

«مقاله پژوهشی»

سازوکارهای تکاملی زمینه‌ساز تنوع متابولیت‌های ثانویه سه گونه براسیکا با استفاده از تحلیل مقایسه‌ای این‌سیلیکو ژن‌های مرتبط

شادی حیدری^{۱*}، پیوند حیدری^۱

۱. دکتری، گروه اصلاح نباتات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

(تاریخ دریافت: ۱۴۰۰/۰۶/۲۱ - تاریخ پذیرش: ۱۴۰۰/۱۰/۲۹)

Evolutionary mechanisms underlying secondary metabolite diversity of the three Brassica species using insilico comparative analysis of the related genes

Shadi Heidari^{1*}, Peivand Heidari¹

1. Ph. D., Department of Plant Breeding, Science and Research Branch, Islamic Azad University, Tehran, Iran

(Received: Sep. 12, 2021 - Accepted: Jan. 11, 2022)

Abstract

Brassica napus field plant, as an important oilseed, has undergone extensive genome reconstruction after interspecies hybridization of its ancestors. To elucidate the evolutionary mechanisms underlying the diversity of secondary metabolites, insilico comparative analysis of different genes between three Brassica species was performed. After assembling the preliminary EST sequence of libraries using EGAssembler software, the contigs were analyzed by X-blast using CLC Protein Workbench software against non-redundant proteins databank. IDEG6 software and Audic-Claverie statistics were used to determine the differential expression of genes. To identify orthologs and paralogs, the Ensembl Plants website and CLUSTALW were used for a pairwise alignment for each pair of proteins. The discovery of the DNA motif is a first step in many systems to study gene function, so the MEME website and STAMP webtool were used to explore the DNA binding motif and determine the similarity of the motif sequences of the paralogs. The results showed a significant difference between 18 genes in the functional groups of secondary metabolism and transcriptional regulation. Most of the genes involved in the glucosinolate diversity in *B. napus* have ortholog genes in the ancestral species and Arabidopsis, which have diverged during evolutionary processes. While most transcriptional regulatory genes, including *MYB28* and *bHLH*, have paralog genes that have been functionally altered within *B. napus* as a result of duplication and mutation following changes in allopolyploidy. The ancestral genome of *B. napus* provides valuable resources for insilico analysis in understanding the genetic consequences of polyploidy, evolution and breeding of *B. napus*.

Keywords: Allopolyploidy, *Brassica napus*, Comparative analysis of insilico, EST, Ortholog groups and paralog genes.

چکیده

گیاه زراعی *Brassica napus* به‌عنوان یک دانه روغنی مهم، پس از هیبریداسیون بین گونه‌های اجدادش تحت بازسازی گسترده ژنوم قرار گرفته است. به‌منظور تبیین ساز و کارهای تکاملی زمینه‌ساز تنوع متابولیت‌های ثانویه، تحلیل مقایسه‌ای این‌سیلیکو ژن‌های افتراقی بین سه گونه براسیکا انجام شد. بعد از هم‌گذاری توالی اولیه EST کتابخانه‌ها با استفاده از نرم‌افزار EGAssembler، کانتیگ‌ها به‌وسیله جستجوگر بلاست X توسط نرم‌افزار CLC Protein Workbench در مقابل پروتئین‌های غیر تکراری بانک ژن واکاوی شدند. نرم‌افزار IDEG6 و آماره Audic-Claverie برای تعیین بیان افتراقی ژن‌ها استفاده شد. برای شناسایی ارتولوگ‌ها و پارالوگ‌ها، از تارنمای Ensembl Plants استفاده شد و هم‌ردیفی دو به دو برای هر جفت پروتئین توسط CLUSTALW انجام شد. کشف موتیف DNA یک گام اولیه در بسیاری از سیستم‌ها برای مطالعه عملکرد ژن است. بنابراین وب‌سایت MEME و وب‌ابزار STAMP برای کاوش موتیف اتصال به DNA و تعیین شباهت توالی‌های موتیف پارالوگ‌ها استفاده شد. نتایج، تفاوت معنی‌داری را بین ۱۸ ژن درگروه‌های کارکردی متابولیسم ثانویه و تنظیم رونویسی نشان داد. اکثر ژن‌های دخیل در تنوع گلوکوزینولات‌ها در *B. napus* دارای ژن‌های ارتولوگ در گونه‌های اجدادی و آرابیدوپسیس بودند که طی فرایندهای تکاملی واگرا شده‌اند. درحالی‌که بیشتر ژن‌های تنظیم رونویسی شامل *MYB28* و *bHLH* دارای ژن‌های پارالوگ بودند که در درون گونه *B. napus* و در نتیجه تکثیر و جهش، به دنبال تغییرات حاصل از آلوپلی‌پلوئیدی تغییر عملکرد یافته‌اند. ژنوم اجداد *B. napus* منابع ارزشمندی برای تحلیل این‌سیلیکو در درک پیامدهای ژنتیکی پلی‌پلوئیدی، تکامل و اصلاح *B. napus* فراهم می‌کند.

واژه‌های کلیدی: آلوپلی‌پلوئیدی، تحلیل مقایسه‌ای این‌سیلیکو،

گروه‌های ارتولوگ و ژن‌های پارالوگ، *Brassica napus*, EST.

مقدمه

علاقه روزافزون به‌نژادگرها به *B. napus* ناشی از استفاده گسترده از آن برای تولید روغن خوراکی، خوراک دام و همچنین برای مصارف صنعتی است. در ارقام دو صفر (۰۰) کلزا، به دنبال تغییر ژنتیکی، اسید اروسیک از محتوای روغن بذر حذف شده و مقدار گلوکوزینولات در بذرها به‌طور قابل توجهی کاهش یافته است، بنابراین روغن و وعده غذایی حاصل از آن منابع ایمن در خوراک انسان و دام است (سازمان خواربار و کشاورزی ملل متحد، ۲۰۲۰). این دستکاری‌های ژنتیکی برای دستیابی به ویژگی مفید در ارقام دو صفر به‌طور قابل توجهی تنوع ژنتیکی کلزا را کاهش داده است. بنابراین، در خانواده براسیکا به‌منظور ایجاد تغییرات ژنتیکی جهت استفاده از متابولیت‌های ثانویه آن به‌عنوان ترکیبات زیست فعال در مصارف صنعتی به منابع جدید نیاز است. یکی از راه‌های گسترش منابع ژنتیکی، به‌دست آوردن آلپلی‌پلوئیدهای مصنوعی به‌وسیله تلاقی با گونه‌های مولد *Brassica oleracea* (وود و تاکبایاشی و بارکر، ۲۰۰۹). *B. napus* دومین دانه روغنی مهم در جهان یک گیاه آلوتراپلوئید ($2n=4x=38$, AACC) و مهم‌ترین گونه زراعی جنس براسیکا می‌باشد که از هیبریداسیون اتفاقی و دو برابر شدن طبیعی کروموزوم‌ها بین *B. rapa* (AA, $2n=2x=20$) و *B. oleracea* (CC, $2n=2x=18$) حاصل و با تغییرات پیچیده محیطی روبرو شده است. این تغییرات، در نتیجه انتخاب طبیعی، در ژنوم آن هارمزگذاری شده و سرخ‌هایی برای واگرایی ژنتیکی آن از اجدادش فراهم می‌کند (وود و تاکبایاشی و بارکر، ۲۰۰۹). مقایسه ژنوم کلزا با اجداد آن و استنباط ترکیبات موجودی ژن در شناسایی تغییرات کارکردی ژنوم بین گونه‌ها در طی تکامل مؤثر می‌باشد تا بدین وسیله بتوان از آنها در گسترش تنوع ژنتیکی کلزا اقدام نمود. وجود تنوع در ارقام و گونه‌ها نه تنها روند

انتخاب را تسریع می‌کند بلکه باعث افزایش کارایی گزینش صفات مطلوب می‌شود (حیدری و عزیزی نژاد و حق پرست، ۲۰۱۹). پلی‌پلوئیدی پدیده‌ای رایج در تکامل گیاهان گل‌دهنده است که منجر به تنوع زیستی و شکل‌گیری سریع گونه‌های جدید می‌شود. ارتولوگ‌ها به‌عنوان ژن‌هایی در گونه‌های مختلف در نظر گرفته می‌شوند که فقط از طریق حوادث گونه‌زایی تکامل یافته‌اند و فرض بر این است که ارتولوگ‌ها عملکردهای بیولوژیک یکسانی را در گونه‌های مختلف دارند. پارالوگ‌ها ضمن اینکه در یک گونه با رویدادهای تکثیر ایجاد می‌شوند، عملکردهای جدید را تکامل می‌دهند. شناسایی ارتولوگ‌ها شجره‌نامه ژن‌ها را برای بررسی نیروها و ساز و کارهای روند تکاملی و ایجاد گروه‌هایی از ژن‌ها با عملکردهای بیولوژیک یکسان مشخص می‌کند (کنین و موشگیان و بورک، ۱۹۹۶). در این پژوهش به‌منظور تبیین سازوکارهای تکاملی زمینه‌ساز تنوع متابولیت‌های ثانویه، تحلیل مقایسه‌ای این‌سیلیکو ژن‌های با بیان افتراقی بین سه گونه براسیکا و بررسی روابط تکاملی ژن‌های همولوگ انجام شد.

روش مقایسه‌ای تجزیه و تحلیلی EST^۱ ابزاری مؤثر، کارآمد، نسبتاً ارزان و سریع است که می‌تواند در بررسی تکامل واگرا برای تشخیص صفات توسط توالی‌های پروتئینی که از دو یا چند ژن همولوگ مشتق شده‌اند به کار برده شود. توالی‌های EST، توالی‌های کوتاه و ویرایش نشده‌ای هستند که به صورت تصادفی از انتهای ۳' یا ۵' یک کلون cDNA انتخاب شده و درک درستی از ناحیه پویا رونویسی در گیاهان فراهم می‌کند. در حال حاضر، از روش تجزیه و تحلیل EST برای تجزیه الگوی بیان تعداد بسیار زیادی ژن به‌طور همزمان در مراحل تکوینی و تکاملی مختلف استفاده

1. Expressed Sequence Tags

پایگاه داده‌ها به‌طور کارآمد توسط محققین مورد بررسی قرار می‌گیرند (نرینکس و لونیسن، ۲۰۰۵). از نرم‌افزار EGassembler برای هم‌گذاری توالی cDNA به‌صورت خوشه‌ای استفاده شد (<http://egassembler.hgc.jp>). در مرحله پیش پردازش، خالص‌سازی توالی‌های EST شامل پاکسازی توالی، پوشاندن تکرار، پوشاندن وکتور، پوشاندن اندامک و هم‌گذاری توالی با درصد تطابق $N \geq 95$ cutoff انجام شد. در مرحله هم‌گذاری، توالی‌های EST در کانتینگ‌ها شامل دو یا چند EST و سینگلتون شامل فقط یک EST دسته‌بندی شدند (مسعودی نژاد و تونومورا و کاواشیما، ۲۰۰۶). برای یافتن شباهت‌های سه کتابخانه، تمام توالی‌های EST توسط نرم‌افزار Egassembler هم‌گذاری شد. سپس همه کانتینگ‌ها با استفاده از X-blast توسط نرم‌افزار CLC bio در برابر پایگاه داده پروتئین‌های غیر تکراری با E-value $\leq 10^{-5}$ مورد تجزیه و تحلیل قرار گرفتند (مک‌گینیس و مدن، ۲۰۰۴). برای شناسایی ژن‌ها با بیان افتراقی در کتابخانه‌ها، از نرم‌افزار IDEG6 و آماره Audic-Claverie (<http://telethon.bio.unipd.it/ideg6>) استفاده شد (رومولدی و بورتولوزی و دانیلی، ۲۰۰۳). توالی‌های مربوط به کانتینگ و سینگلتون هر کتابخانه توسط برنامه X-blast در برابر منبع اطلاعاتی آرآبیدوپسیس بارگیری شده از پایگاه داده TAIR (<ftp://ftp.arabidopsis.org>) مورد تجزیه و تحلیل قرار گرفت (باسل و گلب و مارکز، ۲۰۱۱). برای دسته‌بندی کاتالوگ‌های عملکردی، ابزار طبقه‌بندی مقایسه‌ای و کاربردی GoMapMan (<http://www.gomapman.org>) استفاده شد (رامسک و بابلر و روتر، ۲۰۱۴). از خروجی‌های Mapman برای توصیف کاتالوگ‌های مختلف در کتابخانه‌ها استفاده شده است که می‌توانند کاتالوگ‌های کاربردی را در چندین آزمایش نمونه کشف کنند. از نرم‌افزار IDEG6 و آزمون Audic-Claverie برای یافتن

می‌شود (واسیلف و لونیسن و آتاناسوف، ۲۰۰۵). تاون و چونگ و مایتی، (۲۰۰۶) با استفاده از ژنومیکس مقایسه‌ای و تحلیل این‌سلیکو، از دست‌دادگی و پراکندگی ژن پس از پلی‌پلوئیدی را در *B. oleracea* و آرآبیدوپسیس آشکار کردند و با استفاده از تفسیر توالی، ژن‌های حفظ شده در قطعات ژنوم *B. oleracea* را شناسایی کردند و نشان دادند که قطعات ژنومی براسیکای سه برابر شده، پس از واگرایی تبار آرآبیدوپسیس و براسیکا از یک جد مشترک جدا شدند. حیدری و ملکی زنجانی و حیدری، (۲۰۱۲) با استفاده از تجزیه و تحلیل EST داده‌های جمع‌آوری شده از پایگاه داده دانشگاه هاروارد، الگوی بیان ژن‌های مربوط به تغییر شرایط محیطی را در چند گونه گیاهی بررسی کردند و از نرم‌افزار IDEG6 برای شناسایی ژن‌هایی با بیان افتراقی بین کتابخانه‌ها استفاده کردند.

مواد و روش‌ها

داده‌های اولیه EST برگ و بذر سه کتابخانه شامل *B. napus* واریته pabularia (BN238-Na) با EST ۹۷۶۸، *B. rapa* (BR304-Ra) با EST ۹۰۲۷ و *B. oleracea* (BO51-OI) با EST ۷۹۲۱ مورد تجزیه و تحلیل قرار گرفت. داده‌ها از پایگاه داده دانشگاه هاروارد با آدرس اینترنتی (<http://compbio.dfci.harvard.edu/tgi>) جمع‌آوری شد. گیاهان هر سه گونه در شرایط نرمال (بدون اعمال تنش) و تحت تیمارهای استاندارد در مزرعه رشد کردند. غلاف‌ها، ۱۹ روز بعد از گرده افشانی (DAP) جمع‌آوری شدند. بذرهای نابالغ و برگ‌های جوان برای ساخت کتابخانه‌های cDNA استفاده شدند. کتابخانه‌های زیست محاسباتی و ژنومیکس عملکردی پروژه‌های شاخص ژن‌ها در طیف گسترده‌ای از ارگانسیم‌ها هستند و داده‌های زیستی موجود در این

MEME داندود شد و برای شناسایی شباهت توالی‌های موتیف پارالوگ‌ها، از وب ابزار STAMP (<http://www.benoslab.pitt.edu/stamp>) استفاده شد. هنگامی که در STAMP بیش از یک موتیف وارد شود، هم‌ردیفی‌های چندگانه، نمایه‌های اتصال خانوادگی و درختان شباهت ایجاد می‌کند. این رویکردها، مطالعه‌های تکاملی روی مجموعه‌ای از موتیف‌های مرتبط و ماژول‌های تنظیمی و همچنین تشریح شباهت‌ها و افزونگی‌ها در موتیف‌های ورودی را امکان‌پذیر می‌کند (ماهونی و بنوس، ۲۰۰۷).

نتایج و بحث

نتایج هم‌گذاری توالی‌های EST شامل تعداد کانتیگ‌ها و سینگلتن‌های به‌دست آمده با میانگین طول EST ۸۴۷-۵۷۱ از سه کتابخانه *B. napus*، *B. rapa* و *B. oleracea* در جدول ۱ آمده است. ۲۴۵۷ (۲۵/۱۵ درصد)، ۱۸۴۵ (۲۰/۴۴ درصد) و ۱۸۹۲ (۲۳/۸۹ درصد) در کتابخانه *B. rapa*، *B. napus* و *B. oleracea* سهم خیلی ضعیفی در همولوژی با توالی‌های موجود در پایگاه داده‌ها داشتند ($E \geq 1 * 10^{-5}$ value) و یا هیچ توالی مشابهی با آن‌ها وجود نداشت. در حالی که ۷۳۱۱، ۷۱۸۲ و ۶۰۲۹ EST به ترتیب همولوژی بالا یا متوسطی را نشان دادند ($E < 1 * 10^{-5}$ value). بر پایه نتایج تطبیق و تفسیر توالی، توالی‌هایی که در نشان دادن همولوژی معنی‌دار برای هر پروتئین در پایگاه داده‌های عمومی شکست خوردند، کاندیدای مناسبی برای ژن‌های جدید هستند. مقایسه بیان ژنو ژنومیکس کارکردی *B. napus*، *B. rapa* و *B. oleracea* با استفاده از تجزیه و تحلیل منابع EST این سه گونه نشان‌دهنده تفاوت معنی‌دار ۱۸ ژن دخیل در گروه کارکردی متابولیت ثانویه و تنظیم رونویسی مرتبط با آن‌ها بودند (جدول ۲).

کاتالوگ‌های افتراقی عملکردی در سه کتابخانه استفاده شد. توالی‌های شناسایی شده به سه دسته هستی‌شناسی ژن (GO)^۱ (بیولوژیک، سلولی و مولکولی) طبقه‌بندی شدند. برای شناسایی ژن‌های ارتولوگ با عملکرد یکسان در میان گونه‌ها از وب سایت Ensembl Plants (<http://plants.ensembl.org>) بر اساس روابط فیلوژنی استفاده شد. این سایت، بر اهمیت استفاده از هم‌ردیفی توالی جهانی در شناسایی بهترین هیت دو جهت (BBH)^۲ تأکید می‌کند و به دلیل اشتراک زمین‌های مشترک پروتئین، احتمال مثبت بودن کاذب BBH را کاهش می‌دهد. به‌منظور شناسایی جفت‌های پارالوگ، برای هر جفت پروتئین (با استفاده از CLUSTALW) یک هم‌ردیفی دو به دو انجام شد. پروتئین‌ها تنها در صورتی پارالوگ در نظر گرفته شدند که e-value هم‌ردیفی جفتی $< 1e^{-20}$ و حداقل ۵۰ درصد باقی‌مانده‌ها هم‌ردیف بودند (پرز-برکوف و ماکینو و مک لیزگت، ۲۰۱۰). کشف موتیف DNA یک گام اولیه در بسیاری از سیستم‌ها برای مطالعه عملکرد ژن است و نقش حیاتی در شناسایی مکان‌های اتصال فاکتور رونویسی (TFBSs)^۳ دارد که به یادگیری مکانیسم‌های تنظیم بیان ژن کمک می‌کند. برای کاوش موتیف اتصال به DNA پارالوگ‌ها، از تارنمای MEME (<https://www.meme-suite.org/>) استفاده شد که یک مدل احتمالی به نام ماتریس وزن اختصاصی- موقعیت (PSWM)^۴ یا ماتریس موتیف ایجاد می‌کند که توزیع بازها را برای هر موقعیت در (TFBS) مشخص می‌کند تا موتیف‌ها را از غیرموتیف‌ها متمایز کند و به پارامترهای جستجوی کمی نیاز دارد (سین‌ها و تومپا، ۲۰۰۳). ۱۰ موتیف شناسایی شده با فرمت ماتریس احتمال^۵ از سایت

1. Gene Ontology
2. Bidirectional best hits
3. Transcription Factor Binding Sites
4. Position-specific weight matrix
5. Probability matrix

جدول ۱. تعداد کانتینگ و سینگلتن‌های هر کتابخانه بر اساس خروجی نرم افزار E.Gassembler.

<i>B. oleracea</i>		<i>B. rapa</i>		<i>B. napus</i>		کتابخانه‌ها
برگ	بذر	برگ	بذر	برگ	بذر	
۹۹۸	۶۹۲۳	۱۴۹۳	۷۵۳۴	۱۸۷۳	۷۸۹۵	تعداد کل توالی‌های EST
۹۸	۹۱۲	۱۳۶	۹۴۸	۳۲۴	۱۰۶۱	تعداد کانتینگ‌ها
۴۵۶	۳۸۹۶	۶۹۷	۴۳۶۲	۵۴۶	۴۶۱۳	تعداد EST در کانتینگ
۵۴۲	۳۰۲۷	۷۹۶	۳۱۷۲	۱۳۲۷	۳۲۸۲	تعداد سینگلتن
۱۷	۲۱۴	۳۲	۱۵۶	۴۱	۱۹۸	کانتینگ‌هایی که hit مشخصی ندارند
۱۳۴	۳۷۵	۱۴۷	۳۲۶	۱۲۳	۶۵۴	سینگلتن‌هایی که hit مشخصی ندارند

جدول ۲. تفسیر و سطح بیان EST ها با اختلاف معنی‌دار بین کتابخانه‌های *B. rapa*, *B. napus* و *B. oleracea* و تعداد ESTها برای هر کتابخانه.

p-value	<i>B. oleracea</i>	<i>B. napus</i>	<i>B. rapa</i>	فرایندهای سلولی	اجزای سلولی	Putative identity	فرآیندها	متابولیت‌های ثانویه
۳E ^{-۷۵}	۳۳ ^۰	۱۳ ^۰	۰	فاکتور رونویسی	هسته	MYB28	AG	تعداد توالی‌های EST در تنوع گلوکوزینولات
۵E ^{-۱۲}	۱۵ ^۰	۱۴ ^۰	۰	فاکتور رونویسی	هسته	MYB29	AG	
۲E ^{-۱۷}	۱۱	۸	۲۵ ^۰	فاکتور رونویسی	هسته	MYB34	IG	
۷E ^{-۸۲}	۱۴ ^۰	۱۶ ^۰	۲	طول‌سازی زنجیره آمینو اسید	کلروپلاست	MAM1	AG	
۱E ^{-۶۹}	۱۰ ^۰	۱۳ ^۰	۴	تشکیل ساختار هسته	شبکه آندوپلاسمی	CYP79F1	AG	
۲E ^{-۱۹}	۱۹ ^۰	۵	۱۴ ^۰	تشکیل ساختار هسته	غشا	ST5a	IG (گلوکوبراسین)	
۳E ^{-۱۷}	۱۵ ^۰	۰	۱۷ ^۰	تشکیل ساختار هسته	غشا	CYP81F4	IG (گلوکوبراسین)	
۳E ^{-۲۰}	۱۱	۴	۱۵ ^۰	تشکیل ساختار هسته	غشا	CYP81F2	IG (۴-هیدروکسی گلوکوبراسین)	
۳E ^{-۸}	۲۸ ^۰	۷	۳	تغییرات زنجیره جانبی	-	FMOGS-OX1	AG (گلوکورافانین)	
۱E ^{-۲۵}	۹	۳۴ ^۰	۱۱	تغییرات زنجیره جانبی	-	AOP2	AG (گلوکوتانین)	
۲E ^{-۹}	۲۸	۲۹	۳۹ ^۰	تغییرات زنجیره جانبی	سیتوزول، واکوئل	GSL-OH	AG (پروگوتیرین)	
۱E ^{-۱۲}	۸ ^۰	۴	۲	نقل و انتقال	غشا	GTR	نقل و انتقال گلوکوزینولات	
	۱۹۱	۱۴۰	۱۳۲					تعداد توالی‌های EST دخیل در تنوع گلوکوزینولات
۲E ^{-۱۶}	۰	۶ ^۰	۰	فاکتور رونویسی	هسته	MYC3		تعداد توالی‌های EST دخیل در تنوع ترکیبات فنلی
۸E ^{-۷}	۰	۴ ^۰	۰	فاکتور رونویسی	هسته	MYC4		
۵E ^{-۱۹}	۹	۳۳ ^۰	۲۱ ^۰	فرآیند بیوسنتزی فلاونوئید	غشا	CHI	کوئرستین	
۲E ^{-۲۶}	۷	۲۰ ^۰	۱۸ ^۰	فرآیند بیوسنتزی فلاونوئید	شبکه آندوپلاسمی	Fe2OG dioxygenase	کامپفرول	
۲E ^{-۵}	۱	۱۳ ^۰	۱۰ ^۰	فرآیند متابولیک فنیل پروپانوئید	غشا		اسیدهای هیدروکسی سینامیک	
۵E ^{-۲۲}	۲	۱۱ ^۰	۴	فرآیند بیوسنتزی لیگنان	پلاسمودسما	C4H PLR	لیگنان	
	۱۹	۷۷	۵۳					تعداد توالی‌های EST دخیل در تنوع ترکیبات فنلی

AGs، گلیکوزینولات‌های آلیفاتیک، IGs، گلیکوزینولات‌های ایندولیک. p-value به آماره Audic-Claverie برای بیان متفاوت اشاره دارد.

آلیفاتیک، در *B. rapa* مربوط به گلوکوزینولات ایندولیک و در *B. oleracea* هر دو رونوشت ژن‌های دخیل در بیوسنتز گلوکوزینولات آلیفاتیک و ایندولیک شایع بودند (جدول ۲).

ژن‌های دخیل در بیوسنتز گلوکوزینولات گلوکوزینولات‌ها گروه متفاوتی از متابولیت‌های ثانویه غنی از نیتروژن و گوگرد هستند که از اسیدهای آمینه و قندها به دست می‌آیند (هایز و کلهر و اگلستون، ۲۰۰۸). گلوکوزینولات‌ها به سه گروه، آلیفاتیک

ژن‌های افتراقی در گروه کارکردی متابولیسم ثانویه در دو زیرگروه اصلی گلوکوزینولات‌ها و ترکیبات فنلی (فلاونوئیدها و فنیل پروپانوئیدها) قرار گرفتند. بیشترین میزان رونوشت‌های بیان شده متعلق به زیرگروه کارکردی گلوکوزینولات‌ها به ترتیب مربوط به *B. oleracea*، *B. napus* و *B. rapa* بود. رونوشت‌های شایع در بیوسنتز گلوکوزینولات‌ها (GSLs) در *B. napus* مربوط به گلوکوزینولات

MYB51 و *MYB122* برای ایندول گلوکوزینولات تعیین می‌شود (سوندری و بورو و رو، ۲۰۱۰). گزارش‌های اخیر تایید کرده‌اند که ژن‌های *AtMYB28* و *AtMYB29* در آراییدوپسیس به‌عنوان تنظیم‌کننده اصلی رونویسی بیوستنز آلیفاتیکی گلوکوزینولات عمل می‌کنند (گیگولاشویلی و انگکوئیست ویاتوسویچ، ۲۰۰۷). هارپر و تریک و هیگینز، (۲۰۱۲) در تجزیه و تحلیل ترنسکرپتوم ارتباطی، فاکتور رونویسی *MYB28* را که در بیوستنز آلیفاتیکی *GSL* مشارکت دارد در *B. napus* شناسایی نمودند. از ژن‌های درگیر در متابولیسم طولیل شدن زنجیره جانبی *GSL* که تفاوت معنی‌داری در ۳ کتابخانه نشان داد شامل (*MAM*) بود که در *B. napus* و *B. oleracea* اختلاف معنی‌داری با کتابخانه دیگر نشان داد. *GSL* آلیفاتیکی دارای زنجیره‌های جانبی مختلف با طول‌های مختلف است. یکی از مراحل این کشیدگی توسط ژن *MAM* حاصل می‌شود و تراکم آلدول که باعث افزایش طول مولکول آغازی می‌شود را کاتالیز می‌کند که در تشکیل گلوکوزینولات‌ها ایندول اتفاق نمی‌افتد. کومار و لی و آگوستین، (۲۰۱۹) تعیین کردند که جهش از دست دادن عملکرد ژن *MAM* با محتوای کم گلوکوزینولات و بیان بیش از حد، منجر به افزایش کل *GSL* آلیفاتیکی در لاین‌های تراریخته آراییدوپسیس می‌شود. در ادامه‌ی مرحله طولیل‌سازی اسیدهای آمینه کشیده، مرحله تشکیل ساختار گلوکوزینولات هسته‌ای با عملکرد ژن‌های *CYP* اتفاق می‌افتد. ژن‌ها با بیان افتراقی دخیل در تشکیل ساختار هسته‌ای *GSL* در متابولیسم هر دو گروه گلوکوزینولات آلیفاتیکی و ایندولیک تفاوت معنی‌داری در ۳ کتابخانه نشان دادند. ژن‌های درگیر در متابولیسم تشکیل ساختار هسته‌ای گلوکوزینولات‌ها آلیفاتیکی شامل *CYP79F1*

مشتق شده از *Met*، *Ala*، *Leu*، *Ile* یا *Val*، ایندولیک مشتق شده از *Trp* و آروماتیک مشتق شده از *Phe* یا *Tyr* طبقه‌بندی می‌شوند (کلیبنشتاین و کرویمن و براون، ۲۰۰۱). محصولات مهم براسیکا از نظر اقتصادی و غذایی منابع غنی *GSL* به‌ویژه گلوکوزینولات آلیفاتیکی هستند. ترکیبات تولید شده توسط مسیر بیوستنز *GSL* اخیراً به‌دلیل فعالیت‌های مفید مختلف از جمله فعالیت ضد سرطان و آنتی‌اکسیدانی، دفاع در برابر آفات و عوامل بیماری‌زا در گیاهان و در برخی دیگر به‌منظور حذف خواص ضد تغذیه‌ای و گواتروژنیک در کنجاله دانه مورد توجه بالایی از تحقیقات قرار گرفته‌اند (یونکورا و سایتو، ۲۰۰۹). بنابراین، درک از تنظیم بیوستنز *GSL* اطلاعات مفیدی برای مطالعه متابولیسم ثانویه گیاه و در نتیجه بهبود ارزش محصولات کشاورزی فراهم می‌کند و اهداف بالقوه برای دستکاری ژنتیکی این محصول را نیز مشخص می‌کند. مسیرهای بیوستنز *GSL* شامل سه مرحله مستقل، مرحله طولیل شدن زنجیره، تشکیل یک ساختار *GSL* اصلی و اصلاح ثانویه است. ژن‌های درگیر در متابولیسم گلوکوزینولات‌ها که تفاوت معنی‌داری در سه کتابخانه نشان داد به پنج دسته فاکتور رونویسی (سه ژن)، طولیل شدن زنجیره جانبی (یک ژن)، تشکیل ساختار هسته‌ای (چهار ژن)، اصلاح زنجیره جانبی (سه ژن) و ترنسپورتهای گلوکوزینولات (یک ژن) تقسیم شدند. از فاکتور رونویسی دخیل در بیوستنز گلوکوزینولات‌ها که اختلاف بیان نشان دادند شامل *MYB28*، *MYB29* و *MYB34* بودند. رونوشت‌های *MYB28* و *MYB29* در کتابخانه *B. napus* و *B. oleracea* نسبت به *B. rapa* افزایش بیان نشان دادند. رونوشت *MYB34* در کتابخانه *B. rapa* افزایش بیان نشان دادند. سطح و مشخصات گلوکوزینولات‌ها در آراییدوپسیس برای گلوکوزینولات آلیفاتیکی با فاکتورهای رونویسی مانند *MYB28*، *MYB29* و *MYB76* و توسط *MYB34*

افزایش بیان داشتند و در *B. rapa* ۴-هیدروکسی گلوکوبراسیسین، گلوکوبراسیسین و نئوگلوکوبراسیسین و در *B. oleracea* ژن‌های دخیل در مسیر بیوسنتز گلوکورافانین و گلوکوبراسیسین افزایش بیان داشتند. نوگروهو و هان و پرویتاساری، (۲۰۲۰) رابطه بین تولید ترکیبات آلیفاتیک GSL و مشخصات بیان ژن‌های مرتبط با GSL را به این صورت نشان دادند که گلوکورافازاتین و گلوکورافانین در *Raphanus sativus* غالب بودند، درحالی که گلوکوناپین، گلوکوبراسیکاناپین و گلوکورافانین در *Brassica chinensis* فراوان بود. *B. rapa* حاوی سطح متوسطی از تمام گلوکوزینولات‌ها مورد مطالعه بود، که نشان‌دهنده وراثت از *R. sativus* و *B. chinensis* است. الگوی بیان ژن، تغییرات اساسی در زنجیره جانبی، به ویژه *GRS1*، *FMOgs-ox1* و *AOP2* در میان این گیاهان را نشان داد. افزون بر این، در مطالعه حاضر، افزایش بیان رونوشت (*GTRs*)^۱ در *B. oleracea* نسبت به دو کتابخانه دیگر مشاهده شد که احتمال دارد با افزایش رونوشت گلیکوزینولات‌ها در این کتابخانه همبستگی دارد. مطالعه‌های ژنتیکی نشان داده است که تنوع گلوکوزینولات در محصولات دانه‌های روغنی براسیکا از الگوی پیچیده وراثت پیروی می‌کند و توسط چندین مکان کنترل می‌شود (بیشتر و گوپتا و رامچاری، ۲۰۰۹). روشن کردن اساس مولکولی تکامل ژن‌های کدکننده گلوکوزینولات در گونه‌های براسیکا، در مقایسه با آرآبیدوپسیس توسط پلی‌پلوئیدی ذاتی و معماری ژنومی انواع مختلف براسیکاسه پیچیده شده است. گونه‌های براسیکا به‌عنوان مثال *B. rapa* پس از تقسیم آن‌ها از آرآبیدوپسیس، یک واقعه سه برابر سازی ژنوم را تجربه کردند و پس از آن تکثیر ژنو بازآرایی

(اختلاف بیان در *B. napus* و *B. oleracea* با *B. rapa*) و در تشکیل ساختار هسته‌ای گلوکوزینولات‌ها ایندولیک شامل *CYP81F4* و *ST5a* (اختلاف بیان در *B. oleracea* و *B. rapa* با *B. napus*) و *CYP81F2* (اختلاف بیان در *B. rapa* با دو کتابخانه دیگر) بودند. سیتوکروم P450 (*CYP79*) اسید آمینه‌ها را به آلدوکسیم تبدیل می‌کند، سپس توسط سایر پروتئین‌های *CYP* به فرم‌های دیگر اکسید می‌شود. *CYP81F4*، *CYP81F2* و *ST5a* به‌ترتیب در تولید نئوگلوکوبراسیسین، ۴-هیدروکسی گلوکوبراسیسین و گلوکوبراسیسین نقش دارند. پس از تشکیل ساختار، زنجیره‌های جانبی به‌طور عمده توسط واکنش‌های اکسیژناسیون و هیدروکسیلاسیون اصلاح می‌شوند. ژن‌های درگیر در اصلاح زنجیره جانبی GSL که تفاوت معنی‌داری در ۳ کتابخانه نشان داد شامل *FMOGS-OX*، *AOP2* و *GSL-OH* بودند. اکسیژناسیون S گلوکوزینولات آلیفاتیک یک تغییر متداول است که توسط فلاوین مونوکسیژناز *FMOGS-OX* انجام می‌شود (اختلاف بیان در *B. oleracea* با دو کتابخانه دیگر) و در ایجاد گلوکورافانین نقش دارد. GSL گلوکوناپین توسط دی‌اکسیژنازهای وابسته به ۲ اکسوجلوتارات (*AOPs*) از گلوکوزینولات‌ها اکسیژن دار S مانند گلوکورافانین تولید می‌شوند. ژن *AOP2* در کتابخانه *B. napus* افزایش بیان معنی‌دار نشان داد. ژن *GSL-OH* که برای هیدروکسیلاسیون بوتنیل گلوکوزینولات به ۲-هیدروکسی بوتنیل گلوکوزینولات (پروگویتترین) لازم است در کتابخانه *B. rapa* نسبت به دو کتابخانه دیگر افزایش بیان نشان داد. بنابراین، نتایج نشان می‌دهد درحالی‌که افزایش رونوشت‌های درگیر در مسیر بیوسنتز گلوکوزینولات‌های آلیفاتیک پروگویتترین در هر سه کتابخانه حفاظت شده است، در *B. napus* ژن‌های درگیر در بیوسنتز گلوکوناپین

1. Glucosinolate transporter

برخی از ژن‌ها آنقدر شدید و همیشگی است که وجود همولوگ‌های متعدد از هر ژن کاندید گلوکوزینولات در ژن‌های آلوتتراپلوئید براسیکا انتظار می‌رود. مرحله طویل سازی بیوستنز گلوکوزینولات مشتق شده از Met برای تولید تنوع گلوکوزینولات و اتصال متابولیسم اولیه و تخصصی حیاتی است. بنابراین، ژنوم‌های گونه‌های حاصل شامل چندین هومولوگ از هر ژن بیوستنز گلوکوزینولات است که نشان می‌دهد فعل و انفعالات ژنتیکی می‌تواند تنوع گلوکوزینولات را شکل دهد و تکامل مسیرهای متابولیکی اختصاصی از متابولیسم اصلی، زمینه را برای تنوع شیمیایی محصولات طبیعی و فعالیت‌های بیولوژیک ناشی از آن‌ها فراهم می‌کند. این حقایق اهمیت شناسایی ارتولوژیک عملکردگر را برجسته می‌کند. از این رو، مطالعه گروه‌های ارتولوگ عملاً مطالعه چگونگی تکامل، حفظ و نگهداری عملکردهای ژن‌ها و یا واگرایی بیشتر آن‌هاست. از طرفی تجزیه و تحلیل هم‌ردیفی به‌منظور شناسایی ژن‌های پارالوگ، دو ژن *MYB28* مشخص نمود که در *B. napus* بیان و تنظیم شدند (شکل ۱).

کروموزومی در مقیاس بزرگ در طول دیپلوئیداسیون انجام شده است (لیزاک و کوچ و پسینکا، ۲۰۰۵). از این رو ژنوم‌های آلوتتراپلوئید *B. napus* و *B. juncea* حتی پیچیده‌تر هستند و شناخته شده است که تا شش بخش محافظت‌شده از ژنوم اجدادی را حفظ می‌کند (پنجابی و جاگانات و بیشت، ۲۰۰۸). نتایج بررسی روابط تکاملی ژن‌های ارتولوگ از طریق سایت Ensembl Plants نشان داد که اکثر ژن‌های درگیر در بیوستنز گلوکوزینولات‌ها در *B. napus* دارای ارتولوگ‌هایی در سایر گونه‌های اجدادی و آرابیدوپسیس هستند که احتمال دارد به دنبال واگرایی ژنوم از گونه‌های اجدادی و آرابیدوپسیس رخ داده است (جدول ۳).

ارتولوگ ژن‌های درگیر در بیوستنز گلوکوزینولات‌ها از *B. rapa* (زانگ و کیم و کیم، ۲۰۰۹) و *B. juncea* (شارما و موخوپادهایای و گوپتا، ۲۰۱۶) در آرابیدوپسیس نشان داده شده است. حفظ یک ژن در همه گونه‌های موجود از اجداد مشترک با تشابه توالی بالا که از طریق مقایسه ژنوم شناسایی می‌شود، احتمال دارد نشان‌دهنده این است که فشار انتخابی خاص روی

جدول ۳. تجزیه و تحلیل گروه‌های ژنی ارتولوگ در ۳ کتابخانه براسیکا و آرابیدوپسیس.

ژن‌های ارتولوگ	<i>B. napus</i>	<i>B. rapa</i>	<i>B. oleracea</i>	آرابیدوپسیس
MYB28	BnaC09g05300D	-	Bo9g014610	AT5G61420
MYB29	-	-	Bo9g175680	AT5G07690
MYB34	BnaA03g39790D	Bra035954	Bo2g161170	AT5G60890
MAM1	BnaC02g41790D	Bra029355	Bo2g161100	-
CYP79F1	BnaC05g12520D BnaA06g11000D BnaA06g11010D	Bra026058	Bo5g021810	AT1G16400 AT1G16410
CYP81F4	BnaC01g01530D	Bra011759	Bo1g004730	AT4G37410
CYP79A2	BnaC09g50060D	Bra009100	Bo02541s010	AT5G05260
FMOGS-OX1	BnaA10g25130D	-	Bo9g177260	AT5G35917
AOP2	BnaC09g13360D	Bra027035	Bo9g037180	AT1G65860
GSL-OH	BnaC07g36630D BnaA04g17890D BnaA04g17880D BnaC04g41490D BnaA03g15200D	Bra038780	Bo7g107340	AT4G21200
		Bra022920	Bo4g173530	AT2G25450


```

CLUSTAL W (1.81) multiple sequence alignment

CDX87215/1-139      MSRKPCCVGE-GLKKGAWTTEEDKLLISYIHEHGEGGWRDIPQK-AGLKRCGKSCRLRWT
CDX90721/1-224      MGRSPCCEKAH-TNKGAWTKEEDQRLIDYIRIHGEGSWRSIP-KSAGLLRCGKSCRLRWT
                    *.*.***          :*****.***:*.**.*: *****.*:* * * * * *****

CDX87215/1-139      NYLKPEIKRGEFSSEEEQIIIMLHASRGNKWSVIARHLPRRTDNEIKNYWNTHLKKRLIE
CDX90721/1-224      NYLRPDLKRGNFSDDEEDKVI IKLHSLGKNSLIAGRLPGRTDNEIKNYWNTHIKRLLS
                    ***.*:***.*.*.*:*** ** : *****.* * * * * *****.*:***:..

CDX87215/1-139      QGTFVTHKPLASNTN-----PSYCI-----
CDX90721/1-224      HGIDPQTH--R-----PIKDSIA----VENSFTVRPKTESSGNVASTSGTTTDEDIRQ
                    :* ** **

CDX87215/1-139      -----
CDX90721/1-224      NGDCYNDNSGEKEVNLDTLGSSTSLVRSGGSGNGSSADSKPWWDFVMAARLSLL

```

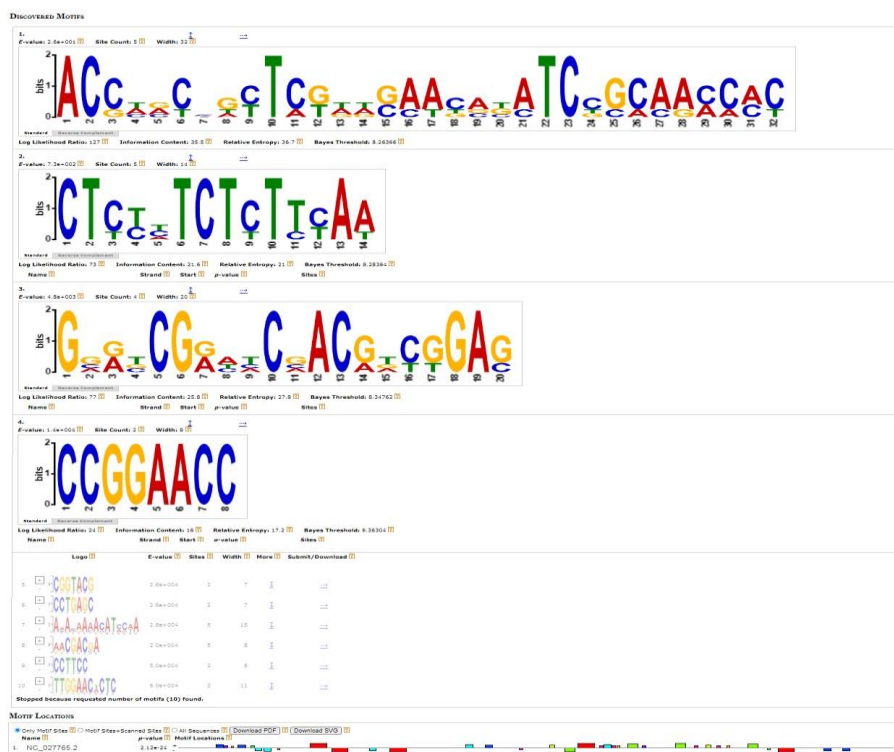
شکل ۱. هم‌ردیفی توالی‌های اسید آمینه دو پروتئین *MYB28* تنظیم کننده گلوکوزینولات آلیفاتیک با استفاده از *CLUSTALW*.

در ژنوم‌های مختلف ممکن است عملکردهای مختلفی داشته باشند (کنین و موشگیان و بورک، ۱۹۹۶).

ژن‌های دخیل در بیوستنز ترکیبات فنلی

از فاکتور رونویسی دخیل در بیوستنز فلاونوئیدها که اختلاف بیان نشان دادند شامل (*MYC3*) بود که در *B. napus* افزایش بیان نشان داد. در گیاهان، فاکتورهای اصلی رونویسی (*bHLH*) نقش اصلی را در فرآیندهای مختلف بیولوژیک بازی می‌کنند. تجزیه و تحلیل جامع و سیستماتیک ژنوم پروتئین *bHLH* بخوبی در آرکیدوپسیس و سایر گونه‌های گیاهی انجام شده است و تعداد زیادی از ژن‌های خانواده *bHLH* در براسیکا شناسایی شده است و نقش برخی *bHLH1* را به‌عنوان تنظیم‌کننده مثبت بیوستنز فلاونوئیدها گزارش کردند. با این وجود عملکرد اصلی این ژن‌ها کاملاً شناخته نشده است. از طرفی تجزیه و تحلیل ژن‌های همولوگ نشان داد که *MYC3* در *B. napus* دارای دو پارالوگ است که در بافت‌های متمایز بذر و برگ بیان شدند و نتایج سایت MEME و ابزار STAMP نشان داد که موتیف اتصال DNA آنها مشابه است (شکل ۲).

نتایج بررسی موتیف توالی‌ها با استفاده از سایت MEME و وب ابزار STAMP، نشان داد که موتیف اتصال DNA دو ژن پارالوگ *MYB28* که هر دو در بافت برگی بیان شدند، غیر مشابه است. این می‌تواند به این علت باشد که دو ژن بایستی عملکردهای مجزا و هم‌پوشانی را در کنترل بیوستنز گلوکوزینولات آلیفاتیک ایفا کنند و احتمال دارد به دنبال وقایع هیبریداسیون و تکثیر تکامل یافته‌اند. تکثیر ژن یک محرک برای نوآوری تکاملی است زیرا به ارگانسیم اجازه می‌دهد تا عملکردهای بیولوژیک موجود را از طریق تنوع بخشیدن و تخصصی کردن به پارالوگ‌های ژنی اولیه اضافی شرح دهد. در مطالعه حاضر، افزایش سطح بیان *MYB28* با وجود نسخه‌های متعدد از ژن در گونه‌های براسیکا منطقی است. پلی‌پلوئیدی ژنوم‌ها یک رویداد بسیار مهم برای تکامل و تنوع گیاه است که ممکن است منجر به تقویت یا افزونگی ژن‌ها شود و ژن‌های تکراری با الگوهای مختلف بیان، مانند خاموش شدن یا بیان قوی در مقایسه با عملکرد اجداد به وجود آورد. گیاهان آلوپلی‌پلوئید همچنین الگوهای مختلف بیان خاص بافت-شرایط را از ژن‌های کپی شده توسط پلی‌پلوئیدی نشان می‌دهند (اودال و سوانسون و نتلتون، ۲۰۰۶). تعداد زیادی تکرار و یا حذف در طول تاریخ تکاملی یک ژن می‌تواند عملکرد ناقص و یا جابجایی ژن غیر ارتولوژیک را نشان دهد و در نتیجه همولوگ‌ها



شکل ۲. نتیجه کاوش موتیف اتصال به DNA پارالوگ *MYC3* با استفاده از وب سایت MEME. ۱۰ موتیف شناسایی شده با فرمت ماتریس احتمال (Probability matrix) از سایت MEME دانلود شد و به‌عنوان ورودی وب ابزار STAMP برای شناسایی شباهت توالی‌های موتیف پارالوگ‌ها استفاده شد.

کتابخانه‌ها ممکن است مربوط به رنگ زرد پوسته بذر براسیکا باشد. دانه زرد از صفات با کیفیت مطلوب گونه دانه‌های روغنی براسیکا است. پیش از این، چندین ژن رنگ پوسته دانه در گونه براسیکا نقشه‌برداری شده است، اما سازوکار مولکولی هنوز ناشناخته است. کی و و و و ژو، (۲۰۲۰) با مطالعه توزیع کروموزوم و تجزیه و تحلیل ژن‌های هم‌مکان^۳ نشان دادند که تکثیر ژن بعد از هیبریداسیون بین *B. rapa* و *B. oleracea* ساز و کار اصلی گسترش *BnabHLH* است. تجزیه و تحلیل بیان نشان داد که *BnabHLHs* به‌طور گسترده‌ای در بافت‌های مختلف گیاه وجود دارد و هفت الگوی اصلی تشکیل می‌دهد که نشان می‌دهد آنها ممکن است در جنبه‌های مختلف توسعه *B. napus* شرکت کنند. در

وجود چندین ژن پارالوگ در ژنوم پلی‌پلوئیدی نشان می‌دهد که آنها احتمالاً در گونه *B. napus* در روند تکثیر ایجاد شده‌اند. مشخص شده است که فاکتور رونویسی *MYC3* در *B. rapa* در نتیجه سه برابر سازی ژنوم دارای سه پارالوگ است (زانگ و کیم و کیم، ۲۰۰۹). لی و چن و هونگ، (۲۰۱۲) یک ژن (*TT8*)^۱ را در براسیکا شناسایی کردند. *BrTT8* که تجمع پروانتوسیانیدین‌ها و تجمع رنگدانه‌ها در پوسته دانه را تنظیم می‌کند، یک پروتئین *bHLH* را کد می‌کند که شباهت زیادی به سایر پروتئین‌های *bHLH* در براسیکا دارد. با تجزیه و تحلیل بیشتر بیان ژن آنها همچنین نشان داد که *BrTT8* در کنترل ژن‌های بیوستتر دیررس (*LBGs*)^۲ مسیر فلاونوئید نقش دارد. تفاوت در بیان این ژن در

1. Transparent TESTA8
2. Late Biosynthetic Genes

سنتز لیگنان‌ها ژن (PLR)^۱ را می‌توان نام برد که در کتابخانه *B. napus* نسبت به سایر کتابخانه‌ها افزایش بیان داشت. لیگنان‌ها گروه بزرگی از پلی‌فنل‌های با وزن مولکولی کم هستند که اثرات ضد تغذیه در دفاع از بذرها و گیاهان در برابر گیاه‌خواران دارند. فرید و اسنودون، (۲۰۰۹) گزارش داده‌اند که محتوای گلوکوزینولات و فلاونول‌های گیاه بسته به ژنوتیپ و توسط عوامل محیطی شدت تغییر می‌کنند. بنابراین تلاش‌ها زمانی مفید خواهند بود که ارزیابی ژنوتیپ‌های مختلف در برنامه‌های اصلاح گیاهان زراعی انجام شود. افزون بر این باید این موضوع را مد نظر قرار داد که اکثر داده‌های بیولوژیک که با سرعت فوق‌العاده‌ای به پایگاه‌های داده سپرده می‌شود، در پاسخ به نیازهای ویژه دانشمندان و با صرف هزینه پدید آمده‌اند، که اغلب به مسائل اختصاصی می‌پردازند. این وضعیت به‌طور جدی سودمندی داده‌های بیولوژیک را محدود می‌کند. لذا مزیت تجزیه و تحلیل چنین داده‌های باز در مقیاس بزرگ، این است که می‌تواند برای کشف وقایع و دیگر اهداف مورد استفاده قرار گیرند و به سؤالات بیولوژیک متفاوت پاسخ دهند. از سوی دیگر، استفاده مناسب از این داده‌ها، تنها با بررسی کامل کیفیت و یکپارچگی داده‌ها قابل اطمینان است و ضمانت کیفیت و کامل بودن داده‌ها یک موضوع مهم در مباحث بیولوژی است.

تنوع ژنتیکی محدود یک مشکل عمده در پرورش کلزا است چراکه انتخاب طولانی مدت و یک طرفه منجر به کاهش تنوع ژنتیکی آن شده است. منابع ژنتیکی مورد نیاز اصلاح کلزا با تأکید بر محتوای روغن زیاد و ویژگی با کیفیت بذر بیشتر کاهش یافته است. افزون بر این، کلزا اشکال وحشی طبیعی ندارد که بتواند منابع طبیعی تنوع باشد.

این تحقیق پارالوگ‌بندی بیان شده در بذر می‌تواند به درک ساز و کار مولکولی و تکاملی دخالت *bHLH* در تشکیل رنگ پوسته بذر در گونه *B. napus* که یک صفت مهم در کیفیت روغن است کمک کند. از دیگر متابولیت‌های ثانویه که در کتابخانه *B. napus* نسبت به کتابخانه *B. rapa* و *B. oleracea* افزایش بیان داشت می‌توان به ژن‌های دخیل در سنتز فلاونوئیدها شامل کوئرستین و کامپفرول و همچنین اسیدهای هیدروکسی سینامیک (HCAA)^۱ اشاره کرد. در کتابخانه *B. rapa* و *B. napus* ژن *CHI*^۲ در بیوسنتز کوئرستین، ژن *Fe2OG dioxygenase domain-containing protein* در بیوسنتز کامپفرول و ژن *C4H*^۳ در بیوسنتز اسیدهای هیدروکسی سینامیک افزایش بیان نشان دادند (جدول ۲). کوئرستین و کامپفرول فلاونول از گروه پلی‌فنل‌های فلاونوئیدی هستند. چندین فرآیند فیزیولوژیکی گیاه مانند جوانه‌زنی بذر، رشد گرده، سیستم آنتی‌اکسیدانی و فتوسنتز را تسهیل می‌کنند (سینگ و ارف و باجور، ۲۰۲۱) اسیدهای هیدروکسی سینامیک دسته‌ای از اسیدهای معطر یا فینیل پروپانوئیدها هستند که به بسیاری از فرآیندهای رشد و همچنین پاسخ گیاهان در برابر پاسخ‌های تنش‌های زیستی و غیر زیستی کمک می‌کنند. رنو ژانگ و لی، (۲۰۲۱) نشان دادند که سطح بیان ژن‌ها و محتوای نسبی متابولیت‌های دخیل در مسیر بیوسنتز فلاونوئیدی تشکیل رنگ‌دانه زرد را در *B. rapa* سبب می‌شود. از آنجایی که سطح بیان ژن‌های دخیل در سنتز فلاونوئیدها در *B. napus* و *B. rapa* بالا بود این مطالعه زمینه‌ای برای تحقیقات بیشتر در مورد ساز و کار مولکولی این فلاونوئیدها در تشکیل رنگ پوسته دانه فراهم می‌کند. از ژن‌های دخیل در

1. Hydroxycinnamic acids
2. Chalcone-flavonone isomerase
3. Cinnamate-4-hydroxylase

بیان شدند، موتیف اتصال DNA متفاوتی داشتند که می‌تواند عملکردهای مجزا و همپوشان آن‌ها را رقم بزند، در حالی که دو پارالوگ ژن *MYC3* که در دو بافت متفاوت بذر و برگ بیان شدند، موتیف اتصال یکسانی داشتند. این یافته‌ها می‌تواند به درک ساز و کار تکامل مولکولی عملکرد *bHLH* در گونه‌های براسیکا که یک صفت مهم در تشکیل رنگ پوسته بذر و کیفیت روغن است کمک کند. مطالعه در مورد بیان و واریانس عملکردی ژن‌های همولوگ ناشی از پلی‌پلوئیدی از اهمیت اساسی برای درک بهتر ساز و کارهای پیچیده کنترل متابولیت‌های ثانویه در این محصولات برخوردار است و امکان دستکاری ژن‌های خاص برای مهندسی هدفمند را فراهم می‌کند.

شناخت خوب از دامنه تنوعی که بر یک ویژگی مشخص اثر می‌گذارد، درک درستی از مجموعه ژن‌های موجود را فراهم می‌کند و برای بهبود تنوع از طریق انتخاب ضروری است (حیدری و حیدری و عزیزی نژاد، ۲۰۲۰). چنین انتخابی از گیاهان بر پایه واگرایی ژنتیکی در بسیاری از محصولات موفقیت‌آمیز بوده است. یافته‌های این پژوهش نشان داد که اکثر ژن‌های درگیر در بیوستنز گلوکوزینولات‌ها در *B. napus* دارای ارتولوگ‌هایی در سایر گونه‌های اجدادی و آراییدوپسیس هستند که احتمالاً به دنبال واگرایی ژنوم از گونه‌های اجدادی و آراییدوپسیس تکامل یافته‌اند. از طرفی مشخص شد که دو پارالوگ ژن *MYB28* که فقط در بافت برگی

REFERENCES

- Bassel, G. W., Glaab, E., Marquez, J. & Bacardit, J. (2011). Functional Network Construction in Arabidopsis Using Rule-Based Machine Learning on Large-Scale Data Sets. *Plant Cell*, 23 (9), 3101-3116.
- Bisht, N.C., Gupta, V., Ramchiary, N., Sodhi, Y.S., Mukhopadhyay, A., Arumugam, N., Pental, D. & Pradhan, A. (2009). Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. *Theoretical and Applied Genetics*, 118 (3), 413-421.
- FAO (2020). Oilseeds oils and meals to contribute to world food security. America, New York.
- Friedt, W. & Snowdon, R. (2009). *Oilseed rape in Oil Crops*. New York, Vollmann press.
- Gigolashvili, T., Engqvist, M., Yatusевич, R., Muller, C. & Flugge, U. I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant Journal*, 50 (5), 886-901.
- Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Werner, P. & Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology*, 30, 798-802.
- Hayes, J.D., Kelleher, M.O. & Eggleston, I.M. (2008). The cancer chemopreventive actions of phytochemicals derived from glucosinolates. *European Journal of Nutrition*, 47 (2), 73-88.
- Heidary, P., Maleki Zanjani, B. & Heidary, S. (2012). A study of gene expression and functional genomics of wheat, rice, cotton and festuca plants under drought stress by analyzing expressed sequence tags (EST). *Modern Genetics Journal*, 7(2 (29)), 129-140. (in Persian).
- Heidari, Sh., Azizinezhad, R., Haghparast, R. & Heidari, P. (2019). Evaluation of the association among yield and contributing characters through path coefficient analysis in advanced lines of durum wheat under diverse conditions. *Journal of Animal and Plant Sciences*, 29 (5), 1325-1335.
- Heidari, Sh., Heidari, P., Azizinezhad, R., Etmnan, A. & Khosroshahli, M. (2020). Assessment of variability

- heritability and genetic advance for agro-morphological and some in-vitro related-traits in durum wheat. *Bulgarian Journal of Agricultural Science*, 26(1), 120-127.
- Ke, Y. Z., Wu, Y. W. & Zhou, H. J. (2020). Genome-wide survey of the bHLH super gene family in *Brassica napus*. *BMC Plant Biology*, 20, 115.
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J. & Mitchell-Olds, T. (2001). Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiology*, 126(2), 811-825.
- Koonin, E. V., Mushegian, A. R. & Bork, P. (1996). Non-orthologous gene displacement. *Trends in Genetics*, 12(9), 334-336.
- Kumar, R., Lee, S. G., Augustine, R., Reichelt, M., Vassão, D. G., Palavalli, M. H., Allen, A., Gershenzon, J., Jez, J. M. & Bisht, N. C. (2019). Molecular Basis of the Evolution of Methylthioalkylmalate Synthase and the Diversity of Methionine-Derived Glucosinolates. *The Plant Cell*, 31(7), 1633-1647.
- Li, X., Chen, L., Hong, M., Zhang, Y., Zu, F., Yi, B., Shen, J., Tu, J. & Fu, T. (2012). A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in *Brassica rapa*. *PloS One*, 7(9), e44145.
- Lysak, M. A., Koch, M. A., Pecinka, A. & Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. *Genome Research*, 15 (4), 516-525.
- Mahony, S. & Benos, P. V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Research*, 35, W253-W258.
- Masoudi Nejad, A., Tonomura, K., Kawashima, Sh., Moriya, Y., Suzuki, M., Itoh, M., Kanehisa, M., Endo, T. & Goto, S. (2006). EGAssembler: online bioinformatics service for large-scale processing clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Research*, 34, 459-462.
- McGinnis, S. & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32: 20-25.
- Neerinx, P. & Leunissen, J. (2005). Evolution of web service in bioinformatics. *Briefings in Bioinformatics*, 6 (2), 178-188.
- Nugroho, A.B., Han, N. & Pervitasari, A.N. (2020). Differential expression of major genes involved in the biosynthesis of aliphatic glucosinolates in intergeneric Baemoochae (Brassicaceae) and its parents during development. *Plant Molecular Biology*, 102, 569.
- Panjabi, P., Jagannath, A., Bisht, N. C., Padmaja, K. L., Sharma, S., Gupta, V., Pradhan, A. K. & Pental, D. (2008). Comparative mapping of Brassica juncea and Arabidopsis thaliana using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassica genomes. *BMC Genomics*, 9, 113.
- Pérez-Bercoff, A., Makino, T. & McLysaght, A. (2010). Duplicability of self-interacting human genes. *BMC Evolutionary Biology*, 28(10), 160.
- Ramsak, Z., Baebler, S., Rotter, A., Korbar, M., Mozetic, I., Usadel, B. & Gruden, K. (2014). GoMapMan: integration consolidation and visualization of plant gene annotations within the MapMan ontology. *Nucleic Acids Research*, 42(1), D1167-D1175.
- Ren, Y., Zhang, N., Li, R., Ma, X. & Zhang, L. (2021). Comparative transcriptome and flavonoids components analysis reveal the structural genes responsible for the yellow seed coat color of *Brassica rapa* L. *PeerJ*, 9, e10770.

- Romualdi, C., Bortoluzzi, S. & Danieli, G. A. (2003). IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiological Genomics*, 12, 159-162.
- Sharma, M., Mukhopadhyay, A., Gupta, V., Pental, D. & Pradhan, A. K. (2016). BjuB.CYP79F1 Regulates Synthesis of Propyl Fraction of Aliphatic Glucosinolates in Oilseed Mustard *Brassica juncea*: Functional Validation through Genetic and Transgenic Approaches. *PLoS One*, 11(2), e0150060.
- Singh, P., Arf, Y., Bajuz, A. & Hayt, S. (2021). The role of quercetin in plant. *Plant Physiology*, 166, 10-19
- Sinha, S. & Tompa, M. (2003). YMF: a program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Research*, 31(13), 3586-3588.
- Sønderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J. & Halkier, B. A. (2010). A Complex Interplay of Three R2R3 MYB Transcription Factors Determines the Profile of Aliphatic Glucosinolates in *Arabidopsis*. *Plant Physiology*, 153 (1), 348-363.
- Town, C.D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., Hine, E. E., Althoff, R., Arbogast, T. S., Tallon, L. J., Vigouroux, M. T., Rick, M. & Bancroft, I. (2006). Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell*, 18(6), 1348-1359.
- Udall, J. A., Swanson, J. M., Nettleton, D., Percifield, R. J. & Wendel, J. F. (2006). A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics*, 173, 1823-1827.
- Vassilev, D. J., Leunissen, A., Atanassov, A., Nenov, A. & G, Dimov. (2005). Application of bioinformatics in plant breeding. *Biotechnology and Biotechnological Equipment*, 19, 139-152.
- Wood, T., Takebayashi, N., Barker, M., Mayrose, I., Greenspoon, P. & Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences*, 106 (33), 13875-13879.
- Yonekura, K. & Saito, K. (2009). Functional genomics for plant natural product biosynthesis. *Natural Product Reports*, 26 (11), 1466-1487.
- Zang, Y. X., Kim, H. U. & Kim, J. A. (2009). Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. *The FEBS Journal*, 276 (13), 3559-3574.