



## OPEN ACCESS

## EDITED BY

Yogeshwar Vikram Dhar,  
Ruhr University Bochum, Germany

## REVIEWED BY

Sarita Srivastava,  
University of Allahabad, India  
Mengran Yang,  
University of Missouri, United States

## \*CORRESPONDENCE

Ning Li

✉ lining@xaas.ac.cn

Yong Qin

✉ xjndqinyong@126.com

<sup>†</sup>These authors share first authorship

RECEIVED 17 December 2024

ACCEPTED 05 June 2025

PUBLISHED 30 June 2025

## CITATION

Xu R, Wang Z, Chen Z, Wang Z, Meng Q, Li N  
and Qin Y (2025) The Evolution and  
expression analysis of USP gene family  
in *Solanum*.  
*Front. Plant Sci.* 16:1546640.  
doi: 10.3389/fpls.2025.1546640

## COPYRIGHT

© 2025 Xu, Wang, Chen, Wang, Meng, Li and  
Qin. This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums  
is permitted, provided the original author(s)  
and the copyright owner(s) are credited and  
that the original publication in this journal is  
cited, in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# The Evolution and expression analysis of USP gene family in *Solanum*

Ruiqiang Xu<sup>1,2†</sup>, Zhongyu Wang<sup>2†</sup>, Zhaolong Chen<sup>1,2</sup>,  
Zepeng Wang<sup>1,2</sup>, Qingyuan Meng<sup>1,2</sup>, Ning Li<sup>2\*</sup> and Yong Qin<sup>1\*</sup>

<sup>1</sup>College of Horticulture, Xinjiang Agricultural University, Urumqi, Xinjiang, China, <sup>2</sup>Biological Breeding Laboratory, Xinjiang Uygur Autonomous Region Academy of Agricultural Sciences, Urumqi, China

As natural environments degrade and extreme weather events become more frequent, humanity increasingly faces the challenge of producing crops under various complex and adverse conditions. Improving crop adaptability has become crucial. Universal stress proteins (USPs) are a class of small molecular proteins widely found in plants, capable of withstanding various biotic and abiotic environmental stresses, including temperature stress, drought, nutrient deficiency, oxidative imbalance, salt and heavy ion toxicity, and pathogenic infections. Enhancing our understanding of USPs holds significant potential for improving plant stress resilience. This study focuses on 13 species of *Solanum*, including cultivated and wild tomatoes, and systematically identified 438 members of the USP gene family through bioinformatics approaches. Phylogenetic analysis reveals that major USP members are conserved within *Solanum*, with interspecies differences in USP numbers primarily attributed to copy number variation (CNV). Through synteny and homology analyses, we found that *USP27* and *USP28* are unique to tomatoes, while the homologous gene of *USP19* is absent in cultivated tomatoes. Notably, five unique USP genes are present in *S. pennellii*, which is characterized by its early differentiation and resistance advantages. *Ka/Ks* analysis indicates that only the *USP10/21* homologous gene pair has undergone positive selection in wild tomatoes, while all other genes are subject to strong negative selection. The USPs in *Solanum* exhibit high consistency in domain characteristics, sequence conservation, and types of promoter regulatory elements, although there are substantial differences in the number of these elements. Utilizing publicly available data, we identified eight USPs that have undergone domestication or improvement selection, particularly noting the tissue-specific expression patterns of domesticated *SolycUSP3/28/30*. Through graph pangenome analysis, we screened 12 USPs covered by high-confidence structural variants, which primarily disrupt the intron regions of USPs, leading to significant differences in their expression responses to salt stress. We anticipate that these findings will provide a theoretical foundation and prior knowledge for further understanding and application of USP in plants.

## KEYWORDS

USP gene family, tomato, domestication, structural variants, expression

## Introduction

The rapid changes in the global environment and the increasing frequency of extreme weather events pose significant challenges to plant growth and crop production. Understanding the stress and defense systems that plants have evolved to cope with mixed and complex adverse environmental conditions is crucial for enhancing and improving plant adaptability. Universal stress proteins (USPs) are small molecular proteins that exhibit responses and resistance to multiple stressors; they were initially identified in *Escherichia coli* (Vanbogaert et al., 1990) and have since been found to be widely present in bacteria, where they can be significantly upregulated in response to various adverse environmental factors, including temperature stress, nutrient deficiency, oxidative imbalance, ion toxicity, and heavy metal exposure (Kvint et al., 2003). A notable characteristic of USP proteins is the presence of one or more USP domains (Foret et al., 2011). Based on their genetic diversity and sequence homology, USPs in *E. coli* can be classified into four categories, which interact synergistically or antagonistically to perform specific functions in mitigating oxidative stress, iron scavenging, cell motility, and adhesion (Kvint et al., 2003; Nachin et al., 2005).

USPs are also present in plants, having been first identified in rice (Sauter et al., 2002). The number of USP proteins in higher plants varies from several dozen to several hundred, exhibiting remarkable structural and functional diversity. In addition to the typical USP domain, plant USP proteins incorporate various structural domains such as kinase, U-box, zinc finger, Homeodomain leucine zipper (HDzip), and cation exchanger (Chi et al., 2019). This rich array of structural features endows plant USPs with highly diversified functional motifs, enabling them to respond to and resist increasingly complex environmental stresses, including extreme temperatures, drought, hypoxia, and salinity, making them crucial components of the plant adaptability system. For instance, the *Arabidopsis thaliana* AtSUP can respond to heat or cold stress, serving as a molecular chaperone to maintain the activity of target proteins or RNAs, thereby enhancing the stress tolerance of *A. thaliana* (Gonzali et al., 2015; Jung et al., 2015; Kang et al., 2013; Melencion et al., 2017). In tobacco, USPs can recognize phosphorylation signals induced by salinity and osmotic pressure, leading to the activation of target gene expression and the accumulation of osmotic regulatory substances, as well as the maintenance of reactive oxygen species (ROS) homeostasis (Udawat et al., 2016). In *Gossypium arboreum*, the expression of two USPs is induced by drought stress, and they play a role in regulating cellular water content (Maqbool et al., 2008; Maqbool et al., 2009). Beyond their well-established roles in resisting abiotic stresses, USPs also participate in plant responses to biotic stresses. Following infection by pathogens or pathogenic factors, the expression levels of certain USP genes significantly increase or undergo phosphorylation, thereby activating plant defense-related signaling pathways (Chou et al., 2007; Lenman et al., 2008; Merkouropoulos et al., 2008).

Tomato is one of the most significant vegetables globally, and both tomatoes and their processed derivatives are indispensable components of the human diet. Enhancing tomato adaptability

holds considerable theoretical and practical value. Recent studies have demonstrated that universal stress proteins (USPs) in tomatoes respond to salt and drought stress, with their overexpression modulating stomatal opening, increasing abscisic acid and chlorophyll content to prevent leaf damage, indicating that USPs contribute to the stress tolerance of tomatoes (Loukehaich et al., 2012; Ijaz et al., 2017).

In contrast to cultivated tomatoes, wild tomatoes typically exhibit higher genetic diversity and enhanced resistance to biotic and abiotic stresses. For instance, wild tomato species such as *Solanum penellii*, *Solanum galapagense*, and *Solanum habrochaites* possess robust salinity, cold, and frost tolerance. Wild tomatoes serve as valuable donors for improving the stress tolerance of cultivated tomatoes and have been repeatedly utilized in breeding programs aimed at enhancing the adaptability of cultivated varieties (Kapazoglou et al., 2023). The assembly and publication of the *Solanum* pangenome have facilitated the integration and utilization of high-adaptability genetic resources from wild tomatoes (Li et al., 2023). Considering the pivotal role of USP genes in plant stress responses, an exploration of the genomic characteristics of the USP gene family within the adaptively diverse *Solanum* genus is of significant value for elucidating the origins of adaptability in wild tomatoes. This study adopts a global perspective at the genus level, identifying 438 USP genes from 13 *Solanum* species. Through phylogenetic analysis and homology inference, we explored the evolutionary trajectory of the USP gene family within *Solanum* and screened four groups of unique USP genes. The USP motifs in *Solanum* are highly conserved, while the number of *cis*-regulatory elements (CREs) in their promoter regions varies. Furthermore, using the pangenome, we identified USP gene groups influenced by structural variants (SVs) and domestication selection. Expression profile data reveals that intronic SVs broadly affect USP expression patterns, leading to alterations in their response capabilities to salt stress. Overall, these findings provide insights into the *Solanum* USP gene family and pave the way for the utilization of high-resistance genetic resources from wild tomatoes.

## Materials and methods

### USP gene family identification

The whole-genome sequences of 13 *Solanum* species were collected from published data (Supplementary Table S1). Each species was independently analyzed for the USP gene family. The hidden Markov model (HMM) for the USP domain was downloaded from the InterPro database (<http://pfam-legacy.xfam.org/>, available on November 10, 2024) with accession number PF00582. We searched for potential USP domains in the whole-genome protein sequences using HMMER (Potter et al., 2018; <http://hmmer.org/>), considering domain scores with an e-value less than 1e-5 as potential USP genes. Additionally, the USP sequences from *Arabidopsis thaliana* were used as a query in a blastp search against the whole-genome protein sequences, classifying proteins with a similarity rate greater than 30% as potential USP genes. All candidate genes were

manually submitted to SMART (<https://smart.embl.de/smart/>, available on November 10, 2024) and the CDD database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/>, available on November 10, 2024) for domain and integrity verification. Candidates lacking the USP domain, having incomplete USP domains, or with an amino acid length of less than 100 were removed, resulting in the final list of potential USP genes in *Solanum*.

## Phylogenetic tree construction

The phylogenetic tree was constructed using all USP amino acid sequences. Multiple sequence alignment (MSA) was performed with Muscle using default parameters (Edgar, 2004). The MSA results were trimmed using trimAL (Capella-Gutiérrez et al., 2009), removing columns with a gap rate greater than 20% and a consistency ratio lower than 0.001. The trimmed MSA was then used to build the phylogenetic tree with IQ-TREE (version 2) employing the maximum likelihood method (Minh et al., 2020). The optimal model was automatically calculated by the software, with 1000 bootstrap replicates. Tree visualization and enhancement were performed using iTOL (<https://itol.embl.de/>, available on November 11, 2024).

## Homology and synteny analysis

We conducted synteny analysis using MscanX (Wang et al., 2012), comparing all species pairwise to identify syntenic regions with an e-value of  $1e-5$  and a minimum of 10 blast hits. The output results were manually organized based on the order of species divergence, and synteny visualization was performed using TBtools (version 2) (Chen et al., 2023). Homology screening and orthogroup classification of all USP genes in *Solanum* were conducted using OrthoFinder2 (Emms and Kelly, 2019; <https://github.com/davidemms/OrthoFinder>). The coding sequences (CDS) of all direct homologous genes in each orthogroup were analyzed for *Ka* and *Ks* values using KaKs\_Calculator (version 3) (Zhang, 2022). The distribution of *Ka/Ks* values was visualized using the R package ggridges (<https://github.com/wilkelab/ggridges>).

## Motif and *cis*-regulatory element screening

The amino acid sequences of USP proteins from wild and cultivated tomatoes were analyzed for motifs using MEME Suite (v4.12.0) (Bailey et al., 2006), with a minimum motif length of 6, a maximum length of 50, a maximum of 10 motifs, and an e-value threshold of  $1e-5$ . A 2 kb upstream region of all *Solanum* USP genes was extracted as the promoter region, and the USP promoters for both wild and cultivated tomatoes were submitted to the PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>, available on November 11, 2024) for *cis*-regulatory element (CRE) prediction.

## Graph pan-genome construction and structural variation visualizations

Utilizing the PGGB pipeline (Garrison et al., 2024, <https://github.com/pangenome/pggb>), each tomato variety was treated as a haplotype, and the graph pan-genome was constructed by chromosome. The percent identity for mapping/alignment was set to 90%, with a minimum block length filter for mapping of 90 bp and a k-mer size for mapping of 19, while all other parameters were maintained at their default settings. The cultivated tomato (SL5) served as the reference path within the graph pan-genome. Structural variants (SVs) were deconstructed using VG (Garrison et al., 2018, <https://github.com/vgteam/vg>), with a maximum variant length of 100 kb, deconstructing both top-level snarls and nested snarls. The ODGI tool (Guarracino et al., 2022, <https://github.com/pangenome/odgi>) was employed to extract the graph structure of the USP gene body region and to generate a 2D visualization. Subsequently, SVs within the filtered USP gene body were displayed using Bandage (Wick et al., 2015).

## The domestication selection of USP genes

Using previously published data from natural tomato populations, analyses of domestication selection and genetic diversity were conducted (National Center for Biotechnology Information accession: SRP045767). BWA-MEM was utilized to align resequencing data from cultivated and wild tomato populations against the cultivated tomato reference genome (SL5.0) (Li, 2013). Samtools was employed for BAM file sorting and index creation (Li et al., 2009). SNP calling was performed using DeepVariant (<https://github.com/google/deepvariant>) in WGS mode on the sorted BAM files (Poplin et al., 2018). VCF merging was accomplished using Glnexus, configured to DeepVariant\_unfiltered (Yun et al., 2021). Multi-allelic variants, with a minor allele frequency (MAF) of less than 0.05, a missing rate greater than 0.2, or a depth exceeding 1.8x or falling below 0.3x of the average depth, were filtered out. Based on previously published results regarding domestication and improvement segments in cultivated tomatoes (Lin et al., 2014), and the positional information of USP genes, USP genes subjected to domestication selection were identified. Candidate domestication selection genes were evaluated for pi values using Vcftools (Danecek et al., 2011), and R was utilized to create a line graph visualizing the pi values for cultivated and wild tomatoes.

## Tissue-specific and salt stress expression analysis

Expression analysis was conducted for cultivated tomato (*Solanum lycopersicum*) utilizing publicly available, standardized transcriptome databases. USP tissue-specific expression data were collected from the Tomato Functional Genomics Database (TFGD):

<http://ted.bti.cornell.edu/cgi-bin/TFGD/digital/home.cgi>, available on November 10, 2024). Additionally, expression levels of USP in the roots of cultivated and wild tomatoes under salt stress were obtained from previously published transcriptomic data (CHINA CITIC BANK CORPORATION LIMITED accession: CRA004289) (Li et al., 2022). All expression levels were initially standardized, followed by visualization through a heatmap, which was generated using the R package complexHeatmap (Gu, 2022).

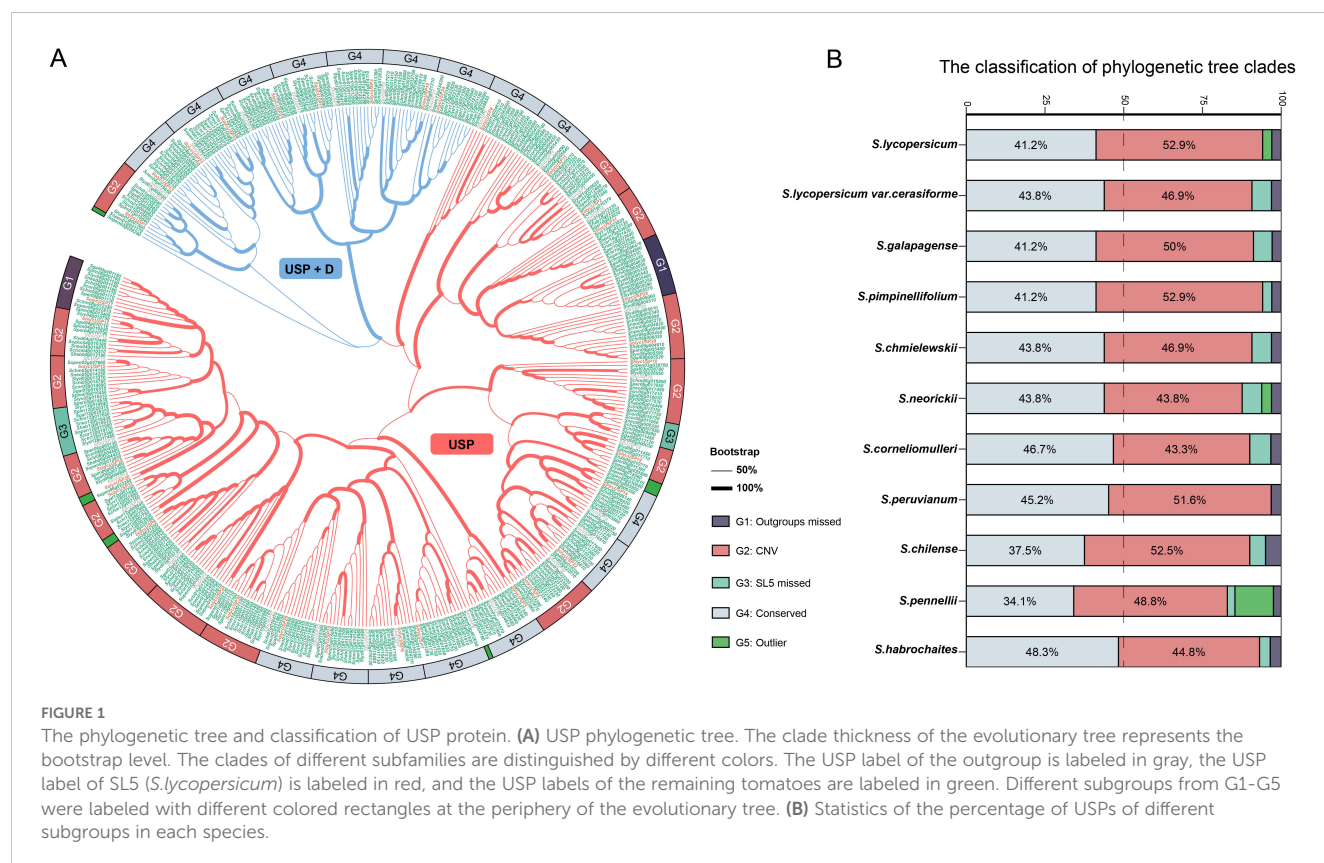
## Result

### The identification and classification of USP gene in *Solanum*

Domain identification and homology comparison methods were simultaneously employed for the identification of USP genes. Following the confirmation of domain integrity, a total of 438 potential USP genes were collected from *Solanum* (Supplementary Table S2), with 31 sourced from potato (considered as outgroups), 66 from *S. lycopersicum* and *S. lycopersicum* var. *cerasiforme* (designated as cultivars), and 300 from eight wild tomato and related species (designated as wild). Among the cultivated tomatoes, 34 potential USP genes were identified, which were renamed as *SolycUSP1* to *SolycUSP34* based on their chromosomal positions and order. All these USPs were assigned to a likelihood tree (Figure 1A), where bootstrap analysis supported the robustness of the major clades and branches. Phylogenetic analysis indicates that the USP genes in

*Solanum* are primarily divided into two categories based on their domain characteristics: 333 USPs (approximately 76% of all USPs) possess a single typical USP domain, occupying the predominant portion of the evolutionary tree (designated as the USP subfamily), while 105 USPs (approximately 24% of all USPs) occupy another clade, possessing one or more additional domain types alongside the USP domain, primarily protein kinases (designated as the USP +D subfamily).

The phylogenetic tree was divided into 39 groups (Figure 1A), with each branch representing a unit based on the respective *SolycUSP*, specifically: Group G1 lacks USPs from the outgroup species (potato); USPs in Group G2 exhibit CNVs; Group G3 is devoid of USPs from *S. lycopersicum*; USPs in Group G4 are highly conserved; and Group G5 comprises outlier USPs. These groupings generally correlate with phylogenetic lineages but also exhibit nesting within different clades. USP27 and USP28 represent two unique groups of USP genes in tomato, with outgroup missing in this group (G1). In the two G3 groups, the absence of cultivated tomato USPs is noted; one G3 group comprises solely wild tomatoes, while the other includes both wild tomatoes and outgroup, with *USP18* and *USP19* being the closest relatives in the phylogenetic tree, suggesting their potential loss during the evolution of cultivated tomatoes. The five G5 group USPs are distributed across isolated branches, with USPs assigned from cultivated tomatoes, wild tomatoes, and outgroup. Sixteen G4 groups (approximately 41% of the total number of groups) exhibit the most conserved phylogenetic characteristics, with each species containing only one USP within these groups. Fifteen G2



groups (approximately 38% of the total groups) show varying degrees of interspecific gene copy number variation (CNV), with some wild tomato USPs exhibiting increases or decreases in gene numbers. G2 may represent USPs in tomatoes that have been subjected to evolutionary perturbation. Based on this grouping, the types and proportions of USPs in each tomato species were statistically analyzed (Figure 1B). The G2 and G4 groups comprise the majority of all tomato USPs, where, except for *S. neorickii*, *S. corneliumulleri*, and *S. habrochaites*, G2 group USPs exceed G4 group USPs in most tomatoes, indicating widespread CNV of USPs in tomatoes. Among the remaining three groups, G5 is distinct, primarily concentrated in the USPs of *S. pennellii*, suggesting a unique differentiation of USPs in this species.

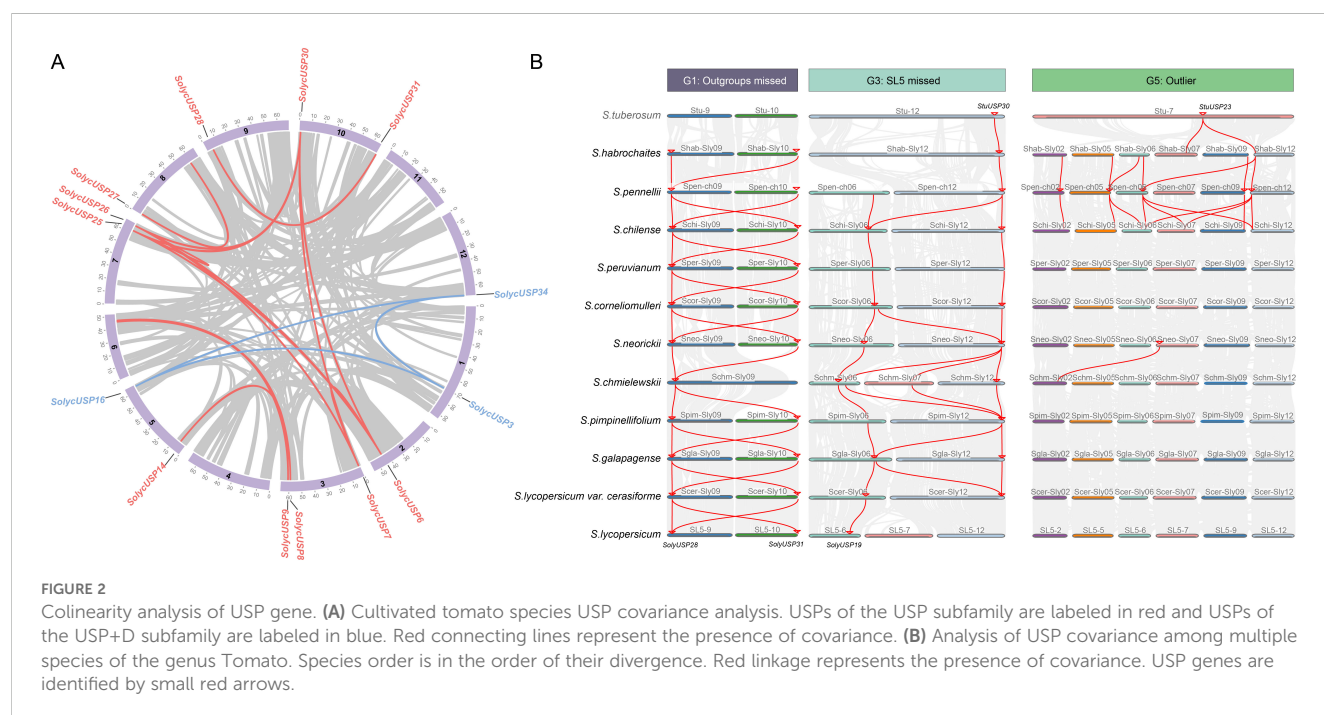
## The collinearity analysis of USP gene in *Solanum*

Collinearity analysis reveals the presence of four pairs of collinear gene pairs in cultivated tomatoes (Figure 2A), among which the USP subfamily includes a pair of collinear genes *USP3/16/34*, while the USP+D subfamily comprises three pairs of collinear genes: *USP6/7/25/26/30*, *USP9/14*, and *USP28/31*. All collinear genes are generally located at different positions on different chromosomes, indicating that segmental duplication predominantly drives the expansion of *SolycUSPs*. Furthermore, the collinearity changes of USPs among tomatoes were explored based on the divergence order of all species. The USPs in G2 and G4 groups exhibited good collinearity across different tomatoes, particularly G4 group USPs, which demonstrated nearly complete conservation of collinearity among species, highlighting the highly conserved nature of USPs in G4 (Supplementary Figure S1). We focused on the G1, G3, and G5

groups, where collinearity exhibited variation among species (Figure 2B). The G1 group includes *USP28* and *USP31*; with the exception of *S. chmielewskii*, these two USPs maintain stable collinear relationships in other tomatoes. In cultivated tomatoes, they form a pair of collinear gene pairs, while potato USP is absent from this group. *SolycUSP28*, located on a branch in the phylogenetic tree, lacks potato USP, whereas *SolycUSP31* shares the same branch as *StuUSP*, suggesting that after the divergence of potato and tomato ancestors, *SolycUSP31* underwent segmental duplication to produce the tomato-specific *SolycUSP28*, which subsequently lost collinearity with potato during further divergence. The G3 group contains *StuUSP30*, which underwent duplication on chromosome 6 after the divergence of *S. habrochaites* and *S. pennellii*, resulting in *USP19*. Both genes were retained in wild tomatoes, while the collinear *SolycUSP* associated with the original *StuUSP30* was lost in cultivated tomatoes, corresponding to the G4 group in the phylogenetic tree, where *USP19*, derived from duplication, remains adjacent to the G4 group branch. G5 group USPs primarily emerged and expanded in *S. habrochaites*, *S. pennellii*, and *S. chilense* tomatoes formed during early divergence, subsequently disappearing in later tomatoes. Although these USPs retained collinear relationships in the early three tomatoes, they were only identified as candidate USPs in *S. pennellii*.

## The Ka/Ks analysis of orthologues USP gene in *Solanum*

To clarify the extent of evolutionary constraints on USPs in *Solanum*, we identified the direct and paralogous relationships among all potential USP genes, categorizing them into 21 orthogroups. We calculated the *Ka* and *Ks* values between all



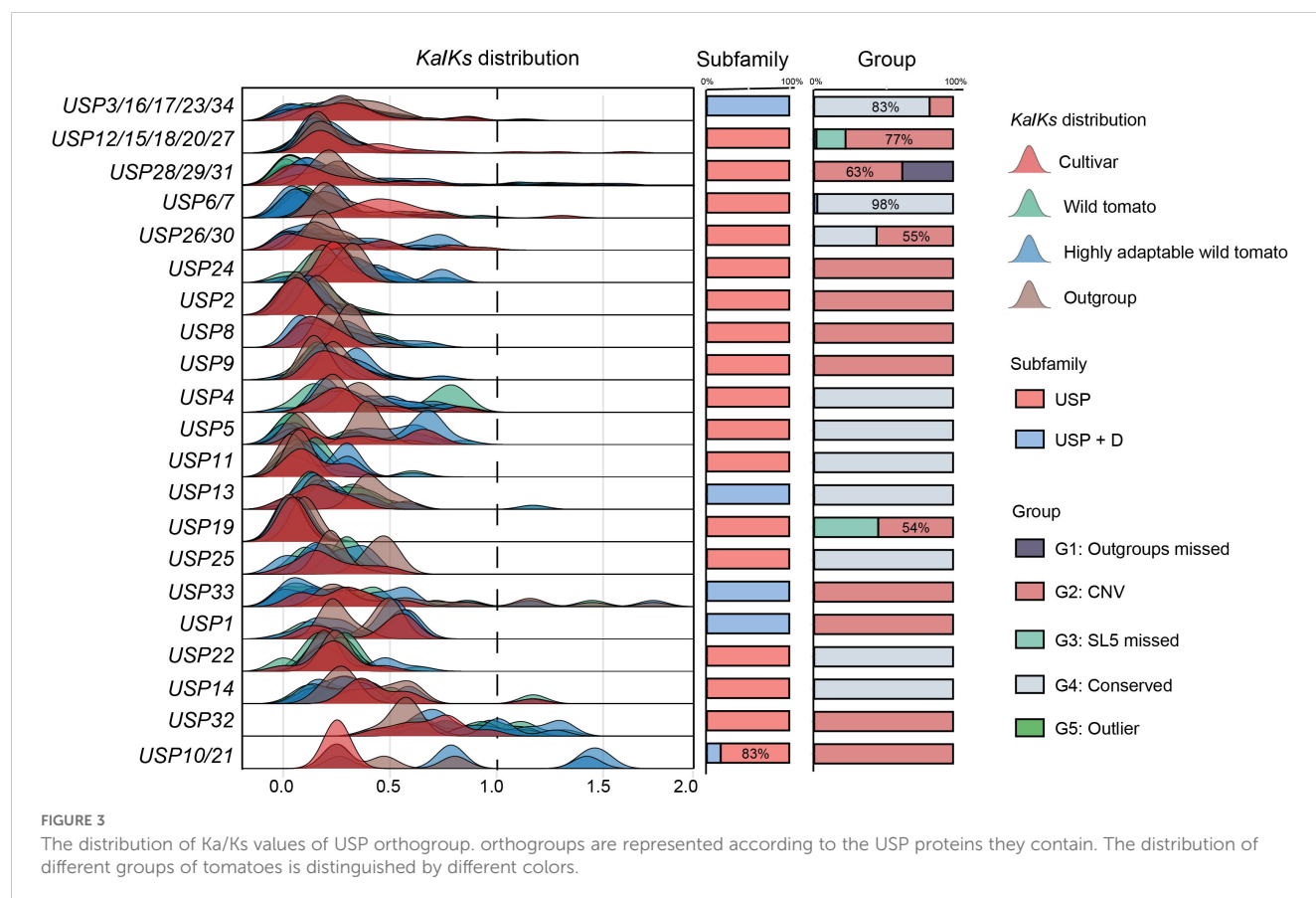
orthologues within each orthogroup (Figure 3). Each orthogroup was designated based on the USP genes it contained from cultivated tomatoes. Notably, several tomato species with broad or strong resistance to environmental stress (*S. pennellii*: tolerant to salt and drought stress; *S. habrochaites*: tolerant to cold stress; *S. galapagense* and *S. chmielewskii*: tolerant to salt stress) were classified as highly adaptable wild tomatoes. The results reveals that in 17 orthogroups, the  $Ka/Ks$  ratio for USPs across all species was significantly less than 1, indicating that the majority of USP genes are subjected to stringent negative selection, leading to a stabilization of gene function throughout evolution. In orthogroups containing USP33, USP32, and USP10/21, some genes exhibited  $Ka/Ks$  values greater than 1, indicating they are under positive selection. These genes belong to the G2 group and are primarily derived from highly adaptable wild tomatoes, suggesting that CNV may have driven the evolution of USPs to some extent, with multi-copy redundancy facilitating gene neofunctionalization or sub functionalization, thus potentially forming the genomic basis for the high adaptability of wild tomatoes.

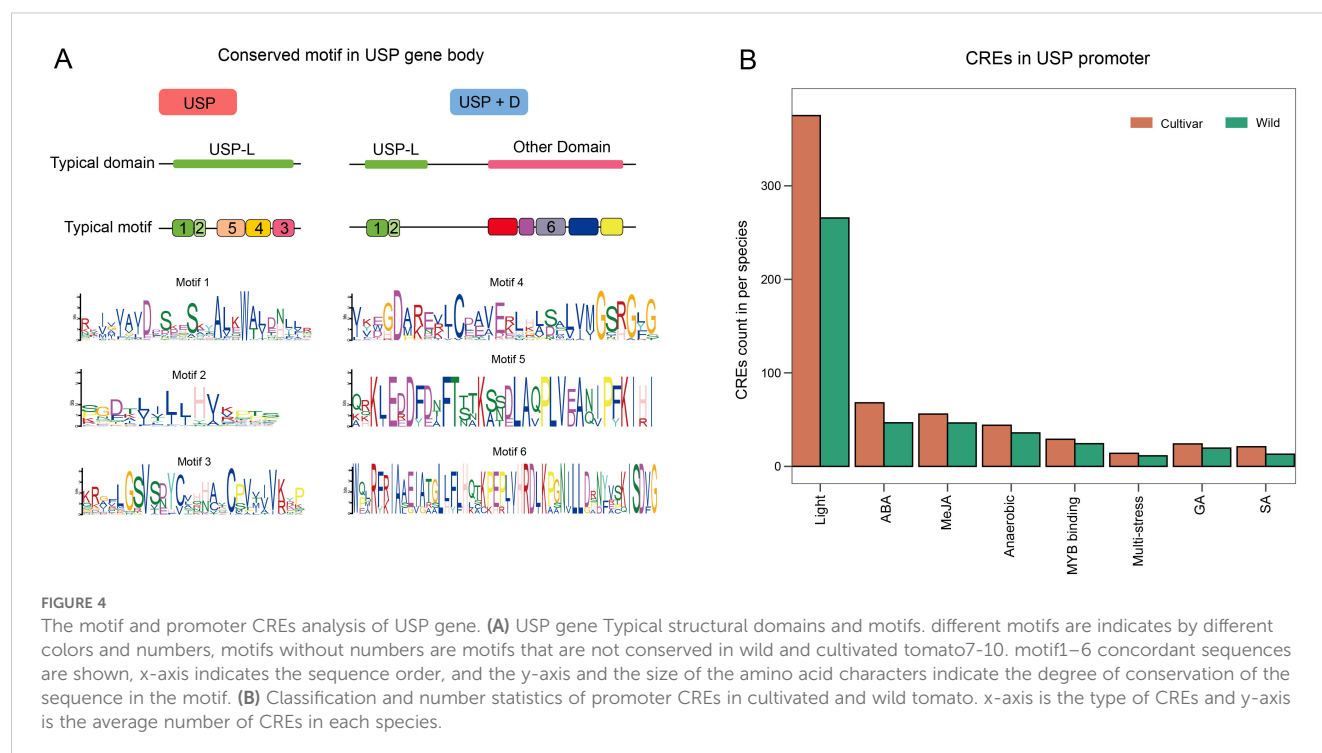
## The motif and CREs analysis of USP gene in tomato

Motifs are closely related to protein function. We independently searched for and predicted 10 motifs from all USPs in both wild and cultivated tomatoes (Figure 4A). The independent prediction results

indicates that six motifs are highly conserved between wild and cultivated tomatoes, while four motifs exhibit considerable variability. The highly conserved motifs 1–5 form the USP domain, whereas the conserved motif 6 and the remaining four variable motifs are concentrated in the variable domains of the USP +D subfamily. This suggests that there may not be a significant functional differentiation of USP genes between wild and cultivated tomatoes, given their stable and consistent motif distribution. The functional differences of USP genes in tomatoes appear to be primarily concentrated in the USP+D subfamily, where the USP proteins possess diverse structures to cope with complex stress environments, potentially leading to substantial variation among different tomato species. Notably, although both subfamilies contain a USP domain, motif analysis reveals that the USP domain characteristics seem to differ; USPs in the USP+D subfamily possess only motifs 1 and 2, lacking motifs 3, 4, and 5.

CREs play a crucial role in triggering TFs to bind and regulate gene expression. We independently searched for and predicted CREs in the promoters of all USP genes from wild and cultivated tomatoes (Figure 4B). Among the top ten ranked CREs, the types of CREs were completely consistent between wild and cultivated tomatoes. Except for those associated with light and MYB binding, all other CREs were related to hormone or environmental responses, aligning with the functional role of USPs in responding to various stress conditions. There are significant differences in the number of CREs among different tomato species, with cultivated tomato USP promoters exhibiting





a greater average number of all CRE types compared to wild tomatoes. This indicates that while USPs in different tomato species may respond to similar environmental factors, their expression levels and patterns differ significantly.

## The SVs and domestication of USP gene in tomato

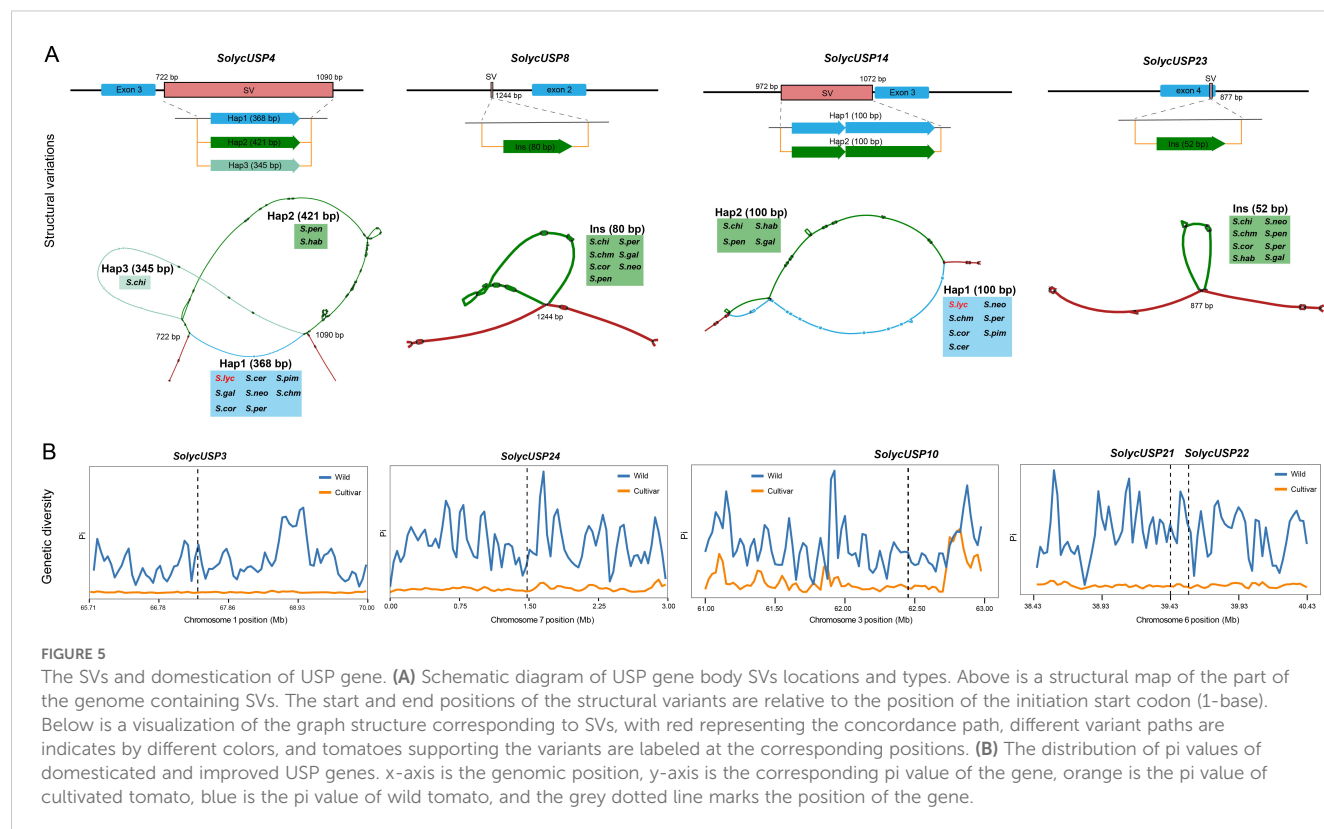
To characterize the variation in the USP gene regions, we followed the PGGB pipeline to generate a graph pangenome that includes the whole genomes of two cultivated tomatoes and nine wild tomatoes (Supplementary Table S3), thereby deconstructing the variations among all tomato species. Due to the high genetic diversity of wild tomatoes, all USPs were found to be covered by variations. SVs typically have greater causal effects, so we focused on SVs supported by multiple tomato species within the USP gene body regions, avoiding overly complex SVs that may arise from graph construction or other factors. Ultimately, 11 USP genes were identified as having stable and reliable SVs between wild and cultivated tomatoes (Supplementary Table S2), with these variations primarily located in the intronic regions of USPs. We illustrate the SVs for four of these genes (Figure 5A): in the *SolycUSP4* region, a 368 bp polymorphic interval exists between the third and fourth exons, supported by *S. pennellii* and *S. habrochaites* for Hap2, and by *S. chilense* for a unique Hap3. In *SolycUSP14*, a 100 bp polymorphic interval is present upstream of the C-terminus of the third exon, with *S. chilense*, *S. habrochaites*, *S. pennellii*, and *S. galapagense* supporting a different Hap2 compared to other tomatoes. Additionally, within the gene regions of *SolycUSP8* and *SolycUSP23*, there are insertions of 80 bp

(supported by seven wild tomato species) and 52 bp (supported by eight wild tomato species), respectively, with the insertion in *SolycUSP23* occurring within its fourth exon.

Previous studies have identified potential genomic segments in cultivated tomatoes that are influenced by domestication and improvement. Among all USPs, *SolycUSP3* and *SolycUSP24* are located within segments under domestication selection, while *SolycUSP10/21/22/28/29/30* are situated within segments under improvement selection. To verify the selection pressures on USPs, we generated Pi value distributions for cultivated and wild tomatoes from natural populations to observe changes in genetic diversity (Figure 5B). Compared to wild tomatoes, the eight USPs subjected to domestication or improvement exhibited a dramatic reduction in genetic diversity in cultivated tomatoes, indicating that these USPs underwent strong selection during the domestication process.

## The expression analysis of USP gene in tomato

Using publicly available expression data, we generated a tissue-specific expression heatmap for *SolycUSPs* (Figure 6A). Among the two genes affected by domestication selection, *SolycUSP3* was expressed in vegetative organs and early fruit development, with particularly high specificity in flowers and buds. *SolycUSP24* exhibited specific expression mainly in vegetative organs, especially in leaves. Domestication selection has led to a reduction in the expression levels of both *SolycUSP3* and *SolycUSP24*. This effect is particularly pronounced among the six genes subjected to improvement selection, with *SolycUSP30* and *SolycUSP28* highly expressed in nearly all cultivated tomato tissues, while only



*SolycUSP20* showed some expression in the leaves of wild tomatoes. *SolycUSP10* and *SolycUSP21*, located in positively selected orthogroups, are also considered affected by improvement selection. Notably, while these two genes exhibit low expression in the fruit of wild tomatoes, their expression increases in cultivated tomatoes. Among the USPs not subjected to domestication or improvement selection, three primary expression patterns were observed: four *SolycUSPs* (*SolycUSP2/26/13/7*) displayed high expression across nearly all vegetative and reproductive organs; six *SolycUSPs* (*SolycUSP1/4/11/17/23/33*) were highly expressed primarily in fruits, particularly in fully mature fruits, with similar expression patterns in both cultivated and wild tomatoes. Additionally, ten *SolycUSPs* were expressed almost exclusively in flowers or flower buds (*SolycUSP15* to *SolycUSP27*), showing little to no expression in other tissues. Furthermore, structural variations did not significantly impact the tissue-specific expression of USPs.

## SV probably enhanced tomato USP response to salt stress

We explored the response of USPs to salt stress using previously published RNA-seq data from the roots of cultivated tomato (M82) and wild tomato (*S. pennellii*) under salt stress (Figure 6B). More than half of the *SolycUSPs* were highly expressed in response to salt stress, with 12 *SolycUSPs* showing significant differences in expression levels between cultivated and wild tomatoes. Under salt stress conditions, nearly all differentially expressed genes (DEGs) exhibited increased expression in wild tomatoes

compared to cultivated tomatoes, indicating that the high salt tolerance of *S. pennellii* may be associated with the increased mRNA levels and protein abundances of *SpenUSPs*. Notably, among the 11 USPs identified as having stable and reliable structural variations, 9 exhibited differential expression between cultivated and wild tomatoes; only *SolycUSP5* and *SolycUSP1* did not contain SVs. This may suggest that haplotype variations, deletions, and insertions of USPs significantly influence their expression responses to stress. Nine *SolycUSPs* (*SolycUSP2* to *SolycUSP10*) were highly expressed but did not respond to salt stress, showing no expression differences between cultivated and wild tomatoes. These 9 high expressed genes exhibited constitutive expression in the roots of tomatoes. Furthermore, 13 *SolycUSPs* showed minimal or low expression in both cultivated and wild tomatoes, and they were also not induced by salt stress.

## Discussion

It is generally accepted that modern tomatoes encompass 16 wild species or closely related wild relatives. Extended evolution, domestication, and breeding efforts have significantly enriched the phenotypic and genotypic diversity of tomatoes (Li et al., 2018; Peralta and Spooner, 2024). The assembly and publication of the pan-genome of the *Solanum* have greatly accelerated the process of utilizing wild tomato resources for the enhancement of modern tomato breeding, facilitating genomic studies across the entire genus (Li et al., 2023). Our research focuses on the USP gene family, which is closely associated with adaptability—the primary

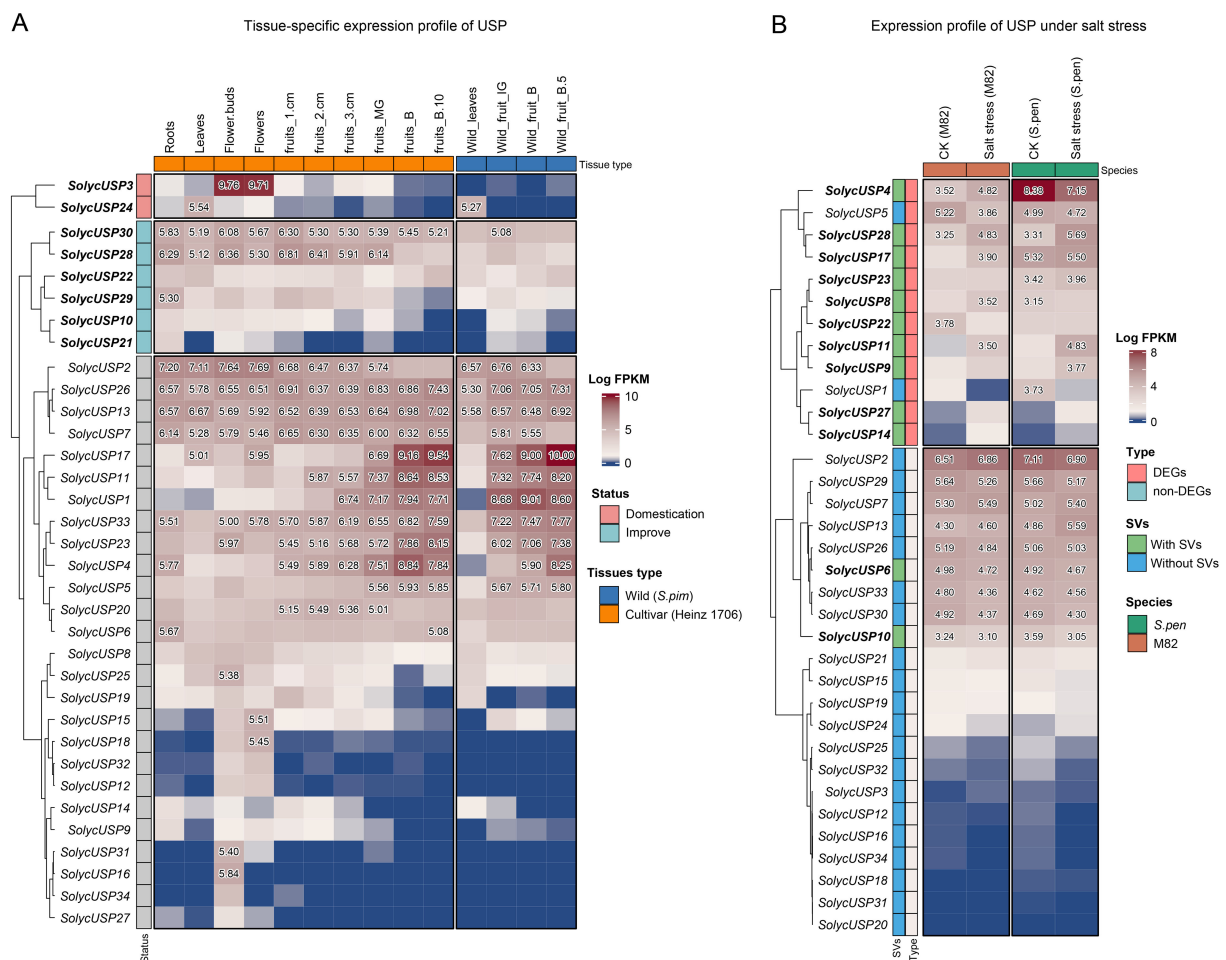


FIGURE 6

The tissue-specific and salt-response expression profile of USP gene. (A) Heatmap of tissue-specific expression of USP. Expression was processed using log2 and numbers greater than 5 are displayed. The status of genes subjected to domestication selection is indicated at gene id. (B) Heatmap of USP expression in response to salt stress. Expression was processed using log2 and numbers greater than 3 are shown. Whether genes were shown significant expression differences and whether they had SV is shown at gene id.

objective of contemporary tomato breeding. Initially identified in bacteria, these genes are characterized by the presence of a single USP domain, whose function is believed to be related to oxidative stress defense and iron scavenging (Nachin et al., 2005). Plants exhibit a vast array of USP types and quantities. We identified over 400 potential USPs from the *Solanum*, with the number of USPs per species ranging from 29 to 40. Notably, a previous study on the potato USP gene family collected 108 USP proteins from the entire genome (Qi et al., 2024), a number considerably exceeding that identified in our study. Although variations in gene family counts may arise from differences in genome annotation, identification methods, and filtering criteria, the substantial discrepancy remains unexplained. Our comparative analysis across species ensures that relevant assessments are conducted within the same dimension, thus we maintain our identified count. Most USP proteins in *Solanum* contain only a single USP domain, representing the primary composition, typical features, and conserved functions of the USP family (Vollmer and Bark, 2018). Another category of USPs includes an additional domain; these USPs may arise from the

fusion of USP domains with various structural domains prompted by diverse and severe environmental conditions (Tkaczuk et al., 2013). This undoubtedly enhances the diversity of the USP gene family and equips plants to confront increasingly complex adverse stresses. In *Solanum*, this subset of USPs, beyond the typical domain, also encompasses PKs, Serine/Threonine kinases, AANH, and FDF domain. These additional domains may confer capabilities for interaction with other proteins, phosphorylation, and stabilization and folding of protein conformations.

Through phylogenetic analysis of all USP proteins, we discovered that copy number variation (CNV) is widely present among USPs of *Solanum*. CNV affects gene numbers through mechanisms such as whole-genome duplication, retention, and segmental duplication. It holds significant implications for species evolution, environmental adaptability, specific gene expression, and regulation, and has been extensively studied (Freeman et al., 2006; Zarrei et al., 2015). As observed in other species, the prevalent CNVs among different tomatoes may be closely related to their adaptive differences (Cho et al., 2023; Kumar et al., 2023; Stalder

et al., 2023). Notably, the USP group G5 is uniquely concentrated in *S. pennellii*, exhibiting collinearity with some genes in *S. habrochaites* and *S. chilense* that lack the typical characteristics of USP domains, which subsequently disappeared during the evolutionary process of tomatoes. *S. pennellii* is renowned for its robust drought, salt, pest, and disease resistance, making it a primary genetic resource and foundation for adaptive breeding in cultivated tomatoes (Bolger et al., 2014). This suggests an intriguing hypothesis that the exceptional resistance traits of *S. pennellii* may be associated with the specific enrichment of unique USPs within this species.

Graph pangenomes represent the forefront of genomic development and serve as a powerful tool for elucidating SVs, having been widely employed in studies of genetic evolution and association mapping (Groza et al., 2024; Khan et al., 2024; Liu et al., 2024). Constructing high-quality and reliable plant graph pangenomes remains challenging due to the complexity of genomic structures and the slow development of corresponding tools. Rather than exploring the global perspective of graph pangenomes, focusing on specific important genomic regions or gene bodies may be more suitable given the current state of graph pangenome development. Such methodologies have been successfully applied in areas such as genomic structure, locus evolution, adaptive breeding, and evolutionary studies, yielding significant advancements (Bolognini et al., 2024; Jamsandekar et al., 2024; Jayakodi et al., 2024). Our research similarly concentrates on variations within the USP gene body regions. To avoid false-positive variations arising from structural complexities, we excluded USP genes with highly convoluted graph structures, ensuring that SVs are supported by multiple haplotypes. These SVs were confirmed through 2D visualization of graph regions. We identified eight USPs covered by stable SVs, which include haplotype variations, insertions, and deletions, with each SV also containing several minor nested variations. This undoubtedly enhances the genetic diversity and heterogeneity of USPs, offering opportunities for their utilization. Notably, SVs predominantly reside within the intron regions of USPs. Analysis of expression profiles from cultivated and wild tomatoes under salt stress indicates that these SVs significantly influence the expression patterns of the corresponding USP genes. Similar phenomena have been observed in other studies, where SVs may affect alternative splicing processes or alter the status of regulatory elements such as enhancers, thereby significantly impacting mRNA abundance of target genes, particularly in complex quantitative traits (Pagani and Baralle, 2004; Chiang et al., 2017; Beird et al., 2024). We hypothesize that these SVs may arise from the balance between evolution and natural selection, adjusting the SVs within USPs to enhance or diminish gene function, thus better adapting to complex environments while avoiding the pressure to generate new genes. It is essential to note that accurately resolving SVs using graph structures constructed from multiple haplotype genomes remains challenging. The breakpoint positions and chimeric variations of these SVs may not align with true biological realities. The conclusions presented in this study are predominantly descriptive. Therefore, we strongly recommend that

subsequent research endeavors aimed at validating these hypotheses incorporate rigorous empirical validation through molecular experimentation or genetic transformation.

## Conclusion

This study systematically identified 438 USP genes across 13 species of the *Solanum*. USP genes among diverse *Solanum* species exhibit widespread CNVs yet are predominantly evolutionarily conserved. The domain architecture and motif composition of *Solanum* USP genes demonstrate considerable consistency, notwithstanding significant variations observed in the abundance of their promoter-associated regulatory elements. Six *SolycUSPs* manifested reduced genetic diversity throughout the domestication process, accompanied by discernible modifications in their tissue-specific expression patterns. Eleven USP genes possess gene bodies affected by high-confidence SVs, which predominantly impact the intronic regions and precipitate widespread alterations in their transcriptional responses to salt stress. *S. pennellii* possesses a greater number of unique USPs, which may be associated with its robust adaptability. Overall, these findings provide valuable insights into the USP gene family within the *Solanum*.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author/s.

## Author contributions

RX: Data curation, Formal analysis, Methodology, Writing—original draft. ZhW: Data curation, Methodology, Writing—review & editing. ZC: Project administration. ZeW: Soft-ware. QM: Visualization. NL: Conceptualization. YQ: Conceptualization, Data curation, Writing—review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was financially supported by the Project of Fund for Stable Support to Agricultural Sci-Tech Renovation (xjnkywdzc-2025001-05); National Natural Science Foundation of China (32260763).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product

that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1546640/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Colinearity analysis of G2 and G4 group.

## References

- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Beird, H. C., Lin, D., Lazar, A. J., and Futreal, P. A. (2024). Patterns of structural variants within introns and relocation of the promoter: a commentary. *J. Pathol.* 263, 131–134. doi: 10.1002/path.6270
- Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., et al. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* 46, 1034–1038. doi: 10.1038/ng.3046
- Bolognini, D., Halgren, A., Lou, R. N., Raveane, A., Rocha, J. L., Guarracino, A., et al. (2024). Recurrent evolution and selection shape structural diversity at the amylase locus. *Nature* 634, 617–625. doi: 10.1038/s41586-024-07911-1
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010
- Chi, Y. H., Koo, S. S., Oh, H. T., Lee, E. S., Park, J. H., Phan, K. A. T., et al. (2019). The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00750
- Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., et al. (2017). The impact of structural variation on human gene expression. *Nat. Genet.* 49, 692–699. doi: 10.1038/ng.3834
- Cho, C. H., Park, S. I., Huang, T.-Y., Lee, Y., Ciniglia, C., Yadavalli, H. C., et al. (2023). Genome-wide signatures of adaptation to extreme environments in red algae. *Nat. Commun.* 14, 10. doi: 10.1038/s41467-022-35566-x
- Chou, M.-X., Wei, X.-Y., Chen, D.-S., and Zhou, J.-C. (2007). A novel nodule-enhanced gene encoding a putative universal stress protein from *Astragalus sinicus*. *J. Plant Physiol.* 164, 764–772. doi: 10.1016/j.jplph.2006.05.009
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* 5, 113. doi: 10.1186/1471-2105-5-113
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y
- Foret, S., Seneca, F., de Jong, D., Bieller, A., Hemmrich, G., Augustin, R., et al. (2011). Phylogenomics reveals an anomalous distribution of USP genes in metazoans. *Mol. Biol. Evol.* 28, 153–161. doi: 10.1093/molbev/msq183
- Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., et al. (2006). Copy number variation: New insights in genome diversity. *Genome Res.* 16, 949–961. doi: 10.1101/gr.3677206
- Garrison, E., Guarracino, A., Heumos, S., Villani, F., Bao, Z., Tattini, L., et al. (2024). Building pangenome graphs. *Nat. Methods* 21, 2008–2012. doi: 10.1038/s41592-024-02430-3
- Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., et al. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* 36, 875–879. doi: 10.1038/nbt.4227
- Gonzali, S., Loreti, E., Cardarelli, F., Novi, G., Parlanti, S., Pucciariello, C., et al. (2015). Universal stress protein HRU1 mediates ROS homeostasis under anoxia. *Nat. Plants* 1, 1–9. doi: 10.1038/nplants.2015.151
- Groza, C., Schwendinger-Schreck, C., Cheung, W. A., Farrow, E. G., Thiffault, I., Lake, J., et al. (2024). Pangenome graphs improve the analysis of structural variants in rare genetic diseases. *Nat. Commun.* 15, 657. doi: 10.1038/s41467-024-44980-2
- Gu, Z. (2022). Complex heatmap visualization. *iMeta* 1, e43. doi: 10.1002/imt2.43
- Guarracino, A., Heumos, S., Nahnsen, S., Prins, P., and Garrison, E. (2022). ODGI: understanding pangenome graphs. *Bioinformatics* 38, 3319–3326. doi: 10.1093/bioinformatics/btac308
- Ijaz, R., Ejaz, J., Gao, S., Liu, T., Imtiaz, M., Ye, Z., et al. (2017). Overexpression of annexin gene *AnnSp2*, enhances drought and salt tolerance through modulation of ABA synthesis and scavenging ROS in tomato. *Sci. Rep.* 7, 12087. doi: 10.1038/s41598-017-11168-2
- Jamsandekar, M., Ferreira, M. S., Pettersson, M. E., Farrell, E. D., Davis, B. W., and Andersson, L. (2024). The origin and maintenance of supergenes contributing to ecological adaptation in Atlantic herring. *Nat. Commun.* 15, 9136. doi: 10.1038/s41467-024-53079-7
- Jayakodi, M., Lu, Q., Pidon, H., Rabanus-Wallace, M. T., Bayer, M., Lux, T., et al. (2024). Structural variation in the pangenome of wild and domesticated barley. *Nature* 636, 654–662. doi: 10.1038/s41586-024-08187-1
- Jung, Y. J., Melencion, S. M. B., Lee, E. S., Park, J. H., Alinapon, C. V., Oh, H. T., et al. (2015). Universal stress protein exhibits a redox-dependent chaperone function in arabidopsis and enhances plant tolerance to heat shock and oxidative stress. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01141
- Kang, H., Park, S. J., and Kwak, K. J. (2013). Plant RNA chaperones in stress response. *Trends Plant Sci.* 18, 100–106. doi: 10.1016/j.tplants.2012.08.004
- Kapazoglou, A., Gerakari, M., Lazaridi, E., Kleftogianni, K., Sarri, E., Tani, E., et al. (2023). Crop wild relatives: A valuable source of tolerance to various abiotic stresses. *Plants* 12, 328. doi: 10.3390/plants12020328
- Khan, A. W., Garg, V., Sun, S., Gupta, S., Dudchenko, O., Roorkiwal, M., et al. (2024). Cicer super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea. *Nat. Genet.* 56, 1225–1234. doi: 10.1038/s41588-024-01760-4
- Kumar, H., Panigrahi, M., Strillacci, M. G., Nayak, S. S., Rajawat, D., Ghildiyal, K., et al. (2023). Detection of genome-wide copy number variation in Murrah buffaloes. *Anim. Biotechnol.* 34 (8), 3783–3795. doi: 10.1080/10495398.2023.2227670
- Kvint, K., Nachin, L., Diez, A., and Nyström, T. (2003). The bacterial universal stress protein: function and regulation. *Curr. Opin. Microbiol.* 6, 140–145. doi: 10.1016/S1369-5274(03)00025-0
- Lenman, M., Sörensson, C., and Andreasson, E. (2008). Enrichment of phosphoproteins and phosphopeptide derivatization identify universal stress proteins in elicitor-treated arabidopsis. 21 (10), 1275–1284. doi: 10.1094/MPMI-21-10-1275
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi: 10.48550/arXiv.1303.3997
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, N., He, Q., Wang, J., Wang, B., Zhao, J., Huang, S., et al. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* 55, 852–860. doi: 10.1038/s41588-023-01340-y
- Li, N., Wang, Z., Wang, B., Wang, J., Xu, R., Yang, T., et al. (2022). Identification and characterization of long non-coding RNA in tomato roots under salt stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.834027
- Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

- Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226. doi: 10.1038/ng.3117
- Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., et al. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat. Genet.* 56, 2804–2814. doi: 10.1038/s41588-024-01967-5
- Loukehaich, R., Wang, T., Ouyang, B., Ziaf, K., Li, H., Zhang, J., et al. (2012). SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *J. Exp. Bot.* 63, 5593–5606. doi: 10.1093/jxb/ers220
- Maqbool, A., Zahur, M., Husnain, T., and Riazuddin, S. (2009). GUSP1 and GUSP2, two drought-responsive genes in gossypium arboreum have homology to universal stress proteins. *Plant Mol. Biol. Rep.* 27, 109–114. doi: 10.1007/s11105-008-0049-0
- Maqbool, A., Zahur, M., Irfan, M., Younas, M., Barozai, K., Rashid, B., et al. (2008). Identification and expression of six drought-responsive transcripts through differential display in desi cotton (*Gossypium arboreum*). *Mol. Biol.* 42, 492–498. doi: 10.1134/S002689330804002X
- Melencion, S. M. B., Chi, Y. H., Pham, T. T., Paeng, S. K., Wi, S. D., Lee, C., et al. (2017). RNA chaperone function of a universal stress protein in arabidopsis confers enhanced cold stress tolerance in plants. *Int. J. Mol. Sci.* 18, 2546. doi: 10.3390/ijms18122546
- Merkouropoulos, G., Andreasson, E., Hess, D., Boller, T., and Peck, S. C. (2008). An arabidopsis protein phosphorylated in response to microbial elicitation, atPHOS32, is a substrate of MAP kinases 3 and 6\*. *J. Biol. Chem.* 283, 10493–10499. doi: 10.1074/jbc.M800735200
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Nachin, L., Nannmark, U., and Nyström, T. (2005). Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J. Bacteriology* 187, 6265–6272. doi: 10.1128/jb.187.18.6265-6272.2005
- Pagani, F., and Baralle, F. E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.* 5, 389–396. doi: 10.1038/nrg1327
- Peralta, I. E., and Spooner, D. M. (2000). *Classification of wild tomatoes: a review*. Available online at: <https://ri.conicet.gov.ar/handle/11336/152176> (Accessed December 16, 2024).
- Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983–987. doi: 10.1038/nbt.4235
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). doi: 10.1093/nar/gky448
- Qi, T., He, F., Zhang, X., Wang, J., Zhang, Z., Jiang, H., et al. (2024). Genome-wide identification and expression profiling of potato (*Solanum tuberosum* L.) universal stress proteins reveal essential roles in mechanical damage and deoxynivalenol stress. *Int. J. Mol. Sci.* 25, 1341. doi: 10.3390/ijms25021341
- Sauter, M., Rzewuski, G., Marwedel, T., and Lorbiecke, R. (2002). The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *J. Exp. Bot.* 53, 2325–2331. doi: 10.1093/jxb/erf096
- Stalder, L., Oggenfuss, U., Mohd-Assaad, N., and Croll, D. (2023). The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Mol. Ecol.* 32, 2443–2460. doi: 10.1111/mec.16435
- Tkaczuk, K. L., Shumilin, A., Chruszcz, M., Evdokimova, E., Savchenko, A., and Minor, W. (2013). Structural and functional insight into the universal stress protein family. *Evolutionary Appl.* 6, 434–449. doi: 10.1111/eva.12057
- Udawat, P., Jha, R. K., Sinha, D., Mishra, A., and Jha, B. (2016). Overexpression of a cytosolic abiotic stress responsive universal stress protein (SbUSP) mitigates salt and osmotic stress in transgenic tobacco plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00518
- Vanbogelen, R. A., Hutton, M. E., and Neidhardt, F. C. (1990). Gene-Protein database of *Escherichia coli* K - 12: Edition 3. *Electrophoresis* 11, 1131–1166. doi: 10.1002/elps.1150111205
- Vollmer, A. C., and Bark, S. J. (2018). Twenty-five years of investigating the universal stress protein: function, structure, and applications. *Adv. Appl. Microbiol.* 105, 1–36.
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Yun, T., Li, H., Chang, P.-C., Lin, M. F., Carroll, A., and McLean, C. Y. (2021). Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics* 36, 5582–5589. doi: 10.1093/bioinformatics/btaa1081
- Zarrei, M., MacDonald, J. R., Merico, D., and Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat. Rev. Genet.* 16, 172–183. doi: 10.1038/nrg3871
- Zhang, Z. (2022). KaKs\_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002