ژنتیک نوین
دوره شانزدهم، شماره ٤، زمستان ١٤٠٠
صفحه ٣٦١ - ٣٤٩

استفاده از الگوریتم CTAnalyzer به‌منظور شناسایی توالی‌های کاندید pri-microRNA در ژنوم *Azadirachta indica*

Utilization of CTAnalyzer algorithm to identify pri-microRNA candidate sequences in the genome of *Azadirachta indica*

سبحان عطائی[1]، جعفر احمدی*[2]، سید امیر مرعشی[3]

۱- دانشجوی دکتری، گروه ژنتیک و به‌نژادی گیاهی، دانشگاه بین‌المللی امام‌خمینی (ره)، قزوین، ایران
۲- استاد، گروه ژنتیک و به‌نژادی گیاهی، دانشگاه بین‌المللی امام‌خمینی (ره)، قزوین، ایران
۳- دانشیار، گروه بیوتکنولوژی، پردیس علوم، دانشگاه تهران، تهران، ایران

Ataei S[1], Ahmadi J[*2], Marashi SA[3]
1- PhD Student, Department of Genetics and Plant Breeding, Imam Khomeini International University, Qazvin, Iran
2- Professor, Department of Genetics and Plant Breeding, Imam Khomeini International University, Qazvin, Iran
3- Associate Professor, Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran

* نویسنده مسئول مکاتبات، پست الکترونیکی: j.ahmadi@eng.ikiu.ac.ir

(تاریخ دریافت: ١٤٠٠/٠١/٢٤ - تاریخ پذیرش: ١٤٠٠/٠٧/٢٨)

چکیده

ساختارهایی که مولکول‌های microRNA بالغ از آن‌ها حاصل می‌شوند (pri-microRNA و pre-microRNA) دارای ویژگی‌های به‌خصوصی هستند که آن‌ها را از مولکول‌های مشابه، متمایز ساخته و برای آنزیم‌ها و پروسسورها قابل شناسایی می‌نماید. این خصوصیات عمدتاً متاثر از ویژگی‌های ساختار ثانویه مولکول RNA است. به‌همین دلیل، تا به امروز الگوریتم‌های متعددی به‌منظور پیش‌بینی ساختار ثانویه برای توالی‌های پلی‌نوکلئوتید طراحی شده است. نرم‌افزار CTAnalyzer الگوریتمی است که با هدف رقومی‌سازی ویژگی‌های ساختارهای ثانویه مولکول RNA طراحی شده و طبقه‌بندی این ساختارها را با محوریت ویژگی‌های مهم در شناسایی مولکول‌های microRNA انجام می‌دهد. در تحقیق حاضر، نرم‌افزار CTAnalyzer با تعریف فیلترها و ضوابط مناسب، برای بررسی و طبقه‌بندی ساختارهای ثانویه استخراج شده از توالی ژنوم گیاه *Azadirachta indica* مورد استفاده قرار گرفت و هدف از این پژوهش، بررسی کارامدی این نرم‌افزار در شناسایی جزئیات ساختاری RNA دو رشته‌ای و دسته‌بندی توالی‌ها برمبنای ویژگی‌های ساختارهای ثانویه آن‌ها بود. برای این منظور همردیفی بین ژنوم *A. indica* و تمام توالی‌های miR گیاهی به‌وسیله الگوریتم BLASTn انجام گرفت و تعداد ٨٠٢١٧ ناحیه در سراسر ژنوم *A. indica* با هومولوژی قابل قبول شناسایی شدند. پس از گسترش طولی نواحی شناسایی شده در هم‌ردیفی و حذف توالی‌هایی که در نواحی کدکننده ژنوم قرار داشتند، برای ٣٨٠٦٧ توالی باقی‌مانده پیش‌بینی ساختار ثانویه انجام گرفت. مقادیر مورد توافق برای ضوابط مربوط به ساختارهای ثانویه microRNA در فیلتر نرم‌افزار CTAnalyzer تعریف شد و این نرم‌افزار موفق به شناسایی و معرفی توالی‌های کاندید microRNA شد. بر اساس نتایج نرم‌افزار CTAnalyzer از کل ٥٦٦٧٥٤ ساختار ثانویه مورد بررسی برای شناسایی microRNAهای ژنوم *A. indica*، تعداد ٤٨٢ ساختار ثانویه (٠/٠٨ درصد) با دارا بودن تمام ویژگی‌های مورد تایید برای microRNAها شناسایی شدند.

واژه‌های کلیدی
بیوانفورماتیک
ساختار ثانویه
Azadirachta indica
CTAnalyzer
microRNA

[DOR: 20.1001.1.20084439.1400.16.4.6.0]

مقدمه

مولکول RNA سلولی، در بخش عمده‌ای از تاریخ زیست‌شناسی مولکولی تنها به‌عنوان واسطه‌ای بین DNA و پروتئین مورد توجه قرار گرفته است. کشف خاصیت کاتالیزوری RNA و توانایی آن در ایفای وظیفه‌ای جدید که تصور می‌شد منحصر به مولکول‌های پروتئین است، این دیدگاه را به‌کلی تغییر داد و جایگاه متمایزی برای مولکول RNA در دنیای زیست‌شناسی مولکولی ایجاد کرد (Mathews et al. 2010). عملکرد RNA مکمل[1]، به‌عنوان فاکتوری فعال در ممانعت از ترجمه RNAهای کدکننده، در اوایل دهه ۱۹۸۰ مورد توجه قرار گرفت (Izant and Weintraub 1984). همچنین fire et al. (1998) اثبات کردند که RNA دو رشته‌ای در گونه *C. elegans* در فرایند خاموشی ژن[2] نقش دارد (Fire et al. 1998). این یافته‌ها سنگ بنای مطالعات بعدی در خصوص نقش مولکول‌های RNA کوچک (sRNA) در تنظیم بیان ژن بودند. تا کنون انواع مختلفی از مولکول‌های RNA کوچک شناسایی شده است که microRNAها و RNAهای کوچک مداخله‌گر[3] از مهم‌ترین انواع آن‌ها هستند. این مولکول‌ها که دارای توالی‌هایی به‌طول تقریبی ۱۹ الی ۲۵ نوکلئوتید هستند، اغلب در نتیجه ایجاد برش بر روی مولکول‌های دو رشته‌ای بلندتر با ساختار سنجاق‌سری ایجاد می‌شوند (Knip et al. 2014). فرآیند خاموشی ژن که از مسیرهای تنظیم بیان ژن است، به‌واسطه اندرکنش RNAهای کوچک (miRNAs وsiRNAs) با پروتئین‌های آرگونات و تشکیل ساختاری تحت عنوان ریسک (RISC)[4] رخ می‌دهد. در این فرایند، انطباق مولکول RNA کوچک موجود در این مجموعه با قسمتی از توالی هدف، در انجام این وظیفه نقش کلیدی دارد. مکانیزم خاموشی ژن بر مبنای RNA، چه در مرحله پیش از رونویسی و چه در مرحله پس از رونویسی، در یوکاریوت‌ها دارای ماهیت حفاظت شده است (Chang et al. 2012; Axtell 2013). فرآیند خاموشی ژن با هدایت RNAهای کوچک[5] در هسته، از طریق ایجاد اختلال در فعالیت RNApol II (Allo and Kornblihtt 2010)، القاء تغییرات اپی‌ژنتیک بر روی هیستون‌ها (Castel and Martienssen 2013) و متیلاسیون DNA (Agrawal et al. 2003; Martínez de Alba et al. 2013) اتفاق می‌افتد. از طرف دیگر، خاموشی ژن پس از رونویسی با مداخله RNA[6] در سیتوزول از طریق شکافت در mRNA و یا از طریق ممانعت از ترجمه با هدایت مولکول‌های RNA کوچک صورت می‌گیرد (Agrawal et al. 2003). در تمامی مسیرهای فوق‌الذکر که تحت عنوان مداخله RNA (RNAi)[7] طبقه‌بندی می‌شوند (Knip et al. 2014) ساختار سنجاق‌سری RNA[8] نقش کلیدی ایفا می‌کند.

ساختار سنجاق‌سری یکی از ساختارهای مهم RNA بوده و در تاخوردگی مولکول RNA، اندرکنش‌های مرتبط با ریبوزوم، حفاظت از mRNA در مقابل عوامل تجزیه کننده، عملکرد به‌عنوان موتیف قابل شناسایی برای پروتئین‌های متصل شونده به RNA و عملکرد به‌عنوان سوبسترا در واکنش‌های آنزیمی نقش بنیادی دارد. در این ساختار که متشکل از یک ساقه دو رشته‌ای و یک حلقه انتهایی می‌باشد، نوکلئوتیدهای قسمت ساقه به‌طور کامل مکمل یکدیگر نبوده و اغلب در این ناحیه تعدادی عدم انطباق، بیرون زدگی‌های یک طرفه (Bulges) و حباب‌های داخلی (Internal loops) متقارن یا غیر متقارن مشاهده می‌شود (Tian et al. 2004; Svoboda and Di Cara 2006; Popenda et al. 2010). ساختارهای سنجاق‌سری ممکن است در نتیجه رونویسی از نواحی تکراری معکوس شده‌ی DNA توسط آنزیم‌های رونوشت بردار وابسته به DNA[9] ایجاد شوند. در مسیری دیگر، ممکن است الگوی قرارگرفتن ناحیه‌ی پیچ خورده‌ی RNA، برای سنتز ادامه‌ی رشته توسط آنزیم رونوشت بردار وابسته به RNA[10]، موجب تولید رشته‌ی مکمل و شکل‌گیری ساختار سنجاق‌سری شود. البته حالت دوم که رخدادی غیر متداول است، منجر به تولید دوبلکس‌های بلند RNA با انطباق بالا شده و منحصراً در برخی ویروس‌ها مشاهده می‌شود (Behrens et al. 1996).

[1] anti-sense RNA
[2] Gene silencing
[3] small-interfering RNAs
[4] RISC: RNA induced silencing complex
[5] sRNA-guided transcriptional Gene silencing
[6] PTGS/RNAi: post-transcriptional Gene silencing RNA interference
[7] RNAi: RNA interference
[8] RNA hairpin structure
[9] DNA-dependent RNA polymerase
[10] RNA-dependent RNA polymerase

[DOR: 20.1001.1.20084439.1400.16.4.6.0]

طبقه‌بندی ساختارهای سنجاق‌سری کار دشواری است؛ چرا که ساختارهایی از این دست به آسانی و با فراوانی بالا به‌وجود آمده و از جنبه‌های ساختاری متعددی مانند موقعیت بر روی رشته RNA، طول ساقه، بزرگی حلقه انتهایی، محل و موقعیت بیرون زدگی‌ها و در نهایت توالی نوکلئوتیدی با یکدیگر تفاوت دارند. از لحاظ عملکردی نیز ساختارهای سنجاق‌سری دارای قابلیت تنظیم بیان ژن به‌صورت سیس و ترنس هستند که در حالت سیس، ساختار تشکیل شده روی مولکول mRNA، صرفاً بیان همان مولکول را تحت تأثیر قرار می‌دهد؛ ولی در حالت ترنس، این ساختار در القای اثر کنترلی بر روی مولکول ژنتیکی دیگری نقش دارد. ساختارهای سنجاق‌سری به‌عنوان جایگاه اتصال برای بسیاری از پروتئین‌ها و به‌عنوان پیش ماده در بسیاری از واکنش‌های آنزیمی نقش داشته و بعضاً به‌صورت مستقیم عملکرد آنزیمی از خود نشان می‌دهند (Svoboda and Di Cara 2006). اولین مشاهدات مبنی بر مداخله مولکول RNA با ساختار سنجاق سری در تنظیم بیان ژن، به ساختار سنجاق سری ٦١ نوکلئوتیدی یا RNA تک رشته‌ای ٢٢ نوکلئوتیدی Lin-4 که یک توالی غیرقابل ترجمه است مربوط می‌شود که این مولکول RNA با اتصال به ناحیه 3′UTR موجب کاهش بیان ژن *lin-14* در *C. elegans* می‌شود (Lee et al. 1993; Wightman et al. 1993; Bartel 2004). علی‌رغم وجود یافته‌هایی از این دست، تشخیص نقش microRNAها به‌عنوان یکی از شایع‌ترین ابزارهای تنظیم بیان ژن در یوکاریوت‌ها طی فرایند RNAi و نقش ساختار سنجاق سری در بیوسنتز microRNA بیش از یک دهه به‌طول انجامید (Fire et al. 1998; Zamore et al. 2000). فرآیند تولید microRNA در هسته و با رونویسی microRNA اولیه توسط RNA pol-II آغاز می‌شود (Du and Zamore 2005; Ha and Kim 2014). شایان ذکر است با وجود شباهت‌های بسیار زیاد در مسیر بیوسنتز و مکانیزم‌های اثر مولکول‌های RNA کوچک در کنترل بیان ژن، هیچ‌یک از اجزای فرآیند در بین موجودات مختلف به‌خصوص گیاهان، در سطح توالی کاملاً حفاظت‌شده نبوده (Lagos-Quintana et al. 2001; Lee and Ambros 2001; Auyeung et al. 2013) و بیشترین سطح حفاظت شدگی در توالی miRNA بالغ دیده می‌شود (Reinhart et al. 2002). در تأیید این مطلب نتایج آزمایشی نشان داد که بیش از یک سوم microRNAهای بالغ در *C. elegans* با microRNAهای انسان هومولوگ هستند (Lim et al. 2003)؛ ولی با انتقال پیش‌ساز این microRNAهای هومولوگ به سلول انسانی، فرایند تبدیل این پیش‌سازها به microRNA بالغ صورت نمی‌گیرد (Auyeung et al. 2013). این مطلب، شاهدی بر لزوم توجه بیشتر بر روی ساختارهای فراتوالی در این‌گونه مطالعات است. اساسا مطابق با دستورالعمل ارائه شده توسط Axtell and Meyers (2018)، حفاظت شده بودن توالی جزو ضوابط و ملزومات شناسایی بیوانفورماتیکی ساختارهای مسیر بیوسنتز microRNA نیست؛ چرا که اولاً شواهدی بر این الزام وجود ندارد و دوم اینکه مولکول‌های RNA متفرقه فراوانی در سلول وجود دارد که علی‌رغم شباهت در توالی، در مسیر بیوسنتز microRNA قرار نمی‌گیرند (Axtell and Meyers 2018).

در سال‌های اخیر با توجه به سهولت به‌وجود آمده در توالی‌یابی اسیدهای نوکلئیک و وفور داده‌های NGS، رویکرد غیر آزمایشگاهی و به‌کارگیری توابع و الگوریتم‌های رایانه‌ای به‌منظور بررسی ساختار و پیش‌بینی عملکرد توالی‌های اسیدهای نوکلئیک به‌خصوص microRNA مورد اقبال گسترده محققین این زمینه قرار گرفته است. این قبیل مطالعات می‌تواند به‌عنوان مقدمه‌ای بر آزمایشات مولکولی، موجب صرفه‌جویی در زمان و هزینه شده و بر احتمال حصول نتایج مطلوب بیفزاید. در حالت کلی، شناسایی microRNA در موجودات زنده با استفاده از اطلاعات حاصل از توالی‌یابی به دو روش انجام می‌گیرد. روش اول که بر مبنای بررسی هومولوژی بین توالی‌های کاندید و miRهای شناسایی شده در سایر موجودات می‌باشد، از ماهیت حفاظت‌شدگی miRNA بهره جسته و با شناسایی این توالی‌ها در نواحی غیر کدکننده و همچنین شناسایی مکمل توالی‌های فوق الاشاره (به‌عنوان ناحیه هدف miRNA) در نواحی کدکننده ژنوم یا ESTهای موجود، اقدام به معرفی miRهای کاندید می‌نماید. بدیهی است این روش که بر مبنای هومولوژی با miRهای حفاظت‌شده درون گونه‌ای و بین گونه‌ای عمل می‌کند، منجر به شناسایی miRهای جدید نمی‌شوند (Mendes et al. 2009; Kleftogiannis et al. 2013). در روش دوم، الگوریتم‌های

machine learning با دریافت مجموعه‌ای از داده‌ها (بر اساس ویژگی‌های مربوط به miRهای تایید شده و همچنین اطلاعات مربوط به توالی‌هایی که علی‌رغم برخورداری از ساختارهای مشابه microRNA، در این گروه طبقه‌بندی نمی‌شوند)، قادر به شناسایی miRهای غیر هومولوگ به روش غیرمقایسه‌ای خواهند بود (Ni et al. 2010).

ویژگی‌های ساختاری و عملکردی miRهای گیاهی و جانوری با یکدیگر متفاوت بوده (Alptekin et al. 2017) و اغلب نرم‌افزارهای مربوط به شناسایی بیوانفورماتیکی miRها برای جانوران و به‌خصوص برای انسان طراحی شده است؛ که این امر متاثر از اهمیت و نقش RNAهای غیر کدکننده من‌جمله miRها در بیماری‌های انسانی می‌باشد (Esteller 2011). همچنین با توجه به تنوع بالا در طول و ساختار پیش‌ساز miRNA بین گونه‌های مختلف گیاهی و فراوانی توالی‌های تکراری به‌ویژه در ژنوم گیاهان اقتصادی نظیر غلات (Mehrotra and Goyal 2014)، عمده توجه محققین در شناسایی miRهای گیاهی بر روی ارتباط بین دو رشته miR و مکمل آن یعنی miR* (Reinhart et al. 2002) و همچنین ویژگی‌های ساختاری پیش‌ساز miRNA متمرکز است (Reinhart et al. 2002; Mendes et al. 2009; Axtell and Meyers 2018). البته این مطالب، مانند آن‌چه که در mir319 دیده می‌شود، وجود توالی حفاظت‌شده در سایر نواحی پیش‌ساز miRNA را نفی نمی‌کند. در mir319 یک توالی حفاظت‌شده در بالادست دوبلکس miR/miR* دیده می‌شود (Palatnik et al. 2003; Warthmann et al. 2008; Addo-Quaye et al. 2009; Bologna et al. 2009; Li et al. 2011; Sobkowiak et al. 2012) و بنابراین می‌توان انتظار داشت توالی‌های حفاظت شده در خارج از محدوده miR/miR* نیز وجود داشته باشد (Chorostecki et al. 2017).

با توجه به ماهیت حفاظت‌شده توالی microRNA، همچنین اهمیت ویژگی‌های ساختاری در رونوشت‌هایی که منجر به سنتز microRNA بالغ می‌شوند (pri-miRNA و pre-miRNA)، بررسی توالی و ساختار ثانویه قطعات مورد مطالعه، دو گام مقدماتی در شناسایی مولکول‌های microRNA است. انواع الگوریتم‌های بررسی هومولوژی و مهم‌ترین آن‌ها، الگوریتم BLASTn از مجموعه BLAST+ (Sayers et al. 2019) با تنظیمات مناسب برای توالی‌های کوتاه، برای بررسی هومولوژی بین قطعات مورد استفاده قرار می‌گیرد. تاکنون الگوریتم‌های متعددی به‌منظور پیش‌بینی ساختار ثانویه توالی‌های پلی‌نوکلئوتید طراحی شده و مورد استفاده قرار گرفته است که نرم‌افزار و وب سرویس‌های mFOLD (Zuker 2003)، RNAstracture (Reuter and Mathews 2010)، SPOR-RNA (Singh et al. 2019)، RNAsoft (Andronescu et al. 2003) و RNAfold از مجموعه Vienna RNA package (Lorenz et al. 2011) از این جمله هستند. صرف‌نظر از نحوه انجام محاسبات، این نرم‌افزارها حالات مختلف تشکیل ساختار دوم را برای توالی مورد نظر پیش‌بینی کرده و نتیجه محاسبات را به‌صورت فایل‌های تصویری و یا در قالب‌هایی نظیر جدول اتصال[1] یا نقطه-کمانک[2] در اختیار کاربر قرار می‌دهند. نظر به این‌که ساختار ثانویه مولکول‌های microRNA اولیه (primary microRNA) و پیش‌ساز microRNA (pre-miRNA) اهمیت ویژه و کلیدی در عملکرد این مولکول‌ها دارد، نرم‌افزارهای پیش‌بینی کننده ساختار ثانویه RNA، از مهم‌ترین ابزارهای مورد استفاده در شناسایی microRNA به‌صورت نرم‌افزاری (in silico) هستند.

تا به امروز نرم‌افزارها و الگوریتم‌های توانمند و کارآمدی هم برای تشخیص هومولوژی بین توالی‌ها و هم به‌منظور پیش‌بینی ساختارهای ثانویه، طراحی و در پژوهش‌ها و مطالعات متعددی به‌کار گرفته شده است. با استخراج مجموعه‌ای از توالی‌های اعتبار سنجی شده miRNA از منابعی نظیر miRBase (Ambros et al. 2003; Griffiths-Jones 2004; Griffiths-Jones et al. 2006; Griffiths-Jones et al. 2008; Meyers et al. 2008; Kozomara and Griffiths-Jones 2011; Kozomara and Griffiths-Jones 2014; Kozomara et al. 2019)، به‌عنوان query، و با در دست داشتن توالی‌های ژنوم یا ترانسکریپتوم موجودات مورد نظر به‌عنوان subject و تنظیمات مناسب برای شناسایی هومولوژی توالی‌های کوتاه، هم‌ردیفی‌های معتبر بین این دو مجموعه شناسایی شده و با گسترش این نواحی در دو جهت، رشته‌ای به‌دست می‌آید که در صورت تایید غیر کدکننده بودن آن می‌توان

[1] Connectivity table
[2] Dot-bracket file format

[DOR: 20.1001.1.20084439.1400.16.4.6.0]

آن را به‌عنوان توالی کاندید microRNA اولیه به الگوریتم‌های بررسی ساختار ثانویه معرفی کرد.

الگوریتم‌های پیش‌بینی کننده ساختار ثانویه، ساختار پیش‌بینی شده را در قالب فایل‌های متنی و تصویری در اختیار کاربر قرار می‌دهد. در مواردی که تعداد توالی‌های مورد نظر و همچنین تعداد ساختارهای پیش‌بینی‌شده محدود و انگشت‌شمار است، امکان بررسی بصری تصاویر شماتیک وجود داشته و با توجه به مشخص بودن ویژگی‌های مهم ساختار از قبیل عدم انطباق‌ها، حباب‌های یک‌طرفه (Asymmetric bulges)، حلقه‌های داخلی (Internal loops)، تعداد شاخه‌های جانبی و سایر ویژگی‌های ساختاری، تایید یا رد ساختار بسته به‌نظر محقق و ضوابط مورد قبول وی صورت می‌گیرد. حال آن‌که در عصر انبوه اطلاعات، که در اغلب مطالعات ژنتیک مولکولی حجم عظیمی از داده‌های حاصل از توالی یابی ژنوم و ترنسکریپتوم موجودات در دسترس بوده و مورد استفاده قرار می‌گیرد، تعداد توالی‌ها و به تبع آن تعداد ساختارهای پیش‌بینی شده ابعاد میلیونی به‌خود گرفته که این امر بهره‌گیری از نرم‌افزارها و الگوریتم‌های مناسب به‌منظور مدیریت و پالایش داده‌ها را اجتناب ناپذیر می‌نماید.

در این راستا در پژوهشی که به‌منظور شناسایی بیوانفورماتیکی miRNome گیاه *Azadirachta indica* صورت گرفت، حجم انبوه داده‌های تولید شده در مرحله‌ی پیش‌بینی ساختار ثانویه، موجب ایجاد انگیزه در مؤلفین اثر برای تهیه و تدوین ابزاری به منظور گروه‌بندی، تحلیل و پالایش این اطلاعات با محوریت ساختار سنجاق سری و به‌طور خاص microRNA اولیه شد. این الگوریتم که با هدف رقومی‌سازیِ ویژگی‌های مهم در ساختارهای دوم مولکول‌های RNA، با نام CTAnalyzer توسط مؤلفین اثر حاضر طراحی و معرفی شده است، با زبان برنامه نویسی PHP نوشته شده و در محیط لینوکس اجرا می‌شود. در این مقاله، با استفاده از توالی ژنوم گیاه *Azadirachta indica*، توانایی نرم‌افزار CTAnalyzer در شناسایی ویژگی‌های ساختاری در Hairpin RNA مورد ارزیابی قرار خواهد گرفت. در صورت مکاتبه با مولفین اثر، این نرم‌افزار با حفظ مالکیت معنوی در اختیار پژوهشگران این زمینه قرار داده خواهد شد.

مواد و روش‌ها

به‌منظور شناسایی توالی‌های مشابه با miRهای شناسایی شده، هم‌ردیفی بین ژنوم *A. indica* به‌عنوان subject و توالی‌های miR گیاهی منحصر به‌فرد به‌عنوان query با تنظیمات مناسب برای توالی‌های کوتاه[1]، به‌وسیله الگوریتم BLASTn از مجموعه standalone BLAST+ نسخه ۲.۱۱.۰ در سیستم عامل لینوکس انجام گرفت (Sayers et al. 2019). طی این فرایند، با درنظر گرفتن ضوابط مورد توافق در رابطه با هومولوژی microRNA در گیاهان (Axtell and Meyers 2018) تعداد ۸۰۲۱۷ ناحیه در سراسر ژنوم *A. indica* با هومولوژی قابل قبول شناسایی شد. پس از گسترش نواحی شناسایی شده در هم‌ردیفی به اندازه حداکثر ۲۰۰ نوکلئوتید در بالادست و پایین‌دست قطعه با استفاده از الگوریتم getFASTA از مجموعه BEDtools (Quinlan and Hall 2010) و حذف توالی‌هایی که در ناحیه کدکننده ژنوم قرار داشتند، در نهایت ۳۸۰۶۷ توالی باقی‌مانده جهت مرحله پیش‌بینی ساختار ثانویه آماده سازی شدند.

برای استفاده از نرم‌افزار CTAnalyzer، لازم است توالی‌های مورد نظر پیش از معرفی به نرم‌افزار پیش‌بینی ساختار ثانویه، با شیوه به‌خصوصی نام‌گذاری شوند. اطلاعات مندرج در نام توالی، توسط نرم‌افزار CTAnalyzer مورد استفاده قرار می‌گیرد. پس از انجام هم‌ردیفی بین ژنوم هدف و miRهای شناسایی شده در سایر گونه‌ها و گسترش دو جهته ناحیه شناسایی شده، هریک از توالی‌های چندصد نوکلئوتیدی قابل معرفی به نرم‌فزار پیش‌بینی ساختار ثانویه، دارای ناحیه‌ای با شباهت معنی‌داری با miRهای بالغ شناخته شده، با طول تقریبی ۱۶ الی ۲۸ نوکلئوتید بودند. در ساختار ثانویه‌ی پیش‌بینی شده، محل قرارگیری این قطعه کوتاه (ناحیه Hit در هم‌ردیفی‌های انجام شده) و همچنین وضعیت آن نسبت به توالی مکملش که در ادامه با نام *Hit شناخته خواهد شد، در بررسی ساختار ثانویه حائز اهمیت است. در صورت تایید اعتبار ساختار ثانویه مورد مطالعه به‌عنوان miRNA، دو ناحیه Hit و *Hit به miR و *miR تغییر نام خواهد داد. در روش نام گذاری قطعات کاندید، نام هر توالی پس از علامت **">"** از چهار

[1] word size= 7, mismatch penalty= -3, match reward= 2, gap opening penalty= -5, gap extension penalty= -2

جزء اصلی تشکیل می‌شود. این اجزا به‌ترتیب عبارت است از: نام کانتیگ یا کروموزوم، رشته‌ای از کانتیگ یا کروموزوم که ناحیه Hit روی آن شناسایی شده است (+ یا -)، شماره اولین و آخرین جایگاه رشته مورد نظر در کانتیگ یا کروموزوم (شماره جایگاه در رشته +) و شماره اولین و آخرین جایگاه ناحیه Hit در توالی مورد بررسی، که این اجزا با علامت "|" از یکدیگر تفکیک می‌شوند. به‌عنوان نمونه، یک توالی با نام >AMWY02010195.1|+|2847-3265|201-219، برشی از جایگاه ۲۸٤۷ الی ۳۲٦۵ از رشته مثبت در کانتیگ AMWY02010195.1 بوده و در این رشته ٤۱۹ نوکلئوتیدی، دامنه بین جایگاه ۲۰۱ و ۲۱۹ (ناحیه Hit) از نظر توالی دارای شباهت قابل قبول با یک microRNA در بانک اطلاعات miRNA سایر موجودات می‌باشد. پس از نام‌گذاری، کلیه توالی‌ها در یک فایل multi FASTA ذخیره شدند.

در مرحله بعد، بسته به‌نظر پژوهشگر، هر نرم‌افزاری که قابلیت ارائه ساختارهای دوم پیش‌بینی شده را در فرمت CT داشته باشد، قابل استفاده است. در این تحقیق از نسخه ۳/٦ نرم‌افزار mFold (Zuker et al. 1999; Zuker 2003) برای پیش‌بینی ساختارهای ثانویه استفاده شد. الگوریتم CTAnalyzer نیز برای خروجی همین نرم‌افزار بهینه‌سازی شده است. نرم‌افزار mFOLD تعداد n ساختار پیش‌بینی شده برای هر توالی را با درج اعداد ۱ الی n در انتهای نام توالی، در پوشه‌ای که هم‌نام با توالی مورد نظر است (با حذف علامت ">" و جایگزینی "|" با "_") ذخیره می‌نماید. این نرم‌افزار، پذیرنده فایل‌های FASTA بوده و در صورت معرفی فایل multi FASTA به نرم‌افزار، تنها اولین توالی توسط نرم‌افزار مورد تجزیه و تحلیل قرار می‌گیرد. برای رفع این مشکل، دستوری نوشته شد که با بازخوانی فایل multi FASTA، توالی‌ها را به‌صورت یک به یک به نرم‌افزار mFold معرفی کرده و فایل‌های خروجی مربوط به هر توالی را در پوشه‌های هم‌نام با توالی مورد نظر ذخیره نماید. با طی این مرحله در تحقیق حاضر برای ۳۸۰٦۷ توالی کاندید مورد نظر، در مجموع ٥٦٦۷٥٤ ساختار ثانویه پیش‌بینی و ذخیره شد. در سطر اول هر فایل CT تولید شده توسط نرم‌افزار mFold به‌ترتیب طول توالی، مقدار dG و نام توالی مندرج است. از سطر دوم، در شش ستون به‌ترتیب شماره جایگاه (در جهت ′۵ به ′۳ مطابق با فایل FASTA)، علامت اختصاری نوکلئوتید، شماره جایگاه منهای یک، شماره جایگاه به اضافه یک، شماره نوکلئوتیدی که با جایگاه مورد نظر جفت شده است و در نهایت شماره‌گذاری طبیعی جایگاه‌ها ثبت شده است. این اطلاعات به‌منظور تشخیص ویژگی‌های ساختار دوم پیش‌بینی شده مورد استفاده قرار می‌گیرد.

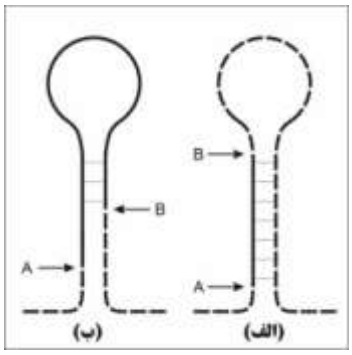
به‌منظور انجام آنالیز، پوشه اصلی حاوی تمام زیر پوشه‌ها و فایل‌های ساختارهای ثانویه پیش‌بینی شده در مرحله قبل، جهت بررسی و تحلیل به نرم‌افزار CTAnalyzer معرفی شد. در حالت کلی تحلیل وضعیت ساختار دوم با محوریت ناحیه Hit، بر پایه بررسی روند سلسله اعداد ستون پنجم (ستون اتصال) در مقابل اعداد متوالی ستون ششم (شماره هر جایگاه در توالی) استوار است. با توجه به اهمیت وضعیت متقابل ناحیه Hit و مکمل آن از نظر تعداد و پراکنش عدم انطباق‌ها، بیرون‌زدگی‌های یک طرفه، حلقه‌های داخلی، وضعیت ساختار ثانویه در محل برش توسط آنزیم، ابعاد و فاصله حلقه انتهایی از ناحیه Hit، طول ساختار دو رشته‌ای در پایین دست ناحیه Hit، طول دنباله‌های تک رشته‌ای در انتهای ساقه و نیز وجود یا عدم وجود شاخه‌های متعدد در ساختار ثانویه، هریک از این موارد به شرحی که در ادامه به آن پرداخته می‌شود توسط نرم‌افزار CTAnalyzer مورد شناسایی و بررسی قرار گرفتند.

نرم‌افزار CTAnalyzer برخی اطلاعات پایه از قبیل طول رشته، تعداد انواع نوکلئوتید، نسبت نوکلئوتیدهای G و C در ناحیه Hit و کل توالی و نیز مقدار محاسبه شده dG که اهمیت آن‌ها در منابع مختلف مورد اشاره قرار گرفته است (Jin et al. 2008) را به‌صورت مستقیم از هر فایل CT استخراج و در جدول نتایج ارائه می نماید. سایر شاخص‌های ساختاری مهم (Cuperus et al. 2011; Bologna et al. 2012; Bologna et al. 2013; Chorostecki et al. 2017) که طی بررسی ترتیب و توالی اعداد مربوط به هر جایگاه و رابطه متقابل بین نوکلئوتیدهای رشته RNA توسط الگوریتم CTAnalyzer شناسایی و محاسبه می‌شود، به شرح زیر می‌باشند:

بررسی استقلال توالی Hit از *Hit: در ساختار اولیه miRNA، توالی miR و *miR دو ناحیه مستقل در یک توالی واحد هستند

[DOR: 20.1001.1.20084439.1400.16.4.6.0]

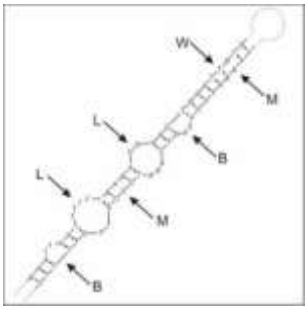
که با یک ساختار ساقه- حلقه انتهایی از یکدیگر جدا می‌شوند (Starega-Roslan et al. 2011; Zhu et al. 2011). در این نرم‌افزار، وجود همپوشانی بین ناحیه Hit و *Hit شناسایی می‌شود. همچنین در صورت استقلال، طول توالی بین این دو ناحیه محاسبه و ثبت می‌شود. دو حالت مربوط به استقلال و عدم استقلال ناحیه Hit و *Hit در شکل ۱ نشان داده شده است.



شکل ۱- حالت استقلال (الف) و عدم استقلال (ب) ناحیه Hit از *Hit. در هر دو شکل، ناحیه Hit با خطوط پیوسته و پیوند هیدروژنی بین نوکلئوتیدها با نقطه‌چین مشخص شده است. ابتدا و انتهای ناحیه Hit با حروف A و B مشخص شده است.

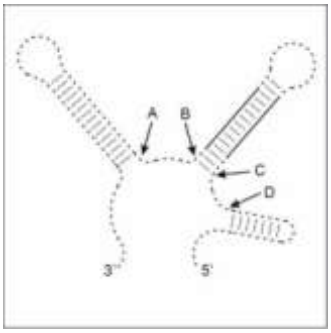
در اغلب موارد، توالی miR و *miR به‌طور کامل مکمل یکدیگر نیستند (Lee et al. 2015). تعداد و بزرگی نواحی غیرمکمل مشتمل بر عدم انطباق، بیرون‌زدگی و حلقه‌های داخلی، در ناحیه Hitتوسط نرم‌افزار شناسایی شده و در جدول نتایج درج می‌شود (شکل ۲). مطابق با ضوابط تعریف شده در شناسایی microRNA، در طول دوبلکس miR/*miR علاوه بر تعداد کلی جایگاه‌های جفت نشده، پراکندگی این جایگاه‌ها نیز حائز اهمیت است (Meyers et al. 2010; Lee et al. 2015). در نتیجه هر دو فاکتور برای هر توالی مشخص شده و در جدول نتایج ثبت می‌گردد.

نظر به اینکه جایگاه‌های مجاور با ابتدا و انتهای دوبلکس miR/*miR سوبسترای انواع آنزیم‌های RNAse III و کمپلکس‌های حاوی این آنزیم بوده (Zhu et al. 2011; Auyeung et al. 2013) و هرگونه عدم انطباق بین دو رشته RNA در این ناحیه می‌تواند منجر به جلوگیری از عملکرد صحیح آنزیم برشی گردد (Alptekin et al. 2017)، وضعیت شش جایگاه نوکلئوتیدی در هر یک از نواحی ابتدا و انتهای دوبلکس و شش جایگاه منتهی به ناحیه Hit در ناحیه پایه ساختار سنجاق سری از نظر وضعیت اتصال مورد بررسی قرار می‌گیرد.



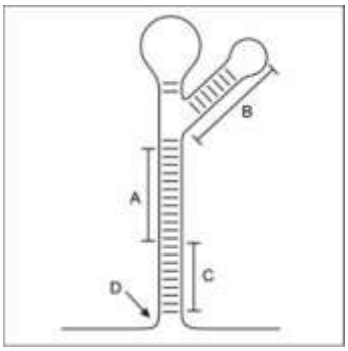
شکل ۲- انواع حالت‌های عدم برقراری رابطه مکملی بین نوکلئوتیدها در dsRNA. بزرگی و جایگاه عدم انطباق‌ها (M: mismatch)، بیرون زدگی‌های یک‌طرفه (B: bulge)، و حلقه‌های داخلی (L: loop) توسط نرم‌افزار CTAnalyzer شناسایی می‌گردد. بسته به تنظیمات نرم‌افزارهای پیش‌بینی کننده ساختار ثانویه، ممکن است جفت شدگی‌های نامتعارف (W: wobble base pairing) به‌عنوان match یا mismatch در نظر گرفته شود.

معمولاً ساختارهای ثانویه پیش‌بینی شده به‌صورت چند شاخه‌ای بوده و در چنین وضعیتی فقط شاخه حاوی توالی Hit مد نظر پژوهشگر است. نرم‌افزار CTAnalyzer ضمن مشخص کردن جایگاه آغازین و پایانی شاخه اصلی، با توجه به اهمیت وجود دو دنباله تک رشته‌ای قبل و بعد از ساختار سنجاق سری (Zeng and Cullen 2005; Han et al. 2006)، طول RNA تک رشته‌ای (ssRNA) در دو انتهای شاخه اصلی را نیز مشخص و در جدول نتایج منعکس می‌کند (شکل ۳).



شکل ۳- شناسایی شاخه اصلی در یک ساختار چند شاخه‌ای. در این تصویر ابتدا و انتهای شاخه اصلی با حروف B وC مشخص شده است. ناحیه Hit و مکمل آن با خطوط ممتد نمایش داده شده است. همچنین قطعات AB و CD به‌عنوان نواحی ssRNA قبل و بعد از شاخه اصلی، توسط نرم‌افزار CTAnalyzer شناسایی می‌شود.

در ساختار اغلب pre-miRNAهای تایید شده، وجود شاخه‌های جانبی و حلقه‌های انتهایی متعدد مشاهده نمی‌شود (Alptekin et al. 2017). نرم‌افزار CTAnalyzer پس از شناسایی شاخه اصلی به بررسی تعداد حلقه‌های انتهایی در این شاخه پرداخته و نتیجه را در جدول نتایج مندرج می‌نماید. در بیشتر مطالعات شاخه‌هایی که دارای بیش از یک حلقه انتهایی هستند از ادامه بررسی حذف می‌شوند (شکل ٤).



شکل ٤- ساختار دوم با شاخه اصلی منشعب. در این شکل ناحیه Hit (A)، شاخه فرعی (B)، ناحیه دو رشته‌ای پایه (C) و تقاطع RNA تک رشته‌ای-دو رشته‌ای (D) مشخص شده است. چنین ساختاری به‌واسطه وجود شاخه فرعی جانبی، فاقد قابلیت تبدیل به microRNA بالغ می‌باشد.

علاوه بر طول ناحیه پایه (فاصله بین ssRNA-dsRNA junction و دوبلکس miR/miR*) (Han et al. 2006; Chorostecki et al. 2017) وضعیت این منطقه از نظر انواع عدم انطباق حائز اهمیت است (Bologna et al. 2012). نرم‌افزار طراحی شده با بررسی این ناحیه موقعیت و بزرگی انواع عدم انطباق را شناسایی و در جدول نتایج ثبت می‌نماید.

همان‌طور که پیش‌تر اشاره شد برای پیش‌بینی pre-miRNA به فرادست و پایین دست ناحیه شناسایی شده دنباله‌هایی به طول بیش از ۱۰۰ نوکلئوتید افزوده می‌شود. پس از تشخیص قسمتی از این توالی چندصد نوکلئوتیدی که ساختار سنجاق سری مناسبی تشکیل داده و می‌تواند به‌عنوان pre-miRNA مد نظر قرار گیرد، حذف نواحی غیر درگیر به‌منظور محاسبه مجدد سطوح انرژی رویکردی منطقی به‌نظر می‌رسد (Alptekin et al. 2017). در این مرحله نرم‌افزار اولین جایگاه در شاخه اصلی را به‌عنوان نقطه مرجع در نظر گرفته و با گسترش دادن توالی از این جایگاه و جایگاه مکمل آن به سمت ابتدا و انتهای توالی، دامنه جدیدی از نوکلئوتیدها را استخراج می‌نماید. میزان گسترش توالی به طرفین توسط کاربر قابل تعریف است. در ادامه نام توالی جدید با توجه به جایگاه آن روی کروموزوم یا کانتیگ و همچنین با توجه به موقعیت ناحیه Hit در توالی جدید تعریف می‌شود.

این امکان برای نرم‌افزار CTAnalyzer در نظر گرفته شده است که با تعریف مقادیر و محدوده قابل پذیرش برای برخی از پارامترها و ضوابط کلیدی در شناسایی ساختارهای قابل قبول، فقط ساختارهای مورد تایید را در جدول نهایی نتایج منعکس نماید. این مزیت نه تنها موجب کاهش چشم‌گیر در حجم داده‌های خروجی می‌شود، بلکه به‌کارگیری ضوابط صحیح در تعریف فیلتر، این نرم‌افزار را به‌طور خاص در زمره نرم‌افزارهای توانمند در زمینه شناسایی microRNA قرار می‌دهد.

نتایج و بحث

برای تایید صلاحیت ساختارهای ثانویه به‌عنوان کاندیدای ایفای نقش به‌عنوان pri-microRNA، ویژگی‌های ساختاری در دو گروه طبقه‌بندی می‌شوند: گروه اول عبارت است از شاخص‌های تیپیک و الزامی برای ساختار سنجاق سری، که در صورت نقض هریک از آن‌ها، نیازی به انجام بررسی‌ها بیشتر و جزئی‌تر برای آن ساختار ثانویه وجود ندارد. در نتایج حاصل از مطالعه ساختارهای ثانویه استخراج شده از ژنوم *A. indica*، تعداد و درصد ساختارهای ثانویه پیش‌بینی شده که مطابق با تحلیل صورت گرفته توسط الگوریتم CTAnalyzer از ویژگی‌ها و شاخص‌های اولیه و اصلی ساختار سنجاق‌سری تبعیت ننمودند، در جدول ۱ نشان داده شده است.

مطابق با خروجی نرم‌افزار (جدول ۱)، مشاهده شد که در ٤٥٨٩ ساختار (۰/۸۱٪ از کل ساختارهای مورد بررسی)، ناحیه Hit در تشکیل dsRNA مشارکت ننموده است. در ۱۱۹۲٤۲ ساختار (۲۱/۰٤٪)، عدم استقلال Hit از *Hit (اشتراک توالی بین Hit و *Hit) مشاهده شد. به‌عبارتی در این ساختارهای ثانویه، بین برخی از نوکلئوتیدهای ناحیه Hit، جفت‌شدگی داخلی وجود دارد (مانند حالتی که در قسمت ب از شکل ۱ نشان داده شده است). در ۱۳۳٤۲ ساختار (۲/۳۵٪)، بین ناحیه Hit و *Hit هیچ نوکلئوتید دیگری وجود نداشته و در نتیجه ساختار فاقد ناحیه ساقه- حلقه

[DOR: 20.1001.1.20084439.1400.16.4.6.0]

انتهایی است. در ۱۲۶۱۷۹ ساختار (۲۲/۲۶٪)، ناحیه *Hit منقطع بوده و به‌عبارتی قسمتی از توالی مکمل Hit، مربوط به پایانه ′۵ و قسمت دیگر مربوط به پایانه ′۳ توالی می‌باشد. در ۶۷۰۹۹ ساختار (۱۱/۸۴٪)، بیش از یک ناحیه ساقه- حلقه انتهایی شناسایی شد. علی‌رغم وجود موارد استثنا، چنین وضعیتی در اغلب miRهای تایید شده، مشاهده نمی‌شود. همچنین در ۷۷۲۵۹ ساختار (۱۳/۶۳٪)، ناحیه بین Hit و *Hit فاقد ساقه شناسایی شد. در مجموع مطابق با خروجی نرم‌افزار CTAnalyzer (جدول ۱) مشاهده شد که ۴۰۳۲۰۹ ساختار از کل ۵۶۶۷۵۴ ساختار ثانویه پیش‌بینی شده توسط نرم‌افزار mFold (۷۱/۱۴ درصد)، حداقل از یکی از ویژگی‌های اصلی ساختار سنجاق سری برخوردار نبوده و نمی‌توانند به‌عنوان ساختار کاندید pri-microRNA مد نظر قرار گیرند و بنابراین از ادامه بررسی‌ها حذف شدند.

گروه دوم شاخص‌های مورد بررسی، بر جزئیات ساختار سنجاق سری تمرکز دارد. منطقه دوبلکس *Hit/Hit، نواحی بالادستی و پایین دستی دوبلکس، مواضع برش و دنباله‌های تک رشته‌ای بر روی ساختارهای انتخاب شده و باقی مانده از مرحله قبل برای ساختارهای ثانویه ژنوم گیاه *A. indica* توسط نرم‌افزار CTAnalyzer بررسی شد. برای بررسی وضعیت حاکم در ناحیه *Hit/Hit، ضوابط عنوان شده توسط Axtell and Meyers (2018) ملاک ارزیابی نرم‌افزار CTAnalyzer قرار گرفت (Axtell and Meyers 2018).

مطابق با این ضوابط، تعداد عدم انطباق مجاز بین Hit و *Hit پنج جایگاه است که از این تعداد، نهایتا سه جایگاه می‌تواند در بیرون زدگی‌های یک‌طرفه یا حلقه‌های داخلی قرار گیرد. همچنین مطابق با نظر برخی محققین، هیچ نوع عدم انطباق در محل برش آنزیم، مجاز نمی‌باشد (Alptekin et al. 2017). بنابراین مطابق با استانداردهای اشاره شده، پس از حذف ساختارهای نامناسب مطابق با جدول ۱، نتایج و اطلاعات آماری ساختارهای ثانویه‌ای که به‌واسطه دارا نبودن ضوابطی که در بالا به آن اشاره شد نامناسب شناخته شدند، در جدول ۲ نشان داده شده است.

جدول ۱- تعداد و درصد ساختارهای ثانویه پیش‌بینی شده که مطابق با تحلیل صورت گرفته توسط الگوریتم CTAnalyzer از شاخص‌های اولیه ساختار سنجاق سری تبعیت نکردند.

شاخص های پایه ای ساختار سنجاق سری	تعداد ساختارهای شناسایی شده	درصد از کل ساختارها
عدم شرکت توالی Hit در ناحیه دو رشته‌ای RNA	۴۵۸۹	٪ ۰/۸۱
عدم استقلال Hit از *Hit (جفت شدگی داخلی توالی Hit)	۱۱۹۲۴۲	٪ ۲۱/۰۴
عدم وجود توالی linker بین Hit و *Hit	۱۳۳۴۲	٪ ۲/۳۵
وجود انقطاع در توالی *Hit	۱۲۶۱۷۹	٪ ۲۲/۲۶
وجود بیش از یک شاخه انتهایی	۶۷۰۹۹	٪ ۱۱/۸۴
فقدان ناحیه دو رشته‌ای (ساقه) بین دوبلکس *Hit/Hit و حلقه انتهایی	۷۷۲۵۹	٪ ۱۳/۶۳
ساختارهای مشمول حداقل یکی از حالت‌های فوق	۴۰۳۲۰۹	٪ ۷۱/۱۴

جدول ۲- تعداد و درصد ساختارهای سنجاق سری که مطابق با ارزیابی صورت گرفته توسط الگوریتم CTAnalyzer از ضوابط مورد تایید microRNA در ناحیه *miR/miR تبعیت ننمودند.

وضعیت متناقض با ضوابط ناحیه *miR/miR	تعداد ساختارهای شناسایی شده	درصد از کل ساختارهای باقی‌مانده
شناسایی بیش از پنج عدم انطباق بین Hit و*Hit	۲۶۶۲۴	٪ ۱۶/۲۸
درگیری بیش از سه جایگاه در بیرون زدگی‌های یک‌طرفه	۲۴۹۹۲	٪ ۱۵/۲۸
درگیری بیش از سه جایگاه در حلقه‌های داخلی	۱۱۰۸۰۱	٪ ۶۷/۷۵
وجود حلقه داخلی، عدم انطباق یا بیرون زدگی در جایگاه برش آنزیم	۱۳۲۷۵۹	٪ ۸۱/۱۸
ساختارهای مشمول حداقل یکی از حالت‌های فوق	۱۵۷۹۰۸	٪ ۹۶/۵۵

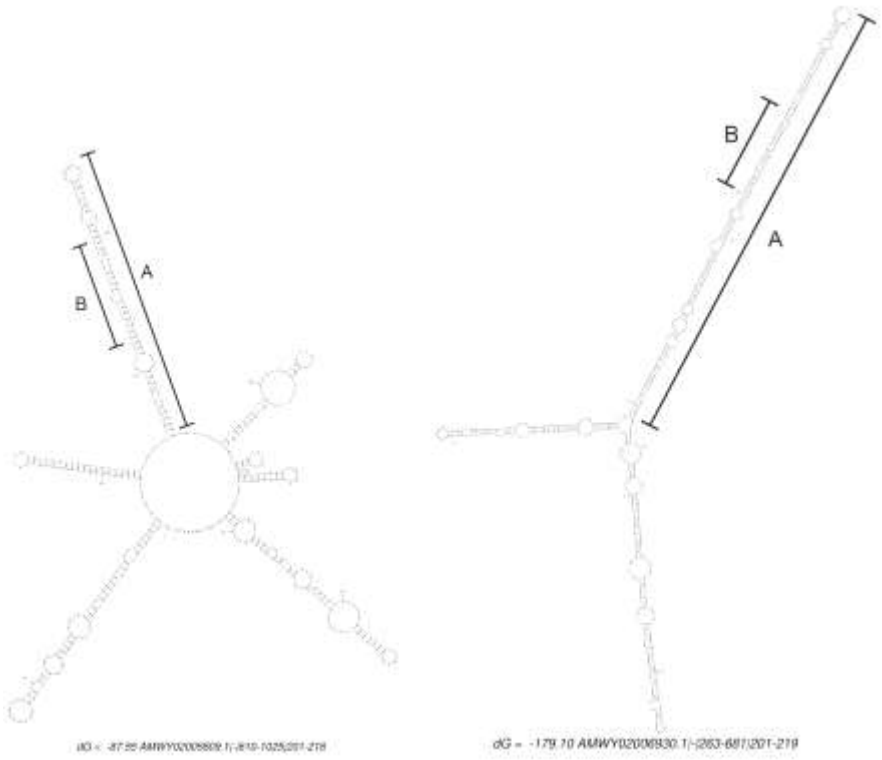
همان‌طور که در جدول ۲ مشاهده می‌شود در ۱۵۷۹۰۸ ساختار از ۱۶۳۵۴۵ ساختار باقی‌مانده مورد بررسی(۹۶/۵۵ درصد)، حداقل از یکی از ضوابط لازم برای ناحیه *miR/miR تبعیت نکردند؛ بدین شرح که وجود بیش از پنج عدم انطباق بین توالی Hit و*Hit در ۲۶۶۲۴ ساختار، درگیری بیش از سه جایگاه در بیرون زدگی‌های یک‌طرفه در ۲۴۹۹۲ ساختار، درگیری بیش از سه جایگاه در حلقه‌های داخلی در ۱۱۰۸۰۱ ساختار و وجود حلقه داخلی، عدم انطباق یا بیرون زدگی در جایگاه برش آنزیم در ۱۳۲۷۵۹ ساختار شناسایی شد.

با توجه به اهمیت وضعیت ساختار سنجاق سری در محل استقرار ریزپردازنده (ناحیه دو رشته‌ای در پایین دست دوبلکس *Hit/Hit)، طول این ناحیه و وضعیت جفت‌شدگی دو رشته در آن مورد بررسی قرار گرفت. وجود عدم انطباق وسیع و بیرون زدگی‌های بزرگ در این ناحیه، مانع استقرار ریزپردازنده می‌شود. مطابق با نظر Bologna et al. (2012) این ناحیه در miRهای جانوری bp ۱۱ طول داشته و به دو دنباله بلند ssRNA متصل می‌شود؛ حال آن‌که در miRهای گیاهی، طول این ناحیه حدود bp ۱۵بوده و در انتهای آن یک حلقه نسبتاً بزرگ قرار دارد (Bologna et al. 2012). این طول مطابق با نظر Unver et al. (2009) معادل bp ۱۷ می‌باشد. با در نظر گرفتن ضوابط فوق، ساختارهای ثانویه باقی‌مانده از نظر وضعیت محل استقرار ریزپردازنده، توسط نرم‌افزار CTAnalyzer مورد بررسی قرار گرفتند که نتایج حاصل در جدول ۳ گزارش شده است. در این مرحله از بین ۵۶۳۷ ساختار ثانویه باقی‌مانده از بررسی‌های مرحله اول و دوم ساختارها (جداول ۱ و۲)، وجود عوارض ممانعت کننده در محل برش (cut site) در ۴۰۳۴ ساختار، کوتاه‌تر بودن طول ناحیه دو رشته‌ای در پایین دست ناحیه *Hit/Hit نسبت به طول مورد نیاز جهت استقرار مجموعه ریزپردازنده در ۲۹۷۳ ساختار و وجود عوارض ممانعت کننده در محل استقرار ریزپردازنده در ۱۸۳۹ ساختار مشاهده شد. بنابراین مطابق با نتایج مندرج در جدول ۳ مشاهده شد که مجموعاً در ۵۱۵۵ ساختار از ۵۶۳۷ ساختار ثانویه باقی‌مانده (حدود ۹۱/۴۵ درصد)، در محل استقرار ریزپردازنده یا محل برش آنزیم، ممانعت فضایی (شامل عدم انطباق‌های متوالی، بیرون زدگی و حلقه داخلی با ابعاد بزرگ) وجود داشته و یا ناحیه dsRNA در پایین دست ناحیه *Hit/Hit از طول کافی برخوردار نبوده و به دلایل فوق جزو ساختارهای ثانویه مورد تائید برای pri-microRNA قرار نگرفته و حذف شدند. بررسی فایل گرافیکی (شکل ۵) چند نمونه از ساختارهای ثانویه ترسیم شده توسط نرم‌افزار mFOLD برای توالی‌های باقی‌مانده شناسایی شده توسط نرم‌افزار CTAnalyzer، نتایج به‌دست آمده از خروجی نرم‌افزار CTAnalyzer را تایید نمود. با درج مقادیر مورد توافق استاندارد در فیلتر نرم‌افزار CTAnalyzer برای ضوابطی که طی مراحل قبل مورد اشاره قرار گرفتند، مشاهده شد که خروجی نهایی نرم‌افزار با نتایجی که طی مراحل سه‌گانه قبل (جداول ۱، ۲ و۳) به‌دست آمد، به‌طور کامل منطبق بوده و الگوریتم CTAnalyzer با دریافت ضوابط مورد نظر در فیلتر نرم‌افزار ضمن بررسی کلیه ساختارهای ثانویه، موفق به شناسایی و معرفی توالی‌های مناسب شده است.

جدول ۳- نسبت ساختارهایی که علی‌رغم وضعیت مناسب در ناحیه *Hit/Hit، دارای موضع مناسب برای استقرار مجموعه ریزپردازنده نبودند.

وضعیت ممانعت کننده از استقرار ریز پردازنده	تعداد ساختارهای شناسایی شده	درصد از کل ساختارهای باقیمانده
وجود عوارض ممانعت کننده در محل برش ناحیه *Hit/Hit	۴۰۳۴	٪ ۷۱/۵۶
کوتاه بودن طول ناحیه دو رشته‌ای در پایین دست *Hit/Hit	۲۹۷۳	٪ ۵۲/۷۴
وجود عوارض ممانعت کننده در محل استقرار ریزپردازنده	۱۸۳۹	٪ ۳۲/۶۲
ساختارهای مشمول حداقل یکی از حالت‌های فوق	۵۱۵۵	٪ ۹۱/۴۵

[DOR: 20.1001.1.20084439.1400.16.4.6.0]



شکل ۵- دو نمونه از ساختارهای ثانویه ترسیم شده توسط نرم‌افزار mFOLD برای توالی‌های کاندید pri-microRNA حاوی شاخه اصلی ساختار (A) و ناحیه دوبلکس (B) Hit/Hit*

بدین ترتیب در نتایج خروجی نرم‌افزار CTAnalyzer با ضوابط در نظر گرفته شده، برای شناسایی microRNAهای ژنوم *A. indica* مشاهده شد که از بین کل ٥٦٦٧٥٤ ساختار ثانویه مورد بررسی تعداد ٥٦٦٢٧٢ ساختار ثانویه (حدود ۹۹/۹۱٪) فاقد ویژگی‌های مورد تایید برای microRNA اولیه بودند و ٤٨٢ ساختار ثانویه باقی‌مانده، دارای ویژگی‌های مورد تایید برای microRNA شناسایی شدند.

ویژگی‌های ساختار ثانویه در شناسایی و فراوری مولکول‌های microRNA نقشی اساسی و تعیین کننده دارد. چنان‌که به‌نظر می‌رسد علی‌رغم وجود ساختار اولیه مناسب در سطح توالی نوکلئوتیدی، فقدان ساختار ثانویه کارآمد، مانع قرارگیری رشته‌های RNA در مسیر بیوسنتز microRNA می‌شود. با در نظر گرفتن الگوهای شناخته شده در ساختار دوم microRNAهای اعتبار سنجی شده، امکان پیش‌بینی پتانسیل یک رشته RNA برای تبدیل به microRNA وجود دارد. نرم‌افزار CTAnalyzer با رقومی سازی ویژگی‌های تعداد بسیار زیاد ساختارهای ثانویه پیش‌بینی شده برای توالی‌های RNA، امکان گروه‌بندی آن‌ها را فراهم کرده و همچنین با حذف حجم انبوهی از ساختارهای فاقد معیارهای تعریف شده و در نتیجه انتخاب ساختارهای برتر، حجم داده‌های نیازمند اعتبارسنجی را به‌صورت محسوس تقلیل و احتمال شناسایی miRهای حقیقی را افزایش می‌دهد. بدیهی است در اینجا نیز مانند کلیه روش‌های بیوانفورماتیک شناسایی microRNA، لازم است اعتبارسنجی توالی‌های شناسایی شده طی روش‌های آزمایشگاهی نیز انجام گیرد.

منابع

Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R and Axtell MJ (2009) Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. RNA 15:2112-2121.

Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK and Mukherjee SK (2003) RNA

interference: biology, mechanism, and applications. Microbiology and Molecular Biology Reviews 67:657-685.
Alló M and Kornblihtt AR (2010) Gene Silencing: Small RNAs Control RNA Polymerase II Elongation. Current Biology 20:704-707.
Alptekin B, Akpinar BA and Budak H (2017) A Comprehensive Prescription for Plant miRNA Identification. Frontiers in plant science 7:2058.
Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G and Tuschl T (2003) A uniform system for microRNA annotation. RNA 9:277-279.
Andronescu M, Aguirre-Hernández R, Condon A and Hoos H H (2003) RNAsoft: A suite of RNA secondary structure prediction and design software tools. Nucleic Acids Research 31:3416-3422.
Auyeung VC, Ulitsky I, McGeary SE and Bartel DP (2013) Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell 152:844-858.
Axtell MJ (2013) Classification and Comparison of Small RNAs from Plants. Annual Review of Plant Biology 64:137-159.
Axtell MJ and Meyers BC (2018) Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. The Plant Cell 30:272-284.
Bartel D (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116:281-297.
Behrens SE, Tomei L and De Francesco R (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. The EMBO Journal 15:12-22.
Bologna NG, Mateos JL, Bresso EG and Palatnik JF (2009) A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. The EMBO Journal 28:3646-3656.
Bologna NG, Schapire AL and Palatnik JF (2012) Processing of plant microRNA precursors. Brifings in functional genomics 12:37-45.
Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC and Palatnik JF (2013) Multiple RNA recognition patterns during microRNA biogenesis in plants. Genome research 23:1675-1689.
Castel S E and Martienssen R A (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature Reviews Genetics 14:100-112.
Chang S-S, Zhang Z and Liu Y (2012) RNA Interference Pathways in Fungi: Mechanisms and Functions. Annual Review of Microbiology 66:305-323.
Chorostecki U, Moro B, Rojas A, Debernardi J, Schapire A, Notredame C and Palatnik J (2017) Evolutionary Footprints Reveal Insights into Plant MicroRNA Biogenesis. The Plant cell 29:1248-1261.
Cuperus JT, Fahlgren N and Carrington JC (2011) Evolution and functional diversification of MIRNA genes. The Plant Cell 23:431-442.
Du T and Zamore PD (2005) microPrimer: the biogenesis and function of microRNA. Development 132:4645-4652.
Esteller M (2011) Non-coding RNAs in human disease. Nature Reviews Genetics 12:861-874.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391:806-811.
Griffiths-Jones S (2004) The microRNA Registry. Nucleic Acids Research 32:109-111.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A and Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Research 34:D140-D144.
Griffiths-Jones S, Saini HK, van Dongen S and Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Research 36:D154-D158.
Ha M and Kim VN (2014) Regulation of microRNA biogenesis. Nature Reviews Molecular Cell Biology 15:509-524.
Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT and Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125:887-901.
Izant JG and Weintraub H (1984) Inhibition of thymidine kinase gene expression by anti-sense RNA: A molecular approach to genetic analysis. Cell 36:1007-1015.
Jin W, Li N, Zhang B, Wu F, Li W, Guo A and Deng Z (2008) Identification and verification of microRNA in wheat (*Triticum aestivum*). Journal of Plant Research 121:351-355.
Kleftogiannis D, Korfiati A, Theofilatos K, Likothanassis S, Tsakalidis A and Mavroudi S (2013) Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role. Journal of Biomedical Informatics 46:563-573.
Knip M, Constantin ME and Thordal-Christensen H (2014) Trans-kingdom Cross-Talk: Small RNAs on the Move. PLoS Genetic 10:e1004602.
Kozomara A, Birgaoanu M and Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Research 47:D155-D162.
Kozomara A and Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39:D152-D157.
Kozomara A and Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research 42:D68-D73.
Lagos-Quintana M, Rauhut R, Lendeckel W and Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294:853-858.
Lee R C and Ambros V (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science 294:862-864.
Lee RC, Feinbaum RL and Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843-854.
Lee W C, Lu S H, Lu M H, Yang C J, Wu S H and Chen H M (2015) Asymmetric bulges and mismatches determine 20-nt microRNA formation in plants. RNA biology 12:1054-1066.

Li Y, Li C, Ding G and Jin Y (2011) Evolution of MIR159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. BMC evolutionary biology 11:122.
Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP (2003) The microRNAs of *Caenorhabditis elegans*. Genes and Development 17:991-1008.
Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF and Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms for Molecular Biology 6:26.
Martínez de Alba AE, Elvira-Matelot E and Vaucheret H (2013) Gene silencing in plants: A diversity of pathways. Biochimica et Biophysica Acta 1829:1300-1308.
Mathews D, Moss W and Turner D (2010) Folding and Finding RNA Secondary Structure. Cold Spring Harbor perspectives in biology 2:a003665.
Mehrotra S and Goyal V (2014) Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics, Proteomics and Bioinformatics 12:164-171.
Mendes ND, Freitas AT and Sagot MF (2009) Current tools for the identification of miRNA genes and their targets. Nucleic Acids Research 37:2419-2433.
Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D and Zhu J-K (2008) Criteria for Annotation of Plant MicroRNAs. The Plant Cell 20:3186-3190.
Meyers BC, Simon SA and Zhai J (2010) MicroRNA Processing: Battle of the Bulge. Current Biology 20:68-70.
Ni M, Shu W, Bo X, Wang S and Li S (2010) Correlation between sequence conservation and structural thermodynamics of microRNA precursors from human, mouse, and chicken genomes. BMC evolutionary biology 10:329.
Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC and Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature 425:257-263.
Popenda M, Szachniuk M, Blazewicz M, Wasik S, Burke EK, Blazewicz J and Adamiak RW (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. BMC bioinformatics 11:231.
Quinlan AR and Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) 26:841-842.
Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. GENES & DEVELOPMENT 16:1616-1626.
Reuter JS and Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC bioinformatics 11: 129.
Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD and Ostell J (2019) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 47:D23-D28.
Singh J, Hanson J, Paliwal K and Zhou Y (2019) RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nature Communications 10:5407.
Sobkowiak L, Karlowski W, Jarmolowski A and Szweykowska-Kulinska Z (2012) Non-Canonical Processing of Arabidopsis pri-miR319a/b/c Generates Additional microRNAs to Target One RAP2.12 mRNA Isoform. Frontiers in plant science 3:46.
Starega-Roslan J, Koscianska E, Kozlowski P and Krzyzosiak WJ (2011) The role of the precursor structure in the biogenesis of microRNA. Cellular and Molecular Life Sciences 68:2859-2871.
Svoboda P and Di Cara A (2006) Hairpin RNA: a secondary structure of primary importance. Cellular and Molecular Life Sciences 63:901-908.
Tian B, Bevilacqua PC, Diegelman-Parente A and Mathews MB (2004) The double-stranded-RNA-binding motif: interference and much more. Nature Reviews Molecular Cell Biology 5:1013-1023.
Unver T, Namuth-Covert DM and Budak H (2009) Review of current methodological approaches for characterizing microRNAs in plants. International Journal of Plant Genomics 262463.
Warthmann N, Das S, Lanz C and Weigel D (2008) Comparative analysis of the MIR319a microRNA locus in Arabidopsis and related Brassicaceae. Molecular biology and evolution 25:892-902.
Wightman B, Ha I and Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 75:855-862.
Zamore PD, Tuschl T, Sharp PA and Bartel DP (2000) RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell 101:25-33.
Zeng Y and Cullen BR (2005) Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. The Journal of biological chemistry 280:27595-27603.
Zhu S, Jiang Q, Wang G, Liu B, Teng M and Wang Y (2011) Chromatin structure characteristics of pre-miRNA genomic sequences. BMC genomics 12:329.
Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31:3406-3415.
Zuker M, Mathews DH and Turner DH (1999) Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. In RNA Biochemistry and Biotechnology, J. Barciszewski and B.F.C. Clark, eds., NATO ASI Series, Kluwer Academic Publishers, Dordrecht, NL 11-43.