

## بررسی ترانسکرپتوم زیره سیاه ایرانی (*Bunium persicum* Boiss.) در مرحله دانه‌بندی و شناسایی ریزماهوره‌های (SSR) ژنومی در این گیاه

محمدرسول سمندری بهرآسمان<sup>۱</sup>، احمد اسماعیلی<sup>۲\*</sup>، سعید اسماعیلی ماهانی<sup>۳</sup>، اسماعیل ابراهیمی<sup>۴</sup> و اولین لوییت<sup>۵</sup>

۱-دانش آموخته دکتری، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه لرستان، خرم آباد، ایران

\*۲- نویسنده مسئول مکاتبات، استاد گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه لرستان، خرم آباد، ایران

پست الکترونیک: Email: ismaili.a@lu.ac.ir

۳-استاد، گروه زیست شناسی، دانشگاه شهید باهنر، کرمان

۴-استاد، گروه تحقیقات ژنومیک، دانشکده علوم، سلامت و مهندسی، دانشگاه لاتروب، ملبورن، استرالیا

۵- دانشیار دانشگاه علوم حیاتی استونی، دانشگاه علوم زیستی استونی، تارتو

تاریخ پذیرش: ۱۴۰۰/۱۲/۰۷

تاریخ دریافت: ۱۴۰۰/۸/۱۵

### چکیده

زیره سیاه ایرانی از مهمترین گیاهان بومی ایران است. دانه این گیاه در رژیم غذایی روزمره ایرانی‌ها مصرف می‌شود و همچنین دارای خواص دارویی و ترکیبات معطر است. متأسفانه اطلاعات کمی در مورد ژنوم و نشانگرهای مولکولی این گیاه وجود دارد. در این پژوهش، ترانسکرپتوم زیره سیاه ایرانی در مرحله دانه‌بندی توالی‌یابی شد و بعد از استخراج ریزماهوره‌های (SSR Markers) مرتبط با این مرحله رشدی به تفسیر آن‌ها پرداخته شد. گل‌آذین و ساقه این گیاه بعد از استخراج RNA، توالی‌یابی شدند. توالی‌یابی با عمق ۶ گیگ و طول قطعات ۱۵۰ جفت باز و بصورت جفت-انتها انجام شد. داده‌های حاصل از توالی‌یابی توسط نرم‌افزارهای مختلف آنالیز شدند. بازچینش رونوشت زیره سیاه ایرانی توسط نرم افزار Trinity انجام شد. استخراج نشانگرهای ریزماهوره توسط نرم‌افزار MISA انجام شد. در نهایت تفسیر عملکردی توالی‌های دربرگیرنده ریزماهوره توسط پایگاه داده WEGO انجام شد. میزان GC نمونه‌های مختلف ۴۷ تا ۵۰ درصد بود. تعداد ۸۳۸۹ نشانگر ریزماهوره از ۴۵۱۴۶ ژن منفرد حاصل شد. حداقل ۱۱ درصد ژن‌های منفرد حاوی ریزماهوره بودند. نوکلئوتیدهای A/T و AG/CT بیشترین درصد نشانگرهای ریزماهوره را در برداشتند. در میان تکرارهای سه نوکلئوتیدی، ATC/ATG و AAG/CTT بیشترین فراوانی ریزماهوره‌ها را به خود اختصاص داد. نتایج تفسیر عملکردی نشان داد که ژن‌های منفرد حاوی نشانگرهای ریزماهوره طیف وسیعی از فعالیت‌های بیولوژیکی گیاه زیره سیاه ایرانی را پوشش می‌دهند و ژن‌های درگیر در تکامل و رشد دانه، حد بالایی از بیان را دارند. این تحقیق برای اولین بار به بررسی نشانگرهای ریزماهوره گیاه زیره سیاه ایرانی پرداخته است. نتایج این تحقیق می‌تواند در پیداکردن نشانگرهای مختلف برای برنامه‌های اصلاحی و توسعه ارقام با اهداف مختلف مفید باشد.

واژه‌های کلیدی: نسل جدید توالی‌یابی، نشانگر، بازچینش، تفسیر عملکردی

### مقدمه

چندساله دارویی و متعلق به خانواده Apiaceae و تعداد

گیاه زیره سیاه ایرانی (*Bunium persicum*) یک گیاه کروموزوم‌های سوماتیک دیپلوئید آن ۱۴ است (Sheidai et

خواب بذر، کاهش مرحله رویشی و توسعه ارقام با عملکرد بالا با کیفیت غذایی بالاتر برخی از اهداف برنامه‌های اصلاحی آینده در زیره سیاه خواهد بود و این تا حد زیادی به در دسترس بودن تنوع ژنتیکی برای ارائه انتخاب‌های بهتر بستگی دارد (Singh & Kumar, 2021).

مطالعات در مورد تنوع در بین جمعیت‌های مختلف، تغییر ژنتیکی ژرم‌پلاسما برای درک عملکرد اهلی‌سازی، بهبود و توسعه ارقام جدید و هیبریدها ضروریست. همچنین مطالعات در مورد ارزیابی تنوع ژنتیکی در سطوح مختلف مولکولی و فنوتیپی مهم است. با این حال، مطالعات در مورد ارزیابی تنوع ژنتیکی در سطح DNA در زیره سیاه بسیار ناچیز است و گزارش‌های کمی در دسترس است. بیشتر مطالعات گزارش شده در مورد تنوع ژنتیکی بر اساس نشانگرهای مولکولی RAPD انجام شده است. مطالعه Pezhmanmehr و همکاران (۲۰۰۹) در ارزیابی تنوع ژنتیکی با استفاده از ۱۵ نشانگر مولکولی RAPD و ۱۷ نشانگر AFLP در بین ۲۰ جمعیت ایرانی زیره سیاه، چندشکلی بالایی را نشان داد. این موضوع نشان‌دهنده یک زمینه ژنتیکی متنوع و منحصر به فرد در این گیاه است (Pezhmanmehr et al., 2009). در مطالعه‌های دیگر با استفاده از نشانگرهای ریزماهواره و روش GC-MS، روابط ژنتیکی و فیتوشیمیایی گونه‌های جمع‌آوری شده از کرمان و یزد بررسی شد. نتایج با پراکندگی جغرافیایی گونه‌های بررسی شده مطابقت داشت و ترکیب‌های دارویی مشترک بین گونه‌های مختلف شناسایی شدند (Jafari & saadatfar, 2020).

آشکارسازی توالی ژنومی برای همه شاخه‌های علوم زیستی حیاتی است. تعیین توالی نسل بعدی (NGS) رویکردی متفاوت در این زمینه است، به گونه‌ای که تحول بزرگی در علوم زیست‌شناسی ایجاد کرده و جنبه‌های مختلف مطالعات سطح ژنوم، رونوشت، اپی‌ژنوم و متاژنوم را پوشش می‌دهد. NGS در مقایسه با روش‌های سنتی به دلیل تأمین پوشش ژنومی مناسب، تعیین تک‌تک جفت بازها و از بین بردن معایب توالی نسل اول (تعیین توالی سنگر) به‌عنوان

(al., 1996). این گونه با نام‌های مختلفی مانند کله زیره، زیره کوهی، زیره سیاه، زیره سیاه ایرانی، زیره ایرانی و زیره کرمانی نام‌گذاری می‌شود. این گیاه به‌طور گسترده در مناطق آب و هوایی معتدل تا خشک و در مناطق کوهستانی دیده می‌شود. به‌طور طبیعی در دامنه‌های تپه‌ای سرد کویری و در دامنه‌هایی با ارتفاعی از ۱۸۰۰ تا ۳۵۰۰ متر رشد می‌کند. ارتفاع گیاه از ۳۰ تا ۸۰ سانتی‌متر متغیر است، برگ‌های کوچک دارد و در سال سوم پس از کاشت بذر، گل‌آذین‌های چتری مرکب با گلچه سفید تولید می‌کند (Sofi et al., 2009).

بهربرداری بی‌رویه این گیاه از طبیعت باعث فرسایش ژنتیکی این گیاه شده است و آن را در معرض انقراض قرار داده است. بنابراین، کشت و توسعه تجاری آن برای جلوگیری از جمع‌آوری این گیاه از طبیعت مورد نیاز است (Ved et al., 2003). این گیاه از تنوع ژنتیکی گسترده‌ای در سراسر جهان برخوردار است و تنوع ژنتیکی خوبی را برای بهبود از طریق انتخاب و شروع برنامه‌های اصلاحی ارائه می‌دهد (Mittal et al., 2006). چندین مشکل مانند مرحله جوانی طولانی، خواب بذر، مرحله رویشی خفته و قابلیت عملکرد پایین در رابطه با کشت تجاری زیره سیاه ایرانی وجود دارد. زیره سیاه ایرانی دارای غده زیرزمینی (کورم) می‌باشد که مسئول تکمیل مراحل رویشی و زایشی گیاه است. در واقع، خواب بذر و دوره جوانی ۷ ساله (از کاشت بذر تا تکوین کورم)، از تمایل به اهلی کردن آن کاسته است. خواب در بذرهای زیره کرمانی سبب می‌شود که بذرها به راحتی جوانه نزنند و برای جوانه‌زنی باید بر خواب بذرها غلبه کرد. دوره جوانی مدت زمان لازم برای تولید میوه از دانه‌های ۴-۷ سال می‌باشد که این دوره در این گیاه طولانی است و این موضوع مانع از ترغیت زارعان به کشت و اهلی کردن زیره سیاه ایرانی است. Emamipoor و Maziah (۲۰۱۴) ثابت کردند که با کمک تنظیم‌کننده‌های رشد گیاهی و تیمار سرمادهی می‌توان بر خواب بذرهای زیره سیاه ایرانی غلبه و تا ۹۷٪ جوانه‌زنی بذرها را فراهم کرد (Emamipoor & Maziah, 2014). بنابراین، شکستن

در ترنسکریپتوم این گیاه استخراج، معرفی و تفسیر عملکردی گردیده است.

### مواد و روش‌ها

جمع‌آوری و نمونه‌گیری:

قسمت‌های هوایی گیاه زیره سیاه ایرانی از ارتفاعات کوهپایه کرمان در اوایل خردادماه سال ۱۳۹۹ درحالی‌که میوه‌ها در مرحله فرایند بلوغ بودند جمع‌آوری شدند. سه عدد بوته که شرایط رشدی مشابهی داشتند برای مراحل بعدی انتخاب شد. مختصات جغرافیایی منطقه جمع‌آوری  $57^{\circ}11' E$  و  $30^{\circ}30' N$  بود. در نهایت نمونه‌های جمع‌آوری شده بلافاصله به ازت مایع منتقل شدند.

### استخراج RNA

استخراج RNA توسط کیت استخراج زیست آسیا انجام شد. برای اندازه‌گیری کیفیت و کمیت RNAها از دستگاه NanoDrop استفاده شد.

### توالی‌یابی

نمونه‌های RNA برای توالی‌یابی به شرکت Novogene چین ارسال شدند، بعد از تأیید کیفیت RNAها کتابخانه cDNA آنها طبق دستورالعمل شرکت نوژن و به روش RNA depletion ساخته شد و در پایان توسط دستگاه Illumina و به صورت جفت-انتها با طول ۱۵۰ جفت-باز و عمق توالی‌یابی ۶ گیگابایت داده، توالی‌یابی شدند.

### کنترل کیفی، پیرایش و بازچینش

کنترل کیفیت خوانش‌های حاصل از توالی‌یابی، قبل و بعد از ویرایش توسط نرم‌افزار FastQC (Version 0.11.6) و در محیط سیستم عامل لینوکس انجام شد. بعد از بررسی نتایج FastQC، ویرایش خوانش‌ها توسط نرم‌افزار Trimmomatic (Version 0.36) و در محیط لینوکس انجام گردید. آداپتورها<sup>۱</sup>

یک روش با کارایی بالا برای تجزیه و تحلیل اطلاعات ژنومی و ترانسکریپتومیک در نظر گرفته می‌شود. اخیراً، تعیین توالی RNA برای شناسایی ژن‌های مرتبط با فرایندهای رشد و نمو و الگوهای بیان آنها در پاسخ به تنش‌های مختلف بیولوژیکی و غیر بیولوژیکی، در اندام‌های مختلف و مراحل رشد در موجودات مختلف مورد استفاده قرار می‌گیرد. این به دانشمندان کمک می‌کند تا میزان بیان ژن، تمایز ایزوفرم‌های مختلف ژن‌ها، تشخیص RNAهای کوچک و همچنین تشخیص ژن‌های فیوژن، عناصر تکراری، آگرون‌های ژن‌ها، UTRها، SNPها و ریزماهورها را در وسعت و کارایی بالاتری انجام دهند (Zhang et al., 2012). در تحقیقی Bharti و همکاران (۲۰۱۸) با بررسی ترنسکریپتوم زیره سبز توسط پایگاه داده MISA موفق به استخراج ۸۰۸۶ نشانگر ریزماهوره از این گیاه شدند. Jie و همکاران (۲۰۱۶) با انجام توالی‌یابی روی نمایه تظاهر گیاه *Boehmeria nivea* و در ادامه با استفاده از نرم‌افزار MISA در مجموع ۴۲۳۰ مکان ریزماهوره را شناسایی کردند. در میان ریزماهوره‌های استخراج شده توالی‌های دی‌نوکلئوتیدی و تری‌نوکلئوتیدی بیشترین تعداد را شامل می‌شدند. Hui و همکاران (۲۰۱۶) با بررسی نمایه تظاهر گیاه کنف (*Hibiscus cannabinus*) با استفاده از نرم‌افزار MISA موفق به استخراج ۹۳۲۴ ریزماهوره شدند. در ادامه، با بررسی ۱۰۰ عدد از این ریزماهوره‌ها مشخص شد که از کیفیت بالایی برخوردار هستند و توانایی تفکیک و تمایز بین گونه‌های مختلف کنف را دارند (Hui et al., 2016).

تاکنون مطالعه‌ای برای بررسی ژنوم و ترنسکریپتوم<sup>۱</sup> گیاه زیره سیاه ایرانی انجام نشده است و بررسی ترنسکریپتوم این گیاه می‌تواند شروعی برای مطالعات عمیق‌تر روی این گیاه باشد. در این مطالعه به تحلیل نمایه تظاهر ژن‌های این گیاه در مرحله دانه‌بندی با انجام توالی‌یابی پرداخته شده است. همچنین منبع ارزشمندی از نشانگرهای ریزماهوره<sup>۲</sup> موجود

1 Transcriptome

2 SSR marker

دوتایی، سه‌تایی، چهارتایی، پنج‌تایی و شش‌تایی با حداقل تعداد تکرار به ترتیب ۵، ۶ و ۷ غربالگری شدند.

تحلیل تفسیر عملکردی ژن‌های منفرد حاوی ریزماهوره بعد از مشخص شدن ژن‌های منفرد حاوی ریزماهوره توسط نرم‌افزار MISA، شماره اختصاصی هر ژن منفرد از فایل اصلی حاوی تمام ژن‌های منفرد استخراج شد، سپس ژن‌های منفرد حاوی توالی‌های ریزماهوره توسط پایگاه داده WEGO<sup>۴</sup> تفسیر عملکردی شدند. از پارامترهای پیش‌فرض برای انجام تحلیل استفاده شد.

### نتایج و بحث

کنترل کیفی و اصلاح خوانش‌های حاصل از توالی‌یابی: در این مطالعه ترانسکریپتوم ۴ تکرار از گل آذین به همراه ۲ تکرار از بافت ساقه و برگ توالی‌یابی شدند. خوانش‌های حاصل از توالی‌یابی بعد از کنترل کیفیت توسط نرم‌افزار FastQC از کیفیت بالایی برخوردار بودند. در بیش از ۹۴ درصد خوانش‌های همه تکرارها میزان خطای توالی‌یابی کمتر از ۰/۱٪ بود. مجموع خوانش‌ها حدود ۲۲۰ میلیون و هر نمونه تقریباً ۴۰ میلیون خوانش بود (جدول ۱). بعد از ویرایش کردن خوانش‌ها با نرم‌افزار تریوموماتیک، حذف خوانش‌های با کیفیت پایین و حذف آداپتورها مراحل بعدی تحلیل انجام شد. همان‌طور که در جدول ۱ نمایش داده شده است، میزان GC نمونه‌های مختلف ۴۷ تا ۵۰ درصد بود. کیفیت بالای خوانش‌ها، نشان‌دهنده کیفیت مورد قبول فرایند توالی‌یابی است و این موضوع باعث افزایش اعتماد به نتایج حاصل از آنالیزهای بعدی می‌شود. نتایج بررسی کیفیت داده‌های حاصل نشان داد که گروه‌های مختلف این مطالعه با کیفیت بسیار بالایی توالی‌یابی شده‌اند.

و خوانش‌های با کیفیت پایین حذف شدند. از پارامترهای LEADING:3، TRAILING:3 و LIDINGWINDOW:4:15 و MINLEN:36 برای پیرایش استفاده شد. بعد از پیرایش خوانش‌ها برای اطمینان از کیفیت عمل انجام شده، خوانش‌ها دوباره از لحاظ کیفیت توسط نرم افزار FastQC بررسی شدند. به دلیل عدم وجود ژنوم مرجع برای گیاه زیره سیاه ایرانی، ژنوم نوپدید برای خوانش‌های موجود توسط نرم‌افزار trinity ساخته شد. برای بازچینش خوانش‌ها از کی-مر<sup>۱</sup> ۳۲ استفاده و بقیه پارامترهای آن در حالت استاندارد قرار داده شد.

### مستندسازی و تفسیر عملکردی<sup>۲</sup>

مستندسازی عملکردی توسط سه پایگاه داده KEGG، KOG و Nr انجام شد. Nr (توالی‌های پروتئینی غیرزاید<sup>۳</sup> NCBI) پایگاه داده‌های توالی پروتئین‌های رسمی NCBI است که شامل اطلاعات توالی پروتئینی GenBank، PDB، SwissProt، PIR، PRF و غیره است. KOG (گروه‌های ارتولوگ یوکاریوتی) براساس روابط ارتولوگ ژن‌های موجود در پایگاه داده NCBI است. با توجه به روابط تکاملی خود، KOG ژنهای همولوگ را از گونه‌های مختلف به خوشه‌های مختلف ارتولوگ تقسیم می‌کند. مجموعه KOG فعلی شامل ۴۸۵۲ خوشه ارتولوگ است که شامل ۵۹۸۳۸ پروتئین است. از آنجایی که ژن‌های ارتولوگ عملکرد یکسانی دارند، مستندسازی عملکردی را می‌توان از خوشه‌های KOG برای اعضای دیگر به اشتراک گذاشت. بعد از بارگذاری داده‌ها در این پایگاه داده‌ها آستانه ارزیابی برای هر سه پایگاه داده روی  $1e-5$  قرار گرفت.

### استخراج ریزماهوره‌ها

استخراج ریزماهوره‌ها بوسیله نرم‌افزار MISA توسط پارامترهای پیش‌فرض انجام شد. پس از سرهم‌بندی خوانش‌ها، کانتیگ‌ها برای موتیف‌های تکراری یک‌تایی،

1 K-mer

2 Functional Annotation

3 Non-redundant

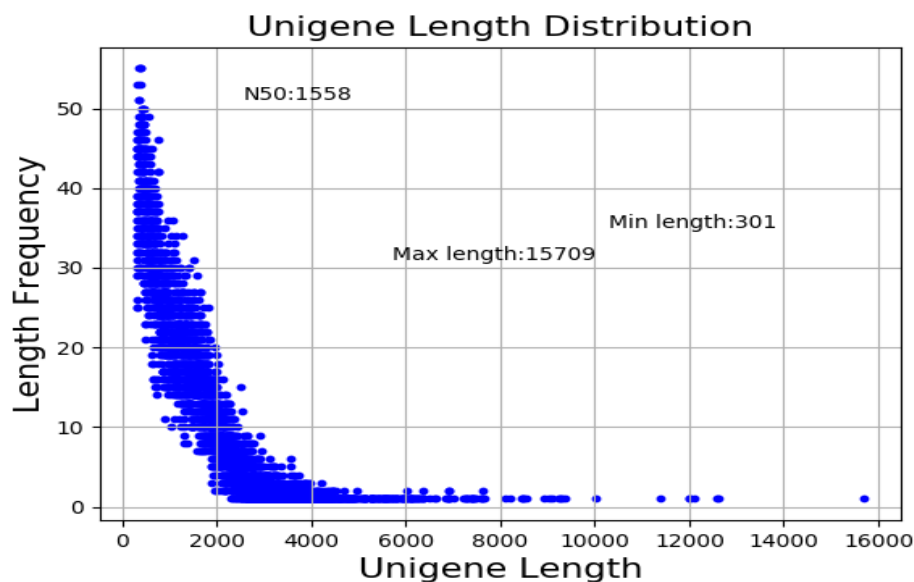
4 Web Gene Ontology Annotation Plot

جدول ۱- نتایج کنترل کیفیت گروه‌های مختلف بعد از ویرایش توسط نرم افزار تریموماتیک

Table 1. Quality control results of different groups after trimming by Trimmomatic software

GC (%)	(%) Q <sub>30</sub>	(%) Q <sub>20</sub>	خوانش‌های باقیمانده بعد از ویرایش (%) Obtained reads after trimming	خوانش‌ها Reads	گروه‌ها Groups
48.19	94.89	98.44	95.73	45646308	گل آذین-۱ Inflorescence-1
48.89	94.13	98.02	98.37	40126114	گل آذین-۲ Inflorescence-2
50.07	94.40	98.13	96.34	45262073	گل آذین-۳ Inflorescence-3
47.10	94.73	98.35	95.66	41679726	گل آذین-۴ Inflorescence-4
49.55	95.37	98.50	98.62	42950712	ساقه-۱ Stem-1
50.55	95.27	98.87	98.90	47193724	ساقه-۲ Stem-2

GC: درصد نوکلئوتیدهای گوانین و سیتوزین تشکیل دهنده توالی. Q<sub>20</sub>: میزان خطای توالی‌یابی کمتر از ۱٪. Q<sub>30</sub>: میزان خطای توالی‌یابی کمتر از ۰.۱٪.  
GC: percentage of guanine and cytosine nucleotides in the sequence. Q<sub>20</sub>: Sequencing error rate less than 1%. Q<sub>30</sub>: Sequencing error rate less than 0.1%.



شکل ۱- توزیع طول قطعات حاصل از بازچینش نوپدید ترنسکریپتوم زیره سیاه ایرانی

Figure 1. The length distribution of the fragments resulting from the *De novo* sequencing of Persian black cumin transcriptome

سلطانی هویزه و همکاران روی گیاه زنیان که از خانواده Apiaceae است میزان  $N_{50}$ ، ۱۲۰۰ گزارش شده است (Soltani Howyzeh *et al.*, 2018). در مطالعه Sadeghi و همکاران (۲۰۱۸) روی گیاه زیره سبز ۴۵۹۱۷ ژن منفرد حاصل شد و  $N_{50}$  حاصل ۱۱۳۶ بود. از آنجایی که این دو گیاه متعلق به خانواده Apiaceae هستند نتایج به دست آمده در این دو تحقیق با نتایج این مطالعه همخوانی دارد (Sadeghi *et al.*, 2018). میزان هم‌ردیفی<sup>۳</sup> خوانش‌های نمونه‌های مختلف با ژنوم مرجع ساخته شده بسیار خوب بود و همه گروه‌های تقریباً بالای ۸۷ درصد با ژنوم مرجع از نو ساخته شده توسط نرم‌افزار Trinity همپوشانی بالایی نشان دادند (جدول ۲).

بازچینش نوپدید و مستندسازی عملکردی<sup>۱</sup>: بررسی ترانسکریپتوم گیاهان در مراحل مختلف می‌تواند اطلاعات ارزشمندی را در اختیار بگذارد تا به فهم دقیق‌تری از ژن‌های دخیل در فرایندهای مختلف بیولوژیکی آن گیاه برسیم. خوانش‌های با کیفیت بالا وارد مرحله یکپارچه‌سازی نوپدید توسط نرم‌افزار Trinity (k-mer : 32) شدند و ژنوم مرجع آنها ساخته شد. بررسی contig ساخته شده نشان داد که در مجموع ۴۵۱۶۱ ژن منفرد<sup>۲</sup> با  $N_{50}$  (طول حداقل ۵۰ درصد از توالی‌های بازچینش شده) ۱۵۵۸ حاصل شده است که کوتاه‌ترین قطعه ۳۰۱ جفت-باز و بلندترین قطعه ۱۵۷۰۹ جفت-باز طول داشتند (شکل ۱). این میزان  $N_{50}$  عدد قابل قبولی است و نشان از کیفیت مناسب یکپارچه‌سازی نوپدید خوانش‌ها دارد. در مطالعه

جدول ۲- میزان هم‌ردیفی خوانش‌های گروه‌های مختلف با ژنوم بازچینش نوپدید توسط نرم‌افزار Trinity

Table 2. The amount of alignment of the reads of different groups with the assembled transcriptome by Trinity software

گروه‌ها Groups	مجموع خوانش‌های هم‌ردیف شده (%)	کی-مر k-mer	خوانش‌های هم‌ردیف نشده (%)	فقط یک‌بار هم‌ردیف شده (%)	بیش از یک‌بار هم‌ردیف شده (%)
	Total aligned reads		Unpaired reads	Only one time aligned	More than on time aligned
ساقه-۱ Stem-1	86.96	32	13.04	51.29	35.67
ساقه-۲ Stem-2	87.69	32	12.31	53.45	34.24
گل آذین-۱ Inflorescence-1	87.10	32	12.9	52.76	34.34
گل آذین-۲ Inflorescence-2	87.86	32	12.14	53.56	34.03
گل آذین-۳ Inflorescence-3	86.98	32	13.03	53.01	33.97
گل آذین-۴ Inflorescence-4	88.89	32	11.11	52.78	36.11

برخلاف نتایج توالی‌یابی گیاه زیره سبز توسط Sadeghi و

تفسیر عملکردی: همان‌طور که در شکل (۲) الف مشخص است نتایج توزیع شباهت داده‌های حاصل از تفسیر عملکردی Nr نشان می‌دهد که بیشترین شباهت گیاه زیره سیاه ایرانی با هویج به میزان ۸۵/۸ درصد است. این نتیجه

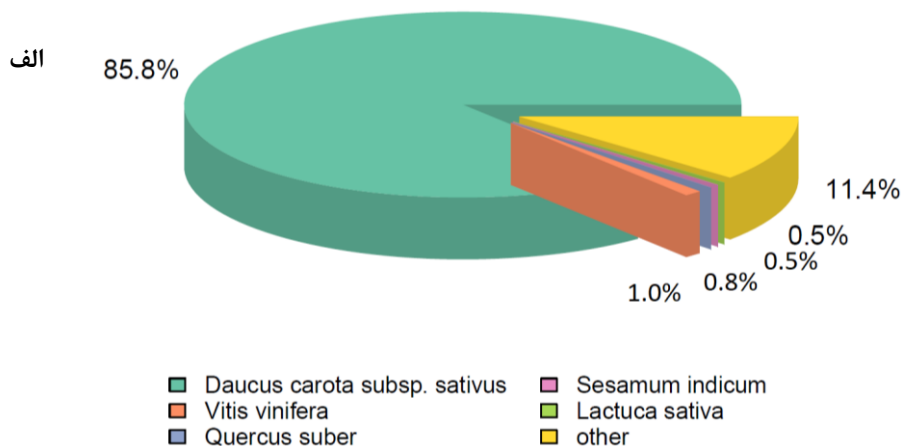
1 Functional annotation

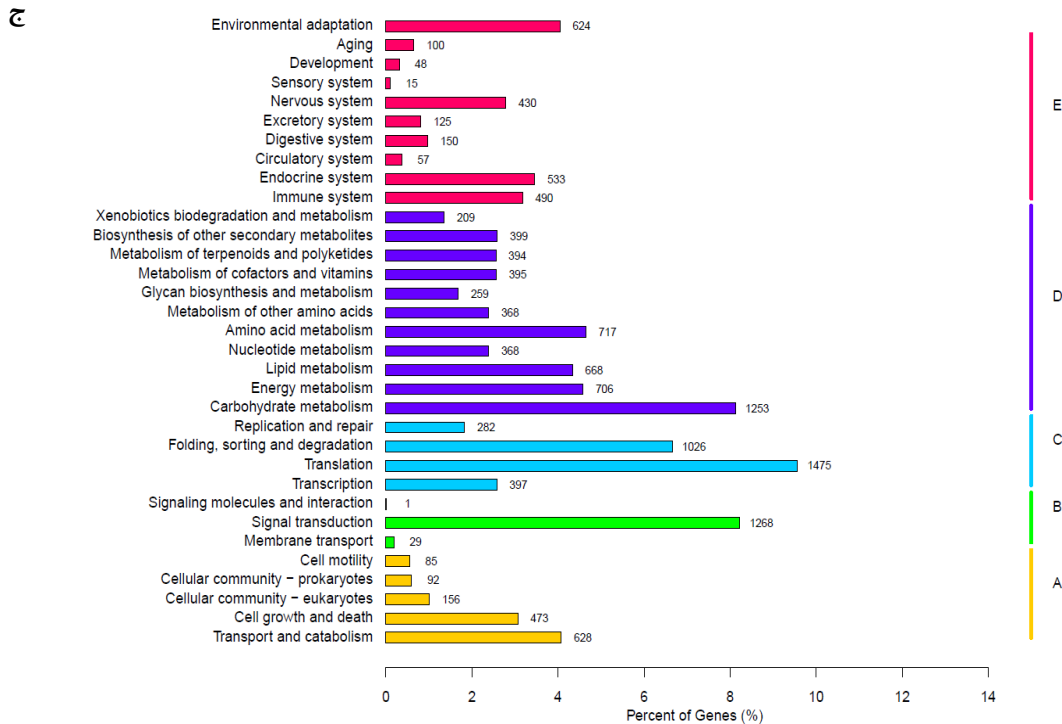
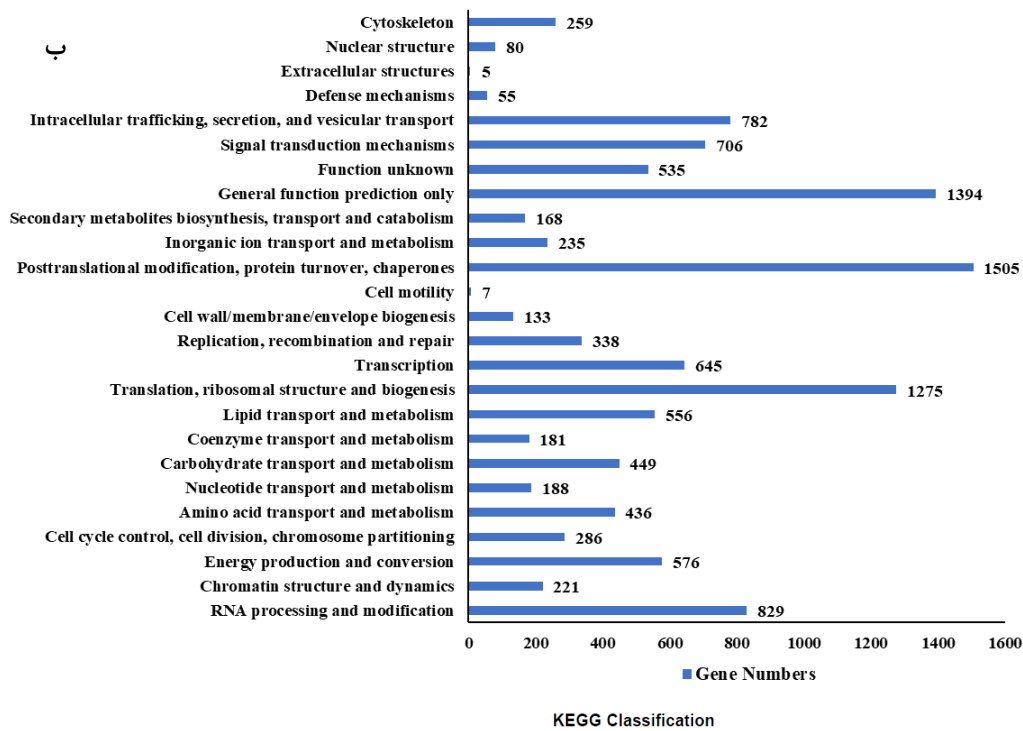
2 Unigene

3 Mapping

فعالیت‌های حاصل شده در این پایگاه داده، احتمالاً مرتبط با شرایط رشدی دانه است و نیاز شدید گیاه به تولید پروتئین را نشان می‌دهد که معمولاً در شرایط نمودی رخ می‌دهد و از آنجایی که نمونه‌های گل‌آذین در مرحله تشکیل دانه نمونه‌گیری شدند این موضوع طبیعی به نظر می‌رسد. از میزان کل ژن‌های منفرد ۱۵۴۱۷ عدد ژن منفرد برابر ۳۴ درصد از کل ژن‌های منفرد در پایگاه داده KEGG تفسیر عملکردی شدند. نتیجه حاصل از تفسیر عملکردی با پایگاه KEGG نیز تقریباً نشان از فعالیت بالای گیاه در تولید پروتئین دارد و حدود ده درصد ژن‌های بیان شده را در قسمت پروتئین‌سازی مشخص کرده است. در رتبه بعدی ژن‌های مرتبط با متابولیسم کربوهیدرات‌ها و انتقال سیگنال قرار دارند (هریک حدود ۸ درصد). ژن‌های مرتبط با فرایندهای تولید متابولیت‌های ثانویه و تریپنوئیدها با هم حدود ۶ درصد از ترنسکرپتوم زیره سیاه ایرانی را تشکیل می‌دهند (شکل ۲-ج).

همکاران (۲۰۱۸) است که حداکثر ۱۵ درصد شباهت باژنوم هویج را نشان می‌دهد (Sadeghi *et al.*, 2018). هویج از خانواده Apiaceae است که توالی‌یابی ژنوم آن انجام شده است. با توجه به هم‌خانواده بودن این گیاه با زیره سیاه ایرانی این میزان شباهت ترنسکرپتوم نشان‌دهنده نزدیکی بودن فعالیت‌های بیولوژیکی این دو گیاه است. نتایج حاصل از مطالعه سلطانی هویزه و همکاران روی گیاه زنیان نیز شباهت بالای ترنسکرپتوم این گیاه را با هویج نشان می‌دهد (Soltani Howyzeh *et al.*, 2018). بررسی تفسیر عملکردی خوانش‌ها توسط پایگاه داده KOG نشان داد که بیشتر ژن‌های تشکیل‌دهنده گیاه زیره سیاه ایرانی در مرحله تشکیل دانه مربوط به فعالیت‌های "تغییرات پس از ترجمه و ترجمه"، "فعالیت‌های عمومی"، "رونوشت‌برداری و ساختار ریبوزومی"، "ویرایش RNA"، "تولید انرژی" و "نقل و انتقالات سلولی" است (شکل ۲-ب). از میزان کل ژن‌های منفرد، ۱۰۵۴۷ عدد برابر ۲۳ درصد از کل ژن‌های منفرد در این پایگاه داده تفسیر عملکردی شدند. این طیف از





شکل ۲- الف) بررسی شباهت ترانسکریپتوم زیره سیاه ایرانی با ژنوم‌های توالی‌یابی شده. ب) تفسیر عملکردی رونوشت سرهم‌بندی شده نوپدید زیره سیاه ایرانی در پایگاه داده KOG. ج) تفسیر عملکردی رونوشت سرهم‌بندی شده نوپدید توسط پایگاه داده KEGG.

Figure 2. a) Examining the similarity of Persian black cumin transcriptome with sequenced genomes. b) Enrichment of the *De novo* assembled transcriptome of Persian black cumin in the KOG database. (c) Enrichment of *De novo* assembled transcriptome by KEGG database.



مربوط به ریزماهوره‌های دوتایی بود (et al., 2014; Functammasan et al., 2015; Bharti et al., 2018). بعد از آن، ریزماهوره‌های تکی و سه‌تایی با تعداد ۲۹۰۳ و ۱۸۷۸ بیشترین تعداد ریزماهوره‌های زیره سیاه ایرانی را تشکیل می‌دادند. در پایان، ریزماهوره‌های چهارتایی، پنج‌تایی و شش‌تایی کمترین میزان را تشکیل می‌دادند (جدول ۴). این نتایج مشابه مطالعات دیگر در گیاهان زیره سبز، *Arabidopsis* انگور و بادام‌زمینی است (Senthilvel et al., 2008; Bharti et al., 2018). بیشترین تکرار تک نوکلئوتیدی A/T بود. در مطالعات انجام شده روی کلزا و عدس نیز این موتیف بیشترین تکرار را داشت (Chen et al., 2017; Sohrabi et al., 2018). در میان تکرارهای دی نوکلئوتیدی تکرار AG/CT و پس از آن AT/AT بیشترین تکرار را به خود اختصاص می‌دهد (شکل ۳). در گیاه زیره سبز بیشترین تکرار در بین ریزماهوره‌های دوتایی متعلق به AT/AT بود و بعد از آن AG/CT بیشترین تعداد ریزماهوره‌های دوتایی را به خود اختصاص داد (Bharti et al., 2018). در مطالعات دیگر نتیجه مشابه با این تحقیق حاصل شده است و توالی A/T بیشترین درصد تشکیل‌دهنده ریزماهوره‌ها را به خود اختصاص داده است (Kumpatla & Mukhopadhyay, 2005; Grubisha et al., 2014; Chen et al., 2017). تکرارهای سه نوکلئوتیدی ATC/ATG و AAG/CTT بیشترین فراوانی ریزماهوره‌ها را به خود اختصاص دادند. همان‌طور که در جدول ۴ نمایان است، فراوانی موتیف‌های ریزماهوره با افزایش طول آنها کاهش می‌یابد و این موضوع در مطالعات قبلی هم دیده شده است (Hosseini et al., 2018).

استخراج نشانگر ریزماهوره: آنالیز ترنسکرپتوم زیره سیاه ایرانی برای پیدا کردن ریزماهوره‌ها با نرم‌افزار MISA انجام شد. از مجموع ۴۵۱۴۶ ژن منفرد حدود ۸۳۸۹ عدد نشانگر ریزماهوره شناسایی شد. تعداد ۶۹۴۶ ژن منفرد حاوی یک نشانگر ریزماهوره بودند و ۱۱۹۲ عدد ژن منفرد بیش از یک نشانگر را دربر داشتند (جدول ۳) که نشان می‌دهد حدود ۱۵ درصد ژن‌های منفرد حاوی نشانگرهای ریزماهوره بودند. در تحقیقات پیشین نشان داده شده است که حدود ۶ تا ۱۱ درصد توالی‌های ترنسکرپتومی دارای نشانگرهای ریزماهوره هستند (Varshney et al., 2002). بنابراین میزان ۱۵ درصد کل ترنسکرپتوم نشان‌دهنده تنوع بالای این نشانگر در ژنوم گیاه زیره سیاه ایرانی است و این موضوع از اهمیت بالایی برای اصلاح نژاد این گیاه برخوردار است. البته در بعضی مطالعات نشانگرهای مونونوکلئوتیدی را به دلیل امکان خطا در فرایند توالی‌یابی لحاظ نمی‌کنند (Fungtammasan et al., 2015). هرچند کیفیت توالی‌یابی نمونه‌های این مطالعه بسیار بالا بود، با این حال اگر این نشانگرهای مونونوکلئوتیدی را هم در این تحقیق لحاظ نکنیم ۱۱ درصد ژن‌های منفرد این گیاه دارای نشانگرهای ریزماهوره‌ای هستند که میزان بسیار مناسبی است (جدول ۳). در مطالعه مشابهی روی زیره سبز، Grubisha و همکاران (۲۰۱۸) تعداد ۸۰۸۶ عدد ریزماهوره از ژنوم این گیاه به دست آوردند که بسیار نزدیک به تعداد ریزماهوره به دست آمده در این تحقیق بود. ریزماهوره‌های دوتایی بیشترین تعداد تشکیل‌دهنده زیره سیاه ایرانی را با تعداد ۳۴۶۷ به خود اختصاص می‌دادند. در مطالعات مشابه دیگر نیز بیشترین تعداد

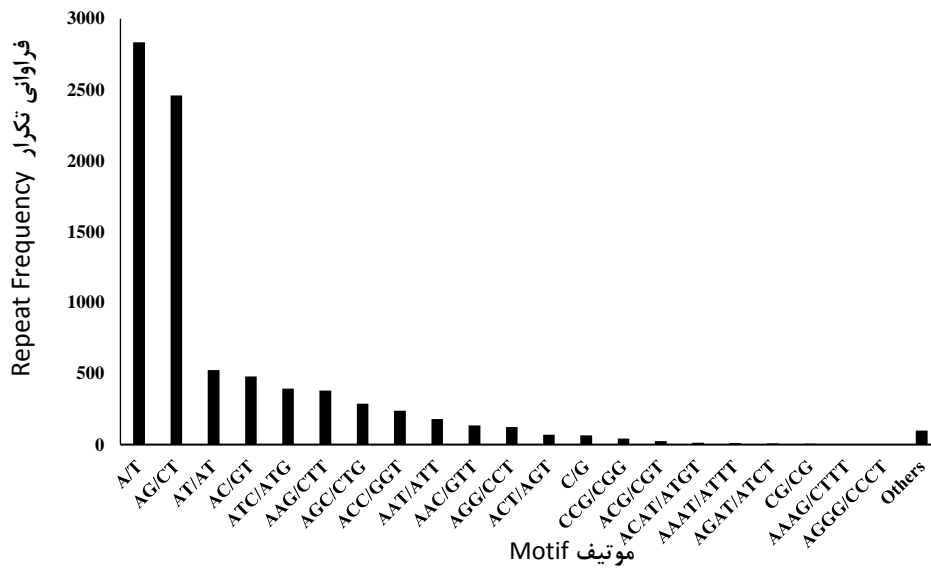
جدول ۳- آمار کلی ریزماهوره‌های استخراج شده توسط MISA  
**Table3. General statistics of microsatellites extracted by MISA**

تعداد number	آماره بررسی شده Evaluated statistics
45146	تعداد ژن‌های منفرد بررسی شده The number of examined unigenes
54646521	طول نهایی مجموع توالی‌های بررسی شده (جفت باز) The final length of the sum of the examined sequences (paired base)
8389	کل نشانگرهای ریزماهوره شناسایی شده Total identified microsatellite markers
6946	تعداد توالی‌های حاوی یک ریزماهوره Number of sequences containing a microsatellite
1192	تعداد توالی‌های حاوی بیش از یک ریزماهوره Number of sequences containing more than one microsatellite
793	تعداد ریزماهوره‌های ترکیبی Number of combined microsatellites

جدول ۴- فراوانی ریزماهوره‌های ترانسکریپتوم حاصل از یکپارچه‌سازی نوپدید زیره سیاه ایرانی

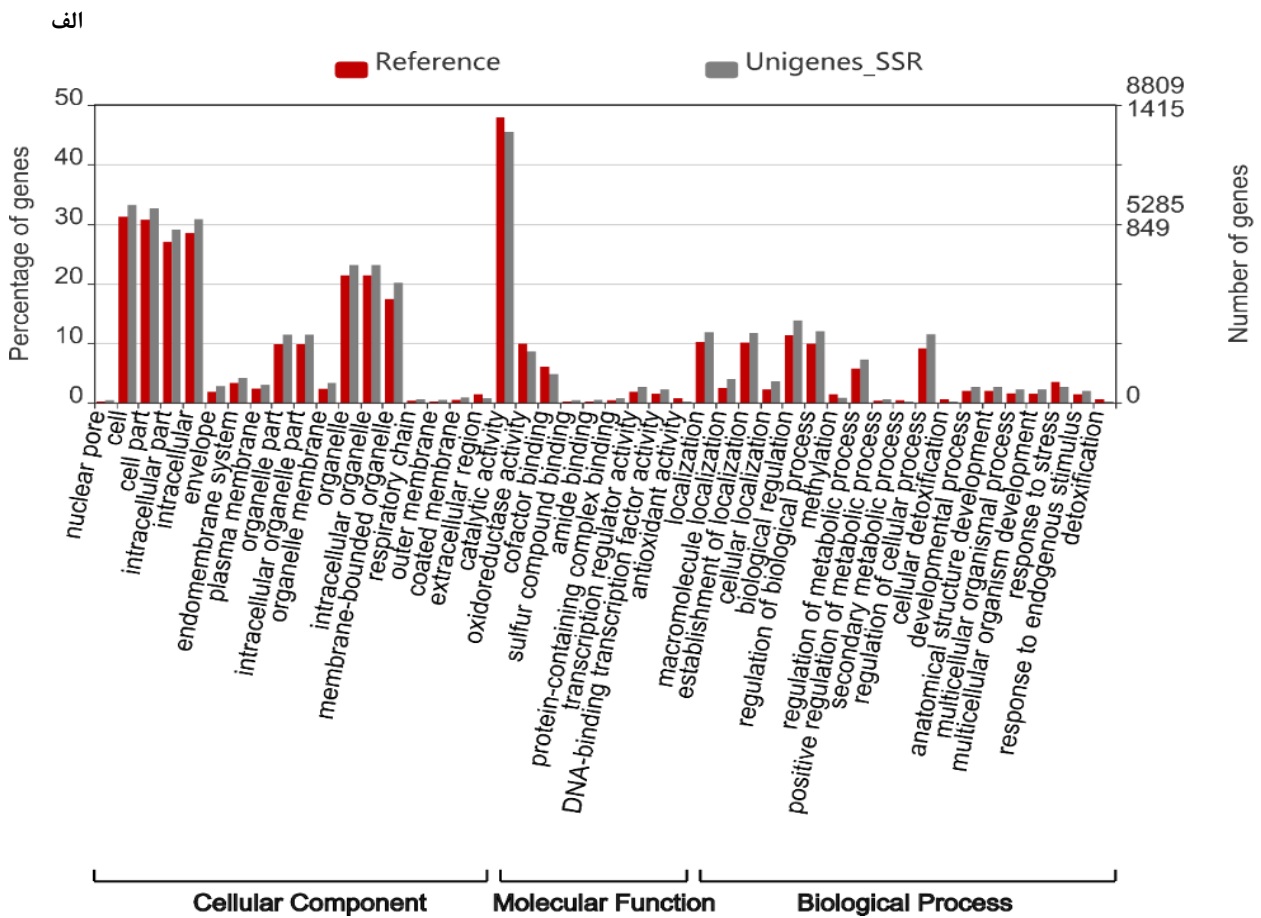
**Table 4. Frequency of microsatellites of the transcriptome resulting from the *De novo* assembled transcriptome of Persian black cumin**

تکرار Repeats	مونو نوکلئوتیدی Mono-nucleotide	دی نوکلئوتیدی Di-nucleotide	تری نوکلئوتیدی Tri-nucleotide	تترا نوکلئوتیدی Tetra-nucleotide	پنتا نوکلئوتیدی Penta-nucleotide	هگزا نوکلئوتیدی Hexa-nucleotide	مجموع Total
5	0	0	1107	34	15	36	1192
6	0	1080	434	9	14	10	1547
7	0	616	177	5	0	7	805
8	0	466	67	1	0	9	543
9	0	307	47	0	0	3	357
10	990	196	13	0	0	0	1199
11	477	138	3	0	1	2	621
12	334	97	4	13	0	0	448
13	237	93	11	0	0	0	341
14	191	111	3	0	0	0	305
15	126	72	4	0	0	0	202
16	100	36	0	0	0	0	136
17	94	32	3	0	0	7	136
18	66	38	0	0	0	0	104
19	60	21	0	0	0	0	81
20	40	29	0	0	0	0	69
≥ 21	180	125	0	0	0	0	305
Total	2903	3467	1878	62	30	74	8392

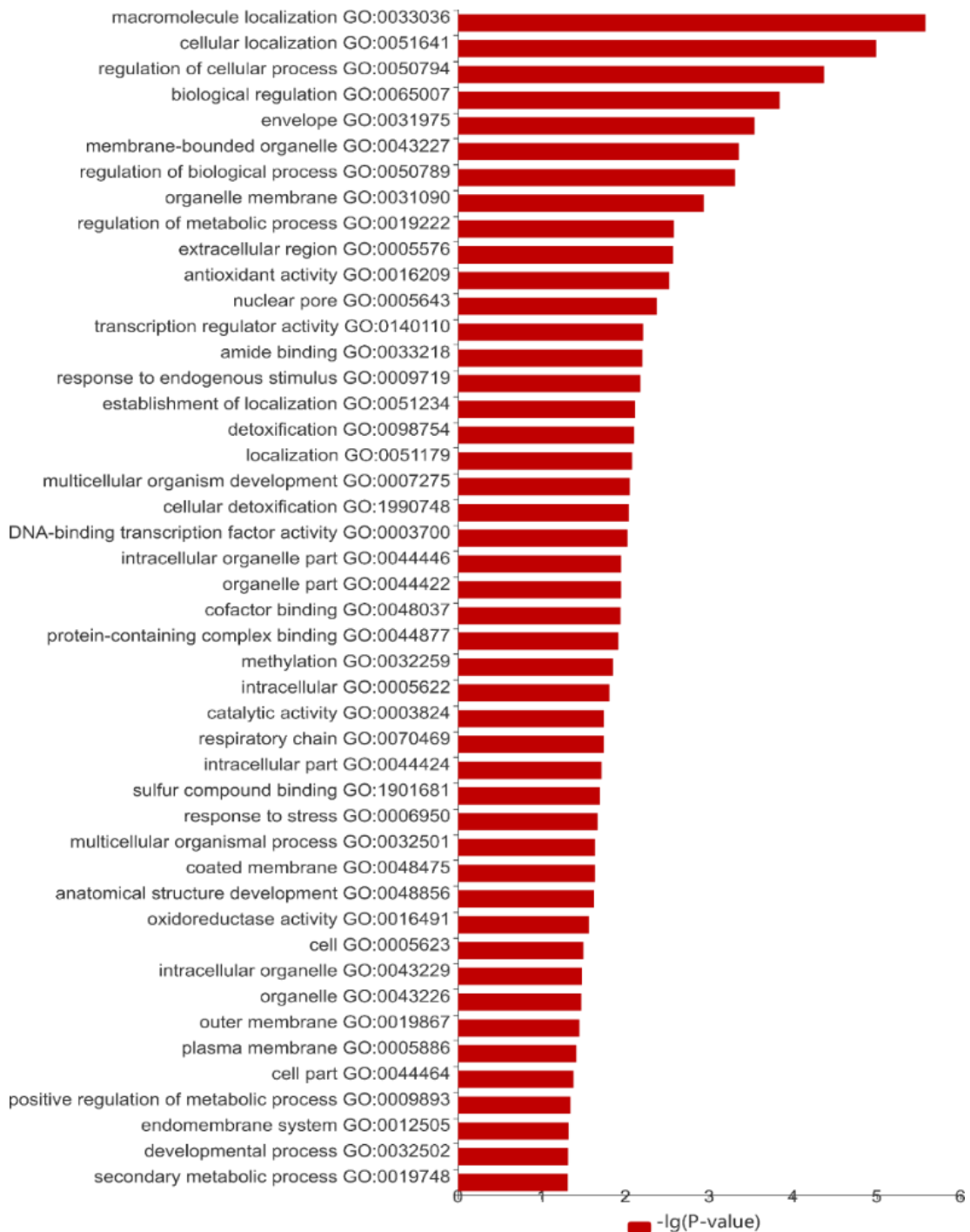


شکل ۳- طبقه‌بندی ریزماهورها بر اساس فراوانی موتیف‌ها

Figure 3. Classification of microsatellites based on the frequency of motifs



ب



شکل ۴- تفسیر عملکردی ریزماهوره‌های حاصل از MISA در پایگاه داده WEGO. الف) طبقه بندی کارکردی ترانسکریپتوم حاوی

ریزماهوره. ب) میزان ارتباط ژن‌های منفرد حاوی ریزماهوره با فرایندهای مولکولی

**Figure 4. Enrichment analysis of MISA-derived microsatellites in the WEGO database. A) Functional classification of microsatellite-containing transcriptome. b) Association of individual genes containing microsatellites with molecular processes**

دیگر نشانگرها، نشانگرهای ریزماهوره انتخاب محققان برای اصلاح گیاهان هستند، زیرا دارای وراثت غالب و چند آلی هستند و فراوانی و تکرارپذیری بیشتری دارند (Li *et al.*, 2002).

### نتیجه‌گیری

تا به امروز، هیچ نشانگر ریزماهوره‌ای در زیره سیاه استخراج و گزارش نشده است، البته تحقیق روی نشانگرهای دیگر این گیاه انجام شده است (Bharti *et al.*, 2018). توسعه نشانگرهای ریزماهوره از طریق غربالگری کتابخانه ریزماهوره یک فرایند نسبتاً وقت‌گیر و پرهزینه است (Wei *et al.*, 2011). این تحقیق با استفاده از توالی‌یابی RNA باعث ایجاد بانک ارزشمندی از نشانگرهای ریزماهوره برای زیره سیاه ایرانی شد. نتایج این تحقیق نشان داد که رونوشت این گیاه از تنوع بالایی برخوردار است و حدود ۱۱ درصد توالی‌های کدکننده این گیاه دارای نشانگرهای ریزماهوره هستند. با توجه به اهمیت فرایندهای اصلاحی برای توسعه کاربردی این گیاه، به نظر می‌رسد نشانگرهای مولکولی به‌دست‌آمده در این مطالعه برای مطالعات آینده مفید باشد و با بررسی دقیق‌تر این نشانگرها می‌توان به شناسایی نشانگرهای مرتبط با صفات کلیدی و اقتصادی این گیاه دارویی پی برد. با توجه به نیاز اساسی و فوری برای نجات این گیاه ارزشمند بومی، می‌توان از اطلاعات این تحقیق برای کمک به فرایند اصلاح و توسعه برنامه‌های اصلاحی استفاده کرد. نشانگرهای ریزماهوره حاصل از توالی‌های بیان‌شونده نسبت به توالی‌های بین ژنی از اهمیت بالایی در طرح‌ریزی برنامه‌های اصلاحی برخوردارند، زیرا به‌طور مستقیم مرتبط با ژن‌های کارکردی هستند (Li *et al.*, 2002).

### منابع مورد استفاده:

- Bharti, R., Kumar, S., and Parekh, MJ., 2018. Development of genomic simple sequence repeat (gSSR) markers in cumin and their application in diversity analyses and cross-transferability.

بررسی عملکردی ژن‌های منفرد حاوی ریزماهوره: نتایج بررسی عملکردی ژن‌های منفرد حاوی ریزماهوره در پایگاه داده WEGO نشان داد که بیشترین میزان تکرار ریزماهوره‌ها به ترتیب در قسمت اجزای سلولی، عملکردهای مولکولی و در آخر فرایندهای بیولوژیکی قرار دارند. در قسمت اجزای سلولی بیشترین تعداد مربوط به سلول، بخش سلول، بخش داخل سلولی، داخل سلولی، ارگانل، ارگانل‌های مربوط به اتصال غشایی بودند. در قسمت فعالیت‌های مولکولی بیشترین درصد مربوط به فعالیت‌های کاتالیزی بود. در قسمت فعالیت‌های بیولوژیکی بیشترین میزان مرتبط با تنظیم فعالیت‌های بیولوژیکی بود (شکل ۴-الف). بررسی فعالیت عملکردی توسط پایگاه داده WEGO نشان داد که به ترتیب جایگزینی غشایی، مکان‌یابی سلولی، تنظیم فعالیت‌های سلولی و تنظیم زیستی سلولی بیشترین میزان معنی‌داری را به خود اختصاص می‌دهند (شکل ۴-ب). نتایج تفسیر عملکردی نشان داد که ژن‌های منفرد حاوی نشانگرهای ریزماهوره طیف وسیعی از فعالیت‌های بیولوژیکی گیاه زیره سیاه ایرانی را پوشش می‌دهند و از آنجایی که نتایج تفسیر عملکردی کل خوانش‌ها نشان می‌دهد که ترنسکرپتوم زیره سیاه ایرانی در مرحله نمو تشکیل دانه است و احتمالاً ژن‌های درگیر در تکامل و رشد دانه حد بالایی از بیان را دارند و این موضوع در برنامه‌ریزی طرح‌های اصلاحی برای کمک به اصلاح ژنتیکی این گیاه مفید خواهد بود. در تحقیقات انجام شده روی این گیاه، از نشانگرهای ریزماهوره شناسایی شده در گیاهان هم‌خانواده آن استفاده شده است (Soltani Howyzeh *et al.*, 2018). استفاده از نشانگرهای ژنتیکی و ارتباط آنها با صفات‌های مورد نظر می‌تواند در انتخاب گیاهان با صفت دلخواه کمک کننده باشد و محققان می‌توانند با دقت بیشتری فرایند اصلاح گیاه را پیش ببرند. توسعه نشانگرهای ریزماهوره از طریق کتابخانه ژنومی یک موضوع زمان‌بر، پرزحمت و پرهزینه است، اما تعیین توالی نسل بعدی (NGS) باعث توسعه سریع‌تر و ارزان‌تر نشانگرهای ریزماهوره شده است (Zhang *et al.*, 2012). در مقایسه با

- distribution, putative functions and mutational mechanisms: a review. *Molecular ecology* 11: 2453-2465.
- Mittal, R., Chahota, S. and Katna, G., 2006. Genetic variability and component analysis in kalazira (*Bunium persicum*) in dry temperate areas of North-western Himalayas. *Crop Improvement-India-* 33, 202.
  - Pezhmanmehr, M., Hasani, M., Jahansouz, F., Najafi, A., Sefidkon, F. and Pirseyedi, M., 2009. Assessment of genetic diversity in some Iranian populations of *Bunium persicum* using RAPD and AFLP markers. *Iranian Journal of Biotechnology* 7: 93-100.
  - Sadeghi, D., Mortazavian, M. and Bakhtiyarizadeh, M., 2018. Transcriptome analysis of cumin (*Cuminum cyminum* L.) using RNA-Seq. *Agricultural Biotechnology Journal* 9: 101-116.
  - Senthilvel, S., Jayashree, B., Mahalakshmi, V., Sathish, P. and 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.
  - Sheidai, M., Ahmadian, P. and Poorseyedy, S., 1996. Cytological studies in Iran zira from three genus: *Bunium*, *Carum* and *Cuminum*. *Cytologia* 61: 19-25.
  - Singh, S. and Kumar, V., 2021. Biology, genetic improvement and agronomy of *Bunium persicum* (Boiss.) Fedtsch. A comprehensive review. *Journal of Applied Research on Medicinal and Aromatic Plants*, 100304.
  - Sofi, P., Zeerak, A. and Singh, P., 2009. Kala zeera (*Bunium persicum* Boiss.): a Kashmirian high value crop. *Turkish Journal of Biology* 33: 249-258.
  - Sohrabi, S., Ismaili, A., Nazarian Firouz-Abadi, F. and Fallahi, H., 2018. Discovery of EST-SSRs Markers in Lentil (*Lens culinaris*) under cold stress. *Crop Biotechnology*. 8:1-14 (In Persian).
  - Varshney, R., Thomas, T., Nils, S., Langridge, P. and Graner, A., 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters* 7: 537-546.
  - Wei, W., Xiaoqiong, W., Linhai, Z., Yanxin, H., Wei, Z. and Xiurong, x., 2011. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC genomics* 12: 1-13.
  - Zhang, H., Wei, L. and Miao, H., 2012. Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC genomics* 13: 1-11.
  - Industrial Crops and Products, 111: 158-164.
  - Emamipoor, Y. and Maziah, M. 2014. An efficient method in breaking of dormancy from *Bunium persicum* (Boiss) Fedtsch seeds: a valuable herb of Middle East and Central Asia. *Asian Pac J Trop Biomed*; 4(8): 642-649.
  - Fungtammasan, A., Guruprasad, A., Suzanne, E., Marcia, S., Chen, S., Robert, H., Paul, M., Kristin, E., and Kateryna, D., 2015. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Research* 25: 736-749.
  - Grubisha, L., Bailey A., Nicholas, J., Steven, L and Matthew, R., 2014. Characterization of microsatellite markers for pinedrops, *Pterospora andromedea* (Ericaceae), from Illumina MiSeq sequencing. *Applications in Plant Sciences* 2: 1400072.
  - Hui, L., Defang, L., Anguo, Ch, Huijuan, T., Jianjun, L. and Siqi, Huang., 2016. Characterization of the Kenaf (*Hibiscus cannabinus*) Global Transcriptome Using Illumina Paired-End Sequencing and Development of EST-SSR Markers. *Polos One*.11: 230-247.
  - Hosseini, S., Ismaili, A., Nazarian Firouz-Abadi, F., Fallahi, H. and Rezaeinejad, A., 2018. Development of est-ssr microsatellite markers related to drought tolerance in lentil (*Lens culinaris*). *Crop Biotechnology*. 8:43-57 (In Persian).
  - Soltani Howyzeh, M., Seyed Ahmad, S., Vahid, S. and Mahboubeh, A., 2018. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. *Scientific reports* 8: 1-19.
  - Jafari, H. and Saadatfar, S., 2020. Genetic relationship and phytochemical assessment among populations of *Bunium persicum* (Boiss.) B. Fedtsch. in the natural habitats of Yazd and Kerman. *Iranian Journal of Medicinal and Aromatic Plants*. 36: 306-316 ( In Persian).
  - Jie, Ch., Runqing, Y., Lijun, L., Bo, W. and Dingxiang, P., 2016. Large-scale developing of simple sequence repeat markers and probing its correlation with ramie (*Boehmeria nivea* L.) fiber quality. *Molecular Genetics and Genomics*. 291:753-761.
  - Kumpatla, S. and Mukhopadhyay S., 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48: 985-998.
  - Li, Y., Abraham, B., Tzion, F., Avigdor, B. and Eviatar, N., 2002. Microsatellites: genomic

## Investigation of Persian black cumin (*Bunium persicum* Boiss.) transcriptome in the granulation stage and identification of its genomic microsatellites (SSR)

M. R. Samandari-Bahraseman<sup>1</sup>, A. Ismaili<sup>2\*</sup>, S. Esmaili-Mahani<sup>3</sup>, E. Ebrahimie<sup>4</sup>, E. Loit<sup>5</sup>

1- Ph.D graduated, Plant Production and Genetic Engineering Department, Faculty of Agriculture, Lorestan University, Khorramabad, I.R. Iran.

2-Corresponding Author. Prof., Plant Production and Genetic Engineering Department, Faculty of Agriculture, Lorestan University, Khorramabad, I.R. Iran. Email: ismaili.a@lu.ac.ir

3- Prof., Department of Biology, Faculty of Sciences, Shahid Bahonar University of Kerman, Kerman, I.R. Iran.

4- Prof., La Trobe Genomics Research Platform, School of Life Sciences, College of Science, Health and Engineering, La Trobe University, Melbourne, VIC 3086, Australia.

5- Prof., Chair of Crop Science and Plant Biology, Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia.

Received: 06.11.2021

Accepted: 26.02.2022

### Abstract

Persian black cumin is one of the most important endemic plants to Iran. Its seed is used in the daily diet of Iranians and also has medicinal properties and aromatic compounds. Unfortunately, there is little information about its genome and molecular markers. In this study, the transcriptome of Iranian black cumin was sequenced in the granulation stage and after extracting the microsatellites (SSR markers) related to this growth stage, their interpretation was done. After RNA extraction, the inflorescences and stems were sequenced. Sequencing depth and length were 6 GB and 150 bp pair-end, respectively. Sequencing data were analyzed using different softwares. De-novo assembly of Persian black cumin was performed by Trinity software. Microsatellite markers were extracted using MISA software. Finally, the functional interpretation of the microsatellite-containing sequences was performed by the WEGO database. The GC content of different samples was 47 to 50%. 8389 microsatellite markers were obtained from 45146 unigenes. At least 11% of unigenes contained microsatellites. A/T and AG/CT nucleotides had the highest percentage of microsatellite markers. Among the three nucleotide repeats, ATC/ATG and AAG/CTT had the highest frequency of microsatellites. The results of functional interpretation showed that unigenes containing microsatellite markers cover a wide range of biological activities of Persian black cumin and the genes involved in seed development and growth have a high level of expression. This study is the first report of microsatellite markers of Persian black cumin. The results of the present study can be useful in finding different markers for breeding programs and developing cultivars with different goals.

**Keywords:** Next Generation Sequencing, Marker, *Denovo Assembly*, Functional Annotation