

Research Paper

Phylogeny Analysis of Some Genes Involved in the Chlorogenic Acid Biosynthesis Pathway in the Medicinal Plant *Sonchus arvensis* L.

Roghaieh Azizyan¹ and Babak Abdollahi Mandoulakani² 

1- Ph.D., Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran
2- Professor, Department of Agricultural Biotechnology, Institute of Biotechnology, Urmia University, Urmia, Iran,
(Corresponding author: b.abdollahi@urmia.ac.ir)

Received: 08 October, 2024

Revised: 10 November, 2024

Accepted: 18 February, 2025

Extended Abstract

Background: *Sonchus arvensis* L. is a medicinal plant from the Asteraceae family. The valuable medicinal properties of this plant are due to the presence of important and valuable phenolic acids, such as chlorogenic acid. Reducing the risk of various diseases, such as heart disease, diabetes, and cancer, following the consumption of chlorogenic acid has been proven in clinical and research studies. Therefore, more research in the field of medicine and pharmacy on this plant will be considered by researchers in the future. Moreover, there is little information about the genetics of secondary metabolism in medicinal plants. Thus, the genes of many secondary metabolite biosynthesis pathways have not been identified, and little information exists about their regulation and function. Therefore, identifying biosynthetic pathways, knowing how the genes involved in these pathways are regulated, and analyzing their phylogeny are particularly important. On the other hand, examining the phylogenetic relationships of genes in different plant species and determining the evolutionary relationships between them can help better understand the evolution of species. Furthermore, knowledge of the nucleotide sequence of genes and the frequency of different nucleotides is essential for estimating the divergence time of species and reconstructing the evolutionary relationships between different species. The present study aimed to sequence part of the coding region of some genes involved in the chlorogenic acid biosynthetic pathway, such as cinnamate 4-hydroxylase (*C₄H*), P-coumarol-ester 3'-hydroxylase (*C₃H*), and hydroxycinnamoyl coashimate/quinate hydroxycinnamoyl transferase (*HCT*), and to investigate their evolutionary and phylogenetic relationships in *S. arvensis*.

Methods: Since the sequences of *C₄H*, *HCT*, and *C₃H* genes were not identified in *S. arvensis*, their sequences were recovered from plants that are evolutionarily close to *S. arvensis*, such as *Cynara cardunculus* var. *scolymus* and *Cichorium. intybus*. First, the sequences of these genes in plants phylogenetically close to *S. arvensis* were collected from GenBank and aligned using Clustal Omega software, followed by identifying conserved regions. Then, specific primers were designed based on the conserved regions using Fast PCR 4.0 and Gen Runner 3.05 software. In the next step, DNA from leaf samples was extracted by the CTAB method, and the studied gene fragments in *S. arvensis* were amplified using the designed specific primers and polymerase chain reaction (PCR). Electrophoresis of the resulting products was performed using a 1.5% agarose gel at a voltage of 80 V for one to two hours. After staining with ethidium bromide, a photograph was taken using a Gel Doc device (INFINITY, France). Given the specificity of the amplified products, the PCR products were sent directly to Microsynth Company (Switzerland) for sequencing part of the coding regions of the *C₄H*, *C₃H*, and *HCT* genes. To ensure reliability, all PCR products were sequenced in both forward and reverse directions, and alignment (BLAST) of the sequenced fragments against DNA, RNA, and protein databases was performed to verify the accuracy of the sequenced fragments. Then, phylogenetic relationships and the frequency of nucleotide mutations were examined and analyzed in these genes. MEGA5 and BLAST software were used to examine phylogenetic relationships (the maximum likelihood method) and analyze nucleotide sequences (including frequency and substitution). Clustal Omega software was used to examine the alignment of nucleotide and protein sequences. Amino acid analysis (the percentage and weight of amino acids and inferred proteins) was performed with Bio Edit and MEGA5 software, and the Neutrality test was performed with MEGA5 software.

Results: For the first time in this research, a portion of the coding sequence of the *C₄H*, *C₃H*, and *HCT* genes was identified in *Sonchus arvensis* and registered in GenBank with accession numbers ON014597.1, ON014595.1, and ON014596.1, respectively. The results of the analysis of nucleotide sequences showed that nucleotide frequency for *C₄H*, *C₃H*, and *HCT* genes were the



highest for adenine (28.1), guanine (28.4), and thymine (1.1), respectively, and the lowest nucleotide frequency belonged to thymine (13.2), thymine (20.1), and cytosine (19.7), respectively. Moreover, the neutrality test results showed the directional selection of these genes during evolution. The transition substitution mutation was more frequent than the transversion substitution mutation. The results of phylogenetic analysis based on *C₄H*, *HCT*, and *C₃H* gene sequences showed a close phylogenetic relationship or a low genetic distance between *S. arvensis* and lettuce (*L. sativa*). The results of the sequence blast at both the nucleotide and protein levels showed a high similarity to evolutionarily close plants, such as *L. sativa*.

Conclusion: The results of this research positively evaluated the process of natural selection during evolution for the studied *C₄H*, *C₃H*, and *HCT* genes, which indicates the effects of genetic drift or the balancing effects of population evolution throughout history, and a low difference between the frequencies of the polymorphisms. The BLAST results of sequenced *C₄H*, *HCT*, and *C₃H* gene fragments in *S. arvensis* at the nucleotide and protein levels showed the highest percentage of similarity (lowest genetic distance) to plants such as *L. sativa* and *Cichorium intybus*, which confirmed the accuracy of the sequence results. Additionally, phylogenetic analysis based on the sequencing of *C₄H*, *C₃H*, and *HCT* genes showed a close phylogenetic relationship between *S. arvensis* and *L. sativa*.

Keywords: Blast, Nucleotide mutations, Phenolic acid, Phylogeny analysis, Sequencing, *S. arvensis*

How to Cite This Article: Azizyan, R., & Abdollahi Mandoulakan, B. (2025). Phylogeny Analysis of Some Genes Involved in the Chlorogenic Acid Biosynthesis Pathway in the Medicinal Plant *Sonchus arvensis* L. *J Crop Breed*, 17(2), 82-93. DOI: 10.61882/jcb.2024.1578

مقاله پژوهشی

تجزیه فیلوژنی برخی ژن‌های مسیر بیوسنتزی اسید کلروژنیک در گیاه دارویی شیرتیغ وحشی (*Sonchus arvensis* L.)رقیه عزیزیان^۱ و بابک عبدالحی مندولکانی^۲ 

۱- فارغ التحصیل دکتری، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران
۲- استاد، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران، (نویسنده مسوول: b.abdollahi@urmia.ac.ir)

تاریخ پذیرش: ۱۴۰۳/۱۱/۳۰

تاریخ ویرایش: ۱۴۰۳/۰۸/۲۰
صفحه: ۸۲ تا ۹۳

تاریخ دریافت: ۱۴۰۳/۰۷/۱۷

چکیده مبسوط

مقدمه و هدف: شیرتیغ وحشی (*Sonchus arvensis* L.) گیاهی دارویی از خانواده کاسنیان است. خواص دارویی ارزشمند این گیاه به دلیل وجود اسیدهای فنلی مهم و با ارزشی مانند اسید کلروژنیک است. کاهش خطر ابتلا به بیماری‌های مختلف مانند بیماری‌های قلبی، دیابت و سرطان به دنبال مصرف اسید کلروژنیک در مطالعات بالینی و تحقیقی به اثبات رسیده است. بنا بر این، تحقیقات بیشتری در زمینه پزشکی و داروسازی در این گیاه در آینده مد نظر محققین است. همچنین، اطلاعات کمی درباره ژنتیک متابولیسم ثانویه در گیاهان دارویی وجود دارد، به طوری که ژن‌های بسیاری از مسیرهای بیوسنتز متابولیت‌های ثانویه شناسایی نشده‌اند و اطلاعات کمی درباره تنظیم و عملکرد انواع آن‌ها وجود دارد. بنا بر این، شناسایی مسیرهای بیوسنتزی و اطلاع از نحوه تنظیم ژن‌های دخیل در این مسیرها و تجزیه فیلوژنی آن‌ها اهمیت ویژه‌ای دارد. از سویی، بررسی روابط فیلوژنتیکی ژن‌ها در گونه‌های مختلف گیاهی و تعیین روابط تکاملی بین آن‌ها می‌تواند در درک بهتر سیر تحول گونه‌ها کمک کند. همچنین، اطلاع از توالی نوکلئوتیدی ژن‌ها و فراوانی نوکلئوتیدهای مختلف جهت برآورد زمان واگرایی گونه‌ها و احیا کردن روابط تکاملی میان گونه‌های مختلف ضروری است. هدف از تحقیق حاضر توالی‌یابی بخشی از ناحیه کدکننده‌ی برخی ژن‌های دخیل در مسیر بیوسنتزی اسید کلروژنیک، مانند سینامات-۴-هیدروکسیلاز (C_4H)، P-کوماریول-استر-3-هیدروکسیلاز (C_3H) و هیدروکسی سیناموئیل کوآ شیکمات/کوبینات هیدروکسی سیناموئیل ترانسفراز (HCT)، و بررسی روابط تکاملی و فیلوژنی آن‌ها در گیاه شیرتیغ وحشی است.

مواد و روش‌ها: از آنجایی که توالی ژن‌های HCT و C_3H در گیاه *S. arvensis* شناسایی نشده بود، توالی‌های آن‌ها از طریق گیاهانی که به لحاظ تکاملی به *S. arvensis* نزدیک بودند، مانند سایر جنس‌های خانواده کاسنی (کنگر فرنگی: *S. arvensis* var. *scolymus*; *Cynara cardunculus* var. *scolymus*) و کاسنی (*Cichorium intybus*) بازیابی شد. ابتدا توالی این ژن‌ها در گیاهانی که از نظر فیلوژنتیکی به *S. arvensis* نزدیک بودند از GenBank جمع‌آوری شد. هم‌ردیفی با استفاده از نرم‌افزار Clustal omega انجام گرفت و مناطق حفاظت‌شده شناسایی شدند. سپس آغازگرهای اختصاصی براساس نواحی حفاظت‌شده با استفاده از نرم‌افزارهای Fast PCR 4.0 و Gen Runner 3.05 طراحی شدند. در مرحله بعد، DNA از نمونه‌های برگ‌ی به‌روش CTAB استخراج شد و با استفاده از آغازگرهای اختصاصی طراحی‌شده و واکنش زنجیره‌ای پلیمراز (PCR) قطعات ژن‌های مورد مطالعه در شیر تیغ وحشی تکثیر شدند. الکتروفورز محصولات حاصل با استفاده از ژل آگارز یک و نیم درصد با ولتاژ ۸۰ به مدت یک الی دو ساعت انجام شد و پس از رنگ‌آمیزی با اتیدیوم بروماید، عکس‌برداری با استفاده از دستگاه ژل‌داک (INFINITY، فرانسه) صورت گرفت. با توجه به اختصاصی بودن محصولات حاصل از تکثیر، محصولات PCR مستقیماً برای توالی‌یابی بخشی از نواحی کدکننده‌ی ژن‌های C_3H ، C_4H و HCT به شرکت میکروسینس سوییس ارسال شدند. برای اطمینان، تمام محصولات PCR در هر دو جهت رو به جلو و معکوس توالی‌یابی شدند و برای تأیید صحت قطعات توالی‌یابی‌شده، هم‌ردیفی (BLAST) قطعات توالی‌یابی‌شده در برابر پایگاه‌های داده DNA، RNA و پروتئین انجام شد. سپس روابط فیلوژنی و همچنین فراوانی جهش‌های نوکلئوتیدی در این ژن‌ها مورد بررسی و تجزیه و تحلیل قرار گرفتند. برای بررسی روابط فیلوژنی (روش حداکثر درست‌نمایی) و تجزیه و تحلیل توالی‌های نوکلئوتیدی (شامل فراوانی و جایگزینی) از نرم‌افزار MEGA5 و BLAST استفاده شد. برای بررسی هم‌ردیفی توالی‌های نوکلئوتیدی و پروتئینی از نرم‌افزار Clustal Omega استفاده شد. تجزیه و تحلیل آمینواسیدها (درصد و وزن آمینواسیدها و پروتئین‌های استنباطی) با نرم‌افزارهای Bio Edit و MEGA5 و تست Neutrality با نرم‌افزار MEGA5 انجام گرفت.

یافته‌ها: برای اولین بار در این تحقیق، بخشی از توالی ناحیه کدکننده‌ی ژن‌های C_3H ، C_4H و HCT در شیرتیغ وحشی شناسایی و در بانک ژن (GenBank) به ترتیب با شماره‌های دسترسی ON014597.1، ON014595.1 و ON014596.1 ثبت شدند. نتایج مربوط به تجزیه و تحلیل توالی‌های نوکلئوتیدی نشان دادند که بیشترین فراوانی نوکلئوتیدی برای ژن‌های C_3H ، C_4H و HCT به ترتیب مربوط به آدنین (۲۸/۱)، گوانین (۲۸/۴) و تیمین (۲۸/۱) و کمترین فراوانی نوکلئوتیدی به ترتیب متعلق به بازهای تیمین (۱۳/۲)، تیمین (۲۰/۱) و سیتوزین (۱۹/۷) بودند. نتایج تست Neutrality انتخاب جهت‌دار بر روی این ژن‌ها را در طول تکامل نشان دادند. جهش جایگزینی انتقالی بیشتر از جهش جایگزینی تقاطعی بود و آنالیز فیلوژنتیکی براساس توالی‌های ژنی C_3H و HCT و رابطه فیلوژنتیکی نزدیکی بین *S. arvensis* و کاهو (*Lactuca sativa*) را نشان داد. نتایج حاصل از بلاست توالی‌ها در سطح نوکلئوتیدی و پروتئینی، شباهت بالای توالی ژن‌های مورد مطالعه در شیرتیغ وحشی به توالی‌های ژن‌های متناظر در گیاه *L. sativa* را نشان دادند.

نتیجه‌گیری کلی: نتایج این پژوهش، روند انتخاب طبیعی در طی تکامل برای ژن‌های C_3H ، C_4H و HCT در گیاه شیرتیغ وحشی را مثبت ارزیابی کردند که نشان‌دهنده اثرات رانش ژنتیکی و یا اثرات متعادل‌کننده تکامل جمعیت در طول تاریخ است. نتایج BLAST قطعات ژنی توالی‌یابی‌شده‌ی مربوط به ژن‌های C_3H و HCT در *S. arvensis* در سطح نوکلئوتیدی و پروتئینی بیشترین درصد شباهت (کمترین فاصله ژنتیکی) را با گیاهانی مانند *L. sativa* و *C. intybus* نشان دادند که تأییدی بر صحت نتایج توالی‌یابی است. همچنین، آنالیز فیلوژنتیکی براساس توالی‌یابی ژن‌های C_3H ، C_4H و $HCT رابطه فیلوژنتیکی نزدیکی را بین *S. arvensis* و *L. sativa* نشان داد.$

واژه‌های کلیدی: تجزیه فیلوژنی، جهش‌های نوکلئوتیدی، ژن سینامات-۴-هیدروکسیلاز اسید، شیرتیغ وحشی، کلروژنیک اسید

مقدمه

گونه‌های خاصی از گیاهان تولید می‌شود (Tajik et al., 2017). کاهش خطر ابتلا به بیماری‌های مختلف مانند بیماری‌های قلبی، دیابت نوع ۲ (Farah et al., 2008)، ایدز، سرطان، هپاتیت B (Gouthamchandra et al., 2017) و محافظت از کبد (Naveed et al., 2018) به دنبال مصرف اسید کلروژنیک در مطالعات تحقیقی و بالینی گزارش شده است و

شیرتیغ وحشی (*S. arvensis* L.) گیاهی دارویی از خانواده کاسنیان و دارای خواص دارویی مختلف مانند کاهش فشار خون، ضد سرطان (به دلیل وجود ترکیباتی مانند اسید کلروژنیک، اسیدشیکوریک) و ضد مالاریا است (Thasa, 2021; Azizyan & Abdollahi Mandoulakani, 2023). اسید کلروژنیک یک ترکیب فنلی مهم است که به‌وسیله

شایانی به نحوه‌ی تکامل این ژن‌ها و گیاهان می‌کند (Nesaj Hosseini & Shamar Bakhsh, 2011). ژن‌های مسیر بیوسنتز اسید کلروژنیک، مانند C_4H و HCT و $C_3'H$ ، در برخی گونه‌های گیاهی شناسایی شده‌اند. به‌عنوان مثال، ژن C_4H در کاهو (*Lactuca sativa*)، ژن HCT در گیاه کنگر فرنگی (*Cynara. cardunculus* var *scolymus*) و ژن $C_3'H$ کنگر فرنگی (*C. cardunculus* var *scolymus*) شناسایی شده‌اند، اما با بررسی‌های انجام‌شده تاکنون هیچ پژوهشی در رابطه با شناسایی ژن‌های در گیر در مسیر بیوسنتز اسید کلروژنیک در گیاه شیرتیغ وحشی صورت نگرفته‌است. بنا بر این، با توجه به اهمیت گیاه شیرتیغ وحشی و نقش کلیدی ژن‌های C_4H ، HCT و $C_3'H$ در مسیر بیوسنتزی اسیدکلروژنیک و با توجه به محدود بودن تولید تجاری و ارزش دارویی این ماده، شناسایی و تعیین توالی این ژن‌ها می‌تواند اولین و مهم‌ترین قدم در راه درک ساختار و مکانسیم‌های بیوسنتزی این ترکیب مهم باشد. همچنین، بررسی روابط فیلوژنتیکی ژن‌ها با سایر گونه‌های گیاهی و تعیین روابط تکاملی توالی‌های نوکلئوتیدی ژن‌ها می‌تواند نقش موثری در درک بهتر سیر تحول گونه‌ها داشته باشد.

مواد و روش‌ها

شناسایی بخشی از توالی ناحیه کدکننده‌ی ژن‌ها

از آنجایی که توالی ژن‌های HCT ، C_4H و $C_3'H$ در *S. arvensis* شناسایی نشده است، توالی‌های آن‌ها از طریق گیاهانی که به لحاظ تکاملی به *S. arvensis* نزدیک هستند، مانند سایر جنس‌های خانواده کاسنی (کنگر فرنگی: *C. cardunculus* var. *scolymus*) و کاسنی (*C. intybus*)، بازیابی شدند. ابتدا توالی این ژن‌ها در گیاهانی که از نظر فیلوژنتیکی به *S. arvensis* نزدیک هستند، از GenBank جمع‌آوری و با استفاده از نرم‌افزار Clustal Omega برای شناسایی مناطق حفاظت‌شده هم‌ردیف شدند. آغازگرهای اختصاصی بر اساس نواحی حفاظت‌شده با استفاده از نرم‌افزارهای Fast PCR 4.0 و Gen Runner 3.05 طراحی شدند. برای جداسازی توالی ژن C_4H ، آغازگرها بر اساس ناحیه نوکلئوتیدی حفاظت‌شده در *C. intybus*، *Lactuca sativa* و *Echinacea angustifolia* طراحی شدند. برای ژن HCT از ناحیه حفاظت‌شده این ژن در *C. cardunculus* var *scolymus*، *C. intybus* و *E. purpurea* برای طراحی آغازگرها استفاده شد. از آنجایی که توالی $C_3'H$ فقط در *C. cardunculus* var *scolymus* شناسایی شد، گیاهانی از جنس‌های دیگر از جمله *Coffea canephora* (از Rubiaceae) و *Trifolium pretense* (از خانواده Fabaceae) نیز برای طراحی آغازگر جهت شناسایی $C_3'H$ استفاده شدند (جدول ۱).

این نکته شایسته تلاش بیشتر در زمینه تحقیقاتی پزشکی و داروسازی در کارهای تحقیقاتی آینده است (Lu et al., 2020). اسید کلروژنیک از مسیر اسید شیکمیک در گیاهان به‌وسیله تنفس هوازی ایجاد می‌شود (Azizyan & Abdollahi, 2024). از جمله آنزیم‌های کلیدی مهم درگیر در بیوسنتز اسیدکلروژنیک $C_3'H$ ، $4CL$ و HCT هستند (Wen et al., 2022). آنزیم سینامات ۴-هیدروکسیلاز (C_4H) دومین آنزیم مهم در مسیر فنیل پروپانوییدی است که هیدروکسیله کردن ترانس اسید سینامیک را به ۴-هیدروکسی سینامات کاتالیز می‌کند (Xu et al., 2010). هیدروکسی سیناموئیل کوآ شیکمات/کوینات هیدروکسی سیناموئیل ترانسفراز (HCT) متعلق به خانواده آسیل ترانسفرازهای BAHD^۱ (این نام از حروف اولیه چهار آنزیم مشخص‌شده از این خانواده مشتق شده است) است که انتقال هیدروکسی سیناموئیل کو آنزیم A را به سوبستراهایی مانند کوئینات، فلاونول گلیکوزید یا آنتوسیانین‌ها کاتالیز می‌کند و در نتیجه باعث تولید اسیدکلروژنیک و یا فلاونول گلیکوزیدهای استیل‌شده می‌شود (Sun et al., 2018). نقش ژن HCT در بیوسنتز اسید کلروژنیک در توتون، گوجه فرنگی (Niggeweg et al., 2004) و کنگر (Comino et al., 2009) ثابت شده است (Peng et al., 2010). $C_3'H$ (P-کوماریول-استر^۲-3-هیدروکسیلاز) یک سیتوکروم P450 وابسته به مونوکسیژناز از خانواده CYP98^۲ (سیتوکروم P450 مونوکسیژناز ۹۸) است. این خانواده به‌عنوان خانواده‌ای از آنزیم‌ها مطرح است که مرحله متا هیدروکسیلاسیون را در مسیر فنیل پروپانوییدی انجام می‌دهند. متا هیدروکسیلاسیون، هم P-کوماریک اسیدهای آزاد را کاتالیز می‌کند و هم با شیکمیک، کوئینیک و یا فنیل اسید لاکتیک ادغام می‌شود (Schoch et al., 2006). یک قدم مهم در سنتز هیدروکسی سیناموئیل اسیدها، هیدروکسیلاسیون 3' حلقه فنلی است. این آنزیم‌ها از P-کوماریک اسیدهای آزاد به‌عنوان سوبسترا استفاده نمی‌کنند، بلکه از شیکمات و از کوئینات و P-اسید کوماریک به‌عنوان سوبسترا استفاده می‌کنند. در نتیجه، این آنزیم‌ها با عنوان $C_3'H$ (P-کوماریول-استر^۲-3-هیدروکسیلاز) شناخته می‌شوند. هیدروکسیلاسیون 3' حلقه فنیل پروپانوییدی یک مرحله مهم در بیوسنتز اسید کلروژنیک است (Mahesh et al., 2007).

شناسایی متابولیت‌های ثانویه نقش مهمی در بهره‌برداری از پتانسیل دارویی گیاهان دارویی دارد (Masoudi Jozchal et al., 2024). مطالعات حاکی از آن هستند که اطلاعات کمی درباره ژنتیک متابولیسم ثانویه گیاهان دارویی وجود دارد به‌طوری که ژن‌های بسیاری از مسیرهای بیوسنتز ترکیبات دارویی مشخص نشده‌اند و اطلاعات کمی درباره تنظیم و عملکرد انواع آن‌ها وجود دارد (Canter et al., 2005). بررسی روابط فیلوژنتیکی میان این ژن‌ها و مطالعه جهش‌های انتقالی (Transition) و تقاطعی (Transversion) در این ژن‌ها کمک

D: Deacetylindoline 4-O-acetyltransferase (DAT)

² Cytochrome P450 monooxygenase 98

¹ B: Benzylalcohol O-acetyltransferase (BEAT)

A: Anthocyanin O- hydroxycinnamoyl transferase (AHCT)

H: Anthranilate N-hydroxycinnamoyl/benzoyltransferase (HCBT)s

جدول ۱- شماره دسترسی، دمای اتصال، توالی آغازگر و اندازه محصول مربوط به آغازگرهای مورد استفاده برای جداسازی ژن‌ها در گیاه *Sonchus arvensis*

Table 1. Accession numbers, annealing temperature, sequences, and product sizes of the primers used to identify the genes in *Sonchus arvensis*

ژن Gene	دمای اتصال (°C) Annealing temperature	توالی آغازگر Primer sequences	اندازه محصول مورد انتظار (bp) Expected product size (bp)
<i>HCT</i> هیدروکسی سیناموئیل کوآشیکمات/ کوینات هیدروکسی سیناموئیل ترانسفراز	62	F: GAGAATCGACGATGGTGAGACCC R: GTGTTGATGAAATGCAGCCCAGA	485
<i>C₄H</i> سینامات ۴-هیدروکسیلاز	64	F: GGTCCATCGAATGGGGAATCGC R: CGGAGGCGGAGGGTTTCTTTGA	166
<i>C₃H</i> P-کوماریل-استر ۳'-هیدروکسیلاز	53	F: CAGTGGCATTCAACAACATTAC R: TGCTCATCCATGATAGCTCG	244

C₄H: سینامات ۴-هیدروکسیلاز، *HCT*: هیدروکسی سیناموئیل کوآشیکمات / کوینات هیدروکسی سیناموئیل ترانسفراز، *C₃H*: P-کوماریل-استر ۳'-هیدروکسیلاز، *C₄H*: cinnamate 4-hydroxylase, *HCT*: hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyl transferase, *C₃H*: coumaroyl ester 3'-hydroxylase

از آنجایی که تمامی قطعات ژنی مشاهده شده روی ژل الکتروفور تک‌باندی و اختصاصی بودند (شکل ۲)، محصولات PCR به همراه آغازگرها برای توالی‌یابی به شرکت میکروسینس سوییچ ارسال شدند. تمام محصولات PCR برای اطمینان در هر دو جهت رو به جلو و معکوس توالی‌یابی شدند و سپس برای تأیید صحت قطعات توالی‌یابی شده، بلاست (BLAST) قطعات توالی‌یابی شده در برابر پایگاه‌های داده DNA، RNA و پروتئین انجام شد.

آنالیز آماری

طراحی آغازگرهای اختصاصی براساس نواحی حفاظت‌شده با استفاده از نرم‌افزارهای Fast PCR 4.0 و Gen Runner 3.05 صورت گرفت. برای بررسی روابط فیلوژنی (روش حداکثر درست‌نمایی) و تجزیه و تحلیل توالی‌های نوکلئوتیدی (شامل فراوانی، جایگزینی) از نرم‌افزار MEGA5 و BLAST و برای هم‌ردیفی توالی‌ها از نرم‌افزار Clustal Omega، برای بررسی تجزیه و تحلیل آمینواسیدها (درصد و وزن آمینواسیدها و پروتئین‌های استنباطی) از نرم‌افزار Bio Edit و MEGA5 و برای بررسی تست Neutrality از نرم‌افزار MEGA5 استفاده شد.

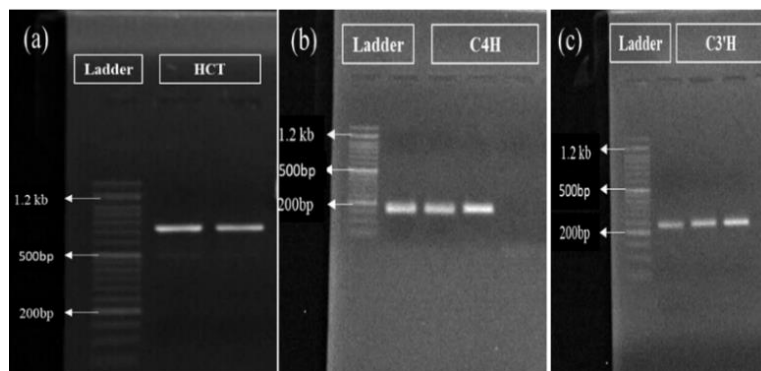
نتایج و بحث

نتایج الکتروفورز محصولات PCR روی ژل الکتروفورز ۱/۵ درصد و وجود باند اختصاصی در نیم‌رخ الکتروفورز نشان دادند که قطعات ژنی مورد نظر به درستی تکثیر شدند (شکل ۱).

تکثیر و توالی‌یابی قطعات ژنی

ابتدا بذور گیاه شیرتیغک وحشی (زرین تاک، ارومیه) کشت شدند و سپس برگ گیاهان مورد نظر در مرحله ۶ الی ۸ برگی برداشت شد. DNA ژنومی نمونه‌های برگ با روش استیل‌تری متیل آمونیوم بروماید Cetyltrimethylammonium bromide (CTAB) استخراج شد (Ausubel *et al.*, 1992). سپس واکنش زنجیره‌ای پلیمرز (PCR) در حجم نهایی ۲۰ میکرولیتر، حاوی ۱۰ میکرولیتر Taq DNA Polymerrase 2X PCR Master Mix (AMPLIQON، دانمارک)، یک میکرولیتر از آغازگر رفت و یک میکرولیتر آغازگر برگشت هر دو در غلظت ۱۰ میکرومولار، سه میکرولیتر DNA با غلظت ۴۰ نانوگرم در میکرولیتر و پنج میکرولیتر آب دوبار تقطیرشده انجام شد. در ادامه، واکنش‌های PCR تحت شرایط یک چرخه در دمای ۹۵ درجه سانتی‌گراد به مدت ۴ دقیقه، ۳۵ چرخه شامل ۹۴ درجه سانتی‌گراد به مدت ۳۰ ثانیه و اتصال مناسب برای هر جفت آغازگر بسته به دمای اتصال آغازگر (۵۳-۶۴ درجه سانتی‌گراد) به مدت یک دقیقه و بسط به مدت یک دقیقه در دمای ۷۲ درجه سانتی‌گراد انجام شدند. الکتروفورز محصولات حاصل با استفاده از ژل آگارز یک و نیم درصد با ولتاژ ۸۰ به مدت یک الی دو ساعت انجام شد و پس از رنگ‌آمیزی با اتیدیوم بروماید، عکس‌برداری با استفاده از دستگاه ژل داک (INFINITY، فرانسه) صورت گرفت.

توالی‌یابی قطعات ژنی



شکل ۱- تکثیر نواحی حفاظت‌شده ژن‌های *HCT* و *C3'H* در گیاه شیرتیغ وحشی. (a): *HCT*، هیدروکسی سیناموئیل کوآ شیکمات/کوینات هیدروکسی سیناموئیل ترانسفراز، (b): *C4H*، سینامات ۴-هیدروکسیلاز، (c): *C3'H*، کوآ-کوماریل-۳-هیدروکسیلاز. نشانگر اندازه مورد استفاده 50 bp EcelBand™ (شرکت Smobio، اسپانیا) است.

Figure 1. Amplification of the conserved regions of *HCT*, *C3'H*, and *C4H* genes in the *S. arvensis* plant. (a): *HCT*, hydroxycinnamoyl- CoA shikimate/ quinate hydroxycinnamoyl transferase, (b): *C4H*, cinnamate 4-hydroxylase, (c): *C3'H*, coumaroyl ester 3'-hydroxylase

به‌عنوان توالی کدکننده جزئی *HCT* با شماره دسترسی ON014595.1 ثبت شد که ترکیب آمینواسیدی مربوط به این توالی (۱۳۸ آمینو اسید)، وزن مولکولی حدود ۵۴۴۶۴/۷۱ دالتون را نشان داد که بیشترین اسیدآمینوها به‌ترتیب به ترتوئین (۲۸/۱۴) و آلانین (۲۷/۵۳ درصد مول) تعلق داشت (جدول ۲ و ۳). در سطح پروتئینی نیز بلاست پروتئینی توالی پروتئین‌های استنباطی *C4H*، *HCT* و *C3'H* با سایر پروتئین‌های گیاهان نزدیک از نظر فیلوژنی انجام شد که برای توالی پروتئینی ژن‌های *C4H* و *C3'H* شباهتی به‌دست نیامد. این عدم تشابه را می‌توان به طول کم توالی پروتئین‌های استنباطی نسبت داد ولی بلاست توالی پروتئینی استنباطی از ژن *HCT* شباهت بسیار بالایی را در سطح اسید آمینه با گیاه *Lactuca sativa* (۹۹/۱۱٪) نشان داد (جدول ۴). هم‌ردیفی توالی‌های پروتئینی استنباطی از ژن‌های *C4H*، *HCT* و *C3'H* با گیاهان نزدیک از نظر فیلوژنی نتایج بلاست را تایید کرد (شکل ۲، ۳ و ۴). پژوهشی که به‌منظور بررسی سطوح بیان ژن‌ها و محتوای ترکیبات فنلی در گیاه شیرتیغ وحشی صورت گرفت، توالی‌یابی و بلاست قطعات ژنی درگیر در مسیر بیوسنتزی کلروژنیک اسید مانند ژن فنیل‌آلانین آمونیلایز (*PAL*)^۳، ۴-کومارات کوآ لیگاز (*4CL*)^۴ و هیدروکسی سینامویل-کوآ شیکمات/کوینات هیدروکسی سیناموئیل ترانسفراز (*HCT*) نیز شباهت بالایی را با گیاه کاهو (*Lactuca sativa*) نشان داد که می‌تواند تاییدی بر نتایج مورد مطالعه ما باشد (Ghaderi & Abdollahi Mandoulakani, 2024).

توالی‌یابی قطعات ژن‌های *HCT*، *C4H* و *C3'H* در *S. arvensis* در سطح نوکلئوتیدی (جدول ۲)، شباهت بالایی را با ژن‌های سایر گیاهانی که از نظر فیلوژنتیکی به گیاه مورد مطالعه نزدیک هستند (مانند *Lactuca sativa* و *C. intybus*) نشان داد که علاوه بر تأیید صحت نتایج توالی‌یابی، نشان‌دهنده توالی حفاظت‌شده این ژن‌ها در گونه‌های گیاهی دیگر است. اندازه مورد انتظار *C4H* با توجه به آغازگرهای طراحی‌شده براساس نواحی حفاظت‌شده، ۱۶۶ جفت‌باز بود که با نتیجه محصول PCR روی ژل مطابقت داشت. ژن *C4H* با طول ۱۲۱ جفت باز در بانک ژن (GenBank) با شماره دسترسی ON014597.1 ثبت شد. بر اساس آمینواسیدهای استنباطی از این قطعه ثبت‌شده (۴۰ آمینو اسید)، قطعه پروتئینی با وزن مولکولی ۱۰۵۳۵/۳۲ دالتون به‌دست آمد که بیشترین درصد مربوط به آمینواسیدهای سیستئین (۴۳/۸۰ درصد مول) و آلانین (۲۸/۱۰) بود. طول مورد انتظار ژن *C3'H* ۲۴۴ جفت بود. این توالی با طول ۱۹۴ جفت‌باز در GenBank با شماره دسترسی ON014/596.1 ثبت شد. تعداد آمینواسید استنباطی از این قطعه حدود ۶۴ آمینو اسید بود و بیشترین فراوانی آمینواسیدها به‌ترتیب مربوط به آمینواسیدهای آلانین (۲۷/۸۴ درصد مول) و گلیسین (۲۸/۳۵ درصد مول) بود. هم‌چنین، قطعه مورد انتظار برای تکثیر *HCT* حدود ۴۸۵ جفت‌باز بود، اما طول توالی ثبت شده در GenBank، ۶۶۱ جفت‌باز بود که نشان‌دهنده یک درج ۱۷۶ جفت‌بازی در *S. arvensis* در این ژن است. این توالی در

^۴ 4-Coumaroyl: coA-ligase

^۳ Phenylalanine ammonialyase

جدول ۲- نتایج بلاست (BLAST) ژن‌های $C_3'H$ ، HCT و C_4H در گیاه *Sonchus arvensis*Table 2. The BLAST results of C_4H , HCT , and $C_3'H$ genes identified from *Sonchus arvensis* L.

ژن Gene	شماره دسترسی Accession number	طول Length (bp)	شماره دسترسی بیشترین شباهت Accession number of the most similar sequence	نام گیاه Plant name	مقدار قابل انتظار E-Value	درصد تشابه (%) Identities	درصد هم‌پوشانی (%) Query Coverage
HCT هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز	ON014595.1	661	KT222891.1	<i>C. intybus</i>	2e-158	96.54	48
C_4H سینامات ۴- هیدروکسیلاز	ON014597.1	121	XM_023882920.3	<i>L. sativa</i>	7e-31	89.26	27
$C_3'H$ P- کواریول - استر 3'- هیدروکسیلاز	ON014596.1	194	XM_023873558.3	<i>L. sativa</i>	2e-53	86.93	90

C_4H : سینامات ۴- هیدروکسیلاز، HCT : هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز، $C_3'H$: P- کواریول - استر 3'- هیدروکسیلاز
 C_4H : cinnamate 4-hydroxylase, HCT : hydroxycinnamoyl- CoA shikimate/ quinate hydroxycinnamoyl transferase, $C_3'H$: coumaroyl ester 3'-hydroxylase

جدول ۳- تعداد و درصد آمینو اسیدهای ژن‌های $C_3'H$ ، HCT و C_4H در گیاه *Sonchus arvensis*Table 3. The number and percentage of amino acids of C_4H , HCT , and $C_3'H$ genes in *Sonchus arvensis*

Table 5: The number and percentage of amino acids of <i>C₄H</i> , <i>HCT</i> , and <i>C₃H</i> genes in <i>Bombus terrestris</i>				ژن (Gene)			
<i>HCT</i>		<i>C₃H</i>		<i>C₄H</i>			
مول % Mol %	تعداد آمینواسید Number of amino acids	مول % Mol %	تعداد آمینواسید Number of amino acids	مول % Mol %	تعداد آمینواسید Number of amino acids	نماد آمینواسید Amino acid symbol	آمینواسید Amino acid
27.53	182	27.84	54	28.1	34	A	Ala (آلانین)
19.67	130	23.71	46	43.8	53	C	Cys (سیستین)
0	0	0	0	0	0	D	Asp (اسپارتیک)
0	0	0	0	0	0	E	Glu (گلوتامیک)
0	0	0	0	0	0	F	Phe (فنیل آلانین)
24.66	163	28.35	55	14.88	18	G	Gly (گلايسين)
0	0	0	0	0	0	H	His (هيسټيدين)
0	0	0	0	0	0	I	Ile (ايزولوسين)
0	0	0	0	0	0	K	Lys (ليزين)
0	0	0	0	0	0	L	Leu (لوسين)
0	0	0	0	0	0	M	Met (متيونين)
0	0	0	0	0	0	N	Asn (اسپاراژين)
0	0	0	0	0	0	P	Pro (پرولين)
0	0	0	0	0	0	Q	Gln (گلوتامين)
0	0	0	0	0	0	R	Arg (ارژنين)
0	0	0	0	0	0	S	Ser (سيرين)
28.14	186	20.1	39	13.22	16	T	Thr (ترونين)
0	0	0	0	0	0	V	Val (والين)
0	0	0	0	0	0	W	Trp (تريپتوفان)
0	0	0	0	0	0	Y	Tyr (تيروزين)

C_4H : سینامات ۴- هیدروکسیلاز، HCT : هیدروکسی سیناموئیل ترانسفراز، $C_3'H$: P- کواریول - استر 3'- هیدروکسیلاز
 C_4H : cinnamate 4-hydroxylase, HCT : hydroxycinnamoyl- CoA shikimate/ quinate hydroxycinnamoyl transferase, $C_3'H$: coumaroyl ester 3'-hydroxylase

جدول ۴- نتایج بلاست (BLAST) پروتئین‌های استنباطی ژن‌های $C_3'H$ ، HCT و C_4H در گیاه *Sonchus arvensis*Table 4. BLAST results of the inferred proteins of C_4H , HCT , and $C_3'H$ genes identified in *Sonchus arvensis* L.

پروتئین Protein	شماره دسترسی Accession number	طول اسید آمینو Amino acid length	شماره دسترسی بیشترین شباهت Accession number of the most similar sequence	نام گیاه Plant name	مقدار قابل انتظار E-Value	درصد تشابه (%) Identities	درصد هم‌پوشانی (%) Query Coverage
HCT هیدروکسی سیناموئیل ترانسفراز	UVV38606.1	138	XP_023744066.1	<i>L. sativa</i>	1e-70	99.11	50
C_4H سینامات ۴- هیدروکسیلاز	UVV38608	40	-	-	-	-	-
$C_3'H$ P- کواریول - استر 3'- هیدروکسیلاز	UVV38607	64	-	-	-	-	-

C_4H : سینامات ۴- هیدروکسیلاز، HCT : هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز، $C_3'H$: P- کواریول - استر 3'- هیدروکسیلاز
 C_4H : cinnamate 4-hydroxylase, HCT : hydroxycinnamoyl- CoA shikimate/ quinate hydroxycinnamoyl transferase, $C_3'H$: coumaroyl ester 3'-hydroxylase

تجزیه فیلوژنی برخی ژن‌های مسیر بیوسنتزی اسیدکلروژنیک در گیاه دارویی شیرتیغ وحشی (*Sonchus arvensis* L.) ۸۹

<i>C. intybus</i>	GPSNGESPNSLITHKSNPNAMNSTPFSAPESK*PSRTFKNCRTFKP*SKKPSSAS	53
<i>s. arvensis</i>	-----PSI-----QLRPNSTQFSDPESKSRNPTSKSSHTSKP*SKKP PPP	39
<i>L. sativa</i>	GQSNGESPPSS*TTLTFSQSSGTSTRNLGPVFRSPSLMFRSLTTSKQSSKKPYVS	54
<i>E. angustifolia</i>	GQSNGESQSL*TI PRSKPSSGTSSSTQSSDPVSKSPSLMSTKTSHTSKP*SKKPSSAS	53

شکل ۲- مقایسه پروتئین استنباطی حاصل از بخشی از ناحیه کدکننده ژن *C₄H* مربوط به گیاه *S. arvensis* با پروتئین‌های سایر گیاهان نزدیک به لحاظ تکاملی (*C. intybus*, *L. sativa* و *E. angustifolia*). حروف نشان دهنده آمینو اسیدها هستند. * نشان دهنده همانند بودن، نشان دهنده جانشینی حفظ‌شده، . نشان دهنده جانشینی نیمه حفظ‌شده و - نشان دهنده فاصله هستند.

Figure 2. Comparison of the inferred protein from part of the coding region of the *C₄H* gene in *S. arvensis* with proteins from other evolutionary close plants (*C. intybus*, *L. sativa*, and *E. angustifolia*). Letters indicate amino acids. * indicates "identity", indicates "conserved substitution", . indicates "semi-conserved substitution", and - indicates a "gap".

<i>S. arvensis</i>	-----TRRV**TSKAKSSKPLWQMG*SSVDL*PWQSTFHGSGGSSLLR	39
<i>C. cardunculus</i>	QWHSTTLGSLGNDL*TRRV*WMTREXXLRQLWLT*SLVHL*PWQSIHSGSDGSSR*K	55
<i>C. canephora</i>	QWHSTTLQDLHLVKGS*IPKE*WMNKERSRR*LQMA*SLVHRSWQSTSHGYGGCFHSM	56
<i>T. pretense</i>	QLHSTTSQGWHLGKDL*TDQV*WMSKEWNSRL*WQMG*S*EHLPLWQSIFLG*GGCSH*K	53

: * . : : . ** . *** * . *

<i>S. arvensis</i>	KTPLPNTENAVTG*PELSGMT	59
<i>C. cardunculus</i>	KKPLQSTEHVVGSHPEPSWMS	76
<i>C. canephora</i>	KLLSRSMVLEGTASLEPSWKS	77
<i>T. pretense</i>	RRLLPSTVLVGTDLLELSWMS	74

: . * * :

شکل ۳- مقایسه پروتئین استنباطی حاصل از بخشی از ناحیه کدکننده ژن *C₃H* مربوط به گیاه *S. arvensis* با پروتئین‌های سایر گیاهان نزدیک به لحاظ تکاملی (*C. cardunculus* var *scolymus* *Coffea canephora* و *Trifolium pretense*). حروف نشان دهنده آمینواسیدها هستند. * نشان دهنده همانند بودن، نشان دهنده جانشینی حفظ‌شده، . نشان دهنده جانشینی نیمه حفظ‌شده و - نشان دهنده فاصله هستند.

Figure 3. Comparison of the inferred protein from part of the coding region of the *C₃H* gene in *S. arvensis* with proteins from other evolutionary close plants (*C. cardunculus* var *scolymus*, *Coffea canephora*, and *Trifolium pretense*). Letters indicate amino acids. * indicates "identity", . indicates "conserved substitution", indicates "semi-conserved substitution", and - indicates a "gap".

<i>S. arvensis</i>	-----NVDLVVPNFHTPSVYFYRPNAGANFFDPKVMKDALSALV	40
<i>E. purpurea</i>	RESTMVRPAETPRVNLWNSNVDLVVPNFHTPSVYFYRPNATNFDPKVMKEALSALV	60
<i>C. intybus</i>	RESTMVRPAETPRILWNSNVDLVVPNFHTPSVYFYRPTGATNFDAKVMKDALSALV	60
<i>C. cardunculus</i>	RESTMVRPAETPRINLWNSNVDLVVPNFHTPSVYFYRPNAGANFFDPKVMKDALSALV	60

***** : * : * : * : * : * : * : * : *

<i>S. arvensis</i>	PFYPMGRLKRDEGRIEIDCQGGVLFVEAESDGMVDDYDGFAPLTLELRKLIPAVDYSQ	100
<i>E. purpurea</i>	PFYPMGRLKRDEGRIEIDCQGGVLFVEAESDGMVDDYDGFAPLTLELRKLIPAVDYSQ	120
<i>C. intybus</i>	PFYPMGRLKRDEGRIEIDCQGGVLFVEAESDGMVDDYDGFAPLTLELRKLIPAVDYSQ	120
<i>C. cardunculus</i>	PFYPMGRLKRDEGRIEIDCQGGVLFVEAESDGMVDDYDGFAPLTLELRKLIPAVDYTL	120

***** : * : * : * : * : * : * : * : *

<i>S. arvensis</i>	GIESYPLLVLQVLLSTIIFNRRRAKIT*QKMELE*YIYKYDSKVRFCFM*TSKRGMTF*	156
<i>E. purpurea</i>	GIESYPLLVLQVTFYKCGGVSGLVGMQHA-----AD-----GASGLHFI	160
<i>C. intybus</i>	GIESYLLVLQVTFYKCGGVSGLVGMQHA-----AD-----GASGLHFI	160
<i>C. cardunculus</i>	GIESYLLVLQVTFYKCGGVSGLVGMQHA-----AD-----GASGLHFI	160

***** : . : : * . : *

<i>S. arvensis</i>	CPNPYIIISQLKFHHRIAVYKLGTRFIFVCWNVGHIL*MWNGFTRSWDAAPRCRWRIWAA	215
<i>E. purpurea</i>	N-----	161
<i>C. intybus</i>	N-----	161
<i>C. cardunculus</i>	N-----	161

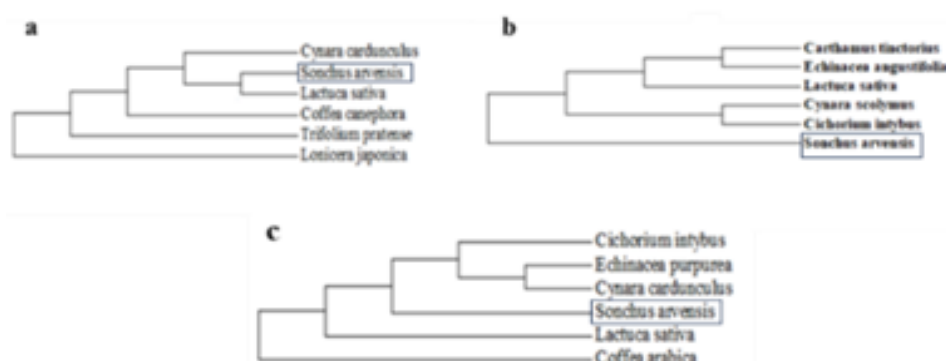
<i>S. arvensis</i>	X	216
<i>E. purpurea</i>	-	161
<i>C. intybus</i>	-	161
<i>C. cardunculus</i>	-	161

شکل ۴- مقایسه پروتئین استنباطی حاصل از بخشی از ناحیه کدکننده ژن *HCT* مربوط به گیاه *S. arvensis* با پروتئین‌های سایر گیاهان نزدیک به لحاظ تکاملی (*C. cardunculus* var *scolymus*، *C. intybus* و *E. purpurea*). حروف نشان دهنده آمینواسیدها هستند. * نشان دهنده همانند بودن، نشان دهنده جانشینی حفظ‌شده، . نشان دهنده جانشینی نیمه حفظ‌شده و - نشان دهنده فاصله هستند.

Figure 4. Comparison of the inferred protein from part of the coding region of the *HCT* gene in *S. arvensis* with proteins from other evolutionary close plants (*C. cardunculus* var *scolymus*, *C. intybus*, and *E. purpurea*). Letters indicate amino acids. * indicates "identity", . indicates "conserved substitution", indicates "semi-conserved substitution", and - indicates a "gap".

L. sativa را برای تمام ژن‌های مورد مطالعه نشان داد و این مطلب بیانگر این است که گیاه *L. sativa* کمترین فاصله ژنتیکی را با گیاه مورد مطالعه ما دارد (شکل ۵).

بررسی روابط فیلوژنی ابزار قدرتمندی برای دانستن بسیاری از ابهامات و مسائل مربوط به مراحل تکامل است (Elliott et al., 2006). آنالیز فیلوژنتیکی بر اساس توالی‌های ژنی *C₄H* و *HCT* و *C₃H* رابطه فیلوژنتیکی نزدیکی بین *S. arvensis*



شکل ۵- درخت فیلوژنتیکی به روش حداکثر درست‌نمایی بر اساس بخشی از توالی ژن‌های C_4H ، HCT و $C_3'H$ مربوط به گیاه *S. arvensis* با سایر جنس‌های موجود. (a): کوماریل-استر 3'-هیدروکسیلاز ($C_3'H$)، (b): سینامات ۴-هیدروکسیلاز (C_4H)، (c): هیدروکسی سینامویل ترانسفراز (HCT).

Figure 5. The phylogenetic tree constructed by the maximum likelihood method using a part of the sequences of C_4H , HCT , and $C_3'H$ genes identified from the *S. arvensis* plant with other existing genera. (a): coumaroyl ester 3'-hydroxylase ($C_3'H$), (b): cinnamate 4-hydroxylase (C_4H), (c): hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyl transferase (HCT)

برآورد زمان واگرایی گونه‌ها و روابط تکاملی میان گونه‌های مختلف استفاده می‌شود (Asghari & Dashab, 2018). نتایج آنالیز توالی‌های نوکلئوتیدی ژن‌های $C_3'H$ ، C_4H و HCT نشان دادند که بیشترین فراوانی نوکلئوتیدی برای ژن‌های $C_3'H$ ، C_4H و HCT به ترتیب به آدنین (۲۸/۱)، گوانین (۲۸/۴) و تیمین (۲۸/۱) و کمترین فراوانی نوکلئوتیدی به ترتیب به بازهای تیمین (۱۳/۲)، تیمین (۲۰/۱) و سیتوزین (۱۹/۷) متعلق بودند (جدول ۵).

برای بررسی آنالیزهای بیوانفورماتیکی نظیر فراوانی‌های نوکلئوتیدی، برآورد جایگزینی نوکلئوتیدی و بررسی اثر تکاملی انتخاب طبیعی، از نتایج حاصل از توالی‌یابی ژن‌ها استفاده شد (Sahraki et al., 2020). نتیجه تغییر در فراوانی ژنی، تکامل است و در گذر زمان برخی آلل‌ها از بین می‌روند و فراوانی بعضی از آن‌ها تا صد درصد افزایش می‌یابد. افزایش این آلل‌ها در جمعیت باعث تثبیت آن‌ها می‌گردد و تثبیت موفق این آلل‌ها سبب تکامل بلندمدت یک گونه می‌شود که بازتابی از پایداری جهش‌ها در طول زمان است (Nesaj Hosseini & Shamar, 2011). از اطلاعات توالی‌های نوکلئوتیدی برای

جدول ۵- فراوانی نوکلئوتیدهای حاصل از توالی ژن‌های $C_3'H$ ، C_4H و HCT در گیاه *Sonchus arvensis*

Table 5. Frequencies of nucleotides obtained from the sequence of C_4H ، $C_3'H$ ، and HCT genes in *Sonchus arvensis*

فراوانی نوکلئوتید Nucleotide frequency					قطعه تکتیری ژن Sequenced gene fragment
G (گوانین)	C (سیتوزین)	T (تیمین) (یوراسیل)	A (آدنین)		
14.9	43.8	13.2	28.1		C_4H سینامات ۴-هیدروکسیلاز
28.4	23.7	20.1	27.8		$C_3'H$ P-کوماریل-استر 3'-هیدروکسیلاز
24.7	19.7	28.1	27.5		HCT هیدروکسی سینامویل کوآ شیکمات/ کوینات هیدروکسی سینامویل ترانسفراز

C_4H : سینامات ۴-هیدروکسیلاز، HCT : هیدروکسی سینامویل کوآ شیکمات/ کوینات هیدروکسی سینامویل ترانسفراز، $C_3'H$: کوماریل-استر 3'-هیدروکسیلاز
 C_4H : cinnamate 4-hydroxylase, HCT : hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyl transferase, $C_3'H$: coumaroyl ester 3'-hydroxylase

هم‌چنین، برآورد جایگزینی نوکلئوتیدی نشان داد که مقادیر بالای جایگزینی نوکلئوتیدی مربوط به جایگزینی از نوع پیریمیدینی بود به‌طوری که این مقدار برای ژن‌های C_4H ، $C_3'H$ و HCT به ترتیب برای تبدیل تیمین به سیتوزین برابر ۲۲/۸۷، ۱۸/۷۶ و ۱۷/۱۱ درصد و برای تبدیل سیتوزین به تیمین به ترتیب برابر ۱۶/۱۰، ۲۰/۱۵ و ۱۷/۴۷ درصد به‌دست آمد. برای بازهای پورینی این مقادیر کمتر بودند و برای ژن‌های $C_3'H$ ، C_4H و HCT به ترتیب برای آدنین به گوانین برابر

۱۲/۴۲، ۱۰/۹۲ و ۱۱/۰۷ درصد و برای تبدیل گوانین به آدنین برابر ۱۱/۶۸، ۱۰/۲۷ و ۱۱/۵۷ درصد برآورد شدند (جدول ۶). این نتیجه با نتایج محققان دیگری (Sheikhi et al., 2019; Sahraki et al., 2020) که بیشتر جایگزینی را جایگزینی پیریمیدینی گزارش کرده‌اند، مطابقت دارد و علت این نوع از جایگزینی را احتمالاً می‌توان به متیله‌شدن سیتوزین نسبت داد (Picoult-Newberg et al., 1999).

تجزیه فیلوژنی برخی ژن‌های مسیر بیوسنتزی اسیدکلروژنیک در گیاه دارویی شیرتیغ وحشی (*Sonchus arvensis* L.) ۹۱

جدول ۶- ماتریس برآورد الگوی جایگزینی نوکلئوتیدی ژن‌های *C₄H*Table 6. The estimation matrix of the nucleotide substitution pattern for the *C₄H* gene

ژن <i>C₄H</i> <i>C₄H</i> gene	از/ به From/To	آدنین A	تیمین/اوراسیل T/U	سیتوزین C	گوانین G
	A (آدنین)	-	3.85	5.46	12.42
	T/U (تیمین/اوراسیل)	4.44	-	22.87	4.72
	C (سیتوزین)	4.44	16.10	-	4.72
	G (گوانین)	11.68	3.85	5.46	-

C₄H: سینامات ۴-هیدروکسیلاز

C₄H: cinnamate 4-hydroxylase

جدول ۷- ماتریس برآورد الگوی جایگزینی نوکلئوتیدی ژن *HCT*Table 7. The estimation matrix of the nucleotide substitution pattern for the *HCT* gene

ژن <i>HCT</i> <i>HCT</i> gene	از/ به From/ To	آدنین A	تیمین/اوراسیل T/U	سیتوزین C	گوانین G
	A (آدنین)	-	5.32	5.21	11.07
	T/U (تیمین/اوراسیل)	5.55	-	17.11	5.31
	C (سیتوزین)	5.55	17.47	-	5.31
	G (گوانین)	11.57	5.32	5.21	-

HCT: هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز

HCT: hydroxycinnamoyl- CoA shikimate/ quinate hydroxycinnamoyl transferase

جدول ۸- ماتریس برآورد الگوی جایگزینی نوکلئوتیدی ژن *C₃'H*Table 8. The estimation matrix of the nucleotide substitution pattern for the *C₃'H* gene

ژن <i>C₃'H</i> <i>gene C₃'H</i>	از/ به From/ To	آدنین A	تیمین/اوراسیل T/U	سیتوزین C	گوانین G
	A (آدنین)	-	4.37	4.07	10.92
	T/U (تیمین/اوراسیل)	5.57	-	18.76	5.93
	C (سیتوزین)	5.57	20.15	-	5.93
	G (گوانین)	10.27	4.37	4.07	-

C₃'H: P - کوماریل - استر 3'-هیدروکسیلاز

C₃'H: coumaroyl ester 3'-hydroxylase

اثرات متعادل کننده تکامل جمعیت در طول تاریخ است و تفاوت پایین بین فراوانی‌های چندشکلی‌ها را بیان می‌کند. در مطالعه ما، برای ژن‌های *C₄H*، *C₃'H* و *HCT* مقادیر D مثبت بودند که بیانگر تفاوت پایین بین فراوانی چندشکلی‌های مختلف است. نتایج این مطالعه انتخاب جهت‌دار بر روی ژن‌های *C₄H*، *C₃'H* و *HCT* را نشان می‌دهند (جدول ۹).

برای شناسایی هر گونه انحراف از فرضیه صفر تکامل خنثی و یا اثر انتخاب طبیعی جهت‌دار بر روی ژن‌های *C₄H*، *C₃'H* و *HCT* در طول تکامل از تست Neutrality استفاده شد. مقادیر D معنی‌دار و منفی بیانگر انتخاب جهت‌دار و نشان‌دهنده تفاوت بالایی میان فراوانی چندشکلی‌های مختلف هستند. مثبت و معنی‌دار بودن مقادیر D نشان‌دهنده اثرات رانش ژنتیکی و یا

جدول ۹- نتایج تست Neutrality در ژن‌های مورد مطالعه

Table 9. The results of the Neutrality test for the studied genes

ژن Gene	M تعداد توالی	S تعداد نواحی جداشده	Ps S/m	Θ Ps/a1	Π تنوع نوکلئوتیدی	D
<i>C₄H</i> سینامات ۴-هیدروکسیلاز	6	81	0.675000	0.295620	0.305000	0.204647
<i>HCT</i> هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز	6	474	0.728111	0.318881	0.338556	0.401585
<i>C₃'H</i> - هیدروکسیلاز 3'-کوماریل - استر P	8	102	0.534031	0.205963	0.263650	1.524637

M: تعداد توالی، S: تعداد نواحی جداشده، Ps/a1: Θ، S/m: Ps، π: تنوع نوکلئوتیدی، D: تست Neutrality با استفاده از روش Tajima، *C₄H*: سینامات ۴-هیدروکسیلاز، *HCT*: هیدروکسی سیناموئیل کوآ شیکمات/ کوئینات هیدروکسی سیناموئیل ترانسفراز، *C₃'H*: P - کوماریل - استر 3'-هیدروکسیلاز

M: sequence number, S: number of isolated regions, S/m: Ps, Ps/a1: Θ, π: nucleotide diversity, D: the neutrality test using the Tajima method, *C₄H*: cinnamate 4-hydroxylase, *HCT*: hydroxycinnamoyl- CoA shikimate/quinate hydroxycinnamoyl transferase, *C₃'H*: coumaroyl ester 3'-hydroxylase

توالی‌یابی شده‌ی *HCT*، *C₄H* و *C₃'H* در *S. arvensis* نشان دادند که در سطح نوکلئوتیدی بیشترین درصد شباهت (کمترین فاصله ژنتیکی) با گیاهانی مانند *L. sativa* و *C. intybus* و در سطح پروتئینی نیز بیشترین درصد شباهت برای ژن *HCT* با گیاه *L. sativa* (۹۹/۱۱) به‌دست آمد که تاییدی بر صحت نتایج توالی‌یابی است. هم‌چنین، آنالیز فیلوژنتیکی بر اساس توالی‌یابی ژن‌های *C₄H*، *C₃'H* و *HCT* رابطه فیلوژنتیکی نزدیکی را بین *L. sativa* و *S. arvensis* نشان داد.

نتیجه‌گیری کلی

مطابق با نتایج این پژوهش، روند انتخاب طبیعی در طی تکامل برای ژن‌های *C₄H*، *C₃'H* و *HCT* مثبت بود که این امر نشان‌دهنده چندشکلی پایین ژن‌ها و در نتیجه مستعد نبودن آن‌ها به تغییرات بازها و جهش است. هم‌چنین، نتایج آنالیز نوکلئوتیدها نشان دادند که بیشتر جانشینی بازها از نوع جایگزینی انتقالی بود و بیشترین فراوانی‌های نوکلئوتیدی ژن‌های *C₄H*، *C₃'H* و *HCT* به‌ترتیب به بازهای آدنین، گوانین و تیمین متعلق بودند. نتایج BLAST قطعات ژنی

References

- André, C. M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C. A. A., Nomberto, G., Hoffmann, F., Larondelle, Y & Evers, D. (2009). Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry*, 70(9), 1107-1116. <https://doi.org/10.1016/j.phytochem.2009.07.008>
- Asghari, B., & Dashab, G. R. (2018). Nucleotide sequence analysis of phylogenetic and evolutionary status of BLG gene. *New Cellular and Molecular Biotechnology Journal*, 8(30), 61-71. [In Persian]
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J., Smith, J. A., & Struhl, K. (1992). Short protocols in molecular biology. *New York*, 275, 28764-28773.
- Azizyan, R & Abdollahi Mandoulakani, B. (2023). The effect of drought stress on some morphological, phytochemical, and biochemical characteristics of the medicinal plant field sowthistle (*Sonchus arvensis* L.). *Journal of Crop Breeding*, 15(47), 41-55. [In Persian]
- Azizyan, R & Abdollahi Mandoulakani, B. (2024). Partial coding sequence identification, gene expression analysis, and content of anticancer phenolic compounds in *Sonchus arvensis* L. under drought stress conditions. *Industrial crops and products*, 209: 118030.
- Canter, P. H., Thomas, H., & Ernst, E. (2005). Bringing medicinal plants into cultivation: opportunities and challenges for biotechnology. *TRENDS in Biotechnology*, 23(4), 180-185.
- Comino, C., Hehn, A., Moglia, A., Menin, B., Bourgaud, F., Lanteri, S., & Portis, E. (2009). The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biology*, 9(1), 1-13.
- Elliott, M. B., Irwin, D. M., & Diamandis, E. P. (2006). In silico identification and Bayesian phylogenetic analysis of multiple new mammalian kallikrein gene families. *Genomics*, 88(5), 591-599.
- Farah, A., Monteiro, M., Donangelo, C. M., & Lafay, S. (2008). Chlorogenic acids from green coffee extract are highly bioavailable in humans. *The Journal of Nutrition*, 138(12), 2309-2315.
- Ghaderi, F., & Mandoulakani, B. A. (2024). Investigation of the association between gene expression levels and phenolic compound content in the leaves of *Sonchus arvensis* plants under salinity stress. *Heliyon*, 10(22).
- Gouthamchandra, K., Sudeep, H., Venkatesh, B., & Prasad, K. S. (2017). Chlorogenic acid complex (CGA7), standardized extract from green coffee beans exerts anticancer effects against cultured human colon cancer HCT-116 cells. *Food Science and Human Wellness*, 6(3), 147-153.
- Han, G., Bai, G., Wu, Y., Zhou, Y., Yao, W., & Li, L. (2022). Comparative transcriptome analysis to identify candidate genes related to chlorogenic acid and flavonoids biosynthesis in iridaceae. *Forests*, 13(10), 1632.
- Lu, H., Tian, Z., Cui, Y., Liu, Z., & Ma, X. (2020). Chlorogenic acid: A comprehensive review of the dietary sources, processing effects, bioavailability, beneficial properties, mechanisms of action, and future directions. *Comprehensive Reviews in Food Science and Food Safety*, 19(6), 3130-3158.
- Mahesh, V., Million-Rousseau, R., Ullmann, P., Chabrilange, N., Bustamante, J., Mondolot, L., ... & Campa, C. (2007). Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. *Plant Molecular Biology*, 64, 145-159.
- Masoudi Jozchal, Z., Bagheri, N., Babaeian Jelodar, N., Ranjbar, G., & Farmani, J. (2024). Phytochemical analysis of the medicinal plant terrestrial Orchid (*Orchis Simia*) in the flowering stage. *Journal of Crop Breeding*, 16(2), 53-66. [In Persian]
- Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., Ahmad, F., Babazadeh, D., Modarresi-Ghazani, F., WenHua, L., & Modarresi-Ghazani, F. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomedicine and Pharmacotherapy*, 97, 67-74.
- Nesaj Hosseini, M & Shamar Bakhsh, M. (2011). Methods of phylogenetic analysis. The first edition of Haqshanas Publications, 239 pp. [In Persian]
- Niggeweg, R., Michael, A. J., & Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology*, 22(6), 746-754.
- Peng, X., Li, W., Wang, W., & Bai, G. (2010). Cloning and characterization of a cDNA coding a hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase involved in chlorogenic acid biosynthesis in *Lonicera japonica*. *Planta Medica*, 76(16), 1921-1926.
- Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A., & Boyce-Jacino, M. (1999). Mining SNPs from EST databases. *Genome Research*, 9(2), 167-174.
- Sahraki, H., Mahdinezhad, N., Fakheri, B., & Haddadi, F. (2020). Separation and identification of FEH1 gene in thorny artichoke plant (*Cynara cardunculus*) and its relative expression under the influence of abiotic stresses. [In Persian]
- Schoch, G. A., Morant, M., Abdulrazzak, N., Asnaghi, C., Goepfert, S., Petersen, M., Ullmann, P. . . & Werck-Reichhart, D. (2006). The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. *Environmental Chemistry Letters*, 4, 127-136.

- Sheikhi, M., Fakheri, B., & Mahdinezhad, N. (2019). Study of genetic diversity and phylogenetic relationship of some gourd pumpkin (*Cucurbita pepo*) genotypes using the ITS ribosomal and rbcL chloroplastic genes' loci. *Genetic Engineering and Biosafety Journal*, 8(2), 168-177.
- Sun, C.-H., Yang, C.-Y., & Tzen, J. T. (2018). Molecular identification and characterization of hydroxycinnamoyl transferase in tea plants (*Camellia sinensis* L.). *International Journal of Molecular Sciences*, 19(12), 3938.
- Tajik, N., Tajik, M., Mack, I., & Enck, P. (2017). The potential effects of chlorogenic acid, the main phenolic components in coffee, on health: a comprehensive review of the literature. *European Journal of Nutrition*, 56, 2215-2244.
- Thasa, M. W. (2021). An Overview of the Traditional Uses, Phytochemicals, and Pharmacological Activities of Tempuyung (*Sonchus arvensis* L.). *Journal of Pharmaceutical Sciences and Medicine (IJPSM)*, 6(6), 34-41.
- Wen, H., Wang, W., Jiang, X., Wu, M., Bai, H., Wu, C., & Shen, L. (2022). Transcriptome analysis to identify candidate genes related to chlorogenic acid biosynthesis during development of Korla fragrant pear in Xinjiang. *Food Science and Human Wellness*, 11(4), 854-864.
- Xu, H., Park, N. I., Li, X., Kim, Y. K., Lee, S. Y., & Park, S. U. (2010). Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. *Bioresource Technology*, 101(24), 9715-9722.
- Xu, J., Zhu, J., Lin, Y., Zhu, H., Tang, L., Wang, X., & Wang, X. (2022). Comparative transcriptome and weighted correlation network analyses reveal candidate genes involved in chlorogenic acid biosynthesis in sweet potato. *Scientific Reports*, 12(1), 2770.