Data in Brief 55 (2024) 110583

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of wheat HSP90.2 chaperome

Yue-Ting Guo[a], Yan Yan[a], Guo-Liang Zhang[a], Jin-Ying Gou[b,*]

[a] School of Life Sciences, Fudan University, Shanghai 200438, China
[b] Key Laboratory of Crop Heterosis and Utilization (MOE), Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

ARTICLE INFO

Article history:
Received 2 January 2024
Revised 19 May 2024
Accepted 29 May 2024
Available online 12 June 2024

Dataset link: Dataset of wheat HSP90.2 chaperome (Original data)

Keywords:
Wheat
Molecular chaperone
Protein interactome
Post-translational regulation

ABSTRACT

Wheat (*Triticum aestivum* L.) is one of the world's most important staple crops, whose production is critical to feed the expanding population worldwide. The 90-kDa Heat Shock Protein 90 (HSP90) is a highly abundant chaperone protein involved in multiple cellular processes. It facilitates the folding of nascent preproteins for their maturation and functioning. This data described HSP90.2 clients identified from the whole genome of wheat. The HSP90.2 chaperome contains over 1500 proteins, most detected by the C terminus and full-length of HSP90.2. Over 60 % of the clients reside in the cytosol, nucleus, and chloroplasts. Cytoskeleton-related proteins are enriched in the chaperome of the N terminus of HSP90.2. The clients of the middle part of HSP90.2 contains several factors involved in ethylene biosynthesis and extracellular vesicle or organelle-related activities. Some clients related to plant hypersensitive response are induced by stripe rust. The presented dataset could isolate proteins regulated by HSP90.2 at the post-translational level.

* Corresponding author.
E-mail address: jygou@cau.edu.cn (J.-Y. Gou).

https://doi.org/10.1016/j.dib.2024.110583

Specifications Table

Subject	Agricultural Science
Specific subject area	Agronomy and Crop Sciences
Type of data	Figure and Table, Analyzed
Data collection	The full-length, N terminus, middle part, and C terminus of *TaHSP90.2* were incorporated into the pLAW10 vector for yeast two-hybrid screening by LR clonase II. The aureobasidin (AbA) (CAT: 60231ES03, Yeasen, Shanghai, China) at 125 ng/mL inhibited the autoactivation of different TaHSP90.2 fragments. Potential clients of TaHSP90.2 fragments were isolated using a high coverage cDNA two-hybrid library prepared with mixed RNAs from different tissues of a hexaploid wheat Fielder (OE Biotechnology, Shanghai, China). The clones were selected on the synthetic dropout medium and used to extract the plasmids for the second-generation sequencing.
Data source location	School of Life Sciences, Fudan University, Shanghai, China
Data accessibility	Repository name: Mendeley data. Data identification number: doi: 10.17632/fmyrz84f65.3 Direct URL to data: https://data.mendeley.com/datasets/fmyrz84f65/3
Related research article	HSP90.2 modulates 2Q2-mediated wheat resistance against powdery mildew. Yan Y, Guo YT, Chang CY, Li XM, Zhang MQ, Ding CH, Cui D, Sun C, Ren Y, Wang ML, Xie C, Ni Z, Sun Q, Chen F, Gou JY. Plant Cell Environ. 2023 Jun;46(6):1935–1945. doi: 10.1111/pce.14579. PMID: 36,890,722.

1. Value of the Data

- This data described over 1500 clients of pathogen-inducible TaHSP90.2 in the whole genome of wheat.
- This dataset can facilitate the isolation of fungal-resistant proteins regulated by TaHSP90.2 in the folding process.
- This dataset provides valuable information for elucidating the regulation of proteins by TaHSP90.2 through different domains.

2. Background

Wheat provides about 20 % of protein and energy for humankind, and its production faces numerous environmental stresses. Heat Shock Proteins (HSPs) are the dominant molecular chaperones folding nascent immune receptors. HSP90.2 is found to respond to abiotic stresses and pathogens, representing a common strategy of post-translational regulation of stress-related proteins. We cloned the full-length, N-terminus, middle part, and C-terminus of HSP90.2 (HSP90.2-FL, HSP90.2-N, HSP90.2-M, HSP90.2-C, respectively, hereafter) and screened their interacting proteins by a yeast two-hybrid system [1]. This dataset supports our related publication by further analyzing the clients on their subcellular localizations, functional annotations, and expression patterns in response to stripe rust [2].

3. Data Description

To identify clients of HSP90.2, a hexaploid wheat cDNA library used for yeast two-hybrid assay was utilized directly for yeast two-hybrid screening. Colonies survived in the medium with 125 mg/L Aureobasidin A, and the respective inserts in the library vector were sequenced by the Illumina HighSeq platform [3]. This dataset provides the annotation and sequence IDs for each potential client, specifies the number of GO terms assigned to the clients, and gives further information about the predicted localization of the candidate clients. The raw data was uploaded to the NCBI GEO dataset under (PRJNA827186 ID: 827186) and Mendeley data (doi: 10.17632/fmyrz84f65.1) [3,4].

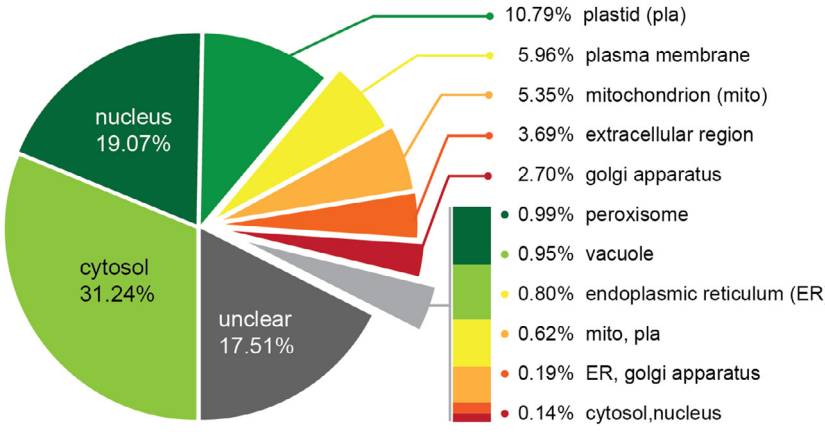

Fig. 1. The predicted subcellular localizations of the TaHSP90.2 chaperome.

The dataset has three levels: subcellular localization prediction, GO enrichment, and expression patterns in response to stripe rust. We predicted the protein structure of HSP90.2 by AlphaFold and identified 1565 unique potential clients using the different fragments of TaHSP90.2. Based on the subcellular-localization prediction, 660 (31.24 %) HSP90.2 clients reside in the cytosol. 403 (19.07 %) HSP90.2 clients have a nucleus localization trend, 25 % higher than the total wheat genome (20,907 out of 137,029, 15.26 %). The ratios of clients with mitochondrion (5.35 %), endoplasmic reticulum (0.8 %), and extracellular (3.69 %) are lower than the whole genome by 31.6 %, 32.4 %, and 45.5 %, respectively (Fig. 1).

According to the GO analysis, the clients of HSP90.2-FL share similar ratios of clients enriched in specific terms with that of HSP90.2-C. The HSP90.2-N has unique clients with annotations of cytoskeletons (microtube, supramolecular fiber, and polymeric cytoskeletal fiber) and proton-transporting ATP synthase complex, consistent with the ATP binding capacity of HSP90.2-N. The clients of HSP90.2-M are enriched in extracellular processes, including extracellular-vesicle, organelle, exosome, and space (Fig. 2).

Additionally, we extracted the RNA-seq data of HSP90.2 clients and analyzed their transcriptional responses to *P. striiformis* Westend. f. sp. *tritici* (*Pst*), the causing pathogen of stripe rust, one of the most devastating wheat fungal diseases. The HSP90.2 clients formed eight gene clusters based on their expression patterns. The pathogen induces the cluster 3 genes at the early stage (1 day post inoculation) with 38 (26.4 %) out of the 144 genes residing in the nucleus, which is higher than the 15.26 % at the genome level (Fig. 3).

4. Experimental Design, Materials and Methods

4.1. Yeast two-hybrid screening of HSP90.2 clients

The *HSP90.2-FL, HSP90.2-N, HSP90.2-M*, and *HSP90.2-C* (HSP90.2 fragments, next after) were incorporated into the pLAW10 vector for yeast two-hybrid screening by LR clonase II [1]. Aureobasidin (AbA) (CAT: 60231ES03, Yeasen, Shanghai, China) at 125 ng/mL inhibited the autoactivation of different *TaHSP90.2* fragments. Potential clients of HSP90.2 fragments were isolated from a high coverage cDNA two-hybrid library prepared with mixed RNAs from different tissues of a hexaploid wheat Fielder (OE Biotechnology, Shanghai, China) [2]. The plates were incubated at a 28 °C incubator to collect the clones to extract the plasmids for the second-generation sequencing.

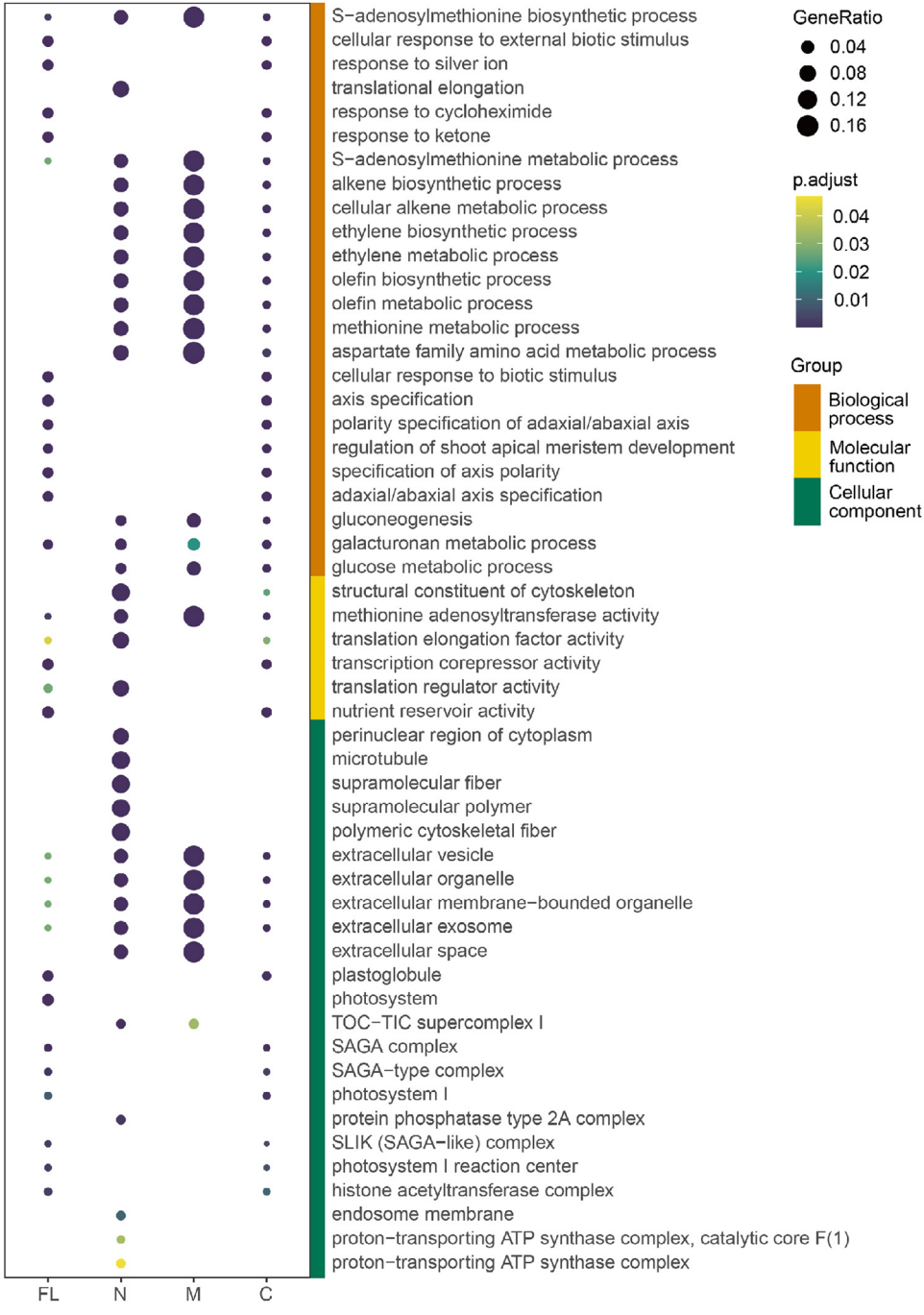

Fig. 2. The Gene Ontology (GO) annotations of the TaHSP90.2 chaperome.

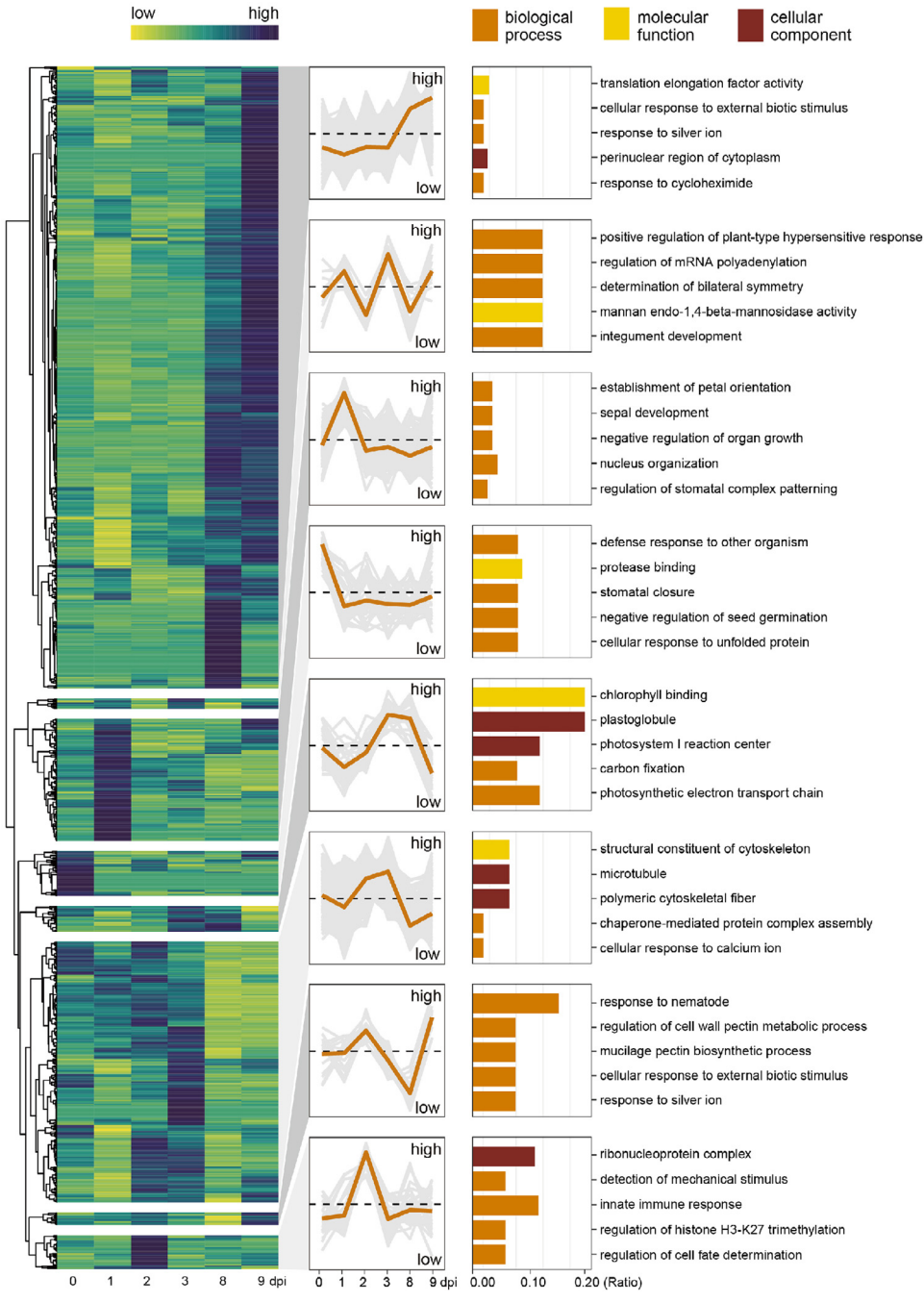

Fig. 3. The expression patterns of the TaHSP90.2 chaperome in response to *Pst*. Data was organized based on the NCBI BioProject accessions PRJNA243835 and PRJNA387101.

4.2. Functional annotation and subcellular localization analysis

R package 'clusterProfiler' v4.6.0 was used to carry out the enrichment analysis of Gene Ontology (GO) annotations for the gene sets [5]. Protein subcellular locations were obtained from the cropPAL2020 data set, including experimental and predicted localizations [6]. Proteins with a high proportion of unknown sequences were excluded.

4.3. Gene expression pattern analysis

Two sets of high-throughput transcriptome data were re-analyzed for gene expression patterns. The raw data of two wheat-*Pst* sample sets were obtained from the NCBI under BioProject accessions PRJNA243835 and PRJNA387101 [7,8]. After evaluating the quality of raw paired-end reads by FastQC v0.11.5, the adapter and low-quality sequences were removed with trimmomatic [9]. The clean reads were mapped to the wheat genome sequences (IWGSC RefSeq v1.1) using hisat2 v2.2.1 [10] with default options. The SAM formatted files were converted to BAM formatted files by samtools v1.14 [11]. After sorting and indexing by samtools, featureCounts v2.0.3 [12] was used to quantify the gene expression level. For each gene in the integrated transcriptome, the number of transcripts per million (TPM) was applied for normalization. R was used to analyze and visualize the genes showing similar expression patterns at 0, 1-, 2-, 3-, 8-, and 9-days post inoculation (DPI). R package 'pheatmap' was used for clustering (clustering_method='average', clustering_distance_rows='correlation'). R package 'reshape2′ and 'ggplot2′ were used for visualization.

Limitations

Not applicable.

Ethics Statement

This study does not involve experiments on humans or animals.

Data Availability

Dataset of wheat HSP90.2 chaperome (Original data) (Mendeley Data).

CRediT Author Statement

Yue-Ting Guo: Conceptualization, Methodology, Software; **Yan Yan:** Data curation; **Guo-Liang Zhang:** Data curation; **Jin-Ying Gou:** Writing – review & editing.

Acknowledgments

The Research described here was supported by the National Natural Science Foundation of China (32372557).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110583.

References

[1] Y. Yan, M.L. Wang, Y.T. Guo, et al., HSP90.2 promotes CO_2 assimilation rate, grain weight, and yield in wheat, Plant Biotechnol. J. (2023), doi:10.1111/pbi.14032.
[2] Y. Yan, Y.T. Guo, C.C. Chang, et al., HSP90.2 modulates 2Q2-mediated wheat resistance against powdery mildew, Plant Biotechnol. J. 46 (2023) (2023) 1935–1945 Plant Cell Environ.
[3] PRJNA827186, BioProject database, National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA827186.
[4] Y.T. Guo, Y. Yan, G.L. Zhang, et al., Dataset of wheat HSP90.2 chaperome, Mendeley Data (2023) V2, doi:10.17632/fmyrz84f65.3.
[5] T. Wu, E. Hu, S. Xu, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation (Camb) 1 (2021) 100141.
[6] C.M. Hooper, I.R. Castleden, et al., CropPAL for discovering divergence in protein subcellular location in crops to support strategies for molecular crop breeding, Plant J. 104 (2020) 812–827.
[7] H. Zhang, Y. Yang, C. Wang, et al., Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew, BMC Genomics 15 (2014) 898.
[8] F. Tao, J. Wang, et al., Transcriptomic analysis reveal the molecular mechanisms of wheat higher-temperature seedling-plant resistance *to Puccinia striiformis f. sp. tritici*, Front. Plant Sci. 9 (2018) 240.
[9] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[10] D. Kim, J.M. Paggi, C. Park, et al., Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915.
[11] P. Danecek, J.K. Bonfield, J. Liddle, et al., Twelve years of SAMtools and BCFtools, Gigascience 10 (2021) giab008.
[12] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014) 923–930.