Indian J. Genet. Plant Breed., (2025); 85(4): 619-628
https://doi.org/10.31742/ISGPB.85.4.10
ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Genome-wide identification, evolution and expression analysis of the HSP20 gene family in lentil [*Lens culinaris* (L.) Medikus]

Shivadarshan S. Jirli[$], Monendra Grover[1], Sharanbasappa D. Madival[1], Dwijesh Chandra Mishra[1], Krishna Kumar Chaturvedi[1], Shruti Sinha[2], Shashi Bhushan Lal[1], Venugopala Gowda R.[2], Lal Dhari Patel[1] and Amit Kumar Singh[2*]

Abstract

Lentil [*Lens culinaris* (L.) Medikus, 2n = 2x = 14] is one of the most important legumes around the world. The crop is sensitive to various abiotic stresses, including heat and drought. Among various classes of proteins, heat shock proteins (HSPs), play a vital role in regulating developmental processes and the responses to environmental stresses. Of these, the HSP20 family has gained attention however, its genomic organization and functions are uncharacterized in *Lens* species. Our study thus identified a total of 47 putative LcHSP20 genes, which were classified into five subfamilies based on predicted subcellular localization and phylogenetic relationships. Sequence analysis revealed that the majority of LcHSP20 genes possess either no introns or only a single short intron, indicating a streamlined gene structure. Phylogenetic clustering further demonstrated that members of the same subfamily exhibit conserved gene structures and motifs, suggesting potential functional similarities. Moreover, expressions analysis revealed that the transcript levels of LcHSP20 genes were significantly induced under salt stress condition. Notably, two genes, *LcHSP20-44* and *LcHSP20-13*, were markedly upregulated in response to these stressors, highlighting their potential roles in enhancing salt stress tolerance in lentil plants. This is the first genome-wide study of the HSP20 gene family in lentils, providing critical insights into their evolutionary and functional roles, laying the groundwork for future research on stress tolerance and improving lentil breeding programs.

Keywords: Abiotic stress, genome-wide analysis, gene duplication, lentil, stress adaptation, transcription factors

Introduction

Lentil [*Lens culinaris* (L.) Medikus] is an important legume crop, highly valued for its nutritional benefits and adaptability to various growing conditions. Its seeds have a rich nutritional profile, offering high levels of protein, dietary fiber, and essential micronutrients such as folate, iron, and zinc (Foodstruct 2023). The crop is grown in around 70 countries and is a staple pulse crop in many regions, particularly in South Asia, the Middle East, and North Africa, where it plays a critical role in combating nutritional deficiencies (Salaria et al. 2022). Canada currently leads the world production of lentil with a 34.72% share, followed by India (18.60%) and Australia (14.65%) (FAOSTAT 2022). Furthermore, regular consumption of lentils has been linked to reduced risks of chronic diseases, such as cardiovascular conditions, diabetes, and metabolic disorders (Singh et al. 2019).

Despite their significance, lentil cultivation is often hampered by various biotic and abiotic stresses. Biotic stresses include pests and diseases (Rubio Teso et al. 2022; Barilli and Rubiales 2023), while abiotic stresses such as drought, salinity, and extreme temperatures can severely limit crop yield (Sehgal et al. 2017; Noor et al. 2024).

The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India

[$,1]Present address: ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

[2]ICAR-National Bureau of Plant Genetic Resources, New Delhi 110 012, India

***Corresponding Author:** Amit Kumar Singh, ICAR-National Bureau of Plant Genetic Resources, New Delhi 110 012, India, E-Mail: amit.singh5@icar.gov.in

How to cite this article: Jirli S.S., Grover M., Madival S.D., Mishra D.C., Chaturved K.K., Sinha S., Lal S.B., Venugopala G.R., Patel L. and Singh A.K. 2025. Genome-wide identification, evolution and expression analysis of the HSP20 gene family in lentil [*Lens culinaris* (L.) Medikus] . Indian J. Genet. Plant Breed., **85**(4): 619-628.

Source of support: Nil

Conflict of interest: None.

Received: Sept. 2025 **Revised:** Nov. 2025 **Accepted:** Nov. 2025

Addressing these challenges is essential to ensuring stable global lentil production, as the crop plays a key role in food security. To cope with these stresses, plants have evolved a range of defense mechanisms, including morphological, molecular, and physiological adaptations (Yang et al. 2021). Among these molecular defenses, heat shock proteins (HSPs), particularly HSP20s, are vital for protecting plants against both biotic and abiotic stressors such as temperature extremes, salinity, drought, heavy metal toxicity, and pathogen attacks (Zhang et al. 2024).

Heat shock proteins (HSPs) are a diverse family of molecular chaperones that are rapidly induced under various abiotic stresses such as high temperature, drought, and salt stress, as well as under biotic stresses like pathogen attack. They play a crucial role in maintaining cellular protein homeostasis by assisting in proper folding, refolding of misfolded proteins, and preventing aggregation (Charng et al. 2006). Based on their molecular weight and sequence homology, HSPs are generally classified into several major families, including HSP100, HSP90, HSP70, HSP60, and HSP20 (Ji et al. 2019). Among these, HSP20s classified as small heat shock proteins (sHSPs) serve as molecular chaperones that stabilize partially folded or denatured proteins, preventing their irreversible aggregation (Eyles and Gierasch 2010). These proteins contain a highly conserved α-crystallin domain (ACD) in the C-terminal region, responsible for substrate interactions and oligomerization. The N-terminal region assists in substrate binding, while the short C-terminal extension facilitates protein-protein interactions (Giese and Vierling 2004; Jaya, Garcia and Vierling 2009). Structurally, the ACD is composed of two hydrophobic β-strand structures separated by a variable-length α-helical region (Bondino et al.2012). Not all ACD-containing proteins are classified as HSP20s, as some perform diverse functions beyond stress tolerance.

Salt stress disrupts cellular homeostasis by causing protein misfolding and oxidative damage. HSP20s bind to misfolded proteins, preventing their aggregation, which stabilizes cellular functions, allowing the plant to better cope with osmotic stress and ion toxicity (Guo et al. 2020). The upregulation of HSP20s during salt stress supports plant resilience, allowing for sustained growth and development in saline environments (Gupta et al. 2010). Understanding the molecular functions of HSP20s in salt stress response is essential for breeding salt-tolerant lentil varieties.

Although HSP20 proteins have been widely studied in model plants like *Arabidopsis thaliana* (Huang et al. 2023), as well as major crops such as rice (Ouyang et al. 2009), wheat (Muthusamy et al. 2017), potato (Zhao et al. 2018), Chickpea (Liu et al. 2024), and soybean (Lopes et al. 2013), the HSP20 gene family remains largely unexplored in lentil. To date, no comprehensive genome-wide study has been conducted to examine the HSP20 gene family in *L. culinaris*.

To address this gap, a study was carried out on the first comprehensive genome-wide identification and characterization of the HSP20 gene family in *Lens culinaris*. The study was also aimed at the systematic identification of HSP20 genes, analyzing their structural features and classifying them based on their phylogenetic relationships. The cis-regulatory elements in the promoter region of the genes were also identified and their potential roles in salt stress tolerance in *Lens* spp. were explored through analysis of expression data. Additionally, synteny and collinearity analyses between lentils and *A. thaliana* provided valuable insights into the evolutionary conservation and expansion of the HSP20 gene family in lentils.

Materials and methods

Genome sequence data retrieval

The genome sequence of *L. culinaris* Medikus was sourced from the KnowPulse database (https://knowpulse.usask.ca/bio_data/2690904), a curated genomic resource for pulse crops maintained by the University of Saskatchewan (Ramsay et al. 2021). This database offers valuable genomic data, and the genome used in this study was retrieved in FASTA format for subsequent bioinformatic analysis.

Identification of HSP20 family members

Members of the HSP20 gene family were identified using a Hidden Markov Model (HMM) profile specific to the alpha-crystallin domain (PF00011) from the Pfam database (Finn et al. 2011). HMMER3.0 software was employed to search for sequences containing the alpha-crystallin domain in the lentil genome. After initial identification, redundant sequences were removed and all hits were screened to confirm the presence of family-specific alpha-crystallin domain using the Simple Modular Architecture Research Tool (SMART). The final validated HSP20 genes were named according to their chromosomal locations (e.g., *LcHSP20-1*, *LcHSP20-2*, etc.), ensuring clear identification for future analysis.

Multiple sequence alignment and phylogenetic analysis

The full-length amino acid sequences of the identified LcHSP20 proteins were aligned using the MUSCLE algorithm. A phylogenetic tree was then constructed using MEGA-X software (Kumar et al. 2018) by applying the Neighbor-Joining method (Saitou et al. 1987), with 1,000 bootstrap replicates to assess the robustness and reliability of the tree. To analyze the evolutionary relationships among LcHSP20 genes and their orthologs from *A. thaliana* and *Glycine max*, the Interactive Tree of Life (iTOL) v5.0 was employed, allowing for a clear visualization of phylogenetic connections and insights into the evolutionary history of HSP20 genes across species.

Gene structure and motif analysis

The exon-intron structures of the LcHSP20 genes were analyzed using the Gene Structure Display Server (GSDS 2.0) (Hu et al. 2015) by comparing the coding sequences with their corresponding genomic sequences to determine the number and arrangement of exons and introns. To identify conserved regions within the LcHSP20 proteins, the Multiple Expectation Maximization for Motif Elicitation (MEME) suite (Bailey et al. 2015) was utilized, with the maximum number of motifs set to 10. The predicted motifs were then visualized using TBtools software (Chen et al. 2020) providing insights into the conserved functional domains across the LcHSP20 family and their potential roles in stress responses.

Chromosomal location of LcHSP20 genes

The chromosomal mapping of HSP20 genes was carried out using MG2c_v2.1 (Chao et al. 2021), an advanced platform for constructing genome-wide visualization plots. This approach enabled the generation of high-resolution maps illustrating the distribution of HSP20 genes across the seven chromosomes of lentil. The physical positions of individual genes were displayed, highlighting clusters in regions with elevated gene density, which may reflect potential gene duplication events or evolutionary hot spots.

Promoter sequence and cis-acting element analysis

To explore the regulation of LcHSP20 genes, the 2000 bp upstream promoter regions of each gene were extracted from the lentil genome. These promoter sequences were analyzed for the presence and composition of cis-acting regulatory elements using the PlantCARE database (Lescot et al. 2002) and the results were then visualized using TBtools(Chen et al. 2020). Key cis-elements associated with hormone responses, such as abscisic acid (ABA), salicylic acid (SA), gibberellin (GA), and methyl jasmonate (MeJA), as well as elements linked to abiotic stress factors like drought, light, and low temperature, were identified.

Physicochemical properties and subcellular localization

The physicochemical properties of the LcHSP20 proteins, such as molecular weight and isoelectric point (pI), were determined using the NaturePred tool (Sharanbasappa et al. 2024). Additionally, the subcellular localization of the LcHSP20 proteins was predicted using WoLFPSORT (Horton et al. 2007), which analyzes amino acid sequences to predict the likely cellular compartments where these proteins function.

Synteny and collinearity analysis

Synteny and collinearity between the *L. culinaris* HSP20 genes and those from *A. thaliana* were analyzed using a multi-step approach. First, HSP20 protein sequences from both lentil and *Arabidopsis* were aligned using Diamond blastp to identify potential orthologous genes. The blastp search was performed in both directions: *L. culinaris* HSP20 proteins were used as the query against the *A. thaliana* HSP20 proteome, and *Arabidopsis* HSP20 proteins were used as the query against the *Lens culinaris* proteome. This bi-directional approach was adopted to identify reversible best hits (RBH), which represent the most likely orthologous genes between the two species.

Once the RBH pairs were identified, the corresponding genomic coordinates were extracted from the gff3 files of both species using rtracklayer (Lawrence et al. 2009) and Genomic Ranges (Lawrence et al. 2013) packages. This step laid the foundation for further analysis of the conserved syntenic relationships and collinearity between the HSP20 genes of lentil and *Arabidopsis*.

Circos plot and duplication analysis

To investigate tandem and segmental duplications within the lentil genome, we employed both phylogenetic tree analysis and sequence similarity approaches. For sequence similarity, we conducted a *DIAMOND* BLASTp analysis, comparing each gene against all other genes, while for the phylogenetic tree, we utilized clade information to differentiate between segmental and tandem duplications. Once these duplication events were identified, we visualized them using a Circos plot created with the *circlize* package in R (Gu et al. 2014). This approach allowed us to map intra- and inter-chromosomal duplications, with segmental duplications represented by syntenic links connecting conserved regions across different chromosomes, and tandem duplications mapped within individual chromosomes. By integrating these analytical and visualization methods, we gained a comprehensive understanding of gene expansion, genomic architecture, and the evolutionary dynamics of gene families in lentils.

Expression profiling of LcHSP20 genes under salt conditions

The RNA-seq raw data for salt-treated and control conditions in lentil (*Lens culinaris*) (https://www.ncbi.nlm.nih.gov/bioproject/?term=SRR23032117 and SRR23032118) were processed following a systematic approach. Initially, the raw sequencing data were downloaded using SRA Toolkit, and adapters were removed using FastQC (Andrews et al. 2010) and trimmed using Trimmomatic (Bolger et al. 2014) to ensure high-quality reads. The cleaned reads were then aligned to the lentil reference genome using Bowtie2-build for accurate mapping (Langdon et al. 2015). After alignment, gene expression quantification was performed using StringTie (Pertea et al. 2015) to obtain FPKM and TPM values specifically for the HSP20 gene family, these values were log2-transformed for comparative analysis.

For downstream analysis, we calculated fold changes between salt-treated and control conditions to identify upregulated and downregulated HSP20 genes. We visualized

the results using heatmaps. The heatmap was generated using Seaborn, representing the expression levels of HSP20 genes across the two conditions, highlighting differentially expressed genes. This comprehensive workflow enabled the identification and visualization of HSP20 gene expression patterns in response to salt stress in lentil (Supplementary Fig. 1).

Results

Identification of HSP20 genes in lentil

A total of 47 HSP20 genes were identified in the *L. culinaris* genome through a search using HMMER3.0, which matched the Hidden Markov Model (HMM) profile of the alpha-crystallin domain (PF00011) against the Pfam database. The presence of the alpha-crystallin domain in these sequences was further validated using the SMART tool, ensuring the accuracy of the identified genes. These 47 non-redundant HSP20 genes were named *LcHSP20-1* to *LcHSP20-47* according to their chromosomal locations. The identified genes showed diverse lengths, ranging from 435 bp to 2334 bp, while the corresponding proteins varied from 78 amino acids (*LcHSP20-9*) to 573 amino acids (*LcHSP20-29*), highlighting the structural diversity of the HSP20 gene family in lentils.

Phylogenetic analysis, classification, and multiple sequence alignment of LcHSP20 proteins

The phylogenetic analysis of the LcHSP20 gene family reveals distinct clustering of *L. culinaris* HSP20 proteins with those from *A. thaliana* (19) and *Glycine max* (44), highlighting both evolutionary conservation and divergence. The LcHSP20 proteins are distributed across 19 distinct subfamilies (Fig. 1), which indicates a mix of shared and specialized functions. Clusters or Subfamilies such as A (11), F (4), G (6), I (5), and K (9) share members from *Lens* and either *Arabidopsis* or *Glycine* or both, indicating cross-species grouping. This suggests these genes are evolutionarily conserved and likely share similar structural features.

Their co-clustering reflects functional conservation, possibly in stress response mechanisms. Such correlation implies they may be orthologous genes retaining ancestral functions across species, whereas subfamilies B (2), C (13), D (6), and E (6) comprise exclusively of lentil HSP20 genes that suggest potential functional diversification, possibly in response to environmental stresses.

Notably, genes like *LcHSP20-4*, *LcHSP20-9*, *LcHSP20-12*, and *LcHSP20-70* exhibit evidence of gene duplication events, a critical evolutionary process that often drives functional diversification. These duplications may have allowed lentils to adapt to a range of environmental conditions, enhancing their resilience to both biotic and abiotic stresses. Overall, the expansion and diversification of the LcHSP20 gene family likely contributed to lentil's ecological success, enabling it to thrive in diverse environments. This analysis provides a foundation for further studies into the functional roles of LcHSP20 genes in stress tolerance.

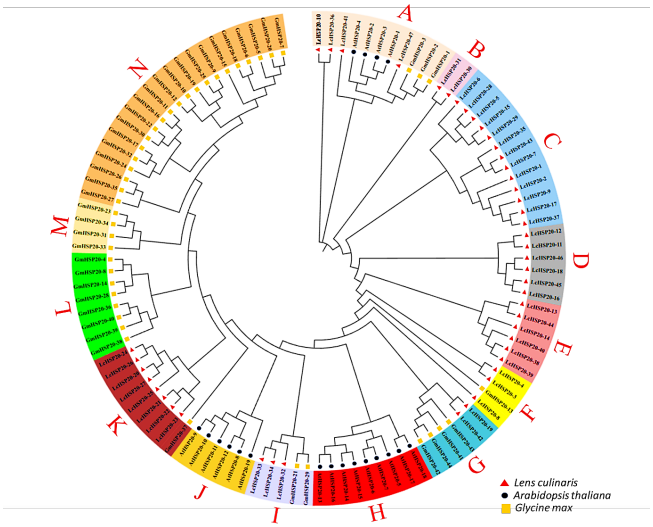


Fig. 1. Phylogenetic tree of LcHSP20 proteins, *A. thaliana* HSP20 proteins and *Glycine max* HSP20 proteins, constructed using the Neighbor-Joining method with 1,000 bootstrap replicates

Gene structure and motif analysis

To investigate the structural composition of the LcHSP20 genes, a detailed gene structure diagram was generated using the GSDS 2.0 server, based on the lentil genomic sequence (Fig. 2). Fig. 2 illustrates the essential genetic elements, including coding sequences (CDS), untranslated regions (UTRs) and introns. Gene architecture helps to identify the regulatory regions of the genes that influence their function, particularly in relation to transcriptional regulation, stress responses, and developmental processes. The comparative analysis of the exon and intron structures of the 47 LcHSP20 genes revealed significant variability, with exon numbers ranging from 1 to 13 across the gene family. 21 genes (44.68%), namely, *LcHSP20-12*, *LcHSP20-13*, *LcHSP20-14,* etc., are intronless, 18 genes (38.29%) possessed one intron and 2 genes (4.2%) contain 2 introns. This suggests an evolutionary simplification that enhances transcriptional efficiency. Seven genes (14.8%), namely *LcHSP20-23*, *LcHSP20-24*, *LcHSP20-25*, *LcHSP20-26*, *LcHSP20-27*, *LcHSP20-38*, and *LcHSP20-41,* lacked UTRs, which could have implications for post-transcriptional regulation and gene expression control.

This comprehensive structural analysis highlights the genetic diversity within the LcHSP20 gene family, reflecting their potential specialization in various biological functions and stress responses in lentils. The variability in exon-intron structures suggests evolutionary adaptation, possibly allowing lentils to better cope with diverse environmental

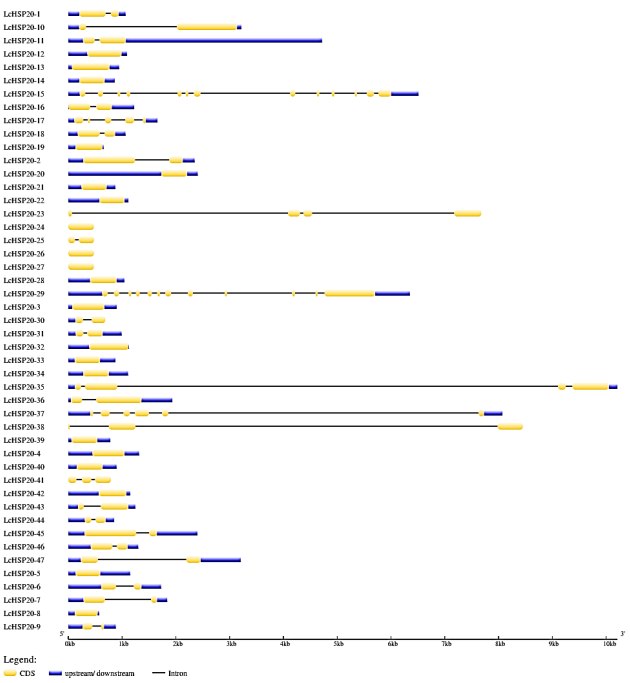


Fig. 2. Exon-intron structure of LcHSP20 genes, with intron lengths and positions represented proportionally for each gene

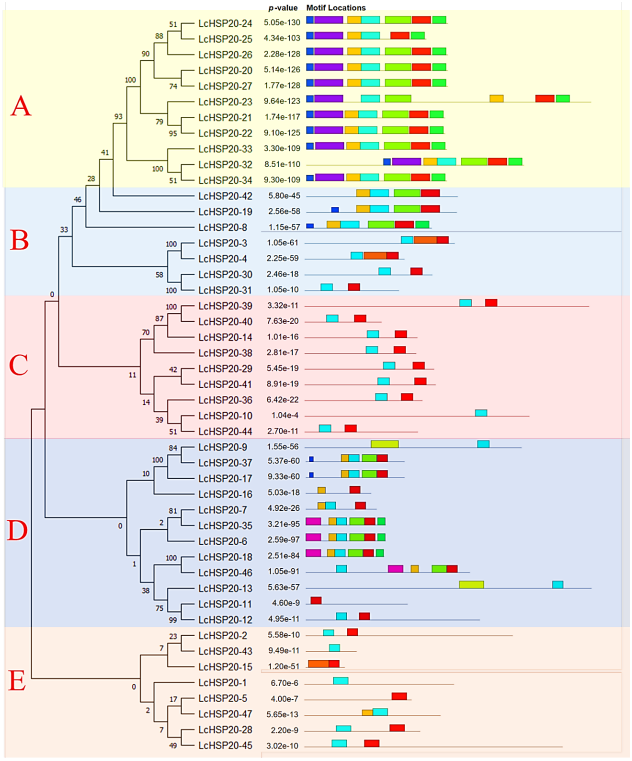


Fig. 3. Distribution of conserved motifs in LcHSP20 proteins, grouped based on their phylogenetic relationships

challenges.

The 10 motifs were identified in HSP20 gene family, with motif lengths ranging from 8 to 49 amino acids. Motif 10 was the longest (49 AAs), while motif 7 was the shortest (8 AAs) (Fig. 3, Supplementary Fig. 2). Each HSP20 gene contained 1 to 7 conserved motifs. Genes such as *LcHSP20-1*, *LcHSP20-10*, *LcHSP20-11*, and *LcHSP20-43* contained only a single motif, indicating structural divergence likely due to evolutionary processes.

Subfamily A includes motifs 1 to 7, associated with transcriptional regulation and stress responses. This suggests broader regulatory roles in response to environmental stressors. Subfamilies B and D feature additional motifs 8, 9, and 10, indicating more complex roles in stress adaptation, signal transduction, and environmental responses. Subfamily C (*LcHSP20-39*–*LcHSP20-44*) has fewer motifs, suggesting a more specialized function, possibly linked to targeted stress responses or developmental processes.

Overall, the diversity in motif compositions among LcHSP20 subfamilies reflects functional divergence, with certain subfamilies evolving unique roles in stress response, signal transduction, and development, enhancing lentil resilience to environmental challenges.

Promoter sequence analyses of LcHSP20 genes in lentil

To investigate potential cis-regulatory elements in the promoter regions of LcHSP20 genes, 2 kb upstream sequences were analyzed using the PlantCARE tool. The analysis identified Core promoter elements located around 30 bp upstream of the transcription start site, along with stress-related cis-elements for light responsiveness, low-temperature response, and anoxic stress. Hormone-responsive elements, including those for abscisic acid (ABA) and methyl jasmonate (MeJA), were also found, suggesting LcHSP20 genes are involved in regulatory pathways that mediate stress responses.

Promoter analysis of seven selected LcHSP20 genes from five phylogenetic clades revealed unique cis-element enrichment patterns (Supplementary Fig. 3). For example, *LcHSP20-20* showed high levels of CPE and MeJARE elements, linking it to jasmonate-mediated stress responses. *LcHSP20-4* had a significant presence of light-responsive elements (LRE), indicating its role in light-dependent processes, while genes like *LcHSP20-29* and *LcHSP20-7* displayed lower enrichment, reflecting the diversity of regulatory mechanisms within the family.

Further cis-element distribution analysis revealed abundant stress and hormone-responsive motifs across all LcHSP20 promoters (Supplementary Fig. 4). Genes like *LcHSP20-20*, *LcHSP20-22*, and *LcHSP20-28* contained dense motif clusters, indicating strong regulatory potential. Enriched motifs like MYB, ABRE, and STRE point to their key roles in abiotic stress responses, positioning certain LcHSP20 genes as candidates for improving stress tolerance in lentil.

Chromosomal distribution of LcHSP20 genes

The 47 LcHSP20 genes are unevenly distributed across seven lentil chromosomes and the unplaced "unitag" scaffold (Fig. 4). Chromosome 5 contains the most genes (14 genes, 29.78%), followed by Chromosome 4 (12 genes, 25.53%), Chromosome 3 (7 genes, 14.89%), Chromosome 1 (5 genes, 10.63%), Chromosome 2 (3 genes, 6.38%), Chromosomes 6 and 7 (2 genes, 4.25% each), and the unitag scaffold (2 genes, 4.26%).

This distribution suggests Chromosome 5 may be a hotspot for gene duplication and diversification, potentially playing a key role in stress tolerance, especially heat stress. The presence of two genes on the unitag scaffold may indicate specialized roles or recent duplication events. The distribution pattern reflects evolutionary dynamics contributing to the functional diversity of the LcHSP20 gene family. These findings underscore the role of gene duplication in lentil's stress adaptation. Future research should focus on genes on Chromosome 5 and the unitag scaffold to better understand their role in stress response and explore potential for crop improvement and stress tolerance breeding.

Primary structure and subcellular localization of LcHSP20 proteins

The physicochemical properties of the LcHSP20 gene family proteins were analyzed using the NaturePred tool, revealing significant diversity. The protein lengths ranged from 78 to 573 amino acids, molecular weights varied from 8.7 kDa (*LcHSP20-9*) to 63.5 kDa (*LcHSP20-29*), and isoelectric points (pI) ranged from 4.84 to 10.6 (Supplementary Table S1). These variations suggest that LcHSP20 proteins may have specialized functions in stress signalling, transcriptional regulation, and other cellular processes.

Subcellular localization predictions using the WoLF PSORT tool showed that most LcHSP20 proteins are localized in the cytoplasm, supporting their role as molecular chaperones in stress conditions like heat and salinity. However, a subset of proteins, including *LcHSP20-5*, *LcHSP20-10*, *LcHSP20-11*, *LcHSP20-12*, *LcHSP20-13*, *LcHSP20-45*, and *LcHSP20-46*, were predicted to localize in the chloroplast (Supplementary Table S1). This suggests their involvement in photosynthesis regulation and light-mediated stress responses, highlighting the broad functional roles of the LcHSP20 gene family in lentil's environmental adaptation.

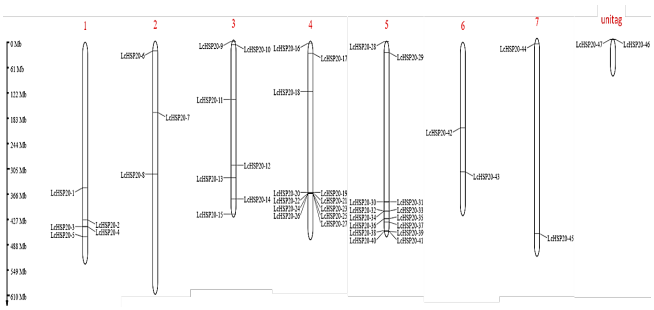


Fig. 4. Chromosomal distribution of 47 LcHSP20 genes across the seven lentil chromosomes, highlighting gene density and evolutionary diversity

Duplication and collinearity analysis of LcHSP20s

Gene duplication events, particularly in HSP20 genes, have been widely recognized as essential for the evolution of new biological functions and adaptation (Cannon et al. 2004; Nei and Rooney 2005). In this study, we utilized both molecular phylogenetic analysis and sequence similarity percentages to investigate the duplication patterns of LcHSP20 genes. Our findings revealed that 26 LcHSP20 genes underwent duplication, with a combination of segmental and tandem duplication events (Fig. 5). Specifically, 15 genes showed segmental duplication, while 11 exhibited tandem duplications,suggesting that segmental duplication played a key role in the evolutionary development of these genes. The integration of phylogenetic relationships and sequence similarity helped distinguish between these duplication types and offered a comprehensive understanding of LcHSP20 gene evolution.

The genome-wide co-llinearity analysis between *L. culinaris* and *A. thaliana* identified 18 collinear gene pairs related to the HSP20 gene family across Chromosomes 1 to 5 in both species (Fig.6, Supplementary Table S2). However, no collinear gene pairs were found on Chromosomes 6

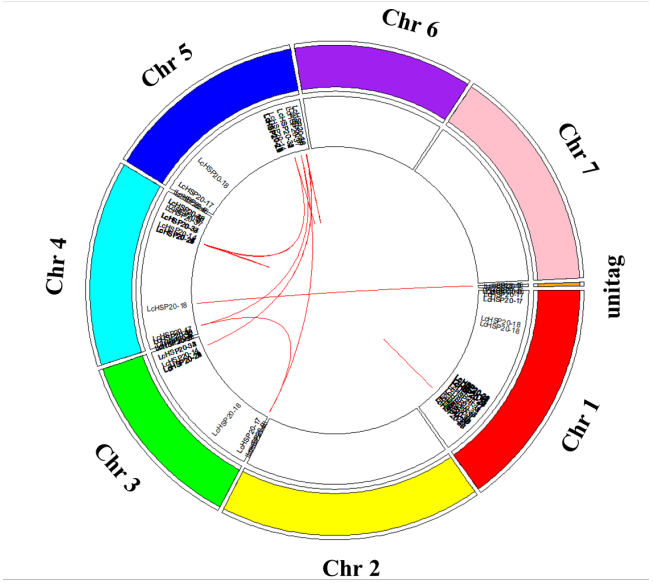


Fig. 5. The distribution of duplicated (Tandem and Segmental) LcHSP20 genes across 7 chromosomes

and 7 in lentil. This indicates that the HSP20 genes on these chromosomes may have undergone independent evolutionary changes or rearrangements, diverging from their counterparts in *Arabidopsis*.

The presence of collinear pairs on Chr1 to Chr5 emphasizes the conserved evolutionary importance of the HSP20 genes in both species. The absence of collinearity on Chr6 and Chr7 suggests that these regions might have experienced different evolutionary pressures, leading to gene divergence.

This pattern of collinearity highlights the evolutionary conservation and functional significance of the HSP20 gene family in lentil and *Arabidopsis*, with potential adaptive roles in abiotic stress tolerance.

Differential expression of HSP20 genes in lentil under salt stress

The RNA-seq analysis of lentil under salt stress highlighted the significant role of the HSP20 gene family in the plant's adaptive response. Nine HSP20 genes were upregulated, with fold changes ranging from 1.28 to 4.56, indicating their involvement in activating stress response pathways. Differential expression was assessed using FPKM/TPM values from both control and salt-treated samples, visualized in a Volcano plot and heat map.

The Volcano plot shows the differential expression, with the x-axis representing log2 fold change (logFC) between salt-treated and control samples, and the y-axis showing the -log10 adjusted p-value (Supplementary Fig. 5). Red-marked genes (logFC> 1) like *LcHSP20-44* and *LcHSP20-13* indicate significant upregulation, suggesting their key roles in defense mechanisms. Conversely, blue-marked downregulated genes (logFC< -1), such as *LcHSP20-42* and *LcHSP20-38*, may be suppressed to conserve energy under stress.

The heatmap illustrates expression profiles of the top 20 differentially expressed HSP20 genes (Fig. 7). Two distinct

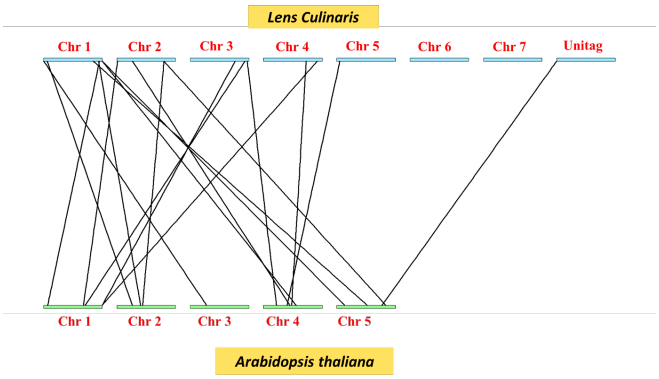


Fig. 6. Collinearity analysis of HSP20 genes between *Lens culinaris* and *Arabidopsis thaliana*

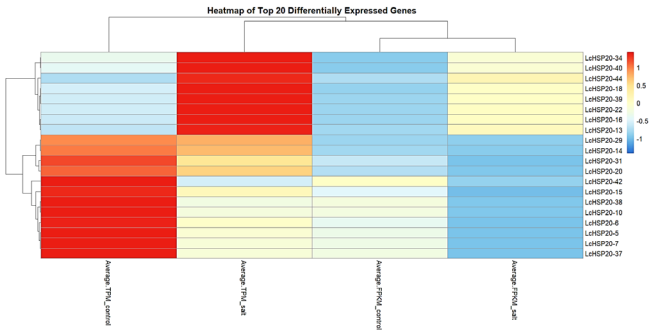


Fig. 7. Heat map of LcHSP20 genes under salt vs control condition

clusters are observed based on their response to salt stress. One cluster, including *LcHSP20-10*, *LcHSP20-6*, and *LcHSP20-5*, shows high expression in control conditions but significant downregulation under salt stress, indicating their roles in normal physiological processes. In contrast, genes like *LcHSP20-34*, *LcHSP20-40*, and *LcHSP20-44* exhibit increased expression under salt stress, suggesting their involvement in stress tolerance mechanisms, possibly via activation of pathways like abscisic acid (ABA).

The heatmap shows a clear separation between control and salt-treated samples, with tight clustering of the salt-treated samples, indicating a consistent transcriptional response to salt stress. This reinforces the critical role of these genes in the plant's adaptation to high salinity. Overall, the differential expression of the HSP20 gene family in lentil highlights strategic modulation in response to salt stress, downregulating non-essential functions while upregulating stress-related genes to enhance survival. These findings provide insights into the functional significance of HSP20 genes in stress responses, offering opportunities for developing salt-tolerant crops.

Discussion

Numerous studies have shown that HSP20s are the most abundant family of heat shock proteins (HSPs), acting as molecular chaperones in plants responses to abiotic stress (Waters et al. 2020). With the completion of lentil (*L. culinaris* Medikus) genome sequencing, we were able to conduct a comprehensive analysis of the HSP20 gene family, identifying and characterizing 47 members for the first time (Supplementary Table S1). This number exceeds those reported in other species such as Coix (32) (Hua et al. 2023), *Arabidopsis* (19) (Scharf et al. 2001), Chickpea (21) (Liu et al. 2024), rice (39) (Ouyang et al. 2009), tomato (42) (Yu et al. 2016), and maize (44) (Qi et al. 2022). However, it is slightly fewer than in grape (48) (Ji et al. 2019) and soybean (51) (Lopes-Caitar et. 2013), and significantly lower than in cotton (94) (Ma et al. 2016). Despite lentil's large genome size (~3.69 Gb) approximately three times that of Coix (1.2 Gb) (Hua et al. 2023), 14.19 times that of *Arabidopsis* (260 Mb), 9.4 times that of rice (389 Mb), and 1.75 times that of maize (B73,

2.106 Gb) the number of HSP20 genes is not proportional to its genome size. This suggests a unique evolutionary trajectory for lentil's HSP20 gene family.

The phylogenetic analysis of LcHSP20 genes in lentil revealed a dynamic evolution, with genes distributed across 9 subfamilies. This distribution mirrors the diversification observed in other plants, such as *A. thaliana*, where some subfamilies are either absent or expanded (Scharf et al. 2001). For instance, *Arabidopsis* lacks the VI and CVII subfamilies (Scharf et al. 2001), while rice and dove tree show the absence of CIV and CVII or CII and CIII subfamilies, respectively (Ouyang et al. 2009; Hou et al. 2022). These patterns suggest that lentil has undergone subfamily-specific expansions and functional diversification (B (2), C (13), D (6), and E (6) clusters), likely driven by environmental selection pressures. Additionally, the clustering of LcHSP20 genes into stress-responsive subfamilies reflects findings from other species, where subfamilies like CI are predominantly involved in heat stress responses, while others exhibit specialized roles.

The intron-exon structure analysis of the LcHSP20 genes provided further insights into their evolutionary and functional differentiation. We observed considerable variation in exon-intron structure, with exon numbers ranging from 1 to 13. Notably, 21 genes, including *LcHSP20-12*, *LcHSP20-13*, and *LcHSP20-14* etc. were intronless, a characteristic that may enhance transcriptional efficiency, especially under stress conditions. This finding aligns with studies in *Arabidopsis*, where genes with fewer introns were shown to respond more rapidly to abiotic stresses (Yan et al. 2017). The exon-intron variation across subfamilies likely reflects adaptive responses to specific stressors.

Additionally, a conserved motif analysis identified 10 conserved motifs across the LcHSP20 proteins, with the alpha-crystallin motif being present in most genes. The absence of this motif in a few genes, such as *LcHSP20-1*, *LcHSP20-9*, and *LcHSP20-13*, suggests functional divergence within the gene family. Similar losses of conserved motifs have been observed in other species, like rice and *Arabidopsis*, where some HSP20 genes exhibit specialized functions beyond chaperone activity (Jeffares et al. 2008). The motif analysis further indicated that genes in different subfamilies possess distinct motif compositions, highlighting their involvement in broader regulatory roles, including stress adaptation and defense mechanisms. These structural and motif variations suggest that LcHSP20 genes perform diverse functions, especially in response to abiotic stresses.

Similar cis-regulatory elements in HSP20 genes, such as those for heat, light, and hormonal regulation, have been found in other crops like rice (Ouyang et al. 2009) and *Arabidopsis* (Scharf et al. 2001), underscoring the conserved role of HSP20 genes in stress adaptation across species.

Subcellular localization prediction revealed that the LcHSP20 proteins are localized across multiple cellular compartments, including the cytoplasm, nucleus, chloroplast, mitochondria, extracellular region,and peroxisome. Notably, no LcHSP20 proteins were found in the endoplasmic reticulum (ER), in contrast to species like Arabidopsis, rice, and maize, which contain ER-localized HSP20s (Qi et al. 2022). This absence of ER-localized HSP20s aligns with findings in Coix (Hua et al. 2023), which also lacks ER-localized HSP20s, further suggesting evolutionary diversification within the gene family. The variation in subcellular localization of LcHSP20 proteins indicates their involvement in a variety of functions, potentially responding to diverse environmental stresses.

Gene duplication, including tandem and segmental duplications, appears to be a key mechanism in the expansion of the LcHSP20 gene family. In lentil, we identified several duplicated gene pairs, such as *LcHSP20-4/LcHSP20-9* (tandem duplication) and *LcHSP20-12/LcHSP20-70* (segmental duplication), etc. These duplications may have contributed to the functional diversification of the LcHSP20 family, enabling lentil to adapt to a range of environmental stresses. Similar patterns of tandem and segmental duplications have been observed in species like rice (Ouyang et al. 2009), maize (Qi et al. 2022), Chickpea 21 (Liu et al. 2024), and dove tree (Hou et al. 2022), where such duplications often lead to conserved or differentiated expression profiles under stress conditions. In lentil, closely related genes within subfamilies A and B exhibit conserved functions, while subfamilies E and D show evidence of functional specialization (Fig. 3). The differential expression patterns of duplicated genes, such as the divergence between *LcHSP20-12* and *LcHSP20-70*, suggest that these genes may have evolved distinct functions, allowing lentil to respond to various environmental challenges.

Expression profiling of LcHSP20 genes under salt stress revealed significant changes in gene expression, confirming their involvement in stress responses. Genes such as *LcHSP20-44* and *LcHSP20-13* were upregulated, with fold changes ranging from 1.28 to 4.56, similar to the findings in soybean and potato, where HSP20 genes are also induced under stress (Lopes-Caitar et al. 2013; Zhao et al. 2018). Conversely, genes like *LcHSP20-42* and *LcHSP20-38* were downregulated, suggesting a role in non-essential processes suppressed during stress, aligning with findings in *Arabidopsis* and sweet pepper (Gonzalez-Gordo et al. 2023).

Overall expression patterns indicate that LcHSP20 genes are modulated in response to salt stresses in lentil, with upregulation of stress-responsive genes (*LcHSP20-34*, *LcHSP20-40*) and downregulation of non-essential genes (*LcHSP20-10*, *LcHSP20-6*) to enhance stress survival. Our findings reinforce the importance of HSP20 genes in plant stress tolerance, as seen in other species like soybean, *Arabidopsis*, and sweet pepper. The present study provides

valuable insights into the LcHSP20 gene family in lentil, highlighting their role in stress tolerance and their potential for improving lentil resilience through breeding programs.

Supplementary materials

Supplementary Tables 1 and 2 and supplementary Figures 1 to 5 are provided, which can be accessed at www.isgpb.org.

Authors' contribution

Conceptualization of research (SSJ, AKS, SDM); Designing of the experiments (SSJ, AKS, MG); Contribution of experimental materials(SSJ, AKS, DCM); Execution of field/ lab experiments and data collection (AKS, DCM, MG);Analysis of data and interpretation (SSJ, SDM, KKC, SBL); Preparation of the manuscript (SSJ, VG, LP, SS).

Acknowledgments

This work was supported by the Department of Biotechnology, Government of India, New Delhi under the project No.BT/PR40161/BTIS/137/32/2021 and No.BT/PR40191/ BTIS/137/82/2023. First author gratefully acknowledges the support given by the ICAR-Indian Agricultural Statistics Research Institutes' through fellowship. The authors are thankful to the Director ICAR-IASRI, New Delhi, and the Indian Council of Agricultural Research, Department of Agricultural Research and Education, Government of India, for providing infrastructural support in terms of the supercomputing facility at ICAR-IASRI, New Delhi.

References

Salaria S., Boatwright J. L.,Thavarajah P., Kumar S. and Thavarajah D. 2022. Protein biofortification in lentils (*Lens culinaris* Medik.) toward human health. Front. Plant Sci., **13**: 869713.

FAOSTAT2022. World Food and Agriculture-Statistical Year Book https://www.fao.org/faostat/en/#data/QCL

Singh D., Singh C. K., Taunk J., Jadon V., Pal M. and Gaikwad K. 2019. Genome wide transcriptome analysis reveals vital role of heat responsive genes in regulatory mechanisms of lentil (*Lens culinaris* Medikus). Sci. Rep., **9**(1): 12976.

Rubio Teso M.L., Lara-Romero C., Rubiales D., Parra-Quijano M. and Iriondo J. M. 2022 Searching for abiotic tolerant and biotic stress resistant wild lentils for introgression breeding through predictive characterization. Front. Plant Sci., **13**: 817849. doi: 10.3389/fpls.2022.817849.

Barilli E. and Rubiales D. 2023. Identification and characterization of resistance to rust in lentil and its wild relatives. Plants, **12**(3): 626. https://doi.org/10.3390/plants12030626.

Sehgal A., Sita K., Kumar J., Kumar S., Singh S., Siddique K. H. and Nayyar H. 2017. Effects of drought, heat and their interaction on the growth, yield and photosynthetic function of lentil (*Lens culinaris*Medikus) genotypes varying in heat and drought sensitivity. Front. Plant Sci., **8**: 1776.

Noor M.M.A.,Tahjib-Ul-Arif M., AlimS.M.A., Islam M.M., Hasan M.T., Babar M.A., Hossain M.A., Jewel Z.A., Murata Y. and Mostofa M.G. 2024. Lentil adaptation to drought stress: response, tolerance, and breeding approaches. Front. PlantSci., **15**: 1403922. doi: 10.3389/fpls.2024.1403922.

Yang X., Lu M., Wang Y., Wang Y., Liu Z. and Chen S. 2021. Response mechanism of plants to drought stress. Hort., **7**(3): 50.

Zhang Q., Dai B., Fan M., Yang L., Li C., Hou G. and Li J. 2024. Genome-wide profile analysis of the HSP20 family in lettuce and identification of its response to drought stress. Front. Plant Sci., **15**: 1426719.

CharngYeeYung C. Y., Liu HsiangChin L. H., Liu NaiYu L. N., Hsu FuChiun H. F.and Ko Swee Suak K. S. 2006. Arabidopsis Hsa32, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. Plant Physiol., **140**(4): 1297-1305.

Ji X. R., Yu Y. H., Ni P. Y., Zhang G. H. and Guo D. L. 2019. Genome-wide identification of small Bheat-shock protein (HSP20) gene family in grape and expression profile during berry development. BMC Plant Biol., **19**(1): 433.

Eyles S. J. and Gierasch L. M. 2010. Nature's molecular sponges: small heat shock proteins grow into their chaperone roles. Proc. Natl. Acad. Sci., **107**(7): 2727-2728.

Giese K. C. and Vierling E. 2004. Mutants in a small heat shock protein that affect the oligomeric state: Analysis and allele-specific suppression. J. Biol. Chem., **279**(31): 32674-32683.

Jaya N., Garcia V. and Vierling E. 2009. Substrate binding site flexibility of the small heat shock protein molecular chaperones. Proc. Natl. Acad. Sci., **106**(37): 15604-15609.

Bondino H. G., Valle E. M.and Ten Have A. 2012. Evolution and functional diversification of the small heat shock protein/α-crystallin family in higher plants. Planta, **235**(6): 1299- 1313.

Guo L. M., Li J., He J., Liu H. and Zhang H. M. 2020. A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. Sci. Rep.,**10**(1): 1383.

Gupta S. C., Sharma A., Mishra M., Mishra R. K. and Chowdhuri D. K. 2010. Heat shock proteins in toxicology: how close and how far. Life Sci., **86**(11-12): 377-384.

Huang S., Huang D., Fu L., Qin Y., Tang J., Sihui L. and Zhou J. 2023. Characterization of *Arabidopsis thaliana* HSP20 gene family and its expression analysis under drought and salt stress. Life Sci. Res., **27**(2): 162-169.

Ouyang Y., Chen J., Xie W., Wang L. and Zhang Q. 2009. Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. Plant Mol. Biol., **70**(3): 341–57.

Muthusamy S. K., Dalal M., Chinnusamy V. and Bansal K. C. 2017. Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in bread wheat. J. Plant Physiol., **211**: 100-113.

Zhao P., Wang D., Wang R., Kong N., Zhang C., Yang C. and Chen Q. 2018. Genome-wide analysis of the potato HSP20 gene family: identification, genomic organization and expression profiles in response to heat stress. BMC Genomics, **19**: 1-13.

Liu S., Wu Y., Li Y., Zhang Z., He D., Yan J. and Liu Y. 2024. Genome-Wide Identification and Expression Analysis of Heat Shock Protein 20 (HSP20) Gene Family in Response to High-Temperature Stress in Chickpeas (*Cicer arietinum* L.). Agronomy, **14**(8): 1696.

Lopes-Caitar V.S., de Carvalho M.C., Darben L.M., Kuwahara M.K., Nepomuceno A.L. and Dias W.P. 2013. Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. BMC Genomics, **14**: 577.

Ramsay L., Koh C. S., Kagale S., Gao D., Kaur S., Haile T. and Bett K. E. 2021. Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies

of wild and cultivated lentil species. Crop Sci., **50**: 1831-1839. bioRxiv 2021.07. 23.453237.

Finn R., Griffiths-Jones S. and Bateman A. 2003. Identifying protein domains with the Pfam database. Curr. Protoc. Bioinformatics, **1**(1): 2-5.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol., **28**(10): 2731-2739.

Saitou N. and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., **4**(4): 406-425.

Hu B., Jin J., Guo A.Y., Zhang H., Luo J. and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics, **31**(8): 1296-1297.

Bailey T.L., Johnson J., Grant C.E. and Noble W.S. 2015. The MEME suite. Nucleic Acids Res.,**43**(W1): W39-W49.

Chen C., Chen H., He Y. and Xia R. 2018. TBtools: a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv 289660.

Chao J., Li Z., Sun Y., Aluko O. O., Wu X., Wang Q. and Liu G. 2021. MG2C: A user-friendly online tool for drawing genetic maps. Mol. Hortic., **1**(1): 16.

Lescot M.,Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y. and Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res., **30**(1): 325-327.

Sharanbasappa D. M., Dwijesh C. M., Krishna K. C., Anu S., Neeraj B.,UlavappaA., Pavana B., Mohammed S. F., Sudhir S. and Girish K. J. 2024. NaturePred: A Tool for Revolutionizing Natural Product Classification with Artificial Intelligence. Curr. Proteomics.

Horton P., Park K. J., Obayashi T., Fujita N., Harada H., Adams-Collier C. J. and Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res., **35**(suppl_2): W585-W587.

Lawrence M., Gentleman R. and Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics, **25**(14): 1841.

Lawrence M., Huber W.,Pagès H.,Aboyoun P., Carlson M., Gentleman R. and Carey V. J. 2013. Software for computing and annotating genomic ranges. PLoS Comput. Biol., **9**(8): e1003118.

Gu Z., Gu L.,Eils R., Schlesner M. and Brors B. 2014. " Circlize" implements and enhances circular visualization in R.

Andrews S. 2010. Fast QC: a quality control tool for high throughput sequence data.

Bolger A. M., Lohse M. and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, **30**(15): 2114-2120.

Langdon W. B. 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Mining., **8**(1): 1.

Pertea M., Pertea G. M., Antonescu C. M., Chang T. C., Mendell J. T. and Salzberg S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol., **33**(3): 290-295.

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

Hua Y., Liu Q., Zhai Y., Zhao L., Zhu J., Zhang X. and Wang D. 2023. Genome-wide analysis of the HSP20 gene family and its response to heat and drought stress in Coix (*Coix lacryma-jobi* L.). BMC Genomics, **24**(1): 478.

Yu J., Cheng Y., Feng K., Ruan M., Ye Q. and Wang R. 2016. Genome-wide identification and expression profiling of tomato Hsp20 gene family in response to biotic and abiotic stresses. Front. Plant Sci., **7**: 1215.

Qi H., Chen X., Luo S., Fan H., Guo J. and Zhang X. 2022. Genome-wide identification and characterization of heat shock protein 20 genes in maize. Life., **12**(9): 1397

Ji X. R., Yu Y. H., Ni P. Y., Zhang G. H. and Guo D. L. 2019. Genome-wide identification of small heat-shock protein (HSP20) gene family in grape and expression profile during berry development. BMC Plant Biol., **19**(1): 433.

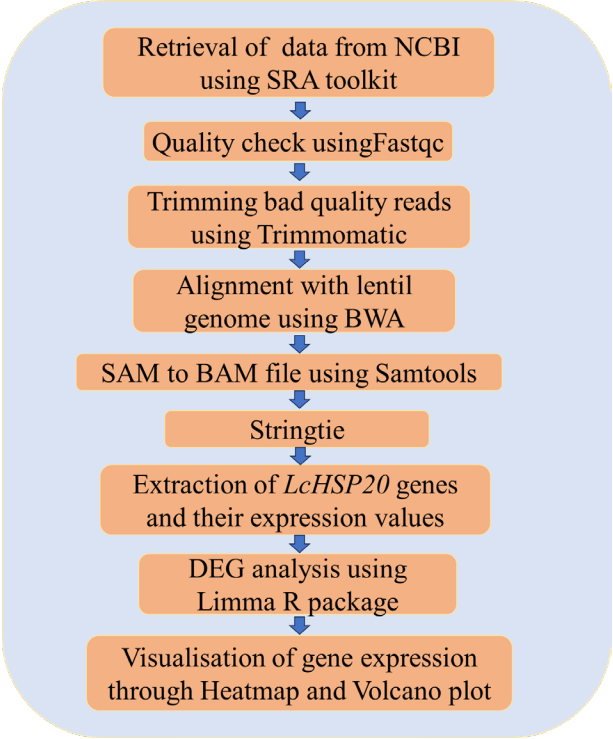
Ma W., Zhao T., Li J., Liu B., Fang L. and Hu Y. 2016. Identification and characterization of the GhHsp20 gene family in *Gossypium hirsutum*. Sci Rep., **6**: 32517.

Hou Z., Li A., Huang C. 2022. Genome-wide identification, characterization and expression of HSP20 gene family in dove. Front. Genet., **13**: 1011676.

Yan H., Zhang A., Chen J., He X., Xu B., Xie G. and Huang L. 2017. Genome-wide analysis of the PvHsp20 family in switchgrass: motif, genomic organization, and identification of stress or developmental-related Hsp20s. Front. Plant Sci., **8**: 1024.

Jeffares D.C., Penkett C.J. and Bahler J. 2008. Rapidly regulated genes are intron poor. Trends Genet., **24**(8): 375–8.

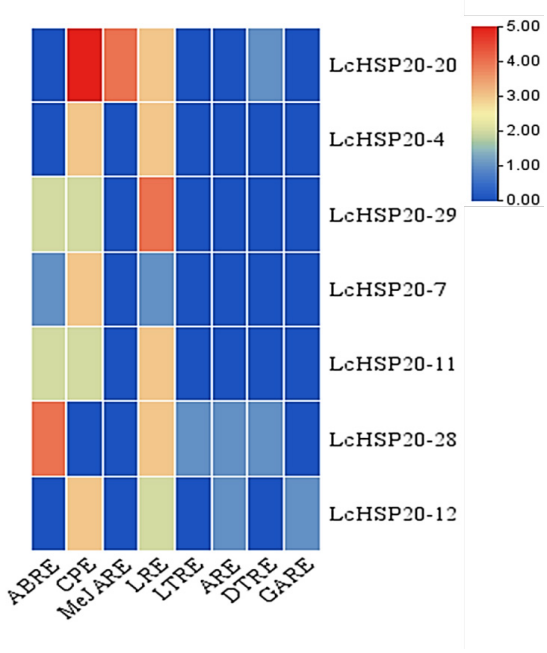
Gonzalez-Gordo S., Palma J.M. and Corpas F.J. 2023. Small heat shock protein (sHSP) gene family from sweet pepper (*Capsicum annuum* L.) fruits: involvement in ripening and modulation by nitric oxide (NO). Plants, **12**(2): 389.



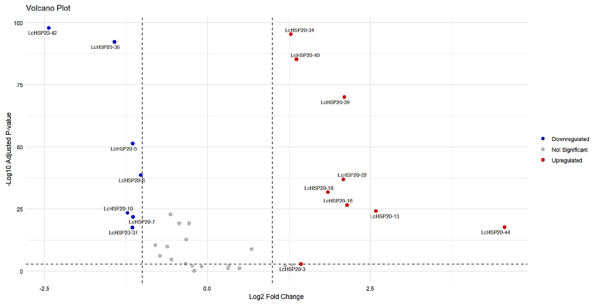
Supplementary Fig. 1. Flowchart of differential expression analysis of LcHSP20 gene under abiotic stress (salt) condition

Motif	Symbol	Motif Consensus
1.		LPENAKLDQVKASMEBGVLTV
2.		HVFKADLPGLKKEEVKVEVED
3.		SGERKVEKEDKNDKWHRVERSSGKFLRRF
4.		GRRSNVFDPFSLDVWDPLKDFPFPSSALSASF
5.		AFVSTRVDWKETPEA
6.		PKEEVKKPEVKSIEI
7.		MSLIPSFF
8.		MDFRLVGLDAPVLHALHQLMDLSDDNSSDK
9.		LSVKGRRSTSSSKVADCPNGCAVSSYLKREIPYGPFEVLWP
10.		LNCLKLWRAVIKLGGYEKVTGCKLWRQVGESFKPPKTCTTVSWTFRGFY

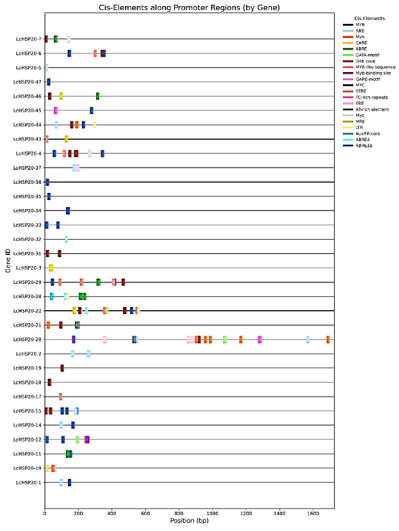
Supplementary Fig. 2. Sequence logos and legend for the 10 conserved motifs identified in LcHSP20 proteins



Supplementary Fig. 3. Count of identified cis-regulatory elements in 7LcHSP20 genes



Supplementary Fig. 5: Volcano plot of LcHSP20 genes under salt stress condition



Supplementary Fig. 4. Distribution of cis-regulatory elements across 47LcHSP20 genes

Supplementary Table S1. A list of LcHSP20genes and their sequence features and subcellular location prediction

Geneid	Assigned name	Chrno	Start	End	Length	Molecular weight	Isoelectric Point	Subcellular localisation
Lcu.2RBY.1g044550	LcHSP20-1	1	353335637	353336701	212	23752.3469	9.10946331	Cytoplasm
Lcu.2RBY.1g057110	LcHSP20-2	1	430292811	430295159	403	45361.9163	5.875128365	Cytoplasm
Lcu.2RBY.1g060540	LcHSP20-3	1	447515005	447515906	198	22782.9153	6.611591148	Extracellular
Lcu.2RBY.1g060580	LcHSP20-4	1	447533693	447535008	198	22772.9469	6.670419502	Extracellular
Lcu.2RBY.1g064150	LcHSP20-5	1	471480851	471481999	152	16833.4362	10.60680599	Chloroplast
Lcu.2RBY.2g009590	LcHSP20-6	2	24038751	24040475	138	15836.9122	5.407116127	Cytoplasm
Lcu.2RBY.2g032660	LcHSP20-7	2	173185742	173187582	169	19193.0137	9.498336601	Cytoplasm
Lcu.2RBY.2g050480	LcHSP20-8	2	322162822	322163393	140	16078.0475	5.668632317	Cytoplasm
Lcu.2RBY.3g000990	LcHSP20-9	3	2179991	2180873	78	8739.0651	9.300354195	Nucleus
Lcu.2RBY.3g002810	LcHSP20-10	3	11554463	11557679	415	47001.9782	9.385259056	Chloroplast
Lcu.2RBY.3g021960	LcHSP20-11	3	144875034	144879750	235	27013.6104	9.316406822	Chloroplast
Lcu.2RBY.3g048230	LcHSP20-12	3	304303099	304304188	211	24130.4761	9.263349342	Chloroplast
Lcu.2RBY.3g054180	LcHSP20-13	3	334944638	334945581	232	26041.4384	7.760193443	Chloroplast
Lcu.2RBY.3g064450	LcHSP20-14	3	385012228	385013093	157	17704.2739	6.163984108	Cytoplasm
Lcu.2RBY.3g073330	LcHSP20-15	3	423147440	423153952	389	43198.4546	6.352575874	Nucleus
Lcu.2RBY.4g002500	LcHSP20-16	4	4886774	4887998	229	25820.6832	8.525831032	Nucleus
Lcu.2RBY.4g007460	LcHSP20-17	4	30026875	30028534	179	19497.9977	5.069151497	Nucleus
Lcu.2RBY.4g020020	LcHSP20-18	4	122308745	122309808	202	22744.4839	6.035243797	Mitochondria
Lcu.2RBY.4g055500	LcHSP20-19	4	367001607	367002262	168	18708.8943	5.690174294	Nucleus
Lcu.2RBY.4g055520	LcHSP20-20	4	367056833	367059239	158	17977.053	6.187117577	Cytoplasm
Lcu.2RBY.4g055540	LcHSP20-21	4	367171071	367171944	155	17741.7508	5.980451012	Cytoplasm
Lcu.2RBY.4g055790	LcHSP20-22	4	368765432	368766550	155	17778.7694	6.187060738	Cytoplasm
Lcu.2RBY.4g055810	LcHSP20-23	4	368850438	368858117	318	35836.2091	8.987360191	Cytoplasm
Lcu.2RBY.4g056020	LcHSP20-24	4	369728958	369729434	158	18107.2362	5.571039772	Cytoplasm
Lcu.2RBY.4g056060	LcHSP20-25	4	369766781	369767259	134	15019.8231	4.900055504	Cytoplasm
Lcu.2RBY.4g056090	LcHSP20-26	4	369852013	369852489	158	18075.1928	5.571039772	Cytoplasm
Lcu.2RBY.4g056110	LcHSP20-27	4	369874497	369874973	158	17994.1248	6.19200573	Cytoplasm
Lcu.2RBY.5g001050	LcHSP20-28	5	1218947	1219988	164	18384.753	6.035130119	Cytoplasm
Lcu.2RBY.5g012100	LcHSP20-29	5	27686373	27692722	573	63565.9286	4.841738701	Nucleus
Lcu.2RBY.5g053160	LcHSP20-30	5	389965510	389966197	131	14948.9052	8.926630974	Peroxisome
Lcu.2RBY.5g053170	LcHSP20-31	5	390002674	390003665	142	16336.6505	7.989711952	Peroxisome
Lcu.2RBY.5g056130	LcHSP20-32	5	411708424	411709553	243	27677.9899	8.912770271	Cytoplasm
Lcu.2RBY.5g056140	LcHSP20-33	5	411715038	411715913	157	17657.6712	5.484132957	Cytoplasm
Lcu.2RBY.5g056150	LcHSP20-34	5	411872783	411873893	157	17821.9302	6.449031639	Cytoplasm
Lcu.2RBY.5g061120	LcHSP20-35	5	431736939	431747147	510	58993.4192	9.046284294	Plastid
Lcu.2RBY.5g061140	LcHSP20-36	5	431749583	431751518	349	39781.8702	5.226538658	Extracellular
Lcu.2RBY.5g063890	LcHSP20-37	5	438924418	438932490	269	29832.4681	9.271214485	Cytoplasm
Lcu.2RBY.5g069130	LcHSP20-38	5	459738758	459747209	329	36810.5248	6.425216103	Cytoplasm
Lcu.2RBY.5g069290	LcHSP20-39	5	460804933	460805715	160	17685.9823	5.950269508	Cytoplasm
Lcu.2RBY.5g069310	LcHSP20-40	5	460836099	460836996	160	17753.1145	5.954248238	Cytoplasm
Lcu.2RBY.5g069330	LcHSP20-41	5	460877908	460878701	204	23480.1262	8.678556633	Cytoplasm
Lcu.2RBY.6g026770	LcHSP20-42	6	207592125	207593274	169	19621.726	8.558580971	Peroxisome
Lcu.2RBY.6g044730	LcHSP20-43	6	313400069	313401314	203	24232.1118	5.561206627	Cytoplasm
Lcu.2RBY.7g006760	LcHSP20-44	7	13763050	13763901	102	11928.236	5.17021122	Cytoplasm

Lcu.2RBY.7g061430	LcHSP20-45	7	474402329	474404728	366	40996.8885	6.233270836	Chloroplast
Lcu.2RBY.L005340	LcHSP20-46	Lcu.2RBY.unitig1147	860738	862039	200	22768.5714	7.92717762	Chloroplast
Lcu.2RBY.L014150	LcHSP20-47	Lcu.2RBY.unitig3163	68639	71847	193	22211.5764	5.510619926	Nucleus

Supplementary Table S2. One-to-one orthologous relationships between LcHSP20 and Arabidopsis thaliana.

Lentil Gene id	Chrno	Start	End	Arabidopsis Geneid	Chrno	Start	End
Lcu.2RBY.5g063890	5	438924418	438932490	AT1G06460	1	1967087	1969656
Lcu.2RBY.4g002500	4	4886774	4887998	AT1G52560	1	19574538	19575866
Lcu.2RBY.1g057110	1	430292811	430295159	AT1G54840	1	20452593	20454266
Lcu.2RBY.3g073330	3	423147440	423153952	AT1G76510	1	28708339	28712968
Lcu.2RBY.1g044550	1	353335637	353336701	AT1G76770	1	28813454	28814378
Lcu.2RBY.5g012100	5	27686373	27692722	AT2G17410	2	7558916	7563781
Lcu.2RBY.5g061120	5	431736939	431747147	AT2G27140	2	11598305	11599322
Lcu.2RBY.4g056090	4	369852013	369852489	AT2G29500	2	12633065	12634120
Lcu.2RBY.5g001050	5	1218947	1219988	AT3G22530	3	7977604	7978635
Lcu.2RBY.1g060540	1	447515005	447515906	AT4G10250	4	6370338	6371372
Lcu.2RBY.2g009590	2	24038751	24040475	AT4G21870	4	11603481	11604637
Lcu.2RBY.4g020020	4	122308745	122309808	AT4G25200	4	12916926	12918110
Lcu.2RBY.3g054180	3	334944638	334945581	AT4G27670	4	13818825	13819971
Lcu.2RBY.5g069330	5	460877908	460878701	AT4G33200	4	16002466	16014981
Lcu.2RBY.5g069310	5	460836099	460836996	AT5G12020	5	3882231	3883111
Lcu.2RBY.5g053170	5	390002674	390003665	AT5G37670	5	14968758	14969591
Lcu.2RBY.L014150	unitig	68639	71847	AT5G54660	5	22203717	22205152