

Identifying Microsatellite Molecular Markers Using Transcriptome Data Mining in the Medicinal Plant *Citrullus colocynthis* L.

Z.M.J. Alzubaidi¹ and M. Soltani Howyzeh^{2*}

¹ MSc Graduate, Department of Genetics, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

^{2*} Corresponding Author, Assistant Professor, Department of Genetics and Plant Breeding, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran Email: mehdisoltani@iau.ac.ir.

Received: 21.04.2024

Accepted: 14.08.2024

Abstract

Background and Objective

Next-generation sequencing (NGS) technology allows one to analyze microsatellite molecular markers associated with terpenoid pathway genes. Microsatellite molecular markers are a powerful tool for understanding genetic diversity. In this study, transcriptome sequencing of the medicinal plant *Citrullus colocynthis* was used for the first time to identify microsatellite markers.

Methodology

This research utilized transcriptome sequencing data from the fruit of *C. colocynthis*, consisting of 83,807 de novo assembled sequences. The assembled transcripts were analyzed to determine the frequency and distribution of microsatellites. Microsatellite markers were identified using Krait v1.5.1, a robust and ultrafast tool with a user-friendly graphical interface for genome-wide microsatellite screening. Specific primers for the identified microsatellite markers were designed and saved in FASTA format.

For functional annotation and gene similarity comparison, microsatellites from the *C. colocynthis* transcriptome were blasted against nucleotide sequences from five reference genomes: Uniprot, *Arabidopsis thaliana* (Atha), *Citrullus lanatus* (Clan), *Momordica charantia* (Mcha), and Watermelon Charleston Gray (WCG), using an E-value threshold of $<10^{-5}$. Additionally, unigene sequences containing microsatellites were uploaded to the KEGG Automatic Annotation Server (<http://www.Genome.Jp/kegg/kaas>) and analyzed using KEGG Orthology (KO) identifiers. KO-specific assignments were performed using the Single-Directional Best Hit (SBH) method to obtain metabolic pathway information for the predicted unigenes.

Results

Among the 83,807 assembled transcripts analyzed with Krait software, 20,221 potential microsatellites were identified in 15,506 transcripts. The most prevalent repeat types were mononucleotide microsatellites (8,313; 41.1%) and trinucleotide microsatellites (5,816; 28.3%), while hexanucleotide (531; 2.63%) and pentanucleotide (480; 2.37%) repeats were the least frequent. The frequency of microsatellite repeats in the *C. colocynthis* transcriptome was approximately one per 4.14 kilobases of assembled sequence, with microsatellites containing five repeats (3,041 unigenes) being the most common. The distribution of microsatellite repeat length (10–50 bp) showed

that 57.55% were classified as second-class (≤ 20 nucleotides), while 42.45% were first-class (> 20 nucleotides). The most dominant sequence was mononucleotide "A" repeat, with 8,167 occurrences (40.39%). Among the microsatellites, the trinucleotide AAG (2,757 occurrences) and the dinucleotide repeats AG (2,298) and AT (1,178) were the most frequent. In total 3,065 primer pairs were designed from the 20,221 identified microsatellites. Genetic similarity analysis using X-blast revealed 360 shared microsatellites between *C. colocynthis* and WCG. Additionally, 32 genes involved in the terpenoid biosynthesis pathway were identified, of which 18 transcripts contained microsatellites.

Conclusion

Microsatellite markers play a crucial role in the genome, particularly in genes associated with secondary metabolite biosynthesis in medicinal plants. This study provides valuable insights for improving secondary metabolite production, developing molecular markers, assessing genetic diversity, and constructing genetic maps for *C. colocynthis*. The findings of this research mark a significant step forward in genetic studies of this medicinal plant.

Keywords: Cucurbitaceae, Medicinal plant, Molecular markers, Genetic map, Primer design.

شناسایی نشانگرهای مولکولی ریزماهواره حاصل از داده‌کاوی ترانسکریپتوم گیاه دارویی هندوانه ابوجهل

(Citrullus colocynthis)

زمن مطشر جبرت الله الزبیدی^۱ و مهدی سلطانی حویزه^{۲*}

۱- دانش‌آموخته کارشناسی ارشد، گروه ژنتیک، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران

۲- * نویسنده مسئول، استادیار، گروه ژنتیک و به‌نژادی گیاهی، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران

پست الکترونیک: mehdisoltani@iau.ac.ir

تاریخ دریافت: ۱۴۰۳/۰۲/۰۲ تاریخ پذیرش: ۱۴۰۳/۰۵/۲۴

چکیده مبسوط

سابقه و هدف:

تکنولوژی توالی‌یابی نسل جدید (NGS) فرصتی را برای تجزیه و تحلیل نشانگرهای مولکولی ریزماهواره که به ژن‌های مسیرهای بیوشیمیایی مرتبط است، ارائه می‌دهد. نشانگرهای مولکولی ریزماهواره روشی قدرتمند برای درک تنوع ژنتیکی هستند. در این مطالعه، از توالی‌یابی ترانسکریپتوم گیاه دارویی هندوانه ابوجهل (*Citrullus colocynthis*) برای شناسایی نشانگرهای ریزماهواره استفاده شد.

مواد و روش:

در این پژوهش از داده‌های مطالعه توالی‌یابی ترانسکریپتوم گیاه هندوانه ابوجهل که تعداد ۸۳۸۰۷ توالی نوپدید ایجاد کرده بود، استفاده شد. مجموعه رونوشت‌های هم‌گذاری شده گیاه هندوانه ابوجهل برای تعیین فراوانی و توزیع ریزماهواره‌ها بررسی شدند. نشانگرهای ریزماهواره با استفاده از یک ابزار شناسایی قوی و فوق سریع با رابط گرافیکی کاربرپسند برای بررسی گسترده ژنوم ریزماهواره‌ها، به نام Krait v1.5.1 جست‌وجو شد. همچنین آغازگرهای اختصاصی نشانگرهای ریزماهواره شناسایی شده توسط نرم‌افزار Krait طراحی گردید و به صورت فایل Fasta ذخیره شد. برای تفسیر کارکردی بلاست ایکس (BLAST X) و مقایسه شباهت ژنی توسط نرم‌افزار Venny، توالی‌های دارای ریزماهواره ترانسکریپتوم هندوانه ابوجهل علیه توالی‌های نوکلئوتیدی پنج ژنوم مرجع شامل یونی‌پرات (Uniprot)، آرابیدوپسیس *Arabidopsis thaliana* Atha، هندوانه گرد *Citrullus lanatus* subsp.97103 Clan، خیار چنبر تلخ *Momordica charantia* Mcha. و هندوانه چارلستون گری Watermelon Charleston Gray (WCG) با مقادیر خطای کمتر از یکصد هزارم بلاست شدند. در تفسیر کارکردی دیگر توالی تک‌ژن‌های حاوی ریزماهواره در سرور تفسیر اتوماتیک (KEGG) (<http://www.Genome.Jp/kegg/kaas>) بارگذاری شده و با استفاده از شناسه‌های اختصاصی KEGG Orthology (KO) تفسیر شدند. روش اجرایی شناسه‌های اختصاصی KO براساس روش تک راهنمای شناسایی بهتر (Single-directional

(Best Hit) انجام شد. با توجه به تخصیص KO، اطلاعات مسیرهای متابولیکی مربوط به تک‌ژن‌های پیش‌بینی شده حاصل گردید.

نتایج:

از تعداد ۸۳۸۰۷ رونوشت هم‌گذاری شده با استفاده از نرم‌افزار Krait، ۲۰۲۲۱ ریزماهوره بالقوه یافت شد که در ۱۵۵۰۶ رونوشت هم‌گذاری شده قرار داشتند. ریزماهوره‌های مونونوکلئوتیدی (۸۳۱۳ عدد، ۴۱/۱ درصد) و تری‌نوکلئوتیدی (۵۷۱۶ عدد، ۲۸/۳ درصد) به ترتیب نوع غالب پرتکرار و هگزانوکلئوتیدی (۵۳۱ عدد، ۲/۶۳ درصد) و پنتانوکلئوتیدی (۴۸۰ عدد، ۲/۳۷ درصد) به ترتیب دارای کمترین تکرار بودند. فراوانی تکرارهای ریزماهوره در ترانسکریپتوم هندوانه ابوجهل تقریباً یک در هر ۴/۱۴ کیلوباز توالی هم‌گذاری شده بود و تکرارهای پنج‌تایی ریزماهوره‌ها (۳۰۴۱ عدد) فراوان‌ترین تکرارها بودند. توزیع فراوانی طولی (۵۰-۱۰ جفت باز) تکرارهای ریزماهوره هندوانه ابوجهل، ۵۷/۵۵ درصد از نوع کلاس دوم (طول کمتر از ۲۰ نوکلئوتید) و ۴۲/۴۵ درصد از نوع کلاس اول (طول بیش از ۲۰ نوکلئوتید) بودند. توالی غالب، مونونوکلئوتید A با ۸۱۶۷ تکرار (۴۰/۳۹ درصد) بود. در میان ریزماهوره تری‌نوکلئوتید AAG (۲۷۵۷ تکرار) و در میان دی‌نوکلئوتیدهای AG (۲۲۹۸ تکرار) و AT (۱۱۷۸ تکرار) بیشترین فراوانی را به خود اختصاص دادند. برای طراحی آغازگرهای اختصاصی از ۲۰۲۲۱ نشانگر ریزماهوره در مجموع ۳۰۶۵ جفت آغازگر سنتز شد. با تفسیر بلاست ایکس رونوشت‌های هم‌گذاری شده حاوی ریزماهوره‌های هندوانه ابوجهل، شباهت ژنتیکی بیشتر با هندوانه چارلستون گری (WCG) (۳۶۰ ریزماهوره مشترک) مشخص شد. در این مطالعه ۳۲ ژن در مسیر پایه بیوسنتز ترینوئیدها شناسایی شدند که ۱۸ رونوشت مربوط به آنها دارای ریزماهوره بود.

نتیجه‌گیری:

با توجه به اینکه نشانگرهای ریزماهوره دارای نقش کلیدی در ژنوم و مرتبط با ژن‌ها به‌ویژه در بیوسنتز متابولیت‌های ثانویه در گیاهان دارویی هستند، از این‌رو از توالی‌یابی نسل جدید، می‌توان برای بهبود متابولیت‌های ثانویه و توسعه نشانگر، تنوع ژنتیکی و ساخت نقشه‌های ژنتیکی در گیاه هندوانه ابوجهل استفاده کرد و نتایج این تحقیق می‌تواند گامی اساسی در این مسیر بردارد.

واژه‌های کلیدی: تیره کدوئیان، طراحی آغازگر، گیاه دارویی، نشانگر مولکولی، نقشه ژنتیکی.

مقدمه

یافته‌اند و از میان آنها مکان ریزماهوره (Microsatellite loci) یا توالی‌های تکراری کوتاه (Simple Sequence Repeats/SSRs) ترجیح داده می‌شود، زیرا آنها ساده‌تر، تکرارپذیر، فراوان، بسیار متغیر، مؤثر و همباز هستند (Tautz, 1989). ریزماهوره‌ها بن‌مایه‌های ساده‌ای از ۱ تا ۶

نشانگرهای مولکولی ابزار قدرتمند برای مطالعه ترانسکریپت‌ها و ژنوم‌های گیاهی و تغییرات ژنتیکی مربوط به فنوتیپ‌های خاص هستند (Appleby *et al.*, 2009). تعدادی از نشانگرهای مولکولی در طول سال‌ها توسعه

مطالعات طبقه‌بندی، توسعه مولکولی، ژنتیک، ژنومیک و متابولیسم ثانویه در گیاهان دارویی هستند.

گیاه دارویی هندوانه ابوجهل با نام علمی *Citrullus colocynthis*، به عنوان یک گیاه چندساله وحشی در مناطق خشک در سراسر جهان مانند شمال آفریقا و خاورمیانه رشد می‌کند، میوه آن معمولاً با نام‌های کولوسینت، خیار تلخ، خربزه تلخ و تاک سدوم تلخ شناخته می‌شود. بخش‌های مختلف این گیاه به عنوان درمان بیماری‌هایی مانند زردی، دیابت و آسم عمل می‌کند (Li et al., 2022). هندوانه ابوجهل یک منبع غنی از ترکیبات فعال زیستی مانند اسانس‌ها، گلیکوزیدها، فلاونوئیدها، ساپونین‌ها، آلکالوئیدها، کولوسینتین، کوکوربیتاسین‌ها و اسیدهای چرب است که باعث تقویت سیستم ایمنی بدن می‌شوند (Gomes-Carneiro et al., 1998). براساس مطالعات، کوکوربیتاسین‌ها دسته‌ای از تری‌ترپنوئیدهای تتراسایکلیک نوعی کوکوربیتان هستند که عمدتاً توسط گیاهان خانواده کدوئیان به‌ویژه در هندوانه ابوجهل تولید می‌شوند. کوکوربیتاسین‌ها تلخ‌مزه بوده و دارای خواص ضددیابت، ضدسرطان، ضدالتهاب و ضدقارچ هستند (Chen et al., 2005; Kaushik et al., 2015).

در مطالعات قبلی ترانسکریپتوم بافت‌های مختلف گیاه دارویی هندوانه ابوجهل بسیاری از اطلاعات ژنتیکی مسیرهای بیوسنتزی متابولیت‌های ثانویه از جمله اسکلت تریپنوئید، تری‌ترپنوئید، سزکویی‌تریپنوئید و کاروتنوئید شناسایی شده‌اند (Dorafshan et al., 2019b; Dorafshan et al., 2020, 2021). در مطالعات اخیر در مجموع ۸۳۸۰۷ رونوشت از کتابخانه‌های cDNA گیاه هندوانه ابوجهل شناسایی شده است (Dorafshan et al., 2020). تا کنون چندین ژن مربوط به تحمل به خشکی در هندوانه ابوجهل شناسایی شده‌اند که از میان ۲۰۵۸۱ رونوشت بررسی شده، ۲۵۴۵ ژن مرتبط با خشکی بود (Wang et al., 2014). مطالعات پیشین انجام شده برای بررسی تنوع ژنتیکی ژنوم و ژرم‌پلاسم گیاه هندوانه ابوجهل، تنها به نشانگرهای RAPD (Levi et al., 2002; Verma et al., 2013) AFLP (Nimmakayala et al., 2010; Si et al., 2009) ISSR

نوکلئوتید هستند که بیش از دو تا چند ده بار در یک مکان تکرار شده و در همه جا در ژنوم یک گونه توزیع شده‌اند و دارای اطلاعات فراوان، کدگذاری و خصوصیات loci هستند (Hancock, 1996). در برخی مطالعات تأثیر ریزوماهواره‌ها را بر بیان ژن دخیل دانسته‌اند و چندشکلی حاصل از تکنیک‌های ریزوماهواره را به‌عنوان رویداد کلیدی در تنظیم تکامل ژنتیکی گزارش شده‌اند (Jarvis et al., 2008). ریزوماهواره‌ها به عنوان یک نشانگر مهم برای تعیین تغییرات ژنتیکی عملکردی، از جمله تعیین والد، مطالعات ژنتیک جمعیت، ارزیابی تنوع ژنتیکی و برای توسعه نقشه ژنتیکی عمل می‌کنند (Senthilvel et al., 2008). ریزوماهواره‌ها نشانگرهای ارزشمندی برای نقشه‌برداری ژنوم و مطالعات جمعیت هستند (Ma et al., 2020; Wang et al., 2020). در سال‌های اخیر، فناوری‌های توالی‌یابی با توان بالا، فرصت خوبی را برای توسعه هزاران ریزوماهواره پیشنهادی در یک زمان فراهم کرده‌اند. پیشرفت‌های بزرگی توسط ریزوماهواره‌ها در اصلاح بسیاری از گیاهان حاصل شده است (Roshan et al., 2021; Song et al., 2021). تکنیک Next Generation Sequencing (NGS) تعداد زیادی از ریزوماهواره‌ها را با هزینه و تلاش کمتر در مقایسه با روش‌های مرسوم پرهزینه و دشوار شناسایی می‌کند. مزیت اصلی توسعه ریزوماهواره‌های حاصل از توالی‌های ترانسکریپتومی NGS، امکان یافتن ارتباط با ژن‌های عملکردی و افزایش ارتباط با فنوتیپ‌هاست. ریزوماهواره‌ها درون منطقه رمزگذاری ژن‌ها ممکن است در واقع بیان و عملکرد ژن را تنظیم کنند و آنها را یک منبع ارزشمند برای مطالعات ژنتیکی و کاربردهای اصلاحی می‌دانند (Dutta et al., 2011; Zalapa et al., 2012). استخراج ریزوماهواره، به‌ویژه برای گیاهان بدون ژنوم مرجع (هم‌گذاری شده نوپدید) قابل استفاده است. در موجودات غیر مدل بدون ژنوم مرجع، مقدار زیادی از داده‌های توالی بیان شده را می‌توان با استفاده از تکنولوژی RNA-Seq به‌دست آورد (Li et al., 2014; Wei et al., 2016). ریزوماهواره‌ها منابع قدرتمند برای بیشتر

(<https://krait.biosv.com/en/latest/>) شناسایی شدند. Krait یک ابزار قوی و فوق سریع با رابط گرافیکی کاربرپسند برای بررسی گسترده ژنوم ریزماهوره‌ها و طراحی پرایمر ریزماهوره است (Thakur and Randhawa, 2018). پارامترهای جست‌وجو برای حداقل تکرار برای هر نوع ریزماهوره کامل، برای نشانگرهای مونوکلوئوتیدی "۱۲"، دی‌نوکلئوتیدی "۷"، تری‌نوکلئوتیدی "۵"، تترا، پنتا و هگزانوکلئوتیدی "۴" قرار داده شد و سطح استانداردسازی موتیف روی سطح ۳ تنظیم گردید. این ابزار موتیف‌های ریزماهوره، تعداد تکرارها و مختصات توالی برای هر ریزماهوره را گزارش کرد.

طراحی آغازگر برای نشانگرهای ریزماهوره:

آغازگرهای اختصاصی نشانگرهای ریزماهوره شناسایی شده توسط نرم‌افزار Krait v1.5.1 طراحی گردید و به صورت فایل Fasta ذخیره شد. فایل آغازگرهای ریزماهوره شامل مشخصات فیزیکی آغازگرهای مورد بررسی از جمله آیدی ژن، موتیف استاندارد، نوع موتیف غالب، نوع نوکلئوتید، تکرار الیگونوکلئوتید، طول الیگونوکلئوتید، محصول توالی آغازگر و دمای اتصال، محتوای G/C و پایداری آغازگرهای رفت و برگشت می‌باشد.

تفسیر کارکردی رونوشت‌های دارای ریزماهوره:

توالی‌های دارای ریزماهوره ترانسکریپتوم هندوانه ابوجهل علیه توالی‌های نوکلئوتیدی پنج ژنوم مرجع یونی پرات (Uniprot)، آرآی‌دوپیسیس *Arabidopsis thaliana* Atha، خیارچنبر هندوانه گرد *Citrullus lanatus* Clan، تلخ‌هندوانه چارلستون *Momordica charantia* Mcha و گری *Watermelon Charleston Gray* (WCG) با مقادیر خطای (E-Value) کمتر از یکصد هزارم با استفاده از یک سرور محلی (Local Server) با حافظه ۶۴ گیگابایت و CPU با ۲۴ هسته و سیستم عامل لینوکس، بلاست ایکس شدند. برای مقایسه شباهت ریزماهوره‌ای گیاهان مذکور، نرم‌افزار

(Badr *et al.*, 2018; Hassan *et al.*, 2012) و EST-SSR (Hwang *et al.*, 2011) تحت تنش خشکی محدود بوده است.

در طی سال‌های اخیر توالی‌های ترانسکریپتوم و نشانگرهای ریزماهوره در بسیاری از گیاهان دارویی، از جمله زینان (*Trachyspermum ammi*) (Soltani Howyzeh *et al.*, 2019)، زنجبیل (*Zingiber officinale* Rosc.) (Zingiber *et al.*, 2021)، زعفران (*Crocus sativus*) (Crocus *et al.*, 2022) و بومادران (*Acillea millefolium*) (Acillea *et al.*, 2023) بررسی شده است. با توجه به اینکه نشانگرهای ریزماهوره دارای نقش کلیدی در ژنوم و مرتبط با ژن‌ها به‌ویژه در بیوسنتز متابولیت‌های ثانویه در گیاهان دارویی هستند، از این رو در این مطالعه، از توالی‌یابی ترانسکریپتوم گیاه هندوانه ابوجهل برای شناسایی نشانگرهای ریزماهوره استفاده شد. بنابراین می‌توان از ریزماهوره‌های موجود در توالی‌های ژنومی برای سازماندهی ژنومی، نوترکیبی، تنظیم بیان ژن، تنوع ژنتیکی کمی و تکامل ژن‌ها استفاده کرد.

مواد و روش‌ها

مشخصات داده‌های ترانسکریپتوم هندوانه ابوجهل:

برای تهیه توالی‌های ترانسکریپتوم هندوانه ابوجهل از فایل توالی‌های حاصل از مطالعه ترانسکریپتوم این گیاه در مطالعه قبلی که شامل ۸۳۸۰۷ توالی نوپدید بود استفاده شد (Dorafshan *et al.*, 2020). این داده‌ها از نمونه‌های میوه از گیاه هندوانه ابوجهل به صورت تصادفی از منطقه اندیمشک در شمال استان خوزستان در سال ۱۳۹۶ جمع‌آوری شده بود، تهیه شدند. همچنین توالی‌یابی RNA کل از نمونه‌های میوه هندوانه ابوجهل در مؤسسه ژنومی چین (Beijing Genomics Institute BGI) واقع در شهر پکن انجام شده است.

شناسایی نشانگرهای ریزماهوره:

نشانگرهای ریزماهوره با استفاده از ابزار شناسایی ریزماهوره به نام Krait v1.5.1

نتایج

گزارش آماری مجموعه رونوشت‌های هم‌گذاری شده گیاه هندوانه ابوجهل برای تعیین فراوانی و توزیع ریزماهورها در جدول ۱ ذکر شده است. این گزارش به‌طور خودکار توسط نرم‌افزار Krait v1.5.1 تهیه شده است. با استفاده از این نرم‌افزار، از ۸۳۸۰۷ رونوشت هم‌گذاری شده ۲۰۲۲۱ ریزماهوره بالقوه یافت شد (جدول ۱). ابزار Krait موتیف‌های ریزماهوره، تعداد تکرارها و مختصات توالی برای هر ریزماهوره را گزارش کرد (جدول ۲). ریزماهوره مونونوکلوئوتید (۸۳۱۳، ۴۱/۱ درصد) فراوان‌ترین ریزماهوره (نوع غالب تکرار) در مجموعه داده این پژوهش بود. به دنبال مونونوکلوئوتید به ترتیب ریزماهوره‌های تری‌نوکلئوتید (۵۸۱۶، ۲۸/۳ درصد)، دی‌نوکلئوتید (۳۸۲۰، ۱۸/۹ درصد)، تترانوکلئوتید (۱۳۶۱، ۶/۷۳ درصد)، هگزانوکلئوتید (۵۳۱، ۲/۶۳ درصد) و پنتانوکلئوتید (۴۸۰، ۲/۳۷ درصد) بیشترین فراوانی را داشتند (شکل ۱). در مجموعه داده توزیع طولی انواع ریزماهوره، نوع مونونوکلوئوتید (۱۶۲۳۹، ۴۰/۶ درصد) فراوان‌ترین ریزماهوره بود. به دنبال آن، به ترتیب ریزماهوره‌های تری‌نوکلئوتید (۱۱۲۴۶۱، ۲۸/۱ درصد)، دی‌نوکلئوتید (۷۶۰۸۰، ۱۹ درصد)، تترانوکلئوتید (۲۴۶۸۸، ۶/۱۷ درصد)، هگزانوکلئوتید (۱۳۷۹۴، ۳/۴۵ درصد) و پنتانوکلئوتید (۱۰۶۵، ۲/۶۵ درصد) بودند (شکل ۲).

Venny (<https://www.biotoools.fr/misc/venny>) مورد استفاده قرار گرفت و به‌صورت نمودار Venn ترسیم شد.

تفسیر کارکردی نقشه‌های مسیر متابولیتی تک‌ژن‌ها توسط

KEGG:

توالی تک‌ژن‌های حاوی ریزماهوره در سرور تفسیر اتوماتیک (KEGG <http://www.Genome.Jp/kegg/kaas>) بارگذاری شده و با استفاده از شناسه‌های اختصاصی (KO) KEGG Orthology تفسیر شدند. روش اجرایی شناسه‌های اختصاصی KO براساس روش تک راهنمای شناسایی بهتر (Single-directional Best Hit) انجام شد. با توجه به تخصیص KO، اطلاعات مسیرهای متابولیکی شامل آنزیم‌های درگیر به‌مراه کد اختصاصی آنزیم (Enzyme code) مربوط به تک‌ژن‌های پیش‌بینی شده حاصل گردید. شناسه‌های KO مجموعه‌ای از ارتولوگ‌های ژنی لینک شده رکوردی در نقشه‌های مسیر KEGG با هدف پیوند ژنوم با مسیرهای زیستی هستند. نقشه‌های مسیر KEGG به‌صورت گسترده برای تفسیر بیولوژیکی توالی ژنوم و سایر داده‌های با توان بالا مانند ترانسکریپتوم استفاده می‌شود. پیوند ژنوم‌ها به مسیرها از طریق سیستم KO، تعریف و به‌صورت دستی که توسط شماره‌های K تعیین شده، ساخته می‌شود. برای بهینه‌سازی این فرایند، تفسیر ماژول‌های KEGG تعریف شده از عناصر K توسعه داده شده است (Kanehisa *et al.*, 2013).

جدول ۱- گزیده‌ای از اطلاعات نشانگرهای ریزماهوره شناسایی شده در هندوانه ابوجهل

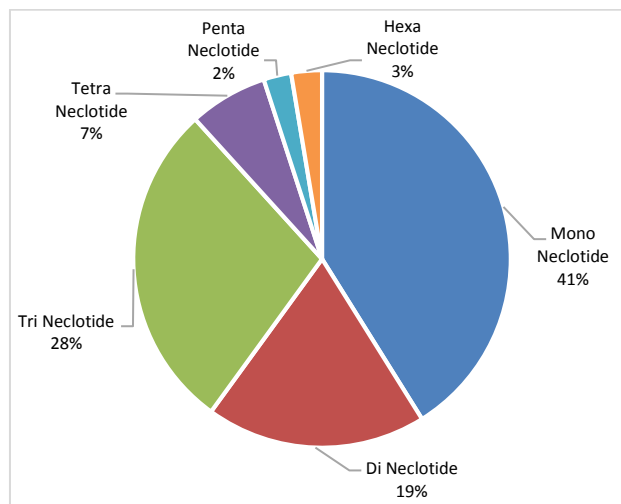
Table 1. A selection of information on microsatellite markers identified of *Citrullus colocynthis*

Item	Description	Number
Total number of perfect SSRs	Counts	20221
Total length of perfect SSRs	Bp	400025
The average length of SSRs	Total SSR length/total SSR counts (bp)	19.79
The percentage of sequence covered by SSRs	Total SSR Length/total sequence length (%)	0.46
Relative abundance	Total SSRs/total valid length (loci/Mb)	231.28
Relative density	Total SSR length/total valid length (bp/Mb)	4575.24

جدول ۲- خلاصه انواع ریزماهوره‌های کامل در هندوانه ابو جهل

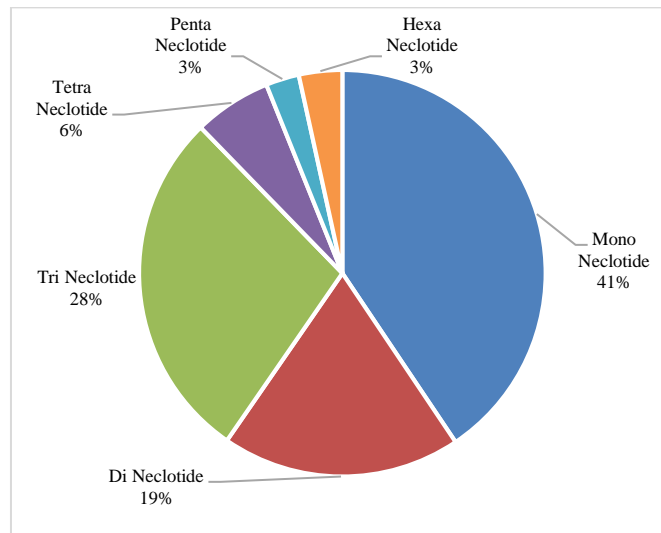
Table 2. The summary of perfect microsatellite types of *Citrullus colocynthis*

Type	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
Mono	8313	162397	41.11	19.54	95.08	1857.4
Di	3820	76080	18.89	19.92	43.69	870.16
Tri	5716	112461	28.27	19.67	65.38	1286.26
Tetra	1361	24688	6.73	18.14	15.57	282.37
Penta	480	10605	2.37	22.09	5.49	121.29
Hexa	531	13794	2.63	25.98	6.07	157.77



شکل ۱- توزیع شمارشی انواع ریزماهوره‌های شناسایی شده در ترانسکریپتوم هندوانه ابو جهل

Figure 1. SSR count distribution for each type identified in the transcriptome of *Citrullus colocynthis*



شکل ۲- توزیع طولی انواع ریزماهوره‌های شناسایی شده در ترانسکریپتوم هندوانه ابو جهل

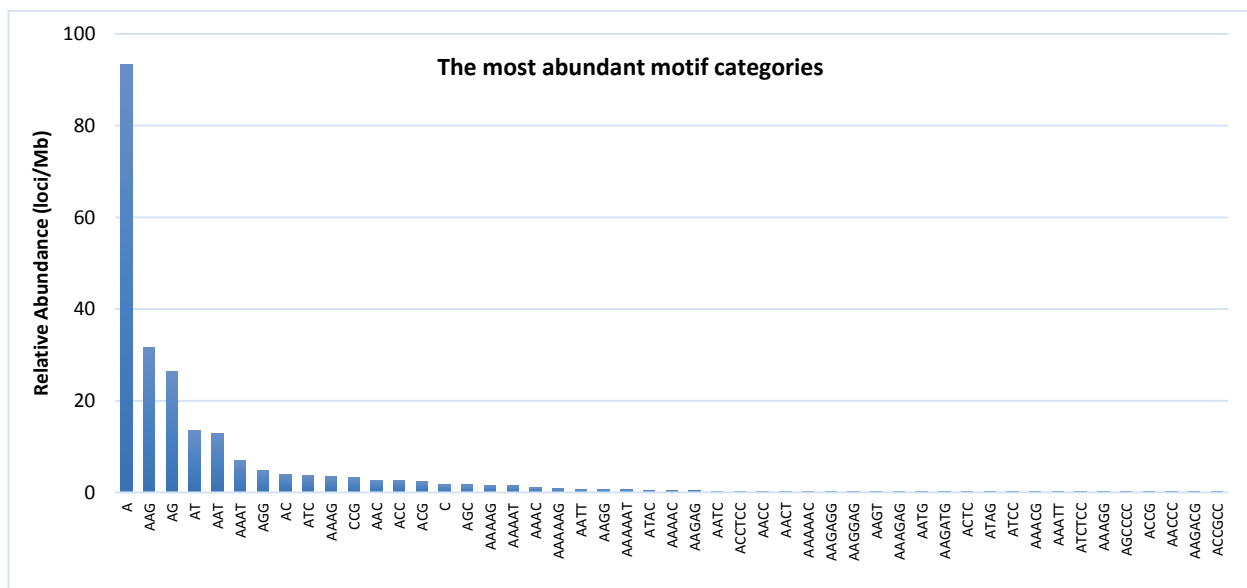
Figure 2. SSR length distribution for each type identified in transcriptome of *Citrullus colocynthis*

فراوانی هر دسته موتیف استاندارد در ترانسکریپتوم:

موتیف‌های استاندارد شناسایی شده به ۱۸۴ نوع تقسیم شدند که توزیع غیرتصادفی ریزماهواره‌ها را در ترانسکریپتوم هندوانه ابوجهل نشان دادند (جدول ۳). فراوان‌ترین ریزماهواره مونونوکلئوتیدی بدست آمده در ترانسکریپتوم هندوانه ابوجهل A با ۸۱۶۷ تکرار (۴۰/۳۹ درصد) بود. در میان ریزماهواره تری‌نوکلئوتید به ترتیب AAG (۲۷۵۷ تکرار) و AAT (۱۱۳۰ تکرار) و در میان دی‌نوکلئوتید AG (۲۲۹۸ تکرار) و AT (۱۱۷۸ تکرار) بیشترین فراوانی را به خود اختصاص دادند (شکل ۳).

فراوانی هر تکرار ریزماهواره در ترانسکریپتوم:

تکرارها به ۹۳ نوع تقسیم شدند که توزیع غیر تصادفی ریزماهواره‌ها را در ترانسکریپتوم هندوانه ابوجهل نشان داد (جدول ۴). فراوانی تکرارهای ریزماهواره در ترانسکریپتوم هندوانه ابوجهل تقریباً یک در هر ۴/۱۴ کیلوباز توالی هم‌گذاری شده بود. در مجموع ریزماهواره‌های با تکرارهای ۵ تا ۳۰۴۱ فراوان‌ترین بودند و پس از آن ریزماهواره‌های با ۷ (۱۷۸۸)، ۱۲ (۱۶۶۱) و ۴ (۱۶۴۴) تکرار فراوان بودند



شکل ۳- فراوان‌ترین دسته‌های موتیف ریزماهواره‌ها در ترانسکریپتوم هندوانه ابوجهل

Figure 3. The most abundant motif categories of microsatellites in the transcriptome of *C. colocynthis*

جدول ۳- فراوانی هر دسته موتیف استاندارد در ترانسکریپتوم هندوانه ابوجهل (تنها فراوانی‌های بیش از ۱۰ نمایش داده شده‌اند)

Table 3. The abundance of each standard motif category in transcriptome of *C. colocythis* (only counts more than 10 are presented).

Motif	Type	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
A	1	8167	160165	40.39	19.61	93.41	1831.87
AAG	3	2757	58389	13.63	21.18	31.53	667.82
AG	2	2298	47984	11.36	20.88	26.28	548.81
AT	2	1178	22186	5.83	18.83	13.47	253.75
AAT	3	1130	22290	5.59	19.73	12.92	254.94
AAAT	4	610	10852	3.02	17.79	6.98	124.12
AGG	3	414	7554	2.05	18.25	4.74	86.4
AC	2	344	5910	1.7	17.18	3.93	67.59
ATC	3	324	5727	1.6	17.68	3.71	65.5
AAAG	4	310	6016	1.53	19.41	3.55	68.81
CCG	3	276	4587	1.36	16.62	3.16	52.46
AAC	3	237	3846	1.17	16.23	2.71	43.99
ACC	3	229	4014	1.13	17.53	2.62	45.91
ACG	3	206	3582	1.02	17.39	2.36	40.97
C	1	146	2232	0.72	15.29	1.67	25.53
AGC	3	143	2472	0.71	17.29	1.64	28.27
AAAAG	5	133	3010	0.66	22.63	1.52	34.43
AAAAT	5	129	2835	0.64	21.98	1.48	32.42
AAAC	4	94	1584	0.46	16.85	1.08	18.12
AAAAAG	6	79	2034	0.39	25.75	0.9	23.26
AATT	4	59	1016	0.29	17.22	0.67	11.62
AAGG	4	57	1128	0.28	19.79	0.65	12.9
AAAAAT	6	50	1266	0.25	25.32	0.57	14.48
ATAC	4	35	600	0.17	17.14	0.4	6.86
AAAAC	5	35	725	0.17	20.71	0.4	8.29
AAGAG	5	28	660	0.14	23.57	0.32	7.55
AATC	4	24	416	0.12	17.33	0.27	4.76
ACCTCC	6	24	660	0.12	27.5	0.27	7.55
AACC	4	22	392	0.11	17.82	0.25	4.48
AACT	4	20	344	0.1	17.2	0.23	3.93
AAAAAC	6	20	486	0.1	24.3	0.23	5.56
AAGAGG	6	20	528	0.1	26.4	0.23	6.04
AAGGAG	6	19	522	0.09	27.47	0.22	5.97
AGT	4	18	324	0.09	18	0.21	3.71
AAAGAG	6	18	510	0.09	28.33	0.21	5.83
AATG	4	14	244	0.07	17.43	0.16	2.79
AAGATG	6	14	378	0.07	27	0.16	4.32
ACTC	4	12	244	0.06	20.33	0.14	2.79
ATAG	4	12	224	0.06	18.67	0.14	2.56
ATCC	4	12	196	0.06	16.33	0.14	2.24
AAACG	5	12	250	0.06	20.83	0.14	2.86
AAATT	5	11	225	0.05	20.45	0.13	2.57
ATCTCC	6	11	324	0.05	29.45	0.13	3.71

جدول ۴- فراوانی هر تکرار ریزماهوره در ترانسکرپتوم هندوانه ابوجهل

Table 4. The abundance of each repeat in the transcriptome of the *C. colocynthis*

Repeat	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
4	1644	31012	8.13	18.86	18.8	354.7
5	3041	49480	15.04	16.27	34.78	565.92
6	1453	27360	7.19	18.83	16.62	312.93
7	1788	30884	8.84	17.27	20.45	353.23
8	1141	21472	5.64	18.82	13.05	245.58
9	795	16749	3.93	21.07	9.09	191.56
10	573	12930	2.83	22.57	6.55	147.89
11	349	8844	1.73	25.34	3.99	101.15
12	1661	24804	8.21	14.93	19	283.69
13	1173	18382	5.8	15.67	13.42	210.24
14	956	16044	4.73	16.78	10.93	183.5
15	839	14565	4.15	17.36	9.6	166.59
16	724	13264	3.58	18.32	8.28	151.71
17	686	13107	3.39	19.11	7.85	149.91
18	589	11952	2.91	20.29	6.74	136.7
19	424	8930	2.1	21.06	4.85	102.14
20	344	7620	1.7	22.15	3.93	87.15
21	220	5250	1.09	23.86	2.52	60.05
22	165	4114	0.82	24.93	1.89	47.05
23	208	4991	1.03	24	2.38	57.08
24	162	4344	0.8	26.81	1.85	49.68
25	117	3025	0.58	25.85	1.34	34.6
26	113	3120	0.56	27.61	1.29	35.68
27	70	2079	0.35	29.7	0.8	23.78
28	60	1680	0.3	28	0.69	19.21
29	47	1392	0.23	29.62	0.54	15.92
30	82	2610	0.41	31.83	0.94	29.85
31	42	1302	0.21	31	0.48	14.89
32	37	1216	0.18	32.86	0.42	13.91
33	22	792	0.11	36	0.25	9.06
34	36	1224	0.18	34	0.41	14
35	30	1050	0.15	35	0.34	12.01
36	23	828	0.11	36	0.26	9.47
37	37	1369	0.18	37	0.42	15.66
38	23	950	0.11	41.3	0.26	10.87
39	34	1326	0.17	39	0.39	15.17
40	17	680	0.08	40	0.19	7.78

مونونوکلویتید (۹۸۸) و تکرار ۸ (۱۱۴۱) و نوع غالب دی‌نوکلئوتید (۱۷۶۳).

توزیع طول برای هر نوع ریزماهوره:

توزیع طول برای هر نوع ریزماهوره در شکل ۵ نشان داده شده است. براساس این نمودار بیشترین تعداد ریزماهوره‌ها برای مونونوکلویتیدها دارای طول ۱۲، ۱۳ و ۱۴ (به ترتیب با تعداد ۱۳۷۱، ۹۸۸ و ۸۰۴)، دی‌نوکلئوتید دارای طول ۱۴، ۱۶ و ۱۸ (به ترتیب با تعداد ۱۰۲۰، ۷۶۳ و ۵۴۱)، تری‌نوکلئوتید دارای طول ۱۵، ۱۸ و ۲۱ (به ترتیب با تعداد ۲۵۳۴، ۱۳۱۰

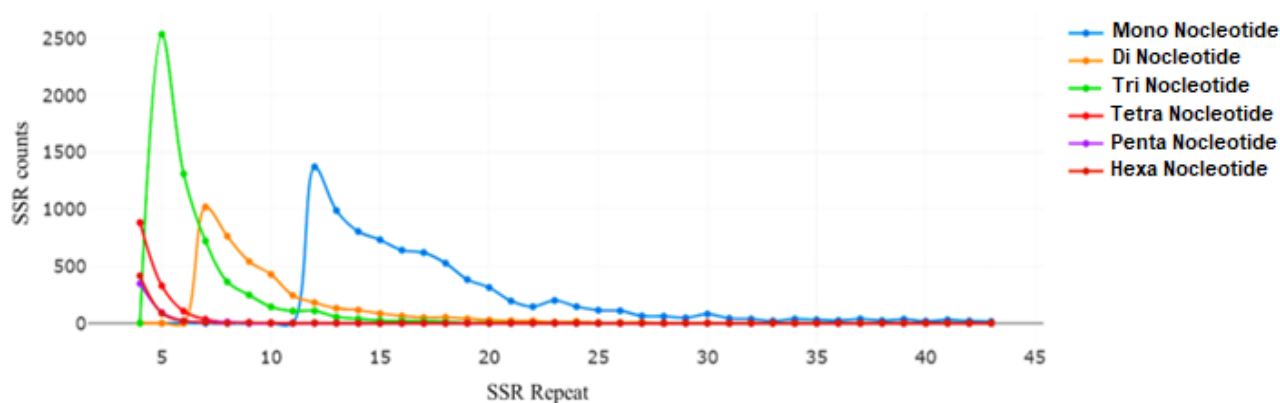
توزیع تکرار برای هر نوع ریزماهوره:

توزیع تکرارهای ریزماهوره هندوانه ابوجهل برای هر نوع در شکل ۴ آمده است. حداکثر فراوانی ریزماهوره در هر تکرار و نوکلئوتید غالب در آن براساس جدول ۴ و شکل ۴ به ترتیب عبارتند از: تکرار ۵ (۳۰۴۱) و نوع غالب تری‌نوکلئوتید (۲۵۳۴)، تکرار ۷ (۱۷۸۸) و نوع غالب دی‌نوکلئوتید (۱۰۲۰)، تکرار ۱۲ (۱۶۶۱) و نوع غالب مونونوکلویتید (۱۳۷۱)، تکرار ۴ (۱۶۴۴) و نوع غالب تترانوکلئوتید (۸۸۲)، تکرار ۶ (۱۴۵۳) و نوع غالب تری‌نوکلئوتید (۱۳۱۰)، تکرار ۱۳ (۱۱۷۳) و نوع غالب

ریزماهوره به کلاس اول (طول بیش از ۲۰ نوکلئوتید) و کلاس دوم (بطول کمتر از ۲۰ نوکلئوتید) تقسیم می‌شوند. در میان نشانگرهای ریزماهوره شناسایی شده در هندوانه ابوجهل، ۷۵/۵۵ درصد از نوع کلاس دوم و ۴۲/۴۵ درصد از نوع کلاس اول بودند.

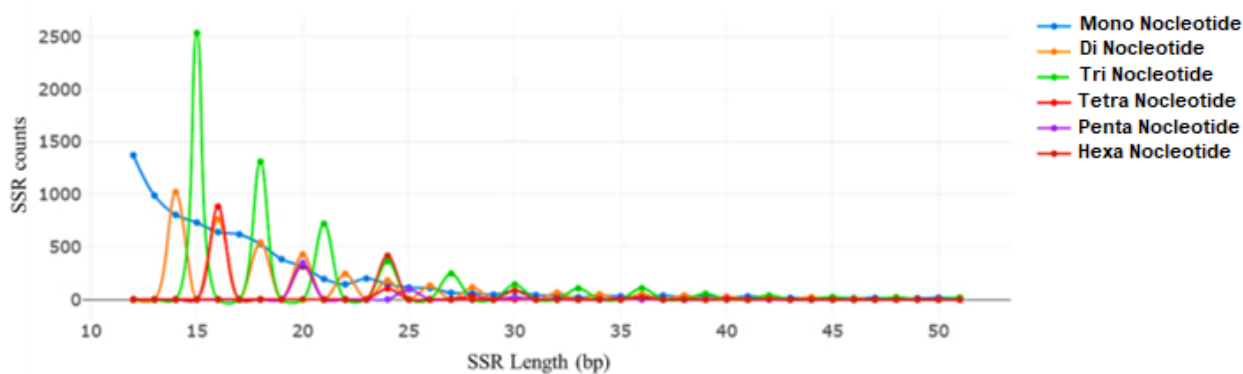
و (۷۲۰)، تترانوکلئوتید دارای طول ۱۶، ۲۰ و ۲۴ (به ترتیب با تعداد ۸۸۲، ۳۲۷ و ۱۰۴)، پتانوکلئوتید دارای طول ۲۰، ۲۵ و ۳۰ (به ترتیب با تعداد ۳۴۷، ۹۴ و ۲۰) و هگزانوکلئوتید دارای طول ۲۴، ۳۰ و ۳۶ (به ترتیب با تعداد ۴۱۵، ۸۶ و ۱۹) بودند.

با توجه به نمودار فراوانی طول تکرارهای ریزماهوره (شکل ۵) در مجموعه طولی ۱۰-۵۰ جفت باز، انواع



شکل ۴- توزیع تکرارهای ریزماهوره شناسایی شده در ترانسکرپتوم هندوانه ابوجهل

Figure 4. Distribution of microsatellite repeats identified in the transcriptome of the *C. colocynthis*



شکل ۵- توزیع طول تکرارهای ریزماهوره شناسایی شده در ترانسکرپتوم هندوانه ابوجهل

Figure 5. Distribution of microsatellite repeats length identified in the transcriptome of the *C. colocynthis*

شناسایی آغازگر حاوی ریزماهوره:

برای طراحی آغازگرهای نشانگرها، از ۲۰۲۲۱ ریزماهوره در مجموع ۳۰۶۵ جفت آغازگر ایجاد شد که از طریق لینک

<https://figshare.com/s/3126fb23814abc19adfc>

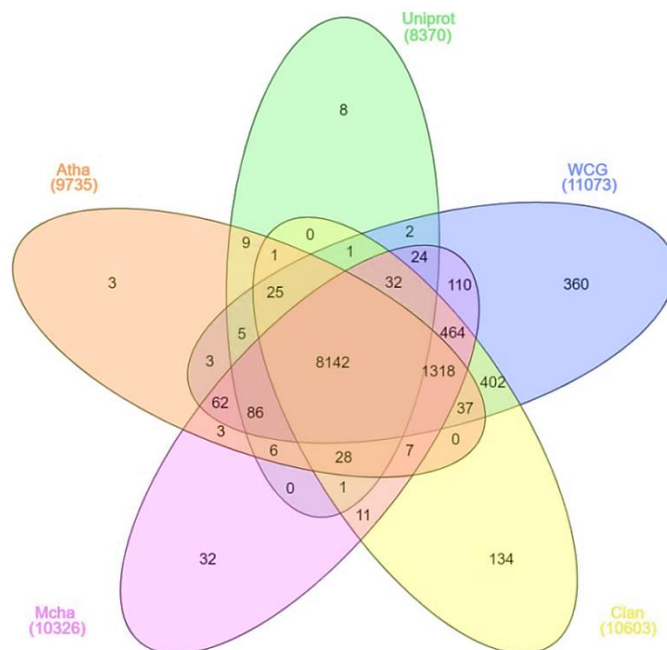
قابل دسترس است.

تفسیر کارکردی رونوشت‌های دارای ریزماهوره:

ترسیم نمودار نتایج بلاست یکس رونوشت‌های هم‌گذاری شده دارای ریزماهوره علیه پایگاه‌های داده مختلف با استفاده از نمودار Venn (شکل ۶) به منظور مقایسه شباهت داده‌ها با ژنوم‌های مرجع مانند یونی‌پروت (Uniprot)، آرآیدوپسیس (Atha) و گونه‌های خانواده کوکوربیتاسه از جمله هندوانه گرد (Clan)، هندوانه رقم چارلستون گری (WCG) و خیار چنبر تلخ (Mcha) ترسیم شد و نتایج همپوشانی رونوشت‌ها با گونه‌های مدل و غیرمدل هندوانه ابوجهل تجزیه و تحلیل گردید. تعداد تک‌ژن‌های حاوی

ریزماهوره به‌طور مشترک در تمام پایگاه‌ها و گونه‌های مدل و غیرمدل در مجموع ۸۱۴۲ رونوشت بود (شکل ۶). در پایگاه Uniprot (۸۳۷۰) و در گونه‌های Atha (۹۷۳۵)، Clan (۱۰۶۰۳)، WCG (۱۱۰۷۳) و Mcha (۱۰۳۲۶) بود (شکل ۶).

علاوه بر این، توالی تک‌ژن‌های حاوی ریزماهوره در سرور تفسیر اتوماتیک KEGG (<http://www.Genome.Jp/kegg/kaas>) بارگذاری شده و با استفاده از شناسه‌های اختصاصی KO تفسیر شدند. با توجه به تخصیص KO، اطلاعات مسیرهای متابولیکی مربوط به تک‌ژن‌های پیش‌بینی شده حاصل گردید. ژن‌های درگیر در بیوسنتز اسکلت ترینوئید هندوانه ابوجهل در جدول ۵ بررسی شده است. تعداد ۱۸ ژن اصلی تولیدکننده آنزیم‌های مسیر بیوسنتز اسکلت ترینوئید شناسایی شده است که دارای ریزماهوره بوده‌اند (جدول ۵).



شکل ۶- نتایج بلاست یکس تفسیر کارکردی تک‌ژن‌های هم‌گذاری شده حاوی ریزماهوره هندوانه ابوجهل علیه پایگاه ژنوم مرجع

Figure 6. X-blast results of the annotated unigenes containing the microsatellite of *C. colocynthis* against reference genome database

جدول ۵- آنزیم‌های حاوی ریزماهوره درگیر در بیوسنتز اسکلت ترپن هندوانه ابوجهل شناسایی شده توسط KEEG

Table 6. Microsatellite-containing enzymes involved in the biosynthesis terpen skeleton of *C. colocynthis* identified by KEEG

Enzyme code	Enzyme description	Unigen ID <i>C. colocynthis</i>	ID K
[EC:2.2.1.7]	dxs; 1-deoxy-D-xylulose-5-phosphate synthase	85625, 95520	K01662
[EC:1.1.1.267]	dxr; 1-deoxy-D-xylulose-5-phosphate reductoisomerase	56519	K00099
[EC:2.7.7.60]	ispD; 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	127222	K00991
[EC:2.7.1.148]	ispE; 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	506184	K00919
[EC:1.17.7.4]	ispH; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	157604	K03527
[EC:5.3.3.2]	idi; isopentenyl-diphosphate Delta-isomerase	1260616	K01823
[EC:2.3.1.9]	ACCAT; acetyl-CoA C-acetyltransferase	1030656, 1021719	K00626
[EC:2.3.3.10]	HMGS; hydroxymethylglutaryl-CoA synthase	1003842	K01641
[EC:1.1.1.34]	HMGCR; hydroxymethylglutaryl-CoA reductase (NADPH)	1002723	K00021
[EC:2.7.1.36]	MK; mevalonate kinase	292084, 506433	K00869
[EC:2.5.1.10]	FDPS; farnesyl diphosphate synthase	7220	K00787
[EC:2.5.1.1]	GPS; geranyl diphosphate synthase	824700	K14066
[EC:2.5.1.87]	DHDDS; ditrans,polycis-polyprenyl diphosphate synthase	502464, 997637	K11778
[EC:2.5.1.58]	FNTB; protein farnesyltransferase subunit beta	1148769	K05954
[EC:2.1.1.100]	ICMT; protein-S-isoprenylcysteine O-methyltransferase	1134250	K00587
[EC:3.1.1.-]	PCME; prenylcysteine alpha-carboxyl methyltransferase	1977, 1173481	K15889
[EC:1.1.1.354]	FLDH; NAD+-dependent farnesol dehydrogenase	1028192	K15891
[EC:2.7.1.216]	FOLK; farnesol kinase	977862, 1159827	K15892

بحث

بسیار متفاوت بود، به عنوان مثال، AG و AT در ریزماهوره‌های دی‌نوکلئوتیدی رایج بودند، در حالی که AAC، ACC و AAG در موتیف‌های تری‌نوکلئوتیدی رایج بودند. فراوان‌ترین ریزماهوره مونونوکلئوتیدی بدست آمده در ترانسکریپتوم هندوانه ابوجهل A با ۸۱۶۷ تکرار (۴۰/۳۹ درصد) بود. در میان ریزماهوره تری‌نوکلئوتید به ترتیب AAG (۲۷۵۷ تکرار) و AAT (۱۱۳۰ تکرار) و در میان دی‌نوکلئوتید AG (۲۲۹۸ تکرار) و AT (۱۱۷۸ تکرار) بیشترین فراوانی را به خود اختصاص دادند (شکل ۳). با توجه به اینکه تکرارهای مونونوکلئوتیدی ممکن است به دلیل خطاهای توالی و اشتباهات موتاز دقیق نباشند، اما با وجود این تکرارهای مونونوکلئوتیدی برای تجزیه و تحلیل بیشتر استفاده شدند. در گیاهان بابونه آلمانی و بابونه رومی نتایج مشابهی گزارش شده است. در بابونه رومی تکرارهای دی‌نوکلئوتیدی، نسبت‌های توزیع نوع AC/CA، GT/TG و AT/TA غالب بودند که به ترتیب ۲۷/۲۵٪، ۲۵/۱۱٪ و ۲۷/۲۵٪ بودند. در تری‌نوکلئوتید، GAT بیشترین توزیع را داشت که ۵/۷۲ درصد بود (Liu et al., 2019). در بابونه آلمانی (GT/TG)_n، (CA/AC)_n و (TA/AT)_n سه نوع موتیف اصلی دیگر بودند که در بین ریزماهوره‌های دی‌نوکلئوتیدی غالب بودند و به ترتیب فرکانس‌های ۴/۲

از نظر غالب‌ترین موتیف (مونو و تری‌نوکلئوتید) نتایج مشابهی در برخی گیاهان خانواده کاسنیان (Asteraceae) مانند بابونه آلمانی (*Matricaria chamomilla*) و بابونه رومی (*Chamaemelum nobile*)، گیاه سارتیا ژاپنی (*Swertia japonica*) از خانواده گل‌سپاسیان (Gentianaceae) و زنجبیل (*Zingiber officinale* Rosc.) گزارش شده است (Liu et al., 2019; Rai et al., 2016; Vidya et al., 2021; Zhang et al., 2018). در گیاه زنجبیل نتایج کاملاً مشابهی در مورد بیشترین و کمترین تکرارهای ریزماهوره‌ها به دست آمد؛ از مجموع ۱۶۷۹۰ ریزماهوره، موتیف‌های تکراری مونونوکلئوتید (۸۵۵۴، ۵۰/۹۴ درصد) فراوان‌ترین ریزماهوره، به دنبال آن ریزماهوره‌های تری‌نوکلئوتید (۴۶۳۹، ۲۷/۶۲ درصد)، دی‌نوکلئوتید (۳۱۷۶ تک‌ژن، ۱۸/۱۹ درصد)، تترانوکلئوتید (۲۵۲ تک‌ژن، ۱/۵ درصد)، هگزانوکلئوتید (۱۰۱ تک‌ژن، ۰/۶۱ درصد) و پنتانوکلئوتید (۲۵۲ تک‌ژن، ۱/۵۰ درصد) بودند (Vidya et al., 2021).

موتیف‌های شناسایی شده در ترانسکریپتوم هندوانه ابوجهل نشان داد که توزیع آنها غیرتصادفی است (جدول ۳) و فرکانس موتیف‌ها حتی در ریزماهوره‌های با اندازه واحد

برای طراحی آغازگرهای نشانگرها، از ۲۰۲۲۱ ریزماهواره در مجموع ۳۰۶۵ جفت آغازگر ایجاد شد (لینک دسترسی:

<https://figshare.com/s/3126fb23814abc19adfc>.

نشانگرهای ریزماهواره همزمان، به راحتی قابل سنجش هستند و با محتوای اطلاعاتی چندشکلی بالا (PIC)، بسیار قابل تکرار هستند. همچنین، بیشتر جایگاه‌های ریزماهواره توسط یک جفت پرایمر منحصر به فرد تعریف می‌شوند (Chen et al., 2006). در مطالعات مشابه، آغازگر ریزماهواره با استفاده از نرم‌افزار Primer3 برای ۱۷۵۴۸ توالی علف سودان (*Sorghum × drummondii*) طراحی شد (Zhu et al., 2017). در مطالعه دیگر ۲۰۰ جفت آغازگر ریزماهواره برای تکثیر ۱۰ نمونه DNA گونه الیموس (*Elymus sp.*) طراحی و انتخاب شدند و آغازگرهای ریزماهواره با موفقیت مقادیر قابل توجهی از پلی‌مورفیسم را شناسایی کردند (Zhang et al., 2019). همچنین برای ۲۴۵۲۲ توالی ریزماهواره زعفران، آغازگر برای توالی‌های ریزماهواره با موفقیت طراحی شد (Rezaei et al., 2022). در گیاه زنجبیل ۱۲۵ جفت آغازگر برای توالی‌های دارای ریزماهواره از تک‌ژن‌ها و توالی‌های کدکننده آن طراحی شدند (Vidya et al., 2021).

نتایج بلاست پوشش گسترده تک‌ژن‌های حاوی ریزماهواره محافظت‌شده را در سراسر هندوانه ابوجهل (۲۰۲۲۱) و ادغام تک‌ژن‌های حاوی ریزماهواره مدل گونه‌های نزدیک مانند Atha, Uniprot, Clan, WCG و Mcha را ارائه دادند. همان‌طور که در شکل ۶ نشان داده شده است، تک‌ژن‌های حاوی ریزماهواره مشترک در پنج داده ژنوم مرجع Atha, Uniprot, Clan, WCG و Mcha تعداد کل ۸۱۴۲ تک‌ژن، در چهار ژنوم مرجع Atha, Uniprot, Clan, WCG و Mcha تعداد کل ۱۳۱۸ تک‌ژن، ۴۶۴ تک‌ژن مشترک برای Atha, Uniprot, Clan, WCG و Mcha، ۴۰۲ تک‌ژن مشترک برای Atha, Uniprot, Clan, WCG و Mcha، ۱۱۰ تک‌ژن مشترک برای Atha, Uniprot, Clan, WCG و Mcha، در نهایت ۳۶۰ مورد برای WCG و ۱۳۴ مورد برای Atha, Uniprot, Clan اختصاص یافت. در این مطالعه مقایسه ژنوم مرجع با ترانسکریپتوم

۳/۴ و ۲/۹ درصد را نشان می‌دادند. در میان ۲۰ نوع ریزماهواره تری‌نوکلئوتیدی، TGA, TCA, GAA و ATC که به ترتیب ۱/۴٪، ۱/۲٪، ۱/۱٪ و ۱/۱٪ شایع‌ترین بودند (Zhang et al., 2018). محققان از بررسی نشانگرهای ریزماهواره ژنوم توالی‌یابی‌شده کرفس (*Apium graveolens*) دریافتند که تکرارهای غالب A/T در بین مونونوکلئوتیدها، AG/CT در بین دی‌نوکلئوتیدها (۴۵/۵۴ درصد) و در بین تری‌نوکلئوتیدها ACT/AGT, ATC/ATG, AAG/CTT و ACC/GGT بیشترین فراوانی را داشتند (Zhou et al., 2014). به‌طور مشابهی فراوانی غالب A/T، AG/CT و AAG/CTT در بین تکرارهای مونو، دی و تری‌نوکلئوتیدی در گیاه (*Sterculia lanceolata*) گزارش شد (Eum et al., 2019).

فراوانی تکرارهای ریزماهواره در ترانسکریپتوم هندوانه ابوجهل تقریباً یک در هر ۴/۱۴ کیلوباز توالی هم‌گذاری‌شده بود. در تحقیقات مشابه مانند گیاه ارکیده دندروبیوم (*Dendrobium nobile*) فراوانی تکرارهای ریزماهواره تقریباً یک در هر ۴ کیلوباز (Xu et al., 2017) و گل صدتومانی (*Paeonia suffruticosa*) فراوانی تکرارهای ریزماهواره تقریباً یک در هر ۴/۶۶ کیلوباز (He et al., 2020) توالی هم‌گذاری‌شده بود. فراوانی تکرارهای ریزماهواره در ترانسکریپتوم گیاه هندوانه ابوجهل کمتر از فراوانی تکرارهای ریزماهواره در زنیان (یک در هر ۱۰/۱ کیلوباز) (Soltani Howyzeh et al., 2019) و گل‌آذین بومادران (یک در هر ۱۰ کیلوباز) (Shahbaz et al., 2023)، جو (یک در هر ۸/۹ کیلوباز) و بیشتر از فراوانی برنج (یک در هر ۳/۶ کیلوباز) (Parida et al., 2010) در توالی‌های هم‌گذاری‌شده بود. در میان نشانگرهای ریزماهواره شناسایی شده در هندوانه ابوجهل، ۵۷/۵۵ درصد از نوع کلاس دوم (به طول کمتر از ۲۰ نوکلئوتید) و ۴۲/۴۵ درصد از نوع کلاس اول (به طول بیش از ۲۰ نوکلئوتید) بودند. ریزماهواره‌ها برای تمایز و گروه‌بندی جدایه‌های CLas مورد استفاده قرار گرفته و توانایی بالایی در تمایز ایزوله‌هایی با روابط نزدیک و ردیابی ژنوتیپ‌های هدف دارد (Zhang et al., 2018).

متابولیت‌های ثانویه و توسعه نشانگر، تنوع ژنتیکی و ساخت نقشه‌های ژنتیکی در گیاه هندوانه ابوجهل استفاده کرد و نتایج

هندوانه ابوجهل هم‌گذاری شده حاوی ریزماهوره انجام شد که مجموعه تک‌ژن‌های حاوی ریزماهوره در نمودار Venn به اشتراک گذاشته شدند. تجزیه و تحلیل مقایسه گسترده‌تری تک‌ژن‌های حاوی ریزماهوره هندوانه ابوجهل می‌تواند به ژنوم‌های مشترک گونه‌های خانواده کدوئیان مرتبط باشد. تک‌ژن‌های حاوی ریزماهوره هندوانه ابوجهل در این تحقیق، به‌طور تقریباً مشابهی علیه ژنوم هندوانه چارلستون گری بیش از سایر مدل‌ها درصد تفسیر بالاتری داشت. این موضوع که هندوانه ابوجهل از خانواده کدوئیان بوده و همانندی بالایی با ژنوم هندوانه چارلستون گری دارد، مورد تأیید بوده که با مطالعه Dorafshan و همکاران (۲۰۱۹a) مطابقت دارد. در تحقیق آنها، تک‌ژن‌های هم‌گذاری شده ترانسکریپتوم هندوانه ابوجهل در مقابل پنج ژنوم مدل هندوانه گرد (۲۳۴۴۰)، هندوانه رقم چارلستون گری (۲۲۵۶۷)، خیار چنبر تلخ (۲۱۶۲۳)، خربزه (۲۹۹۸۰) و خیار چینی بلند (۲۳۲۴۸) با استفاده از نرم‌افزار Venny بررسی شد. در نتیجه ادغام نتایج بلاست ایکس علیه این پنج ژنوم مدل، تعداد کل ۲۹۵۳۲ تک‌ژن تفسیر شده مشترک، ۳۱۹۶ تک‌ژن مشترک برای Clan، WCG و Mcha، ۱۱۶۰ تک‌ژن مشترک برای WCG و Mcha و در نهایت ۲۴۹۴ مورد به WCG متعلق بود (Dorafshan *et al.*, 2019a). مطالعات مشابه تفسیر تک‌ژن‌های حاوی ریزماهوره با استفاده از بلاست ایکس در گیاهان مختلف مانند زنیان (Soltani Howyzeh *et al.*, 2019) و بومادران (Shahbaz *et al.*, 2023) نشان‌دهنده هم‌پوشانی گسترده ژن‌ها با خانواده خود می‌باشد.

به‌طور کلی می‌توان گفت با توجه به اینکه نشانگرهای ریزماهوره دارای نقش کلیدی در ژنوم و مرتبط با ژن‌ها به‌ویژه در بیوسنتز متابولیت‌های ثانویه در گیاهان دارویی هستند، می‌توان از ریزماهوره‌های موجود در توالی‌های ژنومی برای سازماندهی ژنومی، نوترکیبی، تنظیم بیان ژن، تنوع ژنتیکی کمی و تکامل ژن‌ها استفاده کرد. در این مطالعه، نشانگرهای ریزماهوره حاصل از توالی‌یابی ترانسکریپتوم هندوانه ابوجهل در مقیاس گسترده شناسایی و آنالیز شدند. بنابراین از توالی‌یابی نسل جدید، می‌توان برای بهبود

این تحقیق می‌تواند گامی اساسی در این مسیر بردارد.

References

- Appleby, N., Edwards, D., Batley, J., 2009. New technologies for ultra-high throughput genotyping in plants, *Plant genomics*. Springer, pp. 19-39.
- Badr, A., Kamel, M., Zaki, H., 2018. Genetic diversity of Colocynth (*Citrullus colocynthis* Schrader) populations in the Eastern Desert of Egypt as revealed by morphological variation and ISSR polymorphism. *Feddes Repertorium*, 129: 173-184.
- Chen, J.C., Chiu, M.H., Nie, R.L., Cordell, G.A., Qiu, S.X., 2005. Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Natural product reports*, 22: 386-399.
- Chen, C., Zhou, P., Choi, Y.A., Huang, S., Gmitter, F.G., 2006. Mining and characterizing microsatellites from citrus ESTs. *Theoretical and Applied Genetics*, 112: 1248-1257.
- Dorafshan, M., Soltani Howyzeh, M., Shariati, V., 2019a. Comparative analysis of the transcriptome *Citrullus colocynthis* medicinal plant against the reference genome of different species of the Cucurbitaceae family., First National Conference on Agricultural and Environmental Sciences of Iran. Bavi, Khuzestan, Iran. <https://civilica.com/doc/870271/>. (In Persian).
- Dorafshan, M., Soltani Howyzeh, M., Shariati, V., 2019b. Identification of terpenoid backbone biosynthetic pathway genes in the fruit of *Citrullus colocynthis* (L.) Schrad. medical plant by RNA sequencing. *Iranian Journal of Medicinal and Aromatic Plants Research*, 35: 691-702. (In Persian)
- Dorafshan, M., Soltani Howyzeh, M., Shariati, V., 2020. Identifying triterpene and sesquiterpene biosynthetic pathway genes in *Citrullus colocynthis* L. fruit using next-generation sequencing (NGS) technology. *Iranian Journal of Medicinal and Aromatic Plants*, 36. (In Persian)
- Dorafshan, M., Soltani Howyzeh, M., Shariati, V., 2021. De novo transcriptome assembly provides insight into the specific genes of the carotenoid biosynthetic pathway of *Citrullus colocynthis* (L.), The fourth International and 16th National Genetics Congress, Tehran., Iran. <https://civilica.com/doc/1195674/>.
- Dutta, S., Kumawat, G., Singh, B.P., Gupta, D.K., Singh, S., Dogra, V., Gaikwad, K., Sharma, T.R., Raje, R.S., Bandhopadhyaya, T.K., Datta, S., Singh, M.N., Bashasab, F., Kulwal, P., Wanjari, K.B., K Varshney, R., Cook, D.R., Singh, N.K., 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biology*, 11: 17.
- Eum, S.M., Kim, S.-Y., Hong, J.S., Roy, N.S., Choi, S., Paik, J., Lee, S.W., Tran, T.B., Do, V.H., Kim, K.S., 2019. Transcriptome analysis and development of SSR markers of ethnobotanical plant *Sterculia lanceolata*. *Tree Genetics & Genomes*, 15: 1-12.
- Gomes-Carneiro, M.R., Felzenszwalb, I., Paumgarten, F.J., 1998. Mutagenicity testing of (\pm)-camphor, 1, 8-cineole, citral, citronellal, (-)-menthol and terpineol with the Salmonella/microsome assay. *Mutation research/genetic toxicology and environmental mutagenesis*, 416: 129-136.
- Hancock, J.M., 1996. Simple sequences in a 'minimal' genome. *Nature Genetics*, 14: 14-15.
- Hassan, A., Younis, R., Rashed, M., Eldomyati, F., Rashed, N., 2012. Molecular markers associated with drought tolerance in *Citrullus colocynthis*. *Egyptian Journal of Genetics Cytology*, 41: 311-325.
- He, D., Zhang, J., Zhang, X., He, S., Xie, D., Liu, Y., Li, C., Wang, Z., Liu, Y., 2020. Development of SSR markers in *Paeonia* based on De Novo transcriptomic assemblies. *PLoS One*, 15: e0227794.
- Hwang, J., Kang, J., Son, B., Kim, K., Park, Y., 2011. Genetic diversity in watermelon cultivars and related species based on AFLPs and EST-SSRs. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 39(2), pp.285-292.
- Jarvis, D., Kopp, O., Jellen, E., Mallory, M., Pattee, J., Bonifacio, A., Coleman, C., Stevens,

- M., Fairbanks, D., Maughan, P., 2008. Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). *Journal of Genetics*, 87: 39-51.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2013. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, 42: D199-D205.
 - Kaushik, U., Aeri, V., Mir, S.R., 2015. Cucurbitacins—an insight into medicinal leads from nature. *Pharmacognosy reviews*, 9: 12.
 - Levi, A., Thomas, C., Joobeur, T., Zhang, X., Davis, A., 2002. A genetic linkage map for watermelon derived from a testcross population :(*Citrullus lanatus* var. *citroides* × *C. lanatus* var. *lanatus*) × *Citrullus colocynthis*. *Theoretical and Applied Genetics*, 105: 555-563.
 - Li, Q.-Y., Munawar, M., Saeed, M., Shen, J.-Q., Khan, M.S., Noreen, S., Alagawany, M., Naveed, M., Madni, A., Li, C.-X., 2022. *Citrullus colocynthis* (L.) Schrad (Bitter Apple Fruit): Promising traditional uses, pharmacological effects, aspects, and potential applications. *Frontiers in Pharmacology*, 12: 791049.
 - Li, S., Tighe, S.W., Nicolet, C.M., Grove, D., Levy, S., Farmerie, W., Viale, A., Wright, C., Schweitzer, P.A., Gao, Y., Kim, D., Boland, J., Hicks, B., Kim, R., Chhangawala, S., Jafari, N., Raghavachari, N., Gandara, J., Garcia-Reyero, N., Hendrickson, C., Roberson, D., Rosenfeld, J.A., Smith, T., Underwood, J.G., Wang, M., Zumbo, P., Baldwin, D.A., Grills, G.S., Mason, C.E., 2014. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature Biotechnology*, 32: 915.
 - Liu, X., Wang, X., Chen, Z., Ye, J., Liao, Y., Zhang, W., Chang, J., Xu, F., 2019. De novo assembly and comparative transcriptome analysis: novel insights into terpenoid biosynthesis in *Chamaemelum nobile* L. *Plant cell reports*, 38: 101-116.
 - Ma, J., Gao, B., Wang, R., Li, X., Chen, S., 2020. Transcriptome analyses of *Ditylenchus destructor* in responses to cold and desiccation stress. *Genetics and Molecular Biology*, 43: 1-12.
 - Nimmakayala, P., Tomason, Y.R., Jeong, J., Ponniah, S.K., Karunathilake, A., Levi, A., Perumal, R., Reddy, U.K., 2010. Genetic reticulation and interrelationships among *Citrullus* species as revealed by joint analysis of shared AFLPs and species-specific SSR alleles. *Plant Genetic Resources*, 8: 16-25.
 - Parida, S.K., Pandit, A., Gaikwad, K., Sharma, T.R., Srivastava, P.S., Singh, N.K., Mohapatra, T., 2010. Functionally relevant microsatellites in sugarcane unigenes. *BMC Plant Biology*, 10: 1-19.
 - Rai, A., Nakamura, M., Takahashi, H., Suzuki, H., Saito, K., Yamazaki, M., 2016. High-throughput sequencing and de novo transcriptome assembly of *Swertia japonica* to identify genes involved in the biosynthesis of therapeutic metabolites. *Plant Cell Reports*, 35: 2091-2111.
 - Rezaei, M., Sharifi, H., Seifi, A., 2022. Transcriptome assembly and identification of EST-SSR markers in *Crocus sativus*. *Saffron Agronomy and Technology*, 10: 41-49.
 - Roshan, N.M., Ashouri, M., Sadeghi, S.M., 2021. Identification, evolution, expression analysis of phospholipase D (PLD) gene family in tea (*Camellia sinensis*). *Physiology and Molecular Biology of Plants*, 27: 1219-1232.
 - Senthilvel, S., Jayashree, B., Mahalakshmi, V., Kumar, P.S., Nakka, S., Nepolean, T., Hash, C., 2008. Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. *BMC Plant Biology*, 8: 1-9.
 - Shahbaz, B., Howyzeh, M.S., Shariati, V., 2023. "Research Paper" Large Scale Identification of SSR Molecular Markers Using RNA Sequencing in Yarrow (*Acillea millefolium* L.). *Journal of Crop Breeding* 15: 156-165.
 - Si, Y., Zhang, C., Meng, S., Dane, F., 2009. Gene expression changes in response to drought stress in *Citrullus colocynthis*. *Plant cell reports*, 28: 997-1009.
 - Soltani Howyzeh, M., Sadat Noori, S.A., Shariati, V., Amiripour, M., 2019. Large Scale Identification of SSR Molecular Markers in Ajowan (*Trachyspermum ammi*) Using RNA Sequencing. *Plant Genetic Researches*, 6: 31-46.
 - Song, X., Li, N., Guo, Y., Bai, Y., Wu, T., Yu, T., Feng, S., Zhang, Y., Wang, Z., Liu, Z., 2021. Comprehensive identification and

- characterization of simple sequence repeats based on the whole-genome sequences of 14 forest and fruit trees. *Forestry Research*, 1: 1-10.
- Tautz, D. 1989. Hyper variability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17: 6463-6471.
 - Thakur, O., Randhawa, G.S., 2018. Identification and characterization of SSR, SNP and In Del molecular markers from RNA-Seq data of guar (*Cyamopsis tetragonoloba*, L. Taub.) roots. *BMC Genomics*, 19: 1-14.
 - Verma, K.S., Kachhwaha, S., Kothari, S., 2013. In vitro plant regeneration of *Citrullus colocynthis* (L.) Schard. and assessment of genetic fidelity using ISSR and RAPD markers. *Indian J Biotechnol.* 12(3): 409-414.
 - Vidya, V., Prasath, D., Snigdha, M., Gobu, R., Sona, C., Maiti, C.S., 2021. Development of EST-SSR markers based on transcriptome and its validation in ginger (*Zingiber officinale* Rosc.). *PLoS One*, 16: e0259146.
 - Wang, Z., Hu, H., Goertzen, L.R., McElroy, J.S., Dane, F., 2014. Analysis of the *Citrullus colocynthis* transcriptome during water deficit stress. *PLoS One*, 9: e104657.
 - Wang, Z., Wang, S., Yu, C., Han, X., Zou, D., 2020. QTL analysis of rice photosynthesis-related traits under cold stress across multi-environments. *Euphytica*, 216: 1-21.
 - Wei, Z., Sun, Z., Cui, B., Zhang, Q., Xiong, M., Wang, X., Zhou, D., 2016. Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: de novo assembly, annotation and EST-SSR marker development. *Peer J*, 4: e2378.
 - Xu, M., Liu, X., Wang, J.-W., Teng, S.-Y., Shi, J.-Q., Li, Y.-Y., Huang, M.-R., 2017. Transcriptome sequencing and development of novel genic SSR markers for *Dendrobium officinale*. *Molecular Breeding*, 37: 1-7.
 - Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99: 193-208.
 - Zhang, W., Tao, T., Liu, X., Xu, F., Chang, J., Liao, Y., 2018. De novo assembly and comparative transcriptome analysis: novel insights into sesquiterpenoid biosynthesis in *Matricaria chamomilla* L. *Acta Physiologiae Plantarum*, 40: 1-14.
 - Zhang, Z., Xie, W., Zhao, Y., Zhang, J., Wang, N., Ntakirutimana, F., Yan, J., 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: 1-18.
 - Zhou, Y., Chen, L., Fan, X., Bian, Y., 2014. De novo assembly of *Auricularia polytricha* transcriptome using Illumina sequencing for gene discovery and SSR marker identification. *PLoS one*, 9: e91740.
 - Zhu, Y., Wang, X., Huang, L., Lin, C., Zhang, X., Xu, W., Peng, J., Li, Z., Yan, H., Luo, F., 2017. Transcriptomic identification of drought-related genes and SSR markers in Sudan grass based on RNA-Seq. *Frontiers in Plant Science*, 8: 687.