

RESEARCH

Open Access



# GEDG: gene expression database of grape

Wen Sun<sup>1†</sup>, Siyu Chen<sup>2†</sup>, Lina Tan<sup>1</sup>, Yanglin Liu<sup>1</sup>, Xinyuan Zhang<sup>1</sup>, Yan Xiang<sup>1</sup>, Han Zheng<sup>2</sup> and Hanwei Yan<sup>1\*</sup>

## Abstract

**Background** Grapes are one of the oldest cultivated fruit species in the world, known for their high nutritional value. Grapes and grape products, including wine, are significant to many economies worldwide. Both abiotic and biotic stresses significantly impact the growth and reproduction of grapevines. RNA-Seq is a key tool for studying gene expression patterns under various conditions, such as stress responses, growth and developmental stages. In order to optimise the utilisation of these data for comprehensive research and analysis, and to enhance accessibility for users, we have developed a comprehensive database that provides gene expression profiles and traits in grapes.

**Description** We have developed the gene expression database for grapes (GEDG), an online resource that encompasses data on 9 types of abiotic (including heat, cold, hormones, light, climate, nitrogen, frozen) and biotic (disease and pests) stresses and different developmental stages. A total of 32 datasets were related to stress, 20 to growth and development, and one to chromosomal ploidy. The primary objective of GEDG is to expand and utilize transcriptome data to better understand stress responses in grapes. GEDG is the comprehensive database to integrate transcriptome and traits on grape.

**Conclusion** The Gene Expression Database of Grape (GEDG) was created to serve as a free resource platform for researchers, offering a comprehensive collection of genetic expression profiles for grape varieties based on RNA-seq data. The database is available at <https://www.gedg-vitis.cn/index.html>.

**Keywords** Grape, Stress, GEDG, Database, RNA-seq

## Background

The constant exposure of plants to biotic and abiotic stresses can greatly reduce the productivity of crops worldwide [1]. Major abiotic stresses include drought, salinity, heat, cold, freezing, high light intensity and ozone (O<sub>3</sub>) [2]. Additionally, plant face biotic stresses caused by pathogens (including bacteria, fungi, viruses, and nematodes) and herbivorous pests [3]. Numerous

databases related to plant stress have been developed, such as PlaASDB [4], PlantPReS: A database for plant proteome response to stress [5] and Stress combinations and their interactions in plants database (a one-stop resource on combined stress responses in plants) [6]. These are all valuable resources for researchers. With the growing number of omics datasets, particularly the continually updated grape genome, there is a pressing need to create a comprehensive database that can provide gene expression profiles and traits for grapes. This would enable the effective utilisation of these datasets in scientific research.

We propose the establishment of the Gene Expression Database of Grape (GEDG), integrating the publicly available RNA-seq data and collecting the characteristics of different grape varieties. GEDG aims to enhance the

<sup>†</sup>Wen Sun and Siyu Chen contributed equally to this work.

\*Correspondence:

Hanwei Yan

hwyannahau@163.com

<sup>1</sup>Laboratory of Modern Biotechnology, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China

<sup>2</sup>National Engineering Laboratory of Crop Stress Resistance Breeding, School of Life Sciences, Anhui Agricultural University, Hefei 230036, China



accessibility and usability of grape gene expression data, thereby enabling researchers to analyze and select specific gene features efficiently. GEDG not only integrates datasets pertaining to a multitude of stressors, including heat, cold, hormones, light, diseases, and pests, some datasets on growth and development were also collected. GEDG encompasses gene expression-based datasets that facilitate the linkage between genotypes and phenotypes through transcript profiling. Furthermore, each gene within GEDG is functionally annotated to support extensive comparative functional studies, providing a valuable resource for the grape research community. Concurrently, a number of websites pertaining to the grape database were consulted during the construction of this database. Our platform is designed as a free, open resource that requires no registration or login. Users can easily browse the website to locate necessary resources, which are available for free download.

### Construction and content

All data in the GEDG have been downloaded from the Gene Expression Omnibus (GEO) Database (<https://www.ncbi.nlm.nih.gov/>). To illustrate grape gene expression patterns, we obtained 53 high-quality RNA-seq datasets from GEO. The data were then systematically categorized and analyzed. The raw data were classified into different stress types: abiotic (including heat, cold, hormones, light, climate, nitrogen, frozen) and biotic (disease and pests). Additional classification included growth and development, as well as chromosome ploidy level.

First of all, we obtained basic information about the grape genes from Phytozome [7] and Grapedia [8] including coding sequences (CDS), gene sequences, protein sequences, chromosome positions, and protein lengths. Additionally, we incorporated three widely used grape genome versions (12X.v2, v4, and the latest T2T.v5) to provide comprehensive and up-to-date genomic information. Cross-referencing of gene IDs across these genome versions has been established to ensure seamless comparison. Further annotations, such as Pfam [9] and gene ontology (GO) [10], were acquired from Biomart [11] (<https://phytozome.jgi.doe.gov/pz/portal.html>). The molecular weight (MW) and isoelectric point (pI) of the protein sequence were calculated using the pI/MW tool in TBtools-II [12]. We also downloaded grape paralogs annotated in Ensembl via Biomart and orthologs across multiple species (*Arabidopsis thaliana*, *Glycine max*, *Nicotiana attenuata*, *Oryza sativa Japonica Group*, *Populus trichocarpa* and *Solanum tuberosum*).

The primary programming language employed for GEDG is Java, with the Java Development Kit (JDK) version 1.8.0\_412. The database is built on a relational database architecture, and the Basic Local Alignment Search Tool (BLAST) is employed for sequence alignment. To

ensure comparability between samples, expression data were normalized using FPKM values. A DESeq\_edgeR was used for DEG analysis. ClusterProfiler (version 3.18.1) was used for GO and KEGG enrichment analysis.

### Utility and discussion

We collected 32 grape RNA-Seq datasets, comprising 16 samples subjected to abiotic stress and 16 to biotic stress. Additionally, we gathered 20 datasets focused on growth and development and 1 dataset on chromosome ploidy. These RNA-seq data were uniformly processed and integrated into the GEDG. GEDG not only provides a comprehensive collection of gene expression data but also delivers detailed information on individual genes, facilitating in-depth analysis and research.

### Interface

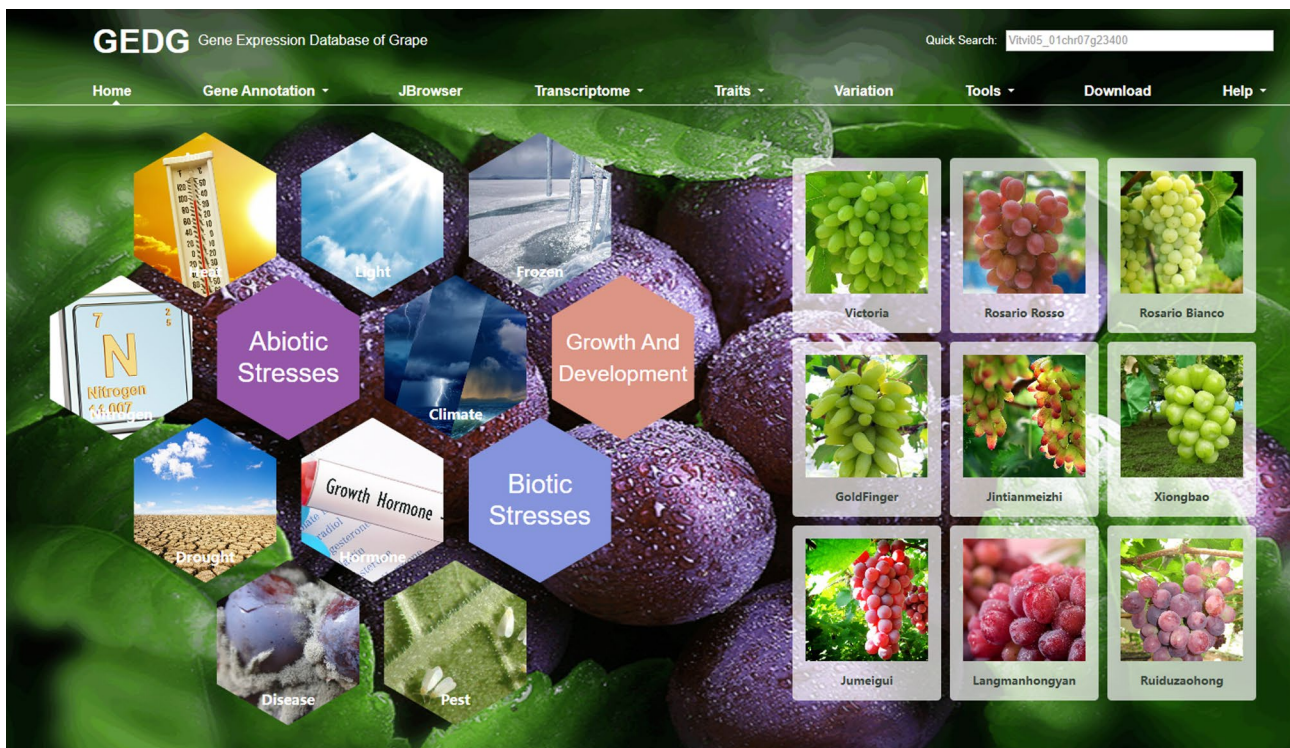
GEDG provides an intuitive, user-friendly and personalized interactive web interface that simplifies searching, browsing, visualization, and data downloading. What's more, GEDG contains the following modules: Home, Gene Annotation, Gene Browse, Transcriptome, Traits, Variation, Tools, Download and Help.

### Home

The home page of the GEDG provides a concise overview of the site. On the left side of the interface, the transcriptional data are divided into three categories: abiotic stress, biotic stress, growth and development. The abiotic stress encompasses factors such as heat, cold, hormones, light, climate, nitrogen and frozen, while the biotic stress category includes diseases and pests. Users can select the modules of interest and subsequently navigate to the corresponding browsing pages. On the right-hand side of the interface, images of various grape varieties are displayed, with a clickable link to provide data on their traits (Fig. 1).

### Gene Annotation

The Gene Annotation comprises two functions: the Search Function and the Reviewed UniprotKB genes. The Search Function allows users to explore gene functions through various search methods, including Gene ID Search, GO Slim Search, Domain Search and KEGG Search. To search for a gene's function, enter the Gene ID, such as *Vitvi05\_01chr07g23400*, into the search box of the locus search module. This query will display the gene's basic functional characteristics and the function of the protein (Fig. 2A). Additionally, there are homologs in other plant species, as well as expression patterns under diverse biotic and abiotic stresses (Fig. 2B). Most results are clickable, allowing for graphical illustration on the genome browser implemented using JBrowse [13]. Users can also search for genes in GEDG using four functional



**Fig. 1** The homepage of GEDG. On the left-hand side of the page, there is a breakdown of stress and growth data, while on the right-hand side, there is a picture of the grape variety

categories: PFAM, InterPro, GO, and KEGG. The search results show a set of genes annotated with functional descriptions (Fig. 2C). Similarly, genetic identifiers are clickable and can be linked to JBrowse for visualisation. Furthermore, users can click on the search number to access external website for more detailed annotations. By combining all of these search methods, we have created a versatile and highly filterable advanced search tool. Another function, Reviewed UniprotKB genes, displays annotations of the protein functions of grapes that we have screened on the Uniprot website [14].

In summary, the search engine offers users a straightforward, adaptable, and accessible tool for systematically retrieving the comprehensive grape data in GEDG. This has significant implications for future grape functional genomics and genetic improvement initiatives.

### JBrowse

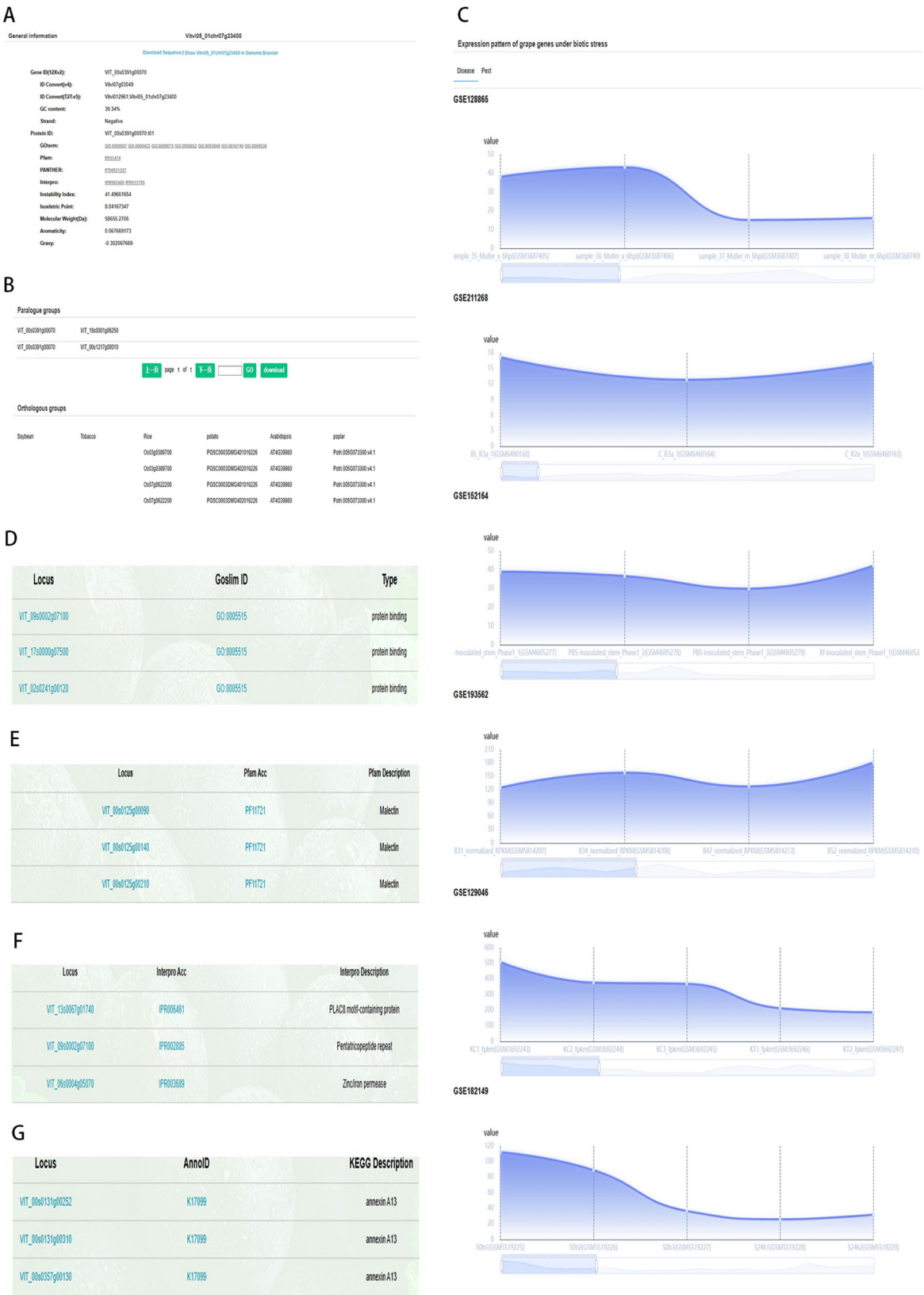
This module functions as a genome browser, enabling users to retrieve and visualize gene's feature. Users can select specific genes to search within the browser or upload their own files for visualization. Additionally, from the Gene ID Search results page, users can click Show ID in Genome Browser to directly access this module. JBrowse provides a navigable view of the grape genome, allowing users to observe gene positions and configurations in detail. By adjusting the view settings,

users can zoom in or out on genomic regions to gain a more detailed understanding of specific gene characteristics. Multiple genes can be viewed simultaneously by adding them to the browser, and additional views, such as GC content, Circular, and Dotplot views, are available for enhanced exploration.

JBrowse also supports various analytical operations, including identifying variant sites and assessing gene expression levels, helping researchers gain deeper insights into genetic variation, gene regulation, and the evolutionary history of the grape genome. Future updates will expand these features, integrating additional genome and annotation data to enhance functionality and user experience.

### Transcriptome

A comprehensive review of the published literature on grape stress was conducted through GEO Datasets, with the relevant transcriptome data subsequently downloaded. A total of 32 transcriptome datasets for stress and 21 for grape growth and development were collected. By visualizing and integrating the data into GEDG, we have created a platform that allows users to easily access the data. Upon selecting the Browse module, a drop-down menu will appear, displaying an Overview, Function Annotation and 9 different stress types, which include disease, pest, drought, hormone, frozen, heat, nitrogen,



**Fig. 2** Search engine of GEDG. Screenshot of the results from (A, B and C) Gene ID Search (e.g., *Vitv05\_01chr07g23400*); (D) GO Slim search (e.g., GO:0005515); (E and F) Domain search (e.g., PF11721, IPR006461); (G) KEGG Search (e.g., K17099)

light, climate. Additional categories such as growth and development and chromosome ploidy are also available. Users can obtain a comprehensive overview of the annotations for each transcriptome dataset through the Overview section. The table and graphs offer the details including the GEO accession number, cultivar, sample number, tissue type, Contributor, a description of the motivations and characteristics (Fig. 3A). These datasets were carefully curated based on their high quality and coverage of grape stress responses and developmental stages, ensuring diversity and usability for grapevine research. By clicking on each GEO accession, users can access the results of our data visualization (Fig. 3B). Furthermore, when the user clicks on Functional Annotation, they can obtain functional details of transcripts assigned by ID Convert, PFAM, PANTHER, GO, KEGG, KOG, Best-hit-arabi-name and arabi-symbol (Fig. 4A). Upon selecting the type of stress of interest, the user can access detailed information and visual data. Each stress management page contains a brief description and images made using the downloaded data (Fig. 3B). To identify differentially expressed genes, we collected all samples from a specific stress condition or developmental stage. Data were first filtered based on quality parameters to exclude low-confidence data, ensuring a reliable dataset for analysis. For DEG identification, we applied the parameter  $|\text{Log}_2\text{FC}| \geq 2$  to compare gene expression between different experimental groups, such as variations in environmental or stress conditions, and across developmental stages. The Venn diagram demonstrates differential expression analysis, which utilized data from multiple datasets, successfully identifying genes with significantly increased or decreased expression. (Fig. 4B). The page also displays the genes that have been identified as differentially expressed, along with the corresponding values, at the bottom of the page.

In conclusion, we have collated datasets that provide users with information in tabular form and also allow the data to be visualized in graphical form. Furthermore, the data has been categorized in a manner that facilitates convenient browsing and consultation.

### Traits and Variation

We collected data on the traits of different grape varieties for the years 2019, 2020 and 2021. In 2019, traits of 44 grape varieties were shown, the highest number of varieties from 2019 to 2021. The data collected included information on various traits, such as cluster weight, cluster length, and cluster width. This information is presented in tabular form, allowing users to click on the variety in the drop-down menu bar to view the desired data (Fig. 5A). Furthermore, an illustration of the grapes is presented at the base of the data table. Users then select

the desired organization of the grapes, which is then accompanied by the corresponding data (Fig. 5B).

Additionally, annotations on variation are provided, with the data presented in a tabular format (Fig. 5C). Users can retrieve the information they need by typing Variation Name into the search box. The information was obtained from Ensembl, which included Variant name, Gene stable ID, Variant consequence, Genotype frequency, etc.

### Tools

We have developed a suite of analytical tools to enable users to fully explore and analyze grape transcriptome datasets.

**Blast Search.** This interface allows users to input a FASTA query sequence and perform a Blast search for nucleotide or protein sequences in the GEDG. The grape genome can be searched using the BLAST suite of algorithms, which includes blastn, blastp, blastx, tblastn, and tblastx. A threshold of E-values of  $1 \times 10^{-5}$  is used to identify potential homologous sequences.

**Gene ID convert.** Users can use this function to convert three different versions of the gene ID (Fig. 5D). We obtained three versions of the grape genome data and established a linkage between the genetic IDs of these versions based on the downloaded data, enabling ID conversion across different versions. Furthermore, within GEDG, utilizing these linked IDs in Search Functions and Functional Annotation functions yielded corresponding results.

**Enrichment analysis.** This interface enables users to perform gene function enrichment analysis based on two functional catalogs: the Gene Ontology (GO) term and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. The results return a significantly richer set of GO or KEGG function categories that are further cross-linked to specific public databases (Fig. 5E).

**Orthologous groups.** This function allows users to search for homologous groups between grape and six representative plant species (Arabidopsis, Soybean, Tobacco, Rice, Poplar, Potato) using genetic identifiers (Fig. 5F).

### Download

In the Downloads module, users can download the available resources.

### Help

Upon hovering the cursor over the Help module, a drop-down menu will appear, including Manual, About us and Useful Links. The Manual page guides users through the steps and operations of GEDG. Useful link provides access to the data source and software used to construct GEDG.

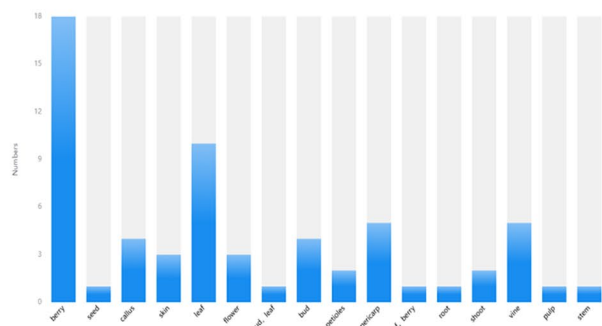
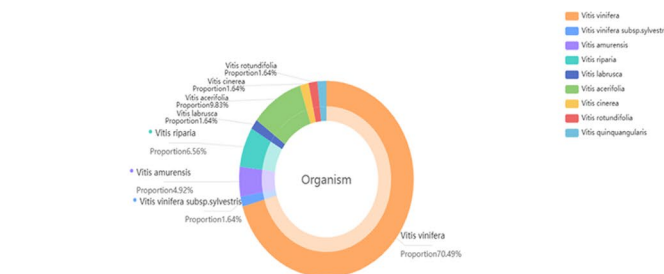
A

Overview Functional annotation Disease Pest Drought Frozen Heat Hormone Nitrogen Light Climate Growth and development Chromosome Ploidy

Overview of the currently released *Vitis vinifera* transcriptomes.

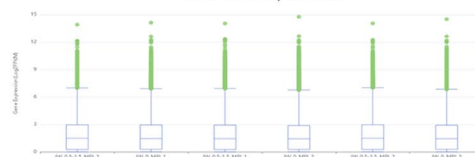
GeoAccession	SampleNumber	Cultivar	Year	Tissues	Contributor	Motivations And Characteristics
GSE87643	99	Cabernet Sauvignon	2018	petioles	Rapicavoli JN, Blanco-Ulate B, Figueroa-Balderas R, Morales-Cruz A, Acosta P, Lukaszynski A, Cantu D, Roper MC	high throughput sequencing; plant defense against invading organisms
GSE128865	10	Mueller Regent	2019	leaf	Boggs J, Eisenmann B, Czernom S	high throughput sequencing; Breeding of resistant V. vinifera cultivars
GSE128966	24		2023	vine	Swaminathan P, Fennell AY	high throughput sequencing; Photoperiod subjected to long day and short day
GSE202924	72	Cabernet Sauvignon	2023	petioles	Castro C, Macconnet M, Her N, diSalvo B, Jaboniska B, Jeske DR, Cantu D, Roper MC	high throughput sequencing; Grape genes in bioprospecting-primed petioles during Pierce's Disease
GSE58653	12	Cabernet Sauvignon	2015	leaf	Czernom S, Galametz ER, Trivison R, McDone AJ, Cramer GR, Baumgartner K	high throughput sequencing; Development of an early detection tool; stems were examined to monitor spread of the infection
GSE175475	29	Pinot noir UCDS	2022	callus	Ting-Hsuan C, Christopher W	high throughput sequencing; Exposed the callus to live cultures of an endemic grapevine yeast Hanseniaspora uvarum
GSE87932	16	Sémillon	2015	berry	Blanco-Ulate B, Amrine KC, Collins TS, Rivers RM, Vicente AR, Morales-Cruz A, Doyle CL, Ye Z, Allen G, Heymann H, Eder SE, Cantu D	high throughput sequencing; In case of a noble rot, to the synthesis of important metabolites for grape berry flavor and aroma
GSE118274	9	Vitis vinifera cv. Garganega	2019	berry	Lozano A, Colombo T, Tomiello GB, Polverani A, Vandelio E, Zenoni S	high throughput sequencing; Identification of the genes deployed by B. cinerea during grape berries colonization in the late rot form
GSE76206	24	Cabernet Sauvignon, Duriel eng-2	2015	skin	Wang Y, Li R	high throughput sequencing; Different development stages of Vitis vinifera cv.
GSE95812	52	Zinfandel	2019	berry	Blanco-Ulate B, Figueroa-Balderas R, Cantu D	high throughput sequencing; Red blotch disease causes major reprogramming of grape berry metabolism and development leading to the inhibition of ripening pathways and stress responses

page 1 of 6 GO

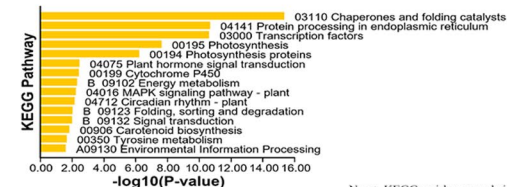
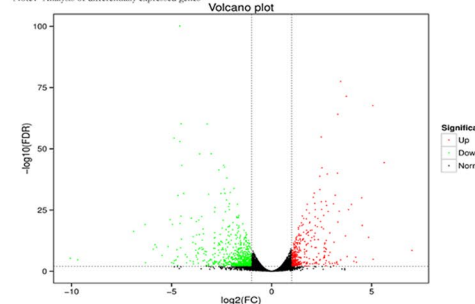


B

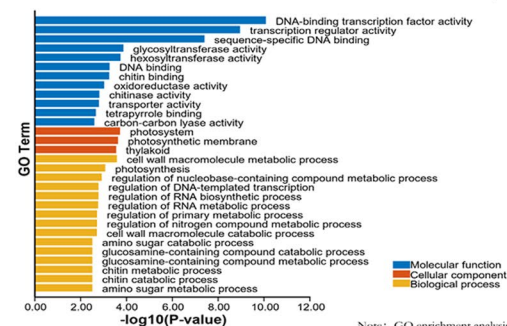
### Distribution of Gene Expression Values



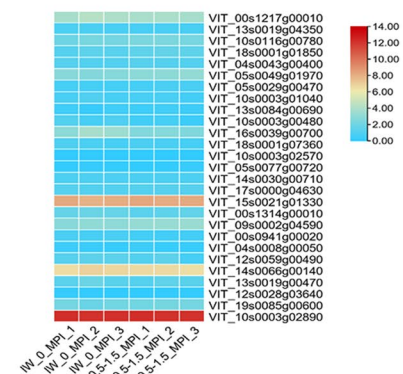
**Note:** Analysis of differentially expressed genes



Note: KEGG enrichment analysis



Note: GO enrichment analysis



Note: Heatmap

**Fig. 3** Overview of transcriptome data and visualization of the data. **(A)** A detailed overview of each dataset and visualization of species proportion and number of tissues; **(B)** The analysis of Boxplots, Heatmaps, and differential gene analysis using transcriptome data, as well as KEGG enrichment and GO enrichment using differential genes, were conducted

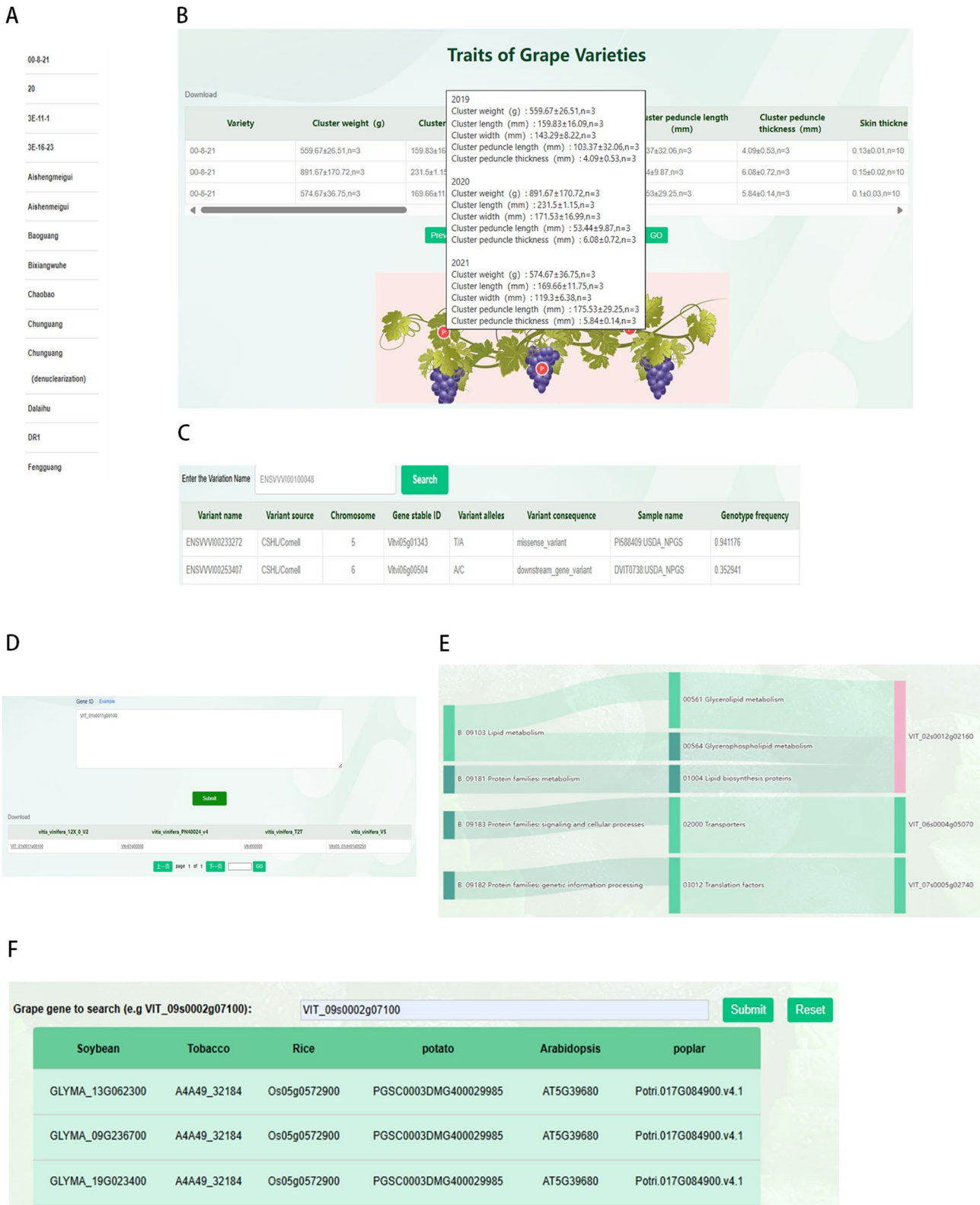
**A**



B



**Fig. 4** Functional annotation and visualization of data on the dataset classification page of GEDG. **(A)** An overview of the Functional annotation module; **(B)** The charts present the gene numbers for each dataset, as well as Venn plots, GO enrichment analysis, KEGG enrichment, and charts of the number of up-regulated and down-regulated genes derived from the differentially expressed genes. The list of differentially expressed genes and the value of  $|\log_2FC| \geq 2$  are presented at the bottom



**Fig. 5** Traits, Variations module and flexible tools implemented in GEDG. **(A)** A display of grape varieties in the drop-down menu of the Traits module; **(B)** Visualization of grape variety trait information and grape representative tissue trait information. (e.g. 00-8-21); **(C)** Information about the Variations module; **(D)** A tool for gene ID conversion among 4 *Vitis* genomes; **(E)** Gene set functional enrichment analysis using KEGG pathway; **(F)** Orthologous groups between grape and six other representative plant species

## Conclusions

We have constructed a comprehensive expressional patterns database of grapes. The database offers a plethora of versatile analytical tools for users to employ in their analyses. GEDG represents the inaugural database of grape gene expression profiles under conditions of stress and different growth and developmental stages. As the number of generated sequences continues to grow, and a multitude of resources will be made available. It is therefore imperative that GEDG prioritise the integration of supplementary transcriptome data obtained under stress conditions, in addition to other pertinent grape-related data. Concurrently, we will enhance the functionality of the system to facilitate more comprehensive queries and analytical processes pertaining to grapes. It is our hope that the implementation of timely and continuous maintenance will facilitate greater awareness of GEDG and encourage contributions to the ongoing development of horticultural grapes.

It is anticipated that GEDG will prove to be a valuable resource for the research community. Specifically, in the context of stress, this approach can be of significant benefit to researchers. A plethora of features and tools have been constructed with the objective of enhancing the user-friendliness of the site. We anticipate receiving a substantial quantity of constructive feedback, which will assist us in further enhancing the functionality of GEDG in the future.

## Abbreviations

GEDG	Gene Expression Database of Grape
GEO	Gene Expression Omnibus
GO	Gene ontology
CDS	Coding sequence
NCBI	National Center for Biotechnology Information

## Author contributions

WS and SYC collected, organized, and analyzed the data, and wrote the paper. LNT, XYZ and YLL provide ideas for building a website, as well as help with data for graphing. YX, HZ and HWY conceived the project and supervised this study. All of the authors discussed the results and commented on the manuscript.

## Funding

The Science and Technology Project of Anhui Province, Grant/Award Number: 2022j11020010; University Outstanding Talents Program of Anhui Province, Grant/Award Number: gxyqZD2022020.

## Data availability

GEDG is accessible from <https://www.gedg-vitis.cn/index.html>. The Download module provides access to a range of data, including transcriptome data and other data.

## Declarations

### Ethics approval and consent to participate

All authors read and approved the manuscript.

### Consent for publication

Consent and approval for publication was obtained from all authors.

### Competing interests

The authors declare no competing interests.

Received: 31 July 2024 / Accepted: 5 May 2025

Published online: 13 May 2025

## References

1. Boyer JS. Plant Productivity and Environment. 1982; 218(4571):443–448.
2. Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. Abiotic and biotic stress combinations. *New Phytol.* 2014;203(1):32–43.
3. Atkinson NJ, Urwin PE. The interaction of plant biotic and abiotic stresses: from genes to the field. *J Exp Bot.* 2012;63(10):3523–43.
4. Guo XK, Wang TP, Jiang LY, Qi H, Zhang ZD. PlaASDB: a comprehensive database of plant alternative splicing events in response to stress. *BMC Plant Biol* 2023; 23(1):225.
5. Mousavi SA, Pouya FM, Ghaffari MR, Mirzaei M, Ghaffari A, Alikhani M, Ghareyazie M, Komatsu S, Haynes PA, Salekdeh GH. PlantPreS: A database for plant proteome response to stress. *J Proteom.* 2016;143:69–72.
6. Priya P, Patil M, Pandey P, Singh A, Babu VS, Senthil-Kumar M. Stress combinations and their interactions in plants database: a one-stop resource on combined stress responses in plants. *Plant J.* 2023;116(4):1097–1117.
7. Rhee SY, Beavis W, Berardini TZ, Chen GH, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al. The Arabidopsis information resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 2003;31(1):224–8.
8. Shi XY, Cao S, Wang X, Huang SY, Wang Y, Liu ZJ, Liu WW, Leng XP, Peng YL, Wang N et al. The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic Res.* 2023; 10(5):uhad061.
9. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, et al. The Pfam protein families database. *Nucleic Acids Res.* 2004;32:D138–41.
10. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004;32:D258–61.
11. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, Vilella AJ, Searle SMJ, Amodio R, Brent S et al. Ensembl comparative genomics resources (vol 2016, bav096, 2016). Database-the Journal of Biological Databases and Curation 2016.
12. Chen CJ, Wu Y, Li JW, Wang X, Zeng ZH, Xu J, Liu YL, Feng JT, Chen H, He YH, et al. TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. *Mol Plant.* 2023;16(11):1733–42.
13. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. JBrowse: A next-generation genome browser. *Genome Res.* 2009;19(9):1630–8.
14. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Cukura A, Denny P, et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2023;51(D1):D523–31.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.