Sari Agriculture Sciences and Natural Resources University

Journal of Crop Breeding Vol. 17, Issue 4, 2025 p: 32-43

OPEN ACCESS

Research Paper

Comparison of Machine-Learning Algorithms for Identifying Key Genes Involved in Sorghum's Response to Temperature Stress

Leyla Nazari[1] and Zahra Zinati[2]

1- Department of Crop and Horticultural Science Research, Fars Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shiraz, Iran.
2- Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran, (Corresponding author: zahrazinati@shirazu.ac.ir)

Received: 12 March, 2025 Revised: 20 June, 2025 Accepted: 22 July, 2025

Extended Abstract

Background: Sorghum, a C4 plant, is relatively tolerant to various abiotic stresses. However, its performance is significantly affected by temperatures above 32 °C and below 15 °C. Identifying key genes through gene expression data and feature selection methods is a valuable approach to understanding stress tolerance. Feature selection filters out the most relevant genes using statistical and computational algorithms. Filter-based methods, which are independent of machine-learning algorithms, offer a fast and efficient way to identify relevant features. Combining multiple filter methods allows for a more precise and robust selection of key genes involved in sorghum's response to temperature stress. Therefore, this study aimed to identify key genes involved in cold and heat stress response in sorghum using transcriptomic data and filter-based methods, including Information Gain, Gain Ratio, and Relief.

Methods: In this study, gene expression count data were extracted from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) with accession number GSE225632. The analysis focused on sorghum shoot data under control conditions and temperature stress at four different times of day. The data were divided into two classes, control and stress (cold and heat), and differentially expressed genes were identified using the DESeq2 package in R. Subsequently, top genes were selected using three feature selection algorithms (Information Gain, Gain Ratio, and Relief), and the Venn diagram was used to examine the overlap of genes identified by the algorithms. Two machine-learning algorithms, Bayes Net and Random Forest, were employed for validation. These algorithms were run in WEKA 3.7, and their performance was compared in classifying samples based on the identified features. The classification algorithms were evaluated and compared using metrics, including True Positive Rate (TP Rate), False Positive Rate (FP Rate), Precision, Recall, F1 score, Matthews Correlation Coefficient (MCC), Area Under the ROC Curve (ROC AUC), and Area Under the PRC Curve (PRC AUC). A confusion matrix was used to display classification errors.

Results: In this study, gene expression changes were first analyzed by comparing control conditions with temperature stress (cold and heat). Among 34,129 genes examined, 2,136 genes with significant expression changes at a 0.05 level and a log2FoldChange greater than 1 were selected and used in subsequent feature selection and machine-learning analyses. Key genes responsive to temperature stress were identified using three feature-selection algorithms, with the top 50 genes extracted by each algorithm based on ranking. Across all methods, nine genes were consistently identified by all three feature selection approaches. The performance of two classification models was evaluated in classifying three classes (control, cold stress, and heat stress). The Bayes Net algorithm showed high discriminative accuracy; a TP Rate of 1, FP Rate of 0.21, and Precision of 0.980 were obtained for the control class; a Precision of 1 and a TP Rate of 0.958 were achieved for the cold-stress class; both accuracy and TP Rate were 1 for the heat-stress class. The Random Forest algorithm also demonstrated strong discriminative power. A correct classification rate of 1 and a Precision of 0.96 were observed for the control class; a correct classification rate of 0.958 and a Precision of 1 were obtained for the cold-stress and heat-stress classes, indicating robust performance in accurately identifying stressed samples.

Conclusion: This study demonstrates that identifying and analyzing key genes involved in sorghum's temperature stress response can provide insights into the biological pathways and regulatory networks activated under such conditions. Nine out of the 2,136 differentially expressed genes were consistently identified by three different selection algorithms. These genes

Copyright ©2025.Nazari and Zinati. Published by Sari Agricultural Sciences and Natural Resources University.
This work is licensed under a Creative Commons Attribution-Non-Commercial 4.0 Unported License, which allows users to read, copy, distribute, and make derivative works for non-commercial purposes from the material, as long as the author of the original work is cited properly.

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]

[DOI: 10.61882/jcb.2025.1611]

can serve as potential molecular markers, but further biological validation is necessary across different sorghum varieties. The high accuracy of Bayes Net and Random Forest confirms the strength of these models in distinguishing gene expression patterns between stressed and control conditions. Homology analysis of genes, such as Sobic.004G283300, Sobic.010G100600, and Sobic.006G093500, in Arabidopsis and maize supports their role in heat stress response. However, six genes (Sobic.010G128900, Sobic.001G093100, Sobic.007G168100, Sobic.002G269100, Sobic.006G183701, and Sobic.002G047800) remain uncharacterized, with no documented molecular function. Further research is required to explore the roles of these genes in physiological and stress-related processes. Understanding their functions could contribute to breeding sorghum varieties that are more resilient to environmental stresses, ultimately supporting sustainable agriculture. Field-based and experimental validation of these molecular markers is also recommended to confirm their applicability under real-world farming conditions.

Keywords: Feature selection, Gene expression, Sorghum, Temperature stress

How to Cite This Article: Nazai, L., & Zinati, Z. (2025). Comparison of Machine-Learning Algorithms for Identifying Key Genes Involved in Sorghum's Response to Temperature Stress. *J Crop Breed,* 17(4), 32-43. DOI: 10.61882/jcb.2025.1611

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]

[DOI: 10.61882/jcb.2025.1611]

دانشگاه علوم کشاورزی و منابع طبیعی ساری

مقاله پژوهشی

مقایسه الگوریتم‌های یادگیری ماشین جهت تعیین ژن‌های کلیدی سورگوم در پاسخ به تنش‌های دمایی

لیلا نظری[۱] و زهرا زینتی[۲]

۱- بخش تحقیقات زراعی و باغی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی فارس، سازمان تحقیقات، آموزش و ترویج کشاورزی، شیراز، ایران
۲- بخش اگرواکولوژی، دانشکده کشاورزی و منابع طبیعی داراب، دانشگاه شیراز، شیراز، ایران، (نویسنده مسوول: zahrazinati@shirazu.ac.ir)

تاریخ دریافت: ۱۴۰۳/۱۲/۲۲ تاریخ ویرایش: ۱۴۰۴/۰۳/۳۰ تاریخ پذیرش: ۱۴۰۴/۰۴/۳۱
صفحه ۳۲ تا ۴۳

چکیده مبسوط

مقدمه و هدف: سورگوم به‌عنوان یک گیاه چهار کربنه، اگرچه نسبت به تنش‌های غیر زیستی مختلف نسبتاً متحمل است، اما عملکرد آن به‌طور قابل توجهی تحت تأثیر دماهای بالاتر از ۳۲ درجه سانتی‌گراد و پایین‌تر از ۱۵ درجه سانتی‌گراد قرار می‌گیرد. انتخاب ژن‌های کلیدی داده‌های بیان ژن با استفاده از روش‌های انتخاب ویژگی، فرآیندی است که در آن به کمک الگوریتم‌های آماری و محاسباتی، مهم‌ترین ژن‌های موثر در تحمل به تنش شناسایی می‌شوند. نتیجه این فرآیند، فیلتر شدن ژن‌هایی است که بیشترین ارتباط را با تنش مورد مطالعه دارند. روش فیلتر یکی از روش‌های انتخاب ویژگی است که مستقل از الگوریتم‌های یادگیری ماشین عمل می‌کند. ترکیب روش‌های مختلف فیلتر، امکان انتخاب بهینه و دقیق ویژگی‌های موثر در پاسخ ژنتیکی سورگوم به تنش‌های دمایی را فراهم می‌آورد. با توجه به این موضوع، هدف از این مطالعه شناسایی ژن‌های کلیدی مؤثر در پاسخ به تنش‌های دمایی (سرما و گرما) در گیاه سورگوم از طریق داده‌های ترانسکریپتومی و به‌کارگیری روش‌های فیلتر شامل کسب اطلاعات (Information Gain)، نسبت کسب اطلاعات (Gain Ratio) و Relief است.

مواد و روش‌ها: در این مطالعه، داده‌های شمارش مربوط به بیان ژن از پایگاه GEO (https://www.ncbi.nlm.nih.gov/geo/) با شماره دسترسی GSE225632 استخراج شدند. تجزیه و تحلیل بر روی داده‌های برگ گیاه سورگوم در شرایط کنترل و تحت تنش دما در چهار زمان مختلف روز انجام شد. داده‌ها به دو کلاس کنترل و تنش (سرما و گرما) تقسیم شدند و با استفاده از بسته DESeq2 در نرم‌افزار R، ژن‌های با تغییرات بیان شناسایی گردیدند. سپس با استفاده از سه الگوریتم انتخاب ویژگی (Information Gain، Gain Ratio و Relief) ژن‌های برتر انتخاب شدند و با استفاده از نمودار ون، تعداد ژن‌های مشترک بین الگوریتم‌ها بررسی شد. برای اعتبارسنجی نتایج، از الگوریتم‌های یادگیری ماشین Bayes Net و Random Forest استفاده شد. این الگوریتم‌ها در نرم‌افزار Weka 3.7 اجرا شدند و عملکرد آن‌ها در طبقه‌بندی نمونه‌ها بر اساس ویژگی‌های شناسایی‌شده مقایسه شد. برای مقایسه و ارزیابی الگوریتم‌های طبقه‌بندی از معیارهای نرخ مثبت واقعی (TP Rate)، نرخ مثبت کاذب (FP Rate)، دقت (Precision)، بازخوانی (Recall)، معیار F، ضریب همبستگی متیوز (MCC)، مساحت زیر منحنی نرخ مثبت واقعی در مقابل نرخ مثبت کاذب (ROC) و مساحت زیر منحنی دقت در مقابل یادآوری (PRC) استفاده شد. جهت نمایش اشتباهات طبقه‌بندی از ماتریس درهم‌ریختگی استفاده شد.

یافته‌ها: در این پژوهش، ابتدا تحلیل تغییرات بیان ژن‌ها با مقایسه بین شرایط کنترل و تنش‌های سرمایی و گرمایی انجام شد. از مجموع ۳۴۱۲۹ ژن بررسی‌شده، ۲۱۳۶ ژن با تغییر بیان معنادار در سطح ۰/۰۵ که دارای log2FoldChange بیشتر از یک بودند، انتخاب شدند و در مراحل بعدی برای انتخاب ویژگی و تحلیل‌های یادگیری ماشین مورد استفاده قرار گرفتند. برای استخراج ژن‌های کلیدی پاسخ‌دهنده به تنش‌های دمایی، سه الگوریتم انتخاب ویژگی اجرا شد. در هر الگوریتم، ۵۰ ژن برتر بر اساس رتبه‌بندی استخراج شدند. در مجموع، نه ژن به‌طور مشترک توسط سه الگوریتم انتخاب ویژگی شناسایی شدند. کارایی دو مدل طبقه‌بندی برای تفکیک سه کلاس کنترل، تنش سرما و تنش گرما ارزیابی شد. الگوریتم Bayes Net دقت بالایی در تفکیک کلاس‌ها داشت. برای کلاس کنترل، نرخ حساسیت (TP Rate) برابر با ۱، نرخ مثبت کاذب (FP Rate) برابر با ۰/۰۲۱ و دقت (Precision) برابر با ۰/۹۸۰ بود. در کلاس تنش سرما، دقت ۱ و نرخ حساسیت ۰/۹۵۸ بدست آمد. در تنش گرما، تمامی نمونه‌ها به‌درستی طبقه‌بندی شدند (دقت و نرخ حساسیت برابر با ۱).
الگوریتم Random Forest نیز دقت بالایی در تشخیص کلاس‌ها از خود نشان داد. در کلاس کنترل، نرخ صحیح طبقه‌بندی برابر با ۱ و دقت برابر با ۰/۹۶ بود. برای کلاس‌های تنش سرما و تنش گرما، نرخ صحیح طبقه‌بندی ۰/۹۵۸ و دقت ۱ به‌دست آمد که بیانگر عملکرد قوی الگوریتم در شناسایی صحیح نمونه‌های تحت تنش بود.

نتیجه‌گیری: نتایج این مطالعه نشان می‌دهند که شناسایی و تحلیل ژن‌های کلیدی موثر در پاسخ گیاه سورگوم به تنش‌های دمایی زمینه‌ساز مطالعه دقیق‌تر مسیرها و شبکه‌های بیولوژیکی دخیل در تنظیم پاسخ‌های سلولی است و می‌تواند به افزایش فهم ما از مکانیسم‌های مولکولی این گیاه تحت شرایط تنش کمک شایانی نماید. از میان ۲۱۳۶ ژن شناسایی‌شده با تغییرات بیان معنادار، نه ژن به‌طور مشترک توسط سه الگوریتم انتخاب ویژگی شناسایی شدند. برای تایید این ژن‌ها به عنوان شاخص‌های مولکولی مناسب، نیاز است که آزمایش‌های بیولوژیکی مستقل بر روی رقم‌های مختلف سورگوم انجام شوند تا تغییرات بیان آن‌ها در شرایط تجربی تأیید گردد. عملکرد مطلوب الگوریتم‌های یادگیری ماشین شامل Bayes Net و Random Forest در طبقه‌بندی نمونه‌ها نشان‌دهنده قدرت و دقت این مدل‌ها در تفکیک بیان ژن‌ها در شرایط تنش دما و کنترل است. بررسی همولوگ ژن *Sobic.004G283300* در آرابیدوپسیس و ذرت، و همولوگ‌های ژن‌های *Sobic.010G100600* و *Sobic.006G093500* در آرابیدوپسیس بر نقش آن‌ها در تنش‌های گرمایی دلالت دارد. علاوه بر این، بررسی همولوگ‌های ژن‌های شناسایی‌شده در گونه‌های مدل مانند برنج و ذرت در پایگاه‌های داده ژنومی و مقالات علمی نشان می‌دهد که تاکنون گزارش معتبر و مستندی در خصوص نقش دقیق و عملکرد زیستی ژن‌های *Sobic.010G128900*، *Sobic.001G093100*، *Sobic.007G168100*، *Sobic.002G269100*، *Sobic.006G183701* و *Sobic.002G047800* در پاسخ به تنش‌های دمایی وجود ندارد. این شش ژن به‌عنوان ژن‌های ناشناخته (*uncharacterized*) معرفی شده‌اند و عملکرد آن‌ها در سطح مولکولی تاکنون مورد مطالعه قرار نگرفته است. بنا بر این، انجام تحقیقات بیشتر برای مشخص کردن نقش این ژن‌ها در فرآیندهای فیزیولوژیکی و پاسخ به تنش‌های دمایی در سورگوم ضروری است تا درک بهتری از عملکرد آن‌ها در واکنش به تنش‌های محیطی به دست آید. این اطلاعات می‌تواند زمینه‌ساز توسعه ارقام جدید و مقاوم در برابر شرایط نامساعد زیست‌محیطی باشد و به پایداری کشاورزی در آینده کمک کند. پیشنهاد می‌شود که مطالعات بیشتری بر روی ژن‌های شناسایی‌شده و همچنین ژن‌های ناشناخته انجام شوند تا عملکرد دقیق آن‌ها در زمینه‌های مختلف تنش‌های زیستی و غیر زیستی بیشتر مورد بررسی قرار گیرد. همچنین، اقدامات تجربی و میدانی برای ارزیابی کارایی این شاخص‌های مولکولی در شرایط واقعی کشاورزی ضروری است. انجام چنین عملیاتی می‌تواند به بهبود مستمر و پایداری تولیدات زراعی کمک کند.

واژه‌های کلیدی: انتخاب ویژگی، بیان ژن، سورگوم، تنش دمایی

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]

[DOI: 10.61882/jcb.2025.1611]

مقدمه

گیاهان چهار کربنه (C4) به‌طور کلی نسبت به تنش غیر زیستی مقاوم‌تر در نظر گرفته می‌شوند (Pardo & VanBuren, 2021). مطالعات ترانسکریپتومیک در گیاهان چهارکربنه مانند ذرت و سورگوم در پاسخ به گرما، سرما یا خشک‌سالی، تنوع قابل‌توجهی را در پاسخ‌های بیان ژن بسته به ژنوتیپ نشان می‌دهند (Abdel-Ghany *et al*., 2020; Frey *et al*., 2015; Shi *et al*., 2017; Sunoj *et al*., 2017). اگرچه سورگوم به تنش‌های غیر زیستی مختلف نسبتاً متحمل شناخته شده است، افزایش دما بالاتر از حد مطلوب (۳۲ درجه سانتی‌گراد) می‌تواند عملکرد سورگوم را کاهش دهد (Prasad *et al*., 2017; Tack *et al*., 2017). در مقایسه با سایر محصولات، سورگوم که گیاهی گرمسیری محسوب می‌شود، نسبت به تنش سرما، به‌ویژه در اوایل فصل رشد، حساسیت بالایی دارد. دماهای پایین‌تر از ۱۵ درجه سانتی‌گراد می‌توانند منجر به کاهش استقرار گیاه و افت عملکرد آن شوند (Moghimi *et al*., 2019). به‌طور خلاصه، سورگوم، اگرچه در شرایط نامطلوب رشد می‌کند، اما دمای کمتر یا بالاتر از حد مطلوب به‌ترتیب باعث ایجاد تنش سرما و گرما می‌شود که بر فیزیولوژی و رشد کلی تأثیر منفی می‌گذارد.

در بیشتر مطالعات تنش در گیاهان، شناسایی ژن‌ها یا آلل‌های کلیدی درگیر در تحمل به تنش از گونه‌ها یا ژنوتیپ‌هایی که تحمل یا سازگاری بالایی را نشان می‌دهند، به‌عنوان یک رویکرد مؤثر و کارآمد در نظر گرفته می‌شود (Rai & Rai, 2020). بنا بر این، سورگوم به‌عنوان یک گونه هدف با توجه به تحمل بالاتر آن به تنش‌های غیر زنده نسبت به دیگر محصولات اصلی مانند برنج، ذرت و نیشکر برای چنین اهداف شناسایی ژنی شناخته شده است (Tu *et al*., 2023).

تحلیل افتراقی بیان ژن‌ها روشی مؤثر برای بررسی مسیرهای زیستی سلولی است که با شناسایی ژن‌های کلیدی، درک بهتری را از مکانیسم‌های مولکولی مرتبط با فرآیندهای فیزیولوژیکی، انتقال پیام، متابولیسم‌های اولیه و ثانویه و پاسخ به تنش فراهم می‌کند (Bordbar *et al*., 2023). تحلیل الگوی بیان ژن‌ها در بافت‌ها و مراحل مختلف رشد گیاه، در شرایط متفاوت تنش‌های محیطی، می‌تواند بینش‌های ارزشمندی را در باره سازوکارهای مولکولی و نحوه واکنش گیاهان به این تنش‌ها ارائه دهد (Rezaei Musa Dargh *et al*., 2024).

انتخاب ویژگی‌ها (ژن‌ها) در مجموعه‌داده‌های بیان ژن در دهه‌های اخیر توجه زیادی را به‌خود جلب کرده است (Abdelwahab *et al*., 2022; Nazari *et al*., 2023). نظری و همکاران (۲۰۲۳) الگوریتم‌های مبتنی‌بر یادگیری ماشین و یادگیری عمیق را برای طبقه‌بندی ژن‌ها در حالت کنترل و تنش زیستی در ذرت مورد استفاده قرار دادند و به‌منظور افزایش عملکرد این الگوریتم‌ها، انتخاب ویژگی‌های کلیدی (ژن‌ها) از طریق الگوریتم Relief اعمال شد (Nazari *et al*., 2023). در مطالعه‌ای، روش ترکیبی انتخاب ژن شامل ReliefF، کسب اطلاعات (Information Gain) و LASSO (least absolute shrinkage and selection operator) جهت شناسایی و پیش‌بینی ژن‌های کلیدی مرتبط با انواع تنش‌های غیر زیستی در آرابیدوپسیس پیشنهاد شد (Nazari *et al*., 2023).

از آن‌جایی که مجموعه داده‌های با بعد بالا معمولاً شامل ویژگی‌های تکراری و غیر مفیدی هستند که به افزایش پیچیدگی محاسباتی و زمان اجرای مدل مورد نظر منجر می‌شوند، بنابراین انتخاب ویژگی‌ها برای انتخاب ویژگی‌های مفید و حذف ویژگی‌های غیرضروری ضروری است. این عمل نه‌تنها زمان اجرا یا زمان آموزش را کاهش می‌دهد بلکه دقت مدل را نیز افزایش می‌دهد. بر اساس این مدل، می‌توان نمونه‌ها در داده‌ها را به کلاس‌های خود دسته‌بندی کرد (James *et al*., 2023).

پیش‌بینی الگوهای بیان ژن یکی از بسیاری از کاربردهای بیولوژیکی است که به‌واسطه یادگیری ماشین تحت‌نظارت (ML)پیشرفت‌های زیادی داشته است (Moore *et al*., 2022). به طور کلی، این حوزه به دو دسته یادگیری ماشین کلاسیک و یادگیری نمایش (representation learning) تقسیم می‌شود. اگرچه یادگیری نمایش اخیراً با کاربردهای یادگیری عمیق توجه زیادی را جلب کرده‌است، یادگیری ماشین کلاسیک همچنان برخی مزایا را دارد، از جمله سهولت در پیکربندی، نیازمندی‌های محاسباتی کمتر، قابلیت اجرا بر روی داده‌های کوچک‌تر و مهم‌تر، و تفسیر آسان‌تر مدل. در سال‌های اخیر، یادگیری ماشین کلاسیک با موفقیت برای مطالعه تنظیم بیان ژن ناشی از استرس‌های غیر زیستی در گیاهان به کار رفته است. برای مثال، آزودی و همکاران از رویکرد ML برای طبقه‌بندی بیان ژن در پاسخ به استرس همزمان گرما و خشکی استفاده کردند (Azodi *et al*., 2020). در مطالعه‌ای، از رویکردهای فراتحلیلی، یادگیری ماشین و همچنین یادگیری عمیق برای شناسایی الگوهای مختلف بیان ژن در ذرت تحت شرایط کنترل و استرس زیستی استفاده شد (Nazari *et al*., 2023). به‌کارگیری داده‌های مربوط به بیان ژن در جمعیت‌هایی که برای افزایش مقاومت به تنش انتخاب شده‌اند، می‌تواند به‌عنوان روشی جایگزین برای شناسایی ژن‌های مرتبط با مسیرهای پاسخ به تنش مورد استفاده قرار گیرد. این موضوع از آن جهت اهمیت دارد که سطوح بیان ژن، ویژگی‌هایی قابل‌توارث هستند (Gibson & Weir, 2005). بونوت و همکاران اثر تنش دمایی و زمان روز را روی دینامیک بیان ژن در ژنوتیپ‌های حساس و مقاوم به حرارت در سورگوم بررسی نمودند (Bonnot *et al*., 2023). همچنین، خوشه‌های ژنی (ماژول‌ها) که به‌صورت همبسته و همزمان در پاسخ به شرایط تنش دمایی فعالیت دارند و ساختارهای شبکه‌ای و مسیرهای بیولوژیک بررسی شدند. در مطالعه‌ای، با بهره‌گیری از یادگیری ماشین، داده‌های ترنسکریپتومیک RNA-seq از ۱۸ مطالعه مستقل تجزیه و تحلیل شدند، که شامل یازده نوع تنش غیر زیستی و سه نوع تنش زیستی بود. با انجام انتخاب ویژگی‌ها، با استفاده از پنج الگوریتم وزن‌دهی، ۴۰۰ ژن اصلی اولویت‌بندی شدند و ژن‌های کلیدی به‌عنوان بیومارکر بالقوه شناسایی شدند (Panahi, 2024).

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

استفاده از رویکردهای پیشرفته انتخاب ویژگی و ارزیابی آن‌ها با الگوریتم‌های یادگیری ماشین، روشی نوآورانه برای شناسایی ژن‌های مهم در پاسخ به تنش است. در مطالعه‌ای توسط هایوری و همکاران (Haury *et al*., 2011)، مقایسه ۳۲ روش انتخاب ویژگی بر روی داده‌های بیان ژن نشان داد که روش‌های مختلف تأثیر قابل توجهی بر دقت پیش‌بینی دارند استفاده از یک الگوریتم انتخاب ویژگی ممکن است منجر به سوگیری در نتایج شود و روش‌های مختلف انتخاب ویژگی می‌توانند نتایج متفاوتی ارائه دهند. رویکردهای ترکیبی در انتخاب ویژگی می‌توانند به شناسایی ژن‌هایی منجر شوند که در شرایط و داده‌های مختلف نیز به‌عنوان ژن‌های کلیدی شناسایی می‌شوند. مطالعه‌ای توسط کولومبلی و همکاران (2022) نشان داد که استفاده از روش‌های ترکیبی انتخاب ویژگی توانست به افزایش پایداری و قابلیت تعمیم نتایج در تحلیل داده‌های بیان ژن کمک کند (Colombelli *et al*., 2022).

در همین راستا، هدف از این پژوهش شناسایی ژن‌های کلیدی مؤثر در پاسخ به تنش‌های دمایی (سرما و گرما) در گیاه سورگوم با استفاده از داده‌های ترانسکریپتومی و به‌کارگیری روش‌های انتخاب ویژگی شامل Information Gain، Gain Ratio و ReliefF است. همچنین، برای ارزیابی دقت ژن‌های منتخب در طبقه‌بندی شرایط تنش، از الگوریتم‌های یادگیری ماشین Bayes Net و Random Forest بهره گرفته شد. این رویکرد می‌تواند زمینه‌ساز شناسایی شاخص‌های مولکولی مؤثر در اصلاح ژنتیکی ارقام مقاوم به تنش‌های محیطی باشد.

مواد و روش‌ها

داده‌های خوانش خام RNA_seq (raw_counts) مربوط به GSE225632 (GPL20690)، از پایگاه داده GEO (https://www.ncbi.nlm.nih.gov/geo/) استخراج شدند. در مطالعه فوق، ژنوتیپ‌های سورگوم شامل RTx430 (حساس به سرما و گرما)، Macia (متحمل به گرما) و SC224 (متحمل به سرما) مورد استفاده قرار گرفتند. گیاهان در اتاق‌های با دمای کنترل‌شده رشد داده شدند و شرایط دمایی روزانه بر پایه ۳۰ درجه سانتی‌گراد روز و ۲۰ درجه سانتی‌گراد شب تنظیم شدند. پس از ۷ روز از کاشت، گیاهان در چهار زمان مختلف روز تحت استرس سرما (۱۰ °C) یا گرما (۴۲ °C) به مدت یک ساعت قرار گرفتند تا پاسخ سریع ژن‌های بیان‌شده بررسی شود. به این منظور، گیاهان در ۰، ۵، ۸ و ۱۴ ساعت پس از طلوع خورشید به اتاق‌های تنش سرما یا گرما منتقل شدند و به‌مدت یک ساعت تحت تنش قرار گرفتند. پس از این مدت، برگ‌های کاملا توسعه‌یافته بالایی گیاهان جمع‌آوری و بلافاصله در نیتروژن مایع منجمد و در دمای ۸۰- °C نگهداری شدند.

در مطالعه حاضر، داده‌های خام بیان ژنی ۹۶ نمونه به دو کلاس کنترل و تنش (سرما و گرما) تقسیم شدند (۴۸ نمونه کنترل، ۲۴ نمونه تنش سرما و ۲۴ نمونه تنش گرما) و ژن‌های افتراقی در شرایط تنش و کنترل با استفاده از پکیج DESeq2 (v. 1.34.0) در محیط rstudio شناسایی شدند. ماتریس داده‌های بیان ۹۶ نمونه با ۲۱۳۶ ژن جهت انتخاب ژن‌ها با استفاده از تجزیه و تحلیل انتخاب ویژگی به‌عنوان ورودی به الگوریتم‌های Information gain، Gain ratio و ReliefF داده شد. داده‌ها به سه کلاس کنترل، تنش سرما و تنش گرما تقسیم شدند. برای انتخاب ویژگی از نرم‌افزار Weka 3.7 استفاده شد و ۵۰ ژن برتر که توسط هر الگوریتم رتبه‌بندی شده‌بودند انتخاب شدند. نموار ون (Venn Diagram) جهت یافتن ژن‌های مشترک بین این سه الگوریتم ترسیم شد (https://bioinformatics.psb.ugent.be/webtools/Venn/).

جهت اعتبارسنجی ژن‌های مشترک در طبقه‌بندی نمونه‌های تنش گرما، تنش سرما و کنترل از دو الگوریتم Bayes Net و Random Forest استفاده شد. الگوریتم‌های فوق در محیط نرم افزار Weka 3.7 اجرا شدند.

انتخاب ویژگی

کسب اطلاعات (Information Gain) معیاری است که مقدار اطلاعات فراهم‌شده توسط یک ویژگی را اندازه‌گیری می‌کند. از کسب اطلاعات برای محاسبه مقدار اطلاعات فراهم‌شده توسط هر ویژگی در مجموعه‌های داده‌های با بعد بالا استفاده می‌شود. سپس، K ویژگی اول که بیشترین اطلاعات را فراهم می‌کنند، انتخاب می‌شوند تا یک زیرمجموعه‌ی ویژگی کاندیدا تشکیل دهند. این معیار به عنوان مقدار اطلاعاتی تعریف می‌شود که از ویژگی برای تعیین کلاس به‌دست می‌آید.

$$Gain(A)=H(D)-H(D|A)$$

Gain(A) میزان اطلاعاتی را که از ویژگی A برای پیش‌بینی کلاس‌ها به‌دست می‌آید، اندازه‌گیری می‌کند. H(D) آنتروپی مجموعه داده D است که نشان‌دهنده میزان نااطمینانی (uncertainty) یا اختلاط اطلاعات در کلاس‌ها قبل از تقسیم‌بندی براساس ویژگی A است. هرچه مقدار آنتروپی بیشتر باشد، نااطمینانی بیشتر است. H(D|A) آنتروپی مجموعه داده با توجه به ویژگی A بعد از تقسیم‌بندی است. این مقدار نشان‌دهنده میزان نا اطمینانی در کلاس‌ها درون زیرمجموعه‌هایی است که براساس ویژگی A ایجاد شده‌اند. به‌عبارت دیگر، این آنتروپی پس از آن است که اطلاعات مرتبط با ویژگی در اختیار قرار دارد (Kononenko & Kukar, 2007).

نسبت کسب اطلاعات (Gain Ratio) معیاری است که برای سنجش توانایی یک ویژگی در تقسیم‌بندی داده‌های ورودی به‌کار می‌رود. این معیار گامی فراتر از معیار کسب اطلاعات ساده است و با تقسیم کردن مقدار کسب اطلاعات یک ویژگی به میزان نااطمینانی (ناخالصی) ناشی از آن ویژگی، به‌دست می‌آید. هدف اصلی از استفاده از نسبت کسب اطلاعات، کاهش تمایل به انتخاب ویژگی‌های با مقادیر بالای دامنه (مثل ویژگی‌هایی که تعداد مقادیر منحصربه فرد زیادی دارند) است. فرمول نسبت کسب اطلاعات به صورت زیر بیان می‌شود:

$$\text{Gain Ratio}(A) = \frac{\text{Gain}(A)}{\text{Split Info}(A)}$$

Split Info(A) میزان نااطمینانی ناشی از تقسیم داده‌ها بر اساس ویژگی A است و به‌صورت زیر محاسبه می‌شود:

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

$$\text{Split Info}(A) = -\sum_{v \in Values(A)} \frac{|D_v|}{|D|} \log_2 \frac{|D_v|}{|D|}$$

در این فرمول، $|D_v|$ تعداد نمونه‌ها در زیرمجموعه‌ای است که با مقدار v از ویژگی A تولیدشده و $|D|$ تعداد کل نمونه‌ها در مجموعه داده است (Liu & Motoda, 2012).

الگوریتم ReliefF یکی از روش‌های پیشرفته برای انتخاب ویژگی‌ها در داده‌های با ابعاد بالا است. این الگوریتم یک روش مبتنی‌بر یادگیری است که برای شناسایی ویژگی‌های مهم در پیش‌بینی کلاس‌ها (یا خروجی‌ها) طراحی شده است. الگوریتم ReliefF به‌صورت زیر عمل می‌کند (Kononenko, 1994):

۱- تعیین نمونه‌های همسایه: برای هر نمونه x_i در داده، K همسایه نزدیک (نزدیک‌ترین نمونه‌های مشابه) پیدا می‌شود که شامل همسایه‌های هم‌کلاس (یعنی نمونه‌هایی که کلاس مشابه دارند) و غیر هم‌کلاس (یعنی نمونه‌هایی که کلاس متفاوت دارند) است.

۲- محاسبه وزن ویژگی‌ها: وزن هر ویژگی A_j بر اساس تفاوت آن در نمونه‌های هم‌کلاس و غیر هم‌کلاس محاسبه می‌شود. ویژگی‌هایی که قادر به تفکیک بهتر دسته‌ها هستند، وزن بالاتری دریافت می‌کنند.

۳- تفاوت آن در نمونه‌های هم‌کلاس و غیر هم‌کلاس محاسبه می‌شود. ویژگی‌هایی که قادر به تفکیک بهتر دسته‌ها هستند، وزن بالاتری دریافت می‌کنند.

معیار ارزیابی الگوریتم‌های طبقه‌بندی

کلاس‌های پیش‌بینی‌شده با استفاده از چندین معیار مبتنی بر مقادیر نرخ مثبت واقعی (TP) و نرخ مثبت کاذب (FP) ارزیابی شدند:

نرخ مثبت واقعی = TP Rate = TP/(TP+FN)
نرخ مثبت کاذب = FP Rate = FP/(FP+TN)
دقت (Precision) = TP/(TP+FP)
بازخوانی (Recall) = TP/(TP+FN)
معیار F = ۲*(Precision*Recall)/(Precision+Recall)

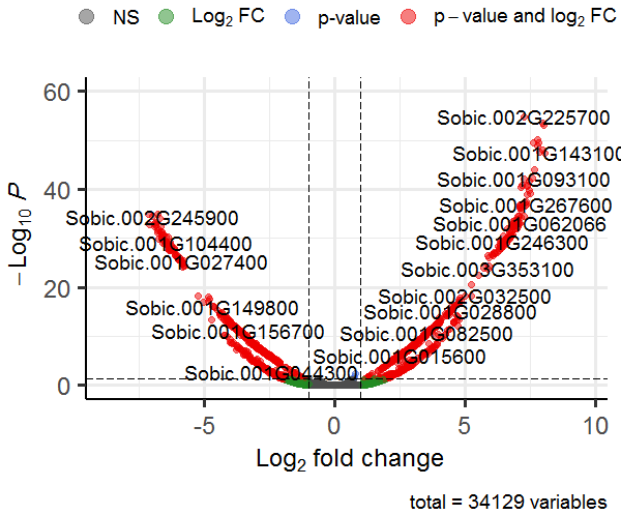
MCC = (TP*TN–FP*FN) / √(TP+FP)(TP+FN)(TN+FP)(TN+FN)
ROC = مساحت زیر منحنی نرخ مثبت واقعی در مقابل نرخ مثبت کاذب
PRC = مساحت زیر منحنی دقت در مقابل یادآوری

کدهای R استفاده‌شده در این مطالعه و توالی پروتئینی ژن‌های کلیدی به‌صورت داده‌های تکمیلی ارائه شده‌اند.

نتایج و بحث

در این مطالعه، ابتدا تحلیل داده‌های بیان ژن با هدف شناسایی ژن‌های دارای بیان افتراقی بین شرایط کنترل و تنش (سرما و گرما) انجام شد. در مجموع، از بین ۳۴۱۲۹ ژن مورد بررسی، تعداد ۲۱۳۶ ژن با مقدار لگاریتم تغییرات بیان بیشتر از یک، به‌عنوان ژن‌های دارای تغییر بیان معنادار انتخاب شدند (شکل ۱).

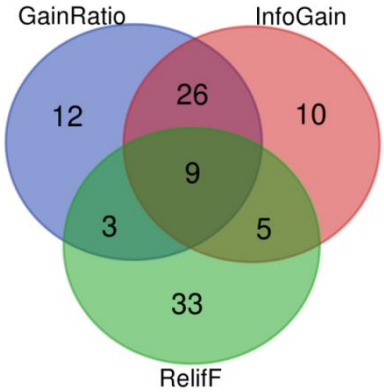
این ژن‌ها در مراحل بعدی برای انتخاب ویژگی و تحلیل‌های یادگیری ماشین مورد استفاده قرار گرفتند. به‌منظور شناسایی ژن‌های کلیدی مرتبط با پاسخ به تنش‌های دمایی، سه الگوریتم انتخاب ویژگی شامل Information Gain، Gain Ratio و ReliefF به کار گرفته شدند. برای هر الگوریتم، ۵۰ ژن برتر بر اساس رتبه‌بندی انتخاب شدند. با استفاده از نمودار ون، تعداد ۴۳ ژن در دو یا چند الگوریتم به‌صورت مشترک شناسایی شدند. از میان آن‌ها، نه ژن (جدول ۱) به‌طور هم‌زمان توسط هر سه الگوریتم انتخاب ویژگی تشخیص داده شدند (شکل ۲). جدول ۱ شامل شناسه‌های ژن در نسخه‌های ۳/۱ و ۱/۴ سورگوم، شناسه‌های UniProt، ژن‌های همولوگ در گونه‌های مذکور و توضیح درباره عملکرد زیستی هر ژن است.



شکل ۱- ژن‌های دارای بیان افتراقی بین شرایط کنترل و تنش (سرما و گرما). از ۳۴۱۲۹ ژن مورد بررسی، تعداد ۲۱۳۶ ژن با مقدار لگاریتم تغییرات بیان بیشتر از یک، به‌عنوان ژن‌های دارای تغییر بیان معنادار انتخاب شدند. محور y مقادیر padj است.

Figure 1. Differentially expressed genes between control and stress conditions (cold and heat). Out of 34,129 genes examined, 2,136 genes with a log2 fold change greater than 1 were selected as significantly differentially expressed genes. The y-axis represents the adjusted p-values (padj).

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]



شکل ۲- نمودار ون نمایش‌دهنده ژن‌های مشترک و اختصاصی شناسایی‌شده توسط سه الگوریتم انتخاب ویژگی Information Gain، Gain Ratio و ReliefF

Figure 2. The Venn diagram illustrating the shared and unique genes identified by the three feature selection algorithms: Information Gain, Gain Ratio, and ReliefF.

جدول ۱- مشخصات ژن‌های منتخب مشترک بین سه الگوریتم انتخاب ویژگی Gain Ratio، Information Gain و ReliefF و ژن‌های همولوگ آن‌ها در گونه‌های ذرت، برنج و آرابیدوپسیس.

Table 1. Characteristics of the common selected genes identified by the three feature selection algorithms: Gain Ratio, Information Gain, and ReliefF, along with their homologs in maize, rice, and Arabidopsis.

شناسه ژن در نسخه ۳.۱ ID V3.1	شناسه ژن در نسخه ۱.۴ ID V1.4	شناسه در پایگاه دادهUniProt UniProt	ذرت Maize	برنج Rice	آرابیدوپسیس Arabidopsis	توضیح عملکرد زیستی ژن Description
Sobic.010G1 28900	*Sb10g01 1733*	*C5Z1J9*	*GRMZM2 G153194*		*AT3G53050*	D-galactoside/L-rhamnose binding SUEL lectin protein پروتئین لک‌تین SUEL که به-D گالاکتوزید-L / رامنوز متصل می‌شود
Sobic.001G0 93100	*Sb01g00 8370*	*C5X176*	*GRMZM2 G074395*	*LOC_Os03g59350*		Scopoletin glucosyltransferase گلوکوزیل‌ترانسفراز اسکوپولیتین
Sobic.004G2 83300			*GRMZM2 G434203*	*LOC_Os02g45420*	*AT1G12630*	A member of the DREB subfamily A-4 of ERF/AP2 transcription factor family عضوی از زیرخانواده DREB نوع A-4 از خانواده عوامل رونویسی ERF/AP2
Sobic.007G1 68100						
Sobic.010G1 00600	*Sb10g00 8750*	*C5Z7K1*	*GRMZM2 G405133*	*LOC_Os06g13580*	*AT3G28370*	SCEP2; Coiled coil protein involved in formation of synaptonemal complex. Localized to the central region of the SC and interacts with SCEP1 and ZYP1. SCEP2؛ پروتئین مارپیچ پیچ‌خورده که در تشکیل کمپلکس سیناپتونمال نقش دارد. این پروتئین در ناحیه مرکزی کمپلکس سیناپتونمال قرار گرفته و با SCEP1 و ZYP1 تعامل دارد.
Sobic.006G0 93500	*Sb06g01 7850*	*C5Y979*	*GRMZM2 G331701*	*LOC_Os04g36750*	*AT4G10250*	Columbia endomembrane-localized small heat shock protein پروتئین کوچک شوک گرمایی واقع در اندوممبران کلومبیا
Sobic.002G2 69100	*Sb02g03 0310*	*C5X625*	*GRMZM2 G124037*	*LOC_Os09g35020*		Dehydration-responsive element-binding protein 1H
Sobic.006G1 83701	*Sb06g02 5800*	*C5YE39*				1Hپروتئین اتصال‌دهنده به عنصر پاسخ‌دهنده به کم‌آبی
Sobic.002G0 47800						

در این مطالعه، عملکرد الگوریتم Bayes Net برای طبقه‌بندی نمونه‌ها به سه کلاس کنترل، تنش سرما و تنش گرما ارزیابی شد. نتایج جدول ۲ نشان می‌دهند که الگوریتم Bayes Net به‌طور کلی در شبیه‌سازی طبقه‌بندی دارای دقت بالایی است. برای کلاس کنترل، نرخ حساسیت (TP Rate) برابر با ۱، نرخ مثبت کاذب (FP Rate) برابر با ۰/۰۲۱ و دقت (Precision) برابر با ۰/۹۸۰ بود. در کلاس تنش سرما، الگوریتم توانست دقت ۱ و نرخ حساسیت ۰/۹۵۸ را به دست آورد. در تنش گرما، الگوریتم Bayes Net عملکرد عالی داشت و تمامی نمونه‌ها به درستی طبقه‌بندی شدند (دقت و نرخ حساسیت برابر با ۱). متوسط عملکرد این الگوریتم در تمام کلاس‌ها، شامل نرخ حساسیت، دقت، و معیار F برابر ۰/۹۹۰ بود که نشان‌دهنده کارایی بالای الگوریتم در طبقه‌بندی نمونه‌ها است.

همچنین، طبق نتایج جدول ۳، ماتریس درهم‌ریختگی طبقه‌بندی به‌طور دقیق‌تری جزئیات اشتباهات طبقه‌بندی را نشان می‌دهد. برای نمونه‌های کنترل، تمام ۴۸ نمونه به درستی شناسایی شدند. در کلاس تنش سرما، از ۲۴ نمونه، ۲۳ نمونه به‌درستی طبقه‌بندی شدند و تنها ۱ نمونه به اشتباه به کلاس کنترل تخصیص یافت. در کلاس تنش گرما، تمامی نمونه‌ها به درستی طبقه‌بندی شدند.

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

جدول ۲- نتایج طبقه‌بندی کلاس‌های کنترل، تنش گرما و تنش سرما با استفاده از الگوریتم Bayes Net
Table 2. Classification results of control, heat stress, and cold stress classes using the Bayes Net algorithm

مدل طبقه‌بندی	کلاس Class	نرخ مثبت واقعی TP Rate	نرخ مثبت کاذب FP Rate	دقت Precision	بازخوانی Recall	معیار F F-Measure	ضریب همبستگی Matthews MCC	مساحت زیر منحنی ROC ROC Area	مساحت زیر منحنی PRC Area
Bayes Net	کنترل Control	1.000	0.021	0.980	1.000	0.990	0.979	0.980	0.944
	تنش سرمایی Cold stress	0.958	0.000	1.000	0.958	0.979	0.972	0.974	0.972
	تنش گرمایی Heat stress	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
	میانگین Average	0.990	0.010	0.990	0.990	0.990	0.983	0.984	0.965
Random Forest	کنترل Control	1.000	0.042	0.960	1.000	0.980	0.959	0.984	0.963
	تنش سرمایی Cold stress	0.958	0.000	1.000	0.958	0.979	0.972	0.970	0.969
	تنش گرمایی Heat stress	0.958	0.000	1.000	0.958	0.979	0.972	1.000	1.000
	میانگین Average	0.979	0.021	0.98	0.979	0.979	0.966	0.984	0.974

جدول ۳- ماتریس درهم ریختگی طبقه‌بندی کلاس‌های کنترل، تنش گرما و تنش سرما با استفاده از الگوریتم Bayes Net
Table 3. Confusion matrix of the classification of control, heat stress, and cold stress classes using the Bayes Net algorithm

		طبقه‌بندی شده Classified		
		کنترل Control	تنش سرمایی Cold stress	تنش گرمایی Heat stress
داده‌های واقعی	کنترل Control	48	0	0
	تنش سرمایی Cold stress	1	23	0
	تنش گرمایی Heat stress	0	0	24

برای ارزیابی بیشتر عملکرد مدل طبقه‌بندی، الگوریتم Random Forestنیز مورد استفاده قرار گرفت. همان طور که در جدول ۲ مشاهده می‌شود، این الگوریتم نیز دقت بالایی در تشخیص کلاس‌ها از خود نشان داد. در کلاس کنترل، نرخ صحیح طبقه‌بندی برابر با ۱ و دقت برابر با ۰/۹۶ بود. برای کلاس‌های تنش سرما و تنش گرما، نرخ صحیح طبقه‌بندی ۰/۹۵۸ و دقت ۱ به‌دست آمد که بیانگر عملکرد قوی الگوریتم در شناسایی صحیح نمونه‌های تحت تنش است .میانگین کلی معیارهای ارزیابی شامل نرخ صحیح طبقه‌بندی ۰/۹۷۹، دقت ۰/۹۸ و معیار F برابر با ۰/۹۷۹ بود که عملکرد الگوریتم را در سطحی بسیار مطلوب نشان می‌دهد.

در ماتریس درهم‌ریختگی (جدول ۴)، مشخص است که از ۴۸ نمونه کلاس کنترل، تمامی نمونه‌ها به‌درستی طبقه‌بندی شدند. در کلاس تنش سرما، تنها یک نمونه به اشتباه به کلاس کنترل اختصاص یافت. در کلاس تنش گرما نیز، از ۲۴ نمونه، یک نمونه اشتباهاً در کلاس کنترل قرار گرفت. در مجموع، نتایج نشان می‌دهند که الگوریتم Random Forest نیز مانند Bayes Netتوانسته است با دقت بسیار بالا، نمونه‌ها را به درستی در کلاس‌های موردنظر دسته‌بندی نماید.

جدول ۴- ماتریس درهم ریختگی طبقه‌بندی کلاس‌های کنترل، تنش گرما و تنش سرما با استفاده از الگوریتم Random Forest
Table 4. Confusion matrix of the classification of control, heat stress, and cold stress classes using the Random Forest algorithm

		طبقه‌بندی شده Classified		
		کنترل Control	تنش سرمایی Cold stress	تنش گرمایی Heat stress
داده‌های واقعی	کنترل Control	48	0	0
	تنش سرمایی Cold stress	1	23	0
	تنش گرمایی Heat stress	1	0	23

نتایج این مطالعه نشان می‌دهند که ترکیب سه الگوریتم انتخاب ویژگی Gain Ratio، Information Gain و Relief با روش‌های یادگیری ماشین، روشی نوین و کارآمد در شناسایی ژن‌های کلیدی پاسخ به تنش دمایی در سورگوم فراهم می‌آورد. در مقایسه با مطالعات مشابه که تنها بر یک الگوریتم تکیه دارند، استفاده از همپوشانی چند الگوریتم دقت و اعتمادپذیری انتخاب ژن‌ها را افزایش داد و احتمال شناسایی کاذب ژن‌ها را کاهش داد. همچنین، اعتبارسنجی بیوانفورماتیکی از طریق مقایسه ژن‌های منتخب با همولوگ‌های شناخته‌شده در گونه‌های مدل مانند برنج و آرابیدوپسیس صورت گرفت. این موضوع موید نقش این ژن‌ها در مسیرهای پاسخ به تنش محیطی است.

علاوه بر این، برای بررسی صحت عملکرد الگوریتم‌ها، معیارهای طبقه‌بندی شامل دقت، حساسیت، و معیار F برای الگوریتم‌های Bayes Net و Random Forest مورد سنجش قرار گرفتند که نتایج رضایت‌بخشی برای تفکیک نمونه‌های

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

کنترل و تنش حاصل شدند. از میان ۲۱۳۶ ژن با بیان افتراقی معنادار، نه ژن به‌طور مشترک توسط سه الگوریتم انتخاب ویژگی Gain Ratio، Information Gain و Relief شناسایی شدند که می‌توانند به‌عنوان شاخص‌های مولکولی قابل اعتماد برای تشخیص شرایط تنشی در سورگوم مطرح شوند. مقایسه این ژن‌ها با ژن‌های همولوگ در گونه‌های مدل مانند برنج، ذرت و آرابیدوپسیس، نقش آن‌ها را در مکانیسم‌های دفاعی و سازگاری گیاه در مواجهه با تنش‌های حرارتی تأیید می‌کند.

همولوگ ژن *Sobic.010G128900* در آرابیدوپسیس *AT3G53050* است. این ژن یک پروتئین D-galactoside/L-rhamnose binding SUEL lectin را کد می‌کند. بر اساس بررسی‌های انجام‌شده، اطلاعاتی درباره نقش پروتئین لکتین متصل‌شونده به -D گالاکتوزید-L/رامنوز (SUEL) در پاسخ به تنش دمایی در گیاهان زراعی گزارش نشده است؛ با این‌حال، مطالعاتی وجود دارند که به نقش کلی لکتین‌ها در تحمل تنش‌های محیطی اشاره می‌کنند (Marothia *et al*., 2023). همچنین، مطالعه‌ای در مورد ژن‌های لکتین *OsEUL* در برنج نشان داده است که بیان این ژن‌ها تحت تأثیر تنش‌های مختلف قرار می‌گیرد (Lambin *et al*., 2020). علاوه بر این، تجزیه و تحلیل ژن‌های کدکننده لکتین در گوجه‌فرنگی نشان داده است که برخی از این ژن‌ها در پاسخ به تنش‌های خشکی و گرما تنظیم می‌شوند (Osman *et al*., 2024). با توجه به این مطالعات، می‌توان نتیجه گرفت که لکتین‌ها در پاسخ به تنش‌های محیطی نقش دارند، اما برای تعیین نقش مشخص پروتئین SUEL در تنش دمایی، تحقیقات بیشتری لازم است.

ژن *Sobic.001G093100* کدکننده scopoletin glucosyltransferase است. براساس بررسی‌های انجام‌شده، گزارشی درباره نقش ژن scopoletin glucosyltransferase در پاسخ به تنش دمایی در گیاهان زراعی موجود نیست. با این‌حال، مطالعاتی در گیاه تنباکو (*Nicotiana tabacum*) نشان داده‌اند که این آنزیم در پاسخ‌های دفاعی گیاه به عوامل بیماری‌زا نقش دارد.

در یک مطالعه، کاهش بیان ژن *TOGT* (scopoletin glucosyltransferase) منجر به کاهش تجمع scopolin (فرم گلوکوزیله شده scopoletin) شد که با افزایش استرس اکسیداتیو و کاهش مقاومت به ویروس موزاییک تنباکو (TMV) همراه بود (Chong *et al*., 2002). همچنین، در مطالعه‌ای دیگر، افزایش بیان همین ژن باعث افزایش فعالیت scopoletin UGT و تجمع بیشتر scopoletin و scopolin پس از آلودگی با TMV شد. این افزایش با تشکیل سریع‌تر ضایعات نکروتیک مرتبط بود، اما تأثیری بر میزان ویروس در هر ضایعه نداشت (Gachon *et al*., 2004). با توجه به این مطالعات، می‌توان نتیجه گرفت که scopoletin glucosyltransferase در پاسخ‌های دفاعی گیاه به عوامل بیماری‌زا نقش دارد، اما نقش دقیق آن در پاسخ به تنش دمایی نیاز به تحقیقات بیشتری دارد.

همولوگ ژن *Sobic.004G283300* در آرابیدوپسیس AT1G12630 است. ژن AT1G12630 در گیاه *Arabidopsis thaliana* یک پروتئین از خانواده DREB زیرگروه A-4 از فاکتورهای رونویسی ERF/AP2 را کد می‌کند. در گیاه آرابیدوپسیس، فاکتور رونویسیDREB2A به‌عنوان عضوی از زیرخانواده A-4 شناخته شده است که در پاسخ به تنش‌های خشکی و گرما نقش دارد. مطالعه‌ای نشان داده است که *DREB2A* با فعال‌سازی بیان ژن‌های مرتبط با تنش گرمایی، به بهبود تحمل گیاه در برابر این تنش کمک می‌کند (Sakuma *et al*., 2006). در ذرت (*Zea mays*)، ژن همولوگ *ZmDREB2A* شناسایی شده است که بیان آن تحت‌تأثیر تنش‌های سرما، خشکی، شوری و گرما افزایش می‌یابد. بیان بیش از حد *ZmDREB2A* در گیاهان تراریخت باعث بهبود تحمل به تنش‌های خشکی و گرما شده است. این نتایج نشان می‌دهند که *ZmDREB2A* ممکن است نقش دوگانه‌ای در تنظیم بیان ژن‌های پاسخ‌دهنده به تنش‌های آبی و گرمایی داشته باشد (Qin *et al*., 2007). در گیاه آرابیدوپسیس، فاکتور رونویسی DEAR4، که عضوی از خانواده DREB/CBF است، در تنظیم پیری برگ و پاسخ به تنش‌های متعدد نقش دارد. این نشان می‌دهد که اعضای مختلف این خانواده می‌توانند در پاسخ به تنش‌های محیطی مشارکت داشته باشند (Zhang *et al*., 2020). همچنین، در گیاه پنبه (*Gossypium hirsutum*)، تحلیل سوپرفامیل AP2/EREBPدر کتان نشان داده است که بسیاری از ژن‌های GhDREB/ERF تحت تأثیر تنش‌های محیطی مانند سرما، گرما و خشکی قرار می‌گیرند. این نشان‌دهنده نقش گسترده این فاکتورهای رونویسی در پاسخ به تنش‌های مختلف است (Liu & Zhang, 2017). با توجه به این مطالعات، می‌توان نتیجه گرفت که اعضای زیرخانواده A-4 از خانواده فاکتورهای رونویسی ERF/AP2، مانند DREB2A، در تنظیم پاسخ گیاهان زراعی به تنش گرمایی نقش مهمی دارند.

همولوگ ژن *Sobic.010G100600* در آرابیدوپسیس AT3G28370 است. این ژن پروتئین مارپیچ–مارپیچ (coiled-coil) را کد می‌کند که در تشکیل ساختار سیناپتونمال (synaptonemal complex) نقش دارد. این پروتئین در منطقه مرکزی این ساختار قرار می‌گیرد و با SCEP1 و ZYP1 تعامل دارد. کمپلکس سیناپتونمال ساختاری پروتئینی است که در طی میوز بین کروموزوم‌های همولوگ تشکیل می‌شود و برای برقراری نوترکیبی ژنتیکی صحیح بسیار حیاتی است. پروتئین‌های دارای ساختار Coiled-coil که در تشکیل کمپلکس سیناپتونمال دخیل هستند، در شرایط تنش گرمایی دچار اختلال عملکرد می‌شوند. این اختلالات می‌توانند منجر به کاهش باروری و عملکرد تولیدمثلی در گیاهان زراعی شوند. مطالعه‌ای توسط (Zhao *et al*., 2025) نشان داد که تنش گرمایی در *Arabidopsis thaliana* باعث کاهش تجمع و پایداری ZYP1 که یک پروتئین مارپیچ–مارپیچ در کمپلکس سیناپتونمال است شد. این موضوع منجربه اختلال در سیناپس کروموزوم‌های همولوگ و کاهش نوترکیبی میوزی می‌گردد.

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

در نتیجه، باروری کاهش می‌یابد و بذرهای کمتری تشکیل می‌شوند (Ning *et al*., 2021). اگرچه بیشتر مطالعات بر *Arabidopsis* انجام شده‌اند، اما به‌دلیل محافظت بالا (conservation) ساختار کمپلکس سیناپتونمال بین گیاهان، انتظار می‌رود که همین اثرات در گیاهان زراعی مانند سورگوم، برنج و ذرت نیز رخ دهد.

همولوگ ژن *Sobic.006G093500* در آرابیدوپسیس *AT4G10250* است. این ژن کدکننده پروتئین کوچک شوک‌حرارتی مستقر در سیستم غشای داخلی است. پروتئین‌های کوچک شوک حرارتی (sHSPs) که در بخش‌هایی از سیستم‌های غشایی داخلی مثل شبکه اندوپلاسمی، گلژی، واکوئل، و ویزیکول قرار دارند و نقش مهمی در پاسخ گیاهان به تنش گرمایی ایفا می‌کنند. این پروتئین‌ها به‌عنوان چاپرون‌های مولکولی عمل کرده، از تجمع و دناتوره شدن پروتئین‌های سلولی در دماهای بالا جلوگیری می‌کنند (Waters & Vierling, 2020).

ژن *Sobic.002G269100* کدکننده پروتئین متصل‌شونده به عنصر پاسخ‌گوی بی‌آبی 1H است. براساس بررسی‌های انجام‌شده، اطلاعات مستقیمی درباره نقش ژن *Dehydration-Responsive Element-Binding Protein 1H (DREB1H)* در پاسخ به تنش دمایی در گیاهان زراعی در دسترس نیست. با این‌حال، اعضای خانواده *DREB1*، از جمله *DREB1C*، *DREB1E* و *DREB1G*، در تحمل تنش‌های محیطی مانند سرما، خشکی و شوری نقش دارند (Wang *et al*., 2022). مطالعه‌ای در برنج نشان داده است که جهش در ژن‌های *OsDREB1C*، *OsDREB1E* و *OsDREB1G* حساسیت به سرما را افزایش می‌دهد، اما این گیاهان توانایی سازگاری با سرما را حفظ می‌کنند (Wang *et al*., 2022). اگرچه نقش دقیق *DREB1H* در پاسخ به تنش دمایی مشخص نشده است، اما با توجه به عملکرد سایر اعضای خانواده DREB در تحمل تنش‌های محیطی، ممکن است این ژن نیز در پاسخ به تنش دمایی نقش داشته باشد. تحقیقات بیشتری برای تأیید این فرضیه مورد نیاز است.

بررسی‌های انجام‌شده در پایگاه‌های داده ژنومی و مقالات علمی نشان می‌دهند که تاکنون گزارشی معتبر و مستندی در خصوص نقش ژن‌های *Sobic.002G047800*، *Sobic.006G183701* و *Sobic.007G168100* در پاسخ به تنش‌های محیطی منتشر نشده است. این سه ژن به‌عنوان ژن‌های ناشناخته (*uncharacterized*) معرفی شده‌اند و عملکرد آن‌ها در سطح مولکولی تاکنون مورد مطالعه قرار نگرفته است. لذا این ژن‌ها نیازمند مطالعات بیشتر برای تعیین نقش آن‌ها در فرآیندهای فیزیولوژیکی و پاسخ به تنش‌های زیستی و غیر زیستی در سورگوم هستند.

در کل، تحلیل ژن‌های کلیدی شناسایی‌شده در پاسخ به تنش دمایی در سورگوم نشان می‌دهد که پاسخ گیاه سورگوم به تنش دما، یک فرآیند پیچیده و چندوجهی است که شامل مسیرهای تنظیمی متنوعی از جمله فعالیت فاکتورهای رونویسی مانند اعضای خانواده‌ی DREB، پروتئین‌های حفاظتی مانند sHSPهای مستقر در غشای داخلی، آنزیم‌های دفاعی مانند scopoletin glucosyltransferase و مولکول‌های اتصال‌دهنده نظیر لکتین‌ها است. همچنین، برخی ژن‌ها با عملکردهای مربوط به میوز و تولیدمثل نظیر پروتئین‌های coiled-coil مرتبط با کمپلکس سیناپتونمال نیز در برابر تنش گرمایی حساس هستند. این تنوع در پاسخ‌های سلولی و مولکولی نشان می‌دهد که سورگوم برای مقابله با دمای بالا از مکانیسم‌های مختلفی بهره می‌برد. با این حال، برای برخی ژن‌ها اطلاعات مستقیمی از نقش آن‌ها در پاسخ به تنش دمایی وجود ندارد و یا عملکرد آن‌ها هنوز ناشناخته باقی مانده‌است؛ این موضوع بر لزوم انجام مطالعات عملکردی بیشتر در سورگوم تأکید دارد. همچنین در این مطالعه، افزایش بیان نه ژن تحت تنش دمایی مشاهده شد که می‌توان آن را در چارچوب مکانیسم‌های دفاعی و تنظیمی گیاهان در برابر دمای بالا توجیه نمود. ژن‌هایی مانند *DREB2A* و *DREB1H* به‌عنوان فاکتورهای رونویسی شناخته‌شده در پاسخ به تنش‌های غیر زیستی، احتمالاً نقش مهمی در فعال‌سازی مسیرهای حفاظتی تحت تنش گرما ایفا می‌کنند. افزایش بیان small HSP به‌عنوان بخشی از پاسخ کلاسیک حرارتی، در حفظ ساختار و عملکرد پروتئین‌ها مؤثر است (Waters, 2013). ژن coiled-coil synaptonemal complex احتمالاً در حفظ ساختارهای میوزی در دمای بالا نقش ایفا می‌کند (Fu *et al*., 2022). همچنین، نقش عمومی لکتین‌ها در تنظیم پاسخ‌های دفاعی و سیگنالینگ در شرایط تنش (Lannoo & Van Damme, 2014)، افزایش بیان ژن *SUEL* را در مواجهه با تنش دمایی توجیه می‌کند. افزایش بیان ژن‌های ناشناخته تحت تنش دمایی، بیانگر احتمال نقش آن‌ها در فرآیندهای دفاعی یا تنظیمی خاص است که نیازمند بررسی‌های بیشتر است. به طور کلی، افزایش بیان این ژن‌ها منعکس‌کننده‌ی پاسخ هماهنگ و پیچیده‌ی گیاه به تنش گرمایی، با هدف حفظ بقاء و تعادل سلولی در شرایط تنش دمایی است.

در بخش طبقه‌بندی، الگوریتم‌های Bayes Net و Random Forest هر دو عملکرد قابل‌توجهی در تفکیک سه کلاس کنترل، تنش گرما و تنش سرما نشان دادند، به‌طوری‌که دقت کلی و معیارهای ارزیابی مانند نرخ حساسیت، دقت، و معیارF در هر دو الگوریتم بسیار بالا و قابل اطمینان بودند. این نتایج نشان می‌دهند که می‌توان با بهره‌گیری از ترکیب داده‌های ترانسکریپتومی و روش‌های پیشرفته تحلیل داده، به درک عمیق‌تری از مکانیسم‌های مولکولی پاسخ به تنش‌های زیستی در گیاهان دست یافت و زمینه را برای توسعه راهکارهای اصلاح ژنتیکی فراهم ساخت. این پژوهش توانست مشخص کند که گروهی از ژن‌ها اعم از فاکتورهای رونویسی، پروتئین‌های حفاظتی، مولکول‌های سیگنالی یا حتی ژن‌های ناشناخته و کم‌بررسی‌شده، در شرایط تنش دمایی در سورگوم فعال‌سازی شده‌اند. با تحلیل جامع همولوگ‌ها در گونه‌های مختلف و مقایسه با پژوهش‌های پیشین، نشان داده شد که بخشی از این پاسخ‌های مولکولی، در سورگوم نیز فعال می‌گردد. همچنین شناسایی ژن‌های ناشناخته با تغییر بیان معنادار می‌تواند مسیر کشف مکانیسم‌ها یا مسیرهای بیوشیمیایی جدید در مقاومت به تنش دمایی را فراهم کند. بنابراین، یافته‌های

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]
[DOI: 10.61882/jcb.2025.1611]

این تحقیق با فراهم کردن یک مجموعه منسجم از ژن‌های تنظیمی، حفاظتی و ناشناخته نقشه‌ای از ارتباط شبکه‌های ژنی مرتبط با بقاء در شرایط دمای بالا ارائه می‌کنند. این امر نه‌تنها به درک دقیق‌تر نحوه تعامل مسیرهای مولکولی نظیر فاکتورهای رونویسی DREB، sHSP، و پروتئین‌های کمپلکس سیناپتونمال کمک می‌نماید بلکه ژن‌های جدیدی را نیز به‌عنوان هدف برای مطالعات عملکردی و اعتبارسنجی تجربی در آینده معرفی می‌کند. با وجود ارائه شواهد قوی در تحلیل داده‌های ترانسکریپتومی و عملکرد مناسب الگوریتم‌های انتخاب ویژگی و یادگیری ماشین، برای تأیید قطعی نقش ژن‌های شناسایی‌شده در پاسخ به تنش دمایی، انجام آزمایش‌های تجربی از قبیل بیان‌سنجی (qPCR) و ارزیابی عملکردی در شرایط واقعی ضروری است. این مطالعات تکمیلی می‌توانند اعتبار نتایج بیوانفورماتیکی را تقویت کرده، زمینه را برای کاربردهای عملی در اصلاح نژاد سورگوم فراهم کنند.

References

Abdel-Ghany, S. E., Ullah, F., Ben-Hur, A., & Reddy, A. S. (2020). Transcriptome analysis of drought-resistant and drought-sensitive sorghum (Sorghum bicolor) genotypes in response to PEG-induced drought stress. *International Journal of Molecular Sciences*, 21(3), 772.

Abdelwahab, O., Awad, N., Elserafy, M., & Badr, E. (2022). A feature selection-based framework to identify biomarkers for cancer diagnosis: A focus on lung adenocarcinoma. *Plos One*, *17*(9), e0269126.

Azodi, C. B., Lloyd, J. P., & Shiu, S.-H. (2020). The cis-regulatory codes of response to combined heat and drought stress in Arabidopsis thaliana. *NAR Genomics and Bioinformatics*, *2*(3), lqaa049 .

Bonnot, T., Somayanda, I., Jagadish, S. K., & Nagel, D. H. (2023). Time of day and genotype sensitivity adjust molecular responses to temperature stress in sorghum. *The Plant Journal*, 116(4), 1081-1096. https://doi.org/10.1111/tpj.16467

Bordbar, M., Darvishzadeh, R., & Pazhouhandeh, M. (2023). Molecular Techniques for Plants Gene Expression Analysis at the Transcriptomics Level [Research]. *Journal of Crop Breeding*, *15*(45), 83-104. https://doi.org/10.61186/jcb.15.45.83 [In Persian]

Chong, J., Baltz, R., Schmitt, C., Beffa, R., Fritig, B., & Saindrenan, P. (2002). Downregulation of a pathogen-responsive tobacco UDP-Glc :phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. *The Plant Cell*, 14(5), 1093-1107.

Colombelli, F., Kowalski, T. W., & Recamonde-Mendoza, M. (2022). A hybrid ensemble feature selection design for candidate biomarkers discovery from transcriptome profiles. *Knowledge-Based Systems*, *254*, 109655.

Frey, F. P., Urbany, C., Hüttel, B., Reinhardt, R., & Stich, B. (2015). Genome-wide expression profiling and phenotypic evaluation of European maize inbreds at seedling stage in response to heat stress. *BMC Genomics*, *16*, 1-15.

Fu, H., Zhao, J., Ren, Z., Yang, K., Wang, C., Zhang, X., & Chen, C. (2022). Interfered chromosome pairing at high temperature promotes meiotic instability in autotetraploid Arabidopsis. *Plant Physiology*, *188*(2), 1210-1228.

Gachon, C., Baltz, R., & Saindrenan, P. (2004). Over-expression of a scopoletin glucosyltransferase in Nicotiana tabacum leads to precocious lesion formation during the hypersensitive response to tobacco mosaic virus but does not affect virus resistance. *Plant Molecular Biology*, *54*, 137-146.

Gibson, G., & Weir, B. (2005). The quantitative genetics of transcription. *TRENDS in Genetics*, *21*(11), 616-623.

Haury, A.-C., Gestraud, P., & Vert, J.-P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *Plos One*, *6*(12), e28210.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In *An introduction to statistical learning: With applications in Python* (pp. 15-67). Springer.

Kononenko, I. (1994, April). Estimating attributes: Analysis and extensions of RELIEF. In European conference on machine learning (pp. 171-182). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kononenko, I., & Kukar, M. (2007). Chapter 6 - Measures for Evaluating the Quality of Attributes. In I. Kononenko & M. Kukar (Eds.), *Machine Learning and Data Mining* (pp. 153-180). Woodhead Publishing. https://doi.org/https://doi.org/10.1533/9780857099440.153

Lambin, J., Demirel Asci, S., Dubiel, M., Tsaneva, M., Verbeke, I., Wytynck, P., . . . & Van Damme, E. J. (2020). OsEUL lectin gene expression in rice: stress regulation, subcellular localization and tissue specificity. *Frontiers in Plant Science*, 11, 185.

Lannoo, N., & Van Damme, E. J .(2014). Lectin domains at the frontiers of plant defense. *Frontiers in Plant Science*, *5*, 397.

Liu, C., & Zhang, T. (2017). Expansion and stress responses of the AP2/EREBP superfamily in cotton. *BMC Genomics*, *18*, 1-16.

Liu, H., & Motoda, H. (2012). *Feature selection for knowledge discovery and data mining* (Vol. 454). Springer science & business media.

Marothia, D., Kaur, N., Jhamat, C., Sharma, I., & Pati, P. K. (2023). Plant lectins: Classical molecules with emerging roles in stress tolerance. *International Journal of Biological Macromolecules*, *244*, 125272.

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]

[DOI: 10.61882/jcb.2025.1611]

Moghimi, N., Desai, J. S., Bheemanahalli, R., Impa, S. M., Vennapusa, A. R., Sebela, D., . . .& Jagadish, S. K. (2019). New candidate loci and marker genes on chromosome 7 for improved chilling tolerance in sorghum. *Journal of Experimental Botany*, *70*(12), 3357-3371.

Moore, B. M., Lee, Y. S., Wang, P., Azodi, C., Grotewold, E., & Shiu, S.-H. (2022). Modeling temporal and hormonal regulation of plant transcriptional response to wounding. *The Plant Cell*, *34*(2), 867-888.

Nazari, L., Aslan, M. F., Sabanci, K., & Ropelewska, E. (2023). Integrated transcriptomic meta-analysis and comparative artificial intelligence models in maize under biotic stress. *Scientific Reports*, *13*(1), 15899. https://doi.org/10.1/038s41598-023-42984-4

Nazari, L., Ghotbi, V., Nadimi, M., & Paliwal, J. (2023). A Novel Machine-Learning Approach to Predict Stress-Responsive Genes in Arabidopsis. *Algorithms*, *16*(9), 407.

Nazari, L., Zinati, Z., & Bagnaresi, P. (2023). Identification of biomarker genes from multiple studies for abiotic stress in maize through machine learning. *Journal of Biosciences*, *49*(1), 1. https://doi.org/10.1007/s12038-023-00392-w

Ning, Y., Liu, Q., Wang, C., Qin, E., Wu, Z., Wang, M., . . . & Liu, H. (2021). Heat stress interferes with formation of double-strand breaks and homolog synapsis. *Plant Physiology*, *185*(4), 1783-1797.

Osman, M. E. M., Osman, R. S. H., Elmubarak, S. A., Ibrahim, M. A., Abakar, H. B. M., Dirar, A. I., & Konozy, E. H. E. (2024). In silico analysis of L-and G-type lectin receptor kinases in tomato: evolution, diversity, and abiotic responses. *BMC Genomics*, *25*(1), 1143.

Panahi, B. (2024). Transcriptome signature for multiple biotic and abiotic stress in barley (Hordeum vulgare L.) identifies using machine learning approach. *Current Plant Biology*, *40*, 100416. https://doi.org/https://doi.org/10.1016/j.cpb.2024.100416

Pardo, J., & VanBuren, R. (2021). Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *The Plant Cell*, *33*(11), 3391-3401.

Prasad, P. V., Bheemanahalli, R., & Jagadish, S. K. (2017). Field crops and the fear of heat stress—opportunities, challenges and future directions. *Field Crops Research*, *200*, 114-121.

Qin, F., Kakimoto, M., Sakuma, Y., Maruyama, K., Osakabe, Y., Tran, L. S. P., . . . & Yamaguchi-Shinozaki, K. (2007). Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in *Zea mays* L. *The Plant Journal*, *50*(1), 54-69.

Rai, K. K., & Rai, A. C. (2020). Recent transgenic approaches for stress tolerance in crop plants. *Sustainable Agriculture in the Era of Climate Change*, 533-556.

Rezaei Musa Dargh, S., Abdollahi Mandoulakani, B., & Ghasemzadeh, R. (2024). The Effect of Iron Deficiency Stress on the Relative Expression of ZIP3, ZIP6, and ZIP7 Genes in Bread Wheat (Triticum aestivum L.) Cultivars [Research]. *Journal of Crop Breeding*, *16*(3), 52-63. https://doi.org/10.61186/jcb.16.3.52 [In Persian]

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proceedings of the National Academy of Sciences*, *103*(49), 18822-18827.

Shi, J., Yan, B., Lou, X., Ma, H., & Ruan, S. (2017). Comparative transcriptome analysis reveals the transcriptional alterations in heat-resistant and heat-sensitive sweet maize (*Zea mays* L.) varieties under heat stress. *BMC Plant Biology*, *17*, 1-10.

Sunoj, V. J., Somayanda, I. M., Chiluwal, A., Perumal, R., Prasad, P. V., & Jagadish, S. K. (2017). Resilience of pollen and post-flowering response in diverse sorghum genotypes exposed to heat stress under field conditions. *Crop Science*, *57*(3), 1658-1669.

Tack, J., Lingenfelser, J., & Jagadish, S. K. (2017). Disaggregating sorghum yield reductions under warming scenarios exposes narrow genetic diversity in US breeding programs. *Proceedings of the National Academy of Sciences*, *114*(35), 9296-9301.

Tu, M., Du, C., Yu, B., Wang, G., Deng, Y., Wang, Y., ... & Li, Y. (2023). Current advances in the molecular regulation of abiotic stress tolerance in sorghum via transcriptomic, proteomic, and metabolomic approaches. *Frontiers in Plant Science,* 14, 1147328. https://doi.org/10.3389/fpls.2023.1147328

Wang, H., Lu, S., Guan, X., Jiang, Y., Wang, B., Hua, J., & Zou, B. (2022). Dehydration-responsive element binding protein 1C, 1E, and 1G promote stress tolerance to chilling, heat, drought, and salt in rice. *Frontiers in Plant Science*, *13*, 851731.

Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany*, *64*(2), 391-403.

Waters, E. R., & Vierling, E. (2020). Plant small heat shock proteins–evolutionary and functional diversity. *New Phytologist*, *227*(1), 24-37.

Zhang, Z., Li, W., Gao, X., Xu, M., & Guo, Y. (2020). DEAR4, a member of DREB/CBF family, positively regulates leaf senescence and response to multiple stressors in Arabidopsis thaliana. *Frontiers in Plant Science*, *11*, 367.

Zhao, J., Fu, H., Wang, Z., Zhang, M., Liang, Y., Cui, X., . . . & Zhang, Y. (2025). Genetic variation in Arabidopsis thaliana reveals the existence of natural heat resilience factors for meiosis. *Plant Physiology*, *197*(1), kiae671.

[Downloaded from jcb.sanru.ac.ir on 2026-01-29]

[DOI: 10.61882/jcb.2025.1611]