



Transcriptome Assembly and Identification of EST-SSR Markers in *Crocus sativus*

Mohammadreza Rezaei¹, Hamidreza Sharifi², and Alireza Seifi^{3*}

Submitted: 8 December 2021

Revised: 17 January 2022

Accepted: 10 March 2022

Available Online: 7 March 2022

How to cite this article:

Rezaei, M., Sharifi, H., and Seifi, A. 2022. Transcriptome Assembly and Identification of EST-SSR Markers in *Crocus sativus*. Saffron Agronomy & Technology, 10(1): 41-49.

DOI: 10.22048/jsat.2022.318852.1446

Abstract

Crocus sativus is a triploid plant propagating by vegetative propagation. Therefore, trait segregation and genetic diversity are limited in this plant. EST-SSR markers have some priority, for example, co-dominant inheritance, locus-specific, and highly polymorphic against all other markers. Due to the availability of transcriptome data, it is possible to develop EST-SSR markers and polymorphism studies in saffron. The development of EST-SSR markers in *C. sativus* makes it possible to study genetic diversity and molecular polymorphism in different genotypes. In order to develop a EST-SSR marker for *C. sativus*, we downloaded public available *C. sativus* RNA-seq data. The quality control and preprocessing of raw reads were done using FastQC and Trimmomatic tools. We performed *de novo* transcriptome assembly using RNA-Bloom. CD-HIT-EST was used in order to reduce redundancy in transcriptome assembly. The assembly quality was evaluated using the BUSCO software, and the completeness of transcriptome assembly was 90%. After achieving a high-quality transcriptome assembly of *C. sativus*, EST-SSRs were identified by the MISA tool. The EST-SSRs primers were designed using Primer3. 35459 SSR-containing sequences were detected, and primer pairs were designed for them. Ten EST-SSR primer pairs were randomly selected to amplify *C. sativus* DNA. Seven pairs of the primers (70%) generated clear and reproducible bands with the expected size. These EST-SSR markers can be functional and useful for *C. sativus* genetic studies.

Keywords: EST-SSR, Genetic diversity, Bioinformatics, Transcriptome.

1- MSc in Agricultural Biotechnology, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran

2- Associate Professor, Crop and Horticultural Science Research Department, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran

3- Assistant Professor, Department of Biotechnology and Plant Breeding, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran

(*- Corresponding author Email: arseifi@um.ac.ir)



مقاله پژوهشی

سرهم‌بندی ترنسکریپتوم و شناسایی نشانگرهای EST-SSR در زعفران

محمدرضا رضائی^۱، حمیدرضا شریفی^۲ و علیرضا سیفی^{۳*}

تاریخ دریافت: ۱۷ آذر ۱۴۰۰

تاریخ بازنگری: ۲۷ دی ۱۴۰۰

تاریخ پذیرش: ۱۹ اسفند ۱۴۰۰

رضائی، م.ر.، شریفی، ح.ر.، و سیفی، ع. ۱۴۰۱. سرهم‌بندی ترنسکریپتوم و شناسایی نشانگرهای EST-SSR در زعفران. زراعت و فناوری زعفران، ۱۰(۱): ۴۹-۴۱.

چکیده

گیاه *Crocus sativus*، گیاهی ترپیلوئید است که به روش رویشی با استفاده از بنه تکثیر می‌شود. به دلیل تولید مثل غیرجنسی، تفرق صفات و تنوع ژنتیکی در این گیاه با محدودیت مواجه است. نشانگرهای EST-SSR مزایایی از جمله هم‌غالیت، اختصاصیت برای لوکوس خاص و پلی‌مورفیسم بالا نسبت به نشانگرهای دیگر دارند. با توجه به دردسترس بودن داده‌های ترنسکریپتوم، امکان شناسایی نشانگرهای EST-SSR برای مطالعات پلی‌مورفیسم در گیاه زعفران وجود دارد. جهت توسعه نشانگرهای EST-SSR، داده‌های RNA-Seq گیاه زعفران از NCBI دریافت شدند. سپس کنترل کیفیت و پیرایش داده‌ها به ترتیب توسط ابزارهای FastQC و Trimmomatic انجام گرفت. با استفاده از این داده‌ها و ابزار RNA-Bloom، سرهم‌بندی ترنسکریپتوم انجام گرفت. ابزار CD-HIT-EST برای حذف ترنسکریپت‌های مشابه و تکراری استفاده شد. کیفیت ترنسکریپتوم با استفاده از BUSCO مورد ارزیابی قرار گرفت و درصد ترنسکریپت‌های کامل، حدود ۹۰ درصد به دست آمد. پس از دستیابی به ترنسکریپتوم با کیفیت در زعفران، از نرم‌افزار MISA برای شناسایی EST-SSR ها در ترنسکریپتوم استفاده شد. طراحی آغازگر با استفاده از نرم‌افزار Primer3 برای توالی‌های EST-SSR، انجام شد. تعداد ۳۵۴۵۹ توالی SSR شناسایی شدند و آغازگر برای آن‌ها طراحی گردید. از تعداد ۱۰ جفت آغازگر که برای تکثیر PCR روی DNA زعفران انتخاب شدند، ۷ جفت آغازگر در اندازه پیش‌بینی شده تکثیر شدند که نشان از کارایی ۷۰ درصدی نشانگر دارد. نشانگرهای EST-SSR شناسایی شده در این پژوهش در مطالعات ژنتیکی زعفران می‌توانند مورد استفاده قرار گیرند.

کلمات کلیدی: بیوانفورماتیک، تنوع ژنتیکی، ترنسکریپتوم، EST-SSR.

۱- کارشناس ارشد بیوتکنولوژی کشاورزی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی خراسان رضوی، سازمان تحقیقات، آموزش و ترویج کشاورزی، مشهد، ایران.
۲- دانشیار، بخش تحقیقات علوم زراعی و باغی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی خراسان رضوی، سازمان تحقیقات، آموزش و ترویج کشاورزی، مشهد، ایران.
۳- استادیار، گروه بیوتکنولوژی و به‌نژادی گیاهی، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران.

(*) نویسنده مسئول: arseifi@um.ac.ir

مقدمه

گیاه *Crocus sativus*، گیاهی نر عقیم است که به روش رویشی با استفاده از بنه تکثیر می‌شود. به دلیل تولید مثل غیرجنسی، تفرق صفات در این گیاه با محدودیت مواجه است (Nemati et al., 2019). برخی مطالعات جهت دستیابی به تنوع ژنتیکی در این گیاه انجام گرفته است، اما تنوع ژنتیکی مشاهده نشده است. در مطالعاتی مانند کاربرد نشانگر IRAP^۱ (Alsayied et al., 2015) و نشانگرهای ISSR^۲ و RAPD^۳ پلی‌مورفیسمی در میان زعفران‌های مناطق مختلف مشاهده نشد. در نتیجه این مطالعات پیشنهاد کردند که زعفران یک گیاه مونومورفیک است و زعفران‌های مناطق مختلف، کلون‌های یکسان هستند (Rubio-Moraga et al., 2009). برخی از پژوهش‌ها مانند کاربرد نشانگر SSR^۴ (Nemati et al., 2012)، نشانگر AFLP^۵ (Siracusa et al., 2013) و نشانگر SRAP^۶ (Babaei et al., 2014) سطوح پایینی از تغییرات ژنتیکی را در این گیاه نشان دادند. با توجه به این که تفاوت‌های فنوتیپی در سطح مزرعه در میان زعفران‌های مناطق مختلف مشاهده می‌شود، پیشنهاد شده است که بخش عمده‌ای از این تفاوت‌ها به دلیل تفاوت‌های اپی‌ژنتیکی و بخش دیگر به دلیل تفاوت‌های اندک ژنتیکی می‌باشد (Busconi et al., 2021). در نتیجه، مطالعات برای یافتن تفاوت‌های ژنتیکی در زعفران اهمیت ویژه‌ای دارد.

نشانگرهای EST-SSRs^۷ مزایایی از جمله هم‌غالبیت، اختصاصیت برای لوکوس خاص و پلی‌مورفیسم بالا نسبت به

نشانگرهای دیگر دارند که آن‌ها را برای مطالعات تنوع ژنتیکی ایده‌آل ساخته است. شناسایی نشانگرهای SSR برای مدیریت مؤثر ژرم‌پلاسم‌های این گیاه با ارزش که اطلاعات ژنتیکی کمی از آن در دسترس است، مورد نیاز است (Nemati et al., 2012). امکان شناسایی تعداد زیادی نشانگر SRR از طریق توالی‌یابی ترنسکرپتوم و سرهم‌بندی *de novo* توالی‌ها وجود دارد. از مزایای این روش جدید می‌توان به صرفه‌جویی در هزینه و استفاده مستقیم از توالی ژن‌های بیان شده اشاره کرد که مطالعات ژنتیکی وابسته به صفات خاص را ممکن می‌کند (Chen et al., 2017). در سال‌های اخیر شناسایی نشانگرهای EST-SSR بر اساس داده‌های توالی‌یابی ترنسکرپتوم در گیاهان مختلفی از جمله *Phyllostachys violascens* (Cai et al., 2019)، *Elymus sibiricus* (Zhang et al., 2019)، *Populus wulianensis* (Wu et al., 2020)، *Sophora japonica* (Lu et al., 2020) و *Stephanandra incisa* (Zhang et al., 2021) انجام شده است و کاربرد موفق این روش برای ارزیابی تنوع ژنتیکی و تجزیه و تحلیل‌های ژنتیکی در این مطالعات نشان داده شده است.

با توجه به اینکه ژنوم زعفران در دسترس نیست، سرهم‌بندی با کیفیت از ترنسکرپتوم و شناسایی نشانگر بر اساس آن از اولویت‌های پژوهشی در مطالعات ژنتیکی زعفران می‌باشد. هدف از این مطالعه، شناسایی نشانگرهای EST-SSR برای زعفران بر اساس سرهم‌بندی ترنسکرپتوم می‌باشد. با استفاده از این نشانگرها می‌توان زعفران‌های ایرانی و خارجی را به صورت مولکولی از هم تفکیک کرد و مطالعات تنوع ژنتیکی را در این گیاه با ارزش از لحاظ اقتصادی انجام داد.

- 1-Inter-retrotransposon-amplified polymorphism
- 2-Inter-simple sequence repeat
- 3-Random amplified polymorphic DNA
- 4-Simple sequence repeat
- 5-Amplified fragment length polymorphism
- 6-Sequence-related amplified polymorphism
- 7-Expressed sequence tag-derived simple sequence repeat

مواد و روش‌ها

جمع‌آوری و کنترل کیفیت داده‌ها

داده‌های RNA-Seq^۱ زعفران از پایگاه داده NCBI دریافت شدند. این داده‌ها شامل خوانش‌هایی از مطالعات مختلف زعفران با کدهای SRR8790835، SRR1140761، SRR8790836، SRR1767299، SRR1767298، SRR1767300 و SRR1767302 می‌باشند که به ترتیب مربوط به کلاله قرمز، جوانه انتهایی در مرحله رویشی، جوانه انتهایی در مرحله گلدهی، بنه، برگ، کاسبرگ، پرچم و کلاله گیاه زعفران هستند. کنترل کیفیت داده‌ها با استفاده از نرم‌افزار FastQC انجام گرفت و توالی‌های با کیفیت پایین، توالی‌های آداپتر و درصد هر نوکلئوتید در توالی توسط این ابزار مشخص شد (Andrews, 2010).

پیرایش، نرمال‌سازی و سرهم‌بندی داده‌ها

پس از تشخیص داده‌های با کیفیت کم توسط FastQC، پیرایش داده‌ها توسط نرم‌افزار Trimmomatic-0.39 انجام گرفت (Bolger et al., 2014). نرمال‌سازی داده‌های با استفاده از ابزار Bignorm انجام شد (Wedemeyer et al., 2017). سرهم‌بندی ترنسکرپتوم توسط ابزار RNA-Bloom انجام شد (Bushmanova et al., 2019).

حذف توالی‌های تکراری و بررسی کیفیت سرهم‌بندی

پس از سرهم‌بندی، برای اینکه ترنسکرپت‌های مشابه و تکراری از سرهم‌بندی حذف شوند، از ابزار CD-HIT-EST v4.8.1 استفاده شد (Fu et al., 2012). بررسی کیفیت ترنسکرپتوم توسط ابزار TrinityStats انجام شد (Haas et al., 2013). سپس بررسی کامل بودن ترنسکرپتوم توسط ابزار BUSCO v4.0.6 بر اساس ژن‌های مجموعه داده

Embryophyta انجام شد (Seppey et al., 2019).

شناسایی EST-SSR ها در ترنسکرپتوم و طراحی آغازگر شناسایی EST-SSR ها در ترنسکرپتوم با استفاده از ابزار MISA انجام شد (Beier et al., 2017). تنظیمات این ابزار برای این آزمایش به صورت توالی‌های ۲ نوکلئوتیدی با حداقل شش تکرار و توالی‌های ۳ تا ۶ نوکلئوتیدی با حداقل پنج تکرار انجام شد. طراحی آغازگر با استفاده از ابزار Primer3 به صورت خط فرمان در لینوکس انجام گرفت (Untergasser et al., 2012). این روش یک روش طراحی آغازگر با خروجی بالا می‌باشد و برای همه توالی‌های SSR شناسایی شده در مرحله قبل، آغازگر طراحی می‌شود. تنظیمات Primer3 به این صورت انجام گرفت که اندازه محصول PCR^۲ بین ۱۰۰ تا ۵۰۰ جفت باز، اندازه آغازگر بین ۱۸ تا ۲۳ جفت باز با اندازه مطلوب ۲۰ جفت باز، حداکثر اختلاف دمای بین جفت آغازگرها ۵ درجه سانتی‌گراد و درصد GC آغازگرها بین ۴۰ تا ۶۰ درصد تنظیم شدند. ده جفت آغازگر EST-SSR برای تکثیر محصول PCR روی DNA زعفران انتخاب شدند (جدول ۱).

استخراج DNA و PCR

استخراج DNA به روش CTAB^۳ (Doyle, 1991) از بنه‌های زعفران جمع‌آوری شده از تربت حیدریه، کشت شده در ایستگاه تحقیقاتی طرق متعلق به مرکز تحقیقات، آموزش و ترویج کشاورزی خراسان رضوی، انجام گرفت. PCR بر اساس چرخه دمایی جدول ۲ به صورت Touch down انجام گرفت. در این نوع PCR، پس از واسرشت سازی آغازین، پنج سیکل با کاهش یک درجه دمای اتصال به ازای هر سیکل وجود دارد، به طوری که دمای ۶۰ درجه سانتی‌گراد در سیکل اول به ۵۵ درجه سانتی‌گراد در سیکل پنجم می‌رسد. سپس ۳۰ سیکل با

2- Polymerase chain reaction

3 - Cetyl trimethylammonium bromide

1- RNA sequencing

دمای ثابت ۵۵ درجه سانتی‌گراد وجود دارد. واکنش PCR در حجم ۲۰ میکرولیتر انجام گرفت.

جدول ۱- آغازگرهای طراحی شده برای EST-SSR انتخاب شده در این مطالعه

Table 1- EST-SSR primers used in this study

آغازگر Primer	توالی آغازگر Primer sequence (5'-3')	دمای اتصال T _m	اندازه محصول PCR (جفت‌باز) PCR Product size (bp)	نوع SSR SSR type
CsSSR1	F: CTTTCATCACCTCGCTCTGC R: TGACAGTGACATCCCTACGG	59.35 59.35	449	(AAT)8
CsSSR2	F: TGAATTCTTGCACCATCGCC R: TCCCTCATTTCCGTCTCCTC	57.30 59.35	470	(GA)15
CsSSR3	F: CTCCGCAGTTTCATTCCAGG R: CCACCGAACCTACCAAAGT	59.35 59.35	418	(AAAGA)6
CsSSR4	F: CTCTTCGTCGCCACCAAAG R: CACTCGGCGGATACATACCA	59.35 59.35	486	(TTC)5
CsSSR5	F: GGTTGAGGCTTGAGAGATGC R: AGTCTCCTGGTATTACATGGT	59.35 58.39	498	(AT)6
CsSSR6	F: CGGGTTTCATCGTGCAAAGT R: GAGCCGTTGGAGATGTTAGC	57.30 59.35	423	(TA)6
CsSSR7	F: GTGTACGCTGATGTGGATGC R: CCTCTTAGCTTGTCCCTTGC	59.35 59.35	500	(TCT)16
CsSSR8	F: TGGCCGAAATGAGTTGATCC R: ATTCGTTGCCTCACTTCAGC	57.30 57.30	437	(CA)8
CsSSR9	F: GGCCACCATTCGAATCACTG R: ACAAAACCATGGACCAGGAT	59.35 57.30	449	(ACTC)6
CsSSR10	F: GGATGGTTGGTTATCTACAGCA R: ACTGGACACACATTGCGAGA	58.39 57.30	484	(TACA)6

جدول ۲- چرخه دمایی واکنش PCR

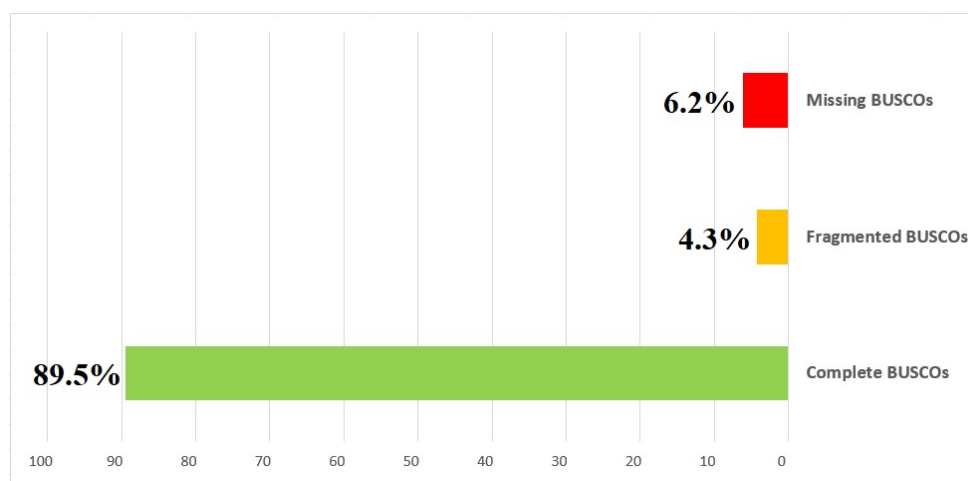
Table 2- Thermal cycling for PCR

مرحله Step	دما Temperature (C)	زمان Time (s)	تعداد سیکل Number of cycles
واسرشت آغازین Initial denaturation	94	240	1
واسرشت Denaturation	94	30	5
اتصال Annealing	60 to 55	30	5
سنتز رشته Extension	72	60	5
واسرشت Denaturation	94	30	30
اتصال Annealing	55	30	30
سنتز رشته Extension	72	60	30
سنتز نهایی Final extension	72	300	1

نتایج و بحث

نتایج سرهم‌بندی با کیفیت و جامع از ترنسکرپتوم زعفران با استفاده از داده‌های RNA-seq بافت‌های مختلف از مطالعات مختلف و توسط ابزار RNA-Bloom انجام گرفت. پس از حذف ترنسکرپت‌های تکراری با استفاده از ابزار CD-HIT-EST در نهایت ۳۳۷۰۱۱ ترنسکرپت نهایی باقی ماندند. با استفاده از

TrinityStats مشخص شد که درصد GC برابر با ۴۳/۸۰ درصد و شاخص N50 برابر با ۱۳۲۳ است که نشان از کیفیت سرهم‌بندی ترنسکرپتوم دارد. درصد ترنسکرپت‌های کامل بر اساس تجزیه و تحلیل نرم افزار BUSCO، ۸۹/۵ درصد به دست آمد (شکل ۱).



شکل ۱- بررسی کیفیت ترنسکرپتوم زعفران با استفاده از ابزار BUSCO
Figure 1- Evaluation of assembly quality using the BUSCO software.

جفت آغازگر برای تکثیر PCR روی DNA زعفران انتخاب شدند. از این تعداد، محصول PCR برای ۷ جفت آغازگر در اندازه پیش‌بینی شده تکثیر شد (شکل ۲). نتایج، نشان‌دهنده کارایی ۷۰ درصدی در تکثیر نشانگر دارد.

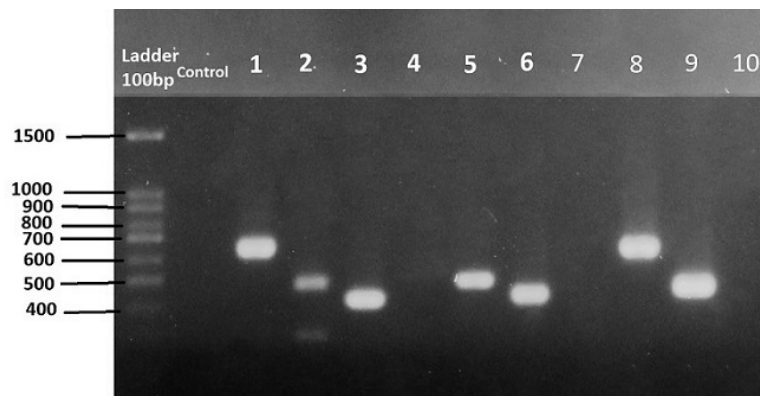
بررسی سرهم‌بندی ترنسکرپتوم با استفاده از ابزارهای TrinityStats و BUSCO نشان می‌دهد که سرهم‌بندی ترنسکرپتوم از کیفیت بیشتری نسبت به سرهم‌بندی بدست آمده در مطالعه قبلی (Yue et al., 2020)، برخوردار است که شامل افزایش ۱۷ درصدی در درصد ترنسکرپتوم‌های کامل نسبت به مطالعه یاد شده می‌شود. در این تحقیق، N50 برابر با ۱۳۲۳، درصد GC برابر با ۴۳/۸۰ و ترنسکرپت‌های کامل حدود

بر اساس آمار و اطلاعاتی که از نتایج شناسایی SSRها توسط نرم‌افزار MISA به دست آمد، تعداد کل SSRهای شناسایی شده ۴۴۶۶۹ و تعداد توالی‌هایی که دارای SSR هستند ۳۵۴۵۹ به دست آمد. بر این اساس، تعداد توالی‌هایی که حاوی بیش از یک SSR بودند، ۷۱۲۹ به دست آمد. از کل تعداد SSRهای شناسایی شده، SSRهای دو جفت‌باز تعداد ۲۳۱۱۶، سه جفت‌باز ۱۸۸۲۶، چهار جفت‌باز ۱۳۰۰، پنج جفت‌باز ۵۴۱ و شش جفت‌باز ۸۸۶ به دست آمدند (جدول ۳). با استفاده از ابزار Primer3 برای ۲۴۵۲۲ توالی، آغازگر با موفقیت طراحی شد و برای ۱۵۵۱۱ توالی، آغازگر مناسبی طراحی نشد. جهت تایید فرایند شناسایی نشانگر EST-SSR، تعداد ۱۰

۹۰ درصد به دست آمد. در جدیدترین مطالعه که بر روی ترنسکریپتوم زعفران صورت گرفته است نیز N50 برابر با ۱۵۳۲ (al., 2020). جفت باز و درصد GC برابر با ۴۳/۵ به دست آمده است (Hu et

جدول ۳- آمار EST-SSR در ترنسکریپتوم سرهم‌بندی شده زعفران
Table 3- EST-SSR statistics in the saffron transcriptome

تعداد	شاخص
Number	Item
337011	تعداد کل توالی‌های ترنسکریپتوم Total number of sequences examined
44669	تعداد کل SSR های شناسایی شده Total number of identified SSRs
35459	تعداد کل توالی‌های دارای SSR Number of SSR-containing sequences
7129	تعداد کل توالی‌های دارای بیش از یک SSR Number of sequences containing more than 1 SSR
4665	تعداد کل SSR های ترکیبی Number of SSRs present in compound form
23116	تکرارهای دو نوکلئوتیدی Dinucleotide
18826	تکرارهای سه نوکلئوتیدی Trinucleotide
1300	تکرارهای چهار نوکلئوتیدی Tetranucleotide
541	تکرارهای پنج نوکلئوتیدی Pentanucleotide
886	تکرارهای شش نوکلئوتیدی Hexanucleotide



شکل ۲- تکثیر آغازگرهای EST-SSR روی ژنوم زعفران. از سمت چپ، چاهک اول نشانگر اندازه 100bp پارس‌توس، چاهک دوم، کنترل و چاهک های سوم تا آخر به ترتیب از CsSSR1 تا CsSSR10 می‌باشند. برای CsSSR4، CsSSR7، CsSSR10 و CsSSR10 باند مشاهده نشد.

Figure 2- EST-SSR primer amplification. Parstoos 100bp DNA size marker and control at the left followed by CsSSR1 to CsSSR10 PCR products. No PCR product was amplified for CsSSR4, CsSSR7 and CsSSR10.

با استفاده از این ترنسکریپتوم کامل و باکیفیت، تعداد ۳۵۴۵۹ توالی دارای SSR شناسایی شد. این تعداد توالی SSR شناسایی شده در جدیدترین مطالعه SSR در زعفران (۱۶۰۵۷)

است و می‌توان از آن برای برای شناسایی نشانگرهای EST-SSR استفاده کرد. با استفاده از این ترنسکرپتوم، SSRهای زعفران شناسایی شدند و آغازگرها با موفقیت برای این توالی‌ها طراحی شدند. تعداد ۷ جفت آغازگر از ۱۰ جفت آغازگر انتخابی، روی ژنوم زعفران تأیید شدند که نشان از کارایی بالای طراحی نشانگر دارد. این تعداد زیاد توالی SSR شناسایی شده در مطالعات مختلف زعفران که ژنوم آن در دسترس نیست، بسیار کاربردی می‌باشد. شناسایی این این نشانگرها به تفکیک مولکولی زعفران‌های ایرانی و خارجی و مطالعات تنوع ژنتیکی در گیاه زعفران کمک خواهد کرد.

توالی (SSR) (Jain et al., 2016) است. بنابراین از نظر تعداد SSR، در این آزمایش دو برابر بیشتر توالی SSR نسبت به آخرین مطالعه زعفران به دست آمد. سپس برای ۲۴۵۲۲ توالی SSR، آغازگر با موفقیت طراحی شد. این تعداد زیاد آغازگر و توالی SSR شناسایی شده، به مطالعات مختلف ژنتیکی در گیاه زعفران کمک می‌کند.

نتیجه‌گیری

کنترل کیفیت سرهم‌بندی انجام شده در این تحقیق نشان می‌دهد که سرهم‌بندی ترنسکرپتوم از کیفیت مطلوبی برخوردار

منابع

- Alsayied, N.F., Fernández, J.A., Schwarzacher, T., and Heslop-Harrison, J.S. 2015. Diversity and relationships of *Crocus sativus* and its relatives analysed by inter-retroelement amplified polymorphism (IRAP). *Annals of Botany* 116: 359-368.
- Andrews, S. 2010. Fast QC: a quality control tool for high throughput sequence data. Available at Web site: www.bioinformatics.babraham.ac.uk/projects/fastqc (verified 5 September 2020).
- Babaei, S., Talebi, M., Bahar, M., and Zeinali, H. 2014. Analysis of genetic diversity among saffron (*Crocus sativus*) accessions from different regions of Iran as revealed by SRAP markers. *Scientia Horticulturae* 171: 27-31.
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33: 2583-2585.
- Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
- Busconi, M., Wischnitzki, E., Del Corvo, M., Colli, L., Soffritti, G., Stagnati, L., Fluch, S., Sehr, E.M., de los Mozos Pascual, M., and Fernández, J.A. 2021. Epigenetic variability among saffron crocus (*Crocus sativus* L.) accessions characterized by different phenotypes. *Frontiers in Plant Science* 12: 642631.
- Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8: 1-13.
- Cai, K., Zhu, L., Zhang, K., Li, L., Zhao, Z., Zeng, W., and Lin, X. 2019. Development and characterization of EST-SSR markers From RNA-Seq data in *phyllostachys violascens*. *Frontiers in Plant Science* 10: 50.
- Chen, J., Li, R., Xia, Y., Bai, G., Guo, P., Wang, Z., Zhang, H., and Siddique, K.H.M. 2017. Development of EST-SSR markers in flowering Chinese cabbage (*Brassica campestris* L. ssp. *chinensis* var. *utilis* Tsen et Lee) based on de novo transcriptomic

- assemblies. PLOS ONE 12: e0184736.
- Doyle, J. 1991. DNA Protocols for Plants. In: Hewitt, G.M., Johnston, A.W.B., Young, J.P.W. (Eds.), Molecular Techniques in Taxonomy. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 283-293.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150-3152.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., and Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494-1512.
- Hu, J., Liu, Y., Tang, X., Rao, H., Ren, C., Chen, J., Wu, Q., Jiang, Y., Geng, F., and Pei, J. 2020. Transcriptome profiling of the flowering transition in saffron (*Crocus sativus* L.). Scientific Reports 10: 9680.
- Jain, M., Srivastava, P.L., Verma, M., Ghangal, R., and Garg, R. 2016. De novo transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. Scientific Reports 6: 22456.
- Lu, Y., Zhang, C., Li, X., Liang, Y., Wang, Y., and Li, W. 2020. Development of EST-SSR markers and their application in the analysis of the genetic diversity of *Sophora japonica* Linn. Trees 34: 1147-1156.
- Nemati, Z., Harpke, D., Gemicioglu, A., Kerndorff, H., and Blattner, F.R. 2019. Saffron (*Crocus sativus*) is an autotriploid that evolved in Attica (Greece) from wild *Crocus cartwrightianus*. Molecular Phylogenetics and Evolution 136: 14-20.
- Nemati, Z., Zeinalabedini, M., Mardi, M., Pirseyediand, S.M., Marashi, S.H., and Khayam Nekoui, S.M. 2012. Isolation and characterization of a first set of polymorphic microsatellite markers in saffron, *Crocus sativus* (Iridaceae). American Journal of Botany 99: e340-e343.
- Rubio-Moraga, A., Castillo-López, R., Gómez-Gómez, L., and Ahrazem, O. 2009. Saffron is a monomorphic species as revealed by RAPD, ISSR and microsatellite analyses. BMC Research Notes 2: 189.
- Seppey, M., Manni, M., and Zdobnov, E.M. 2019. BUSCO: Assessing genome assembly and annotation completeness. In Kollmar M (Ed.), Gene Prediction: Methods and Protocols. Springer New York, New York, NY, pp. 227-245.
- Siracusa, L., Gresta, F., Avola, G., Albertini, E., Raggi, L., Marconi, G., Lombardo, G.M., and Ruberto, G. 2013. Agronomic, chemical and genetic variability of saffron (*Crocus sativus* L.) of different origin by LC-UV-vis-DAD and AFLP analyses. Genetic Resources and Crop Evolution 60: 711-721.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research 40: e115-e115.
- Wedemeyer, A., Kliemann, L., Srivastav, A., Schielke, C., Reusch, T.B., and Rosenstiel, P. 2017. An improved filtering algorithm for big read datasets and its application to single-cell assembly. BMC Bioinformatics 18: 324.

- Wu, Q., Zang, F., Xie, X., Ma, Y., Zheng, Y., and Zang, D. 2020. Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*. *Scientific Reports* 10: 16249.
- Yue, J., Wang, R., Ma, X., Liu, J., Lu, X., Balaso Thakar, S., An, N., Liu, J., Xia, E., and Liu, Y. 2020. Full-length transcriptome sequencing provides insights into the evolution of apocarotenoid biosynthesis in *Crocus sativus*. *Computational and Structural Biotechnology Journal* 18: 774-783.
- Zhang, C., Wu, Z., Jiang, X., Li, W., Lu, Y., and Wang, K. 2021. De novo transcriptomic analysis and identification of EST-SSR markers in *Stephanandra incisa*. *Scientific Reports* 11: 1059.
- Zhang, Z., Xie, W., Zhao, Y., Zhang, J., Wang, N., Ntakirutimana, F., Yan, J., and Wang, Y. 2019. EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC Plant Biology* 19: 235.

Transcriptome Assembly and Identification of EST-SSR Markers in *Crocus sativus*

Mohammadreza Rezaei¹, Hamidreza Sharifi², and Alireza Seifi^{3*}

Submitted: 8 December 2021

Accepted: 10 March 2022

Rezaei, M., Sharifi, H., and Seifi, A. 2022. Transcriptome Assembly and Identification of EST-SSR Markers in *Crocus sativus*. Saffron Agronomy & Technology, 10(1): 41-49.

Abstract

Crocus sativus is a triploid plant propagating by vegetative propagation. Therefore, trait segregation and genetic diversity are limited in this plant. EST-SSR markers have some priority, for example, co-dominant inheritance, locus-specific, and highly polymorphic against all other markers. Due to the availability of transcriptome data, it is possible to develop EST-SSR markers and polymorphism studies in saffron. The development of EST-SSR markers in *C. sativus* makes it possible to study genetic diversity and molecular polymorphism in different genotypes. In order to develop a EST-SSR marker for *C. sativus*, we downloaded public available *C. sativus* RNA-seq data. The quality control and preprocessing of raw reads were done using FastQC and Trimmomatic tools. We performed *de novo* transcriptome assembly using RNA-Bloom. CD-HIT-EST was used in order to reduce redundancy in transcriptome assembly. The assembly quality was evaluated using the BUSCO software, and the completeness of transcriptome assembly was 90%. After achieving a high-quality transcriptome assembly of *C. sativus*, EST-SSRs were identified by the MISA tool. The EST-SSRs primers were designed using Primer3. 35459 SSR-containing sequences were detected, and primer pairs were designed for them. Ten EST-SSR primer pairs were randomly selected to amplify *C. sativus* DNA. Seven pairs of the primers (70%) generated clear and reproducible bands with the expected size. These EST-SSR markers can be functional and useful for *C. sativus* genetic studies.

Keywords: EST-SSR, Genetic diversity, Bioinformatics, Transcriptome.

1- MSc in Agricultural Biotechnology, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran

2- Associate Professor, Crop and Horticultural Science Research Department, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran

3- Assistant Professor, Department of Biotechnology and Plant Breeding, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran

(*- Corresponding author Email: arseifi@um.ac.ir)

DOI: 10.22048/jsat.2022.318852.1446