ISSN: 2476-6917; Modares Journal of Biotechnology. 2020;11(2):185-191

Development of SSR Markers Associated with Biosynthesis Pathway of Steviol Glycosides in Stevia through De Novo Transcriptome Assembly

ARTICLE INFO

Article Type
Original Research

Authors
Khayam Nekoui M.*[1] *PhD*
Moazam Jazi M.[2] *PhD,*
Mardi M.[3] *PhD,*
Kadkhodaei S.[4] *PhD*

How to cite this article
Khayam Nekoui M, Moazam Jazi M, Mardi M, Kadkhodaei S. Development of SSR Markers Associated with Biosynthesis Pathway of Steviol Glycosides in Stevia through De Novo Transcriptome Assembly. Modares Journal of Biotechnology. 2020;11(2):185-191.

ABSTRACT

In stevia (*Stevia rebaudiana*), breeding programs are mainly aimed at developing plants with high Rebaudioside-A (RA) content. To this end, in order to screen stevia plants and selection of varieties with the highest amount of desired sweeteners (RA) using molecular markers, the present study was conducted on RNA-seq data of varieties having different amounts of RA. In order to de novo transcriptome assembly for each variety CLC was used considering k-mer and contig length values of 20 and at least 200bp, respectively. The assembly was annotated using the latest Arabidopsis proteome release. To identify signatures of candidate polymorphic SSRs among the stevia varieties, the assembled sequences were used as an input for CandiSSR, followed by designing primer pairs for identified polymorphic SSRs. 368 potential polymorphic SSRs were identified based on the stevia transcriptome analysis, among which 360 were qualified for primer design. Almost 89% of the contig sequences possessing polymorphic SSRs had the best blast hit against Arabidopsis proteome. In this study, contigs similar to the UDP-Glycosyltransferase protein family and Deoxyxylulose-5-phosphate synthase which are involved in the biosynthesis pathway of steviol glycosides were found. Also, gene set enrichment analysis using PlantGSE through hypergeometric test (FDR<0.05) identified enriched metabolic pathways in the sequences contained polymorphic SSRs. It is, therefore, most likely that such connections exist between the SSRs and biosynthesis of steviol glycosides. Hence, it could conceivably be hypothesized that the SSR markers developed in this study would be reliable in molecular breeding of stevia toward selection of varieties with high content particularly RA.

Keywords Stevia; De Novo Transcriptome Assembly; Marker Development; Steviol Glycoside Biosynthesis Pathway

[1]Bioscience Faculty, Tarbiat Modares University, Tehran, Iran
[2]Institute for Endocrine Science, Shahid Beheshti Medical University, Tehran, Iran
[3]Seed & Plant Certification & Registration Institute, Karaj, Iran
[4]Research Institute for Biotechnology & Bioengineering, Isfahan University of Technology, Isfahan, Iran

Correspondence
Address: Tarbiat Modares University, Nasr Bridge, Jalal-Al-Ahmad Highway, Tehran, Iran. Postal Code: 1411713116
Phone: +98 (21) 66976546
Fax: -
mojtabakhayam@gmail.com

Article History
Received: August 31, 2019
Accepted: October 30, 2019
ePublished: June 10, 2020

CITATION LINKS

[1] A review on the improvement of stevia [Stevia rebaudiana (Bertoni)] [2] Steviol glycosides targeted analysis in leaves of Stevia rebaudiana (Bertoni) from plants cultivated under chilling stress conditions [3] The leaves of Stevia rebaudiana (Bertoni), their constituents and the analyses thereof: A review [4] Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences [5] Genome-wide genetic marker discovery and genotyping using next-generation sequencing [6] Development and validation of genic-SSR markers in sesame by RNA-seq [7] De novo transcriptome sequencing of Camellia sasanqua and the analysis of major candidate genes related to floral traits [8] RNA-Seq for gene identification and transcript profiling of three Stevia rebaudiana genotypes [9] CandiSSR: An efficient pipeline used for identifying candidate polymorphic SSRs based on multiple assembled sequences [10] Critical review on steviol glycosides: Pharmacological, toxicological and therapeutic aspects of high potency zero caloric sweetener [11] Microsatellite markers: What they mean and why they are so useful [12] Potential of RAPD and ISSR markers for assessing genetic diversity among Stevia rebaudiana Bertoni accessions [13] Automated generation of heuristics for biological sequence comparison [14] Genome-wide microsatellite characterization and marker development in the sequenced Brassica crop species [15] Comparative transcriptomics unravel biochemical specialization of leaf tissues of Stevia for diterpenoid production [16] Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids

Copyright© 2020, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.

ردیابی و معرفی نشانگرهای ریزماهواره مرتبط با مسیر بیوسنتز استویول‌گلیکوزیدها در گیاه استویا از طریق بازآرایی ترنسکریپتوم از نو

مجتبی خیام‌نکویی* PhD
دانشکده علوم زیستی، دانشگاه تربیت مدرس، تهران، ایران
مریم معظم‌جزی PhD
پژوهشکده غدد درون ریز و متابولیسم، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
محسن مردی PhD
موسسه تحقیقات ثبت و گواهی بذر و نهال، کرج، ایران
سعید کدخدایی PhD
پژوهشکده زیست فناوری و مهندسی زیستی، دانشگاه صنعتی اصفهان، اصفهان، ایران

چکیده
هدف عمده برنامه‌های به‌نژادی استویا (*Stevia rebaudiana*) ایجاد گیاهانی با میزان بالای ریبودیوزید- آ (RA) است. در این راستا، به‌منظور غربالگری گیاهان استویا و انتخاب واریته‌هایی با بیشترین میزان شیرین‌کننده‌های مورد نظر با استفاده از نشانگرهای مولکولی، مطالعه حاضر روی داده‌های RNA-seq واریته‌های دارای مقادیر مختلف RA انجام شد. به‌منظور بازآرایی ترنسکریپتوم از نو برای هر واریته، از نرم‌افزار CLC با درنظرگرفتن طول k-mer برابر با ۲۰ و حداقل طول کانتیگ برابر با ۲۰۰ جفت باز استفاده شد. برای انجام تفسیر، آخرین نسخه از پروتئوم گیاه مدل آرابیدوپسیس به کار برده شد. برای شناسایی SSRهای چندشکل کاندید در میان واریته‌های استویا، با استفاده از CandiSSR آنالیز توالی‌های بازآرایی و به دنبال آن طراحی جفت آغازگرهای مربوطه انجام شد. حدود ۳۶۸ نشانگر SSR بالقوه شناسایی شدند که در این میان ۳۶۰ نشانگر شرایط لازم برای طراحی آغازگر را داشتند. تقریبا ۸۹% از کانتیگ‌های واجد SSRهای چندشکل دارای بهترین توالی مشابه در برابر پروتئوم آرابیدوپسیس بودند. در این مطالعه، کانتیگ‌های مشابه با خانواده پروتئینی UDP- گلیکوزیل‌ترانسفراز و داوکسی‌دی‌گزیلولوز۵- فسفات که در مسیر بیوسنتز استویول‌گلیکوزیدها دخیل هستند شناسایی شدند. همچنین آنالیز مجموعه ژنی با استفاده از PlantGSE از طریق آزمون هایپرژئومتریک در سطح معنی‌داری ۵% (FDR<0.05) چندین مسیر متابولیکی مرتبط با توالی‌های حاوی SSRهای چندشکل را مورد شناسایی قرار داد. بنابراین می‌توان این فرضیه را مطرح کرد که نشانگرهای SSR چندشکل توسعه‌یافته در این مطالعه با اطمینان در برنامه‌های به‌نژادی مولکولی استویا به‌منظور انتخاب واریته‌های با میزان بالای SG به‌ویژه RA قابل استفاده هستند.
کلیدواژه‌ها: استویا، بازآرایی از نو رونوشت، توسعه نشانگر، مسیر بیوسنتز استویول‌گلیکوزیدها

تاریخ دریافت: ۱۳۹۸/۶/۹
تاریخ پذیرش: ۱۳۹۸/۸/۸
*نویسنده مسئول: mojtabakhayam@gmail.com

مقدمه
استویا (*Stevia rebaudiana*) گیاهی علفی، چندساله با عدد کروموزومی ۲n=۲۲ و اندازه ژنوم (هاپلوئید) برابر با ۱/۳گیگا جفت باز از تیره آستراسه است. گیاهان خودناسازگار هستند و گرده‌افشانی آنها به‌صورت دگرگرده‌افشان است که این امر باعث بروز تنوع بالا در جمعیت گیاهان حاصل از بذر می‌شود[1]. استویا قادر به تولید گروهی از شیرین‌کننده‌ها موسوم به استویول‌گلیکوزید (SG) است که حدود ۳۰۰ بار شیرین‌تر از ساکاروز و به‌علت عدم جذب قند مربوطه برای استفاده بیماران دیابتی بسیار مناسب است. با توجه به ژنوتیپ گیاه و شرایط رشد و برداشت آن، SG حدود ۲۰-۱۰% وزن خشک گیاه را تشکیل می‌دهد. استویول‌گلیکوزیدهای این گیاه مخلوطی از دی‌ترپنوئید گلوکوزیدهای مختلف است که در میان آنها استویوزید (ST) و ریبودیوزید- آ (RA) از همه فراوان‌تر هستند. RA به واسطه داشتن چهار مولکول گلوکز متصل به حلقه استویول شیرین‌تر از استویوزید (با سه مولکول گلوکز) است. از این رو هدف بسیاری از برنامه‌های اصلاحی ایجاد گیاهان استویای واجد میزان بالای RA است[2, 3].

یکی از راهکارهای موثر برای دست‌یابی به واریته‌های مطلوب در گیاهانی با هتروزیگوسیتی بالا مانند استویا، استفاده از چندشکلی و تنوع موجود برای انتخاب واریته‌هایی است که صفت مورد نظر را دارند. برای این منظور، واریته‌های مختلف با استفاده از انواع مختلف نشانگر از جمله نشانگرهای مورفولوژیکی، فیزیولوژیکی، بیوشیمیایی و مولکولی غربالگری و واریته‌های مطلوب گزینش می‌شوند. در میان نشانگرهای مختلف، استفاده از نشانگرهای مولکولی مانند SSR (Simple sequence repeat) و SNP (Single nucleotide polymorphism) به‌سبب دقت بالا، مقرون‌به‌صرفه‌بودن و امکان غربالگری در مراحل اولیه رشد گیاه نسبت به سایر نشانگرها ارجحیت دارند[4]. امروزه با ظهور نسل جدید توالی‌یابی (NGS) تحولی عظیم در حوزه نشانگرهای مولکولی رخ داده است. توالی‌یابی مجدد ژنوم، توالی‌یابی کتابخانه ژنومی کاهش‌یافته و توالی‌یابی کل RNA (RNA-seq) از جمله تکنیک‌های رایج برای شناسایی نشانگرهای مولکولی در مقیاس وسیع هستند[5, 6]. در این میان، تکنیک RNA-seq که قادر به شناسایی نشانگرهای عملکردی است به‌ویژه برای موجودات غیرمدل که توالی ژنومی آنها در دسترس نیست، روش مناسبی برای مطالعه نشانگرها در مقیاس وسیع است[7]. در مطالعه حاضر نیز می‌توان با استفاده از این روش به نشانگرهای چندشکل برای غربالگری و شناسایی گیاهان استویا از نظر میزان قند و به‌ویژه واریته‌های واجد میزان RA بالاتر دست یافت.

مواد و روش‌ها
داده‌کاوی
به‌منظور غربالگری گیاهان استویا و انتخاب واریته (هایی) با بیشترین میزان قند مورد نظر (RA)، در ابتدا به واریته‌هایی که واجد حداکثر و حداقل میزان RA هستند نیاز است. داده‌های مورد نیاز برای این مطالعه از مطالعه چن و همکاران استخراج شد. در مطالعه مذکور سه واریته از استویا که واجد مقادیر متفاوتی از ST و RA بودند انتخاب و با انجام توالی‌یابی RNA کل، ترانسکریپتوم این گیاه و ژن‌هایی با تفاوت بیان در میان این سه واریته بررسی شد (جدول ۱) [8].

جدول ۱) واریته‌های استویا مورد استفاده در مطالعه *چن* و همکاران

نمونه	(%)ST*	(%)RA**
SR-1	2.19	6.91
SR-2	12.87	0.02
SR-3	1.23	9.35

*درصد کل استویول‌گلیکوزیدها، **درصد ریبودیوزید- آ

در این مطالعه، RNA کل از برگ‌های سه واریته مختلف استویا در بلافاصله قبل از مرحله گل‌دهی که میزان SG حداکثر است، استخراج و با استفاده از تکنولوژی Illumina Hiseq 2000 به‌صورت دوطرفه توالی‌یابی و خوانش‌هایی به طول ۱۰۰ جفت باز ایجاد شده است. روش استخراج RNA، تهیه کتابخانه برای آنالیز ترنسکریپتوم و نحوه توالی‌یابی به تفصیل در تحقیق *چن* و همکاران ذکر شده است[8]. خوانش‌های خام این مطالعه با شماره دسترسی SRR1576548 از پایگاه داده NCBI استخراج شد.

بررسی کیفیت و پیرایش داده‌ها

داده‌های حاصل از توالی‌یابی معروف به خوانش (read)، به فرمت FastQ هستند. از آنجایی که توالی‌یابی به‌صورت دوطرفه انجام شد دو سری داده برای هر نمونه وجود دارد. برای بررسی کیفیت خوانش‌ها از نرم‌افزار FastQC 0.11.4 استفاده شد. این نرم‌افزار یک برنامه رایگان به زبان جاوا و روی سیستم عامل‌های لینوکس و ویندوز قابل اجرا است. بازهایی با کیفیت پایین، بازهای ناشناخته (N) و توالی آداپتورها، با استفاده از برنامه BBMAP 37 قابل اجرا روی سیستم عامل لینوکس حذف شد. خوانش‌هایی با نمره بیشتر از ۳۰ (Q≥30) به‌عنوان خوانش‌های مطلوب و با کیفیت بالا در نظر گرفته و خوانش‌هایی با نمره کمتر از ۳۰ حذف شدند.

بازآرایی از نو ترانسکریپتوم و هم‌ردیفی

برای انجام اسمبلی از نو ترانسکریپتوم ابتدا برنامه‌های Trinity و CLC Genomics Workbench با استفاده از پارامترهای پیش‌فرض اجرا شد. با توجه به کارآیی مشابه هر دو برنامه و مدت‌زمان کمتر برای انجام بازآرایی توسط نرم‌افزار CLC Genomics Workbench، از این برنامه برای بازآرایی از نو ترانسکریپتوم هر واریته استویا به‌طور جداگانه استفاده شد. طول k-mer برابر با ۲۰ و حداقل طول کانتیگ برابر با ۲۰۰ جفت باز در نظر گرفته شد. همچنین خوانش‌ها به کمک ترانسکریپتوم‌های اسمبل‌شده با استفاده از نرم‌افزار CLC Genomics Workbench با درنظرگرفتن mismatch/insertion/deletion costs برابر با ۲/۳/۳ و length fraction/similarity برابر با ۰/۹ و ۰/۷۵ هم‌ردیف شدند و درصد خوانش‌های هم‌ردیف‌شده ثبت شد.

خوشه‌بندی توالی‌های مشابه

به‌منظور اجتناب از توالی‌های تکراری یا مشابه در ترانسکریپتوم‌های بازآرایی‌شده، توالی‌های (کانتیگ) تکراری با حداقل ۹۵% یکسانی با استفاده از نرم‌افزار CD-HIT-EST 4.6.1 شناسایی و خوشه‌بندی شدند. الگوریتم خوشه‌بندی این نرم‌افزار که به‌صورت الگوریتم خوشه‌بندی افزایشی گریدی است، تنها توالی‌های بدون اینترون را مورد استفاده قرار می‌دهد. به‌طور خلاصه، توالی‌ها ابتدا براساس کاهش طول دسته‌بندی می‌شوند. طولانی‌ترین دنباله نماینده خوشه اول در نظر گرفته می‌شود. سپس هر توالی باقیمانده با نماینده‌های خوشه‌های موجود مقایسه می‌شود. اگر شباهت با هر نماینده بالاتر از یک آستانه مشخص باشد، در آن خوشه گروه‌بندی می‌شود. و در غیر این صورت، یک خوشه جدید با آن دنباله به‌عنوان نماینده تعریف می‌شود.

شناسایی نشانگرهای SSR چندشکل و ژن‌های حاوی این نشانگرها

برای شناسایی نشانگرهای چندشکل در میان واریته‌های استویا از برنامه CandiSSR استفاده شد[9]. این برنامه تنها روی سیستم عامل لینوکس قابل اجرا است و اجرای آن به برنامه‌های MISA، NCBI Blastall، ClustalW، Bioperl و Primer3 نیاز دارد. این نرم‌افزار به کمک برنامه MISA، SSRهای موجود در ترانسکریپتوم نمونه‌ای که به‌عنوان نمونه مرجع تعیین شده است را شناسایی می‌کند. سپس ۱۰۰ جفت باز از نواحی بالادست و پایین‌دست SSRهای شناخته‌شده را تعیین می‌کند و این نواحی را با نمونه‌های دیگر با استفاده از برنامه Blast هم‌ردیف می‌کند. پس از حذف توالی‌های کم‌کیفیت هم‌ردیف‌شده، در توالی‌های باقی‌مانده SSRها و میزان چندشکلی آنها را تعیین می‌کند و سپس برای نشانگرهای SSR چندشکل به کمک برنامه Primer3، آغازگرهایی طراحی می‌شود.

تفسیر

به‌منظور انجام تفسیر، آخرین نسخه از پروتئوم گیاه *Arabidopsis thaliana* از پایگاه داده TAIR به فرمت FASTA دانلود شد. با استفاده از برنامه ncbi-blast-2.2.30+، ابتدا پایگاه داده پروتئینی از توالی‌های دانلودشده ساخته شد. سپس توالی ترانسکریپت‌های اسمبل‌شده نمونه مرجع (در اینجا نمونه SR3 به‌عنوان نمونه مرجع در نظر گرفته شد) در برابر پایگاه پروتئینی ساخته‌شده با تعیین آستانه e-value برابر با 1e-5 هم‌ردیف شدند (Blastx) و خروجی بلاست به فرمت txt دریافت شد.

همچنین آنالیز مجموعه ژن‌ها (gene set enrichment) برای توالی‌های (کانتیگ) واجد نشانگرهای SSRهای چندشکل با استفاده از ابزار مبتنی بر وب PlantGSEA (http://structuralbiology.cau.edu.cn/PlantGSEA) انجام شد. به‌علت حجم بسیار بالا و مدت زمان طولانی لازم برای انجام تفسیر به کمک پایگاه داده NR یا UniProt، بلاست و تفسیر به کمک پروتئوم گیاه مدل آرابیدوپسیس انجام گرفت.

نتایج و بحث

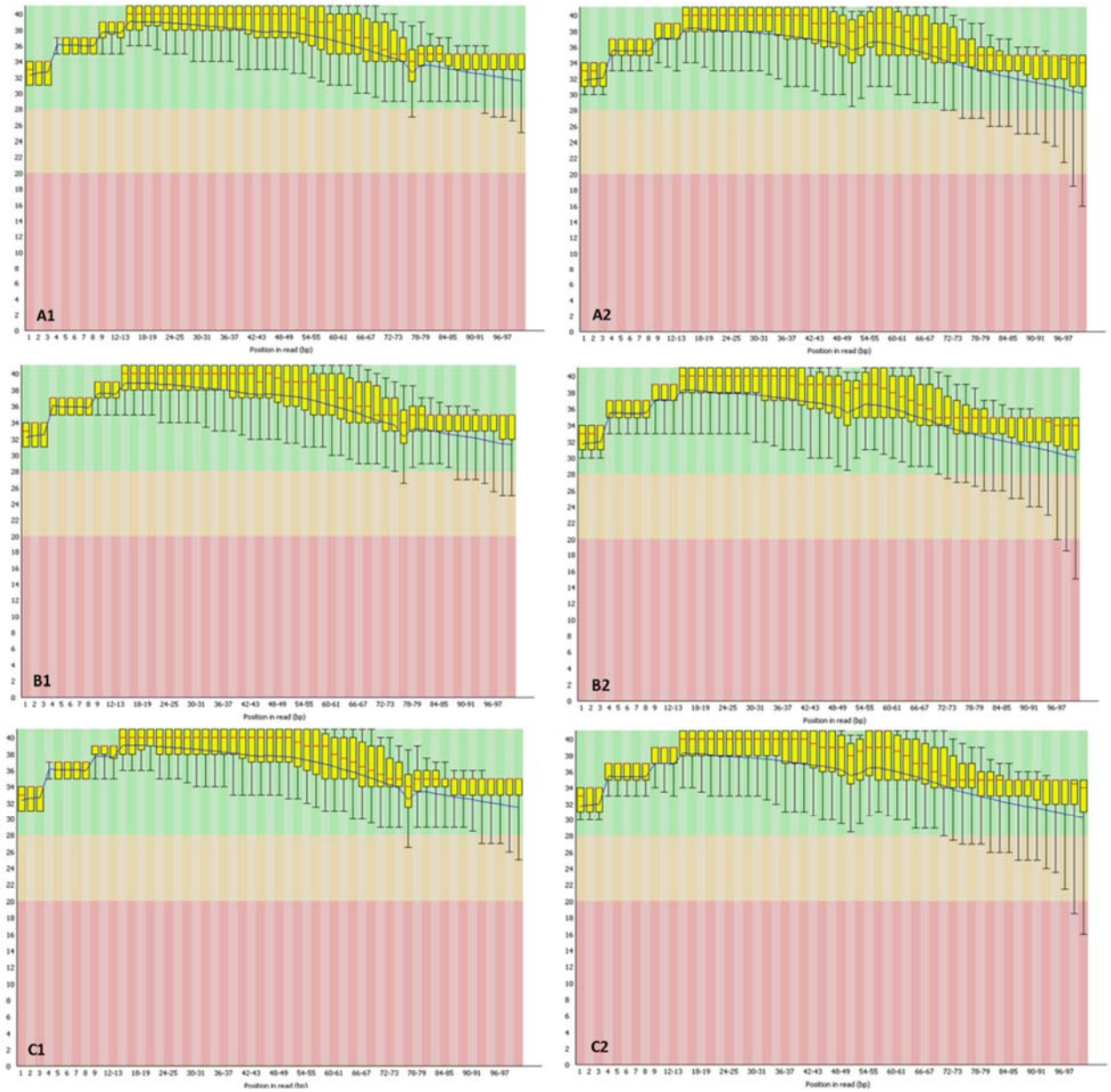
فناوری توالی‌یابی با بازدهی بالا، امکان انجام تجزیه و تحلیل‌های مختلف بیوانفورماتیکی از جمله کشف نشانگرها در ارگانیزم‌های غیرمدل مانند *S. rebaudiana* را فراهم کرده است.

در مطالعه حاضر تعداد خوانش‌های اولیه مورد استفاده در تجزیه و تحلیل‌های بعدی به ترتیب برای واریته‌های SR1، SR2 و SR3 برابر با ۶۱۷۱۰۱۹۴، ۶۸۶۵۲۶۱۴ و ۶۱۲۲۷۴۷۴ بود (جدول ۲). کیفیت

خوانش‌های خام حاصل از توالی‌یابی با استفاده از نرم‌افزار FastQC بررسی شد. شکل ۱، کیفیت بازها را در طول خوانش در موقعیت‌های مختلف نشان می‌دهد. برای ارتقای نتایج آنالیزها از بازهای با کیفیت بالا استفاده و بازهای با کیفیت پایین طی فرآیند پیرایش حذف شدند. به این منظور با استفاده از برنامه BBDuck در پکیج BBMAP 37 اقدام به حذف بازهای با کیفیت پایین (Phred= ۲۰ score cutoff)، بازهای ناشناخته (N) و توالی آداپتورها شد. نهایتاً داده‌های پیرایش‌شده با میانگین امتیاز کیفیت ۳۸ به‌منظور بازآرایی ترنسکریپتوم مورد استفاده قرار گرفتند. میزان خوانش‌ها قبل و بعد از پیرایش در جدول ۲ نشان داده شده است (شکل ۱).

جدول ۲) میزان خوانش‌ها قبل و بعد از پیرایش برای ۳ نمونه استویا

نام نمونه	تعداد اولیه خوانش‌ها	تعداد خوانش‌ها پس از پیرایش	میزان خوانش‌های حذف‌شده طی پیرایش (%)
SR1	۶۱۷۱۰۱۹۴	۶۱۰۰۱۱۶۸	۱/۱۵
SR2	۶۸۶۵۲۶۱۴	۶۷۹۳۰۳۸۱	۱/۰۵
SR3	۶۱۲۲۷۴۷۴	۶۰۰۷۵۶۳۱	۱/۹



شکل ۱) کیفیت خوانش‌های خام در سطح باز؛ (A1 و A2): کتابخانه نمونه SR1؛ (B1 و B2): کتابخانه نمونه SR2؛ (C1 و C2): کتابخانه نمونه SR3

خوانش‌های پیرایش‌شده هر نمونه با استفاده از برنامه CLC Genomics Workbench به‌طور جداگانه بازآرایی شد و در پایان بازآرایی، خوانش‌های اولیه روی ترانسکریپتوم اسمبل‌شده نقشه‌یابی شدند. درصد بالای خوانش‌های نقشه‌یابی‌شده و همچنین N50 بالا حاکی از کیفیت مطلوب بازآرایی انجام‌شده است (جدول ۳). علاوه بر این، درصد خوانش‌های مکان‌یابی‌شده در رونوشت‌های بازآرایی‌شده (RMBT) به‌عنوان معیاری برای ارزیابی کامل‌بودن بازآرایی در نظر گرفته می‌شود که نشان‌دهنده میزان دربرگرفتن خوانش‌ها در ایجاد رونوشت است. همان‌گونه که در جدول ۳ نشان داده شده است، میزان RMBT برای نمونه‌های مختلف حدود ۹۸% است که حاکی از کیفیت بالای بازآرایی انجام‌شده برای آنالیزهای بعدی است. به‌منظور کاهش توالی‌های اضافی و حذف موارد یکسان، خوشه‌بندی رونوشت‌های اضافی با استفاده از CD-HIT 4.6.1 و شاخص ۹۵% انجام شد.

جدول ۳) پارامترهای مرتبط با کیفیت اسمبلی ۳ نمونه استویا

نام نمونه	تعداد کانتیگ	حداکثر طول کانتیگ	سایز N50	درصد خوانش‌های نقشه‌یابی‌شده
SR1	۶۷۵۸۳	۱۵۳۸۴	۱۱۳۷	۸۶
SR2	۷۷۳۰۳	۱۵۵۷۵	۱۰۰۲	۸۶
SR3	۶۰۳۱۴	۱۴۳۸۶	۱۱۰۱	۸۶

در ادامه و پس از حذف توالی‌های مشابه با یکسانی ۹۵% با استفاده از برنامه CD-HIT-EST، از نرم‌افزار CandiSSR برای جست‌وجوی نشانگرهای SSR چندشکل بین ترانسکریپتوم‌های اسمبل‌شده واریته‌های استویا استفاده شد. برای اجرای این برنامه ابتدا باید نمونه‌ای به‌عنوان نمونه مرجع برای نرم‌افزار تعریف شود. در اینجا نمونه SR3 که واجد بیشترین میزان RA است به‌عنوان نمونه مرجع در نظر گرفته شد. این برنامه ۳۶۸ نشانگر SSR را به‌عنوان SSRهای چندشکل بین سه واریته استویا معرفی کرد که در این میان ۳۶۰ نشانگر شرایط لازم برای طراحی پرایمر را دارا بودند (پیوست ۱). پرایمرها توسط برنامه Primer3 که بخشی از نرم‌افزار CandiSSR است برای نشانگرهای چندشکل پیش‌بینی‌شده طراحی شد. برنامه به‌طور پیش‌فرض سه جفت پرایمر برای هر نشانگر طراحی کرد. برای تایید تجربی نشانگرهای چندشکل، پس از انتخاب نشانگرهای مورد نظر بهتر است جفت پرایمرهای مربوط به هر نشانگر با استفاده از برنامه الیگو بررسی شوند و بهترین جفت پرایمر انتخاب شود (پیوست ۱).

۸۹% توالی‌های (کانتیگ) واجد SSRهای چندشکل در برابر پروتئوم آرابیدوپسیس هم‌ردیف شدند و دارای بهترین توالی مشابه با خود بودند (پیوست ۲). در میان توالی‌های هم‌ردیف‌شده، توالی (کانتیگ) مشابه با خانواده پروتئینی UDP- گلیکوزیل‌ترانسفراز و توالی مشابه با داوکسی‌گزیلولوزسنتاز که در مسیر بیوسنتز استویول‌گلیکوزیدها نقش دارند، وجود داشت. علاوه بر این، برای شناسایی مسیرهای زیستی غنی‌شده در توالی‌های واجد SSRهای چندشکل، آنالیز مجموعه ژن‌ها با استفاده از ابزار مبتنی بر وب PlantGSE و با انتخاب آزمون هایپرژئومتریک (Hyper Geometric) در سطح معنی‌داری ۵% (FDR<0.05) انجام شد. براساس این آنالیز، مسیرهای متابولیکی به‌طور معنی‌دار توالی‌های واجد SSRهای چندشکل غنی شده بودند (پیوست ۳) که این امر حاکی از نقش این کانتیگ‌ها در مسیرهای بیولوژیکی فعال است.

ادغام رونوشت بازآرایی‌شده و تفسیر کارکردی یکی از جنبه‌های اصلی داده‌کاوی ترنسکریپتوم در یک ارگانیزم غیرمدل است. در مطالعه حاضر، بررسی تشابه توالی‌های رونوشت‌های بازآرایی‌شده، با استفاده از BLAST و با آستانه e-value معادل 1e-5 در مقابل دو دیتابیس رایج، یعنی NR و TAIR10 (متشکل از توالی‌های دارای تفسیر مناسب گیاه مدل *آرابیدوپسیس تالیانا*) انجام شد. براساس نتایج تفسیر، ۵۳ و ۵۱% کانتیگ‌ها به ترتیب در دو دیتابیس مذکور دارای بهترین هیت با پروتئین‌های همسان شناخته‌شده بودند که این مطابق با نتایج مطالعات قبلی انجام‌شده روی گونه‌های غیرمدل است[8].

ارزش تجاری قابل توجه استویا به‌دلیل شیرین‌کننده‌های طبیعی موجود در آن است که در برنامه‌های به‌زراعی و به‌نژادی بایستی مورد توجه قرار گیرند. در به‌نژادی سنتی از معیارهای فنوتیپی و بیوشیمیایی برای انتخاب واریته‌های خاصی از استویا با میزان بالای SG به‌ویژه ریبودیوزید- آ (RebA) استفاده می‌شود که این معیارها به میزان زیادی تحت تاثیر شرایط محیطی هستند. همچنین این معیارها معمولاً در مراحل پایانی رشد در گیاهان بالغ قابل ارزیابی هستند که در مقیاس وسیع زمان‌بر و پرهزینه خواهد بود[10]. نشانگرهای مولکولی، از جمله توالی‌های ساده تکرارشونده (SSR) به‌دلیل میزان بالای چندشکلی، هم‌بارزبودن و نیز امکان تشخیص در مراحل اولیه رشد و نمو گیاه، ابزار ارزشمندی برای اهداف به‌نژادی گیاهی هستند[11]. توسعه چنین نشانگرهایی و متعاقباً شناسایی جایگاه‌های نشانگرهای مرتبط با صفات مورد نظر، مسیرهای جدیدی را برای متخصصین به‌نژادی گیاهی فراهم کرده است. تنوع ژنتیکی در میان واریته‌های مختلف استویا با استفاده از نشانگرهای مولکولی از قبیل چندشکلی حاصل از تکثیر تصادفی قطعات DNA (RAPD) و توالی‌های بینابینی ساده تکرارشونده (ISSR) مورد بررسی قرار گرفته است[12]. تجزیه و تحلیل مشابهی روی نشانگرهای SSR حاصل از EST انجام شد که به‌وضوح میزان بالایی از چندشکلی را بین تمامی ارقام مورد مطالعه نشان داد[5]. با این حال، در این رابطه مطالعه جامعی با استفاده از داده‌های توالی‌یابی پربازده و با تمرکز بر ایجاد تمایز بین واریته‌های دارای مقادیر مختلف SG به‌ویژه ST و RebA موجود نیست. در حال حاضر توالی‌یابی RNA (RNA-seq) برای تشخیص تعداد زیادی از واریانت‌های ژنتیکی مانند SNPها و SSRها، به‌ویژه برای گونه‌های گیاهی غیرمدل که در آنها عدم وجود اطلاعات ژنتیکی مانع از شناسایی نشانگرهای مولکولی می‌شود، مفید است[7]. در مورد استویا تنها یک مطالعه توسط چن و همکاران برای بررسی ترنسکریپتوم این گیاه با استفاده

از RNA-seq انجام شده و ۱۰/۰۷۰ مورد SSR در میان ۸۰/۱۶۰ یونی‌ژن (unigene) گزارش شده است. در مطالعه حاضر، تلاش شد تا برای اولین‌بار SSRهای چندشکل کاندید بین سه واریته استویا با میزان متفاوت ST و RebA با استفاده از پایپ لاین CandiSSR مورد شناسایی قرار گیرند. در این پایپ لاین، توسط ابزار MISA (MIcroSAtellite identification tool) به ترتیب مراحلی شامل شناسایی توالی‌های تکراری در رونوشت مرجع، هم‌ردیف‌کردن توالی‌های دو طرف SSRهای شناسایی‌شده با استفاده از BLAST در مقابل رونوشت‌های غیرمرجع، جست‌وجوی SSRهای اختصاصی مرجع در توالی غیرمرجع، فیلترکردن PolySSRهای با کیفیت پایین، محاسبه تشابه توالی نواحی دو طرف PolySSRهای شناسایی‌شده و نهایتاً طراحی جفت آغازگرهای هر کدام از PolySSRهای با استفاده از Primer3 انجام می‌شود. میزان دقت این نرم‌افزار تا بیش از ۹۰% گزارش شده است[9]. در این مطالعه، رونوشت بازآرایی‌شده حاصل از واریته SR-3 با بالاترین درصد RebA (۹/۳۵%) به‌عنوان مرجع انتخاب شد. براساس نتایج به‌دست‌آمده، ۳۶۸ نشانگر SSR چندشکل در میان سه رقم استویا شناسایی شد که نشان‌دهنده وجود تنوع ژنتیکی در میان سه واریته مورد بررسی بود و می‌تواند برای تشخیص انواع مختلف استویا با محتوای SG متفاوت مورد استفاده قرار گیرد. چندشکلی شامل موتیف‌های دونوکلئوتیدی با حداقل شش تکرار و سایر موتیف‌ها با حداقل پنج تکرار در رونوشت مرجع است. حداقل و حداکثر تعداد تکرارهای چندشکل به ترتیب ۱ و ۲۹ بود. موتیف‌های AT، TA، TGA و GAT با دربرگرفتن بیش از ۵% کل تکرارهای چندشکل دارای بیشترین فراوانی بودند (پیوست ۴). همچنین با استفاده از Primer3 جفت آغازگرهایی برای ۳۶۰ PolySSR با لحاظ‌کردن میزان کافی از توالی‌های دو طرف با موفقیت طراحی شد (پیوست ۱). به‌منظور حذف جفت آغازگرهایی که چند آمپلیکون تولید می‌کنند، شبیه‌سازی PCR برای تمامی آغازگرها از طریق ماژول IPCRESS در نرم‌افزار Exonerate انجام شد[13] که ۳۵۷ جفت آغازگر تک‌آمپلیکون ایجاد کردند که این نشان‌دهنده اختصاصیت نشانگرهای SSR متناظر است (پیوست ۵). نشان داده شده است که نشانگرهای SSR که در شبیه‌سازی PCR به‌صورت in silico محصول تک‌باند تولید می‌کنند، به‌طور بالقوه نشانگرهای تک‌جایگاهی هستند و می‌توانند مفید واقع شوند[14].

علاوه بر این در مطالعه حاضر اقدام به تفسیر عملکردی کانتیگ‌های حاوی PolySSR شد تا همولوژی آنها با پروتئین‌های شناخته‌شده مشخص شود و عملکردهای احتمالی این توالی‌ها از طریق مقایسه آنها با دیتابیس توالی‌های پروتئینی NR تعیین شود. جالب توجه اینکه، ۸۹% کانتیگ‌های حاوی SSR تشابه معنی‌داری (e-value<1e-5) را در مقایسه با دیتابیس NR نشان دادند (پیوست ۲). در میان توالی‌ها، کانتیگ‌هایی مشابه با ژن‌های کلیدی موجود در مسیر بیوسنتز SGs، از جمله داوکسی‌گزیلولوز۵- فسفات‌سنتاز و UDP- گلیکوزیل‌ترانسفراز وجود داشتند. داوکسی‌گزیلولوز۵- فسفات‌سنتاز در اولین مرحله از مسیر MEP و تبدیل پیروات و دی‌گلیسرآلدهید۳- فسفات به ۱-داوکسی‌دی‌گزیلولوز۵- فسفات نقش دارد. گلیکوزیلاسیون استویول دی‌ترپنوئید توسط UDP- گلیکوزیل‌ترانسفراز آخرین مرحله از بیوسنتز SG است که در آن بقایای قند از قندهای فعال‌شده به یک گیرنده آگلیکون منتقل می‌شود تا انواع SGها تولید شوند[15, 16].

یافته‌های این مطالعه، PolySSRهای عملکردی شناسایی‌شده را به‌عنوان ابزاری مفید برای تسریع مطالعات ژنتیکی و بهبود برنامه‌های به‌نژادی در استویا معرفی می‌کند. علاوه بر این، آنالیز GO با استفاده از ابزار agriGO به‌منظور انتصاب GO termهای متناظر به کانتیگ‌های حاوی PolySSRها انجام شد. برای این آنالیز، رونوشت بازآرایی و تفسیرشده به‌عنوان مرجع در نظر گرفته شد و آزمون هایپرژئومتریک (FDR<0.05) برای انجام آنالیز آماری انتخاب شد. براساس نتایج حاصل، در مجموع ۸۳ مورد GO term به کانتیگ‌های حاوی PolySSR اختصاص یافت (پیوست ۳). آنالیز مذکور مشخص کرد که کانتیگ‌های مورد نظر به‌طور معنی‌داری با چندین مسیر متابولیکی در ارتباط هستند که این نشان‌دهنده نقش این توالی‌ها در مسیرهای مختلف متابولیکی فعال در استویا است و همچنین اینکه به‌طور مستقیم یا غیرمستقیم با مسیر بیوسنتز استویول‌گلیگوزیدها مرتبط هستند. بنابراین با توجه به نتایج حاصل می‌توان این فرضیه را مطرح کرد که نشانگرهای PolySSR توسعه‌یافته در این مطالعه با اطمینان قابل استفاده در برنامه‌های به‌نژادی مولکولی استویا به‌منظور انتخاب واریته‌های با میزان بالای SG به‌ویژه RebA هستند. در فاز دوم مطالعه حاضر، مطالعات آزمایشگاهی و مزرعه‌ای نشانگرها و آغازگرهای معرفی‌شده روی تعدادی از ارقام و واریته‌های استویا در دست بررسی است.

تشکر و قدردانی: مطالعه حاضر در قالب پروژه تحقیق و توسعه گیاه استویا از سوی شرکت فرآورده‌های طعم طبیعت حامی به عنوان کارفرما ابلاغ شده است که بدین وسیله از این مجموعه قدردانی به عمل می‌آید.

تاییدیه‌اخلاقی: موردی توسط نویسندگان ذکر نشده است.

تعارض منافع: موردی توسط نویسندگان ذکر نشده است.

سهم نویسندگان: مجتبی خیام‌نکویی (نویسنده اول)، پژوهشگر اصلی/نگارنده بحث (٤٠%)؛ مریم معظم‌جزی (نویسنده دوم)، روش‌شناس/پژوهشگر اصلی/تحلیلگر آماری (۳۰%)؛ محسن مردی (نویسنده سوم)، پژوهشگر کمکی/تحلیلگر آماری (۲۰%)؛ سعید کدخدایی (نویسنده چهارم)، نگارنده مقدمه/روش‌شناس/نگارنده بحث (۱۰%)

منابع مالی: منابع مالی مربوط به این مطالعه توسط شرکت فرآورده‌های طعم طبیعت حامی تامین شد.

منابع

1- Yadav AK, Singh S, Dhyani D, Ahuja PS. A review on the improvement of stevia [Stevia rebaudiana (Bertoni)]. Can

J Plant Sci. 2011;91(1):1-27.
2- Soufi S, D'Urso G, Pizza C, Rezgui S, Bettaieb T, Montoro P. Steviol glycosides targeted analysis in leaves of Stevia rebaudiana (Bertoni) from plants cultivated under chilling stress conditions. Food Chem. 2016;190:572-80.
3- Wölwer-Rieck U. The leaves of Stevia rebaudiana (Bertoni), their constituents and the analyses thereof: A review. J Agric Food Chem. 2012;60(4):886-95.
4- Zalapa JE, Cuevas H, Zhu H, Steffan Sh, Senalik D, Zeldin E, et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot. 2012;99(2):193-208.
5- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12(7):499-510.
6- Zhang H, Wei L, Miao H, Zhang T, Wang C. Development and validation of genic-SSR markers in sesame by RNA-seq. BMC Genom. 2012;13(1):316.
7- Huang H, Xia EH, Zhang HB, Yao QY, Gao LZ. De novo transcriptome sequencing of Camellia sasanqua and the analysis of major candidate genes related to floral traits. Plant Physiol Biochem. 2017;120:103-11.
8- Chen J, Hou K, Qin P, Liu H, Yi B, Yang W, et al. RNA-Seq for gene identification and transcript profiling of three Stevia rebaudiana genotypes. BMC Genom. 2014;15(1):571.
9- Xia EH, Yao QY, Zhang HB, Jiang JJ, Zhang LP, Gao LZ. CandiSSR: An efficient pipeline used for identifying candidate polymorphic SSRs based on multiple assembled sequences. Front Plant Sci. 2016;6:1171.
10- Mathur Sh, Bulchandani N, Parihar S, Shekhawat GS. Critical review on steviol glycosides: Pharmacological, toxicological and therapeutic aspects of high potency zero caloric sweetener. Int J Pharmacol. 2017;13(7):916-28.
11- Vieira ML, Santini L, Diniz AL, Munhoz CD. Microsatellite markers: What they mean and why they are so useful. Genet Mol Biol. 2016;39(3):312-28.
12- Sharma N, Kaur R, Era V. Potential of RAPD and ISSR markers for assessing genetic diversity among Stevia rebaudiana Bertoni accessions. Indian J Biotechnol. 2016;15(1):95-100.
13- Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinform. 2005;6(1):31.
14- Shi JI, Huang Sh, Zhan JI, Yu JI, Wang XI, Hua WE, et al. Genome-wide microsatellite characterization and marker development in the sequenced Brassica crop species. DNA Res. 2014;21(1):53-68.
15- Kim MJ, Jin J, Zheng J, Wong L, Chua NH, Jang IC. Comparative transcriptomics unravel biochemical specialization of leaf tissues of Stevia for diterpenoid production. Plant Physiol. 2015;169(4):2462-80.
16- Rodrıguez-Concepción M, Boronat A. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. Plant Physiol. 2002;130(3):1079-89.