

"Research Paper"

Large Scale Identification of SSR Molecular Markers Using RNA Sequencing in Yarrow (*Acillea millefolium* L.)

Behnaz Shahbaz¹, Mehdi Soltani Howyzeh² and Vahid Shariati³

1- Department of Genetics and Plant Breeding, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

2- Department of Genetics and Plant Breeding, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran,

(Corresponding author: soltani.m@iauahvaz.ac.ir)

3- Molecular Biotechnology Department, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

Received: 25 March, 2021

Accepted: 24 January, 2023

Extended Abstract

Introduction: The medicinal plant Yarrow (*Achillea millefolium* L.) due to its medicinal and industrial functions is one of valuable pastureland plant in Iran and the world.

Materials and methods: In this study the Illumina Hiseq 2500 platform used to identify the microsatellites markers by transcriptome sequencing of the leaf and inflorescence of Yarrow.

Results: According to the *de novo* sequencing results of the medicinal plant Yarrow cultivated in the Research Institute of Forests and Rangelands, 9450 single transcript sequences (6920 single gene sequences) containing 10,570 potential microsatellites were identified. The most common types of microsatellites were dinucleotides (62%), followed by trinucleotides (2748, 26%), and mononucleotides (7%), respectively. In total, 69% of the microsatellites were classified as second class (10 to 20 nucleotides), and 31% were first class (more than 20 nucleotides). The frequency of microsatellites in the transcriptome of inflorescence was one per 10 kb assembled sequences. According to unigenes annotation results, totally 1,762 (22.3%), 4723 (53.3%), 3714 (47.5%), 4517 (51.3%) and 5189 (60.7%) unigenes annotated from all databases of KAAS, Arabidopsis, UniProt, NCBI database Non-redundant proteins and sunflower respectively. Among 8542 clustered unigenes containing microsatellite, 38440 unigene sequences (45%) were classified as functional and categorized as 52 clusters.

Conclusion: Among all the categorized monomers, 3220 (65%) belonged to the "metabolic process" category, of which 90 (3%) belonged to the "secondary metabolic process" category. The introduction of these markers can be used for future studies of selection using markers, genetic diversity, and genetic maps in this medicinal plant breeding programs.

Key words: De novo, Functional annotation, Genetic variation, Transcriptome



"مقاله پژوهشی"

شناسایی مقیاس گسترده نشانگرهای مولکولی ریزماهوره از طریق توالی‌یابی RNA در گیاه دارویی بومادران

بهناز شهباز^۱، مهدی سلطانی حویزه^۲ و وحید شریعتی^۳

۱- گروه ژنتیک و به‌نژادی گیاهی، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران
 ۲- گروه ژنتیک و به‌نژادی گیاهی، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران (نویسنده مسوول: soltani.m@iauhvaz.ac.ir)
 ۳- گروه زیست فناوری مولکولی، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، تهران، ایران
 تاریخ دریافت: ۱۴۰۰/۱/۵ تاریخ پذیرش: ۱۴۰۱/۱۱/۴
 صفحه: ۱۵۶ تا ۱۶۵

چکیده مبسوط

مقدمه: گیاه دارویی بومادران (*Achillea millefolium* L.) به دلیل داشتن کارکردهای مختلف دارویی و صنعتی یکی از گونه‌های با ارزش در مراتع ایران و جهان است.

مواد و روش‌ها: در این مطالعه، از توالی‌یابی ترنسکرپتوم برگ و گل آذین بومادران با استفاده از پلتفرم Illumina Hiseq ۲۵۰۰ جهت شناسایی نشانگرهای ریزماهوره استفاده شد.

یافته‌ها: طبق نتایج توالی‌یابی نوپدید حاصل از نمونه‌های به دست آمده از گیاه دارویی بومادران کشت شده در موسسه تحقیقات مراتع و جنگل‌های کشور، ۹۴۵۰ توالی تک رونوشت (۶۹۲۰ توالی تک ژن) حاوی ۱۰۵۷۰ ریزماهوره بالقوه یافت شد. فراوان‌ترین نوع ریزماهوره‌ها، دی‌نوکلئوتیدها (۶۲٪)، بعد از آن تری‌نوکلئوتید (۲۷۴۸ عدد، ۲۶٪)، مونوکلئوتیدها (۷٪) بودند. مجموعاً، ۶۹ درصد ریزماهوره‌ها از نوع ریزماهوره کلاس دوم (۱۰ تا ۲۰ نوکلئوتید) و ۳۱ درصد از نوع ریزماهوره کلاس اول (بیش از ۲۰ نوکلئوتید) بودند. فراوانی ریزماهوره‌ها در ترانسکرپتوم گل آذین یک در هر ۱۰ کیلوباز توالی یکپارچه شده بود. در تفسیر تک‌ژن‌ها مجموعاً ۱/۷۶۲ (۲۲/۳٪)، ۴۷۲۳ (۵۳/۳٪)، ۳۷۱۴ (۴۷/۵٪)، ۴۵۱۷ (۵۱/۳٪) و ۵۱۸۹ (۶۰/۷٪) تک‌ژن از همه کتابخانه‌ها به ترتیب به پایگاه‌های داده KAAS، آرآیدوپسیس، UniProt، پروتئین‌های Non-redundant پایگاه NCBI و آفتابگردان گرامان شدند. در بین ۸۵۴۲۱ توالی تک ژن حاوی تکرارهای ریزماهوره، تعداد ۳۸۴۴۰ توالی تک ژن (۴۵٪) دارای دسته‌بندی کارکردی و همه‌ی تک‌ژن‌ها در ۵۲ دسته قرار داشتند.

نتیجه‌گیری: در میان همه تک‌ژن‌های دسته‌بندی شده ۳۲۲۰ تک‌ژن (۶۵ درصد) به‌دسته "فرایند متابولیکی" متعلق بود که از میان آنها ۹۰ تک‌ژن (۳٪) به دسته "فرایند متابولیکی ثانویه" تخصیص داشت. معرفی این نشانگرها می‌تواند برای مطالعات آینده انتخاب به کمک نشانگر، تنوع ژنتیکی و ساخت نقشه‌های ژنتیکی در این گیاه دارویی مورد بهره‌برداری قرار گیرد.

واژه‌های کلیدی: ترنسکرپتوم، تنوع ژنتیکی، گزارمان کارکردی، نوپدید

مقدمه

نشانگرهای مولکولی از ابزارهای کاربردی در بررسی تغییر ژنتیکی در گیاهان هستند. در میان این نشانگرها، ریزماهوره‌ها (SSR) به دلیل تکرارپذیری بالا، طبیعت چند آلی، وراثت‌پذیری هم‌باز، فراوانی نسبی بالا و پوشش خوب ژنوم، کاربردهای فراوانی در ژنتیک و اصلاح گیاهان دارند (۵، ۴۱). توالی تکراری ساده ریزماهوره (SSR) یک توالی تکراری DNA می‌باشد که عموماً شامل ۶-۱ واحد نوکلئوتیدی است. نشانگرهای SSR خاص یک مکان ژنی هستند و به صورت هم‌باز توارث می‌یابند. این نشانگرها علاوه بر تنوع آلی بالایی که نشان می‌دهند، در کل ژنوم نیز پراکنده شده‌اند. اگر چه هزینه استفاده از این نوع نشانگرها زیاد است ولی به واسطه پتانسیل بالای آن‌ها در تشخیص پلی‌مورفیسم و هتروزیگوسیتی در گونه‌های موجودات زنده به فراوانی استفاده می‌شوند (۳۲). مهمترین کاربرد نشانگرهای ریزماهوره، بررسی تنوع ژنتیکی، بررسی تبارزایی (فیلوژنی) و تکاملی، تهیه نقشه ژنومی و نشانگری، همسانه کردن ژنی و نقشه‌برداری ژنتیکی، پژوهش‌های بوم‌شناختی (اکولوژیکی) و تکامل جمعیت‌ها، انتخاب نشانگرهای همراه با صفت مورد نظر، خالص‌سازی مواد ژنتیکی و انجام بررسی‌های سیتوژنتیکی است (۱، ۵). تاکنون استفاده از این نشانگرها در مطالعه‌ی تنوع ژنتیکی تعدادی از گیاهان دارویی نیز گزارش شده است (۲۱، ۲۶).

RNA-Seq یکی از پیچیده‌ترین کاربردهای توالی‌یابی نسل آینده (NGS) است. تکنیک RNA-seq در ابتدا برای آنالیز سطح بیان ژن توسعه‌یافت، اما امروزه برای شناسایی SNP و SSR در نواحی رونویسی‌شده ژنوم گیاهان مورد استفاده قرار می‌گیرد. این تکنیک یک روش بالقوه برای ژنوتایپینگ گونه‌های دارای ژنوم بزرگ است که توالی‌یابی مستقیم آن‌ها پرهزینه است. آزمایش‌های قبلی بر روی گندم وحشی (۲۸) و انسان (۳۱) قابلیت استفاده این تکنیک به‌عنوان یک روش قدرتمند برای شناسایی چندشکلی در نمونه‌های دارای ژنوم بزرگ را اثبات کرده است. قبل از توالی‌یابی RNA، فرایند شناسایی نشانگرهای SSR بسیار پرهزینه، زمان‌بر و به دلیل نیاز به ساخت کتابخانه ژنومی برای موتیف‌های SSR از طریق ایجاد مولکول‌های DNA نوترکیب و آنزیم‌های برشی بسیار سخت بود (۳۹). توالی RNA فقط از توالی اکزون حاصل شده و به همین خاطر می‌تواند برای ایجاد نشانگرهای اختصاصی نواحی ژنومی که عامل احتمالی تغییرات فنوتیپی مفید در برنامه‌های اصلاحی گیاهان هستند مورد استفاده قرار گیرند. علاوه بر این، دقت بالا و اختلاف بسیار کم در بین تکرارهای تکنیکی این روش را سودمندتر ساخته است. تکنیک RNA-Seq در شناسایی نشانگرهای SSR به‌ویژه در گیاهان غیر مدل و بدون ژنوم مرجع بسیار مفید است (۲، ۱۴، ۳۷، ۴۵). حتی در گیاهان غیرمدل بدون ژنوم مرجع حجم عظیمی از

بومادران به‌عنوان یکی از مهم‌ترین گیاهان دارویی در کشور لزوم انجام تحقیقات ژنومیکس و ترانسکریپتومیکس را بیش از پیش ضروری می‌نماید. وجود داده‌هایی از این قبیل برای درک بهتر مسیر بیوسنتزی ترکیبات ثانویه، شناسایی ژن‌های عملکردی و نشانگرهای مولکولی ارزشمند این گیاه حائز اهمیت فراوان است. هدف پژوهش حاضر شناسایی مقیاس گسترده نشانگرهای مولکولی SSR برای نخستین بار از مجموعه رونوشت‌های توالی‌یابی شده گیاه دارویی بومادران می‌باشد.

مواد و روش‌ها تهیه مواد گیاهی

نمونه‌برداری برگ و گل آذین از ۱۲ بوته گیاه دارویی بومادران کشت شده در گلدان در گلخانه مؤسسه تحقیقات مراتع و جنگل‌های کشور در سال ۱۳۹۹ به صورت تصادفی انجام گرفت و سپس هر ۶ نمونه با هم مخلوط و دو نمونه نهایی برگ و گل جهت توالی‌یابی ارسال شدند. پس از نمونه‌برداری، بافت‌ها به نیتروژن مایع انتقال و در دمای -80°C تا زمان استخراج RNA، نگهداری شدند.

استخراج RNA ساخت کتابخانه و توالی‌یابی

از نمونه‌های تهیه شده، استخراج RNA کل به وسیله بافر TRIzol reagent (Invitrogen) صورت گرفت و جهت حذف آلودگی DNA از کیت DNaseI (TURBO DNaseI, Ambion) طبق پروتکل شرکت سازنده تیمار گردید. برای تعیین کیفیت و کمیت RNA استخراج شده به‌ترتیب از ژل آگارز ۱٪ و دستگاه نانودراپ مدل NanoDrop 1000 spectrophotometer (Thermo Scientific) استفاده شد. سنجش کنترل کیفیت (QC) بر روی نمونه‌ها بر روی RNA کل با دستگاه Bioanalyzer مدل (Hørsholm, Agilent Technologies Denmark) توسط شرکت سرویس‌دهنده توالی‌یابی صورت گرفت. همه نمونه‌های دارای عدد RIN (RNA integrity number) بالای ۸ بودند. انتخاب دم پلی A، آماده‌سازی cDNA اتصال آداپتورها، تشکیل کلاسترها و توالی‌یابی در مؤسسه Beijing Genomes Institute (China) طبق توصیه سازنده با کیت استاندارد ایلومینا انجام گردید. توالی‌یابی توسط پلتفرم Illumina HiSeq 2500 platform به صورت دو طرفه (paired-end) و طول ۱۵۰ نوکلئوتید انجام شد. پس از انجام توالی‌یابی توسط شرکت سرویس‌دهنده توالی‌یابی (BGI) در کشور چین، توالی‌ها به صورت خوانش خام نوکلئوتیدی جهت انجام سرهم‌بندی^۲ توسط نرم‌افزارهای یکپارچه‌سازی توالی مورد استفاده قرار گرفتند که در نهایت کانتیگ‌های تولید شده در فرآیند سرهم‌بندی به یونی‌ژن تبدیل گشتند. یونی‌ژن‌ها جهت آنالیزهای بعدی مورد استفاده قرار گرفتند.

آماده سازی داده‌های RNA-Seq و یکپارچه‌سازی نوپدید

آداپتورها و توالی‌های کم کیفیت خوانش‌های خام با استفاده از نرم‌افزار Trimmomatic (ورژن 0.36) و مقدار $QC \geq 30$ حذف شدند. پس از تریم کردن (Trimming)، نرم‌افزار FastQC برای ارزیابی ویژگی‌های کتابخانه و تأیید

داده‌های توالی‌یابی توسط این تکنیک به دست می‌آید، که از خوانش‌های ایجادشده از میلیون‌ها باز در طول یک روز از طریق یک دستگاه می‌تواند برای ایجاد EST-SSR در کل ژنوم مورد استفاده قرار گیرد (۱۲). در تکنیک RNA-Seq در صورت وجود ژنوم مرجع، خوانش‌های حاصل باید با ژنوم یا رونوشت مرجع هم‌تراز شود، اما در نبود ژنوم مرجع، نیازمند نقشه‌یابی نوپدید است (۱۳، ۳۴، ۳۶، ۴۰).

توالی‌یابی ترانسکریپتوم روش بسیار موثری برای کشف و شناسایی مکان‌های SSR در گیاهان بوده و نتایج خوبی تاکنون ارائه کرده است (۳۹). در تحقیق سلطانی حویزه و همکاران (۳۷) از توالی‌یابی ترانسکریپتوم *Trachyspermum ammi* (توسط پلتفرم Illumina ۲۰۰۰ HiSeq برای اولین بار ۱۱۹۹۸ توالی یونی‌ترانسکریپت ۷۹۱۳ توالی یونی ژن) حاوی ۱۳۵۹۳ ریزماهوره بالقوه یافت شد، بطوریکه فراوان‌ترین نوع ریز ماهرها، دی نوکلئوتیدها (۱۷ درصد) و تری نوکلئوتیدها (۲۶ درصد) بودند. هان و همکاران (۱۷) از طریق شناسایی نشانگر SSR براساس توالی‌یابی RNA برای ژرم‌پلاسماهای دیپلوئید و تتراپلوئید گیاه گل داوودی (*Chrysanthemum indicum*) دریافتند که تفاوت معنی‌دار در آنتولوژی ژن (GO) و تفسیر KEGG برای ژن‌های همولوگ اختصاصی با استفاده از روش OrthMCL برای شناسایی خانواده‌ی ژنی همولوگ بین دو سیتوتایپ^۱ بدست می‌آید.

بومادران با نام علمی *Achilla millefolium* که معمولاً به‌عنوان Yarrow شناخته می‌شود، بومی اروپا و آسیاست. یک گیاه علفی، چندساله (۸، ۳۰)، کم توقع و مقاوم به خشکی است که دارای گل‌آذین کوچک به قطر ۷ تا ۱۵ سانتی‌متر و ارتفاع بوته آن ۲۰-۴۰ سانتی‌متر می‌رسد. جنس *Achillea* دارای بیش از ۱۳۰ گونه‌ی گلدار و چندساله است که عمدتاً در اروپا و مناطق معتدل آسیا و برخی مناطق آمریکای شمالی پراکنده هستند. از شاخصه‌های محبوب چندین گونه‌ی این جنس: درمان جراحات، خونریزی، سردرد، التهاب، درد، بیماری‌های تشنج، نفخ و سوءهاضمه می‌باشد. تحقیقات فیتوشیمیایی بر روی گونه‌های *Achillea* آشکارکرد که ترکیبات زیادی از این جنس، زیست‌فعالیت بالایی دارند (۳۳، ۳۸). براساس داده‌های بدست آمده از ترکیبات فرار، ۶ نوع ترکیب شیمیایی از بومادران شناسایی شده است. با توجه به تغییرات ژنتیکی و وجود ژنوتیپ‌های مختلف بومادران، هر ژنوتیپ ترکیب شیمیایی منحصر به فرد خود را دارد، به همین خاطر طیف وسیعی از خصوصیات دارویی به این گیاه نسبت داده می‌شود. در گیاه بومادران سزکوئی‌ترین‌ها و مونوترپن‌ها بیشترین ترکیبات فرار را تشکیل می‌دهند. ترکیبات فعالی مانند کاریوفیلن اکساید، کامفور، گاما-مورولن، گاما اودزمال، برنئول، بتا-اودزما، اسپاتولنول، germacrene D، و γ -gurjunene محتوای ترکیبات اصلی فرار در اکوتیپ *Achillea millefolium* را تشکیل می‌دهند (۱۵، ۳۵).

تاکنون (سال ۱۳۹۹) اطلاعاتی درباره‌ی شناسایی مقیاس گسترده نشانگرهای SSR حاصل از داده‌های توالی‌یابی ترانسکریپتوم گیاه دارویی گزارش نشده است. از سوی دیگر، وجود داده‌های مولکولی بسیار اندک و ناچیز در ارتباط با

(۳٪)، هگزانونوکلئوتیدها (حدود یک درصد) و پنتانونوکلئوتیدها (کمتر از یک درصد) قرار داشتند (شکل ۱). نتایج مشابهی در برخی گیاهان دولپه‌ای دیگر مانند آراییدوپسیس، زنیان، انگور و بادام‌زمینی گزارش شده است (۱۰، ۲۰، ۳۷).

در تحقیقات قبلی ریزماهورهای دی‌نوکلئوتیدی، تکرار غالب در زنیان (*Sonneratia alba* (۳۷)، نوعی شاهپسند (*Sesamum indicum* (۹)، کنجد (*Sa. (22)*، *P. notoginseng* (۲۵) و *Gi. Biloba* (۲۳) بود در حالیکه توالی‌های تکراری کوتاه تری‌نوکلئوتیدی، نوع غالب تکرار در مجموعه داده ترانسکریپتوم *A. tridentate* (درمنه معمولی) (۴)، *Capsicum annuum* (فلفل قرمز) (۲۴)، *Citrullus lanatus* (هندوانه) (۱۶)، *Scabiosa columbaria* (گل داوودی) (۳) و *Polygonum cuspidatum* (هفت بند ژاپنی) (۱۸) بوده است.

در این پژوهش، در مجموع ریزماهورهای با تکرارهای شش‌تایی فراوان‌ترین بودند و پس از آنها به‌ترتیب ریزماهورهایی با پنج، هفت و هشت تکرار فراوان‌ترین بودند (جدول ۱). در گیاه کنجد نتایج کاملاً مشابهی در مورد بیشترین تکرارهای ریزماهورها بدست آمد (۴۳). در گیاه ارکیده دندروبیوم (*Dendrobium officinale*) نیز ریزماهورهای با تکرارهای شش‌تایی فراوان‌ترین بودند و پس از آنها به‌ترتیب ریزماهورهایی با پنج، هفت و هشت تکرار قرار گرفتند (۴۴). سلطانی حویزه و همکاران (۳۷) نتایج مشابهی در خصوص فراوانی ریزماهورها در گیاه دارویی زنیان بدست آوردند.

فراوان‌ترین ریزماهور دی‌نوکلئوتیدی بدست آمده در ترانسکریپتوم گیاه دارویی بومادران AG/CT با ۲۴۹۰ تکرار (۳۸ درصد) بود. در میان ریزماهورهای تری‌نوکلئوتیدی نیز به‌ترتیب AAC/GTT، AGC/GCT، ATC/GAT و AAT/ATT بیشترین فراوانی رو به خود اختصاص دادند (شکل ۲). در گیاهان ارکیده دندروبیوم، زنیان و کنجد نتایج مشابهی گزارش شده است. در این گیاهان نیز فراوان‌ترین ریزماهور دی‌نوکلئوتیدی AG/CT و فراوان‌ترین ریزماهور تری‌نوکلئوتیدی AAG/CTT بود (۳۷، ۴۳، ۴۴). این در حالیست که در گیاه بامبو (*Dendrocalamus latiflorus*) فراوان‌ترین ریزماهور دی‌نوکلئوتیدی AG/CT گزارش گردید اما از سوی دیگر فراوان‌ترین ریزماهور تری‌نوکلئوتیدی CCG/CGG بود (۷).

با استفاده از آنالیز داده‌های RNA-Seq بافت ریشه برای دو وارسته از گیاه گوار^۱ با غربالگری مجموع ۱۰۲۴۷۹ تک‌ژن از طریق نرم‌افزار MISA و ابزار SAM، ۲۵۰۴۰ نشانگر با دقت بالا شناسایی شد که از بین آن‌ها ۱۸۷۹۲، ۵۹۹۹ و ۲۴۹ نشانگر به‌ترتیب برای SSR، SNP و InDel گزارش شدند. بیشترین موتیف‌های SSR مربوط به AG، AAAG، CTTT، ATGAG و TGCTAT بود. حدود ۸۰٪ نشانگرهای SSR با استفاده از PCR تکثیر و مورد تأیید قرار گرفتند (۱۲).

کارایی ترسیم‌کردن مورد استفاده قرار گرفت. خوانش‌های فیلترشده با کیفیت بالا برای تحلیل‌های بعدی مورد استفاده واقع شدند. یکپارچه‌سازی نوپدید خوانش‌های فیلترشده با کیفیت بالای به دست آمده با استفاده از نرم‌افزار Trinity انجام گردید. پس از بهینه‌سازی، Kmer معادل ۳۲ انتخاب و برای سایر پارامترهای نرم‌افزار Trinity از تنظیمات استاندارد استفاده شد. توالی‌های بدست آمده از Trinity تک‌رونوشت نامیده می‌شوند که ایزوفرم‌های رونوشت مورد قبول را نشان می‌دهد. کلاستر تک‌رونوشت‌ها، تک‌ژن نامیده می‌شود که ژن‌های شناسایی شده مورد قبول را نشان می‌دهد، مورد استفاده واقع شدند.

شناسایی نشانگرهای ریزماهور

نشانگرهای ریزماهور با استفاده از ابزار شناسایی ریزماهور به نام MISA (<http://pgrc.ipk-gatersleben.de/misa>) شناسایی شدند. حداقل تکرار برای نشانگرهای دی‌نوکلئوتیدی "ده" و برای نشانگرهای تتری، پنتا و هگزا-نوکلئوتیدی "پنج" قرار داده شد (۶).

تفسیر کارکردی تک‌ژن‌های یکپارچه‌شده بومادران حاوی ریزماهور

برای تفسیر کارکردی تک‌ژن‌های یکپارچه شده ترانسکریپتوم این گیاه (به‌عنوان غیرمدل) علیه پایگاه‌های اطلاعاتی پروتئین‌های NCBI Non-redundant، UniProt، آراییدوپسیس و آفتابگردان با E-Value کمتر از ۱۰^{-۵} به وسیله پایپ لاین تفسیر نرم‌افزار Trinotate^۱ بلاست شد. بدین منظور سرور محلی (Local Server) دارای ۲۵۶ گیگابایت حافظه موقت (RAM) و سی‌پی‌یو با ۸۰ هسته و سیستم عامل لینوکس استفاده شد. تفسیر اتوماتیک KAAS^۲ برای تعیین مسیرهای متابولیتی و توصیف کارکردی هر یک از تک‌ژن‌های بومادران به وسیله دایره‌المعارف ژن‌ها و ژنوم کیوتو (KEGG^۳) بر اساس تخصیص عدد KO^۴ به روش SBH^۵ صورت گرفت. شناسه KO برای تفسیر خودکار توالی‌ها توسط پایگاه KAAS لازم است و هر KO معادل یک گروه ارتولوگ از ژن‌هایی است که به صورت مستقیم به یک رکورد در مسیرهای پایگاه KEGG لینک گردیده است (۲۷). طبقه‌بندی کارکردی ژن شناسی (GO) برای همه یونی ژن‌های یکپارچه شده به وسیله نرم‌افزار تحت وب WEGO^۶ صورت گرفت.

نتایج و بحث

در مجموع تعداد ۹۴۵۰ توالی تک رونوشت (۶۹۲۰ توالی تک ژن) حاوی ۱۰۵۷۰ ریزماهور بالقوه یافت شد. ریزماهورهای دی‌نوکلئوتیدی (۶۲٪) در مجموع بیشترین تکرار را داشتند. پس از دی‌نوکلئوتیدها به‌ترتیب تری‌نوکلئوتیدها (۲۷۴۸ عدد، ۲۶ درصد)، مونونوکلئوتیدها (۷٪)، تترانونوکلئوتیدها

1- <http://trinotate.github.io>

4- KEGG orthology

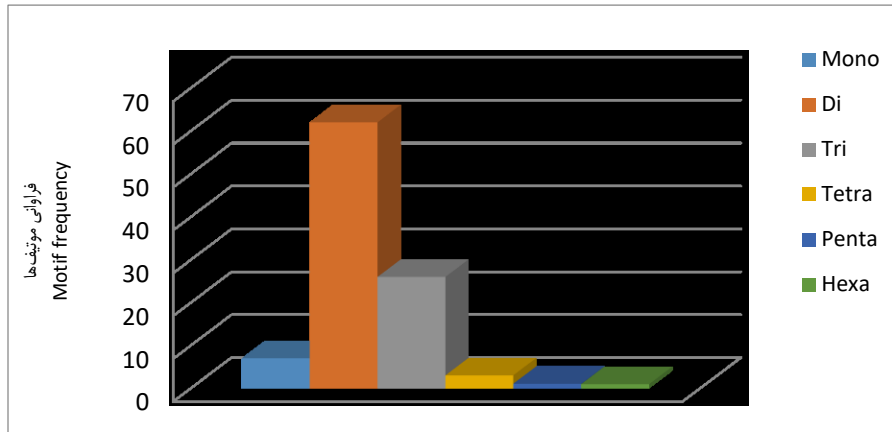
7- Cyamopsis tetragonoloba

2- KEGG automatic annotation server

5- Single-directional best hit

3- Kyoto encyclopedia of genes and genome

6- <http://wego.genomics.org.cn/cgi-bin/wego/index.pl>



شکل ۱- فراوانی نسبی نوع ریزماهواره‌های شناسایی شده در ترانسکریپتوم گیاه دارویی بومادران

Figure 1. Relative frequency of microsatellite repeat motif identified in medicinal plant transcriptome of *A. millefolium*

در میان نشانگرهای ریزماهواره شناسایی شده در بومادران، ۶۹ درصد از نوع ریزماهواره کلاس دوم (به طول ۱۰ تا ۲۰ نوکلئوتید) و ۳۱ درصد از نوع ریزماهواره کلاس اول (به طول بیش از ۲۰ نوکلئوتید) بودند (شکل ۳). فراوانی تکرارهای ریزماهواره در ترانسکریپتوم گل‌آذین گیاه دارویی بومادران تقریباً یک در هر ۱۰/۱ کیلو باز توالی یکپارچه شده بود. در گیاه ارکیده دندروبیوم فراوانی ریزماهواره‌های نوع کلاس دوم ۹۲ درصد و فراوانی ریزماهواره‌های نوع کلاس اول ۸ درصد بود. در این گیاه فراوانی تکرارهای ریزماهواره تقریباً یک در ۴ کیلو باز توالی یکپارچه شده بود (۴۴). فراوانی تکرارهای ریزماهواره در ترانسکریپتوم گل‌آذین گیاه دارویی بومادران کمتر از فراوانی تکرارهای ریزماهواره در گیاهان برنج (یک در ۳/۶ کیلو باز)، سورگوم (یک در ۵/۹ کیلو باز)، جو (یک در ۸/۹ کیلو باز) و ذرت (یک در ۹ کیلو باز) و بیشتر از فراوانی تکرارهای ریزماهواره در ترانسکریپتوم گیاهان گندم (یک در ۱۰/۶ کیلو باز) و نیشکر (یک در ۱۰/۹ کیلو باز) بود (۲۹).

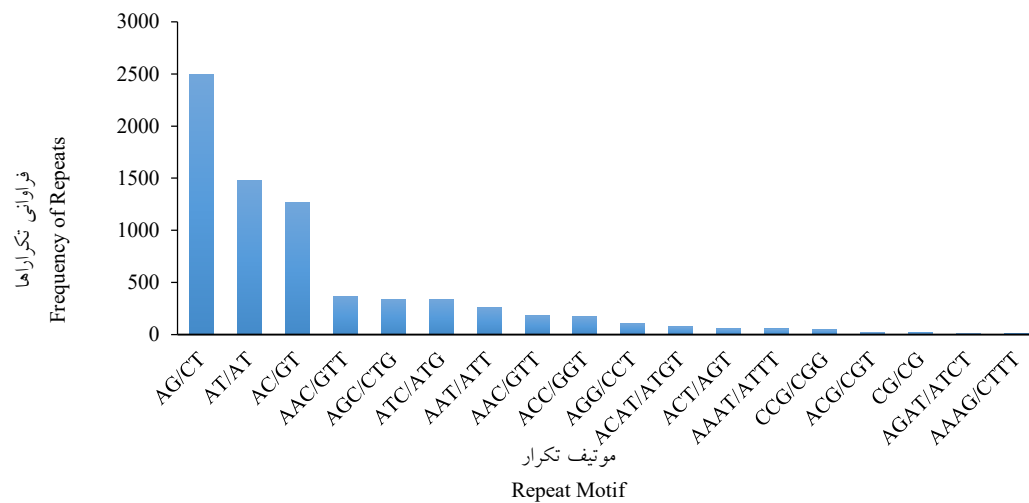
یائو لی^۱ و همکاران (۲۰۱۴) در مجموع ۱۹۳۹ نشانگر SSR از ژنوم توالی‌یابی شده‌ی کرفس^۲ شناسایی کردند که تکرارهای A/T بیشترین در بین مونونوکلوئوتیدها، AG/CT بیشترین در بین دی‌نوکلئوتیدها (۴۵/۵۴٪) و در بین تری‌نوکلئوتیدها ACC/GGT و AAG/CTT، ATC/ATG، ACT/AGT بیشترین فراوانی را داشتند (۴۷). به‌طور مشابهی فراوانی A/T، AG/CT و AAG/CTT در بین تکرارهای مونو، دی و تری‌نوکلئوتیدی بیشترین در گیاه *Sterculia lanceolata* گزارش شد (۱۱).

در بررسی ژنوم ۶ رقم گیاه قیاق^۳ بیشترین تکرار نشانگرهای SSR مربوط به موتیف‌های سه تکراری CCG/CGG به تعداد ۶۸۶۰ (۳۹٪)، بعد از آن موتیف‌های مونونوکلوئوتیدی (A) به تعداد ۶۷۷۱ (۳۸/۵۹٪) و سپس دی نوکلئوتیدی (AG) به تعداد ۳۳۶۶ (۱۹/۱۸٪) گزارش شد. همچنین در بین تکرارهای دارای طیف ۲۳-۵، نشانگرهای SSR با تکرارهای ۸-۵ بیشترین بودند و تکرارهای ۱۲-۹ در جایگاه بعدی قرار داشتند (۴۸).

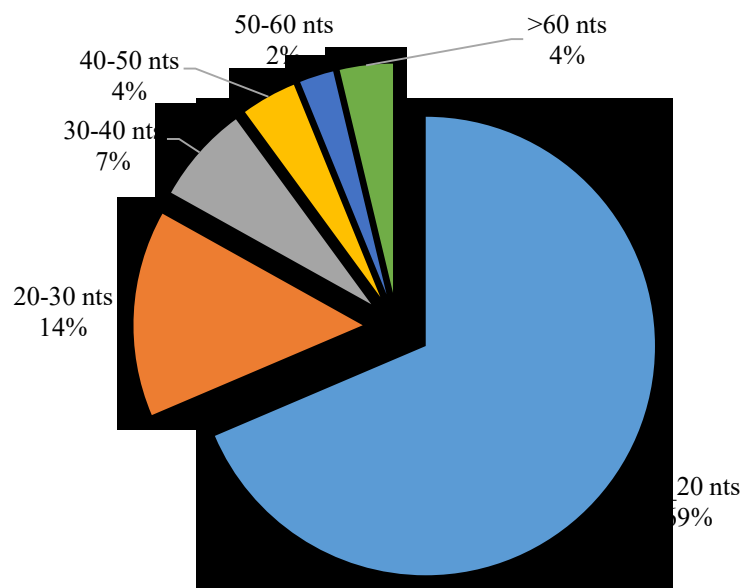
جدول ۱- فراوانی ریزماهواره‌های بومادران براساس نوع و تعداد تکرارها

Table 1. Frequency of *Achillea millefolium* microsatellites based on type and number of repeats

مجموع Total	نوع Type						تکرار Repeat
	هگزا Hexa	پنتا Penta	تترا Tetra	تری Tri	دی Di	مونو Mono	
1760	34	77	202	1447			5
2518	37	14	75	600	1792		6
1501	19	8	11	323	1140		7
1034	4	4	8	191	827		8
613	3	3	7	31	569		9
547	2	2	3	39	376	125	10
508	1	2	2	21	332	150	11
422	1	1	4	19	300	97	12
266	0	0	2	24	147	93	13
227	0	0	2	16	145	64	14
189	0	0	1	5	130	53	15
151	0	0	0	9	119	23	16
152	0	0	0	5	117	30	17
115	0	0	0	6	104	5	18
116	0	0	0	5	101	10	19
451	0	0	0	7	354	90	≥20
10570	101	111	317	2748	6553	740	مجموع Total



شکل ۱- فراوانی موتیف تکرارهای ریزماهواره شناسایی شده در ترانسکریپتوم گیاه دارویی بومادران
Figure2. Frequency of microsatellite repeat motif identified in medicinal plant transcriptome of *A.millefolium*



شکل ۲- فراوانی طول تکرارهای ریزماهواره شناسایی شده در ترانسکریپتوم گیاه دارویی بومادران
Figure3. Frequency of microsatellite repeat length identified in medicinal plant transcriptome of *A.millefolium*

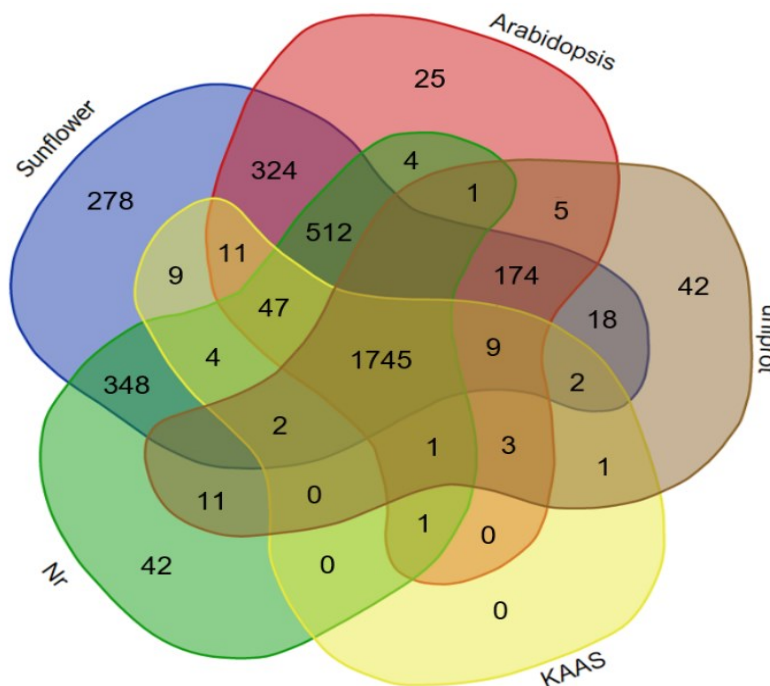


Figure 4. Venn diagram of BLAST annotation results of assembled unigenes containing microsatellites against different databases

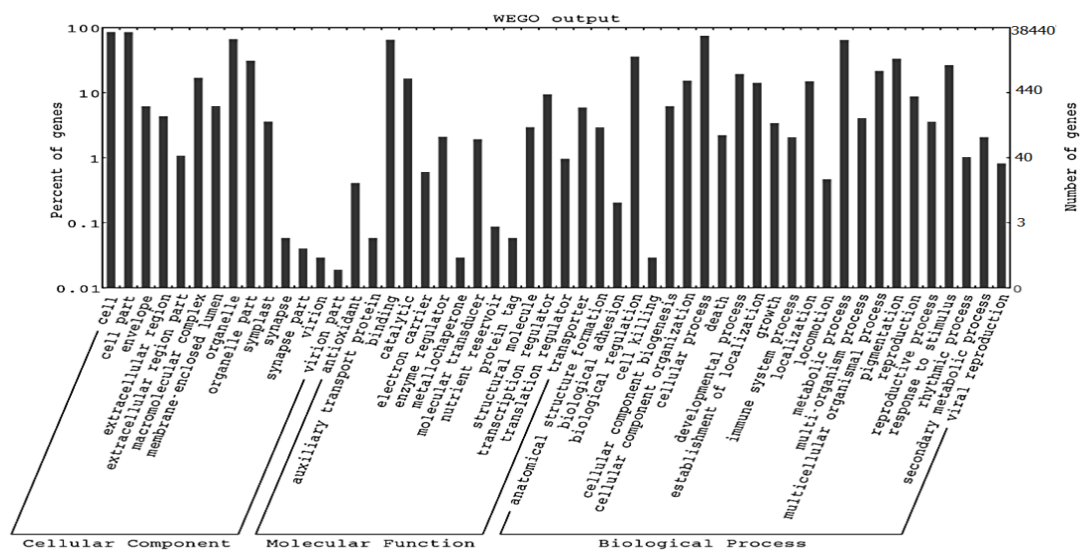


Figure 5. Functional classification of assembled unigenes containing microsatellites. Each bar represents the percentage of unigenes match to each GO term. Plot with Y axis is in log (10) scale

UniProt و آفتابگردان انجام گردید. نتایج بلاست پوشش گسترده ترانسکریپتوم بومادران را نشان دادند.

تک‌ژن‌های بومادران بطور تقریباً مشابهی با پایگاه داده آفتابگردان، بیش از سایر پایگاه‌های داده تفسیر شدند. این امر تأیید کننده این مسئله است که بومادران از خانواده کامپوزیته

تفسیر کارکردی تک‌ژن‌های یکپارچه‌شده بومادران حاوی ریزماهواره

تفسیر تک‌ژن‌های بومادران حاوی ریزماهواره با استفاده از بلاست یکس علیه پایگاه‌های داده KAAS، آرآیدوپسیس، پروتئین‌های Non-redundant پایگاه NCBI،

بیشترین تعداد تک‌ژن کارکردی حاوی SSR را نشان دادند که ۷۲ تک‌ژن از بین آن‌ها در سه مسیر ترپنوئید، فیل‌پروپانوئید و فلاونوئید دخیل بودند (۱۱). نتایج مشابهی در مورد تعداد تک‌ژن‌های حاوی EST-SRR در مسیر بیوستیزی متابولیت‌های ثانویه در ترانسکریپتوم برگ گل داودی (*Chrysanthemum morifolium*) گزارش شد که با نتایج حاصل از این تحقیق متفاوت بود (۴۶). نشانگرهای SSR دارای نقش کلیدی در تحمل به تنش، متابولیسم و تثبیت کربن، بیوستز متابولیت‌های ثانویه، رشد و فتوسنتز شناسایی گردیده‌اند (۷). تاکنون شناسایی و ایجاد SSR ها در گیاه کتان منجر به اصلاح کیفیت فیبری آن در طول چند سال اخیر گشته است (۱۹).

نتیجه‌گیری کلی

نتایج این پژوهش برای نخستین بار توسعه مقیاس گسترده نشانگرهای EST-SSR این گیاه را امکان‌پذیر ساخته است. این نشانگرها می‌توانند برای انجام تجزیه و تحلیل‌های نقشه‌های پیوستگی و ساخت نقشه‌های ژنتیکی با تراکم بالا در گیاه دارویی بومادران بکار روند. شناسایی ریزماهوره‌های موجود در توالی ژن‌های مرتبط به مسیر بیوستز متابولیت‌های ثانویه در رونوشت گیاه دارویی بومادران می‌تواند در شناسایی عوامل موثر تنظیم بیان ژن‌های ستر ترکیبات دارویی این گیاه بسیار موثر باشد.

بوده و همانندی بالایی با ژنوم آفتابگردان دارد. حدود ۴۵٪ (تعداد ۳۸۴۴۰ از ۸۵۴۲۱) از توالی‌های تک ژن حاوی تکرارهای ریزماهوره در دسته‌بندی کارکردی قرار گرفتند. براساس طبقه‌بندی تفسیر ژن‌شناسی اینترنتی (WEGO) ۵۲ دسته کارکردی برای همه تک‌ژن‌ها بدست آمد (شکل ۵). دستجات "سلول"، "بخش سلولی"، "اندامک"، "اتصال"، "کاتالیزوری"، "تنظیم زیستی"، "فرایند سلولی"، "فرایند متابولیکی"، "رنگدانه‌سازی" و "پاسخ به محرک‌ها" بیشترین درصد تخصیص را دارا بودند (شکل ۵).

در مجموع ۱۰۷۶۲ (۲۲/۳٪)، ۴۷۲۳ (۵۳/۳٪)، ۳۷۱۴ (۴۷/۵٪)، ۴۵۱۷ (۵۱/۳٪) و ۵۱۸۹ (۶۰/۷٪) تک‌ژن از همه کتابخانه‌ها به‌ترتیب به پایگاه‌های داده KAAS، آراییدوپسیس، UniProt، پروتئین‌های Non-redundant پایگاه NCBI و آفتابگردان گزارمان شدند (شکل ۴). تعداد تک‌ژن‌های تفسیر شده مشترک بین همه کتابخانه‌ها ۱۷۴۵ بود (شکل ۴).

در طبقه فرایندهای زیستی (BP) در میان همه تک‌ژن‌های دسته‌بندی شده ۳۲۲۰ تک‌ژن (۶۵ درصد) به دسته "فرایند متابولیکی" متعلق بود که از میان آن‌ها ۹۰ تک‌ژن (۳ درصد) به دسته "فرایند متابولیکی ثانویه" تخصیص داشت. به‌طور مشابه در ترانسکریپتوم نوعی ثعلب *Gastrodia elata* بیشترین تعداد تک‌ژن‌ها مربوط به فرایندهای سلولی (۴۷۹۴) و متابولیسم (۴۹۴۴) بود (۴۲). در مطالعه ترنسکریپتوم برگ گیاه *Sterculia lanceolata* مسیر بیوستزی متابولیت‌های ثانویه

منابع

1. Abedi, J., G. Mohammadinejad, A. Baghizadeh. 2019. Mapping of Loci Controlling Some of Agronomic Traits in Bread Wheat Under Drought Stress Conditions using Microsatellite Markers. *Journal of Crop Breeding*, 11(30): 160-7 (In Persian).
2. Amiripour, M., S. A. Sadat Noori, V. Shariati, and M. Soltani Howyze. 2019. Transcriptome analysis of Ajowan (*Trachyspermum ammi* L.) inflorescence. *Journal of Plant Biochemistry and Biotechnology*, 28: 496-508.
3. Angeloni, F., C.A.M. Wagemaker, M.S.M. Jetten, H.J.M. Op den Camp, E.M. Janssen-Megens, K.J. Francoijs, H.G. Stunnenberg and N.J. Ouborg. 2011. De novo transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next-generation sequencing techniques. *Molecular Ecology Resources*, 11(4): 662-674.
4. Bajgain, P., B.A. Richardson, J.C. Price, R.C. Cronn and J.A. Udall. 2011. Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *BMC genomics*, 12: 1-15.
5. Barman, M. and S. Kundu. 2019. Molecular markers and a new vista in plant breeding: a review. *Int J Curr Microbiol Applied Sciences*, 8(12): 1921-1929.
6. Beier S., T. Thiel, T. Münch, U. Scholz, M. Mascher. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33(16): 2583-5.
7. Bhandawat, A., G. Singh, A.S. Raina, J. Kaur and R.K. Sharma. 2016. Development of genic SSR marker resource from RNA-Seq data in *Dendrocalamus latiflorus*. *Journal of Plant Biochemistry and Biotechnology*, 25(2): 179-90.
8. Cavalcanti, A.M., C. H. Baggio, C. S. Freitas, L. Rieck, R. S. de Sousa, J. E. Da Silva-Santos, et al. 2006. Safety and antiulcer efficacy studies of *Achillea millefolium* L. after chronic treatment in Wistar rats. *Journal of ethnopharmacology*. 107(2):277-84.
9. Chen, F., D. Tholl, J. Bohlmann, E. Pichersky. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, 66(1): 212-29.
10. Dorafshan, M., M. Soltani Howyze and V. Shariati. 2019. Identification of terpenoid backbone biosynthetic pathway genes in fruit of *Citrullus colocynthis* (L.) Schrad. medical plant by RNA sequencing. *Iranian Journal of Medicinal and Aromatic Plants Research*, 23;35(4):691-702 (In Persian).

11. Eum, S.M., S.Y. Kim, J.S. Hong, N.S. Roy, S. Choi and J. Paik, et al. 2019. Transcriptome analysis and development of SSR markers of ethnobotanical plant *Sterculia lanceolata*. *Tree Genetics & Genomes*, 15(3): 37.
12. Gao, C., P. Xin, C. Cheng, Q. Tang, P. Chen and C. Wang, et al. 2014. Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. *PloS one*, 9(10): e110638.
13. Garcia-Seco, D., Y. Zhang, F.J. Gutierrez-Mañero, C. Martin and B. Ramos-Solano. 2015. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus* sp. Var. Lochness) fruit. *BMC Genomics*, 16(1): 1-12.
14. Garg, R., R.K. Patel, A.K. Tyagi and M. Jain. 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*, 18(1): 53-63.
15. Gershenzon, J., M.E. McConkey and R.B. Croteau. 2000. Regulation of monoterpene accumulation in leaves of peppermint. *Plant Physiology*, 122(1): 205-14.
16. Guo S., J. Liu, Y. Zheng, M. Huang, H. Zhang, G. Gong, et al. 2011. Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. *BMC genomics*, 12(1):454.
17. Han Z., X. Ma, M. Wei, T. Zhao, R. Zhan and W. Chen. 2018. SSR marker development and intraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis. *BMC genomics*. 19(1):291.
18. Hao D., P. Ma, J. Mu, S. Chen, P. Xiao and Y. Peng, et al. 2012. De novo characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. *Science China Life Sciences*, 55(5): 452-66.
19. Huang C., C. Shen, T. Wen, B. Gao, D. Zhu and X. Li, et al. 2018. SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population. *Molecular Genetics and Genomics*, 293(4): 793-805.
20. Kumpatla S.P. and S. Mukhopadhyay. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome*, 48(6): 985-98.
21. Li, Y.C., A.B. Korol, T. Fahima, E. Nevo. 2004. Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution*, 21(6): 991-1007.
22. Li, Y., C. Sun, H. Luo, X. Li, Y. Niu and S. Chen. 2010. Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX. *Yao xue xue bao= Acta pharmaceutica Sinica*, 45(4): 524-9.
23. Lin, X., J. Zhang, Y. Li, H. Luo, Q. Wu and C. Sun, et al. 2011. Functional genomics of a living fossil tree, *Ginkgo*, based on next-generation sequencing technology. *Physiologia plantarum*. 143(3):207-18.
24. Lu, F.H., M.C. Cho and Y.J. Park. 2012. Transcriptome profiling and molecular marker discovery in red pepper, *Capsicum annuum* L. TF68. *Molecular Biology Reports*, 39(3): 3327-35.
25. Luo, H., C. Sun, Y. Sun, Q. Wu, Y. Li and J. Song, et al. 2011. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC genomics*, 12(S5): S5.
26. Marakli, S. 2018. A Brief Review of Molecular Markers to Analyse Medically Important Plants. *International Journal of Life Sciences and Biotechnology*, 1(1): 29-36.
27. Moriya, Y., M. Itoh, S. Okuda, A.C. Yoshizawa and M. Kanehisa. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research*. 35(suppl_2): W182-W5.
28. Nishijima, R., K. Yoshida, Y. Motoi, K. Sato and S. Takumi. 2016. Genome-wide identification of novel genetic markers from RNA sequencing assembly of diverse *Aegilops tauschii* accessions. *Molecular Genetics and Genomics*, 291(4): 1681-94.
29. Parida, S.K., A. Pandit, K. Gaikwad, T.R. Sharma, P.S. Srivastava and N.K. Singh, et al. 2010. Functionally relevant microsatellites in sugarcane unigenes. *BMC plant biology*. 10(1):251.
30. Paulsen, E. 2002. Contact sensitization from Compositae-containing herbal remedies and cosmetics. *Contact Dermatitis*, 47(4): 189-98.
31. Piskol, R., G. Ramaswami and J.B. Li. 2013. Reliable identification of genomic variants from RNA-seq data. *The American Journal of Human Genetics*, 93(4): 641-51.
32. Powell, W., G.C. Machray, J. Provan. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1(7): 215-22.
33. Saeidnia, S., A. Gohari, N. Mokhber-Dezfuli and F. Kiuchi. 2011. A review on phytochemistry and medicinal properties of the genus *Achillea*. *DARU: Journal of Faculty of Pharmacy, Tehran University of Medical Sciences*, 19(3): 173.
34. Simon, S.A., J. Zhai, R.S. Nandety, K.P. McCormick, J. Zeng and D. Mejia, et al. 2009. Short-read sequencing technologies for transcriptional analyses. *Annual review of Plant Biology*, 60: 305-33.
35. Soltani, Howyzeh M., S. Aslani and O. Pooraskari. 2019. Essential Oil Profile of an Iranian Yarrow (*Achillea millefolium*). *Journal of Essential Oil Bearing Plants*, 22(1): 295-300. doi: 10.1080/0972060X.2019.1586589.
36. Soltani, Howyzeh M., S.A. Sadat Noori, V. Shariati and M. Amiripour. 2018. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal

- plant *Trachyspermum ammi* L. Scientific Reports. 8(1): 13405. doi: <https://dx.doi.org/10.1038%2Fs41598-018-31618-9>.
37. Soltani Howyzeh M., S. A. Sadat Noori, V. Shariati and M. Amiripour. 2019. Large Scale Identification of SSR Molecular Markers in Ajowan (*Trachyspermum ammi*) Using RNA Sequencing. Journal of Plant Genetic Research, 6(1): 31-46 (In Persian).
 38. Taheri Boukani, K. and R. Najafzadeh. 2019. Evaluation of Growth Characteristics, Essential Oil Percentage and Ecological Factors of Some Different Species of Yarrow (*Achillea* spp.). Journal of Crop Breeding, 11(31): 28-41 (In Persian).
 39. Taheri, S., T. Lee Abdullah, M. R. Yusop, M. M. Hanafi, M. Sahebi and P. Azizi, et al. 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules, 23(2): 399.
 40. Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan and M.J. Van Baren, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5): 511-5.
 41. Varshney, R.K., A. Graner and M.E. Sorrells. 2005. Genic microsatellite markers in plants: features and applications. TRENDS in Biotechnology, 23(1): 48-55.
 42. Wang, Y., M.Q. Shahid, F. Ghouri and F.S. Baloch. 2020. De Novo Assembly and Annotation of the Juvenile Tuber Transcriptome of a *Gastrodia elata* Hybrid by RNA Sequencing: Detection of SSR Markers. Biochemical Genetics, 58: 1-21.
 43. Wei, W., X. Qi, L. Wang, Y. Zhang, W. Hua and D. Li, et al. 2011. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC genomics, 12(1): 451.
 44. Xu, M., X. Liu, J.W. Wang, S.Y. Teng, J.Q. Shi and Y.Y. Li, et al. 2017. Transcriptome sequencing and development of novel genic SSR markers for *Dendrobium officinale*. Molecular Breeding, 37(2): 18.
 45. Yates, S.A., M.T. Swain, M.J. Hegarty, I. Chernukin, M. Lowe and G.G. Allison, et al. 2014. De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics, 15(1): 453.
 46. Zhao, X., L. Song, M. Ma, K. Liu, M. Lv and Y. Zhu, et al. 2019. RNA-Seq for excavation of genes involved in the biosynthesis of primary active components and identification of new EST-SSR markers in medicinal chrysanthemum. Archives of Biological Sciences, 71(3): 489-500.
 47. Zhou, Y., L. Chen, X. Fan and Y. Bian. 2014. De novo assembly of *Auricularia polytricha* transcriptome using Illumina sequencing for gene discovery and SSR marker identification. PloS one, 9(3): e91740.
 48. Zhu, Y., X. Wang, L. Huang, C. Lin, X. Zhang and W. Xu, et al. 2017. Transcriptomic identification of drought-related genes and SSR markers in sudan grass based on RNA-Seq. Frontiers in Plant Science. 8: 687.