

## ارزیابی ترانسکرپتوم گیاه باریجه (*Ferula gummosa*) با استفاده از روش RNA-seq

احمد سبحانی نجف آبادی<sup>۱</sup>، محمدرضا نقوی<sup>۲\*</sup>، حمید فرحمنده<sup>۱</sup>، علیرضا عباسی<sup>۳</sup>

۱، ۲ و ۳. دانشجوی دکتری، استاد و دانشیار گروه زراعت و اصلاح نباتات پردیس کشاورزی و منابع طبیعی دانشگاه تهران.

۴. استاد گروه بیوتکنولوژی پردیس کشاورزی و منابع طبیعی دانشگاه تهران

(تاریخ دریافت: ۱۳۹۵/۱۰/۲۸ - تاریخ پذیرش: ۱۳۹۵/۱۲/۱۱)

### چکیده

گیاه باریجه با نام علمی *Ferula gummosa* Boiss از تیره چتریان (Apiaceae) بومی مناطق شرق و غرب ایران است. گیاه باریجه به دلیل داشتن قابلیت استحصال شیرابه-ای با ارزش دارویی و صنعتی، بسیار مورد توجه است. اگر چه این شیرابه از حیث صادرات گیاهان دارویی مرتعی ایران در رتبه اول قرار دارد و خواص دارویی و کاربردهای صنعتی فراوانی دارد، اما اطلاعات اندکی در زمینه مکانیسم های بیوشیمیایی و ژنتیکی تولید این ترکیبات وجود دارد. در همین راستا مطالعات ترانسکرپتومیکس کمک بسیاری به این زمینه از علم گیاهان دارویی کرده است. در حال حاضر تکنیک RNA-Seq جزو پر کاربردترین و دقیق ترین روش های ترانسکرپتومیکس به شمار می رود. در تحقیق حاضر جهت ارزیابی ترانسکرپتوم باریجه از روش RNA-seq برای دو اندام ریشه و گل استفاده گردید. خوانش های بدست آمده از دستگاه Illumina HiSeq2000 پس از تعیین کیفیت، با نرم افزار Trinity مونتاژ و سپس از نظر عملکرد مورد ارزیابی قرار گرفتند. ترانسکرپتوم حاصله دارای ۱۵۸۴۳۶ ترانسکرپت با N50 برابر با ۱۳۴۸ نوکلئوتید بود که پس از بلاست با دیتابیس های مختلف ۱۰۰۲۷۵ ترانسکرپت از نظر عملکردی تفسیر شدند. ارزیابی نتایج بلاست nr از نظر هستی شناسی ژن ها (Gene ontology) منجر به شناسایی ۷۳۶۷۸ ترانسکرپت دارای حداقل یک GO گردید که از بین آن ها از نظر فرایندهای زیستی، فرایندهای سلولی و متابولیکی بیشترین GO ها را به خود اختصاص دادند (۴۴/۵۵ درصد) که این حالت نشان دهنده وضعیت فعال متابولیکی گیاه می باشد. ارزیابی ترانسکرپتوم از نظر ژن های ارتولوگ (دیتابیس KOG) منجر به شناسایی ۴۲۴۵۲ ژن از ۳۹۲۵ خانواده عملکردی در ۲۶ کلاس کلی گردید. بزرگترین کلاس حاصل مرتبط با مکانیسم های انتقال سیگنالی (۱۴/۲۹ درصد) می باشد. کلاس متابولیت های ثانویه، انتقال و کاتابولیسم با فراوانی ۴/۲۵ درصد (۱۸۰۶ ترانسکرپت) از جمله فعالترین کلاس های عملکردی در گل و ریشه باریجه می باشند.

واژه های کلیدی: گیاهان دارویی، توالی یابی، متابولیت های ثانویه، فرایندهای زیستی.

## Evaluation of transcriptome of (*Ferula gummosa*) using RNA-seq method

Ahmad Sobhani Najafabadi<sup>1</sup>, MohammadReza Naghavi<sup>1</sup>, Hamid Farahmand<sup>2</sup>, AliReza Abbas<sup>1</sup>

1. Agronomy and Plant Breeding Dept., Faculty of Agriculture, University of Tehran, Iran,  
2. Biotechnology Dept., Faculty of Agriculture, University of Tehran, Iran.

(Received: January 17, 2017 - Accepted: March 1, 2017)

### ABSTRACT

*Ferula gummosa* Boiss is a herbaceous plant from the Apiaceae family that is indigenous to East and West of Iran. Galbanum is an oleo-gum-resin with high pharmaceutical and industrial values, which is exuded from *F. gummosa* in response to wounds. Although galbanum ranks first in terms of Iranian export of pasture medicinal plants, little information exists on the genetic and biochemical mechanisms of these compounds are existed. In this regard, transcriptomics studies of medicinal plants have much to contribute to this field of science. RNA-Seq techniques are now one of the most used and most accurate methods in transcriptomics. In the present study, RNA-Seq was used to evaluate the root and flower of *F. gummosa* transcriptome. Short reads obtained from Illumina HiSeq2000 were assembled by Trinity software after trimming and filtering for quality scores below 30. The assembled transcriptome with 158,436 transcripts and N50 1,348 bp was annotated using different databases such as NCBI nr-proteins, Plant sub-database of NCBI nr-proteins, Swiss-Prot, TAIR10 and PLAZA 3. In total, 100,275 transcripts were annotated using BLASTX with an E-value cut-off of  $10^{-5}$ . Gene ontology (GO) classification of annotated transcripts was done by Blast2Go and Trinotate pipeline and in total 73,678 transcripts with at least one GO were identified. In terms of biological processes, cellular and metabolic processes are the most GO terms (44.55 %) indicated that the plant is in a metabolically active status. Transcriptome assessment of orthologous genes by NCBI-KOG database led to the identification of 42,452 genes from 3,925 families and 26 functional classes. The largest class of orthologous genes was associated with signal transduction mechanisms (14.29 %), followed by other classes, such as secondary metabolites-transport and catabolism with a frequency of 4.25 % (1,806 transcripts) indicating the most active status of *F. gummosa* roots and flowers for secondary metabolite production and transportation.

\* Corresponding author E-mail: mnaghavi@ut.ac.ir

**Keywords:** Biological processes, pharmaceutical plants, sequencing, secondary metabolites.

با ترکیبات ارزشمند را کشف کند (Spyropoulou et al., 2011; Tang et al., 2014).

طی سال‌های اخیر مطالعات بسیار زیادی از روش RNA-seq برای پی بردن به مکانیسم پیچیده مقاومت‌ها و متابولیسم ترکیبات مؤثره گیاهان دارویی استفاده کرده‌اند. کراری و همکاران با مطالعه روی پروفایل متابولیتی و پروفایل بیان ژن گوجه فرنگی در مراحل مختلف رشد، موفق به شناسایی رفتارهای ژن‌ها طی فرایند رسیدن گوجه فرنگی شدند و از این طریق ژن‌های درگیر در مسیر بیوسنتز لیکوپین در گوجه فرنگی را مورد مطالعه قرار دادند (Carrari et al., 2006). شی و همکاران با همین روش ژن‌های کاندید مسیرهای بیوسنتز متابولیت‌های مهم چای را مورد بررسی قرار دادند (Shi et al., 2011). گوپتا و همکاران با همین روش متابولیسم فلاونوئیدهای بلوبری را مورد مطالعه قرار دادند (Gupta et al., 2015). وایدیا و همکاران با استفاده از این تکنیک‌ها ژن‌های درگیر در مسیر بیوسنتز دیاسونین را شناسایی کردند (Vaidya et al., 2013). کیونگ هیون و همکاران با اندازه‌گیری پروفایل متابولیت‌های توت سیاه در مراحل مختلف و تلفیق داده‌ها با پروفایل RNA-Seq آن، موفق به شناسایی ژن‌های درگیر در بیوسنتز آنتوسیانین‌های توت سیاه شدند (Hyun et al., 2014). پروژه دیگر، ژن‌های درگیر در مسیرهای متابولیسمی کارتنوئیدها، آلکالوئیدها، فلاونوئیدها، ترپنوئیدها و بطور کلی متابولیت‌های ثانویه را مورد مطالعه قرار داده‌اند.

یا توجه به فقدان داده‌های مولکولی روی این گونه و دیگر گونه‌های بومی از خانواده چتریان لزوم انجام تحقیقات ژنومیکس و ترانسکریپتومیکس بیش از پیش ضروری بنظر می‌رسد (تا کنون هیچ داده توالی‌یابی از گونه‌های جنس *Ferula* در آرشیو SRA سایت NCBI گزارش نشده است). وجود داده‌هایی از این دست در درک بیشتر تولید ترکیبات ارزشمندی چون شیرابه باریجه بسیار راه‌گشا می‌باشند. در تحقیق پیش رو ترانسکریپتوم گیاه باریجه توالی‌یابی، بازسازی و از نظر عملکردی تفسیر و دسته‌بندی گردیده است تا با کمک این اطلاعات ارزشمند، مکانیسم‌های بیوسنتزی این ترکیبات بیش از پیش شناسایی گردند و در نهایت راهی برای اصلاح مولکولی این گیاه و دیگر گیاهان

## مقدمه

گیاهان دارویی یکی از مهم‌ترین منابع امیدوارکننده برای تولید انواع داروها و همچنین درمان انواع بیماری‌ها محسوب می‌شوند. به گونه‌ای که تاکنون بسیاری از ترکیبات دارویی مهم و فعال از منابع گیاهی مختلف استخراج و مورد استفاده قرار گرفته است (Julsing et al., 2008) با این وجود، در برخی موارد معرفی منابع اولیه برای تولید چنین ترکیبات دارویی (در مقیاس وسیع و بصورت پایدار) با برخی چالش‌ها و مشکلاتی همراه بوده است (Sarfaraaj Hussain et al., 2012). ایران به عنوان کشوری غنی از نظر پوشش و تنوع گیاهان دارویی، رویشگاه گیاهانی می‌باشد که جنبه‌های دارویی آن‌ها به اثبات رسیده است، ولی مطالعات مولکولی و بیوشیمیایی کمتر روی آن‌ها صورت گرفته است.

گیاه باریجه با نام علمی *Ferula gummosa* Boiss از تیره چتریان، Apiaceae، بومی مناطق شرق و غرب ایران است و در استان‌هایی نظیر سمنان، خراسان، اصفهان، فارس و استان‌های مرکزی کشور پراکنده شده است (Shahbazi et al., 2011). این گیاه علفی، چندساله و مونوکارپیک است که در سال آخر رویش (سال پنجم تا هشتم) به ساقه می‌رود و تشکیل گل و میوه می‌دهد. هرگونه خراش و یا گزش در ریشه و یا ساقه منجر به خروج شیرابه می‌گردد (Ghannadi & Amree, 2002). در طب سنتی این گیاه به منظور تسکین دردها، درمان بیماری‌های داخلی و به عنوان ضدعفونی کننده زخم‌ها مورد استفاده قرار می‌گرفته است. مطالعات مختلف، اثرات دارویی این گیاه را تأیید کرده‌اند (Abbaszadegan et al., 2010; Seyed et al., 2015). در همین راستا مطالعات ترانسکریپتومیکس کمک بسیاری به این زمینه از علم گیاهان دارویی کرده‌اند. به طوری که بسیاری از گیاهان دارویی مهم دنیا به منظور شناسایی رابطه بین ترکیبات ارزشمند و ژن‌های مرتبط با بیوسنتز آنها تحت مطالعه قرار گرفته‌اند (Colebatch et al., 2004; Yang et al., 2007; Zulak et al., 2009). تکنیک RNA-Seq که در حال حاضر جزو پر کاربردترین و دقیق‌ترین روش‌های ترانسکریپتومیکس به شمار می‌رود ابزاری است که می‌تواند با راندمان بالا روابط ذکر شده بین ژن‌های مرتبط

دارویی گشوده شود.

## مواد و روش‌ها

### نمونه‌برداری

نمونه‌های گیاهی تحقیق حاضر در خرداد ۱۳۹۳ از ارتفاعات روستای شهرستانک واقع در جاده چالوس در ارتفاع ۲۲۰۰ متری از سطح دریا، عرض جغرافیایی ۳۵ درجه، ۵۸ دقیقه و ۱۰ ثانیه شمالی و طول جغرافیایی ۵۱ درجه، ۲۱ دقیقه و ۱۳ ثانیه شرقی جمع‌آوری شدند. از آنجا که ریشه و گل از نظر تولید ترپن‌ها (بخش قابل توجه شیرابه باریجه) فعالیت بالایی دارند، این دو اندام جهت مطالعات ترانسکریپتوم مورد استفاده قرار گرفتند. نمونه‌های گل و ریشه از یک گیاه سالم به طول تقریبی ۱۲۰ سانتی‌متر، در اواخر مرحله گلدهی (قبل از ریزش گلبرگ‌های زرد) گرفته شده و با کمک مرکز ملی ذخایر ژنتیک از نظر گیاه‌شناسی مورد تأیید قرار گرفتند. دو اندام ریشه و گل گیاه بطور جداگانه با ازت مایع فریز شدند و برای مراحل بعدی کار در دمای -۲۰ درجه سانتیگراد نگهداری شدند.

### استخراج و ارزیابی RNA

اولین مرحله در آنالیز ترانسکریپتوم استخراج RNA با کیفیت و کمیت مناسب می‌باشد. از آنجا که باریجه گیاهی چند ساله با ریشه‌های چوبی و چوب پنبه‌ای همراه با شیرابه زیاد می‌باشد، استخراج RNA با کیفیت از اندام‌های مختلف آن بخصوص ریشه بسیار سخت و وقت‌گیر می‌باشد. در این تحقیق روش‌های مختلف استخراج RNA از جمله استفاده از کیت‌های ستونی و کیت‌های رسوبی از جمله تریزول و بایوزول و روش‌های دستی مورد ارزیابی قرار گرفتند که از بین این روش‌ها روش استفاده از کیت بایوزول (BioFlux) بهترین و موثرترین روش بوده است (Johnson *et al.*, 2012). RNA استخراج شده با دستگاه‌های نانودراپ ۲۰۰۰، بایوآنالیزر ۲۱۰۰ و الکتروفورز ژل آگارز مورد ارزیابی قرار گرفت.

### ساخت کتابخانه cDNA و توالی‌یابی

اولین مرحله در تولید یک کتابخانه cDNA، جداسازی mRNA کل از یک نوع سلول یا بافت مورد نظر است. از آنجا که mRNAهای کدکننده آنزیم‌های درگیر در تولید متابولیت‌های ثانویه دارای دم پلی‌آدنین می‌باشند، جهت جداسازی آن‌ها از rRNA و tRNAهای موجود در

سلول از زنجیره کوتاهی از تیمیدیلات (الیگو dT) متصل به بستر ستون کروماتوگرافی؛ استفاده شد. پس از جداسازی mRNAها و ساخت cDNA دو رشته‌ای، در نهایت کتابخانه cDNA توسط دستگاه Illumina HiSeq™ 2000 با استفاده از فناوری Paired-end با طول قرائت ۱۵۰ نوکلئوتید طبق دستورالعمل کمپانی BGI توالی‌یابی شدند.

### تعیین کیفیت و مونتاژ خوانش‌ها

نتایج حاصل از دستگاه توالی‌یابی Illumina 2000™ HiSeq خوانش‌های خام نامیده می‌شوند که در قالب فایل فسستکیو ذخیره می‌شوند. داده خام حاوی خوانش‌های نامطلوبی است که می‌تواند در آنالیزهای بیوانفورماتیکی بعدی اثر منفی داشته باشد. بنابراین توالی‌ها با معیار Q20 (معیاری برای سنجش کیفیت توالی‌ها)، سنجیده شده و توالی‌های خام با کیفیت مناسب برای مونتاژ استفاده شدند. لازم به ذکر است که فایل‌های حاوی خوانش‌های ریشه و گل به صورت جداگانه در قسمت آرشیو خوانش‌های توالی‌های (SRA) سایت NCBI ثبت شدند (SRX225025, SRX2253871).

نرم افزار Trinity v2.2.0 روی یک کامپیوتر سرور با رم ۲۵۶ گیگا بایت و دارای ۲۴ هسته سی پی یو ۲ گیگاهرتزی و سیستم عامل لینوکس اوبونتو برای مونتاژ این خوانش‌های خام استفاده گردید (Haas *et al.*, 2013). جهت بررسی ترانسکریپتوم حاصل، تعداد ترانسکریپت و یونی ژن مونتاژ شده، میانگین طول آن‌ها و N50 با نرم افزار Trinitystats.pl محاسبه شدند. و ترانسکریپتوم مونتاژ شده جهت تعیین کیفیت مونتاژ، با نرم افزار Transrate مورد ارزیابی قرار گرفت (Li *et al.*, 2014).

### تفسیر عملکردی ترانسکریپتوم

ترانسکریپتوم باریجه توسط نرم‌فزارهای مختلف از جمله Blast2GO و BLAST+ با دیتابیس‌هایی چون NCBI-nr، PlantCyc، TAIR10، Swiss-Prot، Plaza 3.0، KEGG و همردیفی شدند تا توالی‌های مشابه با آن‌ها شناسایی شوند. BLASTX با حدآستانه‌ی (E-value ≤ 10<sup>-5</sup>) برای تفسیر عملکردی ترانسکریپتوم با دیتابیس‌های مذکور استفاده گردید. از آنجا که دیتابیس nr دارای حجم بسیار زیادی می‌باشد برای بلاست با این دیتابیس از

### ترانسکریپتوم

جهت توالی‌یابی RNA از روش ساخت کتابخانه cDNA با روش oligo-dT استفاده گردید. جهت توالی‌یابی ترانسکریپتوم از دستگاه Illumina HiSeq2000 استفاده گردید که از نظر تولید خوانش‌های با کمیت و کیفیت بالا بسیار مطلوب می‌باشد. از آنجا که برای باریجه ژنوم رفرنسی وجود نداشت و توالی‌یابی RNA آن نیز برای بار اول انجام می‌گرفت، استفاده از پایپ لاین مونتاژ از نو جهت بازسازی ترانسکریپتوم و آنالیزهای بعدی اجتناب‌ناپذیر بود. که در این موارد هر چه طول و تعداد خوانش‌ها بیشتر باشند بهتر است. بنابراین در این تحقیق از روش توالی‌یابی دو سویه با طول خوانش ۱۵۰ نوکلئوتید و مقدار حداقل ۵ گیگا باز برای هر نمونه (به عنوان مثال ۵ گیگا باز برای خوانش‌های سمت راست یکی از نمونه‌ها) استفاده گردید. مشخصات خوانش‌های بدست آمده برای هر یک از نمونه‌ها در جدول ۱ اشاره شده است. جهت تعیین کیفیت خوانش‌ها از نرم افزار FastQC و جهت فیلتر و برش بازهای نامناسب از نرم افزار FASTX-Toolkit استفاده شد. نتایج ارزیابی کیفیت خوانش‌ها نشان داد که کیفیت خوانش‌ها از نظر میانگین کیفیت تک تک نوکلئوتیدها (شکل ۱-۲) و میانگین کیفیت خوانش‌ها (شکل ۲-۲) از وضعیت مطلوبی برخوردار می‌باشند. مونتاژ خوانش‌ها جهت ایجاد ترانسکریپتوم رفرنس با نرم افزار Trinity و تنظیمات پیش فرض روی یک کامپیوتر سرور با سیستم عامل لینوکس انجام گرفت. ترانسکریپتوم حاصل دارای ۱۵۸۴۳۶ ترانسکریپت با طول متوسط ۸۲۱/۵۳ نوکلئوتید می‌باشد. جهت تعیین کیفیت ترانسکریپتوم مونتاژ شده از نرم افزار Transrate استفاده گردید. ارزیابی ترانسکریپتوم با این نرم افزار نشان داد که ۷۳/۱ درصد خوانش‌های گل و ۸۰/۱۱ درصد از خوانش‌های ریشه در تولید ترانسکریپتوم استفاده شده اند. که این نشان از راندمان بالای استفاده از خوانش‌ها می‌باشد. بطور کلی هر چه این عدد بزرگتر باشد نشان از کیفیت بالاتر مونتاژ است، ضمن اینکه کمتر از ۶۰ درصد بودن این اعداد نشان‌دهنده راندمان پایین مونتاژ یا کیفیت پایین خوانش‌ها می‌باشد (Smith-Unna *et al.*, 2016).

کامپیوتر سرور با ۲۴ هسته سی پی یو استفاده شد. از آنجا که بلاست nr کاری زمان بر است در این تحقیق سعی شد راندمان بلاست با زیر مجموعه گیاهی nr نیز مورد ارزیابی قرار بگیرد. نتایج حاصل از بلاست با دیتابیس‌های مختلف توسط نرم افزار R با هم مقایسه گردید (Correa & Silva, 2011).

### ارزیابی GO و KOG ترانسکریپتوم

آنالیز GO کل ترانسکریپتوم با نرم افزارهای Trinotate، Blast2Go و Funrich انجام گرفت و بدین ترتیب کلیه یونیژن‌های دارای GO شناسایی شده و دسته بندی آن‌ها در سه دسته Molecular، Biological process و Cellular component در نرم افزارهای R و Exell انجام گرفت (Inglis *et al.*, 2013; Supek *et al.*, 2011). آنالیز KOG از طریق بلاست (rpsBLAST) ترانسکریپتوم با دیتابیس KOG وبسایت NCBI انجام گرفت. نتایج بلاست با فایل حاوی اطلاعات تکاملی ژن‌ها داده کاوی گردید و یونیژن‌ها در گروه‌های ارتولوگ دسته بندی شدند (Annadurai *et al.*, 2012).

### نتایج و بحث

#### استخراج و ارزیابی کیفی و کمی RNA

همانطور که اشاره گردید در این آزمایش روش‌های مختلفی برای استخراج RNA استفاده گردید که به دلیل خاصیت رزینی و فنولی اندام‌های باریجه، بخصوص ریشه، اکثر آن‌ها با شکست مواجه گردید. در نهایت استخراج با روش بایوزول انجام گرفت و نتایج مطلوب حاصل گردید. در این آزمایش سعی شد RNA بدست آمده کیفیت و کمیت مطلوبی داشته باشد و در نهایت برای توالی‌یابی و سنتز cDNA از RNA هایی استفاده شد که نسبت ۲۶۰ به ۲۸۰ بیشتر از ۲، نسبت ۲۶۰ به ۲۳۰ بیشتر از ۱/۸ (NanoDrop-2000)، نسبت RNA ریبوزومی s ۲۸ به s ۱۸ بیشتر از ۱/۵ (الکتروفورز رو ژل آگارز ۱ درصد) و اینتگریتی (RIN) بیشتر از ۸ (بایو آنالیزر با دستگاه Agilent Bioanalyzer-2100) باشند. نمونه ای از عکس ژل الکتروفورز و نتایج بایوآنالیزر RNA های استفاده شده برای دو اندام ریشه و گل باریجه در شکل آورده شده است.

#### توالی‌یابی، ارزیابی خوانش‌ها و مونتاژ

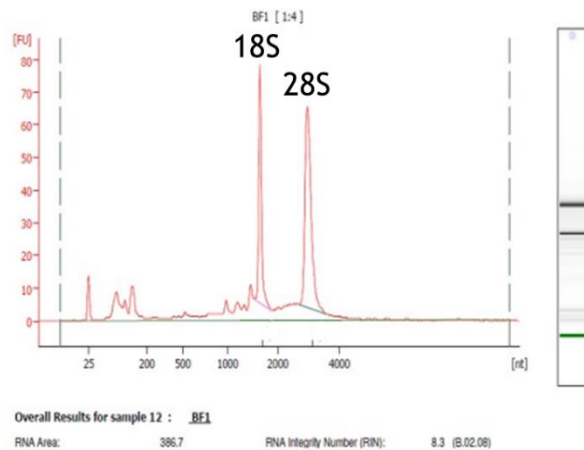


Fig 1-2

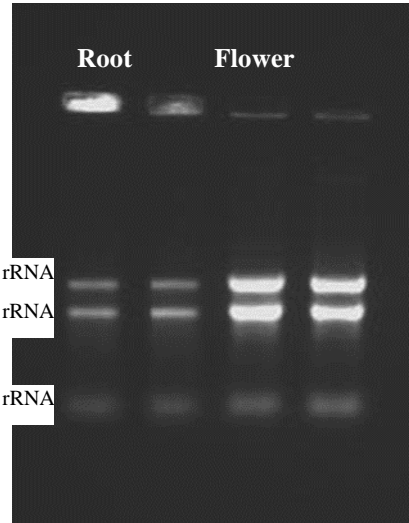


Fig 1-1

شکل ۱: ارزیابی RNA استخراج شده جهت ساخت کتابخانه cDNA، شکل ۱-۱: ارزیابی روی ژل ۱ درصد آگارز، شکل ۱-۲: ارزیابی با دستگاه بیوآنالیزر ۲۱۰۰. Figure 1: Evaluation of extracted RNA for making cDNA library. Fig 1-1: Evaluation applied on 1% agarose gel, Fig 1-2: Evaluation by Bioanalyzer machine 2100.

جدول ۱: کمیت و کیفیت خوانش های بدست آمده از توالی یابی با دستگاه Illumina HiSeq2000  
 Table 1: Short reads quantity and quality from sequencing by Illumina HiSeq2000

Sample	Read length (bp)	Number of Reads	Q20 (%)	GC (%)
<b>Root</b>	150	41,403,916	97.12:96.98	43.87
<b>Flower</b>	150	35,028,552	97.92:95.97	53.55

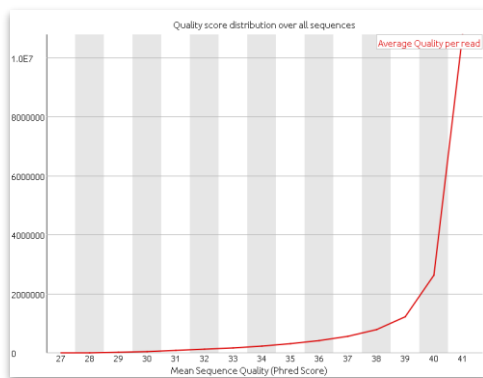


Fig 2-2

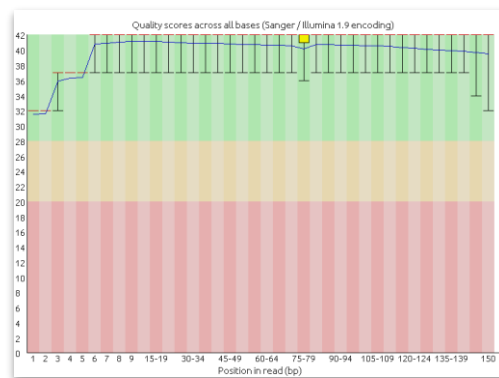


Fig 2-1

شکل ۲: نتایج تعیین کیفیت خوانش ها، شکل ۲-۱: میانگین کیفیت نوکلئوتیدها، شکل ۲-۲: میانگین کیفیت خوانش ها. Figure 2: Short reads quality control by FastQC. Fig 2-1: Quality scores across all bases, and Fig 2-2: Quality score distribution over all reads.

Prot، TAIR10 و PLAZA 3.0 استفاده گردید. بدین منظور فایل دیتابیس های مذکور دانلود شده و پس از ساخت دیتابیس با کمک نرم افزار BLAST+ همردیفی ترانسکرپتوم با این دیتابیس ها با استفاده از BLASTX انجام گرفت (E-value ≤ 10<sup>-5</sup>). برای شناسایی و مستند

**شناسایی و تفسیر عملکردی ترانسکرپتوم با دیتابیس های مختلف**  
 جهت شناسایی و مستندسازی ترانسکرپتوم باریجه از همردیفی ترانسکرپتوم با دیتابیس های پروتئینی رایج از جمله دیتابیس پروتئین های غیر تکراری NCBI، Swiss-

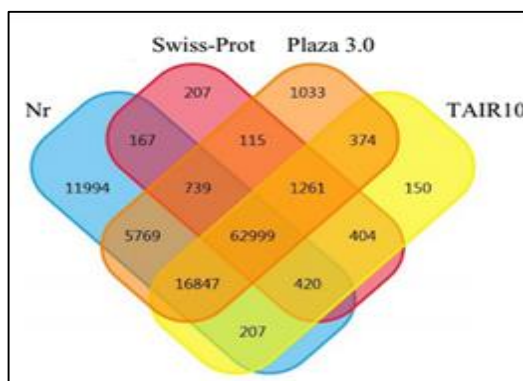
مجموعه ساخته شد و سپس BLASTX مشابه با موارد قبل انجام گرفت. بطور کلی ۱۰۰۲۷۵ (۶۳/۲۹ درصد) ترانسکریپت با موفقیت مستند سازی گردید که سهم هر دیتابیس در این پروسه در جدول ۲ و شکل ۳ آمده است.

سازی ژن‌های درگیر در مسیرهای بیوسنتز متابولیت‌های ثانویه از دیتابیس‌های KEGG و PlantCyc استفاده گردید. جهت بررسی راندمان و سرعت مستند سازی با زیر دیتابیس گیاهی پروتئین‌های غیر تکراری NCBI، این زیر

جدول ۲: نتایج حاصل از BLASTX ترانسکریپتوم با دیتابیس‌های مختلف ( $E\text{-value} \leq 10^{-5}$ )

Table 2: Results of aligning transcriptome with different databases (BLASTX,  $E\text{-value} \leq 10^{-5}$ )

Databases	Number of Hits
NCBI nr-Proteins	99,144 (62.57 %)
Swiss-Prot	66,312 (41.85 %)
TAIR10	82,662 (52.17 %)
PLAZA 3.0	89,137 (56.26 %)
KEGG	43,859 (27.68 %)
PlantCyc	49,053 (30.96 %)
nr زیر دیتابیس گیاهی	89,205 (56.30 %)
Total	(100,275 (63.29 %)



شکل ۳: مقایسه دیتابیس‌های مختلف استفاده شده برای تفسیرسازی ترانسکریپتوم باریجه. نواحی مشترک نشان دهنده توالی‌های تفسیر شده مشترک بین دیتابیس‌های مختلف می‌باشد.

Figure 3: Comparison of databases used to annotate the transcriptome of *F. gummosa*. The overlapping sections represent the shared transcripts that are annotated by different databases.

دیتابیس NCBI-nr از روش ساخت زیر دیتابیس گیاهی این دیتابیس استفاده گردید. زیر دیتابیس گیاهی این دیتابیس حجمی حدود ۱۰ درصد دیتابیس اصلی را دارد. مقایسه زیر دیتابیس گیاهی nr با دیتابیس اصلی (در صورت مشابه بودن سیستم مورد استفاده و پارامترهای بلاست) نشان داد که زمان مورد نیاز برای بلاست کل ترانسکریپتوم با زیر دیتابیس گیاهی nr، ۲۴ برابر کمتر خواهد بود. از نظر تعداد ترانسکریپت دارای هیت، با استفاده از دیتابیس اصلی NCBI-nr تعداد ۹۹۳۷ ترانسکریپت بیشتر از زیر دیتابیس گیاهی آن شناسایی و

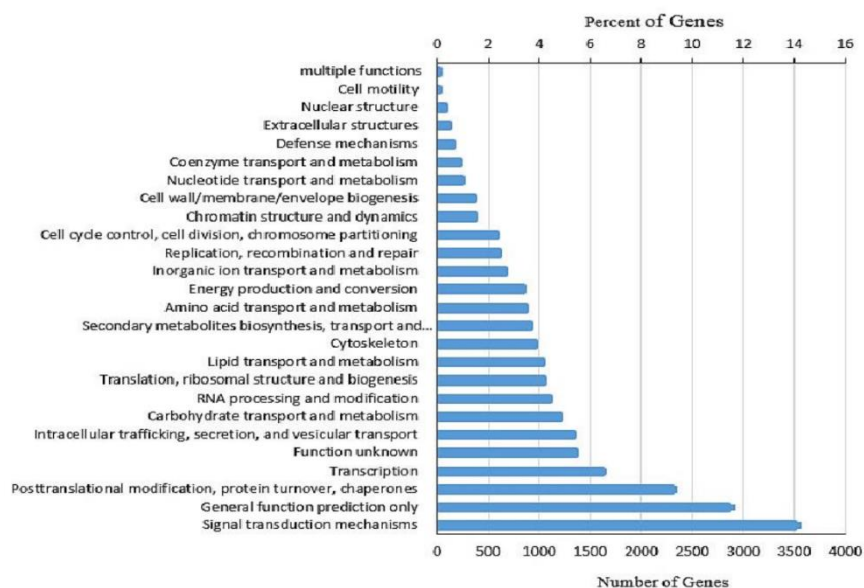
همانطور که انتظار می‌رفت بیشترین تعداد ترانسکریپت دارای هیت از بلاست توالی‌ها با دیتابیس NCBI-nr حاصل گردید. که این نتیجه نشان دهنده کامل تر بودن این دیتابیس نسبت به سایر دیتابیس‌ها می‌باشد. کامل تر بودن دیتابیس NCBI-nr ناشی از زیاد بودن تعداد توالی و حجم دیتابیس می‌باشد که همین موضوع منجر به سرعت پایین تفسیرسازی ترانسکریپتوم با این دیتابیس می‌شود. بنابراین استفاده از این دیتابیس برای تعداد زیادی توالی مثل ترانسکریپتوم باریجه تنها با استفاده از سرور قدرتمند و صرف زمان زیاد مقدور می‌باشد. به منظور کاهش حجم

بالا در ۶۰ گروه و سه دسته اصلی فرایندهای بیولوژیکی، عملکردهای مولکولی و اجزای سلولی طبقه بندی شدند. در دسته فرایندهای بیولوژیکی بیشترین GO مرتبط با پروتئین‌های درگیر در فرایندهای سلولی (۲۵۶۰۷ ژن، ۲۲/۲۸ درصد)، فرایندهای متابولیکی (۲۲/۲۷ درصد)، فرایندهای بیوسنتزی (۸/۵۴ درصد)، تنظیمات بیولوژیکی (۶/۸۷ درصد) و پاسخ به تحریکات (۶/۷۳ درصد) می‌باشد. این نتایج نشان‌دهنده آن است که گیاه در وضعیت فعال متابولیکی قرار داشته است. از نظر عملکردهای مولکولی بیشترین GO ها مرتبط با عملکردهای اتصال (۲۷۵۸۹ ژن، ۲۴/۵۷ درصد)، فعالیت‌های کاتالیکی (۲۰/۴۷ درصد)، عملکردهای اتصال به ممبران اندامک‌ها (۱۰/۹۹ درصد)، اتصال نوکلئوتیدی (۹/۷۲ درصد) و اتصال نوکلئوتیدهای پورین (۸/۴۰ درصد) می‌باشند. در دسته اجزای سلولی بیشترین GO ها در دسته سلولی (۲۳۵۹۹ ژن، ۲۱/۰۲ درصد)، درون سلولی (۱۴/۹۰ درصد)، اندامکی (۱۲/۱۲ درصد)، اندامک‌های درون سلولی (۱۲/۱۱ درصد) و ممبران اندامکی (۱۱/۰۷ درصد) می‌باشند. برای دسته بندی بیشتر ژن‌ها بر اساس عملکرد آن‌ها، از روش مقایسه ژن‌ها با دیتابیس گروه‌های پروتئینی ارتولوگ استفاده گردید (شکل ۴).

تفسیر گردید. ارزیابی بیشتر این ۹۹۳۷ ترانسکرپت نشان داد که این توالی‌ها مربوط به آلودگی‌های باکتریایی، ویروسی، قارچی، حشرات و سایر موجودات زنده دارای روابط پارازیتی یا همزیستی با باریجه می‌باشند. بر اساس نتایج حاصل از تفسیر سازی ترانسکرپتوم با دیتابیس NCBI-nr در مجموع ۱۳۳۱۸ (۸/۳۹ درصد) ترانسکرپت دارای شباهت با پروتئین‌های انگور، ۶۴۱۰ (۴/۰۴ درصد) با قهوه کانفورا و ۶۳۸۹ (۴/۰۲ درصد) با کنجد می‌باشند.

### دسته بندی ترانسکرپتوم باریجه از نظر GO و KOG

در تحقیق حاضر جهت طبقه بندی ترانسکرپتوم (ترانسکرپتوم بدست آمده از گل و ریشه باریجه) بر اساس وظایف ژن‌ها از دو روش طبقه بندی بر اساس هستی‌شناسی ژن‌ها (GO)، و طبقه بندی بر اساس وظایف ژن‌های ارتولوگ (KOG) استفاده گردید. جهت ارزیابی GO برای ترانسکرپت‌های شناسایی شده با دیتابیس NCBI-nr (دارای هیت) GO های موجود استخراج شده و سپس دسته بندی می‌شوند. در کل برای ۷۳۶۷۸ (۴۶/۵۰ درصد) ترانسکرپت‌ها و ۴۲۱۵۱ (۳۷/۵۵ درصد) یونیزن‌های باریجه حداقل یک GO شناسایی شد. GO های با فراوانی



شکل ۴: طبقه بندی ژن‌های باریجه بر اساس عملکرد آن‌ها در مقایسه با پروتئین‌های کلاستر بندی شده (KOG).

Figure 4: KOG classification of *F. gummosa* transcriptome.

ارزشمندی گردید که این اطلاعات راه را برای سایر مطالعات اصلاحی باز می‌نماید. نتایج نشان داد که جهت تفسیر عملکردی ترانسکریپتوم تلفیقی از دیتابیس‌های متفاوت مطلوب می‌باشد اما برای گیاهان بجای استفاده از دیتابیس nr می‌توان از زیر دیتابیس گیاهی nr استفاده کرد که در اینصورت با راندمان مشابه صرفه جویی زیادی در وقت و هزینه خواهد شد. همچنین نتایج دسته بندی ترانسکریپتوم از نظر GO و KOG نشان داد که گل و ریشه باریجه از نظر تولید متابولیت‌های ثانویه در وضعیت فعالی قرار دارند.

### سپاسگزاری

این پروژه با حمایت مالی مرکز مطالعات و همکاری‌های علمی بین‌المللی انجام شده است.

در مجموع ۴۲۴۵۲ ترانسکریپت و ۲۴۵۳۸ ژن در ۳۹۲۵ خانواده عملکردی و ۲۶ کلاس کلی دسته بندی شدند. پروتئین‌هایی که در یک کلاس قرار می‌گیرند از یک جد پروتئینی مشترک بوجود آمده اند. بزرگترین کلاس پروتئینی ترانسکریپت باریجه متعلق به دسته مکانیسم های انتقال سیگنالی (۱۴/۲۹ درصد) می‌باشد. بعد از آن عملکردهای عمومی (۱۱/۶۵ درصد) و تغییرات بعد از ترجمه (۹/۴۰ درصد). همچنین ۱۸۰۶ (۴/۲۵ درصد) ترانسکریپت و ۹۲۱ (۳/۷۵ درصد) ژن در دسته بیوسنتز متابولیت‌های ثانویه، انتقال و کاتابولیسم قرار می‌گیرند که نشان دهنده آن است که مقدار زیاد و متنوعی متابولیت‌های ثانویه در باریجه تولید می‌شود.

### نتیجه گیری کلی

تحقیق حاضر به عنوان اولین مطالعه روی ترانسکریپتوم باریجه منجر به تولید اطلاعات مولکولی بسیار

### REFERENCES

1. Abbaszadegan, A., Gholami, A., Mirhadi, H., Saliminasab, M., Kazemi, A., & Moein, M. R. (2015). Antimicrobial and cytotoxic activity of *Ferula gummosa* plant essential oil compared to NaOCl and CHX: a preliminary in vitro study. *Restor Dent Endod*, 40, 50-57.
2. Annadurai, R. S., Jayakumar, V., Mugasimangalam, R. C., Katta, M. A. V. S. K., Anand, S., Gopinathan, S., Rao, S. N. (2012). Next generation sequencing and de novo transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. *BMC Genomics*, 13, 663.
3. Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanon, M.-I., Nunes-Nesi, A., Fernie, A. R. (2006). Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiology*, 142, 1380-1396.
4. Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Udvardi, M. K. (2004). Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant Journal*, 39, 487-512.
5. Correa, J., & Silva, G. (2011). Parallel BLAST Analysis and Performance Evaluation. BICoB.
6. Ghannadi, A., & Amree, S. (2002). Volatile oil constituents of *Ferula gummosa* boiss. from Kashan, Iran. *Journal of Essential Oil Research*, 14, 5-7.
7. Gupta, V., Estrada, A. D., Blakley, I., Reid, R., Patel, K., Meyer, M. D., Loraine, A. E. (2015). RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *GigaScience*, 4, 5.
8. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494-1512.
9. Hyun, T. K., Lee, S., Rim, Y., Kumar, R., Han, X., Lee, S. Y., Kim, J. Y. (2014). De-novo RNA sequencing and metabolite profiling to identify genes involved in anthocyanin biosynthesis in Korean black raspberry (*Rubus coreanus* Miquel). *PLoS ONE*, 9, 1-13.
10. Inglis, D. O., Binkley, J., Skrzypek, M. S., Arnaud, M. B., Cerqueira, G. C., Shah, P., Sherlock, G. (2013). Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol*, 13, 91.



11. Jalali, H. T., Ebrahimian, Z. J., Evtuguin, D. V., & Neto, C. P. (2011). Chemical composition of oleo-gum-resin from *Ferula gummosa*. *Industrial Crops and Products*, 33, 549-553.
12. Johnson, M. T. J., Carpenter, E. J., Tian, Z., Bruskiwich, R., Burris, J. N., Carrigan, C. T., . . . Wong, G. K. S. (2012). Evaluating methods for isolating total rna and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE*, 7.
13. Julsing, M. K., Quax, W. J., & Kayser, O. (2008). The Engineering of Medicinal Plants: Prospects and Limitations of Medicinal Plant Biotechnology. *Medicinal Plant Biotechnology: From Basic Research to Industrial Applications*, 1-8.
14. Li, S., Tighe, S. W., Nicolet, C. M., Grove, D., Levy, S., Farmerie, W., Mason, C. E. (2014). Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature Biotechnology*, 32, 915-925.
15. Sarfaraj Hussain, M., Fareed, S., Ansari, S., Akhlaquer Rahman, M., Zareen Ahmad, I., & Saeed, M. (2012). Current approaches toward production of secondary plant metabolites. *J Pharm Bioallied Sci.*, 4, 10-20.
16. Seyed, B., Nabavi, F., Ebrahimzadeh, A., & Nabavi, M. (2010). Antioxidant activity of flower , stem and leaf extracts of *Ferula gummosa* Boiss. *Grasasyaceites*, 61, 244-250.
17. Shahbazi, A., Lotfi, M., Mostafavi, K., Asadian, G., & Heidarian, A. R. (2011). Effect of Persian galbanum (*Ferula gummosa* L.) extract on seed germination and growth of some weeds. *African Journal of Agricultural Research*, 6, 5106-5111.
18. Shi, C.-Y., Yang, H., Wei, C.-L., Yu, O., Zhang, Z.-Z., Jiang, C.-J., Wan, X.-C. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics*, 12, 131.
19. Smith-Unna, R., Bournsnell, C., Patro, R., Hibberd, J., & Kelly, S. (2016). TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Research*, gr. 196469.196115.
20. Spyropoulou, E. a., Haring, M. a., & Schuurink, R. C. (2014). RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics*, 15, 402.
21. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, 6.
22. Tang, Q., Ma, X., Mo, C., Wilson, I. W., Song, C., Zhao, H., Qiu, D. (2011). An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics*, 12, 343.
23. Vaidya, K., Ghosh, A., Kumar, V., Chaudhary, S., Srivastava, N., Katudia, K., Chikara, S. K. (2013). De Novo Transcriptome Sequencing in *L.* to Identify Genes Involved in the Biosynthesis of Diosgenin. *The Plant Genome*, 6(2).
24. Yang, S., Tschaplinski, T. J., Engle, N. L., Carroll, S. L., Martin, S. L., Davison, B. H., Brown, S. D. (2009). Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genomics*, 10, 34.
25. Zulak, K. G., Cornish, A., Daskalchuk, T. E., Deyholos, M. K., Goodenowe, D. B., Gordon, P. M. K., Facchini, P. J. (2007). Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. *Planta*, 225, 1085-1106.