بررسی *In silico* روند تکامل مولکولی و گسترش خانوادۀ پروتئینی ژرانیل‌ژرانیل‌دی فسفات‌سنتاز در گیاهان

سارا فرح بخش[۱]، سنبل ناظری[۱*] و زرین مینوچهر[۲]**

[۱] **ایران، همدان، دانشگاه بوعلی سینا همدان، دانشکدۀ کشاورزی، گروه بیوتکنولوژی**

[۲] **ایران، تهران، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، پژوهشکدۀ زیست فناوری صنعت و محیط زیست، گروه زیست فناوری سامانه‌ای**

تاریخ دریافت: ۹۶/۱۱/۳۰ تاریخ پذیرش: ۹۷/۴/۲۷

چکیده

آنزیمهای GGPPS مسئول سنتز ژرانیل‌ژرانیل دی فسفات(GGPP) ، یک پیش‌ماده مهم برای تولید بسیاری از ترپنوئیدها، هستند. در این تحقیق با شناسایی هومولوگهای خانوادۀ GGPPS در پایگاه دادۀ UniProt، روند تکاملی این خانواده با استفاده از آنالیزهای فیلوژنتیک مورد مطالعه قرار گرفت. همچنین، الگوی فرآیندهای مضاعف شدن و از دست دادن ژن در طی تکامل گیاهان، با استفاده ازمعیار بهینه سازی مبتنی بر پارسیمونی بررسی شد. جهت بررسی نقش فرآیندهای مضاعف شدن کل ژنوم (WGD) و تکثیر تکراری ژنها (TD) از پایگاه دادۀ PLAZA استفاده شد. نتایج مشخص نمود که خانوادۀ GGPPS یک خانوادۀ در حال گسترش است و در طی روند تکامل گیاهان، این خانوادۀ پروتئینی در چندین مرحلۀ متناوب فرآیندهای گسترش و انقباض را تجربه نموده است. اغلب این فرآیندهای گسترش در اثر پدیدۀ مضاعف شدن کل ژنوم ایجاد شده‌اند. با این وجود، فرآیندهای تکثیر تکراری ژنها در پدیده‌هایی که منجر به گسترش در یک تبار خاص (LSE) شده است، بیشتر به وقوع پیوسته است. همچنین بررسی تنوع جایابی پروتئین در داخل سلول (PSL) مشخص نمود که فرآیند گسترش منجر به تفکیک عملکرد در این خانوادۀ پروتئینی شده است. لذا به نظر می‌رسد که گسترش خانوادۀ GGPPS در گیاهان باعث ایجاد پاسخهای مناسب‌تر به استرسها و محرکهای محیطی می‌شود و در نهایت به سازگاری بهتر گیاهان با محیط کمک می‌کند. بررسی اثر انتخاب طبیعی در این خانواده نشان داد که اعضای این خانواده در رده‌های مختلف گیاهان تحت تأثیر فشارهای انتخابی متفاوتی قرار دارند و انتخاب طبیعی مثبت فقط در بازدانگان مشاهده شد.

واژه های کلیدی: ترپنوئید، ژرانیل‌ژرانیل‌دی فسفات سنتاز، مضاعف شدن کل ژنوم، تکثیرتکراری ژن، گسترش در یک تبار خاص

* نویسنده مسئول، تلفن: ۳۴۴۲۴۳۶۶-۰۸۱؛ پست الکترونیکی:snazeri@basu.ac.ir

** نویسنده مسئول، تلفن: ۴۴۷۸۷۳۶۲-۰۲۱؛ پست الکترونیکی: minuchehr@nigeb.ac.ir

مقدمه

ترپنوئیدها بزرگترین و متنوع‌ترین گروه متابولیتها در گیاهان می‌باشند. آنها در ساختار مولکولهای متنوعی مانند مولکولهای دفاعی، رنگدانه‌های فتوسنتزی و فیتوکرومها شرکت می‌کنند. همه ترپنوئیدها از واحدهای پنج کربنه ایزوپرن با نامهای؛ ایزوپنتیل دی فسفات (IPP) و دی متیل آلیل دی فسفات (DMAPP) تشکیل شده‌اند. در گیاهان IPP به وسیله مسیر موالونیک اسید (MVA) در سیتوسل تولید می‌شود. مسیر متیل اریترول فسفات (MEP) نیز IPP

را در پلاستیدها تولید می‌نماید. بر اساس تعداد واحدهای ایزوپرنی که در ساختار ترپنوئیدها به کار رفته است، می‌توان آنها را به گروههای مختلفی تقسیم نمود. از جمله: همی‌ترپنوئیدها (C5)، منوترپنوئیدها (C10)، سزکوئی‌ترپنوئیدها (C15)، دی‌ترپنوئیدها (C20)، سزترترپنوئیدها (C25)، تریترپنوئیدها (C30)، تتراترپنوئیدها (C40) و پلی‌ترپنوئیدها با تعداد بیشتری از واحدهای ایزوپرون (۱۱).

در مسیر بیوسنتز ترپنوئیدها آنزیمهای پرنیل‌ترانسفراز (PTs) در نقاط انشعاب مسیر فعالیت می‌نمایند و نقش مهمی در تولید ترپنوئیدهای مختلف دارند. از جمله این آنزیمها می‌توان به ژرانیل‌ژرانیل دی فسفات سنتاز (GGPPS)اشاره نمود. این آنزیم با استفاده از فارنسیل دی فسفات (FPP) و IPP ژرانیل‌ژرانیل دی فسفات (GGPP)، پیش مادهٔ اصلی برای تولید بسیاری از ترپنوئیدها، را تولید می‌نماید. خانوادهٔ GGPPS به طور گسترده‌ای در سلسلهٔ گیاهان پراکنده شده‌اند و در مقایسه با سایر آنزیمهای پرنیل‌ترانسفراز عملکرد پیچیده‌ای دارند. در گیاهان تکامل‌یافته‌تر، معمولاً تعداد پارالوگهای این خانواده افزایش می‌یابد. این پارالوگها در اثر فرآیندهای مختلف مضاعف شدن ژن ایجاد می شوند، مانند؛ مضاعف شدن کل ژنوم، تکثیر تکراری ژنها و مضاعف شدن قطعه‌ای (Block duplication). در ادامه فرآیندهای تفکیک عملکرد (Subfunctionalization)و ایجاد عملکرد جدید (Neofunctionalization) در ژنهای مضاعف شده منجر به ایجاد تنوع بالایی در عملکرد و یا الگوی بیان ژن می شوند (۵ و ۳۰).

فرایند مضاعف شدن ژن یکی از پیش نیازهای اساسی برای ایجاد ژنهای جدید و گسترش خانوادهٔ ژنی است. در طی فرآیند دو برابر شدن، یک کپی از ژن می‌تواند از انتخاب منفی (Purifying selection) رها شود و امکان تجمع موتاسیونها در آن افزایش یابد. حال اگر انتخاب طبیعی به نفع تکثیر ژنها عمل کند، گسترش آن خانوادهٔ ژنی اتفاق می‌افتد (۶ و ۷). به نظر می‌رسد فرآیندهای تکثیر تکراری ژنها نقش موثرتری در تکامل انتخابی داشته باشند. در مقایسه با فرآیندهای غیرتکراری (NTDs)، فرآیندهای تکثیر تکراری ژنها خزانه بزرگتری از تغییرات پایدار را ایجاد می‌نمایند. بنابراین، شانس بیشتری برای حضور در ژنهای مرتبط با پاسخ به تنشهای زنده دارند. از آنجا که هر گیاه تاریخچه و شرایط زندگی خاص خود را دارد، فشار انتخاب طبیعی در تبارهای مختلف گیاهان متفاوت خواهد بود. بنابراین، هنگامی که گروههای ارتولوگ در یک تبار در جهت پاسخ به شرایط محیطی خاص گسترش می‌یابند، معمولاً ژنهای پراساخت (paralogous) آنها در تبارهای دیگر تمایلی به حفظ شدن ندارند (۱۳).

پدیده های ایجاد عملکرد جدید و تفکیک عملکرد در یک خانوادهٔ پروتئینی از نتایج احتمالی واگرایی عملکردی می‌باشند که پس از رویداد تکثیر ژن رخ می‌دهند. ایجاد عملکرد جدید فرآیندی است که در طی آن عملکرد ژن اجدادی بین دوکپی جدید پراساخت تقسیم می‌شود. در مقابل، تفکیک عملکرد هنگامی رخ می دهد که یک کپی از ژن در اثر یک تکامل پرشتاب، عملکرد جدیدی را به دست می آورد و کپی دیگر عملکرد قبلی را حفظ می‌نماید (۱۸). تنوع در جایابی پروتئینها (PSL) می‌تواند منجر به ایجاد یک عملکرد جدید و یا تفکیک عملکرد شود. بنابراین، می‌تواند نقش مهمی در تکامل ژنومهای یوکاریوتی داشته باشد (۳و۲۲). تشخیص جایابی یک پروتئین در اجزای درون سلولی عملکرد بیولوژیکی آن آنزیم را تا حدود زیادی مشخص می کند و بینش مناسبی از مسیر تکاملی خانوادهٔ ژنی فراهم می‌آورد (۳۳).

به دلیل نقش کلیدی آنزیمهای GGPPS در بیوسنتز ترپنوئیدها، در این تحقیق، روند تکاملی این خانواده در گیاهان بررسی گردید و نقش تغییرات مولکولی بر روی عملکرد و خصوصیات آنزیم مورد بررسی قرار گرفت. در

این راستا فرآیندهای مضاعف شدن ژن در خانوادهٔ GGPPS در بین گیاهان مطالعه شد. همچنین سهم فرآیندهای تکثیر تکراری ژنها و مضاعف شدن کل ژنوم در گسترش این خانوادهٔ ژنی مورد بررسی قرار گرفت. در ادامه جهت بررسی تنوع تمرکز درون سلولی پروتئینها، از برنامه های پیشگویی جایابی درون سلولی (subcellular localization prediction) استفاده شد. در نهایت، جهت تعیین نواحی مهم و تأثیرپذیر از نظر تکاملی، فشار انتخاب طبیعی بر روی اعضای این خانواده مورد بررسی قرار گرفت.

مواد و روشها

داده های توالی و همردیفی: توالی پروتئینی GGPPS در گیاه *Arabidopsis thaliana* (p34802) از پایگاه دادهٔ UniProt (http://www.uniprot.org/) انتخاب شد و از آن برای بازیابی دیگر توالیهای خانوادهٔ GGPPS در پایگاه دادهٔ UniProtKB Vridiplantae استفاده شد. به این منظور، یک جستجوی PSI-BLAST با گنجاندن حد آستانه E value 10^{-6} انجام شد. همچنین توالیهایی که طول آنها از ۳۰ درصد طول توالی مورد تقاضا کمتر و یا بیشتر بودند حذف شدند. در نهایت توالیهایی از ۱۳۹ گونه مختلف گیاهی به دست آمد.

همردیفی توالیهای پروتئینی با استفاده از روش MAFFT و با الگوریتم FFT-NS-I (۲۰) در سرور GUIDANCE (http://guidance.tau.ac.il) (۲۷) صورت‌گرفت. در این سرور علاوه بر انجام همردیفی، امکان امتیازدهی به همردیفیها وجود دارد و می‌توان نقاط دارای امتیاز همردیفی پایین را حذف نمود. روش MAFFT در این سرور نسبت به سایر روشها از امتیاز همردیفی بالاتری (امتیاز روش MAFFT ۰/۸۲ و امتیاز روشهای MUSCLE و CLUSTALW به ترتیب (۰/۸۰ و ۰/۶۳) برخوردار بود. پس از انجام همردیفی ستونهایی که دارای امتیاز پایین‌تر از ۰/۷۰ بودند حذف شدند. در نهایت نتایج در نرم‌افزار BioEdit (۱۴) نیز مورد اصلاح قرار گرفت.

تجزیه و تحلیل فیلوژنتیک: نتایج همردیفی برای بازسازی درخت فیلوژنتیکی بدون ریشه مورد استفاده قرار گرفت. به این صورت که ابتدا در نرم‌افزار ProtTest 3.0، مدل JTT+G+I به عنوان بهترین مدل تکاملی برای این مجموعه شناخته شد. سپس با استفاده از نرم‌افزار RaxML 3.0 (۱۳)، رسم درخت به روش حداکثر درست‌نمایی (ML) صورت گرفت. جهت پشتیبانی آماری گره‌ها در درخت فیلوژنتیک، از روش بوت استرپ با ۵۰۰ تکرار استفاده شد. در نهایت، مرحله ریشه‌دار نمودن درخت، در نرم‌افزار Notung 2.9 (۴) انجام شد.

تجزیه و تحلیل فرآیند های مضاعف شدن ژن و بررسی گسترش درخانواده پروتئینیGGPPS: به منظور تعیین درجه گسترش خانوادهٔ پروتئینی، جستجوی گسترده ای در ۵۱ ژنوم موجود در پایگاه دادهٔ Phytozome (http://www.phytozome.net) (۱۲) انجام شد و توالیهای خانوادهٔ پروتئینی GGPPS در آنها شناسایی شدند. بازسازی درخت فیلوژنتیک مطابق با روشی که در بخش قبل توضیح داده شد، صورت گرفت. همچنین، درخت گونهٔ مرتبط با گونه‌های مورد بررسی، براساس اطلاعات موجود در فیتوزوم ساخته شد. از نرم افزار Notung2.9 **(۴)** جهت پی بردن به فرآیندهای از دست دادن و به دست آوردن ژن در طی فرآیند تکامل گیاهان استفاده شد. این فرآیندها به روش تلفیق درخت ژن و درخت گونه (Reconciliation) و با استفاده از روش پارسیمونی مورد شناسایی قرارگرفتند. در اصل پارسیمونی فرض بر این است که فرآیندهای مضاعف شدن و از دست دادن ژن پدیده‌های نادری می‌باشند. بنابراین در نرم افزار Notung تنها در صورت عدم وجود اطلاعات از منابع تکاملی دیگر، از پدیده‌های مضاعف شدن و از دست دادن ژن برای توضیح روابط فیلوژنتیکی استفاده می شود. به طوری که در نهایت حداقل پدیدهٔ مضاعف شدن و از دست دادن ژن در فرآیند تکاملی دیده شود. جهت بررسی زمان تقریبی وقایع

گسترش و انقباض در این خانوادهٔ ژنی، درخت زمان مربوط به گونه های مورد مطالعه، با استفاده از اطلاعات موجود در پایگاه دادهٔ TimeTree (www.timetree.org/) (۱۷) رسم شد. همچنین برای تعیین سهم فرآیند های مضاعف شدن تکراری و مضاعف شدن کل ژنوم در گسترش این خانوادهٔ ژنی، از پایگاه دادهٔ PLAZA (۳۰) نیز استفاده شد.

بررسی جایابی پروتئینهای خانوادهٔ GGPPS در سلول: جهت پیش‌بینی جایابی پروتئینهای خانوادهٔ GGPPS در اندامکهای درون سلولی (PSL) ، از توالی کامل پروتئینهای مورد مطالعه استفاده شد. در این بررسی توالیهایی که فاقد نواحی انتهایی بودند از مجموعه داده‌ها حذف شدند. در مرحله بعد، براساس پروتکل امانوئل‌سون و همکاران محل هدف‌گیری و جایابی پروتئینها در سلول با استفاده از نرم‌افزارهای TargetP (۹)، SignalP (۲۸) و CholoroP (۱۰) مورد تجزیه و تحلیل قرار گرفت. علاوه بر این، برای به دست آوردن نتایج دقیق‌تر، از نرم‌افزار LOCALIZER (۳۱) نیز استفاده شد.

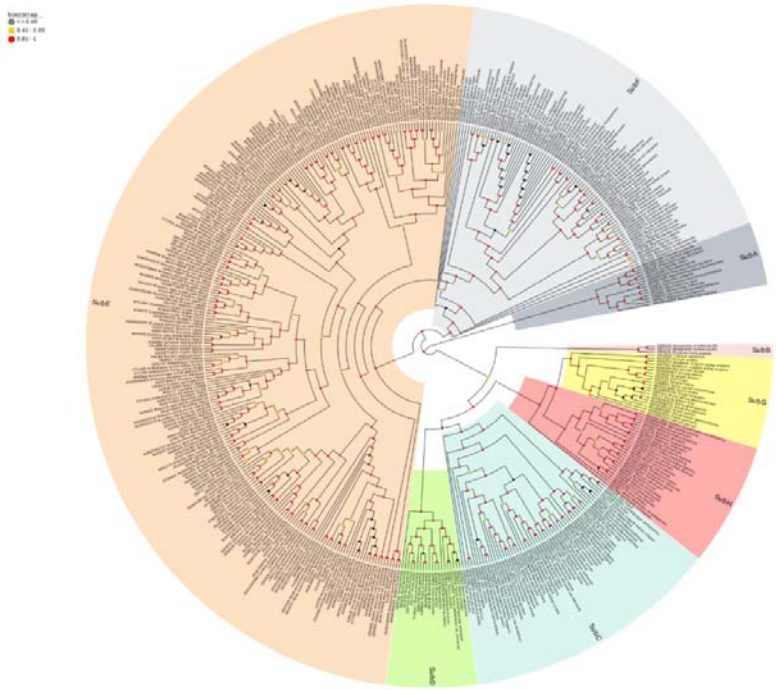
بررسی اثرانتخاب طبیعی مثبت: جهت بررسی اثر انتخاب طبیعی مثبت در خانوادهٔ GGPPS در گیاهان، از برنامهٔ CODEML در بستهٔ نرم‌افزاری PAML 4.8 (۳۶) استفاده شد. به دلیل واگرایی زیاد ژنها در درخت فیلوژنی، توالیهایی از چهار زیرگروه گیاهان آبزی، بازدانگان، گیاهان تک‌لپه وگیاهان دولپه‌ای انتخاب شده و به طور جداگانه مورد بررسی قرار گرفتند. در هر گروه، همردیفی چندگانهٔ کدونهای مرتبط با توالیهای اسیدآمینه‌ای در نرم افزار تحت وب PAL2NAL (http://www.bork.embl.de/pal2nal/) انجام شد. همچنین، درخت فیلوژنی با استفاده از همردیفی چندگانهٔ پروتئینها و با روش حداکثر درست‌نمایی در نرم-افزار MEGA6 بازسازی شد.

فشار انتخاب طبیعی با مقایسهٔ نرخ جایگزینی غیرمترادف (d_N) به نرخ جایگزینی مترادف (d_S) مورد بررسی قرار گرفت. در این روش، اگر یک ژن دارای تکامل خنثی باشد، نسبت d_N به d_S (ω) برابر با ۱ خواهد بود. هنگامی-که $\omega>1$ باشد، ژن انتخاب طبیعی مثبت را تجربه می‌نماید. همچنین، $\omega<1$ به‌معنی وجود انتخاب منفی می‌باشد. در این بررسی از روشهای مبتنی بر حداکثر درست نمایی (ML) برای تعیین فشار انتخاب طبیعی استفاده شد. ابتدا به منظور بررسی متغییر بودن ω در بین جایگاهها، مدل M0 با مدل M3 مورد مقایسه قرار گرفت. در مدل M0 فشار انتخاب طبیعی در تمام جایگاهها یکسان در نظر گرفته می‌شود. در حالی که در مدل M3 فرض بر این است که فشار انتخاب طبیعی در بین جایگاهها متفاوت می‌باشد. در مرحلهٔ بعد، جهت مشخص نمودن جایگاههای انتخاب طبیعی مثبت، مدل M1a با مدل M2a ومدل M7 با M8 مقایسه شدند. مدلهای M1a, M0 و M8 به عنوان فرض صفر درنظر گرفته شدند و از آزمون نسبت درست‌نمایی (LTR) جهت یافتن بهترین مدل برای داده‌ها استفاده شد. سطح معنی‌داری در این مطالعه ۰/۰۵ در نظر گرفته شد.

نتایج و بحث

آنالیز فیلوزنتیک: بررسی درخت فیلوژنتیک نشان داد که تمام هومولوگهای گیاهان آبزی یک کلاد مونوفیلتیک را تشکیل می‌دهند که در شکل۱ به عنوان SubA نشان داده شده است. با این حال، تبارهای دیگر (آنزیمهای GGPPS متعلق به گیاهان خشکی‌زی) تشکیل چندین زیرگروه را داده اند که شامل زیرگروههای B، C، D، E، F، G و H می‌باشند (شکل۱). اعضای زیرگروه B متعلق به گیاهان غیرگلدار مانند قارچها و گیاهان آوندی اولیه می‌باشند. هومولوگهای GGPPS متعلق به زیرگروه C به طور عمده در گیاهان نهاندانه یافت شد، و فقط یک هومولوگ متعلق به گیاهان بازدانه (*Picea sitchensis*) در این گروه مشاهده شد. آنزیمهای متعلق به زیرگروه C به عنوان زیر واحد کوچک (SSUs) مستند سازی شدند. این گروه از آنزیمها می توانند توزیع شار IPP را بین ترکیبات GPP و GGPP

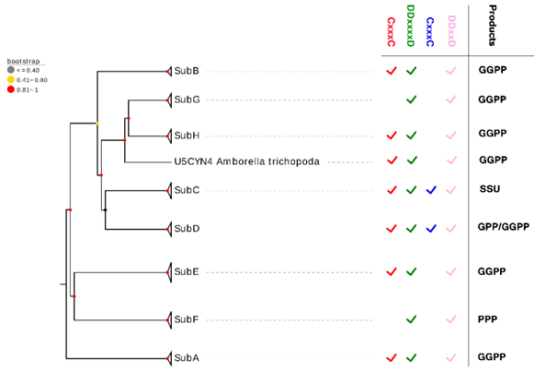
در گیاهان کنترل کند. سطح بیان این ژن در بافتهای مختلف و در شرایط مختلف محیطی و تکاملی گیاه تغییر می کند و به این ترتیب تولید GPP و GGPP کنترل می شود (۷ و۲۶). به نظر می رسد ایجاد چنین آنزیمهایی یک مزیت انتخابی برای گیاهان گلدار است تا آنها بتوانند GGPP یا GPP به طور بهینه تولید کنند. اعضای زیرگروه D متعلق به بازدانگان می‌باشند. پارالوگهای این گروه می توانند انواع مختلفی از پرنیل ترانسفرازها، از جمله GPP، GGPP یا G / GGPP تولید کنند. هومولوگهای GGPPS متعلق به زیرگروه E بیشتر در نهاندانگان دیده می شوند. این گروه خود به زیرگروههای کوچکتر تقسیم می شود، در میان این زیر شاخه ها می توان به E2 و E3 اشاره نمود. آنزیمهایGGPPS مربوط به E2 به طور انحصاری در آستریدها (Asterids) یافت می‌شوند و به عنوان زیرواحدهای بزرگ (GGPPS-LSU) مستندسازی شدند. آنزیمهایLSU نقش مهمی در تنظیم GPP ایفاء می کند که یک پیش مادهٔ اصلی برای مونوترپنها می‌باشد (۱۶ و ۲۶). بنابراین، دسترسی مؤثرتری به مونوترپنها را فراهم می کنند .مونوترپنها در دفاع از گیاه، گرده افشاندن و پراکندگی بذر دخالت می کنند. به نظر می‌رسد حضور هترودایمرها در آستریدها یک مزیت انتخابی برای تولید گردهٔ کارآمدتر می‌باشد. هومولوگهای زیرگروه E3 متعلق به خانوادهٔ Brassicaceae می‌باشند و گسترش در یک تبار خاص (LES) در آنها مشاهده می‌شود. هومولوگهای زیرگروه F که در تک‌لپه‌ایها و دولپه‌ایها دیده می‌شوند؛ محصولی کاملا متفاوت از گروههای دیگر تولید می‌نمایند. این آنزیمها به عنوان Polyprenyl Pyrophosphate Synthase مستندسازی شدند (۵). زیرگروه G متعلق به تک‌لپه‌ایها می‌باشد. در حالی که نزدیکترین گروه به آن یعنی زیرگروه H تنها در دولپه‌ایها یافت می‌شود و شواهد تجربی نشان می‌دهد که این گروه منحصراً GGPP تولید می‌کند (۵).



شکل۱- درخت فیلوژنتیک ریشه‌دار شده متعلق به خانوادهٔ GGPPS در گیاهان: بوت استرپها با نقاط رنگی به نمایش گذاشته شده‌اند. نقاط قرمز رنگ بوت استرپهای بین ۰/۸ تا ۱، نقاط زرد رنگ بوت استرپ ۰/۴ تا ۰/۸ و نقاط خاکستری بوت استرپهای ۰/۴ =< را نشان می‌هند. زیرگروههای A,B,C,D,E,F,G,H با رنگهای مختلف نمایش داده شده‌اند.

جهت فهم بهتر نقش تغییرات مولکولی در تکامل هومولوگهای GGPPS، ساختار نگاره‌های مهم و عملکردی در این خانوادهٔ آنزیمی مورد بررسی قرار گرفت. یکی از مهم ترین نواحی مورد بررسی در دومین ساختاری GGPPS دو ناحیهٔ غنی از آسپارتات می‌باشد. نخستین ناحیهٔ غنی از آسپارتات که با نام FARM شناخته می‌شود، دارای توالی DDxxxxD می باشد. دومین ناحیهٔ غنی از آسپارتات با نام SARM شناخته می‌شود و توالی آن به صورت DDxxD می باشد. این نواحی برای اتصال سوبسترا به آنزیم ضروری می باشند. بررسی نواحی غنی از آسپارتات در توالیها نشان داد که نخستین نگارهٔ غنی از آسپارتات (FARM) ناحیهٔ بسیار حفاظت شده‌ای است. در حالی که دومین نگارهٔ غنی از آسپارتات (SARM) از حفاظت‌شدگی کمتری برخوردار است. هومولوگهای گروهC که به عنوان آنزیمهای زیرواحد کوچک در هترودایمرها (SSUs) شناخته می شوند، فاقد نگارهٔ SARM می‌باشند. این ناحیه جهت عملکرد آنزیم بسیار ضروری می‌باشد، لذا این آنزیمها به تنهایی قادر به فعالیت نمی‌باشند اما در ساختار کمپلکس هترودایمر G(G)PPS شرکت می‌نمایند و تولید محصول را به سمت سنتز GPP سوق می‌دهند. از آنجا که این آنزیمها به طور گسترده‌ای در گیاهان گلدار یافت شدند، احتمالاً موتاسیونهایی که منجر به ایجاد چنین آنزیمهایی شده است در مراحل اخیر تکاملی اتفاق افتاده‌اند و منجر به تولید بهینه‌تر ترپنوئیدها شده‌اند.

یکی دیگر از نواحی مهم مورد بررسی دو نگارهٔ CxxxC می‌باشد. این نواحی برای اتصال زیرواحدها در آنزیمهای هترودایمر G(G)PPS ضروری می‌باشند. در بررسی توالیها مشخص شد که هومولوگهای گروه A و B فاقد یک یا هردو نگارهٔ CxxxC می‌باشند. اما ازآنجا که برخی از هومولوگها در گیاهان آبزی (SubA) دارای یکی از این موتیفها می‌باشند، می‌توان چنین استنباط نمود که این نگاره در مراحل اولیه تفرق گیاهان وجود داشته و در طی فرآیند تکامل برخی از هومولوگها براساس نیاز خود این کپی از نگاره را ازدست داده اند و یا یک کپی دیگر از آن را به دست آورده‌اند. هومولوگهایی که در گروه D دسته بندی شده‌اند، به طور کلی در بازدانگان یافت می‌شوند. در همهٔ این هومولوگها نخستین نگارهٔ CxxxC یافت می‌شود، ولی دومین نگاره به شکل CxxxS تغییر شکل داده است. آنزیمهایی با چنین ساختار دارای عملکرد دوگانه می‌باشند و قادر به تولید هر دو مادهٔ GGPP و GPP می‌باشند. چنین عملکردی سبب تولید بهینهٔ مونوترپنها و دی‌ترپنها در بازدانگان می‌شود. در زیرگروه C هر دو نگارهٔ CxxxC وجود دارد. این گروه از آنزیمها به تنهایی قادر به فعالیت نمی‌باشند و وجود این نگاره‌ها برای فعالیت آنها ضروری می‌باشد. در مقابل، هومولوگهای گروه F که محصولات با طول زنجیره بلند (۲۵ تا ۴۵ کربنه) تولید می‌نمایند، هر دو نگارهٔ CxxxC را از دست داده‌اند. تمام هومولوگهای گروههای E و F فقط دارای نخستین نگارهٔ CxxxC می‌باشند (شکل ۲). به طور کلی، در این نواحی تفاوت قابل ملاحظه‌ای بین هترودایمرها و همودایمرها مشاهده نشد. این نتایج مشخص نمود که ناحیهٔ CxxxC برای اتصال دو زیرواحد ضروری است ولی کافی نیست.



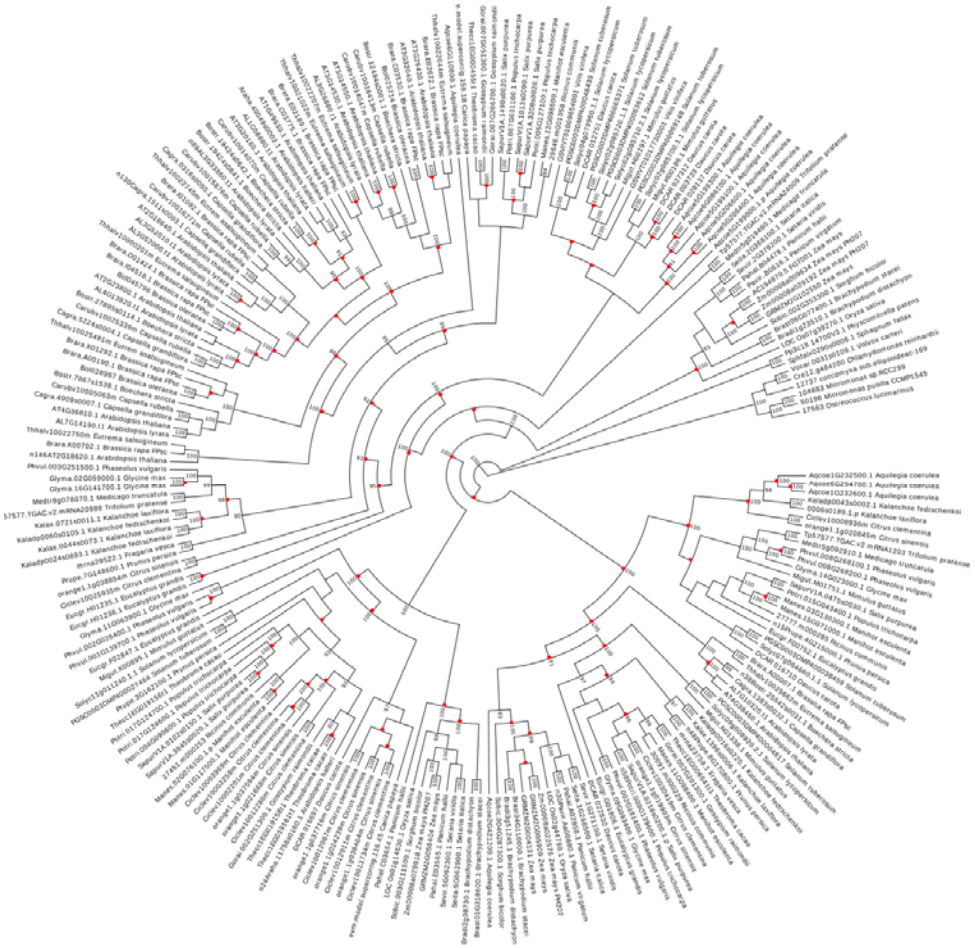
شکل۲ – تکامل مولکولی ناحیهٔ عملکردی پلی‌پرنیل سنتاز: حضور و یا عدم حضور هر نگاره در هر یک از زیرگروهها با علامت مشخص شده است، و محصول نهایی تولید شده در هر یک از زیرگروهها با علایم اختصاری بیان شده است. این محصولات شامل؛ ژرانیل ژرانیل دی فسفات سنتاز (GGPP)، ژرانیل دی فسفات سنتاز (GPP) و سولانسیل دی فسفات سنتاز (PPP).

بررسی روند گسترش و انقباض در خانوادهٔ GGPPS: میزان پدیدهٔ مضاعف شدن ژنها در گیاهان بسیار بیشتر از سایر گونه‌های یوکاریوتی می‌باشد. به کمک این پدیده مواد اولیه برای تکامل گونه‌ها فراهم می‌شود. در نتیجه گیاهان قادر خواهند بود تا سازگاری بهتری در برابر تغییرات محیطی داشته باشند (۳۱). فرآیندهای مضاعف شدن کل ژنوم و تکثیر تکراری ژنها نقش مهمی در گسترش خانواده‌های پروتئینی در گیاهان دارند. فرآیند مضاعف شدن کل ژنوم یکی از مکانیسمهای مهم در تکثیر ژنها است، که سبب افزایش ناگهانی تعداد ژنها در خانواده‌های پروتئینی می‌شود. در مقابل، فرآیند تکثیر تکراری ژنها اغلب سبب گسترش درخانواده‌های پروتئینی‌ای می‌شود که در پاسخ به استرس و محرکهای محیطی نقش دارند. علاوه براین، فرآیندهای تکثیر تکراری ژنها در طی یک نسل دارای فراوانی بالایی هستند. در نتیجه، احتمال بیشتری وجود دارد تا در ژنهای مرتبط با استرس مشاهده و حفظ شوند. در واقع می‌توان گفت که فرآیندهای تکثیر تکراری ژنها در مقایسه با فرآیندهای دیگر باعث به وجود آمدن تنوع بیشتری در خزانه ژنتیکی یک موجود می شوند و از این رو اهداف مناسبی برای انتخاب به حساب می آیند (۱۵ و ۳۲). با توجه به مطالب ذکر شده، برای شناخت کامل خانوادهٔ GGPPS باید درک صحیحی از وقایع مضاعف شدن ژن در این خانوده در طول فرآیند تکامل وجود داشته باشد.

در بررسیهای صورت گرفته بر روی خانوادهٔ GGPPS، ناسازگاری بین درخت ژنی و درخت گونه نشان داد که تاریخچهٔ تکاملی خاصی در این خانواده وجود دارد. در این مطالعه مشخص شد که در طی فرآیند تکامل، خانوادهٔ GGPPS یک خانوادهٔ در حال گسترش را در تبارهای مختلف گیاهی شکل داده است. این خانوادهٔ ژنی حدود ۸۰ فرآیند مضاعف شدن ژن و ۲۱۷ فرآیند از دست دادن ژن را در طی روند تکامل گیاهان تجربه نموده است (شکل ۱). و در همان مراحل اولیه تفرق فرآیند مضاعف شدن ژنها منجر به افزایش در تعداد نسخه‌های ژنی در گیاهان شده‌است. به طوری که، رابطه مستقیمی بین پیچیدگی یک گونهٔ گیاهی وتعداد پارالوگهای GGPPS در آن گونه مشاهده می‌شود. علاوه بر اینها مشاهده می‌شود که فرآیندهای مضاعف شدن ژن در این خانواده باعث گسترش در یک تبار خاص (LSE) شده است، که از جمله مهمترین آنها می‌توان به LSE در خانوادهٔ Brassicaceae اشاره نمود (شکل۴). در مطالعاتی که کومن و همکاران بر روی تکامل این خانوادهٔ پروتئینی انجام دادند نیز به این امر اشاره شده است (۵).

تلفیق درخت ژن و درخت گونه نشان می دهد که جدیدترین جد مشترک (MRCA) در گیاهان تنها دارای یک کپی از ژن GGPPS بود (شکل ۳و ۴). اما در طی فرآیند تکامل، این ژن با نرخ گسترش متفاوتی در چندین مرحله مضاعف شده است. این بررسی نشان می‌دهد که در جلبکهای آبی‌سبز و در گیاهان خشکی‌زی اولیه تنها یک کپی از ژن GGPPS وجود داشته است. در حالی که در همان مراحل ابتدایی تکامل گیاهان و درست پس از جدایی بروفیتها، در MRCA تک‌لپه‌ایها و دولپه‌ایها ۵ فرآیند مضاعف شدن ژن در خانوادهٔ GGPPS اتفاق افتاده است. این روند گسترش پس از جدایی تک‌لپه‌ایها و دولپه‌ایها نیز ادامه داشت. به طوری که در حدود ۱۴۵ میلیون سال پیش، شاهد گسترش قابل ملاحظه‌ای در MRCA گیاهان تک‌لپه می توان بود و پس از آن فرآیندهای پی‌درپی از دست دادن ژن و در نتیجه انقباض آنزیمهای GGPPS در تک‌لپه‌ایها به وقوع پیوسته است.

از میان گیاهان تک‌لپه تنها ذرت (Zea mays) در فرآیندهای اخیر تکاملی، گسترش در این خانوادهٔ ژنی را تجربه نموده است. از سوی دیگر در گیاهان دولپه نیز این فرآیند گسترش تا حدود ۱۱۷ میلیون سال پیش ادامه داشت.



شکل۳- درخت فیلوژنیک ریشه دار شده متعلق به خانوادۀ GGPPS در گیاهان: وقایع مضاعف شدن به صورت نقاط قرمز در گره ها نمایش داده شده است و اعداد نشان دهندۀ بوت استرپ می باشند (بوت استرپهای زیر ۹۰ نشان داده نشده‌اند).

اما فرآیندهای مکرر از دست دادن ژن که پس از آن به وقوع پیوسته، نشان می دهد که اغلب این کپیهای جدید ژنی، قادر به ادامه بقا نبوده و در طی فرآیند تکاملی در اثر انتخاب طبیعی حذف شده‌اند . با این حال از حدود ۷۰ تا ۸۰ میلیون سال قبل فرآیند گسترش آنزیمهای GGPPS دوباره در برخی از تبارها مشاهده می‌شود. برخی از این گسترشها پس از فرآیندهای گونه زایی رخ داده‌اند. این امر می‌تواند نشان دهندۀ نیاز این گونه‌ها به آنزیمهای GGPPS تخصصی برای پاسخگویی به شرایط خاص محیطی باشد (شکل۴و۵). باید توجه داشت که اغلب فرآیندهای گسترش آنزیمهای GGPPS از نظر زمان وقوع با رخدادهای پلی‌پلوئیدی همخوانی دارند (شکل۵).

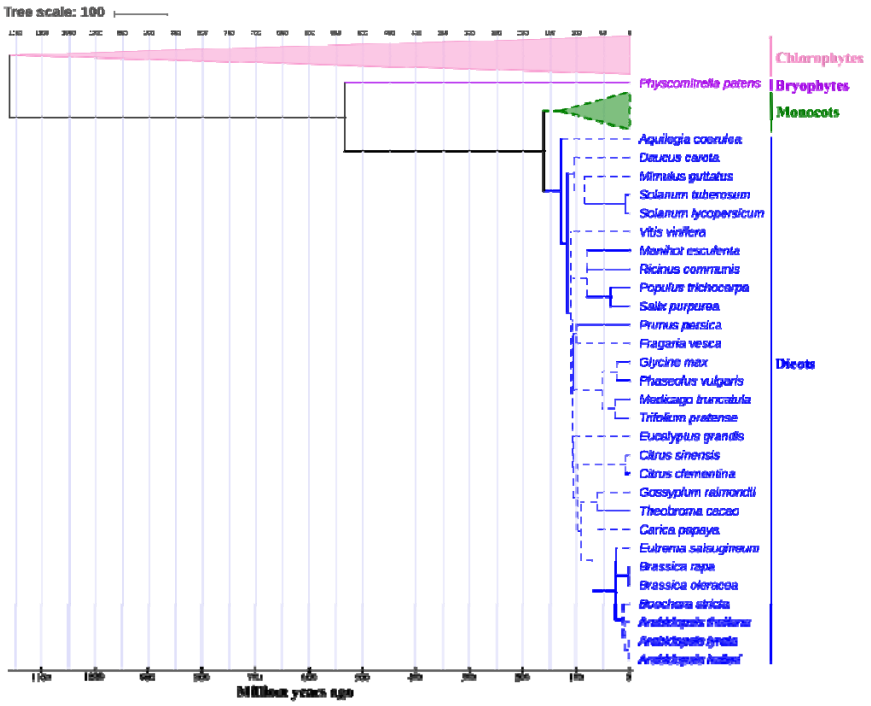
این بدان معناست که اغلب کپیهای ژن در این خانواده در اثر فرآیندهای مضاعف شدن کل ژنوم به دست آمده‌اند. این نتایج با نتایج بررسی در پایگاه دادۀ PLAZA مطابقت دارد (جدول۱). این مطالعه نشان داد که فرآیندهای مختلف مضاعف شدن ژن به گسترش خانوادۀ GGPPS کمک نموده‌اند و به طور کلی پدیده‌های غیر تکراری (NTD) سهم بیشتری در گسترش این خانواده داشته‌اند و این پدیده در تک‌لپه‌ایها مشهودتر است. با این وجود در مواردی که فرآیندهای مضاعف شدن منجر به LSE شده است، نقش فرآیندهای تکثیر تکراری ژنها پررنگ‌تر می‌شود. خانوادۀ Brassicaceae مثالی بر این مدعا است (شکل۴).

شکل۴- گسترش خانوادۀ GGPPS در طی فرآیند تکامل گیاهان: فرآیندهای به دست آوردن ژن در صورت کسر و فرآیندهای از دست دادن ژن در مخرج کسر بر روی شاخه های درخت نمایش داده شده‌اند. تعداد کپیهای ژن در هر گونه و در اجداد مشترک گونه ها در داخل دایره به رنگ زرد نشان داده شده‌اند .

این نتایج تا حدودی با نتایجی که در مطالعات قبلی به دست آمده مطابقت دارد. دراین مطالعات مشخص شده که ژنهای دخیل در تنشهای زیستی عمدتاً در یک تبار خاص قادر به بقا می باشند و اغلب از طریق فرآیندهای تکثیر تکراری ژنها گسترش می‌یابد (۱۵).

بررسی تنوع جایابی پروتئینهای خانوادۀ GGPPS در سلول: به منظور تشخیص فرآیندهای گسترشی که منجربه تغییر در عملکرد شده‌اند، تنوع PSL در این خانواده مورد بررسی قرارگرفت. تغییر در محل جایابی پروتئین در درون سلول (PSL) یکی از منابع مهم بقا و تنوع عملکرد در خانواده‌های پروتئینی است. فرآیند PSL می‌تواند باعث بهبود عملکرد پروتئینها شود و مزایایی را برای جاندار به ارمغان آورد (۱). علاوه بر اینها، تشخیص صحیح جایگاه جایابی پروتئین در درون سلول می‌تواند به تشخیص عملکرد آن پروتئین و مسیر بیولوژیکی که در آن فعالیت

می کند کمک نماید (۲۹). از این رو، پیشگویی جایابی توجه است.
پروتئین در داخل سلول در تحقیقات تکاملی بسیار مورد



شکل ۵- تطبیق فرآیندهای گسترش و انقباض در خانوادهٔ GGPPS با درخت زمان در گونه های مورد بررسی (بر حسب MYA). فرآیندهای گسترش به صورت پررنگ و فرآیندهای انقباض به صورت خط چین نمایش داده شده‌اند.

جدول ۱- سهم انواع فرآیندهای مضاعف شدن در گسترش خانوادهٔ GGPPS

گونه‌های گیاهی	سهم انواع فرآیندهای مضاعف شدن			
	%No	%Block	%Tandem	%Tandem & Block
Green algae:				
Chlamydomonas reinhardtii	۱۰۰	۰	۰	۰
Ostreococcus lucimarinus	۱۰۰	۰	۰	۰
Bryophytes:				
Physcomitrella patens	۰	۱۰۰	۰	۰
Monocots:				
Zea mays	۲۰	۸۰	۰	۰
Sorghum bicolor	۱۰۰	۰	۰	۰
Setaria italica	۱۰۰	۰	۰	۰
Oryza sativa	۱۰۰	۰	۰	۰
Brachypodium distachyon	۱۰۰	۰	۰	۰
Dicots:				
Arabidopsis thaliana	۵۰	۸/۳۳	۲۵	۱۶/۶۶
Arabidopsis lyrata	۵۰	۲۰	۰	۳۰

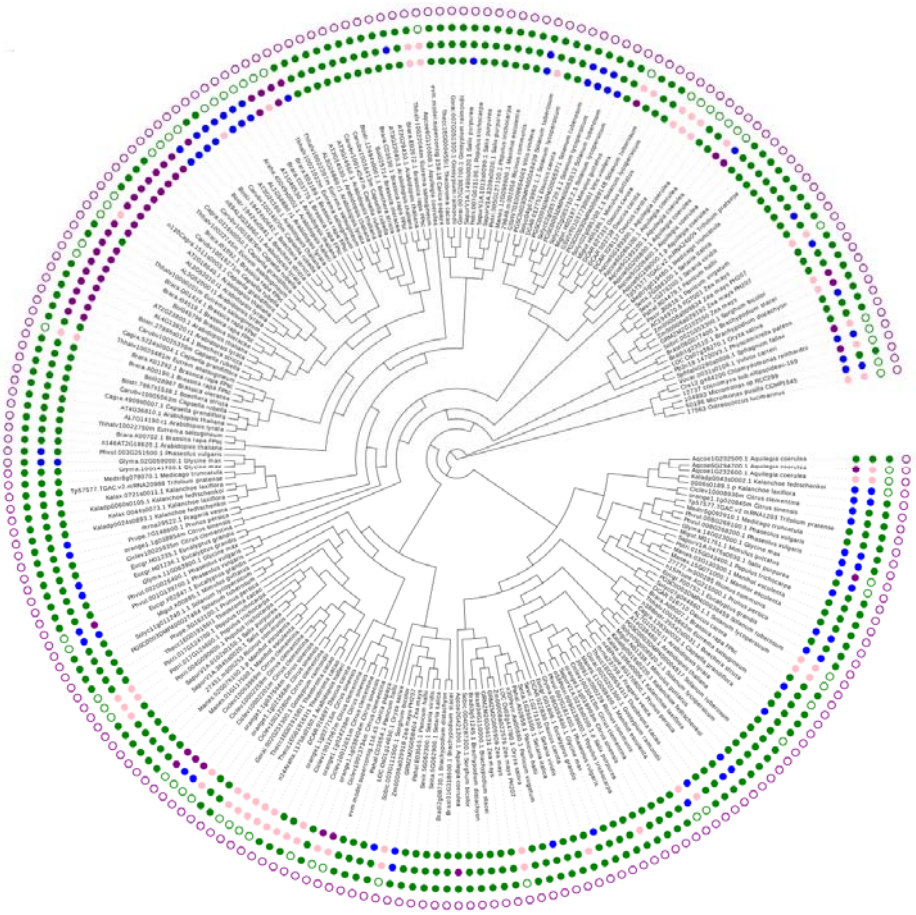
Capsella rubella	۰	۴۲/۸۵	۵۷/۱۴	۰
Brassica rapa	۱۳/۳۳	۶۰	۱۳/۳۳	۱۳/۳۳
Carica papaya	۱۰۰	۰	۰	۰
Gossypium raimondii	۳۰/۷۶	۳۰/۷۶	۱۵/۳۸	۲۳/۰۷
Theobroma cacao	۵۰	۰	۵۰	۰
Citrus sinensis	۶۰	۰	۴۰	۰
Eucalyptus grandis	۶۰	۰	۴۰	۰
Prunus persica	۱۰۰	۰	۰	۰
Fragaria vesca	۱۰۰	۰	۰	۰
Medicago truncatula	۳۳/۳۳	۶۶/۶۶	۰	۰
Glycine max	۶۰	۴۰	۰	۰
Ricinus communis	۱۰۰	۰	۰	۰
Manihot esculenta	۳۳/۳۳	۶۶/۶۶	۰	۰
Populus trichocarpa	۲۸/۵۷	۴۲/۸۵	۰	۲۸/۵۷
Vitis vinifera	۳۳/۳۳	۶۶/۶۶	۰	۰
Solanum lycopersicum	۴۲/۸۶	۱۴/۲۸	۰	۴۲/۸۶
Solanum tuberosum	۴۲/۸۶	۱۴/۲۸	۰	۴۲/۸۶

Source: PLAZA (https://bioinformatics.psb.ugent.be/plaza/)

از آنجا که آنزیمهایی که از IPP به عنوان سوبسترا استفاده می‌کنند در اندامکهای مختلف سلول پراکنده‌اند، منطقی به نظر می‌رسد که آنزیمهای GGPPS نیز در اندامکهای مختلف سلولی مانند کلروپلاست و میتوکندری و یا دستگاه گلژی تجمع یابند (۲۱). مطالعاتی که پیش از این صورت گرفته این فرضیه را تأیید می‌نماید (۲۱ ،۲۷و۳۷). در این مطالعه نیز، تداوم و اهمیت PSL در گسترش خانوادهٔ GGPPS مورد مطالعه قرارگرفت.

آنالیزهای PSL انجام شده بر روی توالیهای خانوادهٔ GGPPS ، نشان داد که تقریباً تمامی این توالیها دارای پپتیدهای انتقالی می‌باشند. در این آنالیزها ناحیهٔ هدف کلروپلاستی، میتوکندریایی و یا شبکهٔ آندوپلاسمی در توالیها شناسایی شد. این بدان معناست که برخی از اندامکهای درون سلولی GGPP را خود تولید می‌نمایند. همانطور که انتظار می‌رفت، اغلب توالیها دارای ناحیهٔ هدف کلروپلاستی بودند (cTP) (شکل۶). آنزیمهای GGPPS فوق می‌توانند در بیوسنتز ایزوپرنوئیدهایی مانند کارتنوئیدها، کلروفیل‌ها و ای‌ان‌تی- کارن و همچنین بیوسنتز برخی از متابولیتهای ثانویه مانند تاکسول شرکت نمایند (۱۶) . تعدادی از توالیها دارای ناحیهٔ هدف میتوکندریایی بودند (mTP) (شکل۶). این آنزیمها در بیوسنتز ایزوپرنوئید کوئینون شرکت می‌کنند. این ماده به عنوان یک ناقل الکترون در زنجیرهٔ انتقال الکترون بسیار ضروری است (۲۵). آنزیمهای با ناحیهٔ هدف میتوکندریایی اغلب در خانوادهٔ Brassicaceae مشاهده شدند. اما در مورد پپتیدهای سیگنالی پیشگوییهای متفاوتی وجود داشت. نرم‌افزار TargetP که می‌تواند سیگنال پپتیدها را با دقت و حساسیت بالایی تشخیص دهد و همچنین نرم‌افزار LOCALIZER توانستند ناحیهٔ پپتیدهای سیگنالی را در برخی از توالیهای خانوادهٔ Brassicaceae مشخص نمایند، ولی SignalP نتوانست هیچ سیگنال پپتیدی را شناسایی نماید (شکل۶). آنزیمهایی که دارای سیگنال پپتید (SP) می‌باشند در واکنشهای پرنیلاسیون پروتئین شرکت می‌کنند، که در مسیرهای پیام-رسان بسیار مهم است. بر اساس این مشاهدات، جایابی در شبکهٔ آندوپلاسمی و میتوکندری در این خانوادهٔ پروتئینی

یک پدیدهٔ نادر است، اما در طی تکامل گیاهان، در برخی از تبارها، آنزیمهای GGPPS برای شرکت در مسیرهای متابولیکی خاصی تکامل یافته‌اند. این پدیده به ویژه در خانوادهٔ Brassicaceae که در فرآیندهای اخیر تکاملی دچار LSE شده ، بیشتر قابل مشاهده است. اما به طور کلی، PSR نقش کوچکی در ایجاد تنوع عملکردی در ژنهای تکثیر شده در این خانواده دارد.



شکل۶- بررسی جایابی درون سلولی پروتئینهای خانوادهٔ GGPPS در گیاهان با استفاده از نرم افزارهایLOCALIZER (ستون داخلی)، TargetP (ستون دوم)، CholoroP (ستون سوم) و SignalP (ستون چهارم).نتایج هر نرم‌افزار به صورت ستونهایی با نقاط دایره ای شکل نمایش داده شده است. cTP (دایرهٔ توپر سبز رنگ)، mTP (دایرهٔ توپر آبی رنگ)، SP (دایرهٔ توپر ارغوانی رنگ)، Others (دایرهٔ توپر صورتی رنگ)، No cTP (دایرهٔ سبز رنگ توخالی)، NO SP (دایرهٔ ارغوانی توخالی).

بررسی فشار انتخاب طبیعی در خانوادهٔ GGPPS در گیاهان: میانگین ω در هر گروه نشان داد که فشار انتخاب طبیعی در رده‌های مختلف گیاهان به طور قابل توجهی متفاوت است. از آنجا که میانگینω در تمام گروهها بسیار پایین بود، به نظر می‌رسد که خانوادهٔ GGPPS در گیاهان به شدت تحت تأثیر انتخاب طبیعی منفی قرار دارد. اما باید توجه داشت که ممکن است انتخاب طبیعی مثبت به صورت محدود در برخی از کدونها اتفاق افتاده باشد. همچنین، مقایسهٔ مدلهای M3 وM0 نشان داد که در تمام گروهها ω به صورت غیریکنواخت درطول توالیها توزیع شده است (جدول۲). بنابراین، آزمونهای تخصصی بیشتری که تنوع بین جایگاهها را بررسی می‌کنند، مورد آزمایش قرار گرفتند. در این آزمونها مدل M2a با مدل M1a و

مدل M8 با مدل M7 مورد مقایسه قرار گرفت. آزمون نسبت درست‌نمایی (LRT) فقط در گروه بازدانگان وجود جایگاههایی با انتخاب طبیعی مثبت را تأیید نمود (جدول۲). سپس این جایگاهها با استفاده از روش بیزین (BEB) مورد شناسایی قرار گرفتند. در گروه بازدانگان ۶ جایگاه با احتمال پسین (posterior probability) بالاتر از ۹۵ درصد به عنوان جایگاههای انتخاب طبیعی مثبت شناسایی شدند (جدول۲). تمامی این نقاط در ناحیهٔ انتهای آمینی قرارگرفته اند. ناحیهٔ انتهای آمینی (حدود ۷۰ تا۹۰ اسیدآمینه) در خانوادهٔ GGPPS مسئول جایابی درون سلولی (subcellular localization) می‌باشد. این نتایج نشان می‌دهد که عملکرد جایابی درون سلولی در بازدانگان تاحدودی تحت تأثیر انتخاب طبیعی مثبت قرار دارد.

جدول۲- خلاصهٔ نتایج بررسی فشار انتخاب طبیعی در خانوادهٔ GGPPS با استفاده از مدلهای مبتنی بر جایگاه.

زیرگروه‌ها	مدل‌ها	ارزیابی پارامترها	لگاریتم درست-نمایی (lnL)	آزمون نسبت درست‌نمایی (LRT)	جایگاه‌های انتخاب طبیعی مثبت
گیاهان آبزی:					
	M0	ω=۰/۰۱۶۷۳	-۷۲۹۰/۹۳۰۰۱۸		
	M3	p_0=۰/۱۳۴۷۱ , p_1=۰/۳۴۷۶۹ , p_2= ۰/۳۳۸۱۰, p_3=۰/۱۷۳۲۸, p_4= ۰/۰۰۶۲۲, ω_0=۰/۰۰۰۰۲, ω_1=۰/۰۰۲۱۱, ω_2=۰/۰۲۱۰۹, ω_3=۰/۰۶۵۵۳, ω_4=۰/۲۷۲۲۹	-۷۲۸۷/۰۳۸۶۸۸	* ۷/۷۸۲۶۶	
	M1a	p_0=۰/۹۷۲۷۷, p_1=۰/۰۲۷۲۳, ω_0=۰/۰۱۶۰۵,ω_1=۱/۰۰۰۰۰	-۷۲۶۲/۳۰۳۷۴۰		
	M2a	p_0=۰/۹۷۲۷۷, p_1=۰/۰۲۷۲۳, p_2=۰/۰۰۰۰۰, ω_0=۰/۰۱۶۰۵, ω_1=۱/۰۰۰۰۰, ω_2=۱۰۳/۰۳۱۵۴	-۷۲۶۲/۳۰۳۷۴۰	۰/۰۰۰۰۰	
	M7	p=۰/۴۲۷۸۷, q=۱۹/۸۶۹۱۶	-۷۰۶۶/۶۵۹۸۸۷		
	M8	p_0=۰/۹۹۹۹۹, p=۰/۴۲۷۸۹, q=۱۹/۸۷۱۶۷, (p_1=۰/۰۰۰۰۱), ω=۱/۰۰۰۰۰	-۷۰۶۶/۶۶۰۶۷۸	۰/۰۰۰۷۹۱	
بازدانگان:					
	M0	ω= ۰/۳۳۰۶۳	-۷۷۶۴/۰۰۴۴۷۵		
	M3	p_0=۰/۴۶۶۶۷۶, p_1=۰/۳۴۰۱۶, p_2=۰/۰۴۳۱۸, p_3=۰/۰۷۵۵۶, p_4=۰/۰۷۴۳۳, ω_0=۰/۰۵۱۶۳, ω_1=۰/۴۱۰۷۸, ω_2=۰/۴۱۴۳۳, ω_3=۱/۴۲۱۲۹, ω_4=۱/۴۲۱۳۰	-۷۴۷۳/۱۰۲۷۸۶	* ۵۸۱/۸۰۳۳۷۸	
	M1a	p_0=۰/۶۸۷۶۹,p_1=۰/۳۱۲۳۱, ω_0=۰/۱۱۸۲۸, ω_1=۱/۰۰۰۰۰	-۷۴۸۸/۲۱۳۸۹۵		
	M2a	p_0=۰/۶۸۳۷۰, p_1=۰/۲۹۰۵۴, p_2=۰/۰۲۵۷۶, ω_0=۰/۱۲۲۰۳, ω_1=۱/۰۰۰۰۰, ω_2=۲/۳۰۳۷۷	-۷۴۸۵/۵۰۱۶۲۶	* ۵/۴۲۴۵۳۸	Y10, F23, S24, G51, V52, G56
	M7	p=۰/۳۶۶۷۵, q=۰/۶۹۳۵۵	-۷۵۷۴/۷۶۵۱۳۷		
	M8	p_0=۰/۸۷۵۷۲, p=۰/۵۹۷۰۲, q=۱/۸۶۴۰۰, (p_1=۰/۱۲۴۲۸), ω=۱/۴۹۷۹۲	-۷۵۶۱/۵۲۴۶۱۷	* ۲۶/۴۸۱۰۴	Y10, F23, S24, G51, V52, G56,
گیاهان تک‌لپه:					
	M0	ω=۰/۱۹۲۶۷	-۵۰۹۵/۲۰۹۲۶۲		
	M3	p_0=۰/۱۲۸۳۷, p_1=۰/۲۲۷۹۱, p_2=۰/۲۹۱۳۳, p_3=۰/۱۸۷۹۱, p_4=۰/۱۶۴۴۷, ω_0=۰/۰۰۴۷۱, ω_1=۰/۰۵۶۵۱, ω_2=۰/۰۵۶۵۱,	-۴۹۰۰/۴۸۹۱۸۱	* ۳۸۹/۴۴۰۱۶۲	

	M1a	ω_3=۰/۱۲۷۸۳,ω_4=۰/۱۹۴۷۹ p_0=۰/۹۸۸۸۰,p_1=۰/۰۱۱۲۰, ω_0=۰/۰۸۰۲۳, ω_1=۱/۰۰۰۰۰	−۴۹۳۲/۴۱۲۵۱۱	
	M2a	p_0=۰/۹۸۸۸۰,p_1=۰/۰۱۱۲۰,p_2=۰/۰۰۰۰۰ ω_0=۰/۰۸۰۲۳, ω_1=۱/۰۰۰۰۰, ω_2=۱۹/۸۹۸۲۶	−۴۹۳۲/۴۱۲۵۱۱	۰/۰۰۰۰۰
	M7	p=۱/۴۰۲۰۷, q=۱۴/۶۰۵۲۷	−۴۹۰۱/۱۶۹۸۵۷	
	M8	p0=۰/۹۹۹۹۹, p=۱/۲۸۴۲۳, q=۱۳/۰۸۷۰۶, (p1=۰/۰۰۰۰۱), ω=۱۵۶/۲۳۵۷۷	−۴۹۰۱/۳۶۱۱۵۴	۲/۰۰۰۰۰
گیاهان دولپه:				
	M0	ω=۰/۱۱۲۵۲	−۹۳۵۲/۰۶۶۱۸۱	
	M3	p_0=۰/۳۴۳۲۳, p_1=۰/۳۴۰۰۰, p_2=۰/۲۳۰۶۵, p_3=۰/۰۶۷۱۶, p_4=۰/۰۱۸۹۶, ω_0=۰/۰۱۲۷۸, ω_1=۰/۰۸۸۹۰, ω_2=۰/۱۹۱۳۳, ω_3=۰/۴۱۶۸۲, ω_4=۰/۷۶۶۲۶	−۹۱۸۸/۱۰۷۵۴۲	*۳۲۷/۹۱۷۲۷
	M1a	p_0=۰/۹۳۷۵۰,p_1=۰/۰۶۲۵۰, ω_0=۰/۰۹۵۵۲, ω_1=۱/۰۰۰۰۰	−۹۲۴۸/۸۴۵۱۷۴	
	M2a	p_0=۰/۹۳۷۵۰,p_1=۰/۰۴۱۴۲,p_2=۰/۰۲۱۱۰۹, ω_0=۰/۰۹۵۵۲, ω_1=۱/۰۰۰۰۰, ω_2=۱/۰۰۰۰۰	−۹۲۴۸/۸۴۵۱۷۴	۰/۰۰۰۰۰
	M7	p=۰/۶۳۹۹۵, q=۴/۲۵۹۷۸	−۹۱۹۲/۹۴۳۵۷۸	
	M8	p0=۰/۹۹۹۹۹, p=۰/۶۳۹۹۵, q=۰/۲۵۹۷۸, (p1=۰/۰۰۰۰۱), ω=۱۶/۸۰۴۲۷	−۹۱۹۲/۹۴۵۳۳۸	۰/۰۰۳۵۲

LRT=2ΔlnL

* P value < 0.05

نتیجه گیری کلی

این مطالعه نشان داد که خانوادهٔ GGPPS در گیاهان یک خانوادهٔ در حال گسترش می باشد. همچنین مشخص شد، فرآیندهای مختلف مضاعف شدن از جمله فرآیندهای مضاعف شدن کل ژنوم و تکثیر تکراری ژنها در این گسترش نقش عمده‌ای دارند. به طور کلی، فرآیند مضاعف شدن کل ژنوم به ویژه در مراحل ابتدایی تکامل سهم بیشتری در این گسترش داشته است. اما فرآیند تکثیر تکراری ژنها در پدیده‌های LSE بیشتر مشاهده می‌شود.

پدیدهٔ گسترش در این خانواده کمک می‌کند تا گیاهان بتوانند به صورت تخصصی‌تر به تولید ترپنوئیدها بپردازند. ترپنوئیدها در طیف وسیعی از گیاهان یافت شده (۱) و اغلب در پاسخ به شرایط محیطی نقش دارند. گسترش در آنزیمهای GGPPS، این خانواده را به یک خانوادهٔ بسیار انعطاف‌پذیر در سازگاری با شرایط محیطی تبدیل نموده است. از آنجا که گونه‌های پیچیده‌تر نیاز به پاسخهای سریعتر و تخصصی‌تر به تغییرات محیطی دارند، منطقی به نظر می رسد که تعداد کپیهای GGPPS در این گونه‌ها افزایش یابد، چراکه با افزایش تعدادکپیهای ژن، هر کپی قادر خواهد بود تا عملکرد اختصاصی‌تری داشته باشد. به بیان دیگر، فشار انتخاب طبیعی که به وسیلهٔ شرایط محیطی به گیاهان تحمیل شده است، باعث گسترش در خانوادهٔ GGPPS در گیاهان شده است.

بررسیهای فشار انتخاب طبیعی نشان داد که آنزیمهای GGPPS در گیاهان به طورکلی تحت تأثیر انتخاب طبیعی منفی (purifying selection) قرار دارند. باید توجه داشت- که تفکیک عملکرد (subfunctionalization) در این خانوادهٔ پروتئینی به وفور یافت می‌شود. لذا وجود انتخاب طبیعی منفی در این خانواده می تواند در اثر وقوع تفکیک عملکرد سریع درست بعد از وقایع مضاعف شدن ژن

باشد. در این حالت بقای هر دوکپی برای حفظ عملکرد ژن اجدادی لازم است و هیچ‌کدام از کپیها نمی‌توانند به مدت بسیار طولانی تحت تأثیر موتاسیونها قرار بگیرند و از فشار انتخاب طبیعی رهایی یابند (۳۵).

آنالیزهایی که برروی PSL انجام شد، نشان داد که اغلب توالیهای مورد مطالعه دارای منطقه هدف‌گیری کلرپلاستی بودند، اما در طول گسترش این خانواده، گاهی هدف‌گیری میتوکندریایی یا ER در آنزیمهای GGPPS رخ داده است. بنابراین، می‌توان نتیجه گرفت که این پروتئینها برای تولید متابولیتهای خاصی، اختصاص داده شده‌اند. به عبارت دیگر، PSR تا حدودی موجب تنوع عملکردی در این خانوادهٔ پروتئینی شده است. بنابراین در مطالعات عملکردی این امر می‌تواند مورد توجه قرار گیرد. همانطور که در برخی از مطالعات نیز بحث شده است، بررسی تنوع در جایابی پروتئین در درون سلول می‌تواند در مطالعات مهندسی ژنتیک و متابولیک مفید باشد (۸ ،۱۹ ، ۲۳، ۲۴و۳۴). همچنین، باید توجه داشت که آگاهی از دومینهای عملکردی و مهم نقش کلیدی در ایجاد سازه‌های جدید و مهندسی پروتئین و به طورکلی مطالعات عملکردی دارد، لذا امروزه توجه فراوانی به این امر وجود دارد و تحقیقات گسترده‌ای در این مورد صورت می گیرد که از جملهٔ آنها می‌توان به تحقیق جلیلی منش و همکاران اشاره نمود (۲).

منابع

۱- بختیاری. ن و احمدزاده. س. (۱۳۹۷). اورسولیک اسید: تری-ترپنوئید ۵ حلقه‌ای موجود در پوست سیب با طیف گسترده‌ای از اثرات درمانی. مجلهٔ پژوهش‌های سلولی و مولکولی. تحت چاپ .

۲- جلیلی منش, م, حداد مشهدریزه, ع, مخدومی, ع, حسین دخت, م. ر. (۱۳۹۷). ارزیابی ویژگی‌های ساختاری وعملکردی مجموعه پروتئینی Xpt به منظور توسعه روش‌های مولکولی در طراحی نسل جدید آفت کش ها, مجله پژوهش‌های سلولی و مولکولی. تحت چاپ.

3- Byun, S. A. and Singh, S. 2013. Protein subcellular relocalization increases the retention of eukaryotic duplicate genes. Genome biology and evolution, 5(12): 2402-2409.

4- Chen, K., Durand, D. and Farach-Colton, M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. Journal of Computational Biology, 7(3-4): 429-447.

5- Coman, D., Altenhoff, A., Zoller, S., Gruissem, W. and Vranová, E. 2014. Distinct evolutionary strategies in the GGPPS family from plants. Frontiers in plant science, 5.

6- Conant, G. C. and Wolfe, K. H. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews Genetics, 9(12): 938-950.

7- Demuth, J. P. and Hahn, M. W. 2009. The life and death of gene families. Bioessays, 31(1): 29-39.

8- Dudareva, N., Klempien, A., Muhlemann, J. K. and Kaplan, I. 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytologist, 198(1): 16-32.

9- Emanuelsson, O., Nielsen, H., Brunak, S. and Von Heijne, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology, 300(4): 1005-1016.

10- Emanuelsson, O., Nielsen, H. and Von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science, 8(5): 978-984.

11- Gershenzon, J. and Dudareva, N. 2007. The function of terpene natural products in the natural world. Nature chemical biology, 3(7): 408-414.

12- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U. and Putnam, N. 2011. Phytozome: a comparative platform for green plant genomics. Nucleic acids research, 40(D1): D1178-D1186.

13- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology, 59(3): 307-321.

14- Hall, T. 2013. BioEdit version 7.2. 5. Ibis Biosciences, Carlsbad, CA, USA.

15- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K. and Shiu, S.-H. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant physiology, 148(2): 993-1003.

16- Hedden, P. and Kamiya, Y. 1997. Gibberellin biosynthesis: enzymes, genes and their regulation. Annual review of plant biology, 48(1):431-460.

17- Hedges, S. B., Dudley, J. and Kumar, S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics, 22(23): 2971-2972.

18- Innan, H. and Kondrashov, F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics, 11(2): 97-108.

19- Kappers, I. F., Aharoni, A., Van Herpen, T. W., Luckerhoff, L. L., Dicke, M. and Bouwmeester, H. J. 2005. Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis Science, 309(5743): 2070-2072.

20- Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4): 772-780.

21- Kuntz, M., Römer, S., Suire, C., Hugueney, P., Weil, J., Schantz, R. and Camara, B. 1992. Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening. The Plant Journal, 2(1): 25-34.

22- Marques, A. C., Vinckenbosch, N., Brawand, D. and Kaessmann, H. 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome biology, 9(3): R54.

23- Moses, T., Pollier, J., Thevelein, J. M. and Goossens, A. 2013. Bioengineering of plant (tri) terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. New Phytologist, 200(1): 27-43.

24- Nevoigt, E. 2008. Progress in metabolic engineering of Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews, 72(3): 379-412.

25- Nowicka, B. and Kruk, J. 2010. Occurrence, biosynthesis and function of isoprenoid quinones. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1797(9): 1587-1605.

26- Okada, K., Saito, T., Nakagawa, T., Kawamukai, M. and Kamiya, Y. 2000. Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. Plant physiology, 122(4): 1045-1056.

27- Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D. and Pupko, T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic acids research, 38(suppl 2): W23-W28.

28- Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods, 8(10): 785-786.

29- Petsalaki, E. I., Bagos, P. G., Litou, Z. I. and Hamodrakas, S. J. 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics, proteomics & bioinformatics, 4(1): 48-55.

30- Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inzé, D., Mueller-Roeber, B. and Vandepoele, K. 2015. PLAZA 3.0: an access point for plant comparative genomics. Nucleic Acids Research, 43(D1): D974-D981.

31- Reams, A. B. and Neidle, E. L. 2004. Selection for gene clustering by tandem duplication. Annu. Rev. Microbiol., 58: 119-142.

32- Rogers, R. L., Cridland, J. M., Shao, L., Hu, T. T., Andolfatto, P. and Thornton, K. R. 2015. Tandem duplications and the limits of natural selection in Drosophila yakuba and Drosophila simulans', PloS one, 10(7): e0132184.

33- Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N. and Taylor, J. M. 2017.LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Scientific Reports, 7.

34- Takahashi, S., Ogiyama, Y., Kusano, H., Shimada, H., Kawamukai, M. and Kadowaki, K. 2006. Metabolic engineering of coenzyme Q by modification of isoprenoid side chain in plant. FEBS letters, 580(3): 955-959.

35- Voldoire, E., Brunet, F., Naville, M., Volff, J.-N. and Galiana, D. 2017. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. PloS one, 12(7), pp. e0180936.

36- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution, 24(8), pp. 1586-1591.

37- Zhu, X., Suzuki, K., Saito, T., Okada, K., Tanaka, K., Nakagawa, T., Matsuda, H. and Kawamukai, M. 1997. Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from Arabidopsis thaliana is localized in mitochondria. Plant molecular biology, 35(3): 331-341.

In silico Analysis of Molecular Evolution and Expansion in Geranylgeranyl diphosphate synthase Protein Family in Plants

Farahbakhsh S.[1], Nazeri S.[1] and Minuchehr Z.[2]

[1] Dept. of Biotechnology, Bu Ali Sina University, Hamedan, I.R. of Iran.

[2] Dept. of Systems Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, I.R. of Iran.

Abstract

GGPPS enzymes are responsible for the synthesis of geranyl geranyl diphosphate (GGPP), an important precursor for the production of variety terpenoids. In this study, by identifying the GGPPSs homologs UniProt database, the evolutionary process of this family was investigated using phylogenetic analysis, and the patterns of gene loss and gain over the course of evolution were inferred by parsimony-based optimization criterion. PLAZA database was used to study the role of whole genome duplication (WGD) and tandem gene duplication (TD). Results revealed that the GGPPS family is an expanding family and during plants evolution, GGPPS family has experienced expansion and contraction events in several alternate stages. Most of these expansions have been occurred as a result of the whole genome duplication (WGD) events. Nonetheless, tandem duplications (TDs) have occurred more in lineage specific expansions (LSEs). Furthermore, Protein Subcellular localization (PSL) analysis revealed that expansion process led to subfunctionalization in this protein family. Therefore, it seems that expansion in GGPPSs improve plants responses to stress and environmental stimuli, and ultimately, lead to a better adaptation to environmental conditions. Our analysis suggested that the selective Pressure in different lineages of plants is variable and the positive selection was observed only in gymnosperms.

Key words: Terpenoid, Geranylgeranyl diphosphate synthase, Whole genome duplication, Tandem duplication, lineage-specific expansion.