

Molecular Phylogeny and Functional Divergence of Uridine Diphosphate Glycosyltransferases (UGTs) in *Crocus* Genus and their Homologues in Other Plants

ABSTRACT

Crocus sativus L. is considered one of the richest sources of apocarotenoids, including crocin, picrocrocin, and safranal. The oxidative breakdown of carotenoids generates apocarotenoids. Glycosylation is the final step of crocin biosynthesis which is crucial due to its role in pigment solubility in water and changes the chemical properties and bioactivity of the molecule. In this study, the protein sequence of UGTs in saffron and their homologs in other plants were analyzed from different aspects, including phylogeny analysis and motif identification, functional divergence analysis, and structural analysis. The present study focused on UGTs responsible for primary and secondary glycosylation in crocin production and picrocrocin glycosylation in plants where crocin is found. Also, the evolutionary relationship of the UGT protein family in *Crocus* and other plants was investigated, including type I and type II functional divergence. Phylogeny analysis showed that UGTs that carry out primary glycosylation and UGTs that carry out secondary glycosylation were placed in two groups with the highest functional distance. Motifs were group-specific and amino acids with a high functional divergence coefficient were identified in those motifs, which can be attributed to the functional difference of these sequences. These findings may facilitate future research aimed at characterizing the function of these genes.

Keywords: Glycosyltransferases, UGTs, Protein phylogeny, Motif identification

فیلوژنی مولکولی و واگرایی عملکردی یوریدین دی فسفات گلیکوزیل ترانسفرازهای (UGT) موجود در جنس زعفران و همولوگ‌های آن در سایر گیاهان

چکیده

زعفران زراعی با نام علمی *Crocus sativus* L. یکی از منابع غنی از آپوکاروتنوئیدها شامل کرو سین، پیکروکروسین و سافرانال است. آپوکاروتنوئیدها از شکست اکسیداتیو کاروتنوئیدها حاصل می‌گردند. در فرآیند بیوسنتز کرو سین، مرحله نهایی گلیکوزیلایسیون است. این مرحله بسیار مهم است زیرا به رنگدانه خصوصیات حلالیت در آب را می‌بخشد و خواص شیمیایی و زیست فعالی مولکول را تغییر می‌دهد. در این مطالعه، توالی پروتئینی UGT های موجود در زعفران، همراه با همولوگ‌های آنها در سایر گیاهان از دیدگاه‌های مختلف شامل آنالیز فیلوژنی و شناسایی موتیف، آنالیز واگرایی عملکردی و آنالیز ساختاری مورد بررسی قرار گرفت. تمرکز مطالعه روی UGT های بود که مسئول گلیکوزیلایسیون اولیه و ثانویه در تولید کرو سین هستند. همچنین رابطه تکاملی خانواده پروتئین UGT در زعفران و گیاهان دیگر مورد بررسی قرار گرفت که واگرایی عملکردی نوع یک و نوع دو را شامل می‌شود. آنالیز فیلوژنی نشان داد که UGT که گلیکوزیلایسیون اولیه را برعهده دارند با UGT که گلیکوزیلایسیون ثانویه را انجام می‌دهند در دو گروه کاملاً جداگانه قرار گرفتند. بیشترین تفاوت عملکردی را نیز این دو گروه داشتند. در هر گروه موتیف‌هایی یافت شد که اختصاصی همان گروه بوده است و در این موتیف‌های اختصاصی، اسید آمینه‌هایی با ضریب واگرایی عملکردی بالا شناسایی شده که می‌توان این واحدها را به تفاوت عملکردی این توالی‌ها نسبت داد. این یافته‌ها ممکن است تحقیقات آینده را با هدف مشخص کردن عملکرد این ژن‌ها تسهیل کند.

کلید واژه‌ها: گلیکوزیل ترانسفراز، UGTs، فیلوژنی پروتئین، شناسایی موتیف

مقدمه

گلیکوزیل ترانسفرازها (GTs) آنزیم‌هایی هستند که مولکول قند را از قندهای نوکلئوتیدی به طیف وسیعی از مولکول‌ها از جمله هورمون‌ها، متابولیت‌های ثانویه و مواد شیمیایی زیستی و غیر زیستی منتقل می‌کنند. هنگامی که گلیکوزیل ترانسفرازها یک بخش قندی در هر مولکول اضافه می‌کنند، آبدوستی آن مولکول و در نتیجه خواص شیمیایی و زیست‌فعالی آن را تغییر می‌دهند (Offen et al., 2006; Vogt & Jones, 2000). این پدیده برای عملکرد مناسب موجودات زنده حیاتی است. گلیکوزیل ترانسفرازها تقریباً در همه موجودات زنده شامل گیاهان، حیوانات و میکروارگانیسم‌ها وجود دارند و در گیاهان بخشی از یک خانواده وسیع چند ژنی هستند (Lim & Bowles, 2004). فرآیند گلیکوزیلاسیون به عنوان ابزاری مهم در صنایع غذایی و داروسازی نیز عمل می‌کند و در تثبیت و حلالیت ترکیبات طبیعی مختلف نقش دارد (Ahmed et al., 2006; Mandai et al., 1992; Weymouth-Wilson, 1997).

این گروه آنزیمی بسیار قدیمی است و اخیراً تحقیقات گسترده‌ای در مورد ساختار، عملکرد و مکانیسم سلولی این دسته از آنزیم‌ها انجام شده است. این آنزیم‌ها در طیف وسیعی از گیاهان شناخته شده‌اند (Le Roy et al., 2016; Rahimi et al., 2019; M. Wang et al., 2023). اگرچه اطلاعات این آنزیم‌ها و ژن‌های مرتبط با آنها در زعفران نیز گزارش شده است با این حال هنوز جنبه‌های بسیاری از این آنزیم‌های مهم در این گیاه ارزشمند ناشناخته مانده است.

زعفران زراعی با نام علمی *Crocus sativus* L. حاوی میزان بالایی از آپوکاروتنوئیدها است. این آپوکاروتنوئیدها شامل کروسین، پیکروکروسین و سافرانال می‌باشند که به ترتیب مسئول رنگ، طعم تلخ و عطر در زعفران هستند (Caballero-Ortega et al., 2007; Winterhalter & Rouseff, 2001). کروسین‌ها (استرهای کروسین) از جمله ترکیبات آپوکاروتنوئیدی گلیکوزیله هستند که در گروه محدودی از گیاهان وجود دارند (Ahrazem et al., 2017). این آپوکاروتنوئیدهای آبدوست به دلیل خواص رنگی و دارویی به طور گسترده در صنایع غذایی و دارویی مورد استفاده قرار گرفته‌اند (Christodoulou et al., 2015). در فرآیند بیوسنتز کروسین، مرحله نهایی گلیکوزیلاسیون است. گلیکوزیلاسیون کروسین (پیش‌ماده کروسین) و کروسین‌ها توسط گلیکوزیل ترانسفرازها انجام می‌شود که بر روی گروه‌های کربوکسیل و گلیکوزیل آنها عمل می‌کنند. در نتیجه این واکنش‌ها، بر اساس تعداد مختلف گلوکوزیل پنج نوع مختلف کروسین به نام‌های کروسین-I، کروسین-II، کروسین-III، کروسین-IV و کروسین-V تشکیل می‌شود. کروسین-I، چهار، کروسین-II، سه، کروسین-III و کروسین-IV، دو و کروسین-V، یک گلوکوزیل دارند (Pu et al., 2020). در این میان کروسین-I با چهار واحد گلوکز، کروسین اصلی موجود در زعفران است (Liang et al., 2014; Xi & Qian, 2006).

پیش‌بینی واگرایی عملکردی در خانواده‌های پروتئینی و شناسایی مکان‌های بالقوه واگرایی عملکردی در مقیاس بزرگ، یکی از بالاترین اولویت‌ها در مطالعات پساژنومیک است (Chagoyen et al., 2016; Nguyen Ba et al., 2014; Swint-Kruse, 2016). به طور گسترده پذیرفته شده است که اهمیت عملکردی یک ژن را می‌توان برحسب محدودیت‌های عملکردی توالی آن پروتئین سنجید (Abhiman et al., 2006; Arnau et al., 2006; Mirny & Gelfand, 2002; Naylor & Gerstein, 2000). برای مثال، موقعیت‌های بسیار حفاظت شده در بین مکان‌های همولوگ نشان می‌دهد که آن واحدها در عملکرد مشترک آن خانواده پروتئینی نقش دارند (Bharatham et al., 2011; del Sol Mesa et al., 2003). شناسایی عملکرد یک واحد اسید آمینه‌ای که منحصر به یک هدف پروتئینی در یک خانواده ژنی است، گامی کلیدی است که می‌تواند منجر به پیشرفت‌های بیوتکنولوژی یا درمانی، مانند طراحی داروهای هدفمند¹ شود (Li et al., 2003; Umesono & Evans, 1989). از این رو توسعه روش‌های *in silico* برای پیش‌بینی

¹ Target specific medical drugs

واحدهای اسید آمینه‌ای که ممکن است در واگرایی عملکردی در میان اعضای ژنی یک خانواده معین دخیل بوده باشند مقرون به صرفه است.

در این مطالعه، توالی پروتئینی UGT های موجود در جنس زعفران، همراه با همولوگ‌های آنها در سایر گیاهان بررسی شد. تمرکز اصلی روی UGT هایی بود که مسئول گلیکوزیلاسیون اولیه و ثانویه در بیوسنتز کرو سین هستند. همچنین رابطه تکاملی خانواده پروتئین UGT در زعفران و گیاهان دیگر مورد بررسی قرار گرفت که ضرایب واگرایی عملکردی نوع یک و نوع دو را شامل می‌شود.

پیشینه تحقیق

مطالعات سیستمی جنبه‌های مهمی را در مکانیسم‌های مختلف گلیکوزیل ترانسفرازها نشان داده است (Breton *et al.*, 2012; UDP-Breton & Imberty, 1999) -گلیکوزیل ترانسفرازها² (UGTs) متعلق به خانواده گلیکوزیل ترانسفراز یک هستند و در گلیکوزیلاسیون طیف گسترده‌ای از مولکول‌های پذیرنده نقش دارند. تصور می‌شود که بیش از صد ژن کد کننده UGT در ژنوم هر گیاه وجود داشته باشد (Le Roy *et al.*, 2016). با این حال تنها تعداد محدودی از UGT ها در چند گونه گیاهی مشخص شده‌اند. برخی از UGT ها، در هر دو فرایند گلیکوزیلاسیون و دگلیکوزیلاسیون مشارکت دارند (Modolo *et al.*, 2007; Zhang *et al.*, 2006).

در گیاه زعفران ژن *CsUGT74AD1* که نسخه‌ای از *UGTCs2* را کد می‌کند، مرحله اولیه گلیکوزیلاسیون کروستین را انجام می‌دهد، که برای تشکیل کروستین ضروری است (Demurtas *et al.*, 2018a). در این گیاه *UGT91P3* نیز گلیکوزیلاسیون ثانویه را در تولید کروستین انجام می‌دهد (López-Jimenez *et al.*, 2021). از دیگر گیاهان شناخته شده برای تجمع کروستین، *Gardenia jasminoides* (Rubiaceae) و *Buddleja davidii* (Scrophulariaceae) هستند (Ahrazem *et al.*, 2017; Pfister *et al.*, 1996). در گیاه *Gardenia* آنزیم *UGT75L6* در گلیکوزیلاسیون کروستین و تبدیل آن به استرهای گلوکوزیل کروستین و *UGT94E5* در گلیکوزیلاسیون ثانویه نقش دارد (Nagatoshi *et al.*, 2012).

تجزیه و تحلیل فیلوژنتیک یک خانواده پروتئینی، روشی ارزشمند جهت شناسایی مناطق حفاظت شده و واگرا است که ممکن است بینش‌هایی را در مورد عملکردهای بالقوه ارائه دهد. مضاعف شدگی ژن و متعاقب آن، واگرایی عملکردی به عنوان منابع اصلی نوآوری تکاملی در نظر گرفته می‌شوند (Ohno, 2013).

در مطالعه‌ای که آنالیز فیلوژنی بر روی مجموعه‌ای از پروتئین‌های شبیه GGT، جهت بررسی تنوع عملکردی صورت گرفت نتایج جالبی در مورد تکامل این پروتئین‌ها بدست آمد. در این مطالعه واحدهای اسید آمینه‌ای مشخص شد که مختص هر زیر خانواده بود. این مکان‌ها نه تنها تفاوت‌هایی در تنوع کاتالیزوری و پایداری GGT های مختلف را روشن‌تر می‌کردند، بلکه می‌توانستند در طراحی بازدارنده‌های خاص در برابر GGT های خاصی کمک کننده باشند (Verma *et al.*, 2015). در مطالعه دیگری که برای بررسی تکاملی مولکولی ویژگی‌های خانواده ژن KNOX در ۱۰ گونه گیاهی مختلف انجام شد، ۱۲۹ عضو از خانواده این ژن شناسایی شد و بر اساس همردیفی چندگانه و بازسازی درخت فیلوژنتیک به دو زیرخانواده طبقه بندی شدند. در تجزیه و تحلیل واگرایی عملکردی ۱۵ مکان

² Uridine diphosphate glycosyltransferases

اسید آمینه به عنوان واگرایی عملکردی نوع یک و ۵۲ مکان به عنوان واگرایی عملکردی نوع دو شناسایی شد. این مکان‌ها می‌توانند نقش مهمی در تکامل مولکولی خانواده ژن KNOX در این گونه‌ها ایفا کنند (Meng *et al.*, 2020).

روش شناسی پژوهش

جمع آوری داده‌ها

دو مجموعه داده جمع‌آوری شدند. مجموعه اول شامل توالی‌های پروتئینی است که در جنس زعفران شناسایی شدند. تعداد ۳۰ توالی UGT پارالوگ در جنس زعفران بدست آمد. توالی‌ها از سایت NCBI (<https://www.ncbi.nlm.nih.gov/>) با استفاده از کلمه کلیدی "گلوکوزیل ترانسفراز" و نام جنس زعفران در سایت مورد جستجو قرار گرفتند (پیوست ۱). مجموعه دوم شامل توالی‌های پروتئینی همولوگ UGT است که بواسطه ابزار BLASTP و با استفاده از تعدادی از توالی‌های پروتئینی UGT موجود در جنس زعفران (UGT74, GLT2, UGT91, UGT85, UGT707) به عنوان توالی مورد سوال^۳ با جستجو در گیاهان دیگر جمع‌آوری شدند. تعداد ۳۷ توالی گردآوری شد (پیوست ۲).

آنالیز فیلوژنی و شناسایی موتیف‌ها

جهت ترسیم درخت فیلوژنی، ابتدا توالی‌های پروتئینی با استفاده از نرم افزار همردیفی چندگانه توالی به نام MAFFT با پارامترها و گزینه‌های پیش فرض، همردیف شدند. آنالیز فیلوژنتیک با استفاده از روش استنباطی بی‌بین^۴ و با کمک نرم افزار MrBayes 3.1.2 انجام شد. از نرم افزار MrModelTest برای انتخاب بهترین مدل استفاده گردید. در مجموعه اول از توالی‌ها، توالی UGT از گیاه *Freesia hybrid* و در مجموعه دوم، توالی UGT از نوعی جلبک سبز به نام *Coccomyxa* به عنوان برون گروه انتخاب گردید. جهت شناسایی نواحی موتیف در توالی‌های پروتئینی مورد بررسی از برنامه MEME (<http://meme.nbcr.net>) استفاده شد.

آنالیز واگرایی عملکردی

نرم افزار DIVERGE 3.0 جهت شناسایی واگرایی عملکردی بین اعضای خانواده پروتئینی UGT مورد استفاده قرار گرفت. همردیفی چندگانه توالی آمینواسیدی UGT، به فرمت FASTA و درخت فیلوژنی که با استفاده از نرم افزار مستربایس رسم گردید بعنوان ورودی به این نرم افزار داده شدند. آنالیز واگرایی عملکردی جهت شناسایی واگرایی عملکردی نوع یک و دو انجام شد. بر اساس ملزومات نرم افزار تنها شاخه‌هایی که چهار یا بیشتر از چهار عضو داشتند جهت مقایسه انتخاب شدند. ضریب واگرایی عملکردی نوع

³ Query

⁴ Bayesian

یک و دو (به ترتیب ΘI و ΘII) بین گروه‌های منتخب محاسبه شد. معمولاً ضریب واگرایی بین صفر و یک است و اگر بزرگتر از صفر باشد بیانگر مناسب بودن توالی جهت شناسایی واحدهای اسید آمینه‌ای مرتبط با واگرایی عملکردی است.

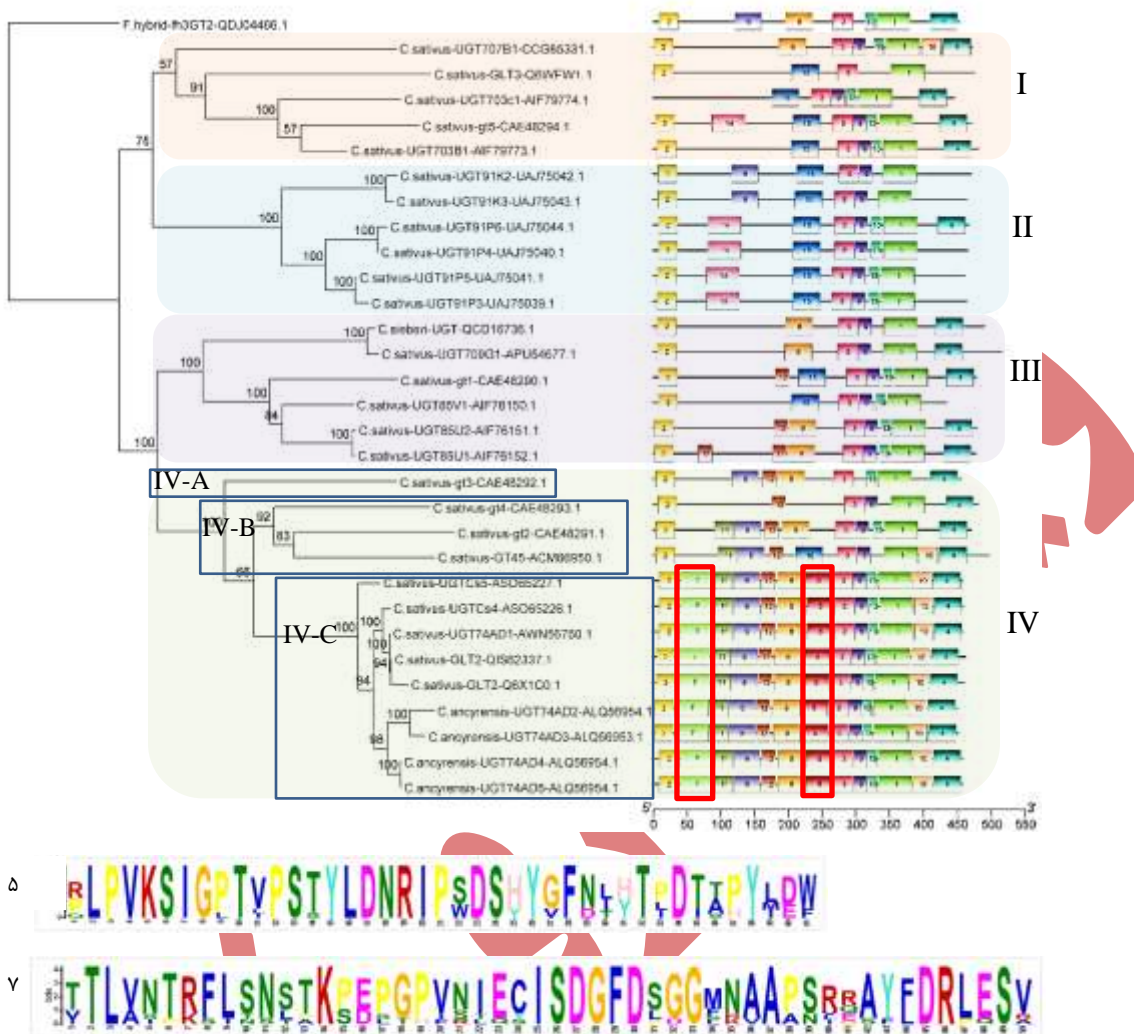
آنالیز ساختاری

از آنجایی که در طی تکامل ساختار پروتئین‌های هومولوگ نسبت به توالی آنها محافظت شده‌تر است، تفسیر عملکردی پروتئین‌ها بر اساس ساختار نسبت به تعیین عملکرد بر اساس توالی قابل اعتمادتر می‌باشد. از این رو در مطالعه حاضر به منظور آنالیز ساختاری از سرور SWISS-MODEL جهت پیش‌بینی و نرم‌افزار PyMOL جهت رسم ساختار و سرور Dali (Holm *et al.*, 2023) به منظور مقایسه میزان شباهت ساختارهای UGT های مورد نظر استفاده گردید.

یافته‌های پژوهش

آنالیز فیلوژنی و موتیف بین توالی های پارالوک UGT در جنس زعفران

در مجموعه اول بر اساس توالی‌هایی که از NCBI بدست آمد، درخت فیلوژنی ۳۰ توالی پروتئینی UGT در جنس زعفران رسم گردید. نتایج نشان داد که UGT های موجود در جنس زعفران در چهار گروه (I, II, III و IV) قرار گرفتند. توالی های UGT707 و UGT703، که هر دو در گلیکوزیلاسیون فلاونونوئیدها دخیل هستند (Trapero *et al.*, 2012; Ahrazem *et al.*, 2014) باهم در گروه I قرار گرفتند. UGT94p3 که در گلیکوزیلاسیون ثانویه نقش دارد با سایر UGT94 ها در گروه II قرار گرفت. در گروه سوم UGT85 که در تنش‌ها نقش دارد (Ahrazem *et al.*, 2015) و UGT709 که در بیوسنتز پیکروکرو سین، دیگر ترکیب اصلی زعفران که عامل طعم تلخ آن است، شرکت دارد (Diretto *et al.*, 2019) باهم گروه بندی شدند. چهارمین گروه به سه زیرگروه (IV-A, IV-B, IV-C) تقسیم شد. زیر شاخه IV-C شامل UGT74AD، GLT2 و UGTCS4 است که در گلیکوزیلاسیون اولیه مشارکت دارند (Demurtas *et al.*, 2018b; Lai *et al.*, 2022). GT45 که در گلیکوزیلاسیون فلاونونوئیدها مشارکت دارد (Moraga *et al.*, 2009) نیز در گروه چهارم در زیرگروه دیگری قرار گرفت (شکل ۱).



شکل ۱: درخت فیلوژنی توالی های پارالوگ UGT در جنس زعفران که با نرم افزار مستربایس رسم گردید به همراه موتیف‌هایی که با استفاده برنامه MEME شناسایی شد. موتیف های ۵ و ۷ که تنها در توالی‌های زیرگروه IV-C که گلیکوزیلاسیون اولیه را انجام می‌دهد یافت شده‌اند.

به منظور رسیدن به درک عمیق‌تر از عملکرد ژن، نرم افزار MEME جهت شناسایی ۱۵ موتیف بین توالی‌های پارالوگ UGT در جنس زعفران به کار برده شد. تعداد موتیف بین چهار گروه حاصل از درخت فیلوژنی متفاوت بود. موتیف‌های ۱ و ۳ در همه توالی‌های پروتئینی پارالوگ UGT جنس زعفران مشاهده شدند. موتیف ۲ و ۹ نیز در اکثر توالی‌ها حضور داشتند. در مجموع این چهار موتیف تقریباً در همه توالی‌ها حضور داشتند. در زیرگروه IV-C از گروه IV که UGT74AD و UGTCS4 در این گروه قرار داشتند، موتیف های شماره ۵ و ۷ مشاهده شد که در هیچ یک از گروه‌های دیگر حضور نداشتند و این دو موتیف مختص این گروه بودند (شکل ۱).

زمانی که تعداد موتیف در تنظیمات اولیه MEME روی ۶۰ تنظیم شد، نشان داده شد که دو موتیف ۳۹ و ۴۶ در UGT91p3 و UGT91p5 حضور دارند که در سایر UGT91 ها یافت نشدند. در مقایسه این دو موتیف بین دو توالی UGT91P3 و UGT91P5

مشخص شد که موتیف ۳۹ کاملاً توالی مشابهی را در هر دو توالی داشت، اما موتیف ۴۶ در دو جایگاه اسید آمینه ۵ و ۱۳ بین این دو توالی متفاوت بود که می‌تواند کاندیدی برای بررسی تفاوت عملکردی بین این دو پروتئین باشد (شکل ۲).

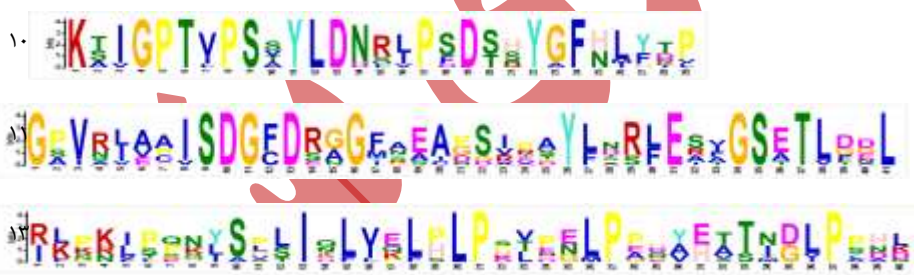
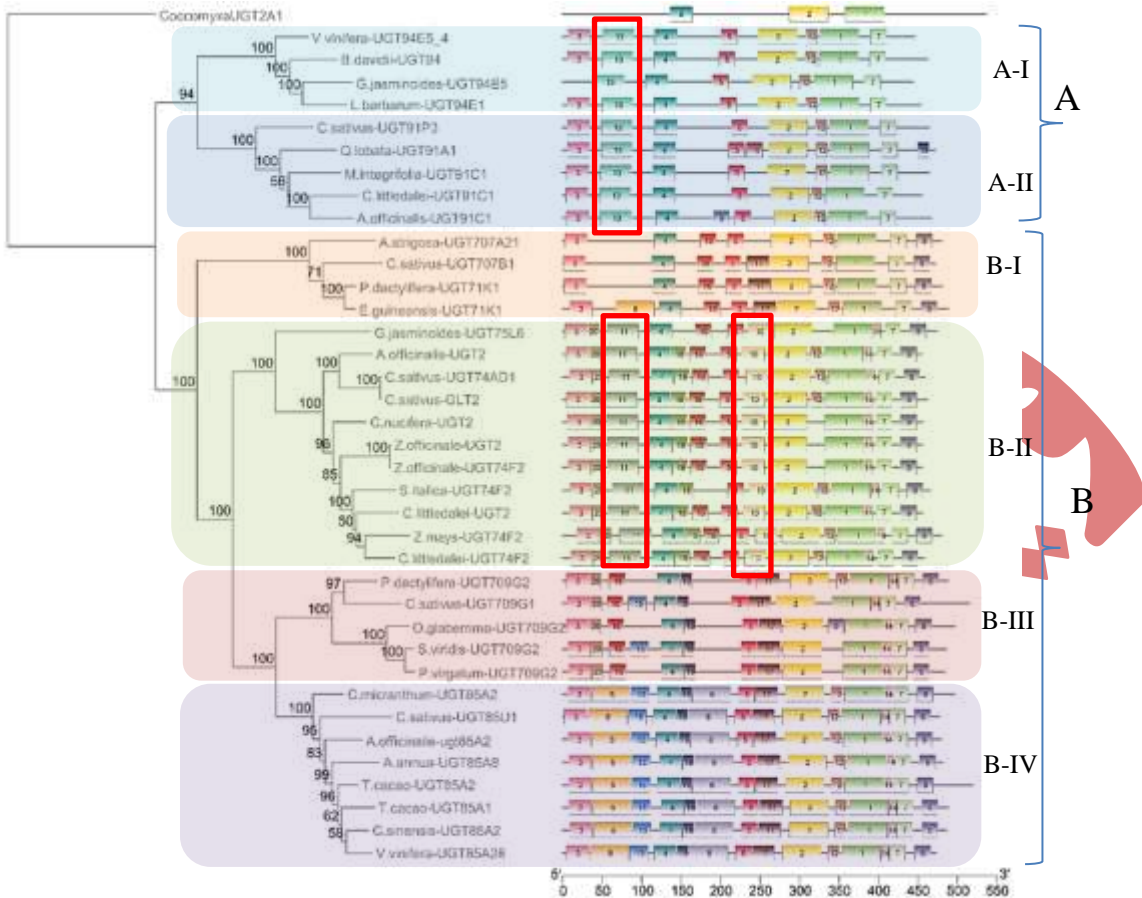


شکل ۲: قسمتی از درخت فیلوژنی مربوط به گروه II توالی‌های پارالوگ UGT در جنس زعفران شامل توالی UGT91P3 که در گلیکوزیلاسیون ثانویه کروسین نقش دارد، به همراه موتیف‌های شناسایی شده زمانی که تعداد موتیف در تنظیمات اولیه MEME روی ۶۰ تنظیم شد.

آنالیز فیلوژنی و موتیف بین توالی‌های همولوگ UGT در جنس زعفران و گیاهان دیگر

هنگامی که درخت فیلوژنی برای گروه دوم داده‌های بدست آمده از NCBI رسم شد، توالی‌ها در دو گروه کلی A و B قرار گرفتند. گروه A شامل دو زیرگروه (A-I و A-II) بود. توالی‌های *G. jasminoides*-UGT94E5 و *C. sativus*-UGT91p3 که گلیکوزیلاسیون ثانویه کروسین را انجام می‌دهند در این گروه اما در زیرگروه جداگانه قرار گرفتند. گروه B شامل چهار زیرگروه (B-I، B-II، B-III و B-IV) بود که توالی‌های *C. sativus*-GLT2، *C. sativus*-UGT74AD1، *C. sativus*-UGT75L6 که گلیکوزیلاسیون اولیه و ثانویه کروسین را انجام می‌دهد در زیر شاخه‌ای جدا ولی در همین گروه قرار گرفت. توالی *C. sativus*-UGT709G1 که گلیکوزیلاسیون پیش‌ماده پیکروکروسین را انجام می‌دهد در گروه B اما در زیرگروه جداگانه‌ای (به ترتیب B-II و B-III) قرار گرفتند.

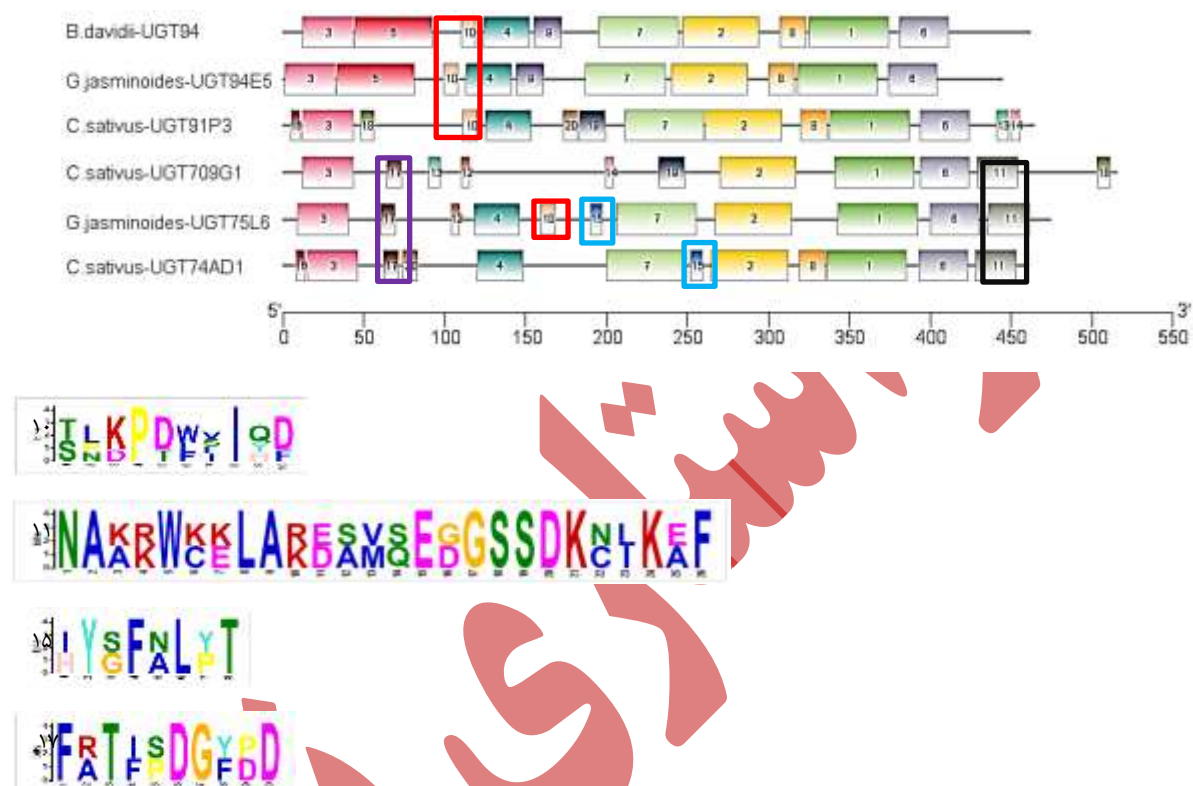
آنالیز موتیف این گروه نشان داد که موتیف‌های یک تا شش تقریباً در همه توالی‌ها حضور داشتند. موتیف‌های ۱۰ و ۱۱ تنها در زیرگروه B-II یافت شدند که UGT‌های مختص گلیکوزیلاسیون اولیه کروسین در این گروه قرار داشتند. این دو موتیف به ترتیب با موتیف‌های ۵ و ۷ حاصل از مقایسه بین پارالوگ‌های UGT در درخت قبلی همپوشانی دارند. موتیف ۱۳ تنها در گروه A یافت شد که UGT‌های مختص گلیکوزیلاسیون ثانویه در این گروه قرار داشتند. موتیف‌های ۶ و ۸ نیز مختص گروه B-IV بودند (شکل ۳).



شکل ۳: درخت فیلوژنی توالی‌های همولوگ UGT در جنس زعفران و گیاهان دیگر که با نرم افزار مستربایس رسم گردید به همراه موتیف‌هایی که با استفاده از برنامه MEME شناسایی شد. موتیف‌های ۱۰ و ۱۱ که تنها در توالی‌های زیرگروه B-II که در گلیکوزیلاسیون اولیه کروستین نقش دارند و موتیف ۱۳ که تنها در توالی‌های گروه A که گلیکوزیلاسیون ثانویه کروستین را انجام می‌دهد، یافت شدند.

هنگامی که ده موتیف فقط بین شش UGT از گیاهان *G. jasminoides*، *C. sativus* و *B. davidi* که توانایی تولید کروسین را داشتند و گلیکوزیلاسیون اولیه و ثانویه کروستین و گلیکوزیلاسیون پیش‌ماده پیکروکروسین را انجام می‌دهند، تعیین گردید، مشخص شد که چهار موتیف با شماره‌های یک، دو، سه و شش بین هر شش UGT مشترک است. در گروهی که گلیکوزیلاسیون ثانویه کروستین را انجام می‌دهند یک موتیف ۱۰ دیده شده که بین آنها مشترک است و همچنین این موتیف در *G. jasminoides*-UGT75L6 که هم در گلیکوزیلاسیون اولیه و هم ثانویه نقش دارد نیز دیده شد، اما در UGT74AD1 که گلیکوزیلاسیون اولیه کروستین را در جنس

زعفران انجام می‌دهد یافت نشد. موتیف‌های ۱۱، ۱۵ و ۱۷ در گروهی که گلیکوزیلاسیون اولیه کروستین را انجام می‌دهند مشترک بوده ولی در گروهی که گلیکوزیلاسیون ثانویه را انجام می‌دهند یافت نشد. البته موتیف ۱۱ و ۱۷ در UGT709G1 که گلیکوزیلاسیون پیش‌ماده پیکروکروسین را انجام می‌دهد نیز یافت شد (شکل ۴).



شکل ۴: موتیف بین شش UGT از گیاهان *B. davidi* و *C. sativus*، *G. jasminoides* که توانایی تولید کروسین را داشتند و گلیکوزیلاسیون اولیه و ثانویه کروسین و گلیکوزیلاسیون پیش‌ماده پیکروکروسین را انجام می‌دهند. تعداد موتیف در تنظیمات اولیه MEME روی ۱۰ تنظیم شد.

آنالیز عملکردی بین توالی‌های همولوگ UGT در جنس زعفران و گیاهان دیگر

در این مطالعه ضریب واگرایی عملکردی نوع یک بین UGT های همولوگ در جنس زعفران و گیاهان دیگر تعیین گردید (جدول ۱). ضریب واگرایی در همه مقایسات جفتی بین گروه‌ها بیشتر از صفر بود که شواهدی را جهت واگرایی عملکردی نوع یک بین گروه‌های UGT فراهم می‌کند

جدول ۱: ضریب واگرایی عملکردی نوع یک بین UGT های همولوگ در جنس زعفران و گیاهان دیگر

P-Value	تعداد واحدهای اسید آمینه مرتبط با واگرایی عملکردی ($Qk > 0.9$)	LRT ^۶	$\Theta I \pm SE^5$	گروه های مورد مقایسه
$P < 0.00001$	۴	۳۰/۷۱	-0.48 ± 0.08	B-I و A
$P < 0.00001$	۳۷	۱۶۰/۰۴	-0.60 ± 0.04	B-II و A
$P < 0.00001$	۵	۳۶/۸۶	-0.46 ± 0.07	B-III ، A
$P < 0.00001$	۱۴	۸۲/۷۴	-0.51 ± 0.05	B-IV و A
$P < 0.00001$	۱۰	۵۸/۳۲	-0.54 ± 0.07	B-II و B-I
$P < 0.00001$	-	۸/۵۵	-0.40 ± 0.10	B-III و B-I
$P < 0.00001$	۲۵	۸۵/۰۶	-0.70 ± 0.07	B-IV و B-I
$P < 0.00001$	۶	۴۴/۷۰	-0.45 ± 0.06	B-III و B-II
$P < 0.00001$	۲۹	۱۲۱/۷۷	-0.59 ± 0.05	B-IV و B-II
$P < 0.00001$	۳	۳۰/۷۲	-0.40 ± 0.07	B-IV و B-III

برخلاف ضریب واگرایی عملکردی نوع یک، ضریب واگرایی عملکردی نوع دو بین همه مقایسات جفتی بین گروه‌ها، بسیار پایین بوده است که بیانگر عدم وجود شواهد واضح از وقوع واگرایی عملکردی نوع دو بین گروه‌ها است. برای تعیین اینکه کدام گروه از نظر عملکردی متفاوت‌تر است، آنالیز فاصله عملکردی^۷ انجام شد که طول شاخه عملکردی در شکل ۵ نشان داده شد. بیشترین طول شاخه مربوط به گروهی بود که گلیکوزیلاسیون اولیه را برعهده دارند.



شکل ۵: شمایی از فاصله عملکردی بین گروه‌های UGT. طول شاخه با استفاده از DIVERGE v.3.0 تخمین زده شده است (Gu, 2013)

⁵ Standard error

⁶ Likelihood ratio test

⁷ Functional distance analysis

واحدهای اسید آمینه مهم که ممکن است مسئول واگرایی عملکردی خانواده UGT باشند با محاسبه پروفایل مختص مکان بر اساس آنالیز احتمال پسین پیش‌بینی شدند (جدول ۲). میزان احتمال پسین بالاتر از $0/9 > (QI(k))$ بعنوان حد آستانه برای شناسایی مکانهای بالقوه واگرایی عملکردی نوع یک در همه مقایسه‌های جفتی بین گروه‌های UGT استفاده شد. در بین ۳۵۰ مکان مورد تجزیه و تحلیل در همه مقایسات جفتی، ۳۷ واحد اسید آمینه ($QI(k) > 0/9$) به عنوان مسئول تفاوت عملکردی بین B-II و A پیش‌بینی شدند.

مکان‌های واگرایی پیش‌بینی شده با موتیف‌های پیش‌بینی شده توسط MEME مقایسه شد تا نقش احتمالی موتیف‌ها در تنوع عملکردی بین گروه‌ها شناسایی گردد. از ۳۷ اسید آمینه پیش‌بینی شده بین دو گروه A و B-II، ۲۱ واحد در موتیف‌ها قرار داشتند. در موتیف‌های اختصاصی شماره ۱۰ و ۱۱ متعلق به گروه B-II، به ترتیب یک (اسید آمینه شماره ۵) و چهار اسید آمینه (اسید آمینه شماره‌های ۱۰، ۱۳، ۳۴، ۳۵) و همچنین در موتیف اختصاصی ۱۳ متعلق به گروه A، ۴ اسید آمینه (اسید آمینه شماره‌های ۱۹، ۲۷، ۳۱ و ۳۷) مرتبط با واگرایی عملکردی نوع یک با احتمال پسین $0/9 <$ شناسایی شدند.

جدول ۴. موقعیت واحدهای اسید آمینه مهم در توالی هم‌ردیف شده در UGT های همولوگ در جنس زعفران و گیاهان دیگر که ممکن است مسئول واگرایی عملکردی خانواده UGT باشند به همراه شماره اسید آمینه‌ای آن در موتیف‌های مشخص شده و نام گروهی که آن موقعیت اسید آمینه‌ای در آن گروه حفاظت شده است. موقعیت‌هایی که برای آنها موتیفی ذکر نشده است در موتیف قرار نداشتند. میزان احتمال پسین بالاتر از $0/9 > (QI(k))$ بعنوان حد آستانه در نظر گرفته شد.

گروه	موقعیت اسید آمینه در توالی هم‌ردیف شده (موتیف، شماره اسید آمینه در موتیف، گروه حفاظت شده)
B-I و A	۱۰۳-۶۰-۳۱-۱۰۳-۶۰-۱۶۴ (A, ۲۷, ۱۳)
BII و A	۳۵-(B-II, ۱۳, ۳)۴۷-(B-II, ۲۴, ۳)۶۰-۶۰-۶۴-(B-II, ۱۳, ۹۵)۹۸-(A, ۱۹, ۱۳)۹۸-(B-II, ۱۰, ۱۱)۱۰۳-(A, ۲۷, ۱۳)۱۰۷-(A, ۳۱, ۱۳)۱۱۳- ۳۷-(A, ۳۵, ۱۱)۱۲۵-(B-II, ۳۵, ۱۱)۱۶۳-۱۵۴-(B-II, ۸, ۴)۱۶۶-(A, ۱۱, ۴)۱۷۵-(A, ۲۰, ۴)۱۷۸-(B-II, ۲۳, ۴)۱۸۶-۲۷۰-۲۰۰- ۲۷۴-(B-II, ۱۳, ۵)۲۹۰-(B-II, ۵, ۱۰)۳۱۸-(B-II, ۹, ۲)۳۶۳-(B-II, ۲, ۲)۳۷۵-(A, ۲۱, ۲)۳۸۱-(A, ۲۷, ۲)۳۹۱-(A, ۲۷, ۲)۴۰۵-۴۰۶- ۴۳۲-(A, ۱۲)۴۴۹-(B-II, ۵, ۱)۴۸۲-(A, ۳۸, ۱)۴۸۲-(B-II, ۵, ۱)۵۰۰-۵۰۰-۵۰۷-۵۲۴-(B-II, ۱۰, ۷)۵۲۴-(B-II, ۴, ۹)۵۵۳-(B-II, ۴, ۹)۵۵۳
B-III و A	۳۱-(A, ۱۰, ۳)۱۲۴-(A, ۲, ۱۲)۴۲۳-(A, ۸, ۱۲)۴۳۴-(A, ۱۳, ۱۲)۴۳۹
B-IV و A	۱۲۱-(A, ۸, ۱۵)۱۲۵-(B-IV, ۱۲, ۱۵)۱۳۲-(B-IV, ۱۸, ۱۵)۱۸۳-(B-IV, ۲۸, ۴)۱۸۳-(B-IV, ۲۳, ۶)۲۴۰-(B-IV, ۴۵, ۶)۲۵۷- ۳۵۵-(B-IV, ۱, ۲)۳۹۵-(A, ۴۱, ۲)۳۹۶-(A, ۴۲, ۲)۳۹۶-(A, ۷, ۱۲)۴۳۳-(A, ۷, ۱۲)۴۴۹-(B-IV, ۵, ۱)۴۴۹-۵۰۷-۵۴۶-۵۶۹
B-II و B-I	۴۴-(B-II, ۲۲, ۲)۴۴-(B-II, ۱۳, ۱۱)۱۰۱-(B-II, ۳۵, ۱۱)۱۲۵-(B-II, ۹, ۴)۱۶۴-(B-II, ۹, ۴)۱۶۴-(B-II, ۴۶, ۲)۴۰۰-(B-II, ۴۶, ۲)۴۰۰-(B-I, ۶, ۱۲)۴۳۲-(B-I, ۶, ۱۲)۴۳۵- ۴۸۰-(B-I, ۹)۴۸۰-(B-II, ۳۶, ۱)۴۹۸-(B-II, ۴, ۱۴)۴۹۸-(B-II, ۴, ۱۴)۴۹۸-۵۰۰
B-III و B-I	-

۴)۱۶۴-۱۵۰-(B-IV، ۲۲، ۱۵)۱۳۶-(B-IV، ۱۲، ۱۵)۱۲۵-(B-IV، ۳۷، ۸)۹۸-(B-IV، ۱۴، ۸)۶۷-۵۴-۵۳-(B-IV، ۱۷، ۳)۳۹ ۹ (B-IV)۱۸۲-(B-I، ۲۷، ۴)۱۸۲-(B-I، ۲۷-۱۹۴-۲۶۷-(B-IV، ۳۴، ۲)۳۸۸-(B-IV، ۳۴، ۲)۳۹۴-(B-I، ۴۰، ۲)۳۹۵-(B-I، ۴۱، ۲)۳۹۶-(B-II، ۴۲، ۲)۴۰۳-۴۰۴ (B-IV، ۲۰، ۹)۵۷۱-(B-IV، ۱۸، ۹)۵۶۹-(B-IV، ۲، ۹)۵۵۱-(B-I، ۱۱، ۷)۵۲۵-(B-IV، ۵، ۷)۵۱۹-۵۰۴-۴۰۶-۴۰۴	B-I و B-IV
۳۳۹-۴۲۳-۲۳۹-(B-II، ۳۴، ۱۱)۱۲۴-۱۰۷-(B-II، ۱۷، ۳)۳۹	B-II و B-III
-(B-II، ۲۷، ۴)۱۸۲-(B-IV، ۲۰، ۴)۱۷۵-۱۱۳-(B-IV، ۴۸، ۸)۱۱۰-(B-IV، ۹، ۱۶)۹۳-(B-IV، ۱۴، ۸)۶۷-(B-IV، ۱۱، ۸)۶۴ ۲۵۵ (B-IV، ۴۳، ۶)۲۸۱-(B-IV، ۴، ۵)۲۸۱-(B-IV، ۳۹۶ (B-II، ۴۲، ۲)۳۹۶-(B-IV، ۴۰۶-۴۳۵-۴۴۳-۴۴۳-(B-II، ۹، ۱)۴۵۳-(B-II، ۴۰، ۱)۴۶۴ ۳۸ (B-IV، ۳۸ (B-IV، ۴۳، ۶)۲۸۱-(B-II، ۴، ۱۴)۴۹۸-(B-IV، ۵، ۷)۵۱۹-(B-IV، ۵، ۷)۵۲۵-(B-II، ۱۱، ۷)۵۲۸-(B-IV، ۱۴، ۷)۵۲۸-۵۵۱-۵۴۶-۵۴۵ ۵۷۱-(B-II)	B-II و B-IV
(B-IV، ۲، ۱۲)۴۲۳-(B-III، ۱۰، ۱۹)۱۹۴-(B-IV، ۱۷، ۳)۳۹	B-III و B-IV

آنالیز ساختار

در مطالعه حاضر نشان داده شده است که ساختارهای پروتئینی ۳ تا ۱۰ برابر بیشتر از توالی‌های اسید آمینه حفظ شده‌اند (Illergård *et al.*, 2009). بررسی شباهت ساختاری شش تا از UGT های منتخب که گلیکوزیل‌اسیون اولیه و ثانویه کروستین و گلیکوزیل‌اسیون پیش‌ماده پیکروکروسین را در سه گیاه *Crocus*، *Budleja* و *Gardenia* بر عهده دارند نشان داد که به طور کلی آنهایی که در گلیکوزیل‌اسیون اولیه و ثانویه در تولید کروستین نقش دارند از نظر ساختاری نیز متفاوت هستند. UGT91P3 در زعفران از نظر ساختاری با UGT94 موجود در *Gardenia* و *Budleja* متفاوت است و در بین این دو گونه، میزان بسیار کمی به *Budleja* شبیه‌تر است و از نظر ساختاری شباهتی با آنزیم‌هایی که گلیکوزیل‌اسیون اولیه را انجام می‌دهند، ندارد. کمترین شباهت بین UGT91P3 در زعفران با UGT709G1 که در گلیکوزیل‌اسیون پیکروکروسین در گیاه زعفران مشارکت دارد وجود دارد. در کل UGT709G1 از نظر ساختاری با آنهایی که گلیکوزیل‌اسیون اولیه را انجام می‌دهند شبیه‌تر از آنهایی هست که در گلیکوزیل‌اسیون ثانویه شرکت می‌کنند (شکل ۶).

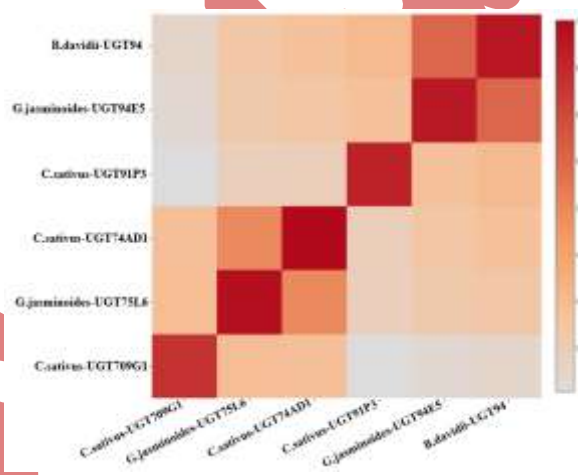
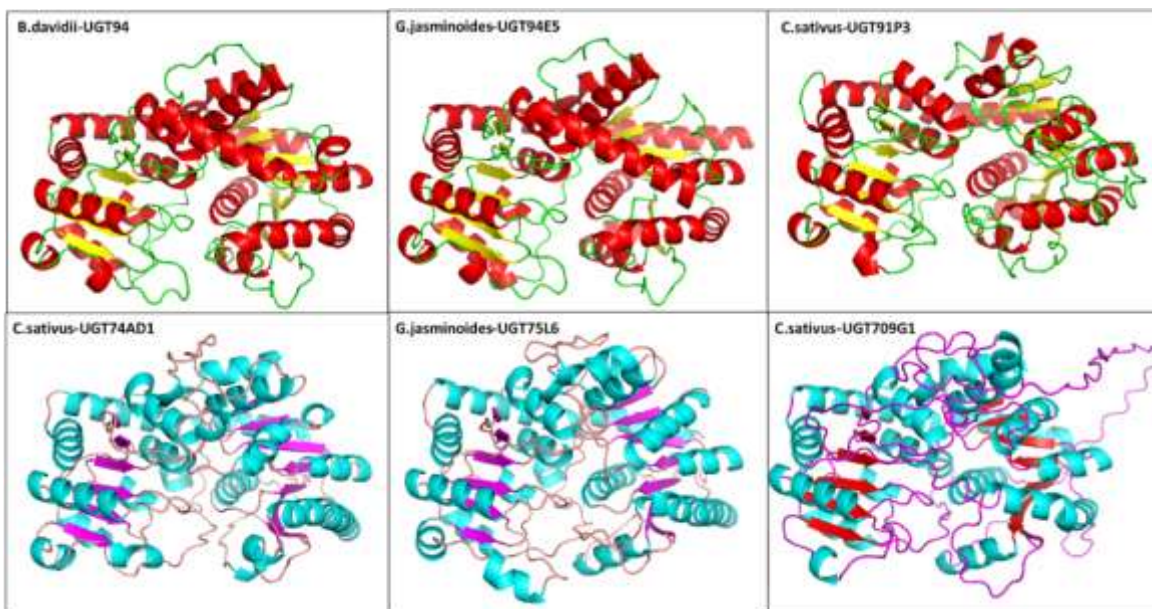
بحث

هزاران نوع متابولیت ثانویه گلیکوزیدی در گیاهان وجود دارند. اما راهبردها و آنزیم‌هایی که در فرآیند گلیکوزیل‌اسیون نقش دارند، هنوز به طور کامل شناخته نشده‌اند. در مورد فیلوژنی UGT ها مطالعات زیادی انجام شده است (Wei et al., 2021; Yu et al., 2017; Wilson & Tian, 2019). مطالعه بر روی UGT جنس زعفران با تمرکز بر تفاوت UGT های مسئول گلیکوزیل‌اسیون اولیه و ثانویه در تولید کروستین تاکنون انجام نشده است. آنالیز فیلوژنی و شناسایی موتیف‌ها در توالی‌های پارالوگ UGT در جنس زعفران و گونه‌های همولوگ دیگر بررسی شده است. نتایج نشان دادند که از رسم درخت فیلوژنی توالی‌های پارالوگ جنس زعفران چهار گروه و از درخت فیلوژنی حاصل از بررسی همزمان توالی‌های UGT جنس زعفران همراه با توالی‌های همولوگ گیاهان دیگر ۶ شش گروه حاصل شد. در هر دو درخت فیلوژنی گروه گلیکوزیل‌اسیون اولیه و ثانویه از همان ابتدا در گروه کاملاً متمایز قرار گرفتند که نشان دهنده تفاوت این دو گروه از هم می‌باشد. با استفاده از نرم‌افزار MEME، ۱۵ موتیف مشترک بین این توالی‌ها شناسایی شده است که در تمایز و عملکرد این پروتئین‌ها نقش دارند. بر اساس موقعیت و حضور موتیف‌ها در توالی‌ها، نتیجه گرفته شده که توالی‌هایی که در گروه‌های مشابه قرار می‌گیرند، احتمالاً عملکردهای مشابهی دارند. از طرف دیگر، توالی‌هایی که در گروه‌های مختلف قرار دارند،

عملکردهای متفاوتی انجام می‌دهند، که این تفاوت‌ها به واسطه وجود موتیف‌های مختلف در آنها می‌تواند توجیه شود. زمانی که UGT های زعفران همراه با سایر گیاهان مورد بررسی قرار گرفت موتیف‌های ۱۰ و ۱۱ در گروه گلیکوزیلاسیون اولیه یافت شدند با موتیف‌های شماره ۵ و ۷ در گروه گلیکوزیلاسیون اولیه حاصل از آنالیز اول همپوشانی داشت که می‌تواند نشان دهنده اهمیت این موتیف‌ها و نقش احتمالی این موتیف‌ها در اختصاصی بودن عملکرد این گروه باشد. بر اساس پژوهش‌های قبلی در حالی که ناحیه C -ترمینال UGT ها اغلب با دهنده گروه قندی در ارتباط است، ناحیه - N ترمینال در تشخیص پیش‌ماده گیرنده گروه قندی تعامل دارد و همچنین ساختار کریستالی UGT ها نشان داد که ناحیه - N ترمینال نسبت به ناحیه - C ترمینال کمتر حفاظت شده است که با تنوع پذیرنده‌های UGT ها همخوانی دارد. (Akere et al., 2020; Albesa-Jové et al., 2014). در مطالعه حاضر موتیف اختصاصی ۷ (آنالیز اول) یا ۱۱ (آنالیز دوم) در گروه گلیکوزیلاسیون اولیه و موتیف اختصاصی ۱۳ (آنالیز دوم) در گروه گلیکوزیلاسیون ثانویه در ناحیه -N ترمینال قرار دارند که با توجه به مطالعات قبلی می‌توانند در اختصاصیت پذیرش پیش‌ماده در مقایسه بین دو گروه گلیکوزیلاسیون اولیه و ثانویه نقش داشته باشند.

مکان‌هایی از توالی که به واگرایی عملکردی نوع یک نسبت داده می‌شوند شامل مکان‌هایی است که در یک گروه بسیار حفاظت شده است اما در گروه دیگر متغیر است. این مکان‌ها نشان دهنده تغییرات در نرخ تکاملی بین دو گروه مورد بررسی هستند. آنالیز واگرایی عملکردی نوع دو، مکان‌هایی که در هر دو گروه حفاظت شده هستند و اما خصوصیات بیوشیمیایی کاملاً متفاوتی دارند را مشخص می‌کند. (Gu, 2003, 2006) بخاطر اینکه در این روش از توالی اسید آمینه استفاده می‌شود، یکی از مزیت‌های آن این است که به اشیاء مکان‌های مترادف - اساس نیست و تغییرات بیشتری که در مکان‌های مترادف قابل تشخیص و تحلیل هستند تأثیر کمتری بر این روش می‌گذارند. (Cao, 2012) در مطالعه حاضر ضریب واگرایی عملکردی نوع یک به میزان قابل توجهی بیشتر از صفر بوده اما ضریب واگرایی عملکردی نوع دو بسیار پایین بوده است. این نتایج بیانگر الگوی غالب واگرایی عملکردی نوع یک برای تمایز بین زیرخانواده‌های مورد بررسی است و نیروی انتخابی مختص مکان قابل توجهی باید در بیشتر ژن‌های UGT نقش داشته باشد که منجر به تکامل عملکرد زیرخانواده خاص پس از واگرایی می‌شود. در مطالعه حاضر برخی از مکان‌های واگرایی پیش‌بینی شده، بخشی از موتیف‌های شناسایی شده هستند. این نتایج اهمیت این مکان‌ها را در ایجاد تفاوت بین ویژگی‌های عملکردی آنها تقویت می‌کند (Meng et al., 2020). در فاصله عملکردی بیشترین طول شاخه مربوط به گروهی بود که گلیکوزیلاسیون اولیه را انجام می‌دادند. طول شاخه بیشتر نشان دهنده محدودیت عملکردی تغییر یافته قابل توجه در آن گروه است و اما هر چه طول شاخه کمتر باشد بیانگر این می‌باشد که نرخ تکامل در هر مکان در آن گروه مضاعف شده با ژن اجدادی شباهت بیشتری دارد (Tine et al., 2012). بنابراین گروه گلیکوزیلاسیون اولیه در مقایسه با ثانویه تفاوت بیشتری با ژن اجدادی دارند.

ساختارهای کریستالی UGT های گیاهی، اساس ساختاری را برای فهم مکانیسم کاتالیزوری و اختصاصیت پیش‌ماده فراهم کرده است. مهندسی UGT بر اساس ساختار، می‌تواند اختصاصیت پیش‌ماده را تغییر دهد و یا کارایی کاتالیزوری را افزایش یا کاهش دهد (Wang, 2009). در مطالعه حاضر نشان داده شد که دو گروه گلیکوزیلاسیون اولیه و ثانویه از نظر ساختاری نیز با هم متفاوت هستند و UGT709G1 که در گلیکوزیلاسیون پیش‌ماده پیکروکروسین نقش دارد به گروه گلیکوزیلاسیون اولیه شباهت بیشتری دارد. بنابراین این احتمال وجود دارد UGT هایی که در گلیکوزیلاسیون اولیه تولید کروسین نقش دارند، بتوانند گلیکوزیلاسیون پیکروکروسین را نیز انجام دهند.



شکل ۶: شباهت ساختاری شش تا از UGT های منتخب که گلیکوزیلاسیون اولیه و ثانویه کروسین و گلیکوزیلاسیون پیش ماده پیکروکروسین را در سه گیاه *Crocus*, *Gardenia*, *Budleja* بر عهده دارند.

نتیجه گیری

گلیکوزیلاسیون به عنوان یکی از مهمترین واکنش های فیزیولوژیکی و بیوشیمیایی، نقش حیاتی و برجسته ای در بسیاری از فرآیندهای ضروری دارد. این اهمیت ذاتی باعث شده تا تحقیقات گسترده ای در زمینه ژن های GT صورت بگیرد تا بتوان از آنها در مهندسی متابولیک و بیوسنتز محصولات طبیعی بهره برداری بهتری داشت. علیرغم شناسایی چندین GT دخیل در گلیکوزیلاسیون، عموماً فعالیت کاتالیزوری کم و ویژگی های پیش ماده سختگیرانه این آنزیم ها به عنوان عاملی محدود کننده در سیستم های صنعتی محسوب می شود. آخرین مرحله تولید کروسین که از ترکیبات اصلی و مهم زعفران می باشد، فرایند گلیکوزیلاسیون است.

کروسین‌ها از نظر تعداد گروه قندی که به آنها متصل می‌شوند متفاوت بوده و این بر روی فعالیت و خاصیت درمانی آنها نیز موثر است. لذا جهت تولید صنعتی این ترکیب نیاز است شناخت عمیق تری نسبت به این آنزیم‌ها، نقش و عملکرد آنها داشته باشیم. مطالعه حاضر پژوهشی بنیادی به روش *in silico* است که از طریق مقایسه ساختار و توالی‌های UGT در گیاه زعفران و سایر گیاهان، رابطه فیلوژنی و نیز واگرایی عملکردی آنها را بررسی می‌کند. تمرکز اصلی پژوهش بر روی UGT هایی بوده است که گلیکوزیلاسیون اولیه و ثانویه در تولید کروسین‌ها را بر عهده دارند و هدف یافتن اسید آمینه‌هایی بوده است که مرتبط با واگرایی عملکردی در این دو گروه هستند. طی این مطالعه موتیف‌های اختصاصی هر گروه و اسید آمینه‌های مرتبط با واگرایی عملکردی که در این موتیف‌ها قرار دارند یافت شد که می‌تواند زمینه‌ساز مطالعات کاربردی بعدی باشد. اهمیت پیش‌بینی واحدهای اسید آمینه عملکردی در یک پروتئین مشهود است، زیرا این واحدها می‌توانند در طراحی عملکردی جدید، تغییر ویژگی‌ها، تعریف خانواده‌ها و زیرخانواده‌های پروتئینی، یا شناسایی وقوع یک نوآوری عملکردی مشارکت کنند. داده‌کاو و مهندسی GT ها هنوز امیدوارکننده ترین رویکردها برای کشف و توسعه آنزیم‌های جدید با ویژگی‌های کاملاً مشخص هستند. ساختارهای کریستالی که با روش‌های آزمایشگاهی تایید می‌شوند یا با روش‌های هوش مصنوعی با دقت بالا مثل AlphaFold2 (Tunyasuvunakool *et al.*, 2021) و Rose TTAfold (Pennisi, 2021) پیش‌بینی می‌شوند، روشن شدن رابطه ساختار-عملکرد را تسریع خواهند کرد. به طور خاص، این شامل شناسایی موتیف‌ها و دامین‌هایی است که بر ساختار پروتئین و فعالیت آنزیمی آن تأثیر می‌گذارد که در طراحی منطقی UGT ها کمک خواهد کرد.

فهرست منابع

- Abhiman, S., Daub, C. O., & Sonnhammer, E. L. L. (2006). Prediction of function divergence in protein families using the substitution rate variation parameter alpha. *Molecular Biology and Evolution*, 23(7), 1406–1413.
- Ahmed, A., Peters, N. R., Fitzgerald, M. K., Watson, J. A., Hoffmann, F. M., & Thorson, J. S. (2006). Colchicine glycorandomization influences cytotoxicity and mechanism of action. *Journal of the American Chemical Society*, 128(44), 14224–14225.
- Ahrazem, O., Diretto, G., Argandoña, J., Rubio-Moraga, Á., Julve, J. M., Orzáez, D., Granell, A., & Gómez-Gómez, L. (2017). Evolutionarily distinct carotenoid cleavage dioxygenases are responsible for crocetin production in *Buddleja davidii*. *Journal of Experimental Botany*, 68(16), 4663–4677.
- Ahrazem, O., Rubio-Moraga, A., Mozos, A. T., & Gómez-Gómez, M. L. (2014). Genomic organization of a UDP-glucosyltransferase gene determines differential accumulation of specific flavonoid glucosides in tepals. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 119, 227–245.
- Ahrazem, O., Rubio-Moraga, A., Trapero-Mozos, A., Climent, M. F. L., Gómez-Cadenas, A., & Gómez-Gómez, L. (2015). Ectopic expression of a stress-inducible glycosyltransferase from saffron enhances salt and oxidative stress tolerance in *Arabidopsis thaliana* while alters anchor root formation. *Plant Science*, 234, 60–73.
- Akere, A., Chen, S. H., Liu, X., Chen, Y., Dantu, S. C., Pandini, A., Bhowmik, D., & Haider, S. (2020). Structure-based enzyme engineering improves donor-substrate recognition of *Arabidopsis thaliana* glycosyltransferases. *Biochemical Journal*, 477(15), 2791–2805.
- Albesa-Jové, D., Giganti, D., Jackson, M., Alzari, P. M., & Guerin, M. E. (2014). Structure–function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology*, 24(2), 108–124.

- Arnau, V., Gallach, M., Lucas, J. I., & Marín, I. (2006). UVPAR: fast detection of functional shifts in duplicate genes. *BMC Bioinformatics*, 7(1), 1–12.
- Bharatham, K., Zhang, Z. H., & Mihalek, I. (2011). Determinants, discriminants, conserved residues—a heuristic approach to detection of functional divergence in protein families. *PLoS One*, 6(9), e24382.
- Breton, C., Fournel-Gigleux, S., & Palcic, M. M. (2012). Recent structures, evolution and mechanisms of glycosyltransferases. *Current Opinion in Structural Biology*, 22(5), 540–549.
- Breton, C., & Imbert, A. (1999). Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology*, 9(5), 563–571.
- Caballero-Ortega, H., Pereda-Miranda, R., & Abdullaev, F. I. (2007). HPLC quantification of major active components from 11 different saffron (*Crocus sativus* L.) sources. *Food Chemistry*, 100(3), 1126–1131.
- Cao, J. (2012). *The pectin lyases in Arabidopsis thaliana: evolution, selection and expression profiles*.
- Chagoyen, M., García-Martín, J. A., & Pazos, F. (2016). Practical analysis of specificity-determining residues in protein families. *Briefings in Bioinformatics*, 17(2), 255–261.
- Christodoulou, E., Kadoglou, N. P. E., Kostomitsopoulos, N., & Valsami, G. (2015). Saffron: a natural product with potential pharmaceutical applications. *Journal of Pharmacy and Pharmacology*, 67(12), 1634–1649.
- del Sol Mesa, A., Pazos, F., & Valencia, A. (2003). Automatic methods for predicting functionally important residues. *Journal of Molecular Biology*, 326(4), 1289–1302.
- Demurtas, O. C., Frusciante, S., Ferrante, P., Diretto, G., Azad, N. H., Pietrella, M., Aprea, G., Taddei, A. R., Romano, E., & Mi, J. (2018a). Candidate enzymes for saffron crocin biosynthesis are localized in multiple cellular compartments. *Plant Physiology*, 177(3), 990–1006.
- Demurtas, O. C., Frusciante, S., Ferrante, P., Diretto, G., Azad, N. H., Pietrella, M., Aprea, G., Taddei, A. R., Romano, E., & Mi, J. (2018b). Candidate enzymes for saffron crocin biosynthesis are localized in multiple cellular compartments. *Plant Physiology*, 177(3), 990–1006.
- Diretto, G., Ahrazem, O., Rubio-Moraga, A., Fiore, A., Sevi, F., Argandoña, J., & Gómez-Gómez, L. (2019). UGT709G1: a novel uridine diphosphate glycosyltransferase involved in the biosynthesis of picrocrocin, the precursor of safranal in saffron (*Crocus sativus*). *New Phytologist*, 224(2), 725–740.
- Gu, X. (2003). Functional divergence in protein (family) sequence evolution. *Origin and Evolution of New Gene Functions*, 133–141.
- Gu, X. (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Molecular Biology and Evolution*, 23(10), 1937–1945.
- Holm, L., Laiho, A., Törönen, P., & Salgado, M. (2023). DALI shines a light on remote homologs: One hundred discoveries. *Protein Science*, 32(1), e4519.
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*, 77(3), 499–508.
- Lai, C., Yang, N., Yusuyin, M., Zhang, D., Yang, Y., Li, C., & Xu, H. (2022). *Characterization of a novel crocetin glycosyltransferase UGTCs4 involved in two steps of glycosylation in crocin biosynthesis from crocus cultured cell*.
- Le Roy, J., Huss, B., Creach, A., Hawkins, S., & Neutelings, G. (2016). Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. *Frontiers in Plant Science*, 7, 735.

- Li, L., Shakhnovich, E. I., & Mirny, L. A. (2003). Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proceedings of the National Academy of Sciences*, *100*(8), 4463–4468.
- Lim, E., & Bowles, D. J. (2004). A class of plant glycosyltransferases involved in cellular homeostasis. *The EMBO Journal*, *23*(15), 2915–2922.
- López-Jimenez, A. J., Frusciante, S., Niza, E., Ahrazem, O., Rubio-Moraga, Á., Direccion, G., & Gómez-Gómez, L. (2021). A new glycosyltransferase enzyme from family 91, UGT91P3, is responsible for the final glucosylation step of crocins in saffron (*Crocus sativus* L.). *International Journal of Molecular Sciences*, *22*(16), 8815.
- Mandai, T., Yoneyama, M., Sakai, S., Muto, N., & Yamamoto, I. (1992). The crystal structure and physicochemical properties of L-ascorbic acid 2-glucoside. *Carbohydrate Research*, *232*(2), 197–205.
- Meng, L., Liu, X., He, C., Xu, B., Li, Y., & Hu, Y. (2020). Functional divergence and adaptive selection of KNOX gene family in plants. *Open Life Sciences*, *15*(1), 346–363.
- Mirny, L. A., & Gelfand, M. S. (2002). Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors. *Journal of Molecular Biology*, *321*(1), 7–20.
- Modolo, L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X., & Dixon, R. A. (2007). A functional genomics approach to (iso) flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Molecular Biology*, *64*, 499–518.
- Moraga, Á. R., Mozos, A. T., Ahrazem, O., & Gómez-Gómez, L. (2009). Cloning and characterization of a glycosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation. *BMC Plant Biology*, *9*(1), 1–16.
- Nagatoshi, M., Terasaka, K., Owaki, M., Sota, M., Inukai, T., Nagatsu, A., & Mizukami, H. (2012). UGT75L6 and UGT94E5 mediate sequential glucosylation of crocetin to crocin in *Gardenia jasminoides*. *FEBS Letters*, *586*(7), 1055–1061.
- Naylor, G. J. P., & Gerstein, M. (2000). Measuring shifts in function and evolutionary opportunity using variability profiles: a case study of the globins. *Journal of Molecular Evolution*, *51*, 223–233.
- Nguyen Ba, A. N., Strome, B., Hua, J. J., Desmond, J., Gagnon-Arsenault, I., Weiss, E. L., Landry, C. R., & Moses, A. M. (2014). Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. *PLoS Computational Biology*, *10*(12), e1003977.
- Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., & Davies, G. J. (2006). Structure of a flavonoid glycosyltransferase reveals the basis for plant natural product modification. *The EMBO Journal*, *25*(6), 1396–1405.
- Ohno, S. (2013). *Evolution by gene duplication*. Springer Science & Business Media.
- Pennisi, E. (2021). *Protein structure prediction now easier, faster*. American Association for the Advancement of Science.
- Pfister, S., Meyer, P., Steck, A., & Pfander, H. (1996). Isolation and structure elucidation of carotenoid-glycosyl esters in gardenia fruits (*Gardenia jasminoides* Ellis) and saffron (*Crocus sativus* Linne). *Journal of Agricultural and Food Chemistry*, *44*(9), 2612–2615.
- Rahimi, S., Kim, J., Mijakovic, I., Jung, K.-H., Choi, G., Kim, S.-C., & Kim, Y.-J. (2019). Triterpenoid-biosynthetic UDP-glycosyltransferases from plants. *Biotechnology Advances*, *37*(7), 107394.

- Swint-Kruse, L. (2016). Using evolution to guide protein engineering: the devil is in the details. *Biophysical Journal*, 111(1), 10–18.
- Tine, M., Kuhl, H., Jastroch, M., & Reinhardt, R. (2012). Genomic characterization of the European sea bass *Dicentrarchus labrax* reveals the presence of a novel uncoupling protein (UCP) gene family member in the teleost fish lineage. *BMC Evolutionary Biology*, 12, 1–15.
- Trapero, A., Ahrazem, O., Rubio-Moraga, A., Jimeno, M. L., Gómez, M. D., & Gómez-Gómez, L. (2012). Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sopherosides in *Crocus sativus*. *Plant Physiology*, 159(4), 1335–1354.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., & Laydon, A. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590–596.
- Umesono, K., & Evans, R. M. (1989). Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell*, 57(7), 1139–1146.
- Verma, V. V., Gupta, R., & Goel, M. (2015). Phylogenetic and evolutionary analysis of functional divergence among Gamma glutamyl transpeptidase (GGT) subfamilies. *Biology Direct*, 10(1), 1–21.
- Vogt, T., & Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in Plant Science*, 5(9), 380–386.
- Wang, M., Ji, Q., Lai, B., Liu, Y., & Mei, K. (2023). Structure-function and engineering of plant UDP-glycosyltransferase. *Computational and Structural Biotechnology Journal*.
- Wang, X. (2009). Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Letters*, 583(20), 3303–3309.
- Wei, Y., Mu, H., Xu, G., Wang, Y., Li, Y., Li, S., & Wang, L. (2021). Genome-wide analysis and functional characterization of the UDP-glycosyltransferase family in grapes. *Horticulturae*, 7(8), 204.
- Weymouth-Wilson, A. C. (1997). The role of carbohydrates in biologically active natural products. *Natural Product Reports*, 14(2), 99–110.
- Wilson, A. E., & Tian, L. (2019). Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. *The Plant Journal*, 100(6), 1273–1288
- Winterhalter, P., & Rouseff, R. L. (2001). *Carotenoid-derived aroma compounds*. ACS Publications.
- Yu, J., Hu, F., Dossa, K., Wang, Z., & Ke, T. (2017). Genome-wide analysis of UDP-glycosyltransferase super family in *Brassica rapa* and *Brassica oleracea* reveals its evolutionary history and functional characterization. *Bmc Genomics*, 18, 1–18.
- Zhang, C., Griffith, B. R., Fu, Q., Albermann, C., Fu, X., Lee, I.-K., Li, L., & Thorson, J. S. (2006). Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science*, 313(5791), 1291–1294.

Abstract

Glycosyltransferases catalyze the transfer of a sugar residue from nucleotide-sugar donors to various acceptor molecules. These enzymes change the hydrophilicity, chemical properties, and bioactivity of the molecules. The process of glycosylation is critical in the food and pharmaceutical industries. Prediction of functional residues in a protein is crucial because these residues can be attributed to new functions, change the protein properties, define

protein families and subfamilies, or identify the occurrence of an innovation. A phylogenetic analysis of a protein family and functional divergence analysis is valuable for identifying conserved and divergent regions that may offer insights into potential functions.

Material and Methods

In the current study, we placed particular emphasis on UGTs (UDP-glucuronosyltransferases), which play a critical role in both primary and secondary glycosylation during crocin biosynthesis. Two distinct datasets were gathered. The first dataset contained UGT protein sequences specifically identified within the *Crocus* genus. The second dataset comprised homologous UGT protein sequences retrieved from NCBI using the *C. sativus* UGTs gene sequence as a query in other plant species. On the first data set, phylogeny and motif analysis and on the second data set, phylogeny, motif, functional divergence, and structural similarity analysis was performed. The phylogeny tree was constructed using MrBayes 3.1.2 software. DIVERGE 3.0 software was utilized to identify functional divergence among members of the UGT protein family. Some UGTs found in plants containing crocin were compared for structural analysis. The SWISS-MODEL server was used for structure prediction, the PyMOL software was employed for structure visualization, and the Dali server was utilized to compare the degree of structural similarity among the desired UGT structures.

Result and discussion

The study focused on UGT enzymes in the *Crocus* genus, specifically comparing those responsible for primary and secondary glycosylation in crocin production. According to the phylogeny analysis UGT proteins were classified into different subfamilies, forming separate branches on the tree. Results revealed four distinct groups within *Crocus* and six when including sequences from other plants in the phylogenetic tree analysis. In both phylogenetic trees, the primary and secondary glycosylation groups were distinctly separated from the beginning, indicating the difference between these two groups.

Using MEME software, 15 shared motifs were identified among these sequences, which likely play crucial roles in protein specificity and function. Based on the position and presence of motifs in sequences, it has been concluded that sequences grouped together likely share similar functions. On the other hand, sequences grouped in different clusters perform distinct functions, which can be justified by the presence of different motifs within them. When UGTs of *Crocus* were studied alongside other plants, motifs 10 and 11 were found in the primary glycosylation group, overlapping with motifs 5 and 7 in the primary glycosylation group identified in the first analysis. This overlap may indicate the significance of these motifs and their potential role in the functional specificity of this group.

Based on previous research, while the C-terminal region of UDP-glucuronosyltransferases (UGTs) often interacts with the sugar donor group, the N-terminal region interacts in substrate recognition of the sugar acceptor group. Additionally, crystal structures of UGTs have shown that the N-terminal region is less protected compared to the C-terminal region, which correlates with the diversity of UGT receptors. In the current study, specific motifs at positions 7 (first analysis) or 11 (second analysis) in primary glycosylation groups, and motif 13 (second analysis) in secondary glycosylation groups located at the N-terminal region. According to previous studies, these motifs may contribute to substrate specificity between primary and secondary glycosylation groups.

In the present study, the functional divergence coefficient of type 1 was significantly greater than zero, indicating a substantial divergence pattern among the subfamilies examined. However, the functional divergence coefficient of type 2 was very low. These results suggest a predominant pattern of functional divergence type 1 for distinguishing between the subfamilies under investigation, and selective pressure specific to certain sites is likely to play a significant role in most UGT genes, leading to the evolution of specific subfamily functions following divergence.

Conclusion

Phylogeny analysis showed that UGTs that carry out primary glycosylation and UGTs that carry out secondary glycosylation were placed in two separate groups. These two groups had high functional distance. In each group, motifs were found specific to the same group. In these specific motifs, amino acids with a high functional divergence coefficient were identified, which can be attributed to the functional difference of these sequences.