was observed between marker density (SNP/Mbp) and chromosome size in three genomes. The principal components analysis and the heatmap dendrogram together with the use of SNP marker information were able to identify and segregate sub-populations from a main population.

Keywords: Bread wheat, DNA, ddRAD-Seq method, NextSeq ™ 500 Sequencing, SNP marker.

* Corresponding author E-mail: mrghanad@ut.ac.ir

مقدمه

گندم یک محصول زراعی مهم، اصلی و عمده است که در بیشتر قسمت‌های جهان رشد می‌کند (Kristensen *et al*, 2018). دو حالت اصلی از آلوپلی‌پوئیدی گندم یعنی هگزاپلوئید (*Triticum aestivum* ssp. *aestivum*) با اندازه ژنوم حدود ۱۷ Gb و تتراپلوئید (*Triticum turgidum* ssp. *durum*) با اندازه ژنوم ۱۲ Gb وجود دارد (Borrill *et al*., 2015). بیش از ۹۵ الی ۹۹ درصد از نواحی کدکننده در کروموزوم‌های همیولوگ‌های گندم مشابه هستند (Krasileva *et al*., 2013). اندازه عظیم ژنوم گندم، درجه شباهت زیاد توالی‌ها و محتوای DNA تکراری آن، مانع جدی در تولید نسخه اولیه توالی ژنوم گندم بود. در نتیجه پژوهشگران در ابتدا روی توالی‌های کد کننده (CDS) و مجموعه‌های بزرگ توالی‌های نشانمند ترجمه شده (ESTs) فعالیت نمودند و یک نقشه مونتاج شده از این توالی‌ها تولید کردند. از طرفی، ظهور فن‌آوری NGS رویکرد جدیدی برای غلبه بر مشکلات توالی‌یابی ژنوم گندم فراهم نمود. فن‌آوری نسل جدید دوم و سوم NGS، باعث دگرگونی اساسی در آنالیز ژنوم و دررنتیجه افزایش فهم ما از رابطه ژنوتیپ - فنوتیپ شده است (Mochida *et al*., 2009). توالی‌یابی DNA یک فرآیند دقیق است که ترتیب صحیح نوکلئوتیدها در یک مولکول DNA را تعیین می‌کند (El-Metwally *et al.,* 2014). ظهور فن‌آوری‌های توالی‌یابی، نقش مهمی در تجزیه و تحلیل توالی‌های ژنومی ارگانیسم‌ها بازی کرده است (Shendure & Ji, 2018). اولین فن‌آوری توالی‌یابی در سال ۱۹۷۷ به وسیله Sanger و Maxam - Gilbert ایجاد شد. فن‌آورهای نسل جدید تحت عنوان فن‌آوری توالی‌یابی NGS یا فن‌آوری های توالی‌یابی‌خودکار با مقیاس بالا، در سال ۲۰۰۵ با فن آوری Roche's 454 معرفی شدند. فن‌آورهای NGS آنالیز حجیم موازی بصورت خودکار از چندین نمونه با هزینه خیلی کم تولید می‌کنند و قادر به توالی‌یابی میلیون‌ها تا میلیاردها خوانش به‌صورت موازی و همزمان در یک مرحله اجرا هستند و زمان مورد نیاز برای تولید خوانش‌های با اندازه گیگابایت اطلاعات، تنها چند روز یا ساعت است (Mardis, 2011).

در زمینه تحقیقات گیاهی، فن‌آوری‌های NGS یکی از ابزارهای مهم برای تشکیل اسمبلی ژنوم‌های مرجع گیاهان زراعی، توالی‌یابی ترانسکریپتوم برای تحقیقات بیان ژن، توسعه نشانگرهای مولکولی در سطح ژنوم و شناسائی نشانگرها در ژن‌های با عملکرد معین شد (Vlk & Repkova, 2016). پلاتفرم Roche/454 در سال ۲۰۰۵، Illumina/Solexa در سال ۲۰۰۶ و ABI/SOLiD در سال ۲۰۰۷ ایجاد شده‌اند (Kchouk et al., 2017). اولین توالی گندم با استفاده از فن آوری 454 و بر اساس توالی‌یابی شات‌گان (WGS) یا شکستن کل ژنوم در سال ۲۰۱۲ منتشر شد (Brenchley *et al*., 2012). تولید توالی‌های مرجع CSS (IWGSC, 2014) با ابعاد نقشه 10.2 Gb و W7984 (Chapman *et al*., 2015) با 9.1 Gb اولین قدم مهم نسبت به کشف تنوع در بین گونه‌های گندم بود (Borrill *et al*., 2015). آخرین ژنوم مرجع معرفی شده از سوی کنسورتیوم بین المللی توالی‌یابی گندم به نام IWGSC RefSeq v1.0 reference با اندازه نقشه 14.5 Gb و عمق ۹۴ درصد بود که از داده های POPSEQ و نقشه HiC برای تهیه آن استفاده شده است (IWGSC, 2018). از طرفی، اندازه بزرگ و پیچیدگی ژنوم گندم، امکان توالی‌یابی مجدد کل ژنوم واریته گندم جدید با توجه به فن‌آوری‌های جاری، از نظر اقتصادی میسر نیست. بنابراین، روش‌های کاهش اندازه ژنوم برای دسترسی به تنوع درون گونه‌ها که تقریبا بر SNP ها متمرکز شده به‌طور گسترده در گندم استفاده شده است (Borrill *et al*., 2015). دو رویکرد GBS (Poland & Rife, 2012) و ddRAD (Shirasawa *et al*., 2016) برای کشف و ژنوتیپ‌سنجی همزمان SNP با استفاده از آنزیم‌های برشی و کاهش اندازه ژنوم بوجود آمده است (Vlk & Repkova, 2016). ژنوتیپ‌سنجی SNP بوسیله NGS به صورت GBS، RAD-Seq و ddRAD-Seq به خاطر انعطاف پذیری و هزینه پایین متداول شده‌اند (Shirasawa *et al*., 2016). روش‌های GBS و RAD-Seq تولید خوانش از یک انتها (single-end) نموده، در صورتیکه روش ddRAD-Seq

تولید خوانش از دو انتها (paired-end) می‌نماید، و در نقشه‌یابی خوانش‌ها نسبت به ژنوم مرجع به ویژه درگیاهان که اغلب دارای ژنوم پلی‌پلوئیدی بزرگ و پیچیده بوده، صحیح‌تر است. در روش ddRAD-Seq از دو آنزیم برشی، که آنزیم دوم به منظور کاهش هزینه و زمان آماده سازی کتابخانه و توالی‌یابی دو طرفه از ژن-های همسان برای تمام نمونه‌ها، استفاده می‌شود. بنابراین از نقطه نظر صحت بالاتر در نقشه‌یابی خوانش ها، حتی در ژنوم‌های گیاهی پیچیده، فن‌آوری ddRAD-Seq سودمندی آن نسبت به GBS و RAD-Seq مناسب‌تر است (Shirasawa *et al*., 2016). روش ddRAD-Seq به‌وسیله Shirasawa *et al*. (2016) در گوجه فرنگی بررسی و شیوه تحلیل داده‌های تجربی و in silico را معرفی کردند. Arafa *et al*. (2017) از روش ddRAD-Seq برای شناسائی ژن‌های کاندید مقاومت به بیماری بادزدگی در جمعیت نسل دوم گوجه فرنگی استفاده نمودند. DaCosta & Sorenson (2016) نتیجه گرفتند که روش ddRAD-Seq یک روش مقرون به صرفه برای تولید داده‌های فیلوژنی قوی، به‌ویژه برای تجزیه و تحلیل گونه و جنس نزدیک، است. تنوع ژنتیکی کلید اصلی برای موفقیت در به‌نژادی و مبنای اساسی برای به‌نژادگران برای انتخاب مداوم ارقام با عملکرد اصلاح شده است. از طرفی انتخاب شدید در دوره اهلی‌سازی و به‌نژادی، باعث حذف تنوع ژنتیکی قابل ملاحضه‌ای در خزانه‌های اصلاحی گیاهان اصلی شده است و فرسایش پتانسیل ژنتیکی برای سازگاری در برابر تغییرات مانند آب و هوا را در پی داشته است. فناوری‌های با توان بالا ژنومی، از طریق ارائه دانش دقیق نسبت به توصیف و پرکردن تنوع ژنتیکی در برنامه‌های به‌نژادی، قادر به حل این معضل هستند. اطلاع از گروه-بندی جمعیت و روابط ژنتیکی مبنائی برای ایجاد گروه-های هتروزیس غیر متقارب برای افزایش میزان هتروزیس در رویکردهای به‌نژادی گندم هیبرید مهم است (Melchinger, 1999). از طرفی، مبادله شدید واریته‌های الیت درون برنامه‌های به‌نژادی گندم که به طور سنتی روی راهبردهای اینبریدینگ به جای تشکیل خزانه هیبرید تکیه کرده‌اند، باعث کاهش شدید تمایز بین مواد الیت شده است (Voss-Fels *et al*., 2015). Wang *et al* (2017)، در بررسی یک پنل ۱۰۵ عددی از واریته‌های گندم و لاین‌های پیشرفته با استفاده از آرایه ۹۰K، چهار زیر جمعیت براساس هشت مولفه اصلی PC و آنالیز شباهت ژنتیکی به‌وسیله نشانگر SNP با روش هیت‌مپ، گروه‌بندی نمودند. بنابراین هدف این تحقیق، تعیین ژنوتیپ، تعداد، پراکنش و چگالی نشانگر SNP و گروه‌بندی و تفکیک زیر جمعیت‌ها بر اساس فاصله ژنتیکی نشانگر SNP برای یک جمعیت پیشرفته اصلاحی است.

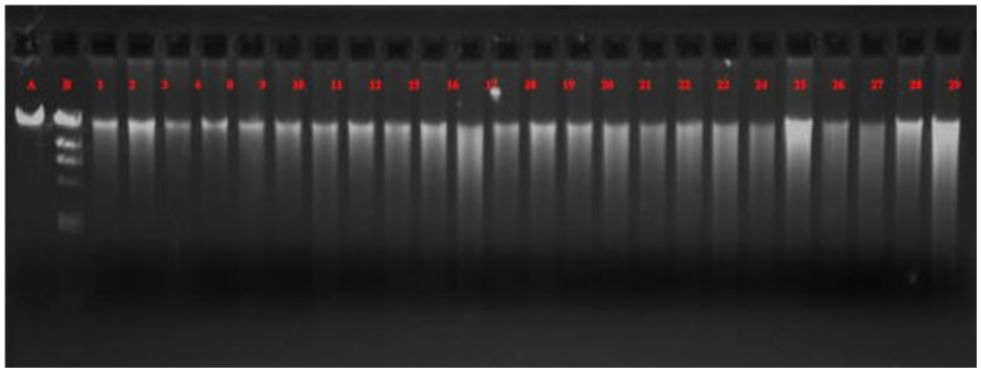
مواد و روش‌ها

مواد ژنتیکی

در این پژوهش، از ۵۰ ژنوتیپ و رقم زراعی حاصل از انتخاب لاین‌های برتر از یک آزمایش مشترک مقدماتی تحت عنوان PRBWYT از ایستگاه‌های گرگان، گنبد، ساری و مغان برای اقلیم گرم و مرطوب شمال موسسه تحقیقات اصلاح و تهیه نهال، استفاده شده است (جدول ۱ ضمیمه).

استخراج، تعیین سلامت و رقیق سازی DNA

به‌منظور استخراج DNA، از بافت برگ گیاهچه‌های با سن چهارده روز نمونه‌برداری شد و نمونه‌ها سریعا به آزمایشگاه موسسه تحقیقات DNA کازوسا (www.kazusa.or.jp/en/) انتقال یافتند و در فریزر ۸۰- درجه ذخیره شدند. از دستگاه Tissuelyser II (Qiagen Inc., Hilden, Germany) برای آسیاب و از پروتکل شرکت کیاژن (The DNeasy Plant Mini kit) برای استخراج DNA **نمونه**ها استفاده شد. غلظت اولیه DNA با دستگاه Nano drop ND-1000 تعیین شد. برای تعیین سلامت DNA از ژل آگارز ۰/۷ درصد، بافر 1X BPB و دو نشانگر لاندا (λ) یک باندی و λ |HinIII چند باندی استفاده شد (Liljeroth & Bryngelsson, 2002). کمیت دقیق‌ترDNA به‌وسیله کیت‌های Qubit® dsDNA BR Assay Kits (Thermo Fisher Scientific Inc., Waltham, USA) و با دستگاهQubit® 30 Fluorometer تعیین شد (شکل ۱).



شکل ۱- تعیین سلامت DNA استخراج شده نمونه‌ها.

Figure1. Health determination of DNA extracted from samples.

پروتکل ddRAD-Seq

پس از استخراج DNA ژنومیک با کیفیت مناسب و برای کاهش اندازه و برش ژنوم، از دو آنزیم برشی FastDiges PstI و FastDiges MspI و بافر 10X FastDigest Buffer استفاده شد (Thermo Fisher Scientific Inc., Waltham, USA) و پلیت DNA ژنومیک به مدت دو دقیقه با سرعت ۱۰۰ × ۱۰۰ rpm سانتریفیوژ و در زمان ۱۶ ساعت با دمای ۳۷ درجه سانتیگراد در دستگاه PCR انکوبات شد. پس از مرحله هضم، آداپتورهای RAD-ad2-PstI ,RAD-ad1-PstI Adaptor(50µm) RAD-ad1-MspI ،Adaptor(50µm) RAD-ad2-MspI Adaptor(50 و Adaptor(50µm) (µm، آنزیم‌های FastDiges PstI و FastDiges MspI، آنزیم T4 DNA ligase (Promega, Madison, WI, USA) و بافر 2x Rapid ligation buffer به DNA ژنومیک اضافه، و سپس پلیت نمونه‌ها در دستگاه PCR برای شرایط ۱۶ درجه سانتیگراد به مدت ۳۰ دقیقه و ۳۷ درجه به مدت ۱۰ دقیقه و برای ۲۵ سیکل انکوبات شد. برای حذف قطعات با طول کمتر از ۳۵۰ bp از محلول گوی‌های مغناطیسی AMPure beads (Beckman Coulter, Brea, CA, USA) استفاده شد. پس از رقیق سازی DNA ژنومیک به نسبت یک به ۱۰۰ و قبل از تشکیل کتابخانه مشترک، اندیکس Read1_N5**/Read2_N7**(0.5uM) و مخلوط ترکیبات -plus- KOD buf 10x، dNTP، MgSO4، KOD –plus– ver.2 (Toyobo, Osaka, Japan) اضافه شد. برنامه ddRAD-Seq برای دستگاه PCR و تکثیر قطعات به صورت؛ آغاز عمل واسرشت‌سازی در دمای ۹۵ °C به مدت سه دقیقه، ۲۰ سیکل بصورت عمل واسرشت: دمای ۹۴ °C در ۳۰ ثانیه، اتصال: دمای ۵۵ °C در ۳۰ ثانیه، تکثیر: دمای ۷۲ °C در یک دقیقه و سه دقیقه انتهایی بعد از اتمام دوره سیکل عمل تکثیر در دمای ۷۲ °C ، انتخاب شد. برای تعیین غلظت DNA ژنومیک از بافر (Qubit® dsDNA HR Buffer) و محلول واکنشگر (Qubit® dsDNA HS Reagent) با حساسیت بالا و دو استاندارد یک (Qubit™ ds DNA HS) و دو (Qubit™ ds DNA HS) استفاده شد. پس از PCR، از هر نمونه حدود چهار میکرولیتر برداشته و در هشت تا نه میکروسانتریفیوژ ۲۰۰ میکرولیتر جمع‌آوری شد و سپس همه مقادیر در یک میکروسانتریفیوژ ۱/۵ میلی‌لیتری جمع شدند و کتابخانه (Pooling) تشکیل شد که به‌وسیله دستگاه کیوبیت، غلظت آن سنجیده شد. با استفاده از دستگاه BluePippin قطعات DNA ژنومیک در رنج 500~800bp انتخاب و خالص‌سازی شدند و کتابخانه آزمایشگاهی برای انجام توالی‌یابی آماده شد. تعیین کمیت و کیفیت DNA ژنومیک کتابخانه به‌وسیله دستگاه‌های Qubit (ds DNA HS) و Tape Station Hs D1000صورت گرفت. برای تعیین غلظت دقیق DNA ژنومیک برحسب پیکومولار، از کیت Kapa Library Quantification Kit(KAPA Kit) استفاده و غلظت با دستگاه 7900HT Fast Real-Time PCR system شرکت Life Technologies تعیین شد. پس از تصحیح رقت کتابخانه به یک nM، عمل واسرشت و رقیق سازی

بوسیله محلول 0.2N NaOH، بافر HT1 و PhiX control انجام گرفت. عملیات بارگذاری کتابخانه در کاتریج واکنشگر آماده سازی و با دستگاه 500 NextSeq ™ Sequencer توالی‌یابی انجام گرفت.

فرآیند محاسبات روش ddRAD-Seq تجربی

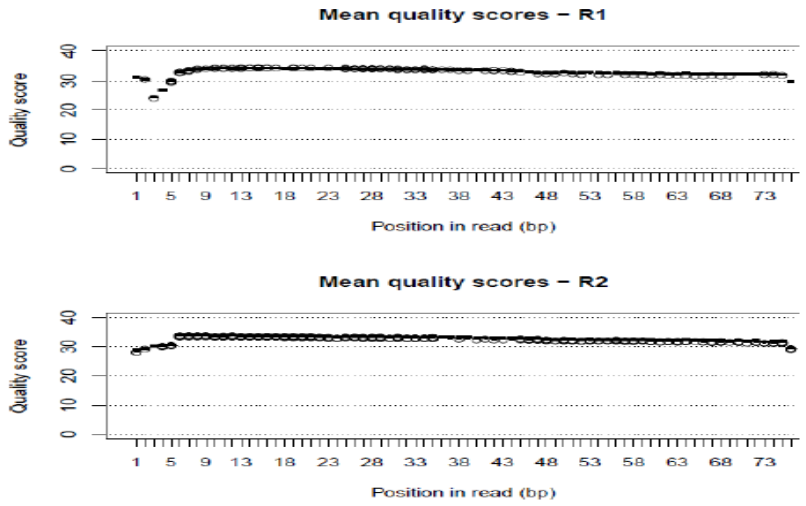
پس از خاتمه عمل توالی‌یابی کتابخانه، اطلاعات در قالب فایل qseq file format(qSEQ) آماده و ارائه شد و بر حسب نوع فن‌آوری هر دستگاه، یک فایل تعیین کیفیت توالی به نام fastQC تهیه شد که معمولا بر مبنای نمره فرد (Phred) تولید می‌شود. نمره کیفیت فرد شاخصی است که کیفیت شناسائی بازهای نوکلئوتیدی که به‌وسیله سکونسرها به صورت خودکار تولید شده است را بررسی می‌کند. پس از این مرحله، برای استخراج توالی‌های صحیح، از پاپلاین‌های بیوانفورماتیک استفاده شد (Shirasawa *et al*., 2016). ابتدا توالی‌های با کیفیت پایین (خوانش یا رید) به‌وسیله نرم افزار PRINSEQ و سپس آداپتورها به‌وسیله fastx_clipper حذف شدند. پس از آن، خوانش‌های فیلتر شده نسبت به توالی‌های مرجع از جمله IWGSC RefSeq v1.0، CSS، W7984 و IWGSC-WGA v0.4 (IWGSC, 2018) با استفاده از نرم افزار bowtie2 نقشه‌یابی شدند. فایل فرمت نتیجه همردیف/ نقشه SAM (Sequence Alignment Map) به فایل فرمت همردیف/ نقشه باینری تبدیل شد و سپس برای قرائت SNP استفاده شد و در نهایت، فایل VCF (variant call format) تولید شد. برای مشاهده موقعیت توالی‌های SNP برای هر ژنوم و کروموزوم و تعداد آن‌ها، از نرم‌افزار TASSEL ورژن 5.2.33 (Bradbury *et al*., 2007)، استفاده شد. روش تجزیه به مولفه‌های اصلی (PCA) بر اساس روش گروه بندی k – means و مقادیر حساب شده بر اساس فاصله تعدیل شده Roger (MRD) برای ۳۳۴۲ نشانگر پلی‌مرفیسم SNP و ۵۰ لاین و رقم زراعی اجرا شد. هیت‌مپ خویشاوندی بین ۵۰ لاین که در آن دندروگرم‌ها با استفاده از روش میانگین جفتی عدم وزنی (UPGMA) بر حسب فاصله اقلیدسی برای ۳۳۴۲ نشانگرها پلی-مرفیسم SNP ترسیم شد و درجه ارتباط به‌وسیله رنگ (قرمز = ارتباط قوی) نشان داده شد. برای اجرای PCA از دستورات (k-means) و هیت‌مپ؛ (heatmap) در محیط نرم‌افزار R ورژن 3.3.2 (R Core Team, 2016) استفاده شده است.

نتایج و بحث

کیفیت باندی حاصل از ازوش استخراج شرکت کیاژن مناسب بود (شکل۱). این قدم اولیه در تهیه کتابخانه مناسب و تامین شرایط برای توالی‌یابی پلاتفرم Illumina® است (Karaaslan *et al.,* 2014). طول خوانش‌های تولید شده از پروتکل ddRAD-Seq معادل ۹۳ bp که پس از پیراش به ۷۶ bp رسید. متوسط نمره کیفیت توالی‌یابی برای هر دو طرف (Paired-end) برای کتابخانه ۳۰ فرد بود (شکل ۲).

از مجموع ۱۷۸۸۱۱۸۴۶ خوانش توالی قرائت شده، ۱۵۰۱۰۸۶۷۸ خوانش صحیح یعنی معادل ۸۴ درصد و به‌طور متوسط به ازای هر لاین ۲۲۰۷۴۸۰ خوانش تولید شد. لاین C13 با ۱۱۰۹۸۵۶ خوانش، کمترین و C40 با ۲۶۷۳۴۱۸ خوانش، بیشترین توالی صحیح تولید نمودند (جدول ۱). با توجه به نرخ همردیفی و کیفیت نقشه، تعداد خوانش‌های صحیح نقشه‌یابی شده به ازای هر ژنوم در توالی‌های مرجع IWGSC Ref Seq v1.0، CSS، W7984 و IWGSC-WGAV0.4 متفاوت، که توالی مرجع IWGSC Ref Seq v1.0 نسبت به بقیه مطلوب‌تر بود (داده‌ها منتشر نشده است). در نقشه‌یابی خوانش‌های صحیح نسبت به ژنوم رفرنس IWGSC RefSeq v1.0 گندم، بالاترین نرخ همردیفی مربوط به ژنوم B و کمترین مربوط به ژنوم D بود. این یافته‌ها نشان می‌دهد که تعداد خوانش‌های صحیح در روش ddRAD-Seq به علت قرائت از دو انتها (paired-end) بیش از دو برابر نسبت به روش GBS گزارش شده توسط Alipour *et al.* (2017) بود. در فیلتر اولیه با عمق توالی مساوی و بیشتر از پنج (DP≥5) و نمره کیفیت مساوی و بیشتر از ۹۹۹ (Quality score=≥999) برای حذف چند آللی و نواحی حذف و اضافه (Indels)، تعداد ۱۰۳۲۷۶۰ نشانگر SNP قرائت شد. با انجام فیلتر دوم برای فراوانی آلل‌های نادر کمتر از پنج درصد (MAF>5%) و میزان هتروزیگوت بالاتر از ۱۰ درصد (Het<10%) برای داده‌های گم‌شده ۲۰ ٪،

۳۰ ٪، ۴۰ ٪ و ۵۰ ٪ انجام شد. که بیشترین SNP قرائت شده برای داده‌های ۵۰ درصد گم‌شده بدست آمد (جدول ۲). تعداد کل SNP های صحیح فراخونی شده برای ۵۰ درصد داده گم‌شده برابر با ۳۳۴۲ عدد، که تعداد ۱۳۲۲ معادل ۳۹/۵۶ درصد روی ژنوم B، ۱۲۵۳ معادل ۳۷/۴۹ درصد روی ژنوم A و ۷۶۷ معادل ۲۲/۹۵ درصد روی ژنوم D شناسائی شد (شکل ۳). این یافته ها با نتایج قبلی مشابه بود (Berkman *et al.*, 2013; ؛ Lai *et al.*, 2015; Alipour *et al.*, 2017). همچنین تعداد نشانگر SNP در ژنوم A و B حدود ۳/۴ برابر ژنوم D بود که با یافته Alipour *et al.* (2017) مطابقت داشت، اما با نتایج تحقیق Cavanagh *et al.* (2013) و Allen *et al.* (2013) که بیش از پنج برابر گزارش نموده‌اند، تفاوت داشت.



شکل ۲- میانگین نمره کیفیت فرد کتابخانه.

Figure2. The average of Pherd's score of the pooling.

جدول ۱- تعداد خوانش‌های قرائت شده

Table 1. The number of reads calling

	Read1	Read2	Total
Total correct sequences	75054339	75054339	150108678
Total wrong sequences	14351584	14351584	28703168
Total	89405923	89405923	178811846

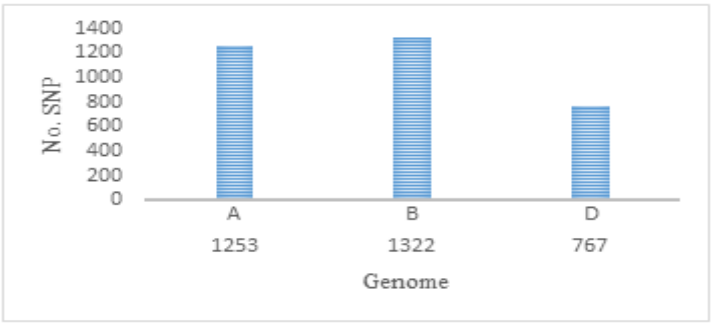
بر اساس تاریخچه تکاملی گندم، ژنوم D پس از سال‌های طولانی به ژنوم گندم نان اضافه شده است و بنابراین میزان موتاسیون، مضاعف‌شدن ژن و چند شکلی در ژنوم‌های قدیمی یعنی A و B که در اولین رویداد هیبریداسیون حدود نیم الی ۳ میلیون سال پیش بین گونه‌های اجدادی دیپلوئید حاصل شده بود، بیشتر است (Alipour *et al.*, 2017). پراکنش توزیع ژنجایSNP ها در روی کرموزوم ژنوم‌ها یکنواخت نبود. بیشترین SNP روی کروموزوم دوم از ژنوم B (2B) و سوم از ژنوم B (3B) و کمترین آن روی کروموزوم چهارم از ژنوم D (4D) شناسائی شد (شکل ۴). این دستاورد با یافته‌های Alipour *et al.* (2017) و Edae *et al.* (2015) مشابه بود. بنابراین منطقی است که با افزایش اندازه کروموزوم و توالی بازهای نوکلئوتیدی روی آن، احتمال موتاسیون و تولید SNP جدید، روی دو ژنوم A و B زیاد شود.

جدول ۲- مقایسه و پراکنش نشانگر SNP تولید شده بر اساس داده‌های گم شده ۲۰، ۳۰، ۴۰ و ۵۰ درصد در سطح ژنوم و کروموزوم

Table 2. The SNP markers Comparison and distribution based on 20, 30, 40, and 50% missing data on genome and chromosome

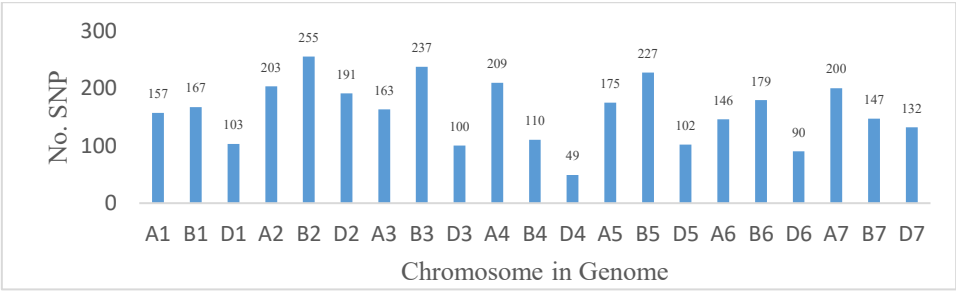
Allele		A1	A2	A3	A4	A5	A6	A7	A Genome	B1	B2	B3	B4	B5	B6	B7	B Genome
Missing data	20%	14	25	18	23	17	11	14	122	19	33	26	11	39	26	21	175
	30%	38	47	33	49	37	31	41	276	43	55	49	22	67	46	35	317
	40%	75	97	73	90	82	70	81	568	100	109	111	46	125	110	69	670
	50%	157	203	163	209	175	146	200	1253	167	255	237	110	227	179	147	1322

Allele		D1	D2	D3	D4	D5	D6	D7	D Genome
Missing data	20%	16	49	19	6	18	8	31	147
	30%	21	64	28	14	30	17	41	215
	40%	51	103	43	22	46	45	82	392
	50%	103	191	100	49	102	90	132	767



شکل ۳- میزان پراکنش SNP در ژنوم های A، B و D در ۵۰ درصد داده گم شده.

Figure3. The SNP markers distribution on A, B and D genomes in 50% missing data.



شکل ۴. میزان پراکنش SNP بر روی هر یک از کروموزوم‌های همیولوگ

Figure 4. The SNP markers distribution on homoeologue chromosomes.

بیشترین چگالی نشانگری روی کروموزوم‌های دوم و پنجم ژنوم B (2B و 5B) و کمترین روی کروموزوم چهارم از ژنوم D (4D) مشاهده شد. همچنین این چگالی در ژنوم B بسیار بیشتر از ژنوم D بود (جدول ۳). برونداد این تحقیق با نتایج Alipour *et al.* (2017) همسان بود. با حذف تاثیر اندازه کروموزوم بر تعداد SNP، در حقیقت علاوه بر اندازه کروموزوم، عوامل دیگری مانند طول دوره تکامل نیز در تعداد SNP روی کروموزوم‌ها دخالت دارند. رابطه خطی بسیار معنی‌دار بین چگالی (تراکم) نشانگری و اندازه کروموزوم در سه

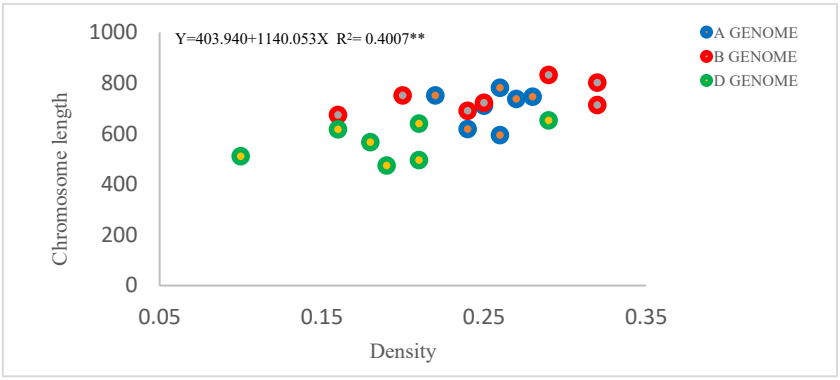
ژنوم مشاهد شد (شکل ۵). این به این معناست که با افزایش اندازه و تراکم نشانگری، تعداد نشانگر SNP افزایش می‌یابد که این مشاهدات با اطلاعات Alipour *et al.* (2017) یکسان بود.

جدول ۳- میزان چگالی نشانگری (SNP/Mbp) روی هر ژنوم و کروموزوم با داده‌های ۵۰ درصد گم شده

Table 3. Marker density (SNP / Mbp) on each genome and chromosome with 50% missing data

Allele	A1	A2	A3	A4	A5	A6	A7	A Genome	B1	B2	B3	B4	B5	B6	B7	B Genome
Chromosome length (Mbp)	594	781	751	745	710	618	737	4935	690	801	831	674	713	721	751	5180
No. SNP	157	203	163	209	175	146	200	1253	167	255	237	110	227	179	147	1322
Density (SNP/Mbp)	0.26	0.26	0.22	0.28	0.25	0.24	0.27	0.25	0.24	0.32	0.29	0.16	0.32	0.25	0.20	0.26

Allele	D1	D2	D3	D4	D5	D6	D7	D Genome
Chromosome length (Mbp)	495	652	616	510	566	474	639	3951
No. SNP	103	191	100	49	102	90	132	767
Density (SNP/Mbp)	0.21	0.29	0.16	0.10	0.18	0.19	0.21	0.19

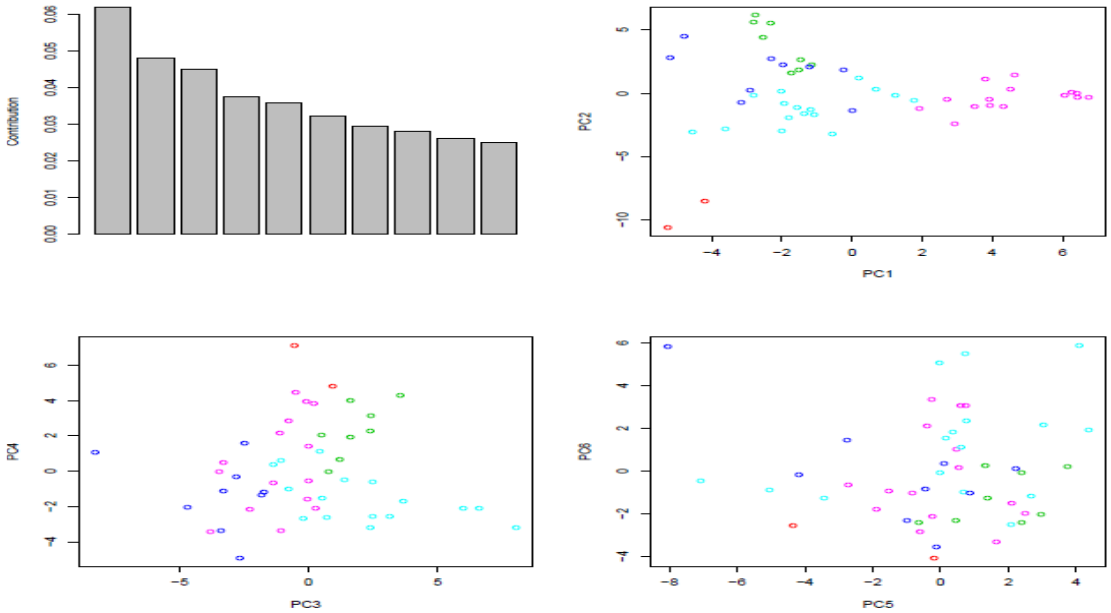


شکل ۵- رابطه خطی بین میزان چگالی (SNP/Mbp) و اندازه کروموزوم هر یک از ژنوم‌های A، B و D

Graph 5. The linear regression between marker density (SNP/Mbp) and chromosome size in each of A, B, and D wheat genomes.

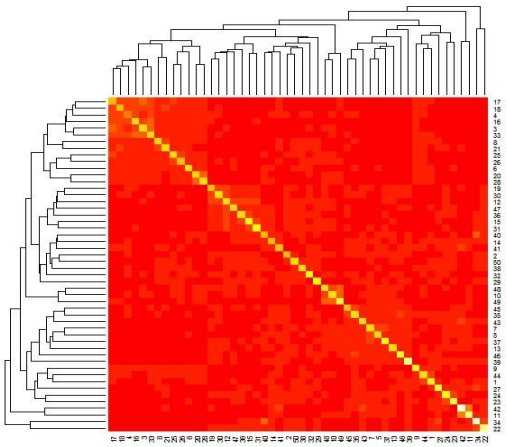
تجزیه به مولفه‌های اصلی نشان داد که بر حسب تعداد و رنگ‌های متفاوت، پنج گروه مختلف بر اساس شش مولفه اول (PC1-PC6) قابل شناسائی است که سهم هر مولفه از تغییرات واریانس کل در اسکری پلات به صورت هیستوگرام نمایان است. بر اساس این یافته، پنج زیر جمعیت به‌صورت SP1 (n=۲)، SP2 (n=۹)، SP3 (n=۸)، SP4 (n=۱۶) و SP5 (n=۱۵) قابل شناسائی بود (شکل۶). این برونداد مطابق با یافته‌های Voss-Fels و همکاران (2015) که از ۴۶۰ ژنوتیپ و آرایه ۹۰K نشانگر SNP برای گروه‌بندی جمعیت استفاده کردند، بود.

علاوه بر آن، نمودار هیت‌مپ همزمان بر اساس ماتریس فاصله ژنتیکی نشانگر SNP و گروه بندی لاین/رقم صورت گرفت که ارتباطی قوی بین ژنوتیپ‌ها برای هر کلاستر جداگانه آشکار کرد، اما تمایز شدید برای گروه-بندی منفرد و متمایز نشان نداد (شکل۷). این دستاورد با یافته Wang *et al.* (2017) مطابقت داشت. گروه بندی بر اساس تجزیه مولفه‌های اصلی و ماتریس تشابه بر اساس نشانگر ژنومیک یعنی SNP، صحت تفکیک زیر جمعیت را تایید می‌کنند، اما همان‌طور که ملاحظه می‌شود، با مقایسه گروه بندی بر اساس تجزیه به مولفه-ها و ماتریس شباهت ژنومی، جمعیت به سه گروه اصلی قابل تفکیک است.



شکل ۶- گروه‌بندی جامعه پیشرفته اصلاحی بر اساس نشانگر SNP.

Figure 6. The advanced breeding population classification based on SNP markers.



شکل ۷- توپولوژی حاصل از ماتریس شباهت بر اساس نشانگر SNP و گروه‌بندی لاین‌ها.

Figure 7. Topology derived from a similarity matrix based on the SNP markers and lines grouping.

نتیجه گیری کلی

متوسط نمره کیفیت فرد برابر با ۳۰ بود که نشان دهنده مناسب بودن استفاده از فن آوری NGS با پلاتفرم Illumina® برای توالی‌یابی گندم است. تعداد خوانش‌های تولید شده در روش ddRAD-Seq به ازای هر لاین برابر با ۲۲۰۷۴۸۰ بود که ۱/۸ برابر روش GBS که توسط Alipour *et al.* (2017) در جمعیت بسیار بزرگ گندم و هشت بار توالی‌یابی بود. بیشترین SNP قرائت شده برای داده‌های ۵۰ درصد گم‌شده به‌دست آمد. تعداد کل SNP های صحیح فراخونی شده برای ۵۰ درصد داده گم‌شده برابر با ۳۳۴۲ عدد SNP بود که بیشترین آن در ژنوم B شناسائی شد. هم میزان پراکنش و هم تراکم SNP به ازای هر کروموزوم در ژنوم متفاوت بود. بین چگالی و اندازه کروموزوم رابطه خطی معنی‌دار مشاهده شد. تجزیه به مولفه‌های اصلی و ماتریس تشابه و گروه‌بندی همزمان با استفاده از اطلاعات نشانگر

SNP، قادر به شناسائی زیر جمعیت‌ها از یک جمعیت اصلی شد، اما ساختار جمعیت بسیار ضعیف بود و بنابراین تاثیری در برآورد ارزش‌های اصلاحی ژنومی (GS) از طریق برازش مدل‌ها و انتخاب بهترین لاین و مطالعات ارتباط نشانگر به صفت (MTA) و استفاده در MAS در برنامه‌های به‌نژادی برای این جمعیت ندارد.

REFERENCES

1. Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G. & Zhang G. (2017). Genotyping-by-Sequencing (GBS) Revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science*, 8, 1-14.
2. Allen, A. M., Barker, G. L., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., Uauy, C., Griffiths, S., Jack, P., Berry, S. & Werner, P. (2013). Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, 11(3), 279-295.
3. Arafa, R. A., Rakha, M. T. Soliman, N. E. K. Moussa, O. M. Kamel, S. M. & Shirasawa, K. (2017). Rapid identification of candidate genes for resistance to tomato late blight disease using next-generation sequencing technologies. *PLoS ONE*, 12(12), 1-15.
4. Berkman, P. J., Visendi, P., Lee, H. C., Stiller, J., Manoli, S., Lorenc, M. T., Lai, K., Batley, J., Fleury, D., Šimková, H. & Kubalakova, M. (2013). Dispersion and domestication shaped the genome of bread wheat. *Plant biotechnology journal*, 11(5), 564-571.
5. Borrill, P., Adamski, N. & Uauy, C. (2015). Genomics as the key to unlocking the polyploidy potential of wheat. *New Phytologist*, 2008, 1008-1022.
6. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
7. Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A. & Bolser, D. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 491, 705– 710.
8. Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A. & See, D. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the national academy of sciences*, 110(20), 8057-8062.
9. Chapman, J. A., Mascher, M., Buluc, A., Barry, K., Georganas, E. and Session, A. (2015). A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*, 16, 26.
10. DaCosta, J. M. & Sorenson, M. D. (2014) Amplification Biases and Consistent Recovery of Loci in a Double-Digest RAD-seq Protocol. *PLoS ONE*, 9(9), 1-14.
11. Edae, E. A., Bowden, R. L. & Poland, J. (2015). Application of Population Sequencing (POPSEQ) for Ordering and Imputing Genotyping-by-Sequencing Markers in Hexaploid Wheat. G3: *Genes, Genomes, Genetics*, *5*(12), 2547-2553.
12. El-Metwally, S., Ouda, M. O., & Helmy, M. (2014). Next Generation Sequencing Technologies and Challenges in Sequence Assembly. Springer Science.
13. International Wheat Genome Sequencing C. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum* L.) *genome. Science*, 345, 6194.
14. International Wheat Genome Sequencing C. (2018). IWGSC *RefSeq assembl*y v1.0. Retrieved from, https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies.
15. Karaaslan, Ç., Akel, H., Ünlü, S. & Perçin, I. (2014). Comparison of Six Commercial DNA Extraction Kits for DNA Extraction from Wheat. *Hacettepe Journal of Biology and Chemistry,* 42 (3), 395-400.
16. Kchouk, M., Gibrat, J. F. & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine* ,(Aligarh), 9(3), 1-8.

17. Krasileva, K., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., Soria, M., Wang, S., Consortium, I. & Akhunov, E. (2013). Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology*, 14(6), R66.
18. Kristensen, P. S., Jahoor, A., Andersen, J. R., Cericola, F., Orabi, J., Janss, L. L. & Jensen, J. (2018). Genome-Wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. Frontiers in Plant Science, 9, 69.
19. Lai, K., Lorenc, M. T., Lee, H. C., Berkman, P. J., Bayer, P. E., Visendi, P., Ruperao, P., Fitzgerald, T. L., Zander, M., Chan, C. K. K. & Manoli, S. (2015). Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Plant biotechnology journal*, *13*(1), 97-104.
20. Liljeroth, E. & Bryngelsson, T. (2002). Earlier onset of DNA fragmentation in leaves of wheat compared to barley and rye. *Hereditas* 136, 108–115.
21. Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. Nature, 470, 198-203.
22. Maxam, A. M. & Gilbert, W. A. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74, 560-564.
23. Melchinger, A.E. (1999). Genetic diversity and heterosis. In: J.T. Gerdes, editor, Genetics and exploitation of heterosis in crops. ASA, CSSA, SSSA, Madison, WI. p. 99–118. doi:10.2134/1999.geneticsandexploitation.c10.
24. Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. & Shinozaki, K. (2009). TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiology*, 150, 1135–1146.
25. Poland, J. A. & Rife, T. W. (2012): Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*, 5, 92–102.
26. R Core Team. 2016. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from, https://www.R-project.org/.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chainterminating inhibitors. *Proceedings of the National Academy of Sciences*,74, 5463–5467.
28. Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26, 135–1145.
29. Shirasawa, K., Hirakawa, H. & Isobe, S. (2016). Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Research*, 23(2), 145-153.
30. VLK, D. & ŘEPKOVÁ, J. (2017). Application of Next-Generation Sequencing in Plant Breeding. Czech J. Genet. *Plant Breeding*, 53(3), 89−96.
31. Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. & Snowdon, R. J. (2015) Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome*, 8(2), 1-13. doi: 10.3835/plant genome2015.03.0013.
32. Wang, S. X., Zhu, Y. L., Zhang, D. X., Shao, H., Liu, P. & Hu, J. B. (2017). Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS ONE*, 12(11), 1-14.

جدول ضمیمه ۱- لاین و ارقام جمعیت اصلاحی.

Appendix 1. Lines and cultivars of the breeding population.

1	Morvarid
2	Gonbad
3	SITTE/MO//PASTOR/3/TILHI/4/WAXWING/KIRITATI
4	REEDLING#1
5	ALTAR 84/AE.SQUARROSA(221)//3*BORL95/3/URES/JUN//KAUZ/4/WBLL1/5/MUTUS
6	NAC/TH.AC//3*PVN/3/MIRLO/BUC/4/2*PASTOR/5/…
7	CHIBIA//PRLII/CM65531/3/SKAUZ/BAV92/4/MUNAL#1
8	KACHU//WBLL*2/BRAMBLING
9	BAJ#1*2/WHEAR
10	PBW343*2/KUKUNA/3/PASTOR//CHIL/PRL/4/GRACK
11	KACHU/BECARD//WBLL1*2/BRAMBLING
12	SUP152*2/TECUE#1
13	WHEAR/KUKUNA/3/C80.1/3*BATAVIA//2*WBLL1/5/…
14	QUAIU*2/KINDE
15	FRNCLN/NIINI #1//FRANCOLIN #1
16	FRNCLN*2/TECUE#1
17	MUTUS*/TECUE#1
18	FRNCOLIN#1/AKURI#1//FRNCLN
19	CHIBIA//PRLII/CM65531/3/SKAUZ/BAV92/4/…
20	KACHU#1//WBLL1*2/KUKUNA
21	CHIBIA//PRLII/CM65531/3/SKAUZ/BAV92*2/4/QUAIU
22	CHIBIA//PRLII/CM65531/3/SW89.5181/KAUZ/4/….
23	KACHU/PVN//KACHU
24	PCAFLR/KINGBIRD#1//KIRITATI/2*TRCH
25	PCAFLR/KINGBIRD #1//KIRITATI/2*TRCH…
26	SUP152*2/TECUE#1
27	SUP152*2/TECUE #1…
28	SITTE/MO//PASTOR/3/TILHI/4/MUNAL#1/5/MUNAL
29	SAAR//INQALAB 91*2/KUKUNA/3/KIRITATI/2*TRCH
30	WHEAR/VIVITSI//WHEAR/3/WHEAR/SOKOLL
31	MILAN/KAUZ//BABAX/3/BAV92/4/WHEAR//2*PRL/2*PASTOR
32	PAURAQ//ND643/2*WBLL1/3/PAURAQUE#1
33	WHEAR/VIVITSI//WHEAR*2/3/KACHU
34	SOKOLL/3/PASTOR//HXL7573/2*BAU/4/PAR.
35	SOKOLL/3/PASTOR//HXL7573/2*BAU/4/SOKOLL/WBLL
36	ND643/2*WBLL1//HEILO
37	SUP152*2/TINKIO#1
38	KACHU*2/3/ND643//2*PRL/2*PASTOR
39	ND643/2*WBLL1/4/CHIBIA//PRLII/CM65531/3/SKAUZ/BAV92/5/BECARO
40	BECARD/3/PASTOR//MUNIA/ALTAR 84
41	PFAU/MILAN//FISCALL/3/VORB/4/MUTUS
42	KIRITATI//ATTILA*2/PASTOR/3/PVN/4/KIRITAI//2*ATTILA*2/PASTOR
43	CHIBIA//PRLII/CM65531/3/SKAUZ/BAV92*2/4/…
44	CHIR3/4/SIREN//ALTAR 84/AE.SQUARROSA(205)/3/3*BUC/5/PFAU/WEAVER/6/VORB
45	PREMIO/BERKUT
46	MILAN/SHA7/3/THB"S"/TON"S"//VEE"S"/6/LUAN/4/V763.23/3/V879.CB//PVN/PICUS/5/OPATA
47	GASPARD//MILAN/SHA7/3/MILAN/SHA7
48	GASPARD//MILAN/SHA7/3/MILAN/SHA7..
49	SW89.3064/STAR//INIA/3/MILAN/SHA7
50	MV17/6/ATRAK/5/4777/FKN/GB/3/VEE"S"/4/BUC"S"…