

Identification of EST-SSR regions in the transcriptome of *Quercus infectoria* using RNA-Seq technology

Forough Joudaki 

Ph.D. Student, Department of Plant Production and Genetic Engineering, Faculty of Agriculture, Lorestan University, Khorramabad, Iran. E-mail: fojoudaki00@gmail.com

Ahmad Ismaili 

*Corresponding author. Professor, Department of Plant Production and Genetic Engineering, Faculty of Agriculture, Lorestan University, Khorramabad, Iran. E-mail: ismaili.a@lu.ac.ir

Seyed Sajad Sohrabi 

Assistant Professor, Department of Plant Production and Genetic Engineering, Faculty of Agriculture, Lorestan University, Khorramabad, Iran. E-mail: sohrabi.s@lu.ac.ir

Abstract

Objective

Gall oak (*Quercus infectoria*), is one of the extraordinary tree species and a valuable medicinal plant for its products, including galls and tannins. It is one of the most important species of the *Quercus* genus, which has been recognized and utilized as a medicinal plant since ancient times. Despite the importance of this species, limited basic genetic information has hindered research on this species. EST-SSR molecular markers, due to their high number of polymorphic loci, can provide specific genetic information. In this study, EST-SSR markers were identified in the *Q. infectoria* transcriptome generated through *de novo* assembly of RNA-Seq reads.

Materials and methods

The RNA sequencing technique was used for *de novo* transcriptome assembly and sequencing was performed by Illumina HiSeq 2500 platform. *De-novo* assembly was performed for high-quality reads using CLC Genomics Workbench (v7.5.0) software. For identification of microsatellite markers, MISA (Microsatellite searching Tool) was used. In this study, the microsatellites from mono-nucleotide to hexa-nucleotide were detected. To functionally group genes, GO analysis was performed. Also, mapping against the Plant Reactome pathway databases was performed to the identification of biological pathways associated with EST-SSR-containing unigenes.

Results

From 96,225 transcripts in the *Q. infectoria* transcriptome, 20,505 EST-SSR markers were identified across 13,368 unigenes, with 4,355 unigenes containing more than one marker locus. Among all types of repeats, mononucleotide repeats (11394, 55.57%) were the most abundant, followed by dinucleotide and three- nucleotide repeats with 23.92 and 18.82%, respectively. The highest percentage of mono-nucleotide motifs was A/T type, the highest percentage of dinucleotide motifs was AG/CT type (68.46%), and the highest percentage of 3, 4, and 5 nucleotide motifs were AAG/CTT, AAAT/ATTT, and AAAAG/CTTTT type, respectively. Identification of biologically active pathways associated with SSR-containing unigenes was carried out using a

plant database. Most unigenes were linked in signal transduction, protein metabolism, cell cycle, and gene expression (transcription) pathways. Analysis of the enriched pathways with SSR-containing unigenes also showed that the highest abundance was related to the membrane, nuclear, and metabolic pathways of compounds containing nucleobases.

Conclusions

This study provides valuable set of EST-SSRs for *Q. infectoria* and can provide the way for further functional and comparative genomic research on this species.

Keywords: EST-SSR marker, *Quercus infectoria*, RNA-Seq

Paper Type: Research Paper.

Citation: Joudaki, F., Ismaili, A., & Sohrabi, S.S. (2026). Identification of EST-SSR regions in the transcriptome of *Quercus infectoria* using RNA-Seq technology. *Agricultural Biotechnology Journal*, 18(4), 23-44.

Agricultural Biotechnology Journal, 18(4), 23-44.

DOI: 10.22103/jab.2026.26408.1820

Received: March 30, 2026.

Received in revised form: May 29, 2026.

Accepted: May 30, 2026.

Published online: August 30, 2026.

Publisher: Shahid Bahonar University of Kerman & Iranian Biotechnology Society.



© the authors

Introduction

Gall oak (*Quercus infectoria*), is one of the extraordinary tree species and a valuable medicinal plant for its products, including galls and tannins. *Q. infectoria* is widely distributed in several regions of the Middle East and Mediterranean areas and plays an important ecological and medicinal role. It is one of the most important species of the *Quercus* genus, which has been recognized and utilized as a medicinal plant since ancient times. The galls produced on this species are rich in tannins and other bioactive compounds that have long been used in traditional medicine, pharmaceutical preparations, and industrial applications. Despite the importance of this species, limited basic genetic information (genomic and transcriptomic resources) for this species remain limited compared with other model plants. The development of molecular markers derived from expressed sequences can therefore provide an effective tool for understanding the genetic structure and functional diversity of plant species. EST-SSR markers, which originate from expressed regions of the genome, are particularly valuable because they are often associated with genes that participate in biological processes and metabolic pathways. These markers are also known for their relatively high transferability among related species and their usefulness in comparative genomic studies. Totally, EST-SSR molecular markers, due to their high number of polymorphic loci, can provide specific genetic information. Hence, in this study, EST-SSR markers were identified in the *Q. infectoria* transcriptome generated through de novo assembly of RNA-Seq reads. Therefore, identifying and characterizing EST-SSR loci in the transcriptome of *Q. infectoria* can contribute to the development of genetic resources that support future research on diversity, evolution, and functional genomics in this medicinal oak species.

Materials and methods

RNA sequencing (RNA-Seq) technology was employed to generate transcriptomic data for *Quercus infectoria* and to facilitate de novo transcriptome assembly. Sequencing was carried out using the Illumina HiSeq 2500 platform, which provides high-throughput sequencing capacity

and reliable read quality for transcriptome analysis. Raw reads obtained from sequencing were subjected to quality control procedures to remove low-quality sequences and potential contaminants. The resulting high-quality reads were then used for de novo assembly using CLC Genomics Workbench (v7.5.0). This assembly approach enabled the reconstruction of transcript sequences without relying on previously available genomic information, which is particularly important for non-model plant species with limited genomic resources. Following transcriptome assembly, the assembled sequences were used as the input dataset for the identification of simple sequence repeats (SSRs). Microsatellite loci were detected using the MISA (Microsatellite Identification Tool) software, which allows systematic scanning of nucleotide sequences for tandem repeat motifs. In this study, SSRs with motif lengths ranging from mono-nucleotide to hexa-nucleotide repeats were identified across the assembled transcripts. The detection of these repeat motifs provided a valuable set of EST-SSR markers derived from expressed regions of the genome. To further investigate the functional relevance of the identified SSR-containing sequences, functional annotation and classification analyses were performed. Gene Ontology (GO) analysis was conducted to categorize the unigenes according to their associated biological processes, molecular functions, and cellular components. In addition, pathway mapping was performed using the Plant Reactome database to identify metabolic and regulatory pathways associated with the EST-SSR-containing unigenes. This integrated analytical workflow enabled the identification of potential molecular markers while also providing insights into the biological roles of the genes represented in the transcriptome of *Q. infectoria*.

Results

From 96,225 transcripts in the *Q. infectoria* transcriptome, 20,505 EST-SSR markers were identified across 13,368 unigenes, with 4,355 unigenes containing more than one marker locus. Among all types of repeats, mononucleotide repeats (11,394, 55.57%) were the most abundant, followed by dinucleotide and three-nucleotide repeats with 23.92 and 18.82%, respectively. The highest percentage of mono-nucleotide motifs was A/T type, the highest percentage of dinucleotide motifs was AG/CT type (68.46%), and the highest percentage of 3, 4, and 5 nucleotide motifs were AAG/CTT, AAAT/ATTT, and AAAAG/CTTTT type, respectively. Identification of biologically active pathways associated with SSR-containing unigenes was carried out using a plant database. Most unigenes were linked in signal transduction, protein metabolism, cell cycle, and gene expression (transcription) pathways. Analysis of the enriched pathways with SSR-containing unigenes also showed that the highest abundance was related to the membrane, nuclear, and metabolic pathways of compounds containing nucleobases. The distribution pattern of SSR motifs in the transcriptome revealed that short repeat units were the most frequent, which is consistent with the SSR profiles reported for many other plant species. The predominance of A/T-rich motifs indicates that these sequences are highly represented within the expressed regions of the genome. In addition, the presence of dinucleotide and trinucleotide motifs within coding sequences may contribute to the development of informative genetic markers that can be used in population genetic studies and molecular breeding programs. The identification of thousands of SSR loci across a large number of unigenes highlights the richness of the transcriptomic dataset generated in this study. Functional annotation results indicated that a considerable proportion of the SSR-containing unigenes were involved in fundamental cellular activities. Many of these genes were associated with transcriptional regulation, protein synthesis, signal transduction pathways, and cell cycle control. These processes play central roles in plant growth, development, and responses to environmental stimuli. The pathway enrichment analysis further revealed that several SSR-associated genes participate in metabolic processes related to nucleobase-containing

compounds, cellular membranes, and nuclear functions. Such findings suggest that microsatellite loci are frequently located in genes that contribute to essential biological functions, highlighting their potential relevance for functional genomics studies.

Conclusions

This study provides valuable set of EST-SSRs for *Q. infectoria* and can provide the way for further functional and comparative genomic research on this species. The identification and characterization of a large number of EST-SSR markers from the *Q. infectoria* transcriptome significantly expands the available molecular resources for this medicinal oak species. The EST-SSR markers developed in this study can therefore serve as valuable tools for a wide range of applications, including genetic diversity analysis, population structure studies, germplasm characterization, and comparative genomic research within the genus *Quercus*. Overall, the dataset generated in this study provides an important foundation for future research aimed at understanding the molecular basis of biologically important traits in *Q. infectoria*. Consequently, the EST-SSR resource reported here represents a useful genomic platform that can support both fundamental and applied studies in plant biotechnology, conservation genetics, and molecular breeding of *Quercus* species.

Author contributions

Conceptualization, Ahmad Ismaili; methodology, Forough Joudaki, Seyed Sajad Sohrabi; formal analysis, Seyed Sajad Sohrabi; investigation, Forough Joudaki; data curation, Seyed Sajad Sohrabi, Forough Joudaki; writing—original draft preparation, Forough Joudaki; writing—review and editing, Ahmad Ismaili, Seyed Sajad Sohrabi; supervision, Ahmad Ismaili; project administration, Ahmad Ismaili. All authors have read and agreed to the published version of the manuscript.

Data availability statement

The data supporting the findings of this study are openly available in the NCBI database, SRA section, with accession number PRJNA877380.

Acknowledgements

The authors express their gratitude to the Faculty of Agriculture at Lorestan University for providing essential laboratory and greenhouse facilities.

Ethical considerations

This investigation did not involve human or animal subjects; thus, ethical approval was not needed. The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

The study was funded by Department of Plant Production and Genetic Engineering, Faculty of Agriculture, Lorestan University, Khorramabad, Iran.

Conflict of interest

The authors declare that there is no conflict of interest.

شناسایی نواحی حاوی نشانگرهای EST-SSR در ترنسکریپتوم بلوط دارمازو (*Quercus infectoria*) با استفاده از فن آوری RNA-Seq

 **فروغ جودکی**

دانشجوی دکتری، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه لرستان، خرم آباد، ایران. رایانامه: fojoudaki00@gmail.com

 **احمد اسماعیلی**

*نویسنده مسئول: استاد، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه لرستان، خرم آباد، ایران. رایانامه: ismaili.a@lu.ac.ir

 **سید سجاد سهرابی**

استادیار، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه لرستان، خرم آباد، ایران. رایانامه: sohrabi.s@lu.ac.ir

تاریخ دریافت: ۱۴۰۵/۰۱/۱۰ تاریخ دریافت فایل اصلاح شده نهایی: ۱۴۰۵/۰۳/۰۸ تاریخ پذیرش: ۱۴۰۵/۰۳/۰۹

چکیده

هدف: دارمازو یا بلوط گال‌زا (*Quercus infectoria*) از درختان دارویی ارزشمند حاوی گال‌ها و تانن‌ها است. این گیاه یکی از گونه‌های جنس *Quercus* است که از زمان‌های قدیم به عنوان یک گیاه دارویی شناخته شده است. با وجود اهمیت فراوان این گونه، محدودیت اطلاعات ژنتیکی، مطالعات در مورد آن را دچار مشکل کرده است. نشانگرهای مولکولی EST-SSR، به دلیل داشتن تعداد بالای مکان‌های چندشکلی، می‌تواند اطلاعات ژنتیکی ویژه‌ای را ارائه دهد. در این مطالعه، نشانگرهای EST-SSR موجود در ترنسکریپتوم *Q. infectoria* که از سرهم‌بندی نوپدید خوانش‌های حاصل از RNA-Seq ایجاد شده، شناسایی شدند.

مواد و روش: از تکنیک توالی‌یابی RNA به منظور ایجاد سرهم‌بندی نوپدید استفاده و توالی‌یابی توسط پلتفرم Illumina HiSeq 2500 انجام شد. سرهم‌بندی نوپدید برای خوانش‌های با کیفیت بالا با استفاده از نرم‌افزار CLC Genomics Workbench (v7.5.0) انجام شد. به منظور تعیین نواحی نشانگرهای ریزماهواره از MISA (ابزار جستجوی ریزماهواره) استفاده شد. در این مطالعه، ریزماهواره‌های تک نوکلئوتیدی تا شش نوکلئوتیدی شناسایی شدند. برای گروه‌بندی عملکردی ژن‌ها، آنالیز هستی‌شناسی ژن‌ها انجام شد. همچنین، نقشه‌برداری در برابر پایگاه‌های داده Plant Reactome به منظور شناسایی مسیرهای بیولوژیکی مرتبط با تک‌ژن‌های حاوی EST-SSR انجام شد.

نتایج: در این مطالعه، نواحی حاوی نشانگرهای EST-SSR در ترانسکریپتوم دارمازو که از سرهم‌بندی نوپدید خوانش‌های حاصل از RNA-Seq به دست آمده، شناسایی شدند. با بررسی ۹۶۲۲۵ یونی‌ژن از ترانسکریپتوم دارمازو، ۲۰۵۰۵ مورد ناحیه نشانگر EST-SSR در ۱۳۳۶۸ یونی‌ژن شناسایی شد که از این تعداد ۴۳۵۵ یونی‌ژن بیش از یک مکان نشانگر را دارا بودند. در بین نشانگرها،

تکرارهای تک (مونو) نوکلئوتیدی (۱۱۳۹۴، ۵۵/۵۷ درصد) و پس از آن تکرارهای دو و سه نوکلئوتیدی به ترتیب با ۲۳/۹۲ و ۱۸/۸۲ درصد دارای بیشترین فراوانی بودند. در مطالعه حاضر، بیشترین درصد موتیف‌های تک نوکلئوتیدی از نوع A/T، بیشترین موتیف‌های دی نوکلئوتیدی از نوع AG/CT (۶۸/۴۶ درصد) و بیشترین موتیف‌های ۴،۳ و ۵ نوکلئوتیدی نیز به ترتیب از نوع AAG/CTT، AAAT/ATTT و AAAAG/CTTTT بودند. شناسایی مسیرهای فعال بیولوژیکی مرتبط با یونی‌ژن‌های دارای نشانگر با استفاده از پایگاه داده گیاهی انجام شد و بیشتر یونی‌ژن‌ها در مسیرهای انتقال سیگنال، متابولیسم پروتئین‌ها، چرخه سلولی، و بیان ژن (رونویسی) قرار گرفتند. بررسی مسیرهای غنی شده با یونی‌ژن‌های حاوی نشانگر نیز نشان داد که بیشترین فراوانی مربوط به مسیرهای غشا، هسته و فرآیند متابولیسمی ترکیبات حاوی نوکلئوبازها است.

نتیجه گیری: این مطالعه مجموعه‌ای از EST-SSRهای مفید را ارائه می‌کند و می‌تواند زمینه تحقیقات عملکردی و مقایسه‌ای ژنومی بیشتر را بر روی دارمازو فراهم کند.

کلمات کلیدی: دارمازو، نشانگر EST-SSR، RNA-Seq

نوع مقاله: پژوهشی.

استناد: جودکی فروغ، اسماعیلی احمد، سهرابی سید سجاد (۱۴۰۵) شناسایی نواحی حاوی نشانگرهای EST-SSR در ترنسکریپتوم بلوط دارمازو (*Quercus infectoria*) با استفاده از فن‌آوری RNA-Seq. *مجله بیوتکنولوژی کشاورزی*، ۱۸(۴)، ۳۳-۴۴.

Publisher: Shahid Bahonar University of Kerman & Iranian

Biotechnology Society.

© the authors



مقدمه

گونه *Quercus infectoria* گیاهی با شاخه‌های فراوان که اغلب با نام بلوط گالزا (Amilah et al., 2022) یا دارمازو (Alaaddin Ahmed et al., 2022) شناخته می‌شود، یک گیاه دارویی متعلق به خانواده Fagaceae است که عمدتاً در یونان، آسیای صغیر و ایران یافت می‌شود (Shrestha et al., 2014). در ایران، جنگل‌های زاگرس به ویژه استان‌های آذربایجان غربی، کردستان و لرستان از زیستگاه‌های اصلی این گونه گیاهی به شمار می‌رود (Askari et al., 2020). این گونه علاوه بر مزایای زیست محیطی، دارای ارزش‌های اجتماعی مختلفی نیز می‌باشد. به عنوان مثال از چوب آن برای ساخت بسیاری از ابزارها استفاده می‌شود و از برگ‌های آن در زمستان به عنوان خوراک دام و از شاخه‌های آن در تابستان برای ساخت سرپناه استفاده می‌شود (Alaaddin Ahmed et al., 2022). این گیاه در بسیاری از نقاط جهان در درمان بیماری‌های مختلف استفاده می‌شود. قسمت‌های مختلف این گیاه دارای خواص معطر و دارویی بوده و حاوی ترکیبات شیمیایی متنوعی مانند آلکالوئیدها، فلاونوئیدها، تانن‌ها و ترکیبات فنولیک هستند. مطالعات نشان داده است که این ترکیبات دارای اثرات هم‌افزایی^۱ هستند بطوریکه ترکیب دو یا نوع از این متابولیتها، اثری قوی‌تر، گسترده‌تر یا متفاوت‌تر از مجموع اثرات کاربرد هر کدام به تنهایی ایجاد می‌کند (Amilah et al., 2022). گال‌های دارمازو سرشار از تانن‌ها و ترکیبات فنولی هستند و ارزش دارویی این گیاه به دلیل فعالیت‌های ضد دیابت، ضد لرزش^۲ و

¹ synergistic effects

² anti tremorin

ضد التهاب و طیف گسترده‌ای از خواص ضد میکروبی مانند فعالیت‌های ضد باکتریایی، ضد ویروسی و ضد قارچی ترکیبات آن است که به عنوان پتانسیل‌های درمانی گال‌ها برجسته شده است. فعالیت‌های متنوع ضد میکروارگانیسم این گیاه به‌ویژه گال‌های آن محققان را به مطالعه بیشتر بر روی آن تشویق کرده است (Zin et al., 2019). به طور کلی گونه‌های مختلف بلوط بر خلاف ارزش اقتصادی و اهمیت زیست محیطی که دارند در مقایسه با سایر درختان جنگلی در مطالعات ژنتیکی کمتر مورد توجه قرار گرفته اند (Alaaddin Ahmed et al., 2022). استراتژی‌های اصلاحی از جمله رویکردهای مولکولی پتانسیلی برای بررسی‌های بیشتر گونه‌های گیاهی ارائه می‌دهند. دانشمندان از فن‌آوری‌های پر بازده که به‌طور گسترده‌ای با عنوان NGS³ نامیده می‌شوند، در جهت مطالعه گونه‌های غیر مدل و گونه‌هایی با اطلاعات ژنتیکی کم یا فاقد اطلاعات ژنتیکی استفاده می‌کنند (da Fonseca et al., 2016). این تکنیک پرکاربرد از طریق تجزیه و تحلیل گسترده ترنسکرپتوم⁴ و با تولید داده‌ها به‌صورت دقیق و با توان بالا، بینشی در مورد موجود زنده ارائه می‌دهد. در حال حاضر، از توالی‌یابی RNA به طور گسترده به عنوان یک رویکرد مهم برای تجزیه و تحلیل سطح بیان ژن و توسعه نشانگرهای مولکولی استفاده می‌شود (Zhou et al., 2022). توالی‌یابی ترنسکرپتوم پر بازده⁵ نه تنها به دلیل تولید حجم عظیمی از داده‌های توالی‌یابی شده به منظور شناسایی نشانگرها، بلکه بیشتر به این دلیل که نشانگرهای کشف شده مبتنی بر ژن هستند می‌تواند منابع برتری برای توسعه چنین نشانگرهایی باشند. چنین نشانگرهایی به دلیل تسهیل در تشخیص تنوع عملکردی در مطالعات ژنتیکی سودمند هستند (Torre et al., 2014). استفاده از این فن‌آوری برای توسعه نشانگرهای مولکولی به‌ویژه برای گونه‌های غیرمدل و بدون ژنوم مرجع مزایای بسیاری دارد (Zhou et al., 2022). نشانگرهای SSR نشانگرهای هم‌توان (codominant) با چندشکلی‌های فراوان در بسیاری از ژنوم‌های یوکاریوتی هستند. این نشانگرها به طور گسترده در محصولات عمده برای آنالیزهای ژنتیکی و ژنومی مانند انگشت نگاری DNA، نقشه برداری پیوند ژنتیکی، ارزیابی تنوع ژنتیکی، نقشه برداری QTL و استخراج ژن استفاده می‌شوند (Wang et al., 2020). این تکرارهای پشت سر هم کوتاه که به‌عنوان ریزماهواره‌ها شناخته می‌شوند و به طور گسترده به‌عنوان نشانگرهای مولکولی انتخابی برای مطالعات ژنتیکی مورد استفاده قرار می‌گیرند (Mohammadabadi et al., 2010) تکرارهای پشت سر هم موتیف‌های DNA کوتاه با یک تا شش جفت باز هستند که به طور گسترده در یوکاریوت‌ها و همچنین در برخی از پروکاریوت‌ها یافت می‌شوند (Zhou & Zhu, 2020). نشانگرهای SSR به دلیل داشتن ویژگی‌هایی از جمله ماهیت چند آلی، تکرارپذیری، پوشش وسیع ژنومی و فراوانی زیاد، اهمیت قابل توجهی در ژنتیک و اصلاح گیاهان پیدا کرده‌اند (Liu et al., 2018) و یکی از پرکاربردترین نشانگرها به‌شمار می‌روند (Sohrabi et al., 2018). بیان ژن‌های یوکاریوتی به‌صورت زمانی و مکانی از طریق مکانیسم‌های کنترلی چندبعدی تنظیم می‌شود (Hajalazadeh et al., 2021) و تنها زیرمجموعه‌ی محدودی از کل ژنوم در هر نوع بافت به‌صورت فعال بیان می‌گردد. همچنین بیان ژن‌ها به مراحل رشد وابسته است (Khabiri et al., 2025) بنابراین، الگوهای بیان ژن در یوکاریوت‌ها به‌صورت اختصاصی بافت‌محور تنظیم می‌شوند و هر بافت بیان ژنی متمایز از خود نشان می‌دهد (Farahvashi et al., 2026). علاوه بر این، سطوح محصولات ژنی سنتز شده در یک بافت معین، و همچنین آن‌هایی که توسط بافت‌های دیگر تأمین می‌شوند، به‌طور جمعی بیان ژن را تنظیم می‌کنند (Shokri et al., 2023). تغییرات اپی‌ژنتیکی به‌عنوان مکانیسمی حیاتی عمل می‌کنند که از طریق آن عوامل

³ NGS: Next-Generation Sequencing

⁴ transcriptome

⁵ High-throughput transcriptome sequencing

محیطی اثری ماندگار بر ژنوم به جا می‌گذارند. این تغییرات می‌توانند اثری باشند و بر نسل‌های بعدی بدون تغییر در توالی DNA تأثیر بگذارند (Mohammadabadi et al., 2025). گروهی از نشانگرهای ریزماهوره‌ای با نام EST-SSR، در نواحی کدکننده ژنی قرار گرفته‌اند که ممکن است با نواحی عملکردی مرتبط باشند و در ارزیابی تکامل ژنی که روی آن قرار دارند موثر باشند (Hosseini et al., 2018; Sajad Sohrabi et al., 2018). EST-SSR ها در انتخاب به‌وسیله نشانگرها کارایی بالایی داشته (Sohrabi et al., 2018) و از طرفی حفاظت شده‌تر بودن نواحی بالا و پایین دست توالی آنها، این نشانگرها را در بررسی تنوع ژنتیکی بین گونه‌ها نسبت به سایر نشانگرهای ریزماهوره ای ژنومی کارآمدتر می‌سازد (Hosseini et al., 2018). برای گونه‌هایی که توالی ژنومی آنها و یا نقشه فیزیکی از آنها در دسترس نیست، اطلاعات به‌دست‌آمده از EST ها وسیله‌ای عملی برای کشف ژن و راهی برای روشن کردن فیزیولوژی و عملکرد ژنوم آنها است (Pereira-Leal et al., 2014). توالی‌یابی ترنسکرپتوم روشی مقرون به‌صرفه و سازنده برای توسعه EST-SSR، به‌ویژه برای گونه‌های بدون ژنوم مرجع ارائه کرد (Sun et al., 2023). در مطالعه بر روی گیاه *Camellia japonica* با استفاده از داده‌های حاصل از توالی‌یابی RNA و ترنسکرپتوم نوپدید مجموعه‌ای از نشانگرهای SSR شناسایی شد که این نتایج اطلاعات ضروری را برای مطالعات بیشتر از جمله مطالعات تاکسونومی، تجزیه و تحلیل گوناگونی^۶ و نقشه برداری همبستگی ژنتیکی^۷ مبتنی بر MAS و SSR را در گیاهان کاملیا فراهم کرده است (Li et al., 2021). Sohrabi et al. (2018) از داده‌های حاصل از RNA-Seq در جهت شناسایی نشانگرهای EST-SSR در گیاه عدس تحت تنش سرما استفاده کردند و در نتیجه ۸۹۰۵ مکان ریزماهوره در ۷۲۱۱ یونی‌ژن حاصل از توالی‌یابی شناسایی شد. مطالعه بر روی گیاه عدس^۸ باعث شناسایی ۱۰۵۴۷ یونی‌ژن حاوی حداقل یک توالی نشانگر EST-SSR شد و در شناسایی EST-SSR های مرتبط با تحمل به تنش خشکی نتایج نشان دادند که ژن‌های حاوی این نشانگرها در اعمال حیاتی مهمی نقش دارند (Hosseini et al., 2018). بررسی نشانگرها در گیاه رودومیرتوس^۹ به‌عنوان یک گیاه دارویی و زینتی با استفاده از توالی‌یابی ترنسکرپتوم انجام شد که در نهایت منجر به شناسایی نشانگر SSR و یونی‌ژن‌های دارای این نشانگرها شد که در آنها ۳۵۴۱ یونی‌ژن حاوی بیش از یک SSR بودند (Sun et al., 2023). بررسی نشانگرها در برخی از گونه‌های *Quercus* نیز انجام شده است. در مطالعه‌ای بر روی بلوط چوب پنبه به منظور تعیین رابطه ژنتیکی و انگشت نگاری DNA^{۱۰} این درختان از نشانگرهای SSR و ISSR استفاده شد (López-Aljorna et al., 2007). توالی‌یابی RNA با استفاده از پلتفرم ایلومینا و سرهم‌بندی نوپدید برای نمونه‌های حاصل از برگ‌های گونه *Quercus pubescens* اطلاعات کاملی در مورد توالی‌های بیان‌شده در بافت برگ نشان داد که منجر به شناسایی ۱۴۲۰۲ نشانگر SSR و ۱۸۴۲۵ نشانگر SNP شد (Torre et al., 2014). توالی‌یابی RNA^{۱۱} بر روی دو گونه بلوط *Q. kerrii* و *Q. austrocochinchinensis* با ادغام بافت‌های برگ، ریشه و ساقه‌ی جدید و قدیمی انجام شد که مجموعه‌ای مفید از نشانگرهای SSR را به منظور تحقیقات عملکردی و مقایسه‌ای ژنومی روی زیرجنس *Quercus* ارائه می‌کند (An et al., 2016). با توجه به اینکه داده‌های حاصل از توالی‌یابی RNA می‌تواند منبع خوبی جهت شناسایی EST-SSR باشد،

⁶ diversity⁷ linkage mapping⁸ *Lens culinaris*⁹ *Rhodomyrtus tomentosa*¹⁰ DNA- Fingerprinting¹¹ RNA-Seq

این تحقیق یک منبع ترنسکرپتومی ارزشمند و پایه ای برای مطالعات عملکردی و مقایسه‌ای آینده فراهم کرده و برای اولین بار مجموعه‌ای از نواحی حاوی نشانگرهای EST-SSR را برای گونه دارویی مهم *Quercus infectoria* ارائه می‌دهد.

مواد و روش‌ها

مواد گیاهی، استخراج RNA، توالی‌یابی: جمع‌آوری نمونه‌های برگ و ریشه از درختان دارمازو در منطقه شینه قلایی از جنگل‌های زاگرس واقع در بخش فیروزآباد از توابع شهرستان الشتر و نهال‌های دو ساله حاصل از کاشت بذر درختان این منطقه در نهالستان خرم‌آباد صورت گرفت. در مجموع ۶ نمونه از دو بافت برگ و ریشه برای توالی‌یابی انتخاب شد به طوریکه هر یک از گروه‌های برگ و ریشه شامل ۳ تکرار و هر تکرار از ادغام پنج نمونه مختلف تشکیل شده بود. سپس RNA مورد نیاز به روش جمی چات چائو (DJAMI-TCHATCHOU & STRAKER, 2012) استخراج شد. کیفیت و کمیت نمونه‌های استخراجی با استفاده از اسپکتروفتومتر Pico200 Microliter و الکتروفورز ژل ۱ درصد مورد بررسی قرار گرفت. از Agilent Bio TruSeq RNA Library Analyzer 2100 به منظور تأیید کیفیت نمونه‌ها استفاده و کتابخانه‌های cDNA با کمک کیت TruSeq RNA Library Preparation Kit و طبق دستورالعمل شرکت سازنده ساخته شد. کیفیت کتابخانه‌ها توسط Agilent Bioanalyzer 2100 system مورد بررسی قرار گرفت و در نهایت فرآیند توالی‌یابی با استفاده از پلتفرم Illumina HiSeq2500 و در موسسه بیوانفورماتیک Novogene (پکن، چین) انجام شد.

کنترل کیفیت و سرهم‌بندی نوپدید خوانش‌های کوتاه: پس از توالی‌یابی، به منظور فیلتر خوانش‌های خام از نرم‌افزار FastQC، برای حذف خطاهای احتمالی در توالی‌های ایجاد شده از نرم‌افزار RCorrector و حذف توالی‌های آداپتور از ابزار Cutadapt (v.0.9.5) استفاده شد. تمام خوانش‌ها با استفاده از نرم‌افزار FastQC نسخه v0.11.6 به منظور کنترل کیفیت پردازش شدند. سپس از نرم‌افزار Trimmomatic (v.0.36) برای پیرایش داده‌ها (با میانگین Phred score کمتر از ۳۰) استفاده شد. پس از کیفیت‌سنجی خوانش‌های کوتاه^{۱۲} سرهم‌بندی نوپدید^{۱۳} برای خوانش‌های باکیفیت بالا با استفاده از نرم‌افزار CLC Genomics Workbench (v7.5.0) انجام شد.

شناسایی نواحی حاوی نشانگرهای SSR: از نرم‌افزار MISA در جهت شناسایی نشانگرهای ریزماهواره‌ای استفاده و حداقل تکرارهای ۱۰، ۵، ۵، ۵ و ۵ به ترتیب برای نشانگرهای مونو، دی، تری، تترا، پنتا و هگزا نوکلئوتیدی تعیین شد. حداکثر فاصله بین دو مکان نشانگر روی یک توالی ۱۰۰bp در نظر گرفته شد.

شناسایی کارکرد یونی‌ژن‌ها: شناسایی کارکرد و تفسیر عملکردی یونی‌ژن‌های حاوی نشانگرهای EST-SSR با استفاده از نرم‌افزار NCBI Blast (v2.6.0) انجام شد و یونی‌ژن‌های حاوی این نشانگرها در مقابل پایگاه داده توالی‌های پروتئینی غیرتکراری NCBI (NCBI non-redundant (Nr) protein sequences database) با در نظر گرفتن حد آستانه معادل $E\text{-Value} \leq 10^{-5}$ مورد Local BlastX قرار گرفتند. هستی‌شناسی ژن‌ها (GO) به منظور گروه‌بندی کارکردی ژن‌ها انجام شد. برای این منظور، بهترین بلاست (بهترین هم‌ترازی بین توالی‌های موردنظر و توالی‌های موجود در پایگاه‌ها) از پایگاه داده NR به

¹² Short Reads

¹³ De novo assembly

Blast2GO وارد شد. مسیرهای بیولوژیکی فعال مرتبط با یونیژن‌های حاوی EST-SSR با نقشه برداری این ژن‌ها در برابر پایگاه داده Plant Reactome، شناسایی شدند.

نتایج و بحث

توالی‌یابی: از توالی‌یابی توسط پلتفرم Illumina HiSeq 2500، ۲۹۲۹۷۲۰۴۸ خوانش ایجاد شد. پس از حذف خوانش‌های با کیفیت پایین در طی فرآیند پردازش^{۱۴}، ۱۵۹۸۴۱۶۳۸ و ۱۳۳۱۳۰۴۱۰ خوانش باکیفیت به ترتیب از نمونه‌های برگ و ریشه و در مجموع ۲۷۶،۸۵۲،۲۵۵ خوانش تمیز برای تجزیه و تحلیل بیشتر به دست آمد. بررسی کیفیت رونوشت‌ها نشان داد که کیفیت پایه به طور متوسط برای ۹۴ درصد از خوانش‌ها بالاتر از Q30 است.

سرهم‌بندی نوپدید: به دلیل عدم وجود ژنوم مرجع، سرهم‌بندی نوپدید برای ایجاد یک ترنسکریپتوم مرجع استفاده شد. پس از حذف توالی‌های اضافی و تکراری، ۹۶۲۲۵ یونیژن با محتوای GC معادل ۴۵/۶۳۶ درصد، طول متوسط ۹۶۱/۱۷۶ جفت باز و پارامتر N50 با طول ۱۱۹۷ جفت باز تولید شد. طول ۳۵/۸۰ درصد یونیژن‌ها بیش از ۱ kb بود و طول ۵ یونیژن بیشتر از ۱۰ kb بود. کوتاه‌ترین و طولانی‌ترین یونیژن‌ها به ترتیب ۲۰۰ و ۲۴۰۵۶ جفت باز بود (جدول ۱). پروفایلی از ترنسکریپتوم برای برخی از گونه‌های بلوط تهیه شده است که تعداد یونیژن‌های به دست آمده در آن مطالعات برای گونه *Q. brantii* (Safari et al., 2022) و برای گونه *Q. rubra* (Soltani et al., 2020) ۵۲۶۶۲ یونیژن بوده است. تعداد یونیژن‌های حاصل از سرهم‌بندی نوپدید برای *Q. ilex* توسط پلتفرم ایلومینا و مونتاژ کننده ترینیتی ۷۷۱۵۹ یونیژن بود (Guerrero-Sanchez et al., 2019). تعداد رونوشت ایجاد شده در سرهم‌بندی نوپدید برای دو گونه *Q. austrocochinchinensis* و *Q. kerrii* به ترتیب ۷۹۳۱۲ و ۸۱۹۲۱ بود (An et al., 2016). مقایسه تعداد رونوشت‌ها و N50 به دست آمده در این تحقیق با نتایج سرهم‌بندی نوپدید گونه‌های مشابه نشان داد که نتایج به دست آمده قابل قبول بوده و می‌توان از آن برای تجزیه و تحلیل‌های بعدی استفاده کرد.

نتایج بلاست x ترنسکریپتوم مرجع دارمازو در مقابل پایگاه داده‌های پروتئین‌های غیر تکراری و هستی

شناسی یونیژن‌ها: نتایج بلاست یونیژن‌ها در برابر پایگاه داده پروتئین‌های غیر تکراری (NR) نشان داد که از مجموع یونیژن‌های به دست آمده، ۶۴۰۸۴ یونیژن دارای رکوردی مشابه در این پایگاه بودند. نتایج بلاست سرهم‌بندی در مقابل پایگاه NR به منظور شناسایی گونه‌هایی با بیشترین تشابه به دارمازو نشان داد که بیشترین میزان شباهت به ترتیب با گونه‌های بلوط قرمز اروپایی^{۱۵} با ۱۶۷۸۰ یونیژن مشابه و گونه *Quercus lobata* با ۱۲۲۱۰ یونیژن مشابه مشاهده شد. پس از آن‌ها نیز بیشترین میزان شباهت با گونه‌های *Quercus suber*، *Carpinus fangiana* و *Quercus rubra* بود (شکل ۱). بیشترین میزان شباهت با چهار گونه بلوط و همچنین با *C. fangiana* که به راسته راش‌سانان تعلق دارد، می‌تواند دلیلی بر درستی سرهم‌بندی انجام شده باشد. در مجموع ۱۵۹۶۴ رونوشت با شماره کمیسیون آنزیم^{۱۶} حاشیه‌نویسی^{۱۷} شدند. این آنزیم‌ها در ۷ کلاس آنزیمی شامل: ترانسفراز (۵۶۷۸ یونیژن)، هیدرولاز (۵۳۵۳)، اکسیدوردوکتاز (۲۵۱۳)، ترنسلوکازها (۶۶۱)، لیگازها (۶۰۱)، ایزومرازا (۵۹۵) و لیازاها (۵۶۳)، معرفی

¹⁴ Trimming

¹⁵ *Quercus robur*

¹⁶ Enzyme Commission

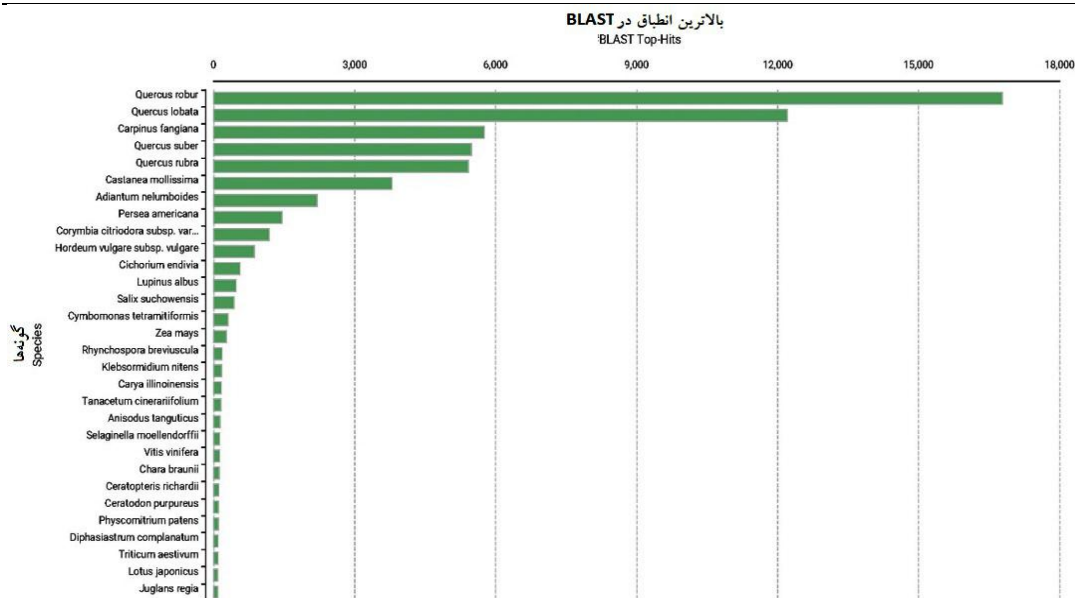
¹⁷ Annotation

شدند (شکل ۲). وجود متابولیت‌های ثانویه مختلف در دارمازو در مطالعات مختلف مورد بررسی قرار گرفته است (Kaur et al., 2004). در توزیع آنزیمی، وجود سه کلاس آنزیمی ترانس فرازا، هیدرولازها و اکسیدوردوکتازها می‌تواند نشان‌دهنده ژن‌های مرتبط با مسیرهای بیوسنتز متابولیت‌های ثانویه باشد (Torre et al., 2014).

جدول ۱. ساختار و معیارهای ارزیابی کیفیت سرهم‌بندی ترنسکریپتوم

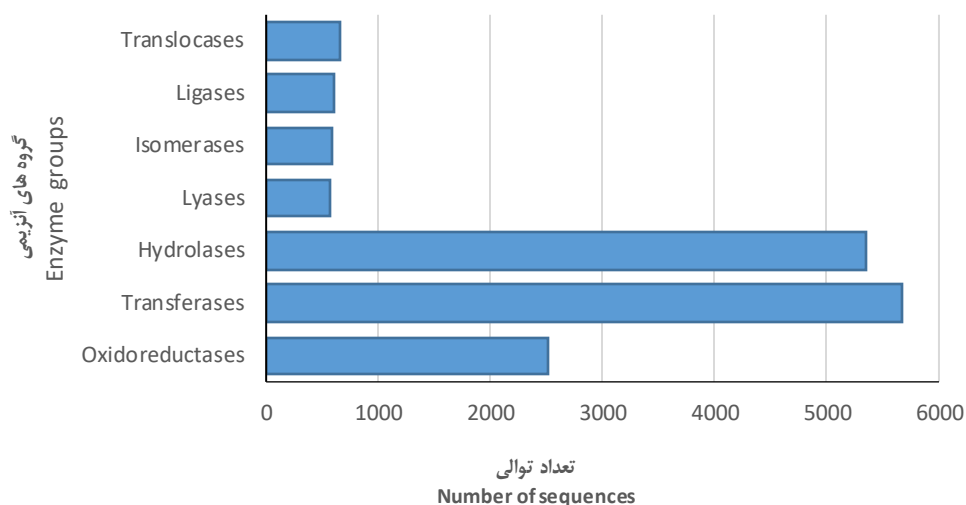
Table 1. Assembly structure and the evaluation metrics of transcriptome assembly quality

Value ارزش	Parameters پارامتر
96225	Number of transcript in the assembly تعداد رونوشت‌های موجود در سرهم‌بندی
92491328	Number of bases included in the assembly تعداد بازهای اسمبلی
200	Shortest transcript (bp) کوتاه‌ترین رونوشت (bp)
24056	Largest transcript (bp) بزرگترین رونوشت (bp)
961.17672	Mean length of the transcript (nt) میانگین طول رونوشت (nt)
34455	transcript >1000 (bp) تعداد رونوشت بیشتر از ۱۰۰۰ b
5	transcript >10000 (bp) تعداد رونوشت بیشتر از ۱۰۰۰۰ b
74651	Number of transcript that had an open reading frame (orf) تعداد رونوشت دارای قاب خوانش باز (orf)
490	N90
1197	N50
1655	N30
2674	N10
45.636	GC% درصد محتوای GC



شکل ۱. توزیع گونه‌هایی با بیشترین شباهت یافته شده در بلاست X سرهم‌بندی مرجع در مقابل پایگاه پروتئین‌های Nr

Figure 1. The results of species distribution by BLAST Top-hit based on BLASTX against the Nr-database



شکل ۲. توزیع کدهای آنزیمی و گروه‌بندی در هفت کلاس آنزیمی

Figure 2. Enzyme code (E.C.) class distribution and grouping into enzyme classes

به منظور توصیف عملکردی یونی‌ژن‌ها، حاشیه‌نویسی^{۱۸} عملکردی GO انجام شد و یونی‌ژن‌ها در سه بخش اصلی با عنوان فرآیند بیولوژیکی^{۱۹}، اجزاء سلولی^{۲۰} و عملکردهای مولکولی^{۲۱} طبقه‌بندی شدند که در شکل ۳ نشان داده شده‌اند. نتایج آزمون فیشرو نمودار هستی‌شناسی اینترنتی AgriGO نشان داد در بخش فرآیندهای بیولوژیکی گروه‌های غالب مربوط به "فرآیند متابولیک اولیه"^{۲۲} (GO:0044238) و "فرآیند متابولیکی ماکرومولکول‌ها"^{۲۳} (GO:0008152) بودند. در رده عملکرد مولکولی زیرگروه -های "اتصال ترکیب حلقوی آلی"^{۲۴}، "فعالیت ترانسفراز"^{۲۵}، و "اتصال پروتئین" بیشترین یونی‌ژن‌ها را به خود اختصاص دادند. در رده اجزای سلولی نیز بیشتر یونی‌ژن‌ها متعلق به "ساختار تشریحی داخل سلولی"^{۲۶} (GO:0005622)، "سیتوپلاسم" و "سیتوسول" بودند. پروژه هستی‌شناسی اصطلاحات تعریف شده‌ای را ارائه می‌دهد که ویژگی‌های محصولات ژنی را نشان داده و ۳ حوزه را پوشش می‌دهد: جزء سلولی (بخش‌هایی از یک سلول یا محیط خارج سلولی آن)، عملکرد مولکولی (فعالیت‌های عنصری یک محصول ژن در سطح مولکولی، مانند اتصال یا کاتالیز)، و فرآیند بیولوژیکی (عملیات یا مجموعه‌ای از رویدادهای مولکولی با شروع و پایان مشخص)، که در واقع فعالیت‌هایی هستند که مربوط به عملکرد واحدهای زنده یکپارچه‌ای مثل سلول‌ها، بافت‌ها، اندام‌ها و موجودات می‌شوند (Kim KiUk et al., 2013). در آنالیز GO در بخش فرآیند بیولوژیکی، زیرگروه "فرآیندهای متابولیک سلولی"^{۲۷} در میان گروه‌هایی با یونی‌ژن‌های بالا بودند که می‌تواند به رشد سریع و فعالیت‌های متابولیکی گسترده در برگ‌ها به‌ویژه برگ‌های

¹⁸ Annotation

¹⁹ Biological Process: BP

²⁰ Cellular Component: CC

²¹ Molecular Function: MF

²² Primary metabolic process

²³ Macromolecule metabolic process

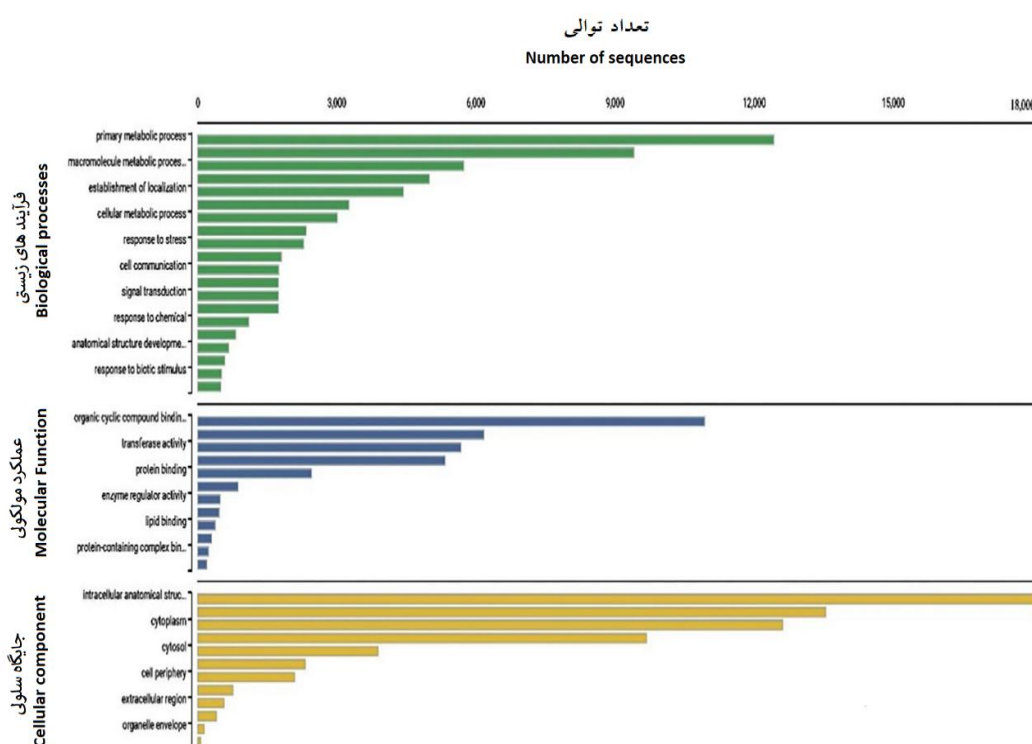
²⁴ Organic cyclic compound binding

²⁵ Transferase activity

²⁶ Intracellular anatomical structure

²⁷ Cellular metabolic processes

جوان مرتبط باشد (Palumbo et al., 2018). به‌طور کلی این یافته‌ها نشان داد که کارکردهای اصلی رونوشت‌های مستند سازی شده مسئول تنظیم اساسی بیولوژیکی و متابولیسم رایج در گیاهان بودند.



شکل ۳. هستی‌شناسی ژن‌ها. طبقه‌بندی یونی‌ژن‌ها در سه گروه اصلی فرآیند بیولوژیکی (BP)، عملکردهای مولکولی (MF) و اجزاء سلولی (CC)

Figure 3. Gene Ontology (GO) functional classification. Unigenes were classified in three major classes as biological process (BP), molecular functions (MF), and cellular component (CC)

شناسایی نواحی حاوی نشانگرهای EST-SSR و تعیین کارکرد رونوشت‌های حاوی نشانگر: به‌طور کلی

تعداد ۲۰۵۰۵ مورد نشانگر EST-SSR در ۱۳۳۶۸ یونی‌ژن شناسایی شد که این تعداد یونی‌ژن ۱۳/۸۹ درصد از کل یونی‌ژن‌ها را شامل می‌شوند. از این تعداد ۶۷/۴۲ درصد (۹۰۱۳ یونی‌ژن) حاوی تنها یک EST-SSR بودند و ۴۳۵۵ یونی‌ژن بیش از یک مکان نشانگر را دارا بودند. از مجموع کل یونی‌ژن‌های شناسایی شده حاوی EST-SSR، تعداد ۲۹۶۸ یونی‌ژن حاوی موتیف‌های پیچیده و ترکیبی بودند. RNA-Seq به‌طور گسترده در گیاهان مدل و غیرمدل به عنوان یک روش موثر برای به دست آوردن تعداد زیادی از توالی‌های EST استفاده شده است و برای توسعه نشانگرهای مولکولی ضروری است (Wang et al., 2019). در مطالعه بر روی گیاه رودومیترتوس توالی‌یابی ترنسکرپتوم برای به دست آوردن EST-SSR در بافت‌های مختلف گیاهی انجام که منجر به شناسایی ۱۸۸۷۹ نشانگر SSR در ۱۴۱۳۲ یونی‌ژن شد که در آنها ۳۵۴۱ یونی‌ژن حاوی بیش از یک SSR بودند (Sun et al., 2023). بررسی ترنسکرپتوم با استفاده از فن‌آوری RNA-Seq در گیاه مگنولیا^{۲۸} منجر به شناسایی ۸۳۰۶ یونی‌ژن حاوی ۹۵۶۷ نشانگر شد

²⁸ *Magnolia wufengensis*

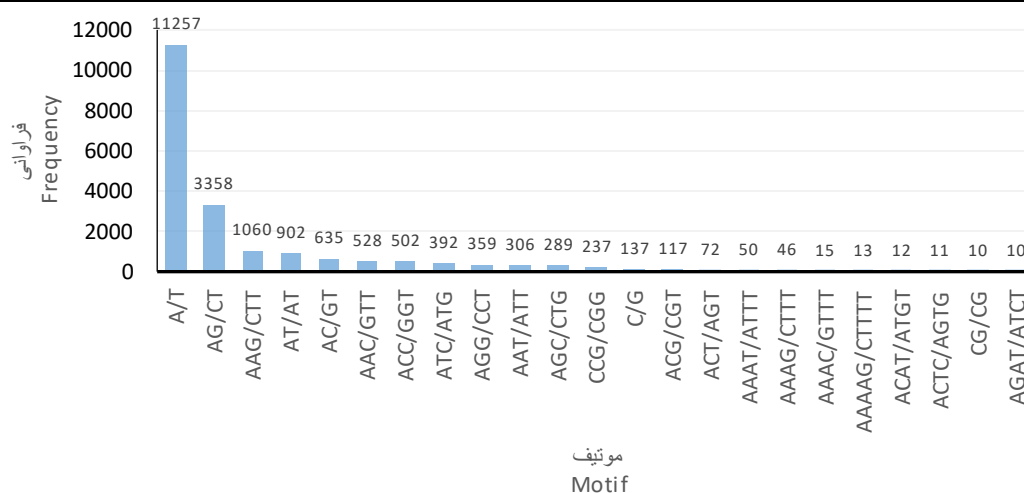
که ۱۰۴۴ (۱۲.۶٪) یونی ژن بیش از یک SSR داشتند و ۶۵۳ رونوشت موتیف‌های ترکیبی را ارائه کردند (Wang et al., 2019). در آنالیز ترنسکرپتوم برگ‌های گونه *Q. pubescens* در مجموع ۱۴۲۰۲ SSR در ۹۶۰۰۶ رونوشت مونتاژ شده شناسایی شد که از این تعداد، ۳۳۶۶ SSR ترکیبی بودند (Torre et al., 2014). فراوانی موتیف‌های مختلف در جدول ۲ نشان داده شده است. در این میان، موتیف‌های تک (مونو) نوکلئوتیدی بیشترین فراوانی را در بین نشانگرها به خود اختصاص دادند (۱۱۳۹۴، ۵۵/۵۷ درصد) و پس از آن موتیف‌های دو و سه نوکلئوتیدی به ترتیب با ۲۳/۹۲ و ۱۸/۸۲ درصد دارای بیشترین فراوانی بودند. در مطالعه Torre et al. (2014) بر روی *Q. pubescens* و بررسی موتیف‌های ۲ تا ۶ نوکلئوتیدی، بیشترین فراوانی برای موتیف‌های دو نوکلئوتیدی بود (۴۹/۳ درصد) (Torre et al., 2014). در گیاه مگنولیا نیز در بین موتیف‌های ۲ تا ۶ نوکلئوتیدی، موتیف‌های ۲ نوکلئوتیدی بیشترین فراوانی را داشتند (۵۰/۸) (Wang et al., 2019). در دو گونه *Q. austrocochinchinensis* و *Q. kerrii* به ترتیب ۱۳۷۶۲ و ۱۳۴۳۰ جایگاه SSR شناسایی شد که بعد از حذف نشانگرهای تک نوکلئوتیدی، موتیف‌های ۲ نوکلئوتیدی فراوانی بالایی را نشان دادند (An et al., 2016). در بررسی ترنسکرپتوم در دو گونه بلوط *Q. liaotungensis* و *Q. mongolica*، ۱۶۶۰۸ نشانگر SSR در ۱۰۷۴۴ توالی در یک کتابخانه ترکیبی از یونی ژن‌های دو گونه شناسایی شد که در این مطالعه نیز بیشترین تعداد نشانگرها مربوط به موتیف‌های ۲ نوکلئوتیدی بود (Sun et al., 2019). مطالعه SSR ها بر روی گیاه ردودندرون^{۲۹} نیز با استفاده از فن آوری RNA-Seq انجام شد و نشانگرهای تک نوکلئوتیدی تا ۶ نوکلئوتیدی شناسایی شدند که در این بررسی نشانگرهای مونونوکلئوتیدی بیشترین میزان فراوانی نشانگرها را به خود اختصاص دادند (Zhou & Zhu, 2020). در مطالعه حاضر، ۹۸/۷۹ درصد موتیف‌های تک نوکلئوتیدی از نوع A/T، و بیشترین موتیف‌های دی نوکلئوتیدی از نوع AG/CT (۶۸/۴۶ درصد) و نادرترین آنها از نوع CG/CG بود. بیشترین موتیف‌های ۳ و ۴، ۵ نوکلئوتیدی نیز به ترتیب از نوع AAG/CTT، AAAT/ATTT و AAAAG/CTTT بودند (شکل ۴). در مطالعه بر روی دو گونه *Q. austrocochinchinensis* و *Q. kerrii* نیز نتایج مشابه دیده شد و بیشترین موتیف‌های ۲ نوکلئوتیدی و نادرترین آن‌ها به ترتیب از نوع AG/CT و CG/CG بود (An et al., 2016). در گیاه چای^{۳۰} نیز نشانگرهای ۲ نوکلئوتیدی، فراوانی بیشتری نسبت به سایر نشانگرها داشتند و موتیف‌های AG/CT، AAAT/ATTT و AAT/ATT رایج‌ترین نوع موتیف‌های ۳، ۲ و ۴ نوکلئوتیدی شناسایی شده بودند (Liu et al., 2018). در گیاه ردودمیرتوس^{۳۱} نیز فراوان‌ترین موتیف‌ها از نوع A/T، AG/CT، AAG/CTT و AAAT/ATTT بود (Sun et al., 2023). شواهد فراوانی نشان می‌دهد که SSR ها به طور غیر تصادفی در مناطق غیر کدکننده پروتئین، UTR ها و اینترون‌ها توزیع می‌شوند. گسترش SSR ها در مناطق کدکننده پروتئین می‌تواند منجر به افزایش یا از دست دادن عملکرد ژن از طریق جهش (تغییر در چهارچوب) یا تولید mRNA مضر شود (Li et al., 2004). با توجه به مطالعات پیشین بر روی گیاهان مختلف بیشترین تعداد نشانگرها مربوط به موتیف‌های ۲ و ۳ نوکلئوتیدی بود. در برخی از مطالعات بررسی تکرارهای مونونوکلئوتیدی نادیده گرفته شده که علت این امر می‌تواند ناشی از خطاهای توالی‌یابی و یا مشکل بودن تشخیص مونونوکلئوتیدها از پلی آدنیلایسون باشد (Taheri et al., 2018).

²⁹ *Rhododendron molle*³⁰ *Camellia sinensis*³¹ *Rhodomyrtus tomentosa*

جدول ۲. توزیع تکرارها / موتیف‌های مختلف نشانگرهای EST-SSR در ترانسکرپتوم دارمازو

Table 1. Distribution of different repeat/motifs of EST-SSR markers in the transcriptome of *Q. infectoria*

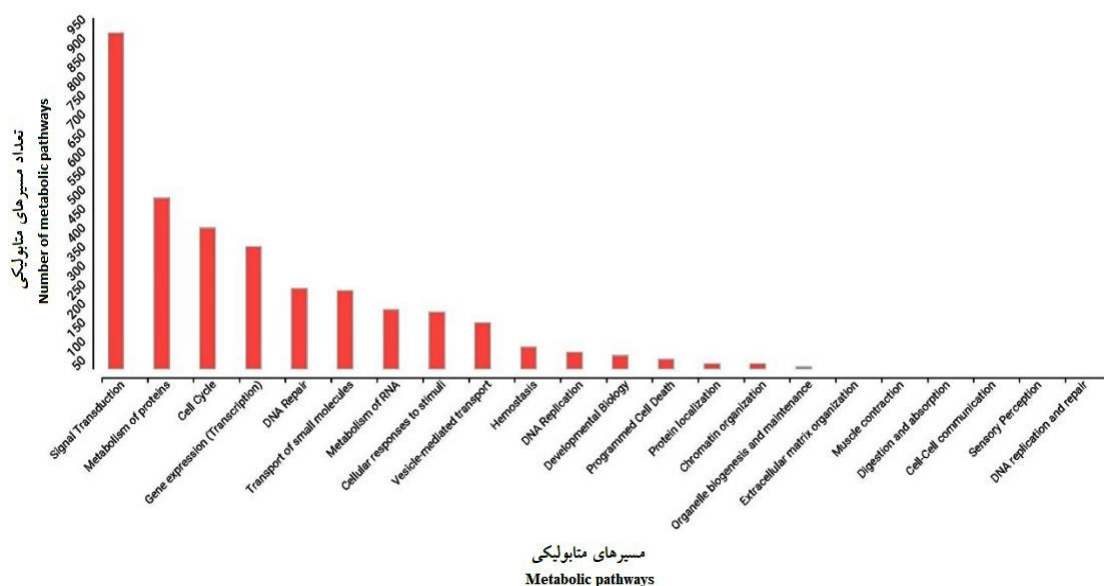
Repeats تعداد تکرار	Mono- تک نوکلئوتیدی	Di- دو نوکلئوتیدی	Tri- سه نوکلئوتیدی	Tetra- چهار نوکلئوتیدی	Penta- پنج نوکلئوتیدی	Hexa- شش نوکلئوتیدی	Total جمع کل	% درصد
5	-	-	223	10	3	3	2205	10.75
6	-	220	122	2	2	1	2176	10.61
7	-	130	84				1298	6.33
8	-	77	38				954	4.65
9	-	72	33	1	1		656	3.20
10	2666	33	14		1		3137	15.30
11	1645	29	7				2008	9.79
12	1163	15	2				1436	7.00
13	928	16					1143	5.57
14	930	8	2				1107	5.40
15	856	10					1007	4.91
16	729	7	1				827	4.03
17	620	5					687	3.35
18	435	6	1				482	2.35
19	312			1			346	1.69
20	245	3					263	1.28
21	198						216	1.05
21>	667	4274	3335	176	55	88	557	2.72
Total	11394	4905	3862	190	62	92	20505	
%	55.57	23.92	18.83	0.93	0.30	0.45	100.00	



شکل ۴. توزیع فراوانی موتیف‌های نشانگر ریزماهوره شناسایی شده در ترانسکرپتوم دارمازو

Figure 4. Frequency distribution of microsatellites marker motifs in *Q. infectoria* based on motif sequence type

در *Arabidopsis thaliana* و غلات تکرار AG/CT نسبت به سایر تکرارها (AT) فراوانی بالاتری را نشان می‌دهد. بر اساس مطالعات پیشین، برخی از موتیف‌های SSR با تأثیر بر شروع رونویسی یا متیلاسیون با بیان ژن مرتبط هستند (Sun et al., 2023). EST-SSR های مشتق شده از رونوشت‌ها، برای ارزیابی تنوع عملکردی و فنوتیپی در جمعیت‌های گیاهی مناسب تر هستند (Zhou et al., 2022). ارتباط SSR ها با تغییرات فنوتیپی آنها را به پرکاربردترین نشانگرهای مولکولی تبدیل کرده که برای کاربردهای مختلفی از جمله اصلاح ژنتیکی گیاهان، نقشه برداری ژن و انتخاب به کمک نشانگر ژنتیکی مورد استفاده قرار می‌گیرند (Zhou & Zhu, 2020). تجزیه و تحلیل مبتنی بر مسیرهای بیولوژیکی به منظور شناسایی یونی‌ژن‌های حاوی EST-SSR درگیر در مسیرهای مختلف بیوسنتزی جهت ارزیابی عملکردهای بیولوژیکی ژن‌ها و شناسایی مسیرهای فعال بیولوژیکی انجام شد. یونی‌ژن‌های حاوی EST-SSR در مقابل پایگاه داده Plant Reactome در مسیرهای بیولوژیکی مختلفی قرار گرفتند. بیشتر یونی‌ژن‌ها در مسیرهای انتقال سیگنال (۹۱۶ یونی‌ژن)^{۳۲}، متابولیسم پروتئین‌ها (۴۸۱ یونی‌ژن)^{۳۳}، چرخه سلولی (۴۰۲ یونی‌ژن)^{۳۴}، و بیان ژن (رونویسی) (۳۵۳ یونی‌ژن)^{۳۵} قرار گرفتند (شکل ۵).

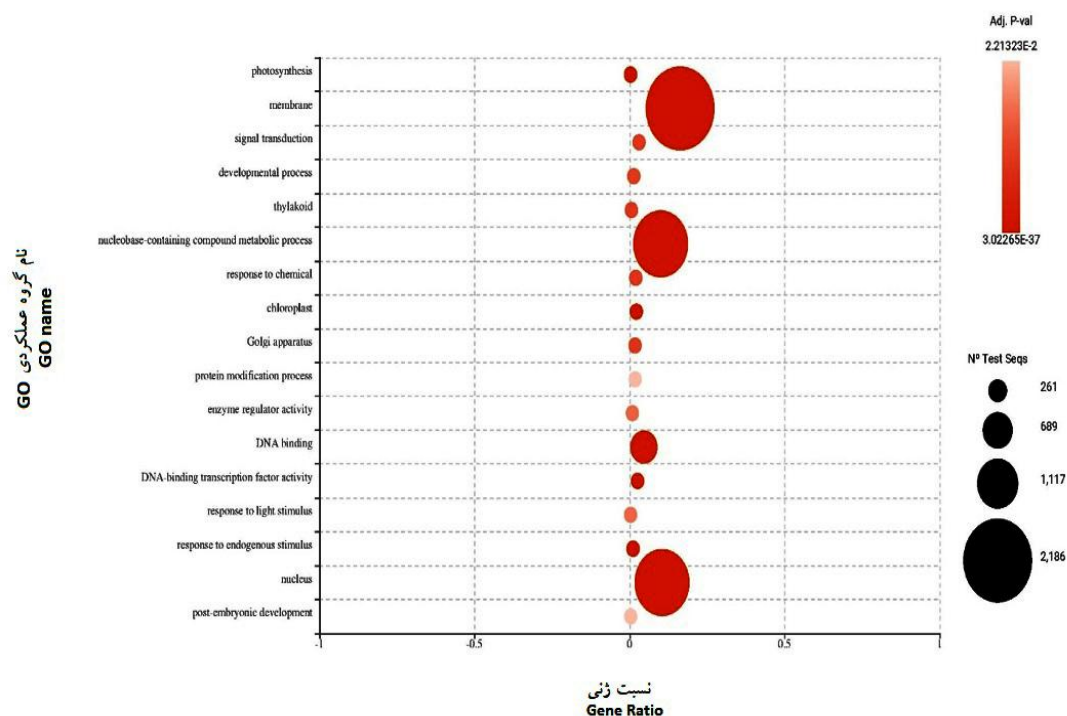


شکل ۵. توزیع یونی‌ژن‌های حاوی EST-SSR در مسیرهای مختلف بیوسنتزی و شناسایی مسیرهای فعال بیولوژیکی با استفاده از پایگاه داده Reactome pathway databases

Figure 5. Distribution of EST-SSR containing unigenes in different biosynthetic pathways and identification of biologically active pathways using Reactome pathway databases

³² Signal Transduction
³³ Metabolism of proteins
³⁴ Cell Cycle
³⁵ Gene expression (Transcription)

نتایج آزمون فیشر و آنالیز هستی شناسی یونی ژن های حاوی EST-SSR برای شناسایی مسیرهای غنی شده^{۳۶} نشان داد که زیرگروه های غشای^{۳۷} با ۲۱۸۶ یونی ژن، هسته^{۳۸} با ۱۴۰۶ یونی ژن و فعالیت های متابولیکی ترکیبات حاوی نوکلئوبازها^{۳۹} با ۱۳۴۷ یونی ژن، و سپس اتصال DNA ترم های غنی شده با یونی ژن های حاوی EST-SSR در این مطالعه بودند (شکل ۶).



شکل ۶. توزیع مسیرهای متابولیکی غنی شده با یونی ژن های حاوی EST-SSR

Figure 6. Distribution of enrichment metabolic pathways of unigenes containing EST-SSRs

گیاهان در طبیعت در معرض انواع مختلفی از تنش های زنده و غیرزنده قرار دارند که در عملکرد و رشد آنان نقش دارند (Filiz & Kurt, 2021). به طور کلی موجودات زنده از جمله گیاهان از انواع مکانیسم های انتقال سیگنال برای درک و پاسخ به تنش های محیطی مختلف استفاده می کنند (Hong et al., 2013). محرک های محیطی به وسیله حسگرهای مختلف غشای سلولی تشخیص داده شده که باعث فعال شدن پیام رسان های ثانویه و انتقال سیگنال های تنش به هسته و در نهایت باعث تغییر در بیان ژن ها می شوند (Filiz & Kurt, 2021). حضور تکرارهای SSR در کلاس های پروتئین های بیولوژیکی مهم و تنوع نسخه های تکراری SSRها موجب انعطاف پذیری سلول و کل ارگانیسم در مواجهه با محیط های درونی و بیرونی می شود که به طور مداوم در حال تغییر هستند (Ahmadi & Ahmadikhah, 2022). مطالعات نشان داده که تعداد SSR های موجود در ناحیه 5' اغلب با بیان ژن همبستگی مثبت داشته است (Sharopova, 2008). تغییرات SSR در 5'-UTR ها می تواند بیان ژن را با تأثیر بر رونویسی و ترجمه تنظیم

³⁶ Enrichment pathways

³⁷ Membrane

³⁸ nucleus

³⁹ nucleobase-containing compound metabolic process

کند (Li et al., 2004) SSR های سه نوکلئوتیدی در ژن های تنظیمی مربوط به رونویسی و انتقال سیگنال^{۴۰} بیشتر حضور داشته در حالیکه در ژن های مربوط به پروتئین های ساختاری حضور کمتری را نشان دادند (Li et al., 2002). این در حالی است که نشانگرهای با ۴ نوکلئوتید یا بیشتر در ناحیه 5'-UTR اغلب با سطوح پایین رونوشت مرتبط بودند (Sharopova, 2008). در مطالعه هستی شناسی و شناسایی مسیرهای غنی شده، مسیر "فعالیت های متابولیکی ترکیبات حاوی نوکلئوبازها" از مسیرهای اشباع شده بود که در رشد سریع و فعالیت متابولیکی گسترده در گیاه دخیل هستند (Kalra et al., 2013). شناسایی یونی ژن های حاوی SSR در سطح هستی شناسی، پتانسیل خوبی برای استفاده از این نشانگرهای مولکولی در جهت شناسایی مسیرهای مختلف است. امروزه از RNA-Seq در گیاهان غیرمدل و یا گیاهانی با اطلاعات توالی محدود، به دلیل تمرکز بر روی ناحیه کدگذاری^{۴۱} به طور گسترده استفاده می شود. EST-SSR های شناسایی شده با کمک RNA-Seq را می توان برای تجزیه و تحلیل تنوع ژنتیکی، ساختار جمعیت و مطالعات عملکردی استفاده کرد. توسعه نشانگرهای EST-SSR بر اساس ژن های کاندید مرتبط با صفات هدف ارزش بالایی دارد. توسعه SSR های مرتبط با فنوتیپ، فعالیت فیزیولوژیکی و مسیرهای متابولیکی از فواید استخراج SSR ها از پایگاه داده ترانسکریپتوم است (Satya et al., 2017).

محدودیت ها و چشم اندازهای آینده: ترنسکریپتوم مرجع این مطالعه با ادغام داده های رونویسی از بافت های برگ و ریشه ایجاد شده است. این رویکرد، اگرچه پوشش ژنی کلی را به حداکثر می رساند، اما مانع از بررسی دقیق تفاوت های بیان ژن در سطح بافت می شود. در نتیجه، صحت و دقت شناسایی برخی از SSR ها (به ویژه آنهایی که بیان مختص بافت دارند) ممکن است تحت تأثیر قرار گرفته باشد. بنابراین، یافته های مربوط به فراوانی و توزیع SSR ها باید به عنوان تخمینی از کل مجموعه ژنی در نظر گرفته شده و نه لزوماً بازتابی دقیق از بیان اختصاصی بافتی تفسیر گردند. از این رو پیشنهاد می شود که تحقیقات آینده بر مواردی از جمله تحقیقات بیان ژن تفکیک شده از هر بافت و اعتبارسنجی SSR های ویژه هر بافت با استفاده از روش های مولکولی متمرکز شوند.

نتیجه گیری: نشانگرهای مولکولی از ابزارهای تشخیص چندشکلی ژنتیکی در جایگاه های خاص و کل سطح ژنوم در گونه های گیاهی هستند. در بین نشانگرهای مولکولی مختلف، SSR ها یکی از مهم ترین نشانگرها در برنامه های ژنتیکی و اصلاحی به شمار می آیند. با این حال، تعداد محدودی از آنان برای گونه های مختلف شناخته شده است که باعث محدودیت رویکردهای اصلاحی شده است. شناسایی ریزماهواره ها و کشف تنوع آنان با کمک تکنیک توالی یابی نسل جدید (NGS) آسان تر شده است. در حال حاضر، استفاده از توالی یابی RNA، که با عنوان RNA-Seq شناخته شده، در مقایسه با فرآیندهای سنتی توسعه SSR، سریع تر و آسان تر بوده و به یک ابزار قابل اعتماد و قوی در شناسایی و توسعه تعداد قابل توجهی از نشانگرهای SSR تبدیل شده است. در مطالعه حاضر از توالی یابی رونوشت ها برای ایجاد یک ترنسکریپتوم مرجع از دارمازو استفاده شد. در مجموع ۹۶۲۲۵ یونی ژن به دست آمد که از این تعداد ۱۳۳۶۸ یونی ژن دارای ۲۰۵۰۵ مورد نشانگر SSR بودند. موتیف های تک نوکلئوتیدی و پس از آن موتیف های دو و سه نوکلئوتیدی دارای بیشترین درصد فراوانی بودند. بیشترین موتیف های تک نوکلئوتیدی از نوع A/T، و بیشترین موتیف های ۲، ۳، ۴ و ۵ نوکلئوتیدی به ترتیب از نوع AG/CT، AAG/CTT، AAAT/ATTT و AAAAG/CTTT بودند. بررسی کارکرد یونی ژن های دارای نشانگر نشان داد که اکثر رونوشت های حاشیه نویسی شده با اعمال اصلی و حیاتی در گیاه در

⁴⁰ Signal transduction

⁴¹ Coding regions

ارتباط هستند. این نتایج یک منبع با ارزش از ترنسکریپتوم را در دارمازو ارائه کرده است که می‌توان از آن برای بررسی ژن‌های مرتبط با سنتز ترکیبات خاص آن استفاده کرد.

سپاسگزاری: از معاونت محترم پژوهشی دانشگاه لرستان که به خاطر پشتیبانی مالی و تجهیزاتی ما را در اجرای هرچه بهتر پژوهش حاضر یاری کردند، سپاسگزاری می‌شود.

References

- Ahmadi, A. J., & Ahmadikhah, A. (2022). Occurrence of simple sequence repeats in cDNA sequences of safflower (*Carthamus tinctorius*) reveals the importance of SSR-containing genes for cell biology and dynamic response to environmental cues. *Frontiers in Plant Science*, 13, Article 991107. <https://doi.org/10.3389/fpls.2022.991107>
- Alaaddin Ahmed, A., Anwar Qadir, S., & Tahir, N. A.-R. (2022). CDDP and ISSR markers-assisted diversity and structure analysis in Iraqi Mazu (*Quercus infectoria* Oliv.) accessions. *All Life*, 15(1), 247–261. <https://doi.org/10.1080/26895293.2022.2042401>
- Amilah, W. A. W. W. N., Mohamad, A. N., Izani, N. J. N., & Arizam, M. F. (2022). Antimicrobial activities of *Quercus infectoria* gall extracts: A scoping review. *Journal of Herbal Medicine*, 32, Article 100543. <https://doi.org/10.1016/j.hermed.2022.100543>
- An, M., Deng, M., Zheng, S.-S., & Song, Y.-G. (2016). De novo transcriptome assembly and development of SSR markers of oaks *Quercus austrocochinchinensis* and *Q. kerrii* (Fagaceae). *Tree Genetics & Genomes*, 12, Article 1. <https://doi.org/10.1007/s11295-016-1060-5>
- Askari, F., Azadi, A., Namavar-Jahromi, B., Tansaz, M., Mirzapour Nasiri, A., & Mohagheghzadeh, A. (2020). A comprehensive review about *Quercus infectoria* G. Olivier gall. *Research Journal of Pharmacognosy*, 7(1), 69–77. <https://doi.org/10.22127/rjp.2019.184177.1494>
- da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrugal, J., Sibbesen, J. A., Marett, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, 30, 3–13. <https://doi.org/10.1016/j.margen.2016.04.012>
- Farahvashi, M., Mohammadabadi, M., Askari-Hesni, M., Amiri Ghanatsaman, Z., & Asadollahpour Nanaei, H. (2026). Genomic differentiation and diversity in Persian Gulf hawksbill turtles (*Eretmochelys imbricata*) revealed by the first whole-genome sequencing study. *Animals*, 16(2), Article 169. <https://doi.org/10.3390/ani16020169>
- Filiz, E., & Kurt, F. (2021). Expression and co-expression analyses of WRKY, MYB, bHLH and bZIP transcription factor genes in potato (*Solanum tuberosum*) under abiotic stress conditions: RNA-seq data analysis. *Potato Research*, 64, 1–21. <https://doi.org/10.1007/s11540-021-09502-3>
- Guerrero-Sanchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., Verardi, A., Jorrín-Novo, J. V., & Rey, M.-D. (2019). Ion Torrent and Illumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PLOS ONE*, 14(1), Article e0210356. <https://doi.org/10.1371/journal.pone.0210356>
- Hajalizadeh, Z., Dayani, O., Khezri, A., Tahmasbi, R., Mohammadabadi, M., Solodka, T., Kalashnyk, O., Afanasenko, V., & Babenko, O. (2021). Expression of calpastatin gene in Kermani sheep using real-time PCR. *Journal of Livestock Science and Technologies*, 9(2), 51–57. <https://doi.org/10.22103/jlst.2021.18165.1381>

- Hong, S.-Y., Roze, L. V., & Linz, J. E. (2013). Oxidative stress-related transcription factors in the regulation of secondary metabolism. *Toxins*, 5(4), 683–702. <https://doi.org/10.3390/toxins5040683>
- Hosseini, S. Z., Ismaili, A., Nazarian Firouzabadi, F., Fallahi, H., & Rezaeinejad, A. (2018). Development of EST-SSR microsatellite markers related to drought tolerance in lentil (*Lens culinaris*). *Crop Biotechnology*, 8(23), 43–57.
- Kalra, S., Puniya, B. L., Kulshreshtha, D., Kumar, S., Kaur, J., Ramachandran, S., & Singh, K. (2013). De novo transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivilianum*. *PLOS ONE*, 8(12), Article e83336. <https://doi.org/10.1371/journal.pone.0083336>
- Kaur, G., Hamid, H., Ali, A., Alam, M. S., & Athar, M. (2004). Antiinflammatory evaluation of alcoholic extract of galls of *Quercus infectoria*. *Journal of Ethnopharmacology*, 90(2–3), 285–292. <https://doi.org/10.1016/j.jep.2003.10.009>
- Khabiri, A., Toroghi, R., Mohammadabadi, M., & Tabatabaeizadeh, S. E. (2025). Whole genome sequencing and phylogenetic relative of a pure virulent Newcastle disease virus isolated from an outbreak in northeast Iran. *Letters in Applied Microbiology*, 78(4), Article ovaf049. <https://doi.org/10.1093/lambio/ovaf049>
- Kim, K. K., Park, S. K., Kang, S. A., Park, M. K., Cho, M. K., Jung, H. J., Kim, K. K., & Yu, H. S. (2013). Comparison of functional gene annotation of *Toxascaris leonina* and *Toxocara canis* using CLC genomics workbench. *The Korean Journal of Parasitology*, 51(5), 525–530. <https://doi.org/10.3347/kjp.2013.51.5.525>
- Li, Q., Su, X., Ma, H., Du, K., Yang, M., Chen, B., Fu, S., Fu, T., Xiang, C., & Zhao, Q. (2021). Development of genic SSR marker resources from RNA-seq data in *Camellia japonica* and their application in the genus *Camellia*. *Scientific Reports*, 11, Article 89350. <https://doi.org/10.1038/s41598-021-89350-w>
- Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, 11(12), 2453–2465. <https://doi.org/10.1046/j.1365-294X.2002.01643.x>
- Li, Y.-C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*, 21(6), 991–1007. <https://doi.org/10.1093/molbev/msh073>
- Liu, S., An, Y., Li, F., Li, S., Liu, L., Zhou, Q., Zhao, S., & Wei, C. (2018). Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Molecular Breeding*, 38, Article 82. <https://doi.org/10.1007/s11032-018-0824-z>
- López-Aljorna, A., Bueno, M. Á., Aguinagalde, I., & Martín, J. P. (2007). Fingerprinting and genetic variability in cork oak (*Quercus suber* L.) elite trees using ISSR and SSR markers. *Annals of Forest Science*, 64(7), 773–779. <https://doi.org/10.1051/forest:2007057>
- Mohammadabadi, M. R., Nikbakhti, M., Mirzaee, H. R., Shandi, M. A., Saghi, D. A., Romanov, M. N., & Moiseyeva, I. G. (2010). Genetic variability in three native Iranian chicken populations of the Khorasan province based on microsatellite markers. *Russian Journal of Genetics*, 46(4), 505–509. <https://doi.org/10.1134/S1022795410040198>
- Mohammadabadi, M., Afsharmanesh, M., Khezri, A., Kheyrodin, H., Babenko, O. I., Borshch, O., Kalashnyk, O., Nechyporenko, O., Afanasenko, V., Slynko, V., & Usenko, S. (2025). Effect of mealworm on GBP4L gene expression in the spleen tissue of Ross broiler

- chickens. *Agricultural Biotechnology Journal*, 17(2), 343–360. <https://doi.org/10.22103/jab.2025.25277.1714>
- Palumbo, F., Vannozzi, A., Vitulo, N., Lucchin, M., & Barcaccia, G. (2018). The leaf transcriptome of fennel (*Foeniculum vulgare* Mill.) enables characterization of the t-anethole pathway and the discovery of microsatellites and single-nucleotide variants. *Scientific Reports*, 8(1), Article 10459. <https://doi.org/10.1038/s41598-018-28775-2>
- Pereira-Leal, J. B., Abreu, I. A., Alabaça, C. S., Almeida, M. H., Almeida, P., Almeida, T., Amorim, M. I., Araújo, S., Azevedo, H., & Badia, A. (2014). A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. *BMC Genomics*, 15(1), Article 371. <https://doi.org/10.1186/1471-2164-15-371>
- Safari, M., Ismaili, A., Sohrabi, S. S., Nazarian-Firouzabadi, F., & Torabi Poteh, H. (2022). Transcriptome analysis of Persian oak (*Quercus brantii* L.) decline using RNA-seq technology. *Biochemical Genetics*, 60, 1–22. <https://doi.org/10.1007/s10528-022-10283-8>
- Sajad Sohrabi, S., Ismaili, A., Azarian Firouz-Abadi, F. N., & Fallahi, H. (2018). Discovery of EST-SSRs markers in lentil (*Lens culinaris*) under cold stress. *Crop Biotechnology*, 8(22), 1–14.
- Satya, P., Chakraborty, A., Jana, S., Majumdar, S., Karan, M., Sarkar, D., Datta, S., Mitra, J., Kar, C. S., & Karmakar, P. G. (2017). Identification of genic SSRs in jute (*Corchorus capsularis*, Malvaceae) and development of markers for phenylpropanoid biosynthesis genes and regulatory genes. *Plant Breeding*, 136(5), 784–797. <https://doi.org/10.1111/pbr.12514>
- Sharopova, N. (2008). Plant simple sequence repeats: Distribution, variation, and effects on gene expression. *Genome*, 51(2), 79–90. <https://doi.org/10.1139/G07-110>
- Shokri, S., Khezri, A., Mohammadabadi, M., & Kheyrodin, H. (2023). The expression of MYH7 gene in femur, humeral muscle and back muscle tissues of fattening lambs of the Kernani breed. *Agricultural Biotechnology Journal*, 15(2), 217–236. <https://doi.org/10.22103/jab.2023.21524.1486>
- Shrestha, S., Kaushik, V. S., Eshwarappa, R. S. B., Subaramaihha, S. R., Ramanna, L. M., & Lakkappa, D. B. (2014). Pharmacognostic studies of insect gall of *Quercus infectoria* Olivier (Fagaceae). *Asian Pacific Journal of Tropical Biomedicine*, 4(1), 35–39. [https://doi.org/10.1016/S2221-1691\(14\)60205-7](https://doi.org/10.1016/S2221-1691(14)60205-7)
- Soltani, N., Best, T., Grace, D., Nelms, C., Shumaker, K., Romero-Severson, J., Moses, D., Schuster, S., Staton, M., & Carlson, J. (2020). Transcriptome profiles of *Quercus rubra* responding to increased O₃ stress. *BMC Genomics*, 21(1), Article 6549. <https://doi.org/10.1186/s12864-020-6549-5>
- Sun, J. J., Zhou, T., Zhang, R. T., Jia, Y., Zhao, Y. M., Yang, J., & Zhao, G. F. (2019). Comparative transcriptomes and development of expressed sequence tag-simple sequence repeat markers for two closely related oak species. *Journal of Systematics and Evolution*, 57(5), 440–450. <https://doi.org/10.1111/jse.12469>
- Sun, L., Li, J., Sun, K., Wang, H., Yang, K., Chen, Q., & Lin, M. (2023). Development and characterization of EST-SSR markers in *Rhodomyrtus tomentosa* Hassk. based on transcriptome. *Genetic Resources and Crop Evolution*, 70(6), 1691–1705. <https://doi.org/10.1007/s10722-022-01528-x>
- Taheri, S., Lee Abdullah, T., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. (2018). Mining and development of novel SSR markers using next generation

- sequencing (NGS) data in plants. *Molecules*, 23(2), Article 399. <https://doi.org/10.3390/molecules23020399>
- Torre, S., Tattini, M., Brunetti, C., Fineschi, S., Fini, A., Ferrini, F., & Sebastiani, F. (2014). RNA-seq analysis of *Quercus pubescens* leaves: De novo transcriptome assembly, annotation and functional markers development. *PLOS ONE*, 9(11), Article e112487. <https://doi.org/10.1371/journal.pone.0112487>
- Wang, D., Yang, T., Liu, R., Li, N., Wang, X., Sarker, A., Zhang, X., Li, R., Pu, Y., & Li, G. (2020). RNA-Seq analysis and development of SSR and KASP markers in lentil (*Lens culinaris* Medikus subsp. *culinaris*). *The Crop Journal*, 8(6), 953–965. <https://doi.org/10.1016/j.cj.2020.04.007>
- Wang, L., Gong, X., Jin, L., Li, H., Lu, J., Duan, J., & Ma, L. (2019). Development and validation of EST-SSR markers of *Magnolia wufengensis* using de novo transcriptome sequencing. *Trees*, 33, 1213–1223. <https://doi.org/10.1007/s00468-019-01853-2>
- Zhou, G.-L., & Zhu, P. (2020). De novo transcriptome sequencing of *Rhododendron molle* and identification of genes involved in the biosynthesis of secondary metabolites. *BMC Plant Biology*, 20(1), Article 2586. <https://doi.org/10.1186/s12870-020-02586-y>
- Zhou, L., Sun, J., Zhang, T., Tang, Y., Liu, J., Gao, C., Zhai, Y., Guo, Y., Feng, L., & Zhang, X. (2022). Comparative transcriptome analyses of different *Rheum officinale* tissues reveal differentially expressed genes associated with anthraquinone, catechin, and gallic acid biosynthesis. *Genes*, 13(9), Article 1592. <https://doi.org/10.3390/genes13091592>
- Zin, N. N. I. N. M., Rahimi, W. N. A. W. M., & Bakar, N. A. (2019). A review of *Quercus infectoria* (Olivier) galls as a resource for anti-parasitic agents: In vitro and in vivo studies. *The Malaysian Journal of Medical Sciences*, 26(6), 19–34. <https://doi.org/10.21315/mjms2019.26.6.3>