



متاآنالیز داده‌های ترنسکریپتومی به منظور شناسایی ژن‌های موثر بر تنش شوری در آرابیدوپسیس

امیرناصر ابراهیمی^{۱*}، رضا شیرزادیان خرم‌آباد^{۱*}، علی جمالیان^۲، زهرا پزشکیان^۳

^۱* گروه بیوتکنولوژی کشاورزی، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت

^۲ گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه گیلان، رشت

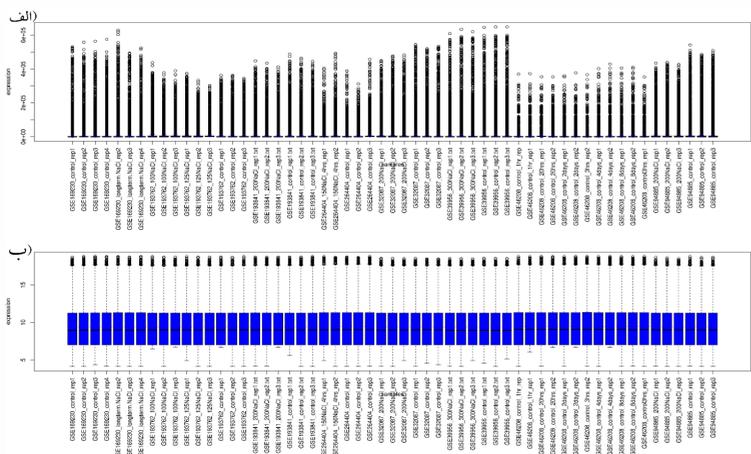
^۳ پژوهشکده آبی‌پروری آب‌های داخلی، موسسه تحقیقات علوم شیلاتی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، بندرانزلی

از اعضای خانواده LEA بوده و این موضوع نشان‌دهنده اثربخشی اعضای این خانواده بر تنش شوری می‌باشد. (Bhardwaj et al., 2013).

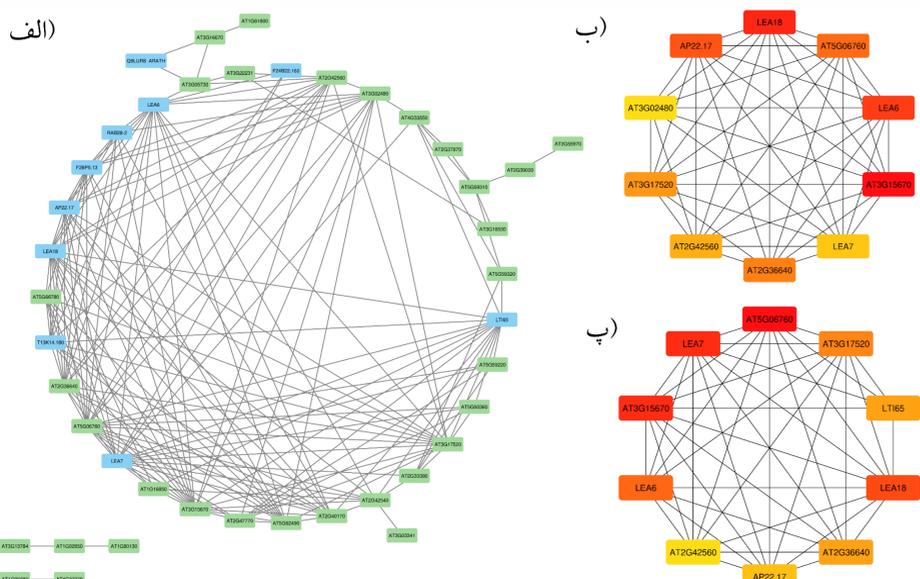
❖ در این مطالعه ۵۰ ژن کلیدی (۴۲ افزایش بیان، ۸ کاهش بیان) در آرابیدوپسیس با دقت ۹۹٪ در واکنش به تنش شوری شناسایی شدند. اعضای خانواده LEA به عنوان ژن‌های با اثربخشی بیشتر شناخته شدند. تحلیل GO فرآیندهای پاسخ به شوری، ABA، کم‌آبی، سرما و اتصال لیپیدی در خارج سلول و سیتوسل را تایید می‌کند.

جدول ۱. ده ژن برتر شناسایی شده موثر بر تنش شوری توسط متاآنالیز داده‌های میکروآرای

شماره دسترسی	تغییر بیان	مقادیر p تعدیل شده	توضیحات
AT1G52690	۵.۲۸	۶.۷۶-۱۲	LEA family protein
AT3G02480	۵.۱۰	۲.۳۶-۱۶	LEA family protein
AT2G37870	۴.۸۸	۵.۶۶-۱۷	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G59310	۴.۸۳	۱.۸۶-۱۳	lipid transfer protein 4
AT2G47770	۴.۴۷	۳.۴۶-۱۵	TSPO(outer membrane tryptophan-rich sensory protein)-like protein
AT3G05730	-۲.۶۷	۲.۲۹-۱۵	defensin-like protein
AT5G42530	-۲.۶۸	۴.۵۶-۱۲	uncharacterized protein
AT3G16530	-۲.۸۶	۱.۲۶-۱۵	Legume lectin family protein
AT3G22231	-۲.۹۳	۴.۲۶-۰۹	pathogen and circadian controlled 1
AT3G16670	-۳.۸۶	۱.۵۶-۱۴	Pollen Ole e 1 allergen and extensin family protein



شکل ۱. نمودار جعبه‌ای داده‌های میکروآرای، (الف) پیش از نرمال‌سازی، (ب) بعد از نرمال‌سازی. قرارگیری دامنه میان چارکی و خط میانه تمامی نمونه‌ها در یک راستا تایید اجرای صحیح عملیات کنترل کیفیت و نرمال‌سازی است.



شکل ۲. شبکه‌های ارتباطات پروتئین-پروتئین (الف) نتیجه جست و جو در پایگاه String با حداقل امتیاز ۰/۴ و حداکثر ده اینتراکشن. (ب) ده پروتئین برتر شناسایی شده در شبکه اولیه وسط الگوریتم MCC (پ) ده پروتئین موثرتر شناسایی شده در شبکه اولیه وسط الگوریتم Degree.

منابع

Bhardwaj, R., Sharma, I., Kanwar, M., Sharma, R., Handa, N., Kaur, H., Kapoor, D., & Poonam. (2013). LEA Proteins in Salt Stress Tolerance. In P. Ahmad, M. M. Azooz, & M. N. V. Prasad (Eds.), *Salt Stress in Plants: Signalling, Omics and Adaptations* (pp. 79–112). Springer. https://doi.org/10.1007/978-1-4614-6108-1_5

Karimi-Fard, A., Saidi, A., Tohidfar, M., & Saxena, A. (2023). Identification of key responsive genes to some abiotic stresses in *Arabidopsis thaliana* at the seedling stage based on coupling computational biology methods and machine learning. *Journal of Applied Biotechnology Reports*, 10(3), 1079–1090.

Kong, W., Zhong, H., Gong, Z., Fang, X., Sun, T., Deng, X., & Li, Y. (2019). Meta-analysis of salt stress transcriptome responses in different rice genotypes at the seedling stage. *Plants*, 8(3), 64.

Singh, A. K., Pal, P., Sahoo, U. K., Sharma, L., Pandey, B., Prakash, A., Sarangi, P. K., Prus, P., Paşcalău, R., & Imbrea, F. (2024). Enhancing Crop Resilience: The Role of Plant Genetics, Transcription Factors, and Next-Generation Sequencing in Addressing Salt Stress. *International Journal of Molecular Sciences*, 25(23), 12537. <https://doi.org/10.3390/ijms252312537>

Xu, F., Peng, Y., He, Z.-Q., & Yu, L.-L. (2023). The role of cyanoalanine synthase and alternative oxidase in promoting salt stress tolerance in *Arabidopsis thaliana*. *BMC Plant Biology*, 23(1), 163. <https://doi.org/10.1186/s12870-023-04167-1>

چکیده

در این مطالعه مجموعه داده‌های ترنسکریپتومی ۵۷ نمونه (۲۳ تنش و ۳۴ کنترل)، پس از نرمال‌سازی quantile، تبدیل log و حذف اثر batch با الگوریتم Combat در محیط R با هدف متاآنالیز داده‌های میکروآرای و شناسایی ژن‌های کلیدی دخیل در پاسخ آرابیدوپسیس به تنش شوری ادغام شدند. ژن‌های با بیان افتراقی با معیار $p < 0.01$ / شناسایی و ژن‌ها با تغییر بیان $|\log_2FC| > 2/5$ به عنوان ژن‌های محوری در پاسخ به شوری معرفی شدند که در میان آن‌ها، اعضای متعدد خانواده LEA برجسته بودند. همچنین، آنالیز نشان داد این ژن‌ها عمدتاً در فرآیندهای پاسخ به کم‌آبی، آبسزیک اسید، تنش شوری، سرما و نمو بذر نقش دارند و در اجزای سلولی مانند ناحیه خارج سلولی، سیتوسل، شبکه آندوپلاسمی و غشاء کلروپلاست قرار دارند. در حالی که از نظر عملکرد مولکولی عمدتاً در اتصالات لیپیدی مشارکت می‌کنند. تحلیل KEGG نیز غنی‌شدن مسیر متابولیسم سیانوآمینواسید را نشان داد که نشان‌دهنده درگیری مسیرهای سمیت‌زدایی مرتبط با بیوسنتز اتیلن در پاسخ به تنش شوری است. شبکه برهم‌کنش پروتئین-پروتئین نیز بر نقش محوری پروتئین‌های خانواده LEA در سازمان‌دهی شبکه پاسخ به تنش شوری در آرابیدوپسیس تأکید کرد.

مقدمه

با پیش‌بینی افزایش جمعیت جهان به ده میلیارد نفر تا سال ۲۰۵۰، تأمین غذای کافی به یکی از چالش‌های اصلی تبدیل شده و لذا، تولیدات کشاورزی و دامپروری باید شصت درصد افزایش یابد. تنش‌های غیرزیستی نظیر خشکی و شوری می‌توانند عملکرد نهایی را سالانه تا هفتاد درصد کاهش دهند. تنش شوری با ایجاد سمیت یونی، برهم‌زدن یکپارچگی غشای سلولی، مهار سنتز پروتئین و کاهش کارایی فتوسنتز، رشد گیاه را محدود می‌کند. مجموعه‌ای از ژن‌ها شامل فسفاتازها، کینازها، هورمون‌ها و فاکتورهای رونویسی در قالب شبکه‌های تنظیمی، پاسخ گیاه به شوری را کنترل می‌کنند. از سوی دیگر، مدیریت نادرست منابع آب، زهکشی ناکافی و آبیاری بیش از حد، منجر به تجمع نمک در خاک و شور شدن اراضی کشاورزی می‌شود (Singh et al., 2024). در چنین بستری، به‌کارگیری متاآنالیز داده‌های میکروآرای به عنوان رویکردی کارآمد برای ادغام مجموعه داده‌های ترنسکریپتومی و شناسایی ژن‌های کلیدی، مسیرهای متابولیکی مهم و ژن‌های با بیان افتراقی (DEGs)، ابزاری قدرتمند برای درک جامع‌تر پاسخ گیاه به تنش و بهره‌گیری از آن‌ها در برنامه‌های اصلاحی و به‌نژادی فراهم کرده است (Kong et al., 2019). در نتیجه، متاآنالیز داده‌های میکروآرای مربوط به تنش شوری در آرابیدوپسیس در شناسایی ژن‌های کلیدی دخیل در پاسخ به این تنش موثر می‌باشد و در این مطالعه از این رویکرد استفاده شد.

مواد و روش‌ها

❖ جهت دریافت پروفایل داده‌های بیانی گیاه *Arabidopsis thaliana* تحت تنش شوری، از پایگاه داده GEO استفاده گردید. این مجموعه حاصل هشت مطالعه و شامل بیست و سه نمونه تحت تنش شوری و سی و چهار نمونه کنترل (مجموعاً پنجاه و هفت نمونه) بود. نرمال‌سازی داده‌های بیانی خام توسط الگوریتم quantile انجام و سپس log₂ روی آن‌ها اعمال شد. برای ساخت ماتریس بیانی، تمامی پروفایل‌ها بر اساس Gene Symbol ادغام شدند. همچنین، batch effect توسط ابزار Combat پیکج SVA حذف گردید. تمامی پیکج‌ها و ابزارها مربوط به زبان R هستند. مقادیر بیانی مربوط به هر ژن خاص، به صورت میانگین قرار گرفتند. پس از آنالیز داده‌های نرمال شده، جهت شناسایی ژن‌ها با احتمال خطا $p < 0.01$ / (اطمینان ۹۹٪) از متد eBayes پیکج SVA با معیار $|\log_2FC| > 2/5$ استفاده شد. در ادامه، از میان ژن‌های شناسایی شده، ژن‌های با $|\log_2FC| > 2/5$ (تغییرات فولدچنگ) به عنوان DEG انتخاب شدند. این دو معیار جهت شناسایی ژن‌های با تغییرات بسیار ($|\log_2FC| > 2/5$) و با دقت بسیار بالا ($p < 0.01$) استفاده شدند (Karimi-Fard et al., 2023).

❖ به منظور بررسی عملکرد مولکولی، جایگاه سلولی و فرآیند بیولوژیکی DEGs شناسایی شده، آنالیزهای GO و KEGG توسط ابزار DAVID انجام شد.

❖ به منظور بررسی ارتباطات پروتئین-پروتئین، DEGs به عنوان ورودی در ابزار String استفاده شدند. جهت ساخت شبکه ارتباطات پروتئینی، برهمکنش‌های با امتیاز بیشتر از ۰/۴ و نرم‌افزار Cytoscape استفاده شد. به منظور رتبه‌بندی ژن‌ها در شبکه از پلاگین نرم‌افزار Cytoscape شامل cytohubba (الگوریتم‌های Degree و MCC) استفاده شد.

نتایج و بحث

❖ ابتدا بر روی تمامی داده‌ها کنترل کیفیت، نرمال‌سازی و در ادامه حذف اثرات batch با استفاده از الگوریتم Combat پیکج SVA انجام گرفت و سپس برای تایید صحت کار، از باکس پلات داده‌ها قبل و بعد از عملیات ذکر شده استفاده شد (شکل ۱). قرارگیری دامنه میان چارکی و خط میانه تمامی نمونه‌ها در یک راستا تایید اجرای صحیح عملیات کنترل کیفیت و نرمال‌سازی است.

❖ ۴۲ ژن افزایش بیان و ۸ ژن کاهش بیان نشان دادند ($p < 0.01$ / $|\log_2FC| > 2/5$) (جدول ۱). ژن *NM_202280* (*AT1G52690*)، عضو خانواده LEA (موثر در پاسخ به تنش شوری)، بیشترین افزایش بیان ($\log_2FC = 5.28$) و ژن *NM_112540* (*AT3G16670*)، عضو خانواده *Pollen Ole e 1 allergen and extensin* (یک سنسور تنش اکسیداتیو)، بیشترین کاهش بیان ($\log_2FC = -3.86$) را نشان دادند.

❖ نتایج آنالیز GO نشان داد که DEGs به طور معناداری در فرآیندهای بیولوژیکی پاسخ به عدم دسترسی به آب، پاسخ به آبسزیک اسید، پاسخ به تنش شوری، پاسخ به سرما و نمو بذر فعالیت می‌کنند. همچنین، این ژن‌ها در جایگاه‌های سلولی شامل ناحیه خارج سلولی، سیتوسل، شبکه آندوپلاسمی و غشاء کلروپلاست قرار دارند. به علاوه، عملکرد مولکولی آن‌ها، اتصال به لیپید را نشان داد. در نهایت، آنالیز KEGG متابولیسم سیانوآمینواسید را نشان داد. براساس تحقیقات قبلی، گیاهان تحت تنش (به خصوص تنش شوری)، به منظور حذف اثرات سوء سیانید تولید شده در مسیر بیوسنتز اتیلن، از مسیر بیوسنتز سیانوآمینواسیدها استفاده می‌کنند (Xu et al., 2023).

❖ به منظور بررسی ارتباطات پروتئین-پروتئین از شماره دسترسی DEGs شناسایی شده در مرحله قبل استفاده شد. بدین منظور ۵۰ شماره دسترسی در پایگاه String جست‌وجو و شبکه با حداقل امتیاز ۰/۴ و حداکثر ۱۰ اینتراکشن تولید و در نرم‌افزار Cytoscape رسم شد (شکل ۲). در ادامه به کمک روش‌های MCC و Degree افزونه Cytohubba، ۱۰ پروتئین برتر شبکه شناسایی شدند. الگوریتم MCC پروتئین‌های *LEA6*، *LEA18*، *AT3G15670*، *AT2G36640*، *AT5G06760*، *AP22.17*، *LEA7*، *AT3G17520*، *AT3G15670*، *AT5G06760*، *AT3G02480* و الگوریتم Degree پروتئین‌های *LEA7*، *AT2G42560*، *AT3G17520*، *LEA6*، *LEA18*، *LEA7*، *AT2G36640*، *AT3G17520*، *LEA6*، *LEA18*، *LEA7*، *AT2G42560* و *AP22.17* را به ترتیب (از رتبه ۱ تا ۱۰) به عنوان پروتئین‌های برتر شناسایی کردند. بیشتر پروتئین‌های فوق به جز *AT3G02480* (embryonic cell protein 63)،