نشریه زراعت و فناوری زعفران
جلد ۷، شماره ۱، بهار ۱۳۹۸ ص. ۸۰-۶۹

Saffron Agronomy & Technology
Vol. 7, No. 1, Spring 2019, P. 69-80

مقایسه عملکرد ابزارهای سرهم‌بندی ترنسکریپتوم در گیاه زعفران زراعی (*Crocus sativus* L.)

مریم واحدی[۱]، سید علیرضا سلامی[۲*]، مجید شکرپور[۳] و حسن رضادوست[۴]

تاریخ دریافت: ۱۱ خرداد ۱۳۹۶ **تاریخ پذیرش:** ۹ آذر ۱۳۹۶

واحدی، م.، سلامی، س.ع.، شکرپور، م.، و رضادوست، ح. ۱۳۹۸. مقایسه عملکرد ابزارهای سرهم‌بندی ترنسکریپتوم در گیاه زعفران زراعی (*Crocus sativus* L.). زراعت و فناوری زعفران، ۷(۱): ۸۰-۶۹.

چکیده

زعفران گیاه دارویی و ادویه‌ای ارزشمند متعلق به خانواده زنبق و به‌عنوان منبع غنی از آپوکاروتنوئیدها در جهان محسوب می‌شود. به دلیل سایز بزرگ و پیچیدگی ژنوم زعفران توالی‌یابی آن به‌عنوان چالش مطرح است. با ظهور تکنیک‌های توالی‌یابی نسل بعد، توالی‌یابی RNA به‌عنوان منبع غنی مطالعات بیولوژیکی توسعه یافته است. سرهم‌بندی ترنسکریپتوم‌ها از تعداد بی‌شمار خوانش‌های کوتاه منبعی غنی برای مطالعه گونه‌هایی که ژنوم مرجع آن‌ها در دسترس نیست فراهم می‌کند. اما سرهم‌بندی قرائت‌ها و رسیدن به نتیجه مطلوب به‌ویژه برای گیاهان پلی‌پلوئید همواره یک چالش بزرگ محسوب می‌شود. در این مطالعه کارایی ابزارهای مختلف سرهم‌بندی با توجه به فاکتورهایی هم‌چون طول N50، تعداد کل یونی‌ژن‌ها و درصد هم‌ردیفی مورد مقایسه قرار گرفتند. نتایج نشان داد که Bridger به‌عنوان ابزاری بهتر جهت سرهم‌بندی قرائت‌های ترنسکریپتوم زعفران است که می‌تواند سرهم‌بندی بر اساس پارامترهای تعداد ترنسکریپت‌ها، طول N50، اندازه کل سرهم‌بندی و درصد هم‌ردیفی قرائت‌ها به ترنسکریپتوم فراهم آورد. Velvet/Oases بالاترین درصد درهم-ریختگی را نشان می‌دهد که منجر می‌شود اعضای مختلف یک خانواده ژنی که شباهت بالایی به یکدیگر دارند در یک ترنسکریپت سرهم‌بندی شوند. نتایج حاصل از این تحقیق می‌تواند به محققان در جهت انتخاب بهتر ابزار سرهم‌بندی و توسعه ابزارهای موجود راهکارهایی را ارائه نماید.

کلمات کلیدی: زعفران، سرهم‌بندی، BinPacker، Bridger، Trinity، Velvet/Oases.

۱- دانشجوی دکتری، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج.
۲- استادیار پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج.
۳- دانشیار پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج.
۴- استادیار، پژوهشکده گیاهان و مواد اولیه دارویی، دانشگاه شهید بهشتی، تهران.
(*- نویسنده مسئول: asalami@ut.ac.ir)

DOI: 10.22048/jsat.2017.87859.1235

مقدمه

زعفران *Crocus sativus* L. از خانواده زنبق (Iridaceae) از جمله گیاهان دارویی و ارزشمند دنیا است که کشت آن از دیرباز به علت کلاله گران‌بهایش در ایران رواج داشته است (Fernandez, 2004; Kafi et al., 2006; Izadpanah et al., 2014; Nemati et al., 2014). نتایج آنالیزهای فیتوشیمیایی و تجزیه شیمیایی زعفران حضور سه ماده مؤثره مهم و ارزشمند در مصارف دارویی و صنعتی به نام‌های کروسین (رنگیزه‌های کاروتنوئیدی محلول در آب)، پیکروکروسین (گلیکوزید تلخ‌مزه) و سافرانال (مسئول عطر و بو) را نشان می‌دهد (Yilmaz et al., 2010).

تریپلوئید و عقیم بودن گیاه زعفران به همراه وجود ژنومی بزرگ با اندازه تقریبی ۱۱/۸ Gb تحقیقات ژنتیکی در خصوص این گیاه محدود و با مشکل مواجه ساخته است (Fiore et al., 2010). کاهش محدودیت‌ها در ایجاد اطلاعات توالی، توصیف ژن‌ها و ژنوم‌ها، عملکرد ژن‌ها و مطالعه بیولوژی گیاه با ظهور تکنولوژی توالی‌یابی نسل بعد (NGS[1]) تسهیل شد. توالی‌یابی ترنسکریپت‌ها (RNA-Seq) به کمک تکنیک‌های نوین امکان شناسایی، بررسی بیان و مطالعه رونوشت ژن‌های درگیر در شبکه‌های تنظیمی و متابولیکی را با دقت و حساسیت بالایی فراهم آورده است (Wang et al., 2009; Marguerat & Bahler, 2010). آنالیز و پردازش داده‌ها مهم‌ترین و حساس‌ترین بخش در این حوزه محسوب می‌شود به‌طوری‌که بازسازی کامل ترنسکریپتوم از تعداد زیادی خوانش‌های کوتاه، چالش‌های محاسباتی بزرگی را به همراه داشته و دارد (Clarke et al., 2013; Zhao et al., 2011; Honass et al., 2016).

تاکنون نرم‌افزارهای زیادی از قبیل Abyss (Simpson et al., 2009)، Trinity (Grabherr et al., 2011)، Velvet/Oase (Schulz et al., 2012)، SOAPdenovo-Trans (Xie et al., 2014)، Bridger (Chang et al., 2015) و BinPacker (Liu et al., 2016) برای سرهم‌بندی[2] خوانش‌ها توسعه یافته است. غالب این نرم‌افزارها از رویکردی مشابه برای سرهم‌بندی قرائت‌ها استفاده می‌کنند اما اختلافی بزرگ در تعداد ترنسکریپت‌ها و ژن‌ها دارند (Wang & Gribskov, 2017). ساخت گراف‌های هم‌پوشان[3] و دی براین[4] دو استراتژی اصلی هستند که نرم‌افزارهای فوق از آن‌ها در ایجاد سرهم‌بندی استفاده می‌کنند (Grabherr et al., 2011; Chang et al., 2015). گراف‌دی براین با استفاده از K-mer های منفرد (گره[5]) که در محل لبه[6] (K-merهای هم‌پوشان) اتصال می‌یابند، ساخته می‌شود. Trinity اولین ابزاری است که به‌طور اختصاصی برای سرهم‌بندی ترنسکریپت‌ها طراحی شده است (Grabherr et al., 2011). این نرم‌افزار ابتدا با استفاده از قرائت‌های کوتاه کانتیگ[7]های بلندتر را می‌سازد و سپس با استفاده از این کانتیگ‌ها تعداد زیادی گراف دی براین ایجاد می‌کند. در نهایت تمامی مسیرهای ممکن به‌عنوان نماینده ایزوفرم‌ها از هر گراف مشتق می‌شود. برخلاف نرم‌افزارهای دیگر، نرم‌افزار Bridger (Chang et al., 2015) بر اساس روش نوین‌تر حداقل پوشش مسیر[8]، گراف‌های اسپلایسینگ را برای ساخت ترنسکریپت‌ها ایجاد می‌نماید (Chang et al., 2015). BinPacker (Liu et al., 2016) جزء جدیدترین ابزارهای سرهم‌بندی است که از دو ویژگی منحصر به فرد

1-Next Generation Sequencing

2-Assembly
3-Overlap graph
4-De brujin
5-Node
6-Edge
7-Contig
8-Minimum path cover

برخوردار است: ۱- تنها اسپلایسنگ جانکشن‌ها[1] در فرایند سرهم‌بندی دخیل هستند. ۲- قرائت‌های pell-mell به‌واسطه حرکت شانه‌ای در طول لبه‌های اتصال در یک گراف سرهم-بندی می‌شوند.

پارامترهای زیادی برای ارزیابی ترنسکریپتوم رفرنس وجود دارد. زمان محاسبات[2]، میزان استفاده از حافظه[3] کامپیوتر، N50 N50 و هم‌پوشانی[4] توالی‌ها، تعداد ترنسکریپت‌ها (کانتیگ‌ها)، هم‌ردیفی خوانش‌ها به ترنسکریپتوم مرجع، میزان خطای سرهم‌بندی[5] و دقت[6] سرهم‌بندی از مهم‌ترین پارامترهایی هستند هستند که برای ارزیابی نتایج سرهم‌بندی و کارایی نرم‌افزار در ایجاد یک نتیجه مطلوب مورد بررسی و مقایسه قرار می‌گیرند (Zhao et al., 2011; Honaas et al., 2016). پارامترهایی از قبیل شمار کانتیگ‌ها، ارزش N50 و طول کانتیگ جز مهم‌ترین این پارامترها در ارزیابی یک ترنسکریپتوم هستند (Baker et al., 2012). خطای سرهم‌بندی انواع مختلفی دارد که متداول‌ترین آن‌ها شامل درهم‌ریختگی[7] (زمانی که اعضای یک خانواده ژنی داخل یک کانتیگ سرهم‌بندی می‌شوند)، کیمریسم[8] کیمریسم[8] (ادغام دو یا تعداد بیشتری ترنسکریپت در یک کانتیگ در طول سرهم‌بندی)، اضافه‌های حمایت نشده[9]، ناتمامی[10]، تکه‌تکه شدن[11] و افزونگی[12] است. ارزش آماری به طول کوتاه‌ترین توالی در ۵۰ درصد کل سرهم‌بندی اطلاق می‌شود که جزء یکی از مهم‌ترین پارامترهای ارزیابی سرهم-بندی محسوب می‌شود. طویل‌تر بودن طول N50 نشان از کارایی بالاتر نرم‌افزار در ایجاد یک سرهم‌بندی بهتر است (Lin et al., 2011). با این وجود برخی مطالعات نشان می‌دهد شاخص‌های ارزیابی کیفیت سرهم‌بندی از قبیل N50 و همچنین اندازه کل سرهم‌بندی که در سرهم‌بندی ژنوم استفاده می‌شود برای بررسی کیفیت سرهم‌بندی ترنسکریپتوم مناسب نیستند چرا که توالی‌های طویل‌تر یا اندازه بزرگ‌تر سرهم‌بندی به‌منزله بهتر بودن سرهم‌بندی نیست درواقع این شاخص‌ها ممکن است سطح بالای کیمریسم را در پی داشته باشد (Wang & Gribskov, 2017). از آنجایی که ارزیابی برخی از پارامترهای کیفیت سرهم‌بندی ترنسکریپتوم مانند شمار کیمریسم‌ها و شمار ترنسکریپت‌ها با طول کامل در مقایسه با ژنوم مرجع امکان‌پذیر است پارامترهایی چون N50 و اندازه کل سرهم‌بندی نیز برای بررسی کیفیت سرهم‌بندی استفاده می‌شود. هناس و همکاران (Honaas et al., 2015) پارامترهای، (۱) نسبت خوانش‌های هم‌ردیف شده به ترنسکریپتوم، (۲) طول N50 و (۳) تعداد یونی ژن‌ها را ازجمله پارامترهای مورد بررسی در ارزیابی سرهم‌بندی معرفی کردند.

تا سال ۲۰۱۴ اطلاعات چندانی از مسیر بیوسنتزی ترکیبات مهم زعفران وجود نداشت تا آنکه تکنیک‌های و ابزارهای فوق محققین را بر آن داشت اطلاعات نسبتاً جامعی از ژن‌های درگیر در تولید این ترکیبات فراهم آوردند، از همین رو بابا و همکاران (Baba et al., 2015) با استفاده از تکنیک Illumina اولین کاوش در ترنسکریپتوم زعفران را رقم زدند. در این مطالعه از نرم‌افزار Trinity برای سرهم‌بندی خوانش‌ها استفاده شد که منجر به تولید ۶۴۴۳۸ ترنسکریپت شد. ۳۲۲۰۴ یونی ژن[13] به ۹۸۵۳ کلاستر و ۲۲۳۵۱ سینگلتون[14] طبقه‌بندی شدند. متوسط طول کانتیگ با ۶۰۹/۵۷ جفت باز، میزان GC ۴۳/۹۹ درصد و N50 با طول ۷۵۳ جفت باز دیگر پارامترهای محاسبه شده برای

1-Splicing junctions
2-Computation time
3-RAM usage
4-Coverage
5-Assembly error rate
6-Accuracy
7-Collapse
8- Chimerism
9- Unsupported insertion
10- Incompleteness
11- Fragmentation
12- Redundancy

13- Unigene
14- Singleton

ارزیابی نتایج سرهم‌بندی در این مطالعه بودند. دومین مطالعه در سال ۲۰۱۶ توسط جین و همکاران (Jain et al., 2016) انجام شد. در این مطالعه از توالی‌یابی ترنسکریپتوم پنج اندام با بافت مختلف برای سرهم‌بندی با ابزارهای متفاوت استفاده شد، از میان نرم‌افزارهای مورد مطالعه، Oases بهترین خروجی را با ۱۱۲۰۳۷ ترنسکریپت با متوسط طول کانتیگ ۶۲۵ جفت باز و N50 ۱۰۳۱ جفت باز داشت.

تلاش برای به حداکثر رساندن ترنسکریپت‌هایی با طول کامل در *Nicotiana benthamiana* از چهار ابزار سرهم‌بندی مختلف شامل TransAbyss، Trinity، SOAPdenovo-Trans و Oases با مقادیر متفاوتی از K-mer ها استفاده شد. نتایج این بررسی نشان داد که برای دست‌یابی به ترنسکریپتومی با بالاترین کیفیت، ترکیب کردن نتایج تعداد زیادی ابزار سرهم-بندی یک مزیت است و می‌تواند نتایج بهتری نسبت به تک‌تک ابزارها فراهم کند (Nakasugi et al., 2014). کاباو و همکاران (Cabau et al., 2017) نشان دادند که ترکیب کردن نتایج Trinity و Oases می‌تواند نتایج بهتری نسبت به دیگر ابزارها فراهم کند.

پژوهش حاضر باهدف بررسی کارایی ابزارهای سرهم‌بندی برای خوانش‌های تولیدی از نمونه زعفران ایران انجام گرفت. چرا که توانایی در تصحیح بازسازی و تشخیص میان ترنسکریپت‌هایی با شباهت بالا حاصل از ژن‌های هومولوگ و پارالوگ چالشی دیگر در سرهم‌بندی ترنسکریپتوم گیاهان پلی-پلوئید است. به‌طوری که این پدیده با حضور ایزوفرم‌ها پیچیده‌تر نیز می‌شود. نتایج حاصل از این تحقیق می‌تواند معیارهای اصولی برای مقایسه ابزارهای سرهم‌بندی را مشخص نماید و درنهایت بهترین و قدرتمندترین ابزار برای سرهم‌بندی خوانش‌های کوتاه گیاه زعفران در حجم وسیع داده را پیشنهاد کند.

مواد و روش‌ها

مواد گیاهی و استخراج RNA

استخراج RNA، از کلاله و کالوس فریز شده در ۸۰- درجه سانتی‌گراد با سه تکرار بیولوژیکی و تکنیکی انجام شد. هاون‌ها با ازت مایع سرد شدند و سپس ساییدن کلاله‌ها به‌سرعت با کمک ازت انجام گرفت. به کمک اسپاتول‌های استریل، بافت پودر شده به میزان ۱۰۰ میلی‌گرم به تیوپ ۲ میلی‌لیتری از قبل سرد شده منتقل شد. استخراج RNA توسط کیت استخراج RNA شرکت سیگما مطابق دستورالعمل شرکت سازنده انجام گرفت. درنهایت به‌منظور بررسی کیفیت RNAهای استخراج شده از الکتروفورز RNA کل استخراج شده و دستگاه نانودراپ با نسبت جذب در طول‌موج ۲۶۰ به ۲۸۰ نانومتر، ۲۶۰ به ۲۳۰ نانومتر و غلظت RNA استفاده شد. تائید نهایی کمیت و کیفیت RNAها با استفاده از دستگاه بایوآنالایزر ۲۱۰۰ انجام شد.

ساخت کتابخانه و توالی‌یابی

ساخت کتابخانه با استفاده از پروتکل بهینه‌شده SMART-Seq (Zhu et al., 2001) در آزمایشگاه Aureliano Bombarely واقع در دانشگاه Virginia Tech آمریکا انجام شد و درنهایت بعد از کنترل کمیت و کیفیت توسط دستگاه بایوآنالایزر، کتابخانه جهت توالی‌یابی به دانشگاه دوک ارسال شد. جهت توالی‌یابی از پلتفرم IlluminaHiseq با خوانش دوطرفه به طول ۱۵۰bp استفاده شد.

آنالیز بیوانفورماتیک

کنترل کیفیت داده‌ها

در مرحله اول کنترل کیفیت داده‌ها با استفاده از نرم‌افزار FastQC v0.11.2 انجام گرفت. سپس از نرم‌افزار Trimmomatic v0.32 (Bloger et al., 2014) برای حذف

توالی‌های آداپتوری سمت ′۳ و ′۵ و حذف خوانش‌های بی‌کیفیت استفاده گردید. در همین راستا خوانش‌های باکیفیت پایین کمتر از ۳۰ و خوانش‌هایی که بعد از پیرایش طولی کوتاه‌تر از ۵۰ نوکلئوتید داشتند، حذف شدند. این مقادیر آستانه‌های حداقل برای کنترل کیفیت توالی‌ها محسوب می‌شوند. خوانش‌های پیرایش شده با استفاده از نرم‌افزار FastQC v0.11.2مورد کنترل کیفیت مجدد قرار گرفتند.

سرهم‌بندی قرائت‌ها

از نرم‌افزارهای Trinity-2.0.6 (Grabherr et al., 2011)، Bridger v2014-12-01 (Chang et al., 2015)، BinPacker v1.0 (Liu et al., 2016) و Velvet v1.2.10/Oases v0.2.0 (Schulz et al., 2012) برای سرهم‌بندی توالی‌ها استفاده شد. از ابزار Trinity برای سرهم‌بندی با اندازه K-mer-25 استفاده شد. این در حالی است که دو اندازه K-mer-25 و K-mer-27 برای نرم‌افزارهای Bridger و BinPacker انتخاب شد. Velvet/Oases برای سرهم‌بندی با اندازه K-mer-25 تا K-mer-105 بافاصله ۴bp مورد استفاده قرار گرفت.

ارزیابی کیفیت سرهم‌بندی‌های به‌دست آمده با استفاده از پارامترهایی چون شمار ترنسکریپت‌ها، N50 و درصد هم‌ردیفی به کمک Bowtie2 (Langmead & Salzberg, 2012)، اندازه کل[1] سرهم‌بندی، میانگین طول کانتیگ‌ها و تعداد ترنسکیپت‌ها با اندازه کامل یا نزدیک به کامل انجام شد. برای شناسایی ژن‌ها و بررسی بهترین نتایج حاصل از سرهم‌بندی به کمک ارزیابی کیفیت از BlastX با اطلاعات موجود در بانک اطلاعاتی UniProt استفاده شد.

نتایج و بحث

توالی‌یابی ترنسکریپتوم

توالی‌یابی کتابخانه‌های cDNA جهت ایجاد ترنسکریپتومی جامع از کلاله و کالوس زعفران، ۱۸۹۲۶۵۵۷۲ قرائت برای کالوس (از ۱۸/۴ تا ۵۷/۴ میلیون قرائت برای سه تکرار) و ۱۷۳۸۴۰۰۴۴ قرائت برای کلاله (از ۲۴/۶ تا ۳۷/۵ میلیون قرائت برای سه تکرار) تولیدکرد. ۲۴٪ قرائت‌ها برای کالوس و ۲۲٪ قرائت‌ها برای کلاله با فیلترینگ توالی‌های آداپتوری و قرائت-هایی باکیفیت پایین و طول کوتاه حذف شدند. درنهایت ۱۴۳۳۷۵۶۹۰ قرائت از کالوس و۱۳۴۸۷۸۹۷۰ قرائت از کلاله جهت سرهم‌بندی مورد استفاده قرار گرفت.

شمار کانتیگ‌ها و اندازه کل سرهم‌بندی

ارزیابی شمار کانتیگ‌های به‌دست آمده از ابزارهای مختلف، عملکرد متفاوت این ابزارها را در ایجاد سرهم‌بندی قرائت‌ها نشان داد. Trinity ، اولین ابزاری است که به‌طور اختصاصی برای سرهم‌بندی قرائت‌های کوتاه جهت ایجاد ترنسکریپتوم رفرنس طراحی شده است و در بسیاری از مطالعات به‌ویژه در حوزه گیاهی از کارایی بالایی نسبت به برخی دیگر از ابزارها برخوردار بوده است (Grabherr et al., 2011; Haas et al., 2013). با این وجود نتایج به‌دست آمده از این ابزار در سرهم‌بندی قرائت‌های ترنسکریپتوم زعفران چندان رضایت‌بخش نیست زیرا تعداد ۶۴۷۸۰۳ ترنسکریپت دسته‌بندی شده در ۵۱۶۹۸۹ یونی‌ژن عدد بسیار بالایی است. حصول شمار بالایی از ترنسکریپت‌ها در دیگر ابزار استفاده‌شده در این مطالعه نیز به چشم می‌خورد. شمار ترنسکریپت‌های به‌دست آمده به کمک Bridger با پارامتر K-mer به طول ۲۵ و ۲۷bpبه ترتیب برابر با ۲۶۲۰۲۴ و ۴۳۴۳۴۰ بودند. ابزار Bridger توانست شما ترنسکریپت‌ها را در K-mer به طول ۲۵bp به‌طور قابل‌توجهی به بیش از نصف در مقایسه با Trinity کاهش دهد. BinPacker و Velvet/Oases نتوانستند نتایجی بهتر از

1 Total size

Bridger فراهم کنند. از دیدگاه مقایسه میان نرم‌افزارهای موردبررسی در این مطالعه کمترین تعداد ترنسکریپت‌های به‌دست آمده محصول نرم‌افزار Bridger با K-mer به طول ۲۵bp بود. سرهم‌بندی‌ها با بیشترین و کمترین ترنسکریپت‌های تولیدشده به این ترتیب قرار گرفتند (جدول ۱).

Trinity<Velvet/Oases-K-mer-25<Velvet/Oases-K-mer-27<BinPacker-K-mer-25<BinPacker-K-mer-27<Bridger-K-mer-25<Bridger-K-mer-27

جدول ۱- شاخصه‌های مربوط به سرهم‌بندی از مجموع خوانش‌های حاصل از کتابخانه، توسط نرم‌افزارهای Trinity، Bridger و BinPacker در گیاه زعفران

Table 1- Assembly statistics from library reads by Trinity, Bridger and BinPacker softwares in saffron plant

نرم‌افزار Software	Trinity	Bridger		BinPacker	
اندازه کامر (جفت باز) K-mer length (bp)	25	25	27	25	27
تعداد ترنسکریپت‌ها Transcripts counts	647803	262024	434340	442603	439676
مجموع کل بازها Total base (million)	294.13	159.08	265.61	385.47	374.32
طول N50 (جفت باز) N50 Length (bp)	461	884	860	1360	1330
میانگین طول یونی ژن‌ها (جفت باز) Average length of uigene (bp)	454.05	607.12	611.54	870.93	851.36
طول بلندترین یونی ژن (جفت باز) Maximum length of unigene (bp)	17066	16500	23320	18510	19114
طول کوتاه‌ترین یونی ژن (bp) Minimum length of unigene (bp)	224	201	201	200	200

سرهم‌بندی‌های خلق‌شده توسط Velvet/oases یک کاهش تدریجی در شمار ترنسکریپت‌ها را با افزایش اندازه K-mer نشان دادند. از میان K-merهای مورد بررسی K-mer با اندازه ۸۵bp توانست شمار پایینی از ترنسکریپت‌ها را با حجم قابل قبولی داده (اندازه کل) را ایجاد نماید (جدول ۲). در کل شمار بالای ترنسکریپت‌های به‌دست آمده، تکه‌تکه بودن[1] ترنسکریپتوم مرجع و شباهت بالای بسیاری از ترنسکریپت‌ها را نشان می‌دهد. پایین بودن میزان پوشش[2] یکی از دلایل اصلی تکه‌تکه بودن سرهم‌بندی است (Smith-Unna et al., 2016). تخمین یک حداقل پوشش موردنیاز برای یک سرهم‌بندی موفق سخت و دشوار است. درواقع رسیدن به یک حداقل عمق برای سرهم‌بندی ژن‌هایی بابیان بالا آسان است (به‌عنوان مثال ۵۰۰Mbp) در حالی که عمق موردنیاز برای پوشش ژن‌هایی با بیان پایین شاید عملاً امکان‌پذیر نباشد. نتایج بررسی‌ها نشان می‌دهد که چیزی در حدود ۵Gbp–۴ اطلاعات می‌تواند نماینده مناسبی از ترنسکریپتوم گیاهی باشد. با این وجود توجه به این نکته بسیار مهم است که اطلاعات بیشتر می‌تواند به‌طور هم‌زمان سرهم‌بندی ژن‌هایی با بیان پایین را بهبود بخشد در حالی که فراوانی ژن‌هایی با بیان بالا را نیز افزایش می‌دهد (Honaas et al., 2016).

از طرفی دیگر، به‌طور قطع شمار ترنسکریپت‌ها و اندازه کل سرهم‌بندی به‌دست آمده متناسب با تعداد خوانش‌های استفاده شده است (Chang et al., 2015). محصول سرهم‌بندی بیش

1- Fragmentation
2- Coverage

از ۶۴ میلیون قرائت از نمونه گل و ۵۱ میلیون قرائت کلاله، ۶۴۴۳۸ ترنسکریپتوم با استفاده از Trinity در مطالعه بابا و همکاران (Baba et al., 2015) بود. این درحالی است که از ۲۲۵/۸ میلیون قرائت استفاده شده در مطالعه جین و همکاران (Jain et al., 2016)، ۳۲۹۹۶ ترنسکریپتوم با اندازه کل ۲۷/۴۷Mp به‌وسیله Trinity به‌دست آمد. هرچند ارتباط مستقیمی میان تعداد خوانش‌ها، تعداد ترنسکریپت‌ها و اندازه کل سرهم‌بندی وجود ندارد اما به‌طور عموم تعداد ترنسکریپ بالاتر ایجاد شده در سرهم‌بندی، اندازه کل بالاتری در سرهم‌بندی ایجاد می‌کند.

جدول ۲- شاخصه‌های مربوط به سرهم‌بندی از مجموع خوانش‌های حاصل از کتابخانه، توسط نرم‌افزارهای Velvet/Oases در گیاه زعفران
Table 2- Assembly statistics from library reads by Velvet/Oases softwares in saffron plant

طول کامر (جفت باز) K-mer length (bp)	تعداد کانتیگ‌ها Number of contigs	مجموع کل بازها Total base (million)	طول N50 N50 length	میانگین طول یونی ژن‌ها (جفت باز) Average length of unigene (bp)	طول بلندترین یونی ژن (جفت باز) Maximum length of unigene (bp)	طول کوتاه‌ترین یونی ژن (جفت باز) Minimum length of unigene (bp)
25	403510	248.24	868	615.2	17665	100
29	409192	253.92	897	620.54	16378	97
33	412643	261.57	940	633.89	19973	110
37	410550	265.19	983	645.96	29372	122
41	405021	267.54	1030	660.57	65452	126
45	399338	268.09	1072	671.33	19998	132
49	390892	264.75	1108	677.31	30841	152
53	383433	263.45	1169	687.09	40300	138
57	373488	261.04	1231	698.94	41100	155
61	362547	256.02	1269	706.18	33283	191
65	388176	364.09	1773	937.95	41480	200
69	353441	360.23	1927	1019.22	18565	200
73	303519	328.2	1978	1081.34	16457	200
77	264298	298.91	2029	1130.97	33836	200
81	230623	265.26	2036	1150.22	20685	200
85	196262	231.51	2020	1179.61	20834	200
89	158164	137.88	1411	871.78	22601	200
93	133824	118.7	1394	887.05	21641	200
97	112600	99.61	1341	884.67	20690	200
101	112738	95.167	1211	844.14	14415	200
105	91609	76.37	1161	833.7	13987	200

با این حال این الگو با نتایج بدست آمده از Trinity و BinPacker در تناقض است، زیرا BinPacker با تولید تعداد ترنسکریپت کمتر نسبت به Trinity توانست اندازه کل بالاتری برای سرهم‌بندی ایجاد کند امّا مقایسه BinPacker با Bidger الگوی فوق را تائید می‌کند. Bridger با اندازه K-mer به طول ۲۵bp تعداد ترنسکریپت کمتر همراه با اندازه کل سرهم‌بندی کمتری ایجاد کرد.

طول ترنسکریپت‌ها و N50

میانگین طول یونی‌ژن‌ها در خروجی نرم‌افزار Trinity در حدود ۴۵۴/۰۵bp تخمین زده شد. بالاترین مقدار برای این پارامتر توسط Bridger با K-mer به طول ۲۷bp به‌دست آمد (۶۰۷/۲۱bp) که تنها در چند نوکلئوتید با K-mer به طول ۲۵bp اختلاف داشت (۶۱۱/۵۴). با این وجود طول کوتاه‌ترین یونی‌ژن در Trinity بیش از Bridger در هر دو K-mer بود. در سرهم‌بندی‌های به‌دست آمده از Velvet/Oases میانگین طول ترنسکریپت‌ها با افزایش اندازه K-mer تا ۸۵bp افزایش تدریجی (ماکزیمم میانگین طول ترنسکریپت ۱۱۷۹/۶۱bp متعلق به K-mer با طول ۸۵bp) و با طویل‌تر شدن اندازه K-mer کاهش نشان داد.

ارزیابی توزیع طول کانتیگ‌ها نشان داد که در ترنسکریپتوم مرجع به‌دست آمده از Trinity، ۹۳/۵٪ کانتیگ‌ها طولی کمتر از ۱Kb و۶/۵٪ طولی مابین ۱ تا ۵Kb داشتند در حالی که نتایج Bridger با K-mer به طول ۲۵bp به‌ترتیب ۸۶/۱٪ و ۱۳/۵٪ و Bridger با K-mer به طول ۲۷bp به‌ترتیب ۸۵/۸٪ و ۱۳/۹٪ بود. نرخ تعداد کانتیگ‌ها در این توزیع در حدود ۱/۶ و ۱/۷ برابر به سود Bridger با K-mer به طول ۲۷bp بود.

ازنظر تئوری سرهم‌بندی‌های بهتر می‌بایست تعداد قرائت‌های بیشتری را در کانتیگ‌های طویل‌تر باارزش N50 بالاتر داشته باشد. با این وجود اهمیت N50 هنوز موردبحث بسیاری از محققین است (Salzberg et al., 2012). مطالعات نشان می‌دهد که ارزش N50 بر پایه استراتژی K-mer یا تعیین حداقل طول کانتیگ توسط کاربر می‌تواند افزایش یابد. نتایج این مطالعه این امر را اثبات کرد چراکه بالاترین ارزش N50 متعلق به ترنسکریپتوم مرجع به‌دست آمده از Velvet/Oases با اندازه K-mer به طول ۸۱bp بود. کانتیگ‌های کوتاه با K-merهای بسیار طویل که قرائت‌های کوتاه ترنسکریپت‌هایی با فراوانی پایین را سرهم‌بندی نمی‌کند یا K-merها کوتاهی که ترنیسکریپت‌های کوتاه تکه‌تکه به دلیل فقدان همپوشانی را سرهم‌بندی نمی‌کند ایجاد می‌شود (Chopra et al., 2014; Surget-Groba et al., 2010). بررسی اثر طول K-mer با دو نرم‌افزار SOAP و Oases نشان داد که K-merهای کوتاه‌تر N50 بالاتری ایجاد می‌کنند. با این وجود ارزیابی ۱۱ پارامتر ترنسکریپتوم نشان داد که خود نرم‌افزار تأثیر بیشتری در ایجاد یک ترنسکریپتوم مرجع بهتر دارد تا متغیر K-mer که با نتایج پژوهش‌هایی (Chopra et al., 2014; Chang et al., 2015; Surget-Groba & Montoya-Burgos, 2010) که مزایا و معایب ارزش‌های بالا و پایین K-mer را برجسته کردند در تضاد بود (Rana et al., 2016).

درصد هم‌ردیفی

میانگین درصد هم‌ردیفی قرائت‌ها با ترنسکریپتوم مرجع به‌دست آمده از Bridger با K-mer به طول ۲۵ برای کالوس و کلاله به ترتیب ۸۸/۷۲ و ۹۳/۲ درصد و برای Bridger با K-mer به طول ۲۷ به‌ترتیب ۹۱/۴۸ و ۹۴/۶ درصد تخمین زده شد. در حالی که میانگین درصد هم‌ردیفی برای سرهم‌بندی

به‌دست آمده از Trinity برای کالوس و کلاله بیش از ۹۱ و ۹۵ درصد محاسبه شد. پایین‌ترین نرخ هم‌ردیفی مربوط به سرهم-بندی به‌دست آمده از Velvet/Oases با K-mer به طول ۸۵bp بود (جدول ۳). نسبت پایین قرائت‌های هم‌ردیف شده نشان می‌دهد که ترنسکریپتوم مرجع مربوطه نمی‌تواند نماینده کامل و جامعی برای انجام آنالیزهای بعدی باشد.

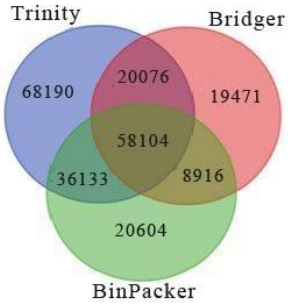
جدول ۳- درصد هم‌ردیفی قرائت‌ها در برابر ترنسکریپتوم سرهم‌بندی شده در زعفران
Table 3- Alignment percent of reads to assembled transcriptome in saffron

نرم‌افزار Software	Trinity	Bridger	BinPacker	Velvet/Oases	
طول کامر (جفت باز) K-mer length (bp)	25	25	27	85	81
کالوس جنین‌زا- تکرار ۱ Embryogenic callus- Rep 1	93.58	87.4	90.61	65.57	72.01
کالوس جنین‌زا- تکرار ۲ Embryogenic callus- Rep 2	94.9	90.04	92.35	67.79	75.06
کالوس جنین‌زا- تکرار ۳ Embryogenic callus- Rep 3	94.51	89.67	92.15	64.25	73.43
کالوس غیرجنین‌زا- تکرار ۱ Non embryogenic callus- Rep 1	94.05	90.92	92.93	64.6	74.14
کالوس غیرجنین‌زا- تکرار ۲ Non embryogenic callus- Rep 2	93.82	87.56	90.75	65.34	72.01
کالوس غیرجنین‌زا- تکرار ۳ Non embryogenic callus- Rep 3	93.16	86.74	90.08	63.69	71.33
گل نرمال با سه کلاله- تکرار ۱ Normal flower with 3 stigmata- Rep1	95.10	92.22	93.82	67.74	76.32
گل نرمال با سه کلاله- تکرار ۲ Normal flower with 3 stigmata- Rep 2	95.7	94.06	95.07	63.8	75.07
گل نرمال با سه کلاله- تکرار ۳ Normal flower with 3 stigmata- Rep 3	94.31	92.65	94.17	66.47	76
گل موتانت با چهار کلاله- تکرار ۱ Mutant flower with 4 stigmata- Rep 1	96.04	94.22	95.45	65.31	76.32
گل موتانت با چهار کلاله- تکرار ۲ Mutant flower with 4 stigmata- Rep 2	94.92	92.53	94.14	66.51	74.62
گل موتانت با چهار کلاله- تکرار ۳ Mutant flower with 4 stigmata- Rep 3	96	93.66	94.97	67.31	74.39

تفسیر ترنسکریپت‌ها به کمک BlastX

برای تفسیر جامع ترنسکریپت‌های به‌دست آمده جستجوی شباهت برای توالی‌ها به کمک BlastX در مقابل پایگاه داده UniProt انجام شد. بر اساس پارامترهای بررسی و ذکرشده، اسمبل به‌دست آمده از Bridger، BinPacker و Trinity با اندازه K-mer به طول ۲۵bp برای انجام این آنالیز انتخاب شدند. بر اساس نتایج به‌دست آمده از BlastX بیش از ۷۰٪ ترنسکریپت‌ها با شباهتی بالای ۸۰٪ به توالی‌های موجود در بانک اطلاعاتی بلاست شدند. نتایج نشان می‌دهد بیش از ۵۰٪ ژن‌های شناسایی‌شده در ۳ سرهم‌بندی مقایسه شده با یکدیگر هم‌پوشانی دارند. در این میان Trinity بالاترین تعداد ژن منحصربه‌فرد را نشان داد که در دیگر اسمبلی‌ها موجود نبودند. از میان ۶۸۱۹۰ ژن منحصربه‌فرد در سرهم‌بندی Trinity تنها

نزدیک به ۲۰ هزار ترنسکریپت درصد شباهت بالای ۷۰٪ درصد نشان دادند. نتایج BlastX نشان می‌دهدسرهم‌بندی به‌دست آمده از ابزار Bridger بالاترین تعداد ترنسکریپت با طول را دارد (شکل ۱). تعداد ترنسکریپت‌ها با طول کامل برای سه سرهم-بندی Bridger، BinPacker و Trinity به‌ترتیب ۴۷۳۴، ۴۰۲۸ و ۳۹۸۶ بود.



شکل ۱- مقایسه نتایج حاصل از BlastX سرهم‌بندی‌های به‌دست آمده از سه ابزار Trinity، Bridger و BinPacker

Figure 1- BlastX comparisons of produced assembled from Trinity, Bridger and BinPacker tools.

نتیجه‌گیری

در حال حاضر استانداردی طلایی برای انجام یک سرهم‌بندی با استفاده از قرائت‌های کوتاه ترنسکریپتوم در گیاهان غیر مدل و بدون ژنوم مرجع وجود ندارد. اثبات شده است که نرم‌افزارهای متفاوت با استفاده از قرائت‌های یکسان و کاربرانی مشابه نتایج متفاوتی را تولید می‌کنند. (Moreton et al., 2014; Chopra et al., 2014; He et al., 2015; O'Neil et al., 2013). هدف اصلی این مطالعه ارزیابی چند نرم‌افزار پرکاربرد در سرهم‌بندی قرائت‌های ترنسکریپتوم گیاه تریپلوئید زعفران با پارامترهای مشخص، جهت تعیین استراتژی بهتر و با کارایی بالاتر بود. مطالعات متعددی حکایت از عملکرد بهتر Bridger نسبت به دیگر ابزارهایی از قبیل SOAP و Trinity بر اساس پارمترهای تعداد ترنسکریپت‌ها، اندازه N50، درصد هم‌ردیفی و تعداد ترنسکریپت‌ها با طول کامل دارد (Rana et al., 2016). مورد دیگری که Bridger و Trinity را از هم متمایز می‌کرد زمان لازم برای سرهم‌بندی قرائت‌ها است. اگرچه ما در این مطالعه سرعت را موردبررسی قرار ندادیم اما نتایج بررسی‌ها نشان می‌دهد که Bridger به علت سریع‌تر بودن محبوبیت بیشتری دارد. روی هم رفته Bridger به‌عنوان ابزاری بهتر جهت سرهم‌بندی قرائت‌های ترنسکریپتوم زعفران شناخته شد که می‌تواند سرهم‌بندی با پارامترهای بهتر فراهم آورد.

منابع

Baba, S.A., Mohiuddin, T., Basu, S., Swarnkar, M.K., Malik, A.H., Wani, Z.A., Abbas, N.A., Singh, A.K., and Ashraf, N. 2015. Comprehensive transcriptome analysis of *Crocus sativus* for discovery and expression of genes involved in apocarotenoid biosynthesis. BMC genomics 16 (1): 698–712.

Baker, M. 2012. De novo genome assembly: what every biologist should know. Nature Methods 9 (4): 333-337.

Cabau, C., Escudie, F., Djari, A., Guiguen Y., Bobe, J., and Klopp, C. 2017. Compacting and correcting Trinity and Oases RNA-Seq *de novo* assemblies. PeerJ 5: e2988.

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., and Huang, X. 2015. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biology 16 (1): 1.

Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C.E., and Burow, M.D. 2014. Comparisons of de novo transcriptome

assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-seq data. PloS One 9 (12): e115055.

Clarke, K., Yang, Y., Marsh, R., Xie, L., and Zhang, K.K. 2013. Comparative analysis of de novo transcriptome assembly. Science China Life Sciences 56 (2): 156-162.

Fernandez, J.A. 2004. Biology, biotechnology and biomedicine of saffron. Recent Research Developments in Plant Science 2: 127–159.

Fiore, A., Pizzichini, D., Diretto, G., Scossa, F., and Spano, L. 2010. Genomics and transcriptomics of saffron: new tools to unravel the secrets of an attractive spice. The Editor 25: 1-14.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, Ch., Lindblad-Toh, K., Friedman N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29 (7): 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J. Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., and Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8 (8): 1494-1512.

Honaas, L.A., Wafula, E.K., Wickett, N.J., Der, J.P., Zhang, Y., Edger, P.P., Altman, N.S., Pires, J.C., Leebens-Mack, J.H., and dePamphilis, C.W. 2016. Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. PloS One 11 (1): e0146062.

Izadpanah, F., Kalantari, S., Hassani, M.E., Naghavi, M.R., and Shokrpour, M. 2014. Variation in Saffron (*Crocus sativus* L.) accessions and Crocus wild species by RAPD analysis. Plant Systematics and Evolution 300 (8): 1941-1944.

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.

Kafi, M. 2006. Saffron (*Crocus sativus*): Production and Processing. Science Publishers, 249 p.

Langmead, B., and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9 (4): 357-359.

Lin, Y., Li, J., Shen, H., Zhang, L., and Papasian, C.J. 2011. Comparative studies of de novo assembly tools for next-generation sequencing technologies. Bioinformatics 27 (15): 2031-2037.

Liu, J., Li, G., Chang, Z., Yu, T., Liu, B., McMullen, R., Chen, P., and Huang, X. 2016. BinPacker: packing-based De Novo transcriptome assembly from RNA-seq data. PLoS Computational Biology 12 (2): e1004772.

Marguerat, S., and Bähler, J. 2010. RNA-seq: from technology to biology. Cellular and Molecular Life Sciences 67 (4): 569-579.

Moreton, J., Dunham, S.P., and Emes, R.D. 2014. A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: assembly of the duck (Anasplatyrhynchos) transcriptome. Frontiers in Genetics 5 (190): 1-6.

Nakasugi, K., Crowhurst, R., Bally, J., and Waterhouse, P. 2014. Combining transcriptome assemblies from multiple De Novo assemblers in the Allo-tetraploid plant *Nicotiana benthamiana*. PloS One 9 (3): e91776.

Nemati, Z., Mardi, M., Majidian, P., Zeinalabedini,

M., Pirseyedi, S.M., and Bahadori, M. 2014. Saffron (*Crocus sativus* L.), a monomorphic or polymorphic species?. Spanish Journal of Agricultural Research 12 (3): 753-762.

O'Neil, S., and Emrich, S.J. 2013. Assessing De Novo transcriptome assembly metrics for consistency and utility. BMC Genomics 14 (465): 1-12.

Rana, S.B., Zadlock IV, F.J., Zhang, Z., Murphy, W.R., and Bentivegna, C.S. 2016. Comparison of De Novo transcriptome assemblers and k-mer strategies using the Killifish, fundulus heteroclitus. PloS One 11 (4): e0153104.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., Marcais, G., Pop, M., and Yorke, J.A., 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Research 22 (3): 557-567.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28 (8): 1086-1092.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J. E., Jones, S.J., and Birol, I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Research 19 (6): 1117-1123.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M., and Kelly, S. 2016. *TransRate: reference-free quality assessment of de novo transcriptome assemblies*. Genome Research 26: 1134–1144.

Surget-Groba, Y., and Montoya-Burgos, J.I. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. Genome Research 20 (10): 1432-1440.

Wang, S., and Gribskov, M. 2017. Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. Bioinformatics 33 (3): 327-333.

Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10 (1): 57-63.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K., and Wang, J.,. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30 (12): 1660-1666.

Yilmaz, A., Nyberg, N.T., Mølgaard, P., Asili, J., and Jaroszewski, J.W. 2010. 1H NMR metabolic fingerprinting of saffron extracts. Metabolomics 6 (4): 511-517.

Zhao, Q.Y., Wang, Y., Kong, Y.M., Luo, D., Li, X., and Hao, P. 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12 (14): S2.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechnique 30: 892–897.

Comparative performance of transcriptome assembly programs for saffron (*Crocus sativus* L.)

Maryam Vahedi[1], Seyed Alireza Salami[2], Majid Shokrpour[3] and Hassan Rezadoust[4]*

Submitted: 1 June 2017 **Accepted:** 30 November 2017

Vahedi, M., Salami, S.A., Shokrpour, M., and Rezadoust, H. 2019. Comparative performance of transcriptome assembly programs for saffron. Saffron Agronomy & Technology 7(1): 69-80.

Abstract

Saffron (*Crocus sativus* L.) belonging to the Iridaceae family as a source of apocarotenoids is one of the most valuable spices and medicinal plants in the world. Because of the large size and high complexity of saffron genome, its sequencing remains a challenge. The arrival of next-generation sequencing technologies (NGS) has allowed rapid and efficient development for RNA sequencing. De novo assembly of transcriptome from short-read RNA-Seq data provides a great resource for the study of species without a reference genome. *De novo* assembly of the transcriptome has some unique challenges, particularly in the case of plants, which possess a large amount of paralogs, orthologs, homoeologs and isoforms. In this research, we attempted to compare the performance of *de novo* assembly tools including BinPacker, Bridger, Oases-Velvet and Trinity through consideration of quality metrics such as N50 length, the total number of contigs and alignment scores. The results of these analyses revealed that assembly using Bridger had a superior performance for saffron transcriptome, Oases suffered from relatively high chimera rates and redundancies which causes genes family with high similarity to be assembled into one transcript, Trinity performs worse than Bridger in the increase of false positives. Our comparison study will assist researchers in selecting a well-suited assembler and offer essential information for the improvement of existing assemblers.

Keywords: Saffron, Assembly, BinPacker, Bridger, Trinity, Velvet/Oases.

1 - PhD. Student, College of Agriculture and Natural resources, University of Tehran, Karaj, Iran.
2- Assistant professor, College of Agriculture and Natural resources, University of Tehran, Karaj, Iran.
3- Associate professor, College of Agriculture and Natural resources, University of Tehran, Karaj, Iran.
4- Assistant professor, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Evin, Tehran, Iran
(*-Corresponding author Email: asalami@ut.ac.ir)
DOI: 10.22048/jsat.2017.87859.1235