

آنالیز داده‌های ترانسکریپتوم زعفران در راستای بیان ژن‌های دخیل در مسیر

بیوسنتزی فنیل پروپانویید

محبوبه دهقان^۱، مسعود توحیدفر^{۲*}

چکیده

زعفران امروزه به عنوان یک گیاه استراتژیک در کشاورزی محسوب می‌شود. این گیاه به عنوان گران‌ترین ادویه جهان شناخته می‌شود و علاوه بر خواص مختلف خوراکی از دیرباز در طب سنتی مورد استفاده قرار گرفته است. یکی از مهم‌ترین خواص دارویی زعفران برای جلوگیری و درمان سرطان است. در این مطالعه ابتدا داده‌های ترانسکریپتوم از سایت NCBI Short read archive دانلود شدند. داده‌ها شامل پنج بافت کلالة، گلبرگ، برگ، پیاز و پرچم هستند. کنترل کیفی داده‌ها توسط نرم‌افزار FastQC و Trimmomatic انجام شده و بعد از آن اسمبلی خوانش‌ها صورت گرفت. پیش‌بینی ژن‌ها توسط نرم‌افزار Trinity صورت گرفت. بیان ژن‌ها با استفاده از نرم‌افزار RSEM سنجیده شده و با استفاده از داده‌ای موجود در مقالات ژن-های ضد سرطان در مسیر بیوسنتز فنیل پروپانویید با استفاده از مشابهت توالی‌های میان آن‌ها و ژن‌های ساخته شده توسط Trinity با نرم‌افزار BLASTx مورد بررسی قرار گرفتند. از میان ژن‌های موجود در مسیر بیوسنتز فنیل پروپانویید در این مطالعه، برای اولین بار ژن CAD1 که کد کننده آنزیم Cinnamaldhyde است، شناسایی شد، این آنزیم در مسیرهای مختلف باعث جلوگیری از سرطان یا فعالیت تومور زایی می‌شود. در بررسی بیان این ژن مشخص شد که این ژن بیشترین بیان را در بافت‌های گلبرگ و کلالة دارد و کلالة این گیاه علاوه بر خواص ادویه آن می‌تواند به عنوان یک داروی ضد سرطان هم استفاده شود.

کلمات کلیدی: زعفران، ترانسکریپتوم، ژن، مسیر بیوسنتزی

^۱ دانشجوی دکتری مهندسی بیوتکنولوژی، گروه بیوتکنولوژی گیاهی، دانشکده علوم و فناوری‌های زیستی، دانشگاه شهید بهشتی.

^۲ * نویسنده مسئول، دانشیار بیوتکنولوژی کشاورزی، گروه بیوتکنولوژی گیاهی، دانشکده علوم و فناوری‌های زیستی، دانشگاه شهید بهشتی. ایمیل:

مقدمه

ژنی برای این گیاه در پایگاه داده مذکور ثبت نشده است (Jain et al., 2016).

جدیدترین و کارآمدترین ابزار برای بررسی ترانسکریپتوم بهره‌گیری از روش‌های توالی‌یابی با توان بالا استفاده از روش RNA-SEQ است. روش RNA-SEQ قادر به شناسایی ایزو فرم‌های مختلف ژن، رخدادهای جابه‌جایی واریانت‌های ژن (شامل انواع موتاسیون‌ها) و تغییرات بعد از رونویسی است. یکی از مهم‌ترین اهداف آزمایش‌ها RNA-SEQ تشخیص تغییرات بیان ژن‌ها در دو یا چند شرایط یا بافت متفاوت است. یکی از مزیت‌های RNA-SEQ نسبت به روش‌های مرسوم پیشین، امکان کار بر روی ترانسکریپتوم موجوداتی است که ژنوم مرجع آن‌ها موجود نیست (Marguerat and Bähler, 2010).

هدف آنالیز ترانسکریپتوم یا RNA-SEQ پیدا کردن اجزا ژنوم نظیر ژن یا رونوشت‌ها و تفسیر عملکرد آن‌ها است. این روش به جهت اینکه برای شروع و آنالیز کار نیاز به اطلاعات ژنوم مرجع و اطلاعات نقشه ژنی ندارد برای بسیاری از موجودات غیر مدل استفاده می‌شود (Liu et al., 2013). این روش بر پایه توالی‌یابی است و برعکس روش‌های سنتی مولکولی نظیر RT-PCR و ریزآرایه و همین‌طور روش‌های پرهزینه توالی‌یابی سنتی مانند روش توالی‌یابی سنگر، کار آزمایشگاهی و هزینه بسیار پایین‌تر است. در این روش برخلاف روش توالی‌یابی سنگر که طول توالی نزدیک به ۱۰۰۰ جفت باز را پوشش می‌دهد، میلیون‌ها قطعه کوچک بین ۷۵ تا ۳۰۰ باز را توالی‌یابی می‌کند و بخش زمان‌بر آن قسمت آنالیز بیوانفورماتیک داده است. در سالهای اخیر به جهت مزایای متعددی که روش توالی‌یابی RNA-SEQ نسبت به روش‌های مرسوم قبلی داشته است، بیشتر مورد استفاده قرار گرفته است. از جمله مزایای این

زعفران به نام علمی *Crocus Sativus* شناخته می‌شود. گیاهی تریپلوئید و متعلق به خانواده تک‌لپه‌ها است. گیاهی چندساله است که به‌طور گسترده در ایران، هند، یونان، اسپانیا و ایتالیا کشت می‌شود. زعفران به عنوان گران‌ترین ادویه جهان شناخته می‌شود. کلاله زعفران با رنگ قرمز و یا نارنجی به عنوان عامل طعم‌دهنده، رنگ‌آمیزی مواد غذایی استفاده می‌شود. به‌علاوه زعفران به عنوان یکی از داروهای مؤثر در طب گیاهی و سنتی شناخته می‌شود. اثرات دارویی زعفران عمدتاً به کروکسین، کروکستین، پیکروکروسین و سافرانال مربوط می‌شود. در مطالعات مختلفی نشان داده شده که به‌واسطه متابولیت‌هایی ذکر شده مسئول رنگ، مزه و بو در گیاه می‌باشند (Khorasany and Hosseinzadeh, 2016; Jain et al., 2016).

کروکسین اثرات درمانی بر روی قسمت‌های مختلف بدن مانند سیستم قلبی عروقی، ایمنی، تنفسی، دستگاه تناسلی و سیستم عصبی مرکز دارد. زعفران نقش کلیدی در درمان اختلالات سیستم گوارش از طریق داروهای شیمیایی، مهار تکثیر سلولی، القا مرگ سلولی، اثرات آنتی‌اکسیدانی و پراکندگی رادیکال را دارند. نتیجه مکانیسم‌های ذکر شده ویژگی‌های بالقوه درمانی زعفران را در برابر سرطان کبد، سمیت کبدی، کبد چرب، سرطان معده، سرطان روده بزرگ، سرطان پانکراس، دیابت و انقباضات یبوست را نشان می‌دهد (Khorasany and Hosseinzadeh, 2016). باوجود اهمیت بالای اقتصادی و دارویی این گیاه تاکنون مطالعات اندکی با روش‌های ارزیابی مولکولی بر روی زعفران انجام شده است. در پایگاه داده NCBI برای این گیاه اطلاعات ژنومی به ثبت نرسیده است. تعداد Expressed sequence tag ثبت شده از زعفران بر روی پایگاه داده NCBI ۶۷۶۸ عدد است و تاکنون

نقشه ژنی و اطلاعات ژنوم می‌بایست نقشه ژنی بر اساس مونتاژ خوانش‌ها و رونوشت‌ها ساخته شود (Grabherr et al., 2011).

مطالعه انجام شده بر روی ترانسکریپتوم زعفران توسط بابا و همکاران در سال ۲۰۱۵ نشان داد که در مسیر کارتنوئید و آپوکارتنوئید زعفران ژن‌های زیادی دخیل هستند (Baba et al., 2015). جین و همکاران در سال ۲۰۱۶ با بررسی اندام مختلف گیاه زعفران نظیر کلاله، برگ، گل برگ، پیاز و پرچم به مطالعه پروفایل بیانی این گیاه پرداختند، در مطالعه مذکور بیان ژن‌های مسیر بیوسنتز آپوکارتنوئید و همین‌طور مراحل مختلف رشد کلاله مورد بررسی قرار گرفته است (Jain et al., 2016).

در مطالعات اندک انجام شده بر روی ترانسکریپتوم زعفران بیشتر به بررسی بیان ژن‌ها در مسیر آپوکارتنوئید پرداخته شده است. در مطالعه هدف بررسی بیان ژن‌های درگیر ضد سرطان در مسیر فنیل پروپانوئید مورد بررسی قرار گرفته است.

با استفاده از نرم‌افزار FastQC (Andrews, 2010) کیفیت خوانش‌ها مورد بررسی قرار گرفت و مشخص شد که بازهای ۱ تا ۱۵ دارای نوسان در توزیع بازهای C, T, A و G است. برای حذف بازهای فوق از نرم‌افزار Trimmomatic (Bolger et al., 2014) استفاده شد. علاوه بر آن بازهای باکیفیت کمتر از Q20 حذف شد، Q20 کیفیت خوانش یک باز توسط دستگاه را نشان می‌دهد و نشان‌دهنده احتمال خطای یک درصد در خوانش باز مورد نظر است.

اسمبلی ترانسکریپتوم

به منظور اسمبلی رونوشت‌ها از نرم‌افزار Trinity V 2.4.0 استفاده شد. این نرم‌افزار با مونتاژ خوانش‌ها بر اساس مشابهت توالی هر خوانش با خوانش دیگر ابتدا توالی رونوشت‌ها را و سپس توالی ژن‌ها پیش‌بینی می‌شود. برای اسمبل کردن خوانش‌ها از پروتکل

روش می‌توان به بررسی بیان ژن در موجوداتی که ژنوم مرجع موجود نیست، پیدا کردن اطلاعات ژن‌ها و ایزو فرم‌های جدید و یافتن جهش‌های رخ داده بر روی توالی رونوشت است (Wang et al., 2009).

در روش توالی یابی ترانسکریپتوم قطعات توالی تولید شده توسط دستگاه یا همان خوانش کوتاه (short read) در ابتدا با ژنوم هم‌ردیف می‌شوند و بر اساس نقشه ژنی تعداد خوانش‌های متصل شده به آن ژن شمرده می‌شود. در گیاهانی که ژنوم مرجع ندارد از روش اسمبلی de novo استفاده می‌شود، به این شرح که ابتدا خوانش‌های کوتاه بر اساس شباهت توالی‌ها با همدیگر مونتاژ شده و بر اساس آن رونوشت‌ها ساخته می‌شوند. از مجموع این رونوشت‌ها، ترانسکریپتوم مرجع ساخته می‌شود. از این مرحله به بعد ترانسکریپتوم مرجع مانند ژنوم مرجع است و خوانش‌های کوتاه علیه آن هم‌ردیف می‌شود و تعداد خوانش‌هایی که به هر رونوشت متصل شده شمرده می‌شود. فرق این روش با زمانی که ژنوم مرجع در دسترس است در این است که به جای بهره‌گیری از

مواد و روش‌ها

ابتدا با استفاده از کلمات کلیدی نظیر Crocus sativus و RNA-SEQ در پایگاه داده NCBI Short read archive (SRA) (et al., 2010) جستجو انجام و تمامی فایل‌ها به شناسه پروژه PRJNA273002 دریافت شد. پس از آن داده‌های خام (فایل‌های SRA) بر روی یک سرور با مشخصات سیستم عامل لینوکس ابونتو ۱۶.۰۴، ۸ هسته و ۶۴ گیگا بایت RAM ذخیره شدند. پروژه ذکر شده دارای ۵ فایل متناظر با بافت‌های کلاله، گلبرگ، پیاز، پرچم و برگ است. پس از دانلود داده‌ها با استفاده از نرم‌افزار fastq-dump موجود در مجموعه ابزار SRA toolkit، فایل‌های SRA به fastq تبدیل شد.

ژن‌ها و پروتئین‌های شناخته‌شده در گیاهان دیگر استفاده شد. به‌منظور این کار با استفاده از نرم‌افزار Blastx و Blastp (Madden, 2013) و HMMScan (Finn et al., 2011) به بررسی مشابهت توالی‌های اسمبل شده با توالی‌های پایگاه داده Pfam و Swiss-PROT اقدام شد. در این بررسی مشابهت Subject اطلاعات توالی‌های Pfam یا Swiss-PROT می‌باشد و Query همان ترانسکریپتوم اسمبل شده مرجع است. برای ترجمه ترانسکریپتوم مرجع به‌توالی اسیدآمینه از نرم‌افزار Transdecoder استفاده شده است. به‌منظور بررسی Gene ontology از سایت WEGO (Ye et al., 2018) (<http://wego.genomics.org.cn>) استفاده شده است.

پیدا کردن ژن‌های دخیل در مسیر فنیل پروپانویید
با استفاده از اطلاعات موجود در مقاله‌ای (Deng et al., 2016)، ژن‌های کاندید در مسیر بیوسنتز فنیل پروپانویید انتخاب شدند. توالی پروتئین آن‌ها در گیاه آرآبیدوپسیس از سایت UniProt (Apweiler et al., 2004) دانلود شد و Blastx ترانسکریپتوم مرجع با توالی‌های پروتئین انجام شد. خروجی این مرحله Blastx که دارای توالی‌هایی با حداقل طول ۱۰۰ اسیدآمینه و حداقل مشابهت ۶۰ درصد بودند به عنوان کاندید انتخاب شد. برای بررسی بیان آن‌ها خروجی Blastx این مرحله با ماتریس بیان ژن توسط RSEM ادغام شد.

برای ترانسکریپتوم اسمبل شده مرجع ۴۲/۷۵ درصد بود (جدول ۲). N50 نشان‌دهنده حداقل طول توالی ژن / رونوشت در حداقل ۵۰ درصد از ژن‌ها یا رونوشت‌ها بود (جدول ۲). میزان N50 در ژن‌ها ۱۰۶۱ باز و در رونوشت‌ها ۱۲۰۶ باز بود. میانگین طول ژن‌های اسمبل شده ۸۶۸ باز و میانگین طول رونوشت‌ها ۹۷۱ باز بود (جدول ۲). نتایج نشان داد

Trinity استفاده شد (Haas et al., 2013)، در این پروتکل پس از ادغام خوانش‌ها از هر نمونه بافت گیاه در دو فایل مختلف چپ و راست، اقدام به اسمبل کردن آن‌ها با استفاده از نرم‌افزار Trinity شد. در این قسمت کانتیگ‌هایی که طول زیر ۳۰۰ باز داشتند فیلتر شدند.

کمی‌سازی بیان ژن‌ها و پیدا کردن بیان ژن‌های افتراقی

در ادامه با استفاده از نرم‌افزار RSEM (Li et al., 2011) پروفایل بیانی هر بافت بر اساس هم‌ردیف کردن نمونه با ترانسکریپتوم مرجع اسمبل شده محاسبه شد. نرم‌افزار RSEM یک ماتریس می‌سازد که شماره هر سطر اسم ژن و به تعداد بافت‌ها ستون دارد. تعداد خوانش‌هایی که به هر ژن متصل می‌شود را شمرده و برای هر بافت در ستون متناظر آن قرار می‌دهد. از ماتریس ساخته‌شده توسط RSEM به عنوان ورودی نرم‌افزار R (Robinson et al., 2013) برای کشف ژن‌های بابیان افتراقی استفاده شد. ژن‌هایی با Pvalue کمتر از ۰/۰۵ و logFC بیشتر از ۱ یا کمتر منفی ۱ به عنوان معنی‌دار در نظر گرفته شد.

تفسیر عملکردی

با توجه به اینکه ژن‌هایی که توسط Trinity پیش‌بینی شده‌اند فقط توالی هستند، در این مرحله به‌منظور پیدا کردن عملکرد ژن یا رونوشت پیش‌بینی شده، از مشابهت توالی‌های پیش‌بینی شده با دیگر

نتایج و بحث

طول خوانش‌های خام دارای ۱۰۱ باز است که پس از کنترل کیفی ۱۵ باز آن حذف شده است، تعداد کل خوانش‌های توالی یابی شده هر بافت بین ۱۷ میلیون تا ۲۴ میلیون خوانش است (جدول ۱). پس از مرحله اسمبلی با نرم‌افزار Trinity تعداد ۸۶۸۴۲ ژن و ۲۴۰۹۴۶ رونوشت به‌دست‌آمده است. میزان GC

که میزان مشابهت ترانسکریپتوم مرجع با Blastp ۸۴/۹۸ درصد است. میزان دومین‌های یافت پروتئین‌های یافت شده Swiss-PROT توسط Blastx برابر با ۸۳/۹۵ درصد و با نرم‌افزار HMMScan ۴۶/۷۱ درصد بوده است (جدول ۲).

جدول ۱- اطلاعات خوانش‌های مربوط به هر بافت

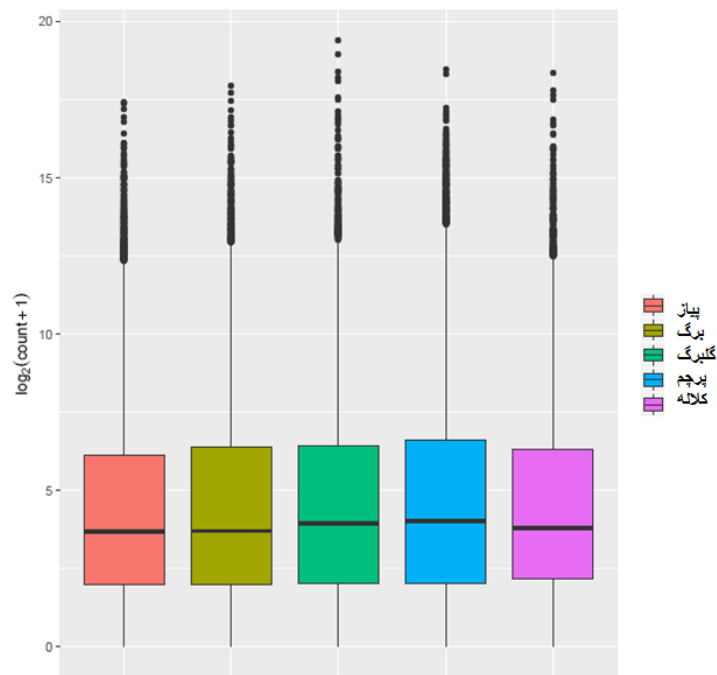
بافت	طول خوانش قبل از کنترل کیفی	طول خوانش پس از کنترل کیفی	تعداد خوانش قبل کنترل کیفی	تعداد خوانش‌ها پس از کنترل کیفی
پیاز	۱۰۱	۸۶	۱۷۹۶۶۱۰۸	۱۷۹۵۵۱۱۵
برگ	۱۰۱	۸۶	۱۸۸۲۵۱۷۲	۱۸۸۱۶۰۷۹
گلبرگ	۱۰۱	۸۶	۲۴۱۵۶۴۹۱	۲۴۱۴۴۳۸۷
پرچم	۱۰۱	۸۶	۱۹۲۶۷۷۹۴	۱۹۲۵۸۸۶۶
کلاله	۱۰۱	۸۶	۲۴۱۵۶۴۹۱	۲۲۹۲۲۳۲۴

جدول ۲- خلاصه اطلاعات اسمبلی و اطلاعات عملکردی با پایگاه داده‌ها

پارامتر	تعداد	درصد
تعداد ژن‌ها	۸۶۸۴۲	-
تعداد رونوشت‌ها	۲۴۰۹۴۶	-
میزان GC	۴۲.۷۵	-
میزان N50 در ژن	۱۰۶۱	-
میانگین طول ژن	۸۶۷	-
تعداد بازهای استفاده‌شده در اسمبلی ژن‌ها	۷۵۳۴۸۳۹۳	-
میزان N50 در رونوشت	۱۲۰۶	-
میانگین طول رونوشت	۹۷۱	-
تعداد بازهای استفاده‌شده در اسمبلی رونوشت	۲۳۳۹۷۲۹۶۴	-
تعداد پروتئین‌های مشابه یافت شده با استفاده از Blastx	۲۰۲۲۹۳	۸۳.۹۵
تعداد پروتئین‌های مشابه یافت شده با استفاده از Blastp	۲۰۴۷۶۵	۸۴.۹۸
تعداد دومین‌های یافت شده با استفاده از HMMScan	۱۱۲۵۵۰	۴۶.۷۱

نمودار جعبه‌ای دارای ۴ چارک است چارک اول که بیان آن‌ها بین ۰ تا ۲/۲۵ بوده‌اند. چارک دوم، سوم و چهارم بیان آن‌ها به ترتیب بین ۲/۲۵ تا ۳/۵، ۳/۵ تا ۵/۵ و ۵/۵ تا ۱۲/۲ است.

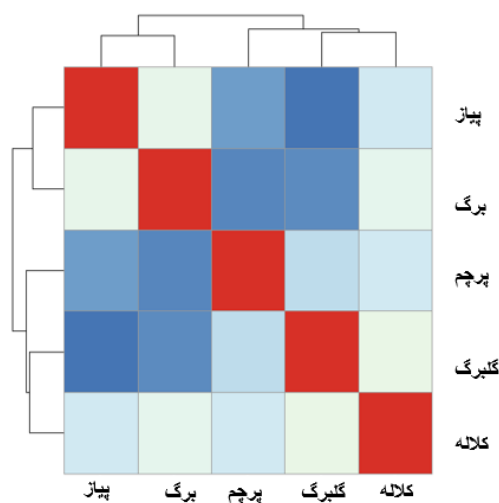
ماتریس بیان ژن‌های ایجادشده توسط RSEM مشخص کرد که ۵ بافت موردبررسی دارای پروفایل بیانی یکنواختی است. این نتیجه با نمودار جعبه‌ای ساخته شده از \log_2 ماتریس مذکور ساخته شد.



شکل ۱- نمودار جعبه‌ای پروفایل بیانی هر بافت

ترتیب، پیاز و برگ پروفایل بیانی مشابهی دارند. کلاله و گلبرگ و پرچم دارای مشابهت بیشتر است. از طرفی به‌طور معنی‌داری پروفایل بیانی کلاله و گلبرگ بیشتر به یکدیگر نزدیک می‌باشند. درنهایت در این نمودار ۵ بافت دو گروه تقسیم‌شده است که پیاز و برگ با یکدیگر در یک گروه هستند و پرچم، گلبرگ و کلاله نیز در گروه دوم قرار گرفته‌اند.

به‌منظور یافتن میزان نزدیکی و دوری پروفایل بیانی هر بافت از نمودار گرافیکی *Correlation heatmap* استفاده شد (شکل ۲). این نمودار برای نشان دادن میزان مشابهت پروفایل بیانی بافت‌های مختلف به‌صورت سلسله‌مراتبی است که بر اساس اطلاعات تعداد خوانش‌های متصل شده به هر ژن رسم شده است. این نمودار بافت‌هایی با پروفایل بیانی مشابه را دوبه‌دو باهم در یک گروه قرار می‌دهد. طبق شکل به



شکل ۲- نمودار *Heatmap* به منظور بررسی همبستگی ترانسکریپتوم بین بافت‌های مورد مطالعه

(down) در هر مقایسه دوبه‌دو نشان داده شده است. آستانه Fold-change بین ۲ و منفی ۲ در نظر گرفته شده است. در مقایسات مشخص شد بیشترین تفاوت بیان ژن‌ها بین بافت پرچم و کلالة با ۹۸۶ ژن که ۸۸ ژن بیان بالاتری در بافت کلالة نسبت به پرچم و ۸۹۹ ژن بیان پایین‌تری در بافت کلالة نسبت به پرچم داشته است. کمترین تفاوت بیان بین بافت های گلبرگ و کلالة با ۲۶۴ ژن است. که ۴۲ ژن در کلالة بیان بالاتری نسبت به بافت گلبرگ و ۲۲۳ ژن در بافت کلالة بیان پایین‌تری نسبت به بافت گلبرگ داشته است. علاوه بر این، بر اساس مقایسات مشخص شد که بافت پرچم بیشترین ژن‌های بابیان بالا را دارد. در این بررسی ژن‌های بابیان بالا در بافت پرچم نسبت به کلالة ۸۹۹ ژن، پرچم نسبت به پیاز ۸۶۴ ژن و پرچم نسبت به برگ ۴۵۷ ژن بیان بالاتری داشته است. به همین خاطر نشان داده می‌شود که بافت پرچم تفاوت بیانی بیشتری نسبت به دیگر بافت های زعفران دارد.

به‌منظور بررسی پروفایل بیان بافت های مختلف می‌بایست خوانش های باکیفیت را با ترانسکریپتوم مرجع هم‌تراز کرد. بین ۹۱ تا ۹۵ درصد از خوانش های پنج بافت مورد بررسی، به ترانسکریپتوم مرجع هم‌تراز شدند. برای این کار از پروتکل نرم‌افزار RSEM استفاده شد. برای هم‌تراز کردن خوانش‌ها، RSEM از نرم‌افزار (Bowtie Langmead et al., 2009) استفاده می‌کند و پس از به دست آمدن فایل خروجی هم‌تراز شده، RSEM شروع به شمارش تعداد خوانش های متصل شده به هر ژن می‌کند. بعد از این مرحله تعداد خوانش های مربوط به هر ژن در هر بافت در یک ماتریس ادغام می‌شود. ماتریس خروجی نرم‌افزار RSEM با استفاده از R و edgeR برای پیدا کردن ژن‌های بابیان افتراقی استفاده شده است.

نتایج بیان افتراقی در جدول ۳ آورده شده است. این جدول شامل اطلاعات تمام ژن‌های زعفران است که به‌صورت متفاوت بین دو بافت بیان شده و به‌علاوه تعداد ژن‌های دارای بیان بالا (up) و بیان پایین

جدول ۳ - بررسی بیان افتراقی بین بافت های مختلف

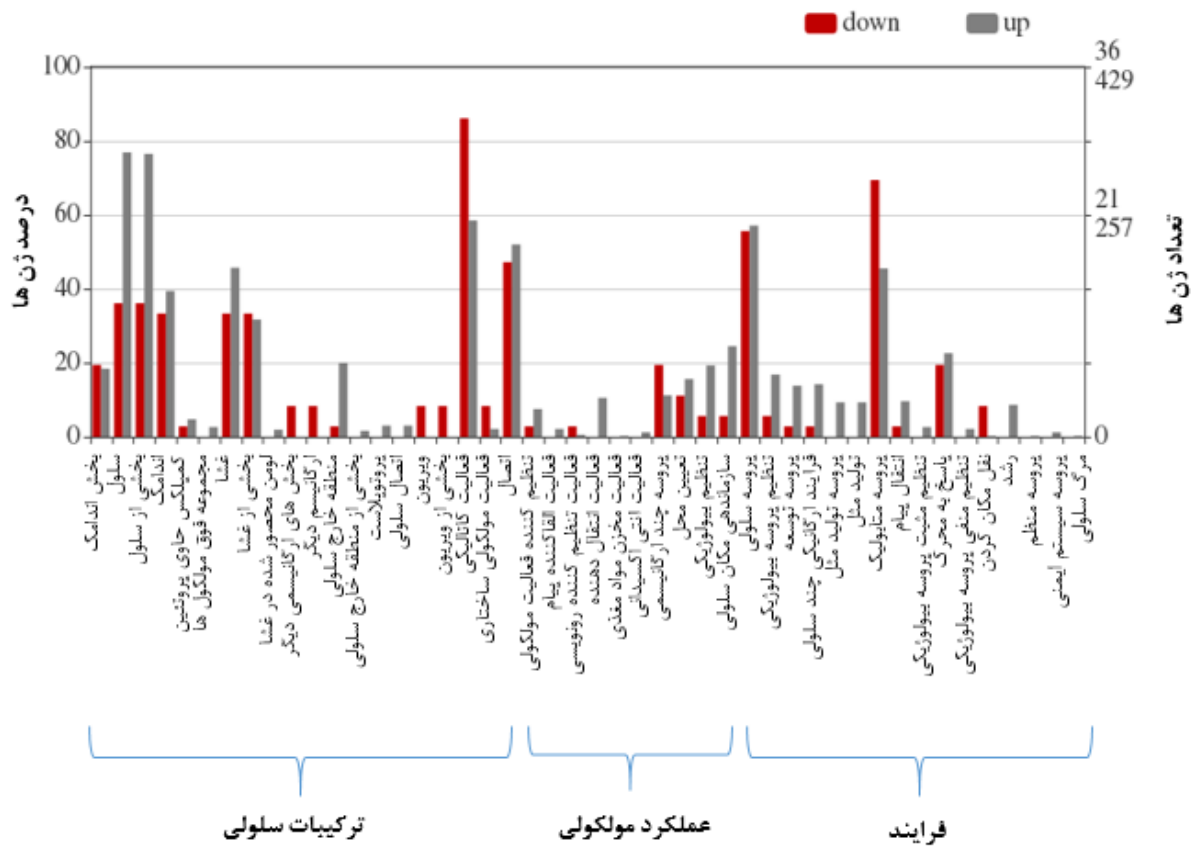
بافت ۱	بافت ۲	تعداد ژن‌های بابیان افتراقی	تعداد ژن بیان بالا در بافت ۲	تعداد ژن بیان پایین در بافت ۲
پیاز	برگ	۴۵۴	۴۳۰	۲۵
پیاز	گلبرگ	۷۰۲	۵۷۷	۱۲۶
پیاز	پرچم	۹۳۱	۸۶۴	۶۸
پیاز	کلالة	۲۸۷	۲۳۷	۵۱
برگ	گلبرگ	۵۴۶	۲۷۱	۲۷۵
برگ	پرچم	۷۴۵	۴۵۷	۲۸۹
برگ	کلالة	۲۹۲	۷۵	۲۱۸
گلبرگ	پرچم	۳۶۴	۲۰۱	۱۶۳
گلبرگ	کلالة	۲۶۴	۴۲	۲۲۳
پرچم	کلالة	۹۸۶	۸۸	۸۹۹

شد بیشتر ژن‌های بابیان بالا در کلالة در قسمت‌های غشا و اجزای آن، منطقه خارج سلولی و اجزا سلول

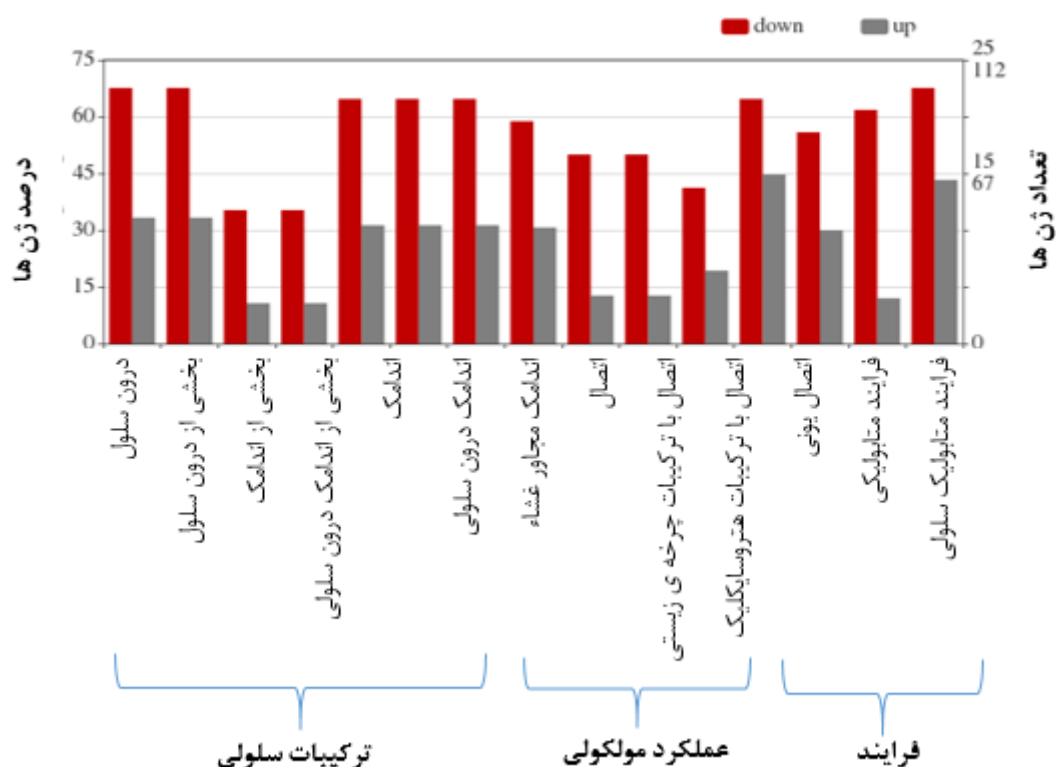
GO بین بافت پرچم و کلالة از طریق سایت WGEO آنالیز شد. در آنالیز اجزا سلولی مشخص

اجزا سلولی ژن‌ها بایان کم در کلاله در قسمت‌هایی نظیر اجزا درون سلولی و اندامک‌های سلولی بوده است. در بررسی عملکرد مولکولی ژن‌های بایان کم در کلاله بیشتر در اتصال ترکیب‌های چرخه زیستی و اتصال یون‌ها بوده است. در آنالیز مسیر زیستی نیز ژن‌های بایان کم در کلاله بیشتر در پروسه‌های سلولی، پروسه‌های متابولیک و پروسه‌های سنتز زیستی نقش داشته‌اند (شکل ۴).

قرارگرفته‌اند. در آنالیز عملکرد مولکولی مشخص شد که ژن‌های بایان پایین در کلاله در فعالیت کاتالیکی نقش داشته است. در آنالیز مسیر زیستی مشخص شد ژن‌های بایان بالا در کلاله بیشتر در پروسه‌های سلولی نظیر پروسه تکاملی سلول نقش داشته‌اند و پروسه‌های متابولیک شامل ژن‌هایی است که در کلاله بیان پایین داشته است (شکل ۳). آنالیز GO بین بافت گلبرگ نسبت به کلاله نیز انجام شد. در آنالیز



شکل ۳- بررسی Gene ontology بافت پرچم نسبت به کلاله (up و down در بافت کلاله است).

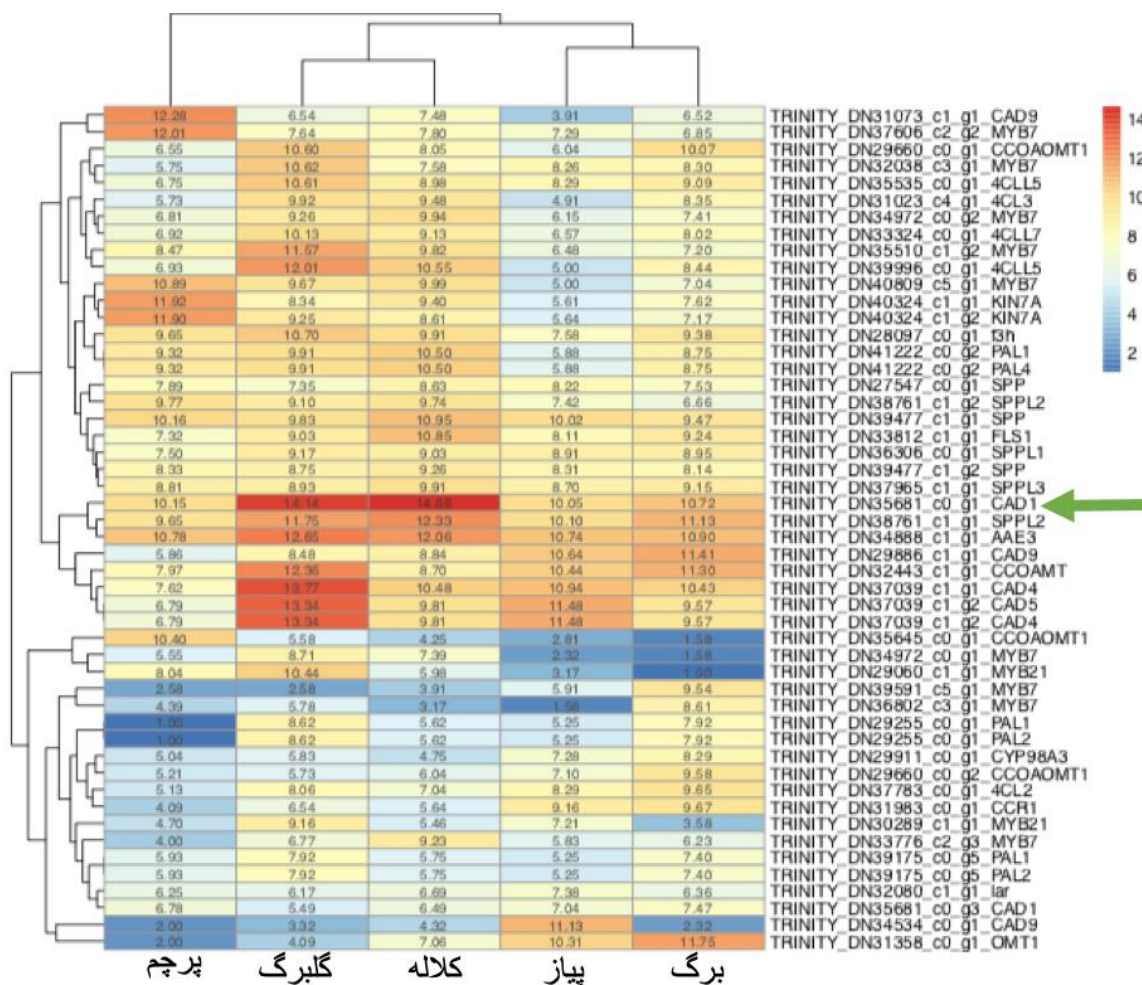


شکل ۴- بررسی Gene ontology بافت گلبرگ نسبت به کلاله (up و down در بافت کلاله است).

اساس \log_2 از تعداد خوانش‌ها است. رنگ آبی و رنگ قرمز به ترتیب مربوط به بیان کم و بیان بالا است. ژن‌هایی که پروفایل بیانی مشابه دارند در یک کلاستر واقع شده‌اند. در این شکل ژن‌ها به دو گروه بابیان پایین و بالا تقسیم شده‌اند، ژن‌هایی که بیان بالایی دارند همانند ژن CAD1 و خانواده CAD بارنگ متمایل به قرمز در یک کلاستر قرار گرفته است. به عنوان نمونه از ژن‌های بابیان پایین، ژن OMT1 بارنگ متمایل به آبی در پایین نمودار مشخص شده است. در بالای نمودار نحوه کلاستر شدن بافت نشان داده شده است. بر همین اساس برگ و پیاز در یک کلاستر، کلاله و گلبرگ در گروه دیگر و پرچم نیز به‌تنهایی در کلاستر دیگر قرار گرفته است.

برای پیدا کردن ژن‌های مؤثر ضد سرطان در ترانسکریپتوم مرجع زعفران از مقاله دنگ و همکاران سال ۲۰۱۶ برای انتخاب ژن‌ها استفاده شد. مقاله یادشده شامل ژن‌های متعدد در مسیر فنیل پروپانویید و فلاونویید بوده است. پس از جستجو در سایت UniProt تعداد ۶۰ توالی پروتئین مشابه به ژن‌های مسیر فنیل پروپانویید و فلاونویید از گیاه آرابیدوپسیس دانلود شد و با ترانسکریپتوم مرجع Blastx شدند. شکل ۵ نشان‌دهنده خروجی Blastx ادغام‌شده با ماتریس بیانی ساخته شده توسط RSEM است.

در این شکل میزان بیان ژن‌های کاندید ضد سرطان در بافت‌های مختلف آورده شده است. بیان ژن‌ها علاوه بر عدد بارنگ نیز مشخص شده است که بر



شکل ۵- شکل Heatmap ژن‌های مؤثر ضد سرطان در مسیر فنیل پروپانویید و فلاونویید و بیان آن‌ها در بافت‌های مختلف زعفران

در مطالعه حاضر، ژن CAD1 برای اولین بار از طریق آنالیز ترانسکریپتوم زعفران به دست آمده است. ژن CAD1 کد کننده سینامالدهید (Cinnamaldehyde) با ترکیب شیمیایی C₉H₈O است. این ژن در گیاه آراییدوپسیس به صورت فراوان در مسیرهای توسعه رشدی گیاه به صورت فعالانه عمل می‌کند (Kim et al., 2004). خانواده این ژن دارای چندین نسخه از این ژن بر روی ژنوم است، و به سه کلاس اصلی تقسیم‌بندی می‌شود. ژن‌های کلاس یک و دو ژن CAD بیشتر در توسعه Lignin در گیاه نقش دارد و باکم شدن بیان ژن CAD میزان Lignin به شکل محسوسی کاهش پیدا می‌کند. دیگر ژن‌های کلاس دو و سه در عملکرد بافت‌های گیاه تحت تنش‌های زیستی نقش دارد (Barakat

et al., 2009; Chabannes et al., 2001). عصاره سینامالدهید از پوست گیاه دارچین استخراج می‌شود و نقش مهمی در درمان سرطان دارد. اثر سمی این ماده بر روی انواع سرطان‌ها نظیر سرطان خون، روده، کبد، رحم و ریه اثبات شده است. سینامالدهید در سطح مولکولی با تغییراتی نظیر جلوگیری از تکثیر القا مرگ سلولی، عدم رگ زایی، جلوگیری از متاستاز و سرکوب التهاب توموری باعث جلوگیری از بروز سرطان می‌شود (Carvalho et al., 2015; Larasati and Meiyanto, 2018; Dutta and Chakraborty, 2018). در مطالعه روی سرطان روده بزرگ از این ماده به عنوان یک داروی جدید و عامل القاکننده مرگ سلولی، چسبندگی سلول و

بیشترین بازده را دارند انتخاب شده و پس از مخلوط کردن تعداد مشخص از بافت کلالة برای بررسی ترانسکریپتوم آن و خصوصاً ژن CAD1 اقدام نمود.

نتیجه گیری کلی

از میان ژن های انتخاب شده در مسیریوسنتز فلاونوئید و فنیل پروپانوئید تعدادی در زعفران بیان شده اند. از میان آنها یکی از ژن هایی که بیان بالایی در بین رونوشت های بیان شده داشته ژن CAD1 است. این ژن که کد کننده پروتئین cinnamyl alcohol dehydrogenase است به ترتیب در گلبرگ و کلالة بیشترین بیان را دارا بوده است. این ژن در مسیر بیوسنتز فنیل پروپانوئید قرار دارد و برای اولین بار است که با استفاده از آنالیز داده های ترانسکریپتوم در زعفران شناخته شده است. ژن CAD1 که کد کننده پروتئین Cinnamaldehyde است، به عنوان یک دارو در طب سنتی در درمان سرطان استفاده می شود (۲۳).

مهارکننده مسیر سیگنال دهی PI3K/Akt نام برده شده است (Li et al., 2016).

اگرچه پیش از این مطالعاتی در زمینه خواص ضد سرطانی زعفران گزارش شده، ولی به طور خاص بیان ژن های ضد سرطان در آنها انجام نشده است. این مطالعه سعی نموده تا علاوه بر بررسی ژن های که در مسیر آپوکارتونوئید هستند، ژن های جدیدی را نیز در این مسیر ضد سرطان کشف کند. یکی از نکات کلیدی در این مطالعه بیان بالای ژن CAD1 در کلالة زعفران است که بیشترین مصرف خوراکی را در بین بافت های دارد. با توجه به بومی بودن این گیاه می توان به عنوان منبعی غنی از ماده سینامالدهید مورد استفاده قرار گرفته شود. با گسترش روزافزون استفاده از طب سنتی و گیاهان دارویی پیشنهاد می شود که به صورت جداگانه مطالعه کامل تری بر روی خواص متابولیکی بافت های مختلف زعفران و همین طور آزمایش عصاره این بافت های بر روی رده های سلولی یا حیوانات آزمایشگاهی انجام شود. برای مطالعه بیشتر پیشنهاد می شود ارقامی که

- generation and analysis. *Nature protocols*, 8, 1494.
12. Li, B. and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 2, p 323.
 13. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
 14. Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI Handbook* [Internet]. 2nd edition. National Center for Biotechnology Information (US)
 15. Finn, R.D., Clements, J. and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, 39, 29-37.
 16. Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A. and Zhang, X. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic acids research*, 46, 71-75.
 17. Deng, N., Chang, E., Li, M., Ji, J., Yao, X., Bartish, I.V., Liu, J., Ma, J., Chen, L., Jiang, Z. and Shi, S. (2016). Transcriptome characterization of *Gnetum parvifolium* reveals candidate genes involved in important secondary metabolic pathways of flavonoids and stilbenoids. *Frontiers in plant science*, 7, p 174.
 18. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. and Martin, M.J. (2004). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 32, 115-119.
 19. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10, p 25.
 20. Kim, S.J., Kim, M.R., Bedgar, D.L., Moinuddin, S.G., Cardenas, C.L., Davin, L.B., Kang, C. and Lewis, N.G. (2004). Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 101, 1455-1460.
 21. Barakat, A., Bagniewska-Zadworna, A., Choi, A., Plakkat, U., DiLoreto, D.S., Yellanki, P. and Carlson, J.E. (2009). The cinnamyl alcohol dehydrogenase gene family
 1. Khorasany, A.R., Hosseinzadeh, H. (2016). Therapeutic effects of saffron (*Crocus sativus* L.) in digestive disorders: a review. *Iranian journal of basic medical sciences*, 19, 455-462.
 2. Jain, M., Srivastava, P.L., Verma, M., Ghangal, R. and Garg, R. (2016). De novo transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Scientific reports*, 6, 22456.
 3. Marguerat, S. and Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and molecular life sciences*, 7, 569-679.
 4. Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y. and Tang, S. (2013). De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of *CesA* gene in ramie (*Boehmeria nivea* L. Gaud). *BMC genomics*, 14, 125-137.
 5. Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10, 57-68.
 6. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A. and Amit, I. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29, 644-658
 7. Baba, S.A., Mohiuddin, T., Basu, S., Swarnkar, M.K., Malik, A.H. and Wani, Z.A. (2015). Comprehensive transcriptome analysis of *Crocus sativus* for discovery and expression of genes involved in apocarotenoid biosynthesis. *BMC genomics*, 16, 698-712.
 8. Leinonen, R., Sugawara, H., Shumway, M. and Collaboration, INSD. (2010). The sequence read archive. *Nucleic acids research*, 39, 19-21.
 9. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
 10. Bolger, A.M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.
 11. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. and MacManes, M.D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference

24. Larasati, Y.A. and Meiyanto, E. (2018). Revealing the Potency of Cinnamon as an Anti-cancer and Chemopreventive Agent. Indonesian Journal of Cancer Chemoprevention, 9, 47-62.
25. Dutta, A. and Chakraborty, A. (2018). Cinnamon in Anticancer Armamentarium: A Molecular Approach. Journal of toxicology.
26. Li, J., Teng, Y., Liu, S., Wang, Z., Chen, Y., Zhang, Y., Xi, S., Xu, S., Wang, R. and Zou, X. (2016). Cinnamaldehyde affects the biological behavior of human colorectal cancer cells and induces apoptosis via inhibition of the PI3K/Akt signaling pathway. Oncology reports, 35, 1501-1510.
- in Populus: phylogeny, organization, and expression. BMC Plant Biology, 9, p 26.
22. Chabannes, M., Barakate, A., Lapierre, C., Marita, J.M., Ralph, J., Pean, M., Danoun, S., Halpin, C., Grima-Pettenati, J. and Boudet, A.M. (2001). Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. The Plant Journal, 28, 257-270.
23. Carvalho, A.A. and Andrade, L.N. (2015). de Sousa ÉBV, de Sousa DP. Antitumor phenylpropanoids found in essential oils. BioMed research international.

Phenylpropanoid gene expression analysis in saffron using transcriptome data

Mahbobeh Dehghan¹, Masoud Tohidfar²

Abstract

Today, saffron is a known strategic plant in agricultural. It knowns as the most expensive spice and also has several medical usage in traditional medicine. One of the most important usage of saffron is in cancer prevention and cancer therapy. In this study data were downloaded from website NCBI short read archive, which contains five different tissue mRNA sequences, stigma, tepal, leaf, corm and stamen. After downloading raw data, quality control and trimming were done. Then, gene expression quantification was done using RSEM software, and finally blastx was applied between reference assembled transcriptome and some of known genes in phenylpropanoid biosynthesis pathway. In this study, for the first time we report expression of CAD1 which codes the Cinnamoldehyde enzyme, it has several reports on its anticancer and anti-tumor effects. The expression level of CAD1 was high in Stigma and Tepal, which means Stigma could be used as anti-cancer drugs as well as its spice.

Keywords: Saffron, Transcriptum, Gene, Biosynthetic pathway

¹ PH.D student of Biotechnology Engineering, Department of Plant Biotechnology, Faculty of Science and Biotechnology, Shahid Beheshti University, Tehran, Iran.

² * Corresponding author, Associate Professor of Agricultural Biotechnology, Department of Plant Biotechnology, Faculty of Science and Biotechnology, Shahid Beheshti University, Tehran, Iran, Email: m_tohidfar@sbu.ac.ir.