

## شناسایی ژن‌های مسیر بیوسنتزی اسکلت ترپنوئید در میوه گیاه دارویی هندوانه ابوجهل (*Citrullus colocynthis* (L.) Schrad.) با استفاده از تکنیک توالی‌یابی RNA

معصومه درافشان<sup>۱</sup>، مهدی سلطانی حویزه<sup>۲\*</sup> و وحید شریعتی<sup>۲</sup>

۱- دانشجوی کارشناسی ارشد، گروه ژنتیک و به‌نژادی گیاهی، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران

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

پست الکترونیک: soltani.m@iauhvaz.ac.ir

۳- استادیار، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری، تهران، ایران

تاریخ پذیرش: خرداد ۱۳۹۸

تاریخ اصلاح نهایی: خرداد ۱۳۹۸

تاریخ دریافت: بهمن ۱۳۹۷

### چکیده

توالی‌یابی RNA در حال حاضر یک انتخاب مؤثر و پرسرعت برای مطالعه ترانسکریپتوم گونه‌های گیاهی غیر مدل است که برای شناسایی شبکه‌های ژنی و الگوهای بیان ژن‌های تولیدکننده متابولیت‌های ثانویه در اندام‌های مختلف گیاهان استفاده می‌شود. اصلی‌ترین ترکیب‌ها در بافت‌های میوه و برگ گیاه دارویی هندوانه ابوجهل (*Citrullus colocynthis* (L.) Schrad.) را ترپنوئیدها، فلاونوئیدها و آلکالوئیدها تشکیل می‌دهند. در این مطالعه، ترانسکریپتوم میوه گیاه هندوانه ابوجهل با استفاده از تکنیک توالی‌یابی RNA، بن‌سازه Illumina HiSeq2500 اجرا گردید. بعد از کنترل کیفیت با استفاده از نرم‌افزارهای FastQC و Trimmomatic تعداد ۲۱,۹۵۲,۸۸۵ خوانش دارای کیفیت بالا تولید شد و با استفاده از برنامه Evidential-gene به صورت نوپدید یکپارچه‌سازی شد که منتهی به تولید ۵۵,۳۱۱ تک‌ژن دارای N50 برابر ۹۲۷ جفت باز گردید. توالی تک‌ژن‌های یکپارچه شده در پایگاه KAAS بارگذاری شد. در مجموع ۱۳,۶۵۷ تک‌ژن تفسیر شد که این تعداد در ۱۳۴ مسیر زیستی قرار گرفتند. مسیر "اسکلت ترپنوئید" با تعداد ۹۳ تک‌ژن از پر تعدادترین مسیرهای شناسایی شده از میان ۱,۵۵۲ تک‌ژن مسیر بیوسنتز متابولیت ثانویه بود. با بررسی تک‌ژن‌های مرتبط با مسیر بیوسنتزی اسکلت ترپنوئید، ۲۹ شناسه ژنی (K) شناسایی شد که تمام ژن‌های دو مسیر اصلی بیوسنتزی اسکلت ترپنوئیدی: مسیر سینتوزولی موالونیک اسید (MVA) و پلاستیدی متیل-ارتريتول-فسفات (MEP) از ابتدای مسیر تا تولید ایزوپنیل دی‌فسفات (IPP) را شامل می‌شود. شناسایی ژن‌های مسیر بیوسنتزی اسکلت ترپنوئید امکان تحقق توسعه تجاری محصول گیاه دارویی که پایه‌ای برای پژوهش‌های آینده مربوط به شناسایی مسیرهای بیوسنتزی سایر متابولیت‌های اختصاصی، مهندسی متابولیت، اصلاح مولکولی و به‌نژادی گیاهان دارویی است را فراهم می‌کند.

واژه‌های کلیدی: گیاه دارویی، توالی‌یابی نسل آینده، ترانسکریپتوم، یکپارچه‌سازی نوپدید، پایگاه اطلاعاتی.

## مقدمه

گیاهان بیوشیمی غنی، پیچیده و متنوعی دارند و به فراوانی متابولیت‌های اختصاصی و یا ثانویه تولید کرده که به عنوان محصولات طبیعی شناخته می‌شوند. این متابولیت‌ها ترکیب‌های آلی هستند که غیرضروری بوده، اما به‌طور غیرمستقیم بر بقاء و رشد و نمو گیاه تأثیر می‌گذارند. متابولیت‌های ثانویه گیاه ترکیب‌هایی با وزن مولکولی کم و ساختار متنوع هستند که منابع طبیعی مهم برای توسعه داروها و محصولات طبیعی، فعالیت‌های مختلف بیولوژیکی و اکولوژیکی از جمله سازگاری با محیط‌های مختلف و دفاع در برابر تنش‌های زیستی به‌شمار می‌روند (Harvey *et al.*, 2015؛ Moore *et al.*, 2014). گزارش شده که حدود ۲۵۰,۰۰۰ متابولیت ثانویه در سلول گیاهی وجود دارد، که براساس مبانی بیوسنتز از لحاظ ساختاری و بیوشیمیایی می‌توانند به سه دسته عمده شامل فلاونوئیدها، ترپنوئیدها و آلکالوئیدها تقسیم شوند. به این ترتیب، بیش از ۳۶,۰۰۰ ترپنوئید، ۱۲,۰۰۰ آلکالوئید و ۱۰,۰۰۰ فلاونوئید کشف شده است، اگرچه این نشان‌دهنده تنها کسری از آنچه که در طبیعت وجود دارد، می‌باشد (Chen *et al.*; Fang *et al.*, 2011). (al., 2016).

ترپنوئیدها از یک یا چند واحد C5 به نام ایزوپنتنیل (isopentenyl) یا ایزوپرن (isoprene) تشکیل شده‌اند. اصطلاح "ترین" از ماده اصلی "ترباتین" درخت بنه یا پسته کوهی (*Pistacia terebinthus*) یا ماده اصلی "سقز" (Turpentine) درخت کاج گرفته شده است. این اصطلاح در ابتدا برای توصیف ترکیبی از هیدروکربن‌های گیاهی ایزومریک دارای فرمول مولکولی  $C_{10}H_{16}$  در اسانس استفاده می‌شد (Croteau *et al.*, 2000). ترپنوئیدها با توجه به تعداد واحدهای ایزوپرن درگیر در بیوسنتز آنها به مونوترپن‌ها ( $C_{10}$ )، سزکویی‌ترین‌ها ( $C_{15}$ )، دی‌ترین‌ها ( $C_{20}$ )، سسترتین‌ها ( $C_{25}$ )، تری‌ترین‌ها ( $C_{30}$ ) و پلی‌ترین‌ها طبقه‌بندی می‌شوند. این متابولیت‌های اختصاصی بزرگترین گروه از محصولات طبیعی بوده و کاربردهای متنوعی در صنعت دارند (Bhat *et al.*, 2007). مونوترپن‌ها و

سزکویی‌ترین‌ها جزء اصلی اسانس بوده که نقش در عطر، طعم، رنگدانه‌های فتوسنتزی و فعالیت‌های آنتی‌اکسیدانی داشته، در حالی‌که سایر ترین‌ها جزء عصاره و رزین‌ها هستند که به‌عنوان مواد سمی به‌منظور سازوکارهای ارتباطی و دفاعی در برابر آفات و بیماری‌های گیاهی فعالیت می‌کنند (Steele *et al.*, 1998؛ Langenheim, 1994). ساختار ترپنوئیدها از دو پیش‌ساز اصلی ایزوپنتنیل پیروفسفات (IPP: isopentenyl pyrophosphate) و ایزومر آن دی‌متیل‌آلیل پیروفسفات (DMAPP: dimethylallyl pyrophosphate) تشکیل شده که از طریق مسیر سیتوزولی موالونیک اسید (MVA: Mevalonic acid) یا پلاستییدی متیل‌ارتیتول فسفات (MEP: Methylerythritol phosphate) سنتز می‌شوند (Carretero-Paulet *et al.*, 2002). مسیر MVA عمده‌تاً مسئول سنتز سزکویی‌ترینوئید و تری‌ترینوئید می‌باشد، در حالی‌که مسیر MEP برای بیوسنتز مونوترپنوئید، دی‌ترینوئید و تتراترپنوئید اعمال می‌شود (Rasulov *et al.*, 2015). فارنزیل دی‌فسفات سنتاز (FPPS: Farnesyl diphosphate synthase) یک آنزیم کلیدی در بیوسنتز ایزوپرنوئیدها است که با اتصال دو واحد ایزوپنتنیل پیروفسفات (IPP) به دی‌متیل‌آلیل دی‌فسفات (DMAPP: dimethylallyl diphosphate) تولید فارنزیل دی‌فسفات (FPP: farnesyl diphosphate) می‌کند (شکل ۱). به‌عنوان ماده متشکله جدید در اولین نقطه چند شاخه‌ای در مسیر بیوسنتز ایزوپرنوئید برای چندین ترپنوئید مانند سزکویی‌ترینوئید، استرول‌ها و پلی‌ایزوپرنوئید عمل می‌کند (Dhar *et al.*, 2013؛ Sando *et al.*, 2008).

در حال حاضر دستیابی به دو تحول کلیدی منجر به انقلابی در کشف مسیرهای جدید تولید مواد طبیعی و بیوشیمیایی گیاهان شده است. اولین تحول، دستیابی به موفقیت در بهبود تکنولوژی توالی‌یابی و کاهش هزینه‌های توالی‌یابی می‌باشد که توانایی کشف مسیرهای جدید محصولات طبیعی به‌وسیله شناسایی ژن‌های اختصاصی را تسریع کرده است و امروزه توالی ژنوم و ترانسکریپتوم گونه‌های متعددی از گیاهان در دسترس است. دومین تحول،

استفاده می‌شود (Hussain et al., 2014). کوکوروبیتاسین‌ها یک دسته از ترپینوئیدهای تتراسایکلیک هستند که به‌طور عمده توسط گیاهان خانواده Cucurbitaceae تولید می‌شود و طیف وسیعی از خواص دارویی مانند ضد التهاب، آنتی‌اکسیدان، ضد ویروس، ضد تب، ضد درد و ضد مالاریا را به همراه دارند. کوکوروبیتاسین‌ها نقش مهمی در تولید دارو به‌ویژه در توسعه داروهای ضد سرطان دارند (Chen et al., 2005؛ Hussain et al., 2014؛ Kaushik et al., 2015). در حالی‌که بسیاری از متابولیت‌های ثانویه در هندوانه ابوجهل شناسایی شده‌است، اما مسیرهای بیوسنتز آنها به علت اطلاعات ژنتیکی محدودی که برای گیاه موجود است، هنوز مشخص نیست. پیش از این، مجموع ۵,۰۳۸ تگ‌های توالی بیان شده (EST) از کتابخانه‌های استاندارد cDNA بافت برگ تحت تنش خشکی، شناسایی شده بود (Wang et al., 2014).

با توجه به اینکه مهمترین خواص دارویی گیاه هندوانه ابوجهل مربوط به ترپنوئیدها (کوکوروبیتاسین‌ها) می‌باشد، از این‌رو این مطالعه به‌منظور شناسایی توالی‌های رونوشت‌های ژنی مرتبط با مسیر بیوسنتزی اسکلت ترپنوئیدی در بافت میوه هندوانه ابوجهل انجام شد. رونوشت‌های شناسایی شده در این مطالعه پایه‌ای برای مطالعات آینده مربوط به انجام مهندسی متابولیک برای افزایش بیوسنتز ترپنوئیدها و تنظیم آن برای داروهای گیاهی می‌باشد.

### مواد و روش‌ها

نمونه‌گیری میوه از یک بوته گیاه دارویی هندوانه ابوجهل خودرو در منطقه اندیمشک استان خوزستان با طول و عرض جغرافیایی ۴۸,۱۳۹E و ۳۲,۴۰۲N در سال ۱۳۹۶ انجام شد. پس از برداشت نمونه بلافاصله در نیتروژن مایع به آزمایشگاه تحقیقاتی دانشکده کشاورزی و منابع طبیعی دانشگاه آزاد اسلامی واحد اهواز انتقال داده شد و تا زمان استخراج RNA در فریزر با دمای ۸۰- درجه سانتی‌گراد نگهداری شد. RNA کل از نمونه بافت گیاهی با استفاده از کیت کیاژن مطابق با

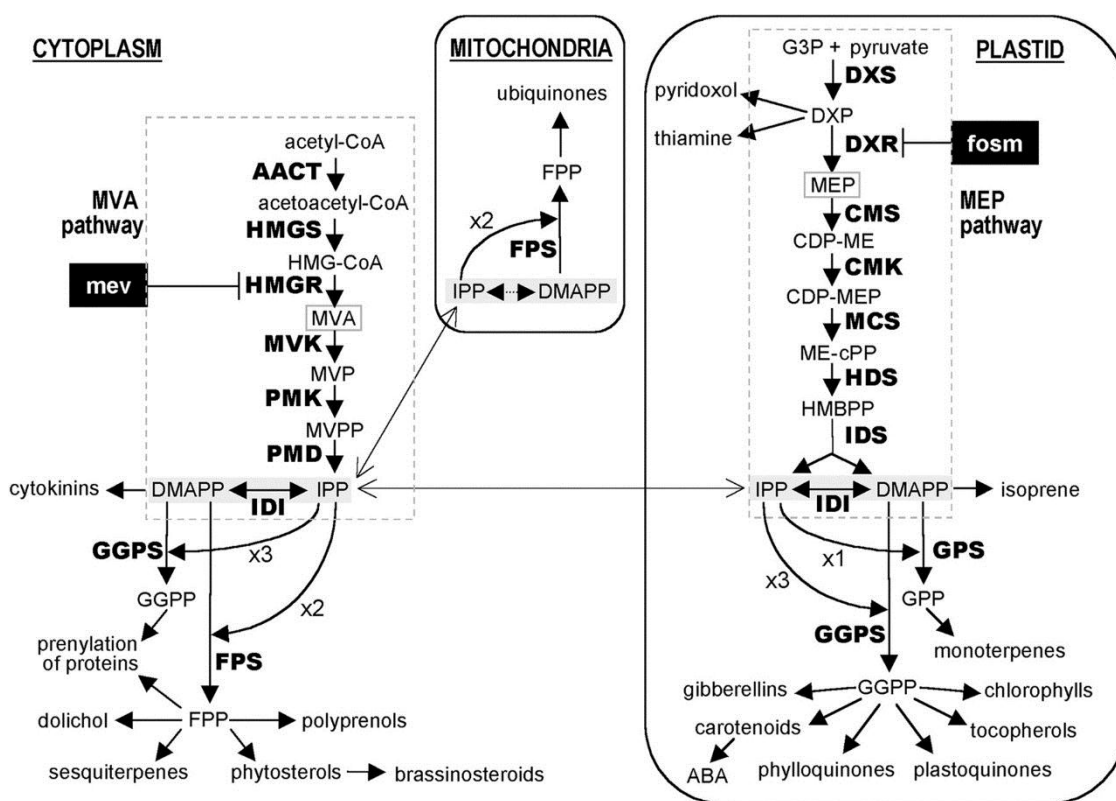
شناخته شدن ژن‌هایی است که کدگذاری برخی از مسیرهای بیوسنتز محصولات طبیعی را به‌صورت خوشه‌ای در ژنوم‌های گیاهی سازماندهی می‌کنند (King et al., 2014؛ Lau & Sattely, 2015).

ظهور فناوری‌های توالی‌یابی نسل جدید، به علت توان بالا، حساسیت و دقت آن، امکان تجزیه و تحلیل ترانسکرپتوم گیاه را فراهم کرده است. توالی‌یابی RNA امکان ایجاد پروفایل بیان ژن، شناسایی رونوشت‌ها و ژن‌های جدید را در فرایندهای مختلف بیولوژیکی ایجاد می‌کند و در حال حاضر مطالعه ترانسکرپتوم گونه‌های گیاهی مدل و غیر مدل که دارای منابع محدود ژنتیکی و ژنومی هستند یک انتخاب مؤثر و قدرتمند است (Dillies et al., 2013؛ Fang et al., 2015). این تکنولوژی برای شناسایی ژن‌های گوناگون از متابولیت‌های ثانویه بر پایه بیوسنتز در گیاهان مختلف استفاده شده است. برای نمونه، بیوسنتز فلاونوئیدها در گیاه گلرنگ (*Carthamus tinctorius*) (Li et al., 2012)، بیوسنتز گلوکوزینولات در تربچه (*Raphanus sativus*) (Wang et al., 2013)، بیوسنتز آلکالوئید در گیاه پنجه‌گره (*Uncaria rhynchophylla*) (Guo et al., 2014)، بیوسنتز جینسنوزید در جنسینگ (*Panax ginseng*) (Jayakodi et al., 2015) و بیوسنتز تیمول در گیاه دارویی زنیان (Soltani Howyzeh et al., 2018) با استفاده از این تکنیک انجام شد.

هندوانه ابوجهل (*Citrullus colocynthis*) گونه یک‌ساله یا چندساله به‌عنوان گیاه دارویی از خانواده کوکوروبیتاسه در مناطق شنی سراسر شمال آفریقا، جنوب‌غربی آسیا و منطقه مدیترانه به‌صورت بومی رشد می‌کند (Dane et al., 2007). این گیاه دارویی مهم یک منبع روغنی با ارزش است، برگ و میوه آن حاوی ترکیب‌های اختصاصی فعال دارویی مانند آلکالوئیدها، تانن‌ها، فلاونوئیدها، ساپونین‌های تری‌ترینی، ترپنوئیدها از جمله کوکوروبیتاسین‌ها و کلوسینتین بوده و به‌طور گسترده‌ای برای درمان بسیاری از بیماری‌ها از جمله دیابت، سرطان، ورم پستان، عفونت‌های میکروبی و التهابی، سقط جنین و غیره

طرفه ۲×۱۵۰ اجرا گردید. کیفیت اولیه خوانش‌ها با استفاده از نرم‌افزار FastQC تعیین شده و توالی‌های دارای طول کوچکتر از ۵۰ نوکلئوتید، نوکلئوتیدهای دارای کیفیت کوچکتر از ۲۰، همچنین آداپتورهای ایلومینا توسط نرم‌افزار Trimmomatic حذف گردید. سپس با استفاده از نرم‌افزار FastQC خصوصیات کتابخانه ارزیابی و کارایی پیرایش داده‌ها تأیید شد.

دستورالعمل شرکت سازنده استخراج شد و در نهایت با آنزیم DNase I تیمار گردید. کیفیت و کمیت نمونه RNA به ترتیب با استفاده از ژل آگارز ۱٪ و اسپکتروفتومتر نانودراپ تعیین شد و نمونه به مؤسسه ژنومی چین (BGI) ارسال گردید. خلوص RNA با استفاده از تکنولوژی Bioanalyzer<sup>®</sup>2100 Agilent مورد تجزیه و تحلیل قرار گرفت؛ سپس توالی‌یابی با استفاده از بن‌سازه Illumina HiSeq2500 به صورت خوانش‌های دو



شکل ۱- دو مسیر بیوسنتزی اسکلت ترپنوئیدها، (برگرفته از: Rodriguez-Concepción & Boronat, 2002)

پایگاه KAAS (<http://www.Genome.Jp/kegg/kaas>) بارگذاری شد و با استفاده از شناسه‌های اختصاصی KO (KEGG Orthology) تفسیر شدند. روش اجرایی شناسه‌های اختصاصی KO روش Single-directional Best Hit (SBH) می‌باشد. شناسه‌های KO مجموعه‌ای از ارتولوگ‌های ژنی لینک شده رکوردی در نقشه‌های مسیر KEGG با هدف پیوند ژنوم با مسیرهای زیستی هستند.

از آنجا که توالی ژنوم گیاه دارویی هندوانه ابوجهل وجود ندارد، از روش یکپارچه‌سازی نوپدید (de novo assembly) ترانسکریپتوم استفاده گردید. به این منظور خوانش‌های دارای کیفیت بالا با استفاده از نرم‌افزار Evidential-gene یکپارچه‌سازی شدند. پس از بهینه‌سازی K-mer استاندارد اجرا شد و برای سایر پارامترهای نرم‌افزار از تنظیمات پیش فرض استفاده گردید. توالی تک‌ژن‌های یکپارچه شده در

## نتایج

متابولیت های ثانویه عمدتاً در میوه گیاه دارویی هندوانه ابوجهل تجمع می یابند (Hussain *et al.*, 2014). بنابراین توالی یابی ترانسکریپتوم میوه این گیاه دارویی با هدف شناسایی ژن های درگیر در بیوسنتز متابولیت های ثانویه انجام شد. بعد از استخراج RNA از بافت میوه و ارزیابی کمیّت و کیفیت نمونه، توالی یابی با استفاده از بن سازه Illumina HiSeq2500 اجرا گردید و حاصل آن تعداد ۲۲،۲۵۵،۲۰۰ خوانش بود که بعد از کنترل کیفیت توالی تعداد ۲۱،۹۵۲،۸۸۵ خوانش دارای کیفیت بالا برای آنالیزهای پردازشی بعدی مورد استفاده قرار گرفت. در گام اول آنالیزهای بیوانفورماتیکی مرحله یکپارچه سازی نوپدید ترانسکریپتوم با استفاده از نرم افزار Evidential-gene، تعداد ۵۵،۳۱۱ تک ژن با میانگین کانتینگ ۱۲۵۸/۵ جفت باز و N50 برابر ۹۲۷ جفت باز تولید شد. در گام بعد توالی تک ژن ها در پایگاه KAAS بارگذاری گردید. مرجع KEGG به عنوان یکی از فعال ترین پایگاه های داده زیستی در جهان از طریق نقشه های مسیر که به صورت گسترده برای تفسیر بیولوژیکی توالی ژنوم و سایر داده های با توان بالا استفاده می شود، مجموعه ژن های ژنوم را به عملکردهای سلول و موجود مرتبط می سازد (Kanehisa *et al.*, 2015).

در این مطالعه از پایگاه اتوماتیک KAAS به منظور نقشه یابی مسیرهای KEGG برای ژن های ارتولوگ اختصاصی گیاه مربوطه استفاده گردید. از تعداد کل ۵۵،۳۱۱ تک ژن یکپارچه شده هندوانه ابوجهل تعداد ۱۷،۳۵۹ مورد از آنها در مسیرهای زیستی و غیرزیستی پایگاه KAAS شناسایی شده که در کل ۱۳،۶۵۷ تک ژن موجود در این پایگاه در ۱۳۴ مسیر گیاهی تفسیر شدند. بر تعدادترین مسیرها به ترتیب "مسیرهای متابولیکی" (۲،۹۹۴ تک ژن، ko01100)، "بیوسنتز متابولیت ثانویه" (۱،۵۵۲ تک ژن، ko01110)، "ریبوزوم" (۵۱۹ تک ژن،

ko03010)، "متابولیسم کربن" (۳۹۲ تک ژن، ko01200)، "بیوسنتز آمینواسید" (۳۶۷ تک ژن، ko01230) و "انتقال سیگنال هورمون گیاهی" (۳۴۷ تک ژن، ko04075) تخمین زده شدند. بیشترین تعداد تک ژن ها از میان تمام مسیرهای متابولیت ثانویه، در مسیر "بیوسنتز فنیل پروپانویید" (۲۰۰ تک ژن، ko00940) و پس از آن "بیوسنتز اسکلت ترپنوئید" (۹۳ تک ژن، ko00900) بودند.

تری ترپنوئیدها به عنوان اجزای اصلی تشکیل دهنده عصاره هندوانه ابوجهل از پیش ماده IPP حاصل از مسیر MVA سنتز می شوند. نتایج حاصل از تفسیر کارکردی تک ژن ها در پایگاه KEGG به شناسایی رونوشت های مختص ۲۹ شناسه (K) مسیرهای بیوسنتزی MEP و MVA در خوانش های ترانسکریپتوم بافت میوه هندوانه ابوجهل منتهی گردید. آنزیم های درگیر در مسیر MVA شناسایی شده در این پژوهش شامل استیل-کوآنزیم آ استیل ترانسفراز، HMG-کوآنزیم آ سنتاز، HMG-کوآنزیم آ ردوکتاز (HMGCR)، موالونات کیناز، فسفوموالونات کیناز و موالونات دی فسفات دکربوکسیلاز و به همین ترتیب آنزیم های مسیر MEP شامل ۱- دئوکسی-دی-زایلولوز ۵-فسفات سنتاز (DXS)، ۱- دئوکسی-دی-زایلولوز ۵-فسفات رداکتوایزومر (DXR)، ۲-سی-متیل-دی-اریتریتول ۴-فسفات سیتیدیل ترانسفراز (ispD)، ۴-دی فسفوسیتیدیل ۲-سی-متیل-دی-اریتریتول کیناز (ispE)، ۲-سی-متیل-دی-اریتریتول ۲و۴-سیکلو دی فسفات سنتاز (ispF)، ۴-هیدروکسی-۳-متیل بوت-۲-انیل-دی فسفات سنتاز (gcpE)، ۴-هیدروکسی-۳-متیل بوت-۲-انیل-دی فسفات ردوکتاز (ispH)، ایزوپنتیل-دی فسفات دلتا ایزومراز (idi) و ژرانیل-دی فسفات سنتاز (GPS) شدند (جدول ۱ و شکل ۲).

جدول ۱- آنزیم‌های درگیر در بیوسنتز اسکلت ترپنوئید در ترانسکریپتوم گیاه دارویی هندوانه ابوجهل

## براساس نتایج پایگاه KAAS

عدد EC*	توصیف آنزیم	شناسه تک‌ژن هندوانه ابوجهل	شناسه K
[EC:2.2.1.7]	dxs; 1-deoxy-D-xylulose-5-phosphate synthase	33740, 95520, 509617, 730548, 8750, 1189983, 1268809	K01662
[EC:1.1.1.267]	dxr; 1-deoxy-D-xylulose-5-phosphate reductoisomerase	1005801, 56519, 1023516	K00099
[EC:2.7.7.60]	ispD; 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	12661, 127222, 1116685, 153036	K00991
[EC:2.7.1.148]	ispE; 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	506184, 1163043	K00919
[EC:4.6.1.12]	ispF; 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	720263, 489723, 993082, 1099967	K01770
[EC:1.17.7.3 1.17.7.1]	gcpE; (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	509657, 1097004, 1259782	K03526
[EC:1.17.7.4]	ispH; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	42158, 179994	K03527
[EC:5.3.3.2]	idi; isopentenyl-diphosphate Delta-isomerase	33154, 93439, 1159413, 1260616, 1159409	K01823
[EC:2.3.1.9]	ACCAT; acetyl-CoA C-acetyltransferase	10324, 20091, 699229, 1030656, 963217, 1120889, 1120895, 1334413, 1352923	K00626
[EC:2.3.3.10]	HMGS; hydroxymethylglutaryl-CoA synthase	1003842	K01641
[EC:1.1.1.34]	HMGCR; hydroxymethylglutaryl-CoA reductase (NADPH)	27801, 913699, 1002723, 194244	K00021
[EC:2.7.1.36]	MK; mevalonate kinase	33952, 128555, 292084, 506433, 1140796, 1331006	K00869
[EC:2.7.4.2]	PMK; phosphomevalonate kinase	39418, 507640	K00938
[EC:4.1.1.33]	MVD; diphosphomevalonate decarboxylase	1289237, 980205	K01597
[EC:2.5.1.10 2.5.1.1]	FDPS; farnesyl diphosphate synthase	15685, 37220, 502293, 1166876	K00787
[EC:2.5.1.1]	GPS; geranyl diphosphate synthase	1203334, 174202, 824700, 1093120	K14066
[EC:2.5.1.29 2.5.1.10 2.5.1.1]	GGPS; geranylgeranyl diphosphate synthase, type II	16349, 64778, 82102, 128527	K13789
[EC:1.3.1.111 1.3.1.83]	chlP, bchP; geranylgeranyl diphosphate/geranylgeranyl-bacteriochlorophyllide a reductase	1034, 506044	K10960
[EC:2.5.1.85 2.5.1.84]	SPS; all-trans-nonaprenyl-diphosphate synthase	506117	K05356
[EC:2.5.1.87]	DHDDS; ditrans.polycis-polyprenyl diphosphate synthase	500780, 502464, 997637	K11778
[EC:2.5.1.58]	FNTB; protein farnesyltransferase subunit beta	6943, 1148769, 1291262, 1291267	K05954
[EC:2.5.1.59 2.5.1.58]	FNTA; protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha	821461	K05955
[EC:3.4.22.-]	RCE1, FACE2; prenyl protein peptidase	191654, 610381, 705453, 1086585	K08658
[EC:3.4.24.84]	STE24; STE24 endopeptidase	620605	K06013
[EC:2.1.1.100]	ICMT; protein-S-isoprenylcysteine O-methyltransferase	1134250	K00587
[EC:3.1.1.-]	PCME; prenylcysteine alpha-carboxyl methylesterase	1513, 1977, 506390, 1173481	K15889
[EC:1.8.3.6 1.8.3.5]	PCYOX1, FCLY; prenylcysteine oxidase / farnesylcysteine lyase	506539, 191826	K05906
[EC:1.1.1.354]	FLDH; NAD+-dependent farnesol dehydrogenase	503459, 1028192	K15891
[EC:2.7.1.216]	FOLK; farnesol kinase	977862, 1159827	K15892

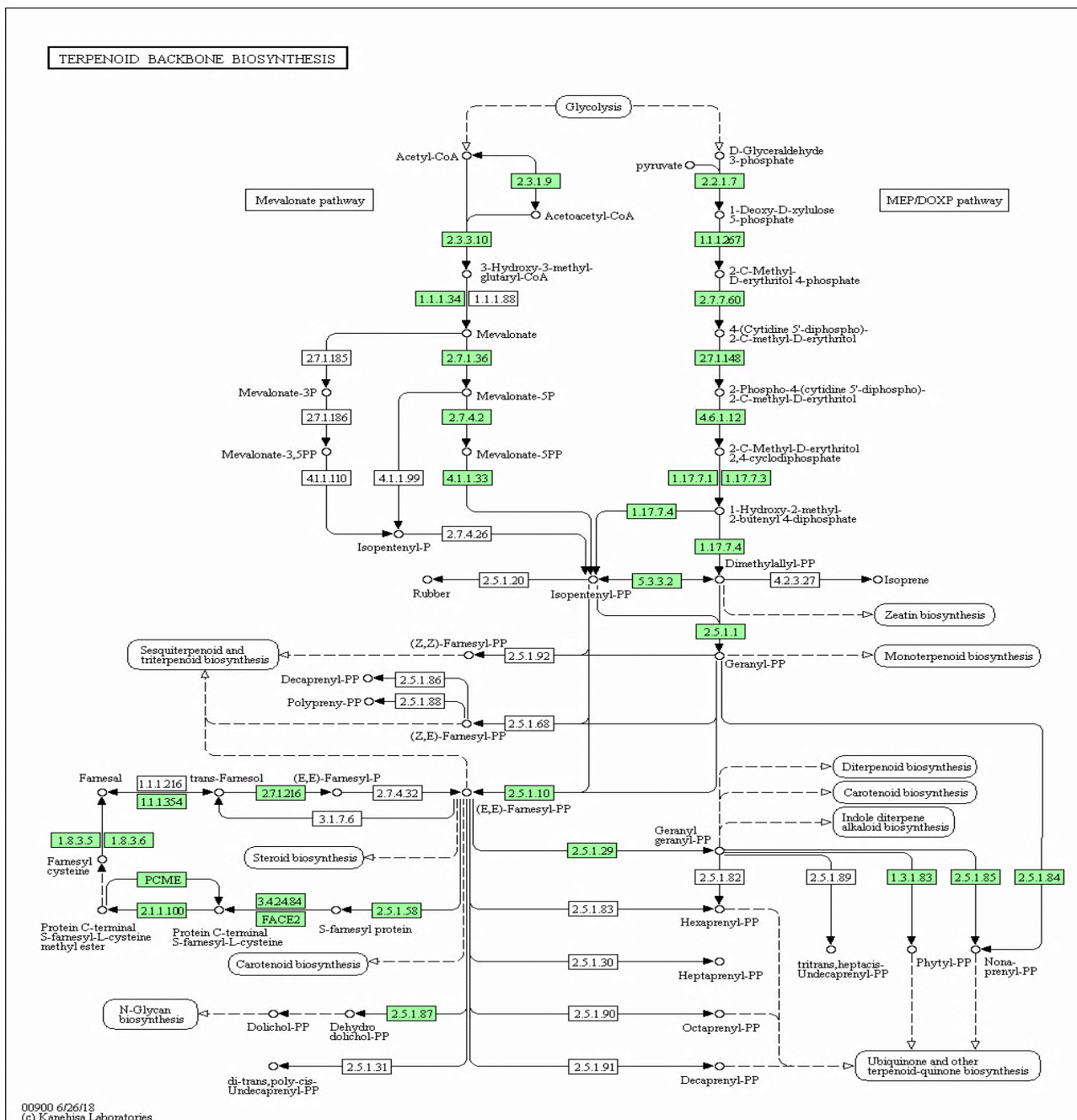
## بحث

استفاده از مطالعه تحلیلی ترانسکریپتوم تاکنون منجر به شناسایی تعداد زیادی از ژن‌های مرتبط با بیوسنتز متابولیت‌های ثانویه با ارزش در گیاهان دارویی از جمله انواع ترپنوئیدها شده است. شناخت سازوکارهای مولکولی بیوسنتز ترپنوئیدها برای بهبود عملکرد و کیفیت آن ضروری می‌باشد. تجزیه و تحلیل RNA-Seq براساس تکنولوژی توالی‌یابی نسل جدید یک روش قدرتمند برای شناسایی ژن‌های کاندید کدکننده آنزیم‌هایی است که مسئول بیوسنتز متابولیت‌های ثانویه جدید در یک گیاه غیر مدل هستند. در ادامه به عنوان نمونه به تعدادی از مطالعات RNA-Seq در خانواده کوکوربیتاسه می‌توان اشاره نمود. میوه راهب یا میوه مونک (*Siraitia grosvenorii*) حاوی پودر شیرین، گوشتی و خوراکی است که عصاره میوه آن دارای عنصر اصلی بیولوژیکی، نوعی از تری‌ترین ساپونین مانند کوکوربیتاسین بوده که به عنوان مورگوسید وی (mogroside V) شناخته می‌شود. مورگوسید تقریباً ۳۰۰ برابر شیرین‌تر از ساکارز است. باین حال در گیاه *S. grosvenori* بیوسنتز این ماده به مقدار کمی در اواخر مراحل مسیر شناخته شده است. براساس الگوی بیان ژن‌ها، آنزیم‌های سیتوکروم P450 (CYP450s) و UDP-glucosyltransferase (UDPGs) به عنوان نامزدهای درگیر در بیوسنتز مورگوسید انتخاب شدند (Tang et al., 2011).

میوه رسیده کدو خاردار (*Momordica cochinchinensis*) که به نام گاک (Gac) معروف است سرشار از کاروتنوئیدهایی مانند بتاکاروتن و لیکوپن می‌باشد. این گیاه یک منبع خوب برای مهندسی متابولیت کاروتنوئید است، تمامی ژن‌های موجود در مسیرهای بیوسنتز اسکلت ترپنوئید و بیوسنتز کاروتنوئید این گیاه به وسیله توالی‌یابی ترانسکریپتوم شناسایی شده‌اند. از تجزیه و تحلیل رابطه بین ژن‌های بیوسنتز احتمالی

کاروتنوئید و تغییرات میزان کاروتنوئید در طی رسیدن میوه، آنالیز بیان ژن بر روی سه مرحله مختلف رسیدن آریل (Aril) - پوشش فرعی دانه - انجام شد. این مطالعه نشان داد که سنتز آنزیم‌های فیتوئن (phytoene)، ۱۵-سیس-فیتون دزاتوراز (15-cis-phytone desaturase)، زتا-کاروتن دزاتوراز (zeta-carotene desaturase)، کاروتنوئید ایزومراز (desaturase carotenoid isomerase) و لیکوپن اپسیلون سایکلاز (epsilon cyclase lycopene)، ممکن است عامل کلیدی برای کنترل محتوای کاروتنوئید در طی رسیدن آریل باشند (Hyun et al., 2012). در گیاه خربزه تلخ یا کارلا (*Momordica charantia*) تجزیه و تحلیل ژن‌های درگیر در مسیر بیوسنتز تری‌ترینوئید از اندام‌های بالغ مختلف، با استفاده از بن‌سازه ایلومینا انجام شد. پانزده ژن کدکننده آنزیم‌های مرتبط با بیوسنتز تری‌ترینوئید شناسایی شد. بیشتر ژن‌ها در مراحل بلوغ در گلها به ویژه گل‌نر و یا میوه‌ها به شدت بیان شده بود که نشان می‌دهد این ژن‌ها نقش مهمی در بیوسنتز کارانتین (Charantin) در کارلا دارند. همچنین شرایط نور مطلوب برای افزایش بیوسنتز کارانتین در کارلا مورد بررسی قرار گرفت که در نتیجه نور قرمز مؤثرترین طول موج بود. کارانتین یک نوع تری‌ترینوئید کوکوربیتان است که عملکرد مفید دارویی مانند فعالیت ضد سرطان، ضد دیابت و ضد باکتری دارد (Cuong et al., 2017).

درخت نفت یا فریون هند (*Euphorbia tirucalli*)، حاوی دو استروئید دارویی فعال اتوفول و بتا-سیتوسترول در لاتکس آن می‌باشد. در مطالعه‌ای آنالیز ترانسکریپتوم این گیاه برای شناسایی ژن‌های بیوسنتز اسکلت استرول انجام شد که منجر به شناسایی ژن‌های ۶ آنزیم در مسیر MVA و ۷ آنزیم در مسیر MEP گردید (Qiao et al., 2018).



شکل ۲- آنالیز مسیر KEGG نمایانگر ژن‌های درگیر در بیوسنتز اسکلت ترپنوئید

مستطیل‌های خاکستری آنزیم‌های موجود در ترانسکریپتوم گیاه دارویی هندوانه ابوجهل را نشان می‌دهد.

شده است. در این مطالعه ۱۹۸ ژن در مسیر MVA و MEP، ژن‌های بیوسنتز اسکلت ترپنوئید و ۶۹ ژن سنتز ترپنوئید شناسایی شده که احتمالاً درگیر در متابولیسم

بیشتر ژن‌های مرتبط با دو مسیر بیوسنتزی MVA و MEP متابولیت‌های ثانویه در گیاه مریم‌گلی ( *Salvia guaranitica* ) با استفاده از توالی یابی RNA شناسایی



و عصاره گیاهان دارویی مؤثر واقع گردد. تحقیقات بر روی مسیره‌های بیوسنتزی متابولیت‌های ثانویه و عملکردهای مرتبط می‌تواند پایه‌ای برای پژوهش‌های بعدی با پیشبرد اهداف مهندسی متابولیت، اصلاح مولکولی، به‌نژادی گیاه و گیاهان مهندسی ژنتیک شده در جهت توسعه محصولات داروهای گیاهی ایجاد نماید.

### منابع مورد استفاده

- Ali, M., Hussain, R.M., Rehman, N.U., She, G., Li, P., Wan, X., Guo, L. and Zhao, J., 2018. *De novo* transcriptome sequencing and metabolite profiling analyses reveal the complex metabolic genes involved in the terpenoid biosynthesis in Blue Anise Sage (*Salvia guaranitica* L.). *DNA Research*, 25(6): 597-617.
- Amiripour, M., Sadat Nouri, S.A., Shariati, V. and Soltani Howyzeh, M., 2018. Identification of terpenoid backbone biosynthetic pathway genes in Ajowan (*Trachyspermum ammi* L.) by RNA-Seq. *Journal of Novin Genetic*, 13(1): 133-141.
- Bhat, S.V., Nagasampagi, B.A. and Sivakumar, M., 2007. *Chemistry of Natural products*. New York: Springer Berlin Heidelberg, 830p.
- Cárdenas-Conejo, Y., Carballo-Uicab, V., Lieberman, M., Aguilar-Espinosa, M., Comai, L. and Rivera-Madrid, R., 2015. *De novo* transcriptome sequencing in *Bixa orellana* to identify genes involved in methylerythritol phosphate, carotenoid and bixin biosynthesis. *BMC Genomics*, 16(1): 877.
- Carretero-Paulet, L., Ahumada, I., Cunillera, N., Rodríguez-Concepción, M., Ferrer, A., Boronat, A. and Campos, N., 2002. Expression and molecular analysis of the Arabidopsis DXR gene encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-c-methyl-d-erythritol 4-phosphate pathway. *Plant physiology*, 129(4): 1581-1591.
- Chen, J.C., Chiu, M.H., Nie, R.L., Cordell, G.A. and Qiu, S.X., 2005. Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Natural product reports*, 22(3): 386-399.
- Chen, Q., Ma, C., Qian, J., Lan, X., Chao, N., Sun, J. and Wu, Y., 2016. Transcriptome sequencing of *Gynostemma pentaphyllum* to identify genes and enzymes involved in triterpenoid biosynthesis. *International Journal of Genomics*, 2016(10): 1-10.
- Croteau, R., Kutchan, T.M. and Lewis, N.G., 2000. *Natural products (secondary metabolites)*.

ترپنوئید هستند (Ali et al., 2018). آنالیز ترانسکریپتوم گیاه دارویی زنیان (*Trachyspermum ammi* L.) از خانواده چتریان که بذره‌های آن دارای اصلی‌ترین ترکیب‌های مونوترپن‌های تیمول، گاما-تریپنن و پاراسیمن هستند، منتهی به شناسایی ۱۲۷ مسیر متابولیکی شامل مسیره‌های MVA و MEP و ۳۰ ژن کلیدی مرتبط با مسیر بیوسنتزی اسکلت‌ترین شد (Amiripour et al., 2018).

صدها ژن جدید در مسیره‌های متابولیکی MEP و MVA مرتبط با بیوسنتز ترپنوئیدها، در گیاهان مختلف مانند جینسینگ (*Panax notoginseng*) (Luo et al., 2011)، شب‌بو (*Litsea cubeba*) (Han et al., 2013)، گیاه *Thapsia laciniata* (Drew et al., 2013)، گوجه‌فرنگی (*Solanum lycopersicum*) (Spyropoulou et al. 2014)، هویج (*Daucus carota*) (Yahyaa et al., 2015)، درخت آناتو (*Bixa Orellana*) (Cárdenas-Conejo et al., 2015)، جینسینگ آمریکایی (*Panax japonicas*) (Rai et al., 2016)، عشقه یا پیچک (*Hedera helix*) (Sun et al., 2017)، سیب‌زمینی شیرین (*Dioscorea zingiberensis*) (Li et al., 2018)، بابونه آلمانی (*Matricaria chamomilla*) (Zhang et al., 2018) و سایر گیاهان دارویی شناسایی شده است.

در این پژوهش استفاده از RNA-Seq منجر به شناسایی ژن‌های مرتبط با مسیر بیوسنتزی اسکلت ترپنوئید در گیاه دارویی هندوانه ابوجهل شد. این مطالعه نه تنها اولین تجزیه و تحلیل نوپدید ترانسکریپتوم در بافت میوه هندوانه ابوجهل را نشان می‌دهد، بلکه اطلاعاتی را در مورد شناسایی و بیان رونوشت‌های مربوط به تری‌ترپنوئید و سزکویی‌ترپنوئید که یکی از مهمترین متابولیت‌های اختصاصی فارماکولوژیک فعال توزیع شده در بافت‌های متفاوت هندوانه ابوجهل هستند، فراهم می‌کند. در مهندسی متابولیک، شناسایی ژن‌های مرتبط با بیوسنتز متابولیت‌های ثانویه می‌تواند در به‌نژادی این مسیره‌ها برای افزایش متابولیت‌های اختصاصی در اسانس

- discovery in the genomics era. *Nature Reviews Drug Discovery*, 14(2): 111-129.
- Hussain, A.I., Rathore, H.A., Sattar, M.Z., Chatha, S.A., Sarker, S.D. and Gilani, A.H., 2014. *Citrullus colocynthis* (L.) Schrad. (bitter apple fruit): A review of its phytochemistry, pharmacology, traditional uses and nutritional potential. *Journal of Ethnopharmacology*, 155(1): 54-66.
  - Hyun, T.K., Rim, Y., Jang, H.J., Kim, C.H., Park, J., Kumar, R., Lee, S., Kim, B.C., Bhak, J., Nguyen-Quoc, B. and Kim, S.W., 2012. *De novo* transcriptome sequencing of *Momordica cochinchinensis* to identify genes involved in the carotenoid biosynthesis. *Plant Molecular Biology*, 79(4-5): 413-427.
  - Jayakodi, M., Lee, S.C., Lee, Y.S., Park, H.S., Kim, N.H., Jang, W., Lee, H.O., Joh, H.J. and Yang, T.J., 2015. Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biology*, 15(1): 138.
  - Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M., 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1): 457-462.
  - Kaushik, U., Aeri, V. and Mir, S.R., 2015. Cucurbitacins-an insight into medicinal leads from nature. *Pharmacognosy reviews*, 9(17): 12.
  - King, A.J., Brown, G.D., Gilday, A.D., Larson, T.R. and Graham, I.A., 2014. Production of bioactive diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. *The Plant Cell*, 26(8): 3286-3298.
  - Langenheim, J.H., 1994. Higher plant terpenoids: a phytocentric overview of their ecological roles. *Journal of Chemical Ecology*, 20(6): 1223-1280.
  - Lau, W. and Sattely, E.S., 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349(6253): 1224-1228.
  - Li, H., Dong, Y., Yang, J., Liu, X., Wang, Y., Yao, N., Guan, L., Wang, N., Wu, J. and Li, X., 2012. *De novo* transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids. *PloS one*, 7(2): e30987.
  - Li, J., Liang, Q., Li, C., Liu, M. and Zhang, Y., 2018. Comparative transcriptome analysis identifies putative genes involved in dioscin biosynthesis in *Dioscorea zingiberensis*. *Molecules*, 23(2): E454.
  - Luo, H., Sun, C., Sun, Y., Wu, Q., Li, Y., Song, J., Niu, Y., Cheng, X., Xu, H., Li, C. and Liu, J., 2011. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics*, 12(5): S5.
  - Biochemistry and Molecular Biology of Plants, 24: 1250-1319.
  - Cuong, D.M., Jeon, J., Morgan, A.M., Kim, C., Kim, J.K., Lee, S.Y. and Park, S.U., 2017. Accumulation of charantin and expression of triterpenoid biosynthesis genes in bitter melon (*Momordica charantia*). *Journal of Agricultural and Food Chemistry*, 65(33): 7240-7249.
  - Dane, F., Liu, J. and Zhang, C., 2007. Phylogeography of the bitter apple, *Citrullus colocynthis*. *Genetic Resources and Crop Evolution*, 54(2): 327-336.
  - Dhar, M.K., Koul, A. and Kaul, S., 2013. Farnesyl pyrophosphate synthase: a key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development. *New Biotechnology*, 30(2): 114-123.
  - Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J. and Guernec, G., 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14(6): 671-683.
  - Drew, D.P., Dueholm, B., Weitzel, C., Zhang, Y., Sensen, C.W. and Simonsen, H.T., 2013. Transcriptome analysis of *Thapsia laciniata* Rouy provides insights into terpenoid biosynthesis and diversity in Apiaceae. *International Journal of Molecular Sciences*, 14(5): 9080-9098.
  - Fang, X., Yang, C.Q., Wei, Y.K., Ma, Q.X., Yang, L. and Chen, X.Y., 2011. Genomics grand for diversified plant secondary metabolites. *Plant Diversity and Resources*, 33(1): 53-64.
  - Fang, Y., Huang, J., Huang, X., Chen, S.H., Zou, P.C., Li, W.S., Yu, K. and Liu, Y.W., 2015. Generation of expressed sequence tags from a cDNA library of *Coleus forskohlii* for identification of genes involved in terpene biosynthesis. *Biologia Plantarum*, 59(3): 463-468.
  - Guo, Q., Ma, X., Wei, S., Qiu, D., Wilson, I.W., Wu, P., Tang, Q., Liu, L., Dong, S. and Zu, W., 2014. *De novo* transcriptome sequencing and digital gene expression analysis predict biosynthetic pathway of rhynchophylline and isorhynchophylline from *Uncaria rhynchophylla*, a non-model plant with potent anti-alzheimer's properties. *BMC Genomics*, 15(1): 676.
  - Han, X.J., Wang, Y.D., Chen, Y.C., Lin, L.Y. and Wu, Q.K., 2013. Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. *PloS one*, 8(10): e76890.
  - Harvey, A.L., Edrada-Ebel, R. and Quinn, R.J., 2015. The re-emergence of natural products for drug

- activate terpene synthase promoters. *BMC Genomics*, 15(1): 402.
- Steele, C.L., Katoh, S., Bohlmann, J. and Croteau, R., 1998. Regulation of oleoresinosis in grand fir (*Abies grandis*): differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. *Plant Physiology*, 116(4): 1497-1504.
  - Sun, H., Li, F., Xu, Z., Sun, M., Cong, H., Qiao, F. and Zhong, X., 2017. *De novo* leaf and root transcriptome analysis to identify putative genes involved in triterpenoid saponins biosynthesis in *Hedera helix* L. *PLoS One*, 12(8): e0182243.
  - Tang, Q., Ma, X., Mo, C., Wilson, I.W., Song, C., Zhao, H., Yang, Y., Fu, W. and Qiu, D., 2011. An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics*, 12(1): 343.
  - Wang, Y., Pan, Y., Liu, Z., Zhu, X., Zhai, L., Xu, L., Yu, R., Gong, Y. and Liu, L., 2013. *De novo* transcriptome sequencing of radish (*Raphanus sativus* L.) and analysis of major genes involved in glucosinolate metabolism. *BMC Genomics*, 14(1): 836.
  - Wang, Z., Hu, H., Goertzen, L.R., McElroy, J.S. and Dane, F., 2014. Analysis of the *Citrullus colocynthis* transcriptome during water deficit stress. *PLoS One*, 9(8): e104657.
  - Yahyaa, M., Tholl, D., Cormier, G., Jensen, R., Simon, P.W. and Ibdah, M., 2015. Identification and characterization of terpene synthases potentially involved in the formation of volatile terpenes in carrot (*Daucus carota* L.) roots. *Journal of agricultural and food chemistry*, 63(19): 4870-4878.
  - Zhang, W., Tao, T., Liu, X., Xu, F., Chang, J. and Liao, Y., 2018. *De novo* assembly and comparative transcriptome analysis: novel insights into sesquiterpenoid biosynthesis in *Matricaria chamomilla* L. *Acta Physiologiae Plantarum*, 40(7): 129.
  - Moore, B.D., Andrew, R.L., Külheim, C. and Foley, W.J., 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist*, 201(3): 733-750.
  - Qiao, W., Li, C., Mosongo, I., Liang, Q., Liu, M. and Wang, X., 2018. Comparative transcriptome analysis identifies putative genes involved in steroid biosynthesis in *Euphorbia tirucalli*. *Genes*, 9(1): E38.
  - Rai, A., Yamazaki, M., Takahashi, H., Nakamura, M., Kojoma, M., Suzuki, H. and Saito, K., 2016. RNA-seq transcriptome analysis of *Panax japonicus*, and its comparison with other *Panax* species to identify potential genes involved in the saponins biosynthesis. *Frontiers in Plant Science*, 7: 481.
  - Rasulov, B., Talts, E., Kannaste, A. and Niinemets, U., 2015. Bisphosphonate inhibitors reveal a large elasticity of plastidic isoprenoid synthesis pathway in isoprene-emitting hybrid aspen. *Plant Physiology*, 168(2): 532-548.
  - Rodríguez-Concepción, M. and Boronat, A., 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. a metabolic milestone achieved through genomics. *Plant Physiology*, 130(3):1079-1089.
  - Sando, T., Takaoka, C., Mukai, Y., Yamashita, A., Hattori, M., Ogasawara, N., Fukusaki, E. and Kobayashi, A., 2008. Cloning and characterization of mevalonate pathway genes in a natural rubber producing plant, *Hevea brasiliensis*. *Bioscience, Biotechnology, and Biochemistry*, 72(8): 2049-2060.
  - Soltani Howyzeh, M., Sadat Nouri, S.A., Shariati, V. and Amiripour, M., 2018. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. *Scientific Reports*, 8(1): 13405.
  - Spyropoulou, E.A., Haring, M.A. and Schuurink, R.C., 2014. RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that

## Identification of terpenoid backbone biosynthetic pathway genes in fruit of *Citrullus colocynthis* (L.) Schrad. medical plant by RNA sequencing

M. Dorafshan<sup>1</sup>, M. Soltani Howyzeh<sup>2\*</sup> and V. Shariati<sup>3</sup>

1- M.Sc. student, Genetic and Plant Breeding Departemant, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

2\*- Corresponding Author, Genetic and Plant Breeding Departemant, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

E-mail: soltani.m@iauahvaz.ac.ir

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

Received: February 2019

Revised: May 2019

Accepted: June 2019

### Abstract

Nowadays, RNA sequencing is an effective and fast choice to study the transcripts of non-model plant species used to identify gene networks and patterns of gene expression producing secondary metabolites in different plant organs. Terpenoids, flavonoids, and alkaloids are the main compounds in fruits and leaves of *Citrullus colocynthis* (L.) Schrad. In this study, the transcriptome of *C. colocynthis* fruit was performed using RNA-Seq technique, Illumina HiSeq2500 platform. After quality control using FastQC and Trimmomatic software, 21,952,885 high-quality reads were produced and the *de novo* assembly with the Evidential-gene program resulted in the production of 55,311 unigenes with an N50 equal to 927 base pairs. The sequence of assembled unigenes was loaded on the KAAS database. A total of 13,657 unigenes were annotated, matched into 134 biosynthetic pathways. The "terpenoid backbone" biosynthetic pathway with 93 unigenes was one of the most numerous identified pathways among 1,552 unigenes of secondary metabolites biosynthetic pathway. By examining the unigenes associated with the biosynthetic pathway of the terpenoid backbone, 29 gene identifiers (K number) of the pathway were detected, which contained all identified genes of the two main biosynthetic pathways, Mevalonic Acid (MVA) and Methylerythritol Phosphate (MEP) pathways, from the beginning of the path to the production of isoprenyl diphosphate (IPP). Identification of the genes in the biosynthetic pathway of the terpenoid backbone makes it possible to realize the commercial development of the medicinal plant products, providing the basis for further research on the identification of biosynthetic pathways of other specific metabolites, metabolite engineering, molecular breeding, and medical plants breeding.

**Keywords:** Medicinal plant, next generation sequencing, transcriptome, *de novo* assembly, database.