

Article

Not peer-reviewed version

---

# Integrative Omic Dissection of Quinoa Hsp20 Genes: From In Silico Re-Annotation to Transcriptional Profiling Under Heat, Drought, and Salinity Stress

---

[Sabrina Costa-Tártara](#)\*, [Débora Pamela Arce](#), [Gabriel Tolosa](#), [Guillermo Pratta](#)

Posted Date: 30 April 2026

doi: 10.20944/preprints202604.2134.v1

Keywords: omic data; abiotic stress; Hsp20; gene family; duplications; *Chenopodium quinoa*



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Integrative Omic Dissection of Quinoa Hsp20 Genes: From In Silico Re-Annotation to Transcriptional Profiling Under Heat, Drought, and Salinity Stress

Sabrina Costa-Tártara<sup>1,2,\*</sup> , Débora Arce<sup>2,3</sup> , Gabriel Tolosa<sup>1</sup>  and Guillermo Pratta<sup>2,4</sup> 

<sup>1</sup> Departamento de Ciencias Básicas, Universidad Nacional de Luján

<sup>2</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

<sup>3</sup> Secretaría de Ciencia y Tecnología, Facultad Regional San Nicolás, Universidad Tecnológica Nacional, San Nicolás, Argentina

<sup>4</sup> IICAR-CONICET, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental Villarino, Zavalla, Argentina

\* Correspondence: scosta@unlu.edu.ar; Tel.: +54-11-5820-8093 (SCT)

## Abstract

The Hsp20 protein family, recognized in all organisms for its chaperone activity in the heat-stress response, is part of the Heat Shock Protein (Hsp) superfamily, defined by a conserved alpha-crystallin domain (ACD). Hsp20s are the smallest proteins in the superfamily (mostly between 15 and 22 kDa) and assist in protein refolding during stress and developmental processes. In this study, we characterize the Hsp20 gene family in *Chenopodium quinoa* ( $2n = 4x = 36$ ) using an integrative omic approach. *C. quinoa* is well known for its global contributions to food production and tolerance to various abiotic stresses. We identified 69 CqHsp20 genes distributed across the nine chromosomes of each subgenome (A and B), organized mainly into homologous pairs, with paralogs on eight chromosomes, likely from tandem duplications, suggesting a well-conserved evolutionary pattern within the species. The phylogenetic analysis grouped CqHsp20 proteins into two main clusters, split into four sub-clusters based on peptides' cellular localization, consistent with a characteristic gene structure and conserved motif analysis. The integration of transcriptomic data from published experiments enabled us to detect a cluster of putatively ubiquitously expressed CqHsp20, as well as other groups that showed differential responses across abiotic stress conditions. The pattern shows more genes with transcriptional activity under drought and salinity than under heat, key adaptive traits underlying quinoa's known ecological versatility. Some of these genes, with null or low transcriptional activity under heat stress, encode organelle-targeting peptides, a phenomenon not reported in other model plant studies. Varying expression within CqHsp20 homologous and paralogs supports the idea that gene duplication creates genomic diversity, facilitating adaptation to variable extreme environments. However, while theoretical and in silico analysis provide valuable insight into quinoa Hsp20 response, empirical data are essential to unequivocally understand how these gene expression variations affect quinoa response to abiotic stressors.

**Keywords:** omic data; abiotic stress; Hsp20; gene family; duplications; *Chenopodium quinoa*

## 1. Introduction

Plants are continuously exposed to fluctuating environmental conditions that may become unfavorable for life, thereby constituting stress [1]. Climate-related factors act as stressors that trigger the synthesis of stress-responsive proteins [2], such as Heat Shock Proteins (HSPs), which play a central role in maintaining cellular homeostasis by coordinating protein folding, repairing damaged proteins, and, ultimately, regulating degradation processes. The HSPs superfamily acts as chaperones, assisting in the folding and refolding of proteins under stress and preventing their irreversible aggregation. Among the five HSP families characterized in plants [3], the Hsp20 subfamily is the smallest, with

proteins ranging in size from 15 to 22 kilodaltons (kDa) mostly. They are ATP-independent and play an intermediate role during stress, forming dynamic oligomeric complexes as a regulatory mechanism [4]. When no environmental stress is present, the Hsp20 gene expression is constitutive. Differential expression of the Hsp20 subfamily has been observed in some growth and development stages, such as embryogenesis, germination [5], pollen development and fruit ripening [6–9]. Based on a comparative sequence analysis among *Arabidopsis thaliana*, *Populus trichocarpa* and *Oryza sativa* genomes, several cytoplasmic, nuclear, mitochondrial, plastidial and endoplasmic reticulum Hsp20 members were identified [3,10,11] in a variety of numbers (e.g., 19 in *A. thaliana* vs. 36 in *P. trichocarpa*). Despite their diversification in sequence and cellular localization [12,13], their structure contains a conserved  $\alpha$ -crystallin domain (ACD), also known as Hsp20 (PF00011), which represents the characteristic signature motif of the subfamily [14]. As expected for genes related to stress response [15,16], processes such as tandem and segmental gene duplications drive the expansion of the Hsp20 subfamily [17], and other HSPs [18], besides other families involved in nutrient transport, phytosanitary resistance and domestication. It is hypothesized that gene duplications promote long-term evolutionary adaptation; however, it may be difficult to assess the evolutionary relationships among genes precisely. Nowadays, more plant genomes have been analyzed including common bean *Phaseolus vulgaris* [19] and pigeonpea (*Cajanus cajan*) [20], which each harbour Hsp20 subfamilies of 46 and 39 members, respectively. These species are considered pulses or vegetable crops for cultivation in marginal areas. For species such as tomato (*Solanum lycopersicum* cv. Caimanta), pepper (*Capsicum annuum*), and rice [7,21,22], deeper studies have allowed a better understanding of the behaviour of specific Hsp20 genes. As a halophytic species, *Chenopodium quinoa* (quinoa) exhibits tolerance to several abiotic stresses and simultaneously adapts to a variety of agro-ecological environments [23,24]. Considered an antique pseudocereal [25] in South America, it is primarily cultivated for its superior protein content and balanced essential amino acid profile in grains [26,27]. Quinoa belongs to the Allotetraploid Goosefoot Complex (ATGC) [28,29] with an AABB genome ( $2n = 4x$ ) with 36 chromosomes. The hybridization might have occurred millions of years ago in North America [30–33] among ancestral A- and B- genome species. As a formerly neglected crop, *C. quinoa* is undergoing a global expansion [34,35], promoting the study of several agronomic aspects, stress physiology, and nutritional characteristics worldwide [23,24,36–38]. The genes related to abiotic stress tolerance, such as *CqSOS1* (Salt Overly Sensitive 1) and *CqNHX*, both sodium transporters, were the first genes cloned [39,40], which also revealed differential expression across ecotypes. The availability of omic data increased in the last decade [41–43], thus enhancing in silico analysis to describe relevant gene families [44–48] and approaches for describing the evolution of physiological processes [49–51]. This study aims to characterize the Hsp20 gene subfamily in quinoa through a completely in silico approach. The specific objectives are to: i) identify the Hsp20 gene family members in the *C. quinoa* genome, ii) describe the gene structure, repetitive motifs and regulatory elements, iii) elucidate phylogenetic relationships among protein sequences, and iv) propose possible duplication events. Additionally, by leveraging publicly available datasets, we integrate transcriptomic data generated under abiotic stresses such as drought, salinity, and high temperature. Drought and salinity tolerance are key adaptive traits underlying quinoa's broad ecological range, whereas high temperatures remain a major constraint for its expansion as a crop. This approach is expected to reveal stress-specific Hsp20 chaperone activity patterns, providing insight into mechanisms of protein homeostasis under those conditions.

## 2. Materials and Methods

### 2.1. Identification and Retrieval Information of Hsp20 Genomic Sequences

Putative Hsp20 genomic sequences were retrieved from the National Center for Biotechnology Information (NCBI) database using keywords like "small heat shock protein" and "*Chenopodium quinoa*", and from the annotation file GFF published with the first version of the reference genome (ASM168347v1, accession PI 614886 - QQ74) [30]. The nomenclature for gene Identification follows the chromosome order determined for the second version of reference genome QQ74\_V2 (CoGe Genome

id 60716) [32]. Gene coordinates were extracted from the annotation file. The corresponding protein sequence for each gene were checked through the InterPro [52] and Uniprot databases [53] to confirm the presence of Alpha-crystallin/Hsp20-domain (PF00011 and IPR002068). In addition, we use the proteome developed by [5] on *C. quinoa* mature seed tissue as a valid sequence to check the identity of a few members of the Hsp20 gene family.

## 2.2. Peptide Physicochemical Features and Subcellular Localization Prediction

Physicochemical properties of each Hsp20 protein sequence were calculated using the ProtParam tool (<https://web.expasy.org/protparam/>; [54]). Subcellular localization of the encoded peptides, a key criterion for Hsp20 classification, was predicted using MULocDeep, a deep learning-based approach for multi-compartment localization inference (<https://www.mu-loc.org/>; [55,56]). The same approach was applied to predict subcellular localization of Hsp20 proteins from the model plant species included in the phylogenetic analysis.

## 2.3. Gene Structure, Protein Motif Prediction and Prediction of Transcription Factor Binding Sites

Gene structure description was performed using GSDS 2.0 [57] based on coordinates reported in QQ74\_V2 (CoGe Genome id 60716). Conserved motifs were identified using MEME Suite tools (<https://meme-suite.org/meme/tools/meme>; [58]) with a maximum of 5 motifs occurred 0 or 1 time per sequence of 8 to 20 residues length, a setting chosen to capture the principal conserved regions of the  $\alpha$ -crystallin domain and subfamily specific signatures. We predicted transcription factor binding site (TFBS) patterns within the 1,000 bp upstream regulatory regions of each gene. The analysis was conducted using the Regulatory Sequence Analysis Tool for Plants (RSAT) available through the ELIXIR platform. TFBS prediction was based on position-specific scoring matrices (PSSMs) from *A. thaliana* curated in the Cistrome database [59]. Motifs corresponding to five TF families associated with stress responses were analyzed, including Heat Shock Elements (HSEs), NAC [44], bZIP [45], bHLH [60], and WRKY [48].

## 2.4. Phylogenetic Analysis

Protein sequences of putative Hsp20 genes for *C. quinoa* were multiple aligned with the MUSCLE algorithm [61] with default parameters (include 16 number of refinement iterations). We infer the maximum-likelihood (ML) tree using IQ-TREE 2 [62] with the Model Finder Plus (MFP) parameter in the first run [63]. The branch's support was obtained with the ultrafast bootstrap after 1000 replications [64]. The scripts are detailed in Supplementary (S1). In order to analyze phylogenetic relationships among other Hsp20 identified in model plants, the protein sequences belonging to *A. thaliana*, *O. sativa*, *S. lycopersicum* and *Spinacia oleracea* annotated as Hsp20 was retrieved from NCBI. Every sequence was checked for the presence of the Hsp20 domain through the Uniprot database (Table S1).

## 2.5. Chromosomal Organization, Gene Duplications and Evolutionary Analysis

Gene chromosomal locations were retrieved from the QQ74\_V2 GFF file deposited in the CoGe database. Duplication events in Hsp20 genes were examined primarily following the criteria reported by Yue et al. [48]: the alignment covers >80% of the longer gene, and the aligned region has an identity of 80% or more. The similarity and identity for each sequence pair were calculated with a global alignment algorithm (Needle) implemented in Emboss [65]. Tightly linked genes were counted as one duplication event. Selective pressure acting on duplicated Hsp20 genes and for each putative homeolog pair were assessed by calculating the ratio of non-synonymous ( $K_a$ ) to synonymous ( $K_s$ ) substitution rates ( $K_a/K_s$ ) using KaKs\_Calculator 2.0 [66]. Coding sequences of each member were first aligned at the amino acid level and subsequently converted into codon-based alignments using PAL2NAL [67]. Substitution rates were estimated under the Li-Wu-Luo (LWL) evolution model suited for closely related sequences [68].

## 2.6. Identification of Hsp20 Under Different Stress Conditions and Quantification Pattern Analysis

We retrieved the raw RNA-seq data from four independent experiments conducted under heat (PRJNA526621 from [42]), drought (PRJNA195391 from [69] and PRJNA305752 from [41]), and salinity (PRJNA306026 from [43]) conditions to assess CqHsp20 transcriptional activity. The experiments differed in tissue type, sequencing depth, read length, and experimental design, as they were performed at different time points (details in supplementary S1 and Table S9), likely reflecting the availability of sequencing platforms. The heat experiment consisted of organ-specific treatments applied during the flowering stage, in which heat was imposed on the root (HR, 30 °C), the stem (HS, 35 °C), or both organs simultaneously (HRS; 30 °C and 35 °C, respectively) for eleven days. Leaf samples were collected on day 1 (HR\_1, HS\_1, HRS\_1) and day 11 (HR\_11, HS\_11, HRS\_11), with corresponding controls maintained at 22 °C at each time point. Drought experiments included root tissue only (D1; [69]) from two cultivars (Ingapirca: drought\_I; Ollague: drought\_O) or both root and shoot tissues (D2; [41]) on R49 cultivar (drought\_R49 vs control\_R49), with sampling at three or four time points under varying field capacities. The salinity experiment involved root (control\_r vs S\_root) and aerial tissues (considered as shoots: S\_shoot vs control\_s) from plants exposed to 300 mM NaCl for seven days in a hydroponic system. Raw reads were downloaded from the Sequence Read Archive (SRA) using `prefetch` and `fasterq-dump` (SRA Toolkit). Quality trimming was performed with Trimmomatic [70], using parameters adjusted for read length and library type (S1). Cleaned reads were aligned to the *C. quinoa* QQ74\_V2 reference genome using HISAT2 v2.2.1 [71] with the `-no-softclip` option. Resulting BAM files were sorted with SAMtools v1.22.1 [72]. Gene-level quantification was performed using `featureCounts` [73] from the Subread package [74], with the QQ74\_V2 GFF3 annotation. Exons were used as counting features and aggregated by parent gene. To reduce ambiguous assignments, only reads that overlap at least one-third of their length against a feature were counted, and multi-overlapping reads were excluded. The quantification results were transformed to transcripts per million (TPM), which standardizes for gene length and the number of reads in each sample (SRR#) and is suitable for visualization and comparative quantification.

## 3. Results

### 3.1. Identification and Descriptions of CqHsp20 Genomic Sequences

A total of 69 non-redundant sequences were identified as members of the CqHsp20 gene family (Table 1), distributed across the nine chromosomes of each haploid subgenome defined in the QQ74\_V2. The labels in the annotation file refer to weights and classify proteins into nineteen categories, such as: HSP14.7, HSP15.4, HSP15.7, HSP17.4, HSP17.5, HSP17.6, HSP17.9, HSP18, HSP18.2, HSP18.5, HSP21, HSP21.7, HSP22, HSP22.7, HSP23, HSP26.2, HSP26.5, hspC2. Five genes (CQ004930, CQ006179, CQ031512, CQ032391, and CQ041411) were identified as novel Hsp20 members through a comparative alignment using the CoGeBlast tool <https://genomevolution.org/coge/CoGeBlast.pl>. This involved cross-referencing contigs from [41] and putative Hsp20 genes from [43] against the QQ74\_V2 genome (CoGe id 60716). Consequently, sequences initially labelled as 'Protein of unknown function' in the GFF file were functionally re-annotated as 'putative Hsp20' for all subsequent analyses. The raw RNA-seq data from these studies were integrated for further downstream analysis. We used the quinoa proteome built by [5] from mature seed tissue as a valid reference for contrasting the sequences of CqHsp20 members. Four out of the 6 Hsp20 reported in that study were identified in this analysis, such as CQ025080 (XP\_021769109.1), CQ054519 (XP\_021744114.1), CQ057039 (XP\_021732430.1) and CQ055373 and CQ055384, whose protein product aligned with the same protein NCBI accession ID corresponding to XP\_021732378.1. For all CqHsp20 members, the equivalence nomenclature between the gene IDs reported in QQ74\_V1 and QQ74\_V2 and their corresponding ID entries in the Uniprot database is detailed in Table S2.

**Table 1.** Hsp20 genes from *C. quinoa* (CqHsp20) with their corresponding nucleotide length (bp), chromosomal location, annotation in reference genome and subcellular localization.

CqHsp20	ID Gene	Chr	Start	End	strand	Length (bp)	Note	Description	Subcellular localization	Suborganellar localization
CqHsp20_1	CQ000416	Cq6B	4760615	4762599	+	1984	HSP21	Small heat shock protein%2C chloroplastic ( <i>Pisum sativum</i> OX%3D3888)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane
CqHsp20_2	CQ001244	Cq6B	29417175	29417741	-	566	HSP22.7	22.7 kDa class IV heat shock protein ( <i>Pisum sativum</i> OX%3D3888)	Endoplasmic	endoplasmic reticulum lumen
CqHsp20_3	CQ002590	Cq6B	78911925	78912416	-	491	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_4	CQ003465	Cq5B	31785	32884	+	1099	HSP23	Small heat shock protein%2C chloroplastic ( <i>Oxybasis rubra</i> OX%3D3560)	Mitochondrion	mitochondrion matrix
CqHsp20_5	CQ004929	Cq5B	19873023	19872162	+	1839	HSP17.5-E	17.5 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3847)	Endoplasmic	endoplasmic reticulum membrane
CqHsp20_6	CQ004930	Cq5B	19873649	19875601	+	1952	Putative Hsp20	Protein of unknown function	Endoplasmic	endoplasmic reticulum membrane
CqHsp20_7	CQ004971	Cq5B	21350497	21352202	+	1705	HSP22.0	22.0 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Membrane	cell membrane
CqHsp20_8	CQ006179	Cq5B	70325544	70331841	+	6297	Putative Hsp20	Protein of unknown function	Nucleus	nucleolus, chromosome
CqHsp20_9	CQ008069	Cq7B	9790716	9792735	-	2019	HSP26.5	26.5 kDa heat shock protein%2C mitochondrial ( <i>Arabidopsis thaliana</i> OX%3D3702)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_10	CQ009410	Cq7B	61723203	61724706	+	1503	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Mitochondrion	mitochondrion matrix
CqHsp20_11	CQ009411	Cq7B	61736730	61738298	+	1568	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Mitochondrion	mitochondrion matrix
CqHsp20_12	CQ010302	Cq8B	2128538	2130966	+	2428	HSP15.4	15.4 kDa class V heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Cytoplasm	cytosol
CqHsp20_13	CQ010717	Cq8B	6525197	6526685	+	1488	HSP26.2	26.2 kDa heat shock protein%2C mitochondrial ( <i>Oryza sativa</i> subsp. <i>japonica</i> OX%3D39947)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_14	CQ014165	Cq2B	20862728	20869174	-	1896	HSP15.7	15.7 kDa heat shock protein%2C peroxisomal ( <i>Arabidopsis thaliana</i> OX%3D3702)	Peroxisome	peroxisome membrane
CqHsp20_15	CQ015602	Cq2B	65554231	65556160	-	1929	HSP17.9	17.9 kDa class II heat shock protein ( <i>Helianthus annuus</i> OX%3D4232)	Nucleus	nucleolus, nucleus speckle
CqHsp20_16	CQ017152	Cq3B	6543782	6544985	+	1203	HSP22	Small heat shock protein%2C chloroplastic ( <i>Petunia hybrida</i> OX%3D4102)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_17	CQ020271	Cq4B	5841595	5843390	-	1795	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane
CqHsp20_18	CQ020272	Cq4B	5845203	5846496	-	1293	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane
CqHsp20_19	CQ020274	Cq4B	5849685	5851242	-	1557	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_20	CQ020275	Cq4B	5853034	5855059	-	2025	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_21	CQ020276	Cq4B	5855447	5858290	-	2843	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma
CqHsp20_22	CQ020278	Cq4B	5859801	5861206	-	1405	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_23	CQ022281	Cq4B	66507791	66510378	+	2587	HSP21.7	21.7 kDa class VI heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_24	CQ023196	Cq1B	9076468	9077151	-	683	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Cytoplasm	cytosol
CqHsp20_25	CQ024659	Cq1B	63625415	63625801	+	386	HSP22.0	22.0 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_26	CQ025070	Cq1B	67633753	67634238	+	485	HSP18	18.3 kDa class I heat shock protein ( <i>Oxybasis rubra</i> OX%3D3560)	Nucleus	nucleolus
CqHsp20_27	CQ025074	Cq1B	67663581	67664066	+	485	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_28*	CQ025080*	Cq1B	67789198	67789677	+	479	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_29	CQ025082	Cq1B	67796702	67797133	-	431	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_30	CQ025097	Cq1B	67976470	67977066	+	596	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Plastid	chloroplast thylakoid lumen, plastid chloroplast thylakoid membrane
CqHsp20_31*	CQ025122*	Cq1B	68262816	68263304	+	488	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_32	CQ025859	Cq6A	5459407	5460351	+	944	HSP21	Small heat shock protein%2C chloroplastic ( <i>Pisum sativum</i> OX%3D3888)	Nucleus	nucleolus
CqHsp20_33	CQ026671	Cq6A	24294981	24295547	+	566	HSP22.0	22.0 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Endoplasmic	endoplasmic reticulum lumen
CqHsp20_34	CQ028022	Cq6A	58683254	58683745	-	491	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_35	CQ028869	Cq5A	240412	241889	+	1477	HSP23	Small heat shock protein%2C chloroplastic ( <i>Oxybasis rubra</i> OX%3D3560)	Mitochondrion	mitochondrion matrix
CqHsp20_36	CQ030247	Cq5A	18530915	18532643	-	1728	HSP22.0	22.0 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Membrane	cell membrane
CqHsp20_37	CQ031384	Cq5A	55145998	55146510	+	512	HSP18.5-C	18.5 kDa class I heat shock protein ( <i>Glycine max</i> OX%3D3847)	Nucleus	nucleolus
CqHsp20_38	CQ031512	Cq5A	56496464	56498160	+	1696	Putative Hsp20	Protein of unknown function-SHSP domain-containing protein	Endoplasmic	endoplasmic reticulum membrane
CqHsp20_39	CQ032391	Cq9B	1045640	1047217	+	1577	Putative Hsp20	SHSP domain-containing protein	Nucleus	nucleolus
CqHsp20_40	CQ032642	Cq9B	3213204	3215697	+	2493	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_41	CQ032647	Cq9B	3248233	3248712	+	479	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_42	CQ033041	Cq9B	6925014	6927875	+	2861	HSP17.4B	17.4 kDa class III heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_43	CQ033691	Cq9B	16252267	16254671	+	2404	HSP14.7	14.7 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_44	CQ033695	Cq9B	16261504	16265597	-	4093	hspC2	Small heat shock protein C2 ( <i>Rickettsia felis</i> (strain ATCC VR-1525 / URRAWCa12) OX%3D315456)	Nucleus	chromosome
CqHsp20_45	CQ035567	Cq4A	5774532	5791524	-	16992	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, chloroplast thylakoid lumen
CqHsp20_46	CQ035568	Cq4A	5797559	5798553	-	994	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_47	CQ035569	Cq4A	5799657	5804216	-	4559	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_48	CQ037843	Cq4A	57885839	57888383	-	2544	HSP21.7	21.7 kDa class VI heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Cytoplasm	cytosol
CqHsp20_49	CQ038186	Cq8A	2335050	2336484	+	1434	HSP15.4	15.4 kDa class V heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_50	CQ038601	Cq8A	6861985	6863544	+	1559	HSP26.2	26.2 kDa heat shock protein%2C mitochondrial ( <i>Oryza sativa</i> subsp. <i>japonica</i> OX%3D39947)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_51	CQ041158	Cq7A	2563866	2564345	-	479	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_52	CQ041161	Cq7A	2584994	2585461	-	467	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_53	CQ041411	Cq7A	4638026	4639642	-	1616	Putative Hsp20	SHSP domain-containing protein	Nucleus	nucleolus
CqHsp20_54	CQ041626	Cq7A	6818599	6821690	+	3091	HSP17.4B	17.4 kDa class III heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_55	CQ043111	Cq7A	46482205	46483574	+	1369	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Mitochondrion	mitochondrion matrix
CqHsp20_56	CQ043112	Cq7A	46507093	46508453	+	1360	HSP22	Small heat shock protein%2C chloroplastic (Fragment) ( <i>Glycine max</i> OX%3D3847)	Mitochondrion	mitochondrion matrix
CqHsp20_57	CQ044478	Cq3A	6321631	6323853	+	2222	HSP22	Small heat shock protein%2C chloroplastic ( <i>Petunia hybrida</i> OX%3D4102)	Plastid	chloroplast stroma, plastid chloroplast thylakoid membrane, chloroplast thylakoid lumen
CqHsp20_58	CQ051072	Cq9A	36959210	36962360	+	3150	HSP14.7	14.7 kDa heat shock protein ( <i>Arabidopsis thaliana</i> OX%3D3702)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_59	CQ051076	Cq9A	36970639	36975236	-	4597	hspC2	Small heat shock protein C2 ( <i>Rickettsia felis</i> (strain ATCC VR-1525 / URRAWCa12) OX%3D315456)	Nucleus	chromosome
CqHsp20_60	CQ051646	Cq9A	44858084	44859935	+	1851	HSP26.5	26.5 kDa heat shock protein%2C mitochondrial ( <i>Arabidopsis thaliana</i> OX%3D3702)	Mitochondrion	inner membrane, mitochondrion matrix
CqHsp20_61	CQ053725	Cq2A	15091148	15093074	-	1926	HSP15.7	15.7 kDa heat shock protein%2C peroxisomal ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_62	CQ054519	Cq2A	40101462	40104471	+	3009	HSP17.9-D	17.9 kDa class II heat shock protein ( <i>Glycine max</i> OX%3D3847)	Nucleus	nucleolus
CqHsp20_63	CQ055373	Cq2A	49930925	49931410	-	485	HSP18	18.3 kDa class I heat shock protein ( <i>Oxybasis rubra</i> OX%3D3560)	Nucleus	nucleolus
CqHsp20_64	CQ055378	Cq2A	49989213	49989705	+	492	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_65	CQ055384	Cq2A	50041297	50041788	-	491	HSP17.6C	17.6 kDa class I heat shock protein 3 ( <i>Arabidopsis thaliana</i> OX%3D3702)	Nucleus	nucleolus
CqHsp20_66	CQ055420	Cq2A	50431690	50432130	+	440	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus
CqHsp20_67	CQ055422	Cq2A	50435461	50435913	-	452	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Cytoplasm	cytosol
CqHsp20_68	CQ057039	Contig2248	9641	10093	+	452	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Cytoplasm	cytoskeleton
CqHsp20_69	CQ057041	Contig2248	13424	13864	-	440	HSP18.2	18.2 kDa class I heat shock protein ( <i>Medicago sativa</i> OX%3D3879)	Nucleus	nucleolus

### 3.2. Physicochemical Properties and Peptide Subcellular Localization

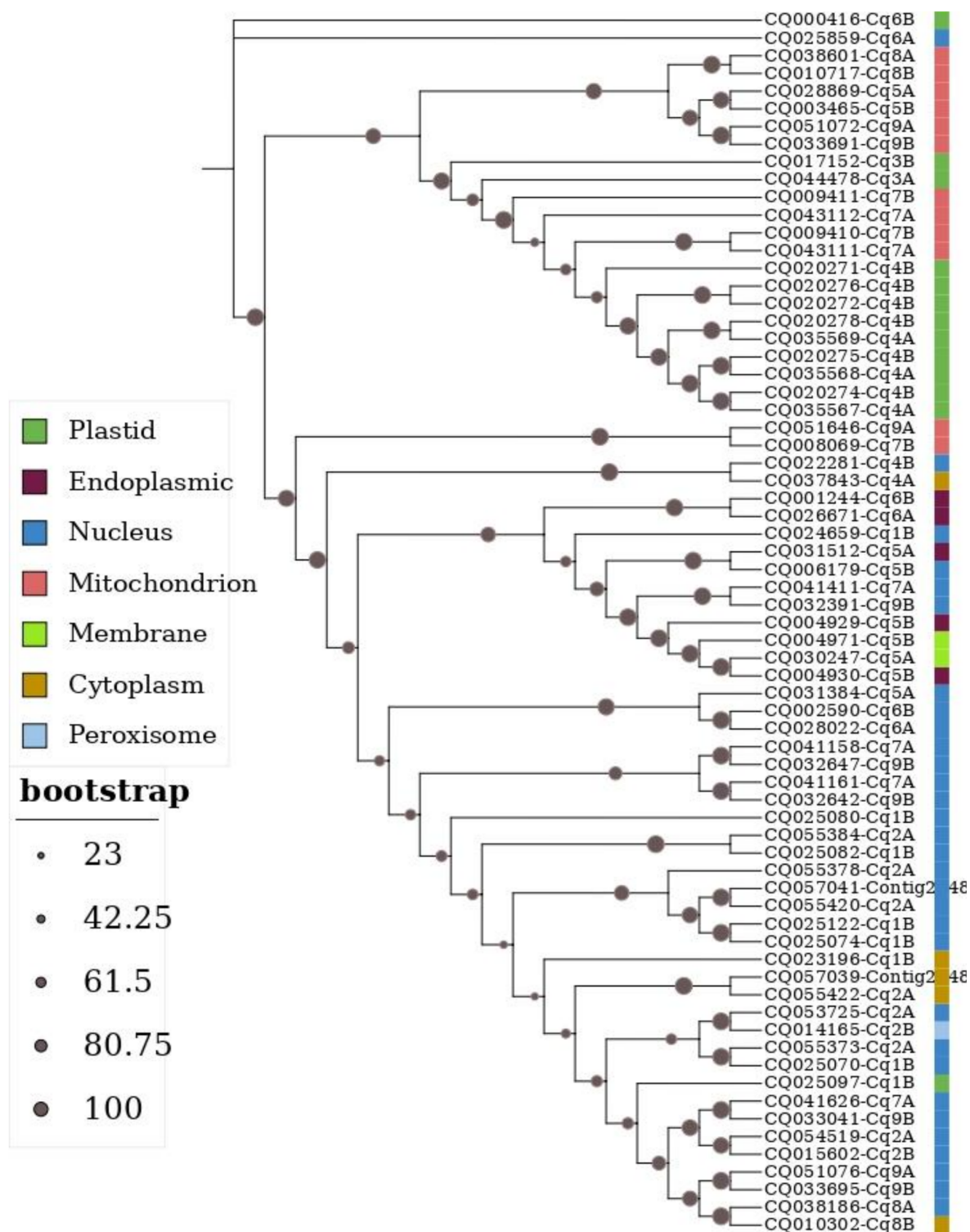
The range of molecular weight (MW) among CqHsp20 peptides was between 12,94 kDa (CQ032642) and 38,41 kDa (CQ035567) (Table S3). Sixty-four Hsp20 members weigh between 13 and 28 kDa, the most frequent MW range reported for the Hsp20 subfamily; meanwhile, only five members weigh higher. The MW predicted for some members does not match the GFF annotation label. The Isoelectric point (pI) showed a bimodal distribution: 35 members presented values between 4.73 and 7.21, considered acid peptides, whereas the rest showed higher values, with a maximum of 9.69, considered basic peptides. The prediction of the subcellular target localization for each CqHsp20 peptide resulted in seven different places, such as the Nucleus (N) (31), Cytoplasm (C) (5), and Plastid (CP) (13), Endoplasmic reticulum (ER) (5), Mitochondria (M) (12), Membrane (2), and Peroxisome (1) composing every classes descriptive for Hsp20 (Table 1). Numbers in parentheses indicate total members. Besides, every CqHsp20 peptide product has, in turn, a suborganellar localization assigned to it. For CqHsp20 localized in C, the majority were cytosolic; in contrast, for CqHsp20 localized in N, the majority (27) showed a nucleolar signal. Only three were predicted to be expressed on the chromosome (with the highest MW) (Table S4).

### 3.3. Gene Structure, Protein Motif Prediction and Prediction of Transcription Factor Binding Sites

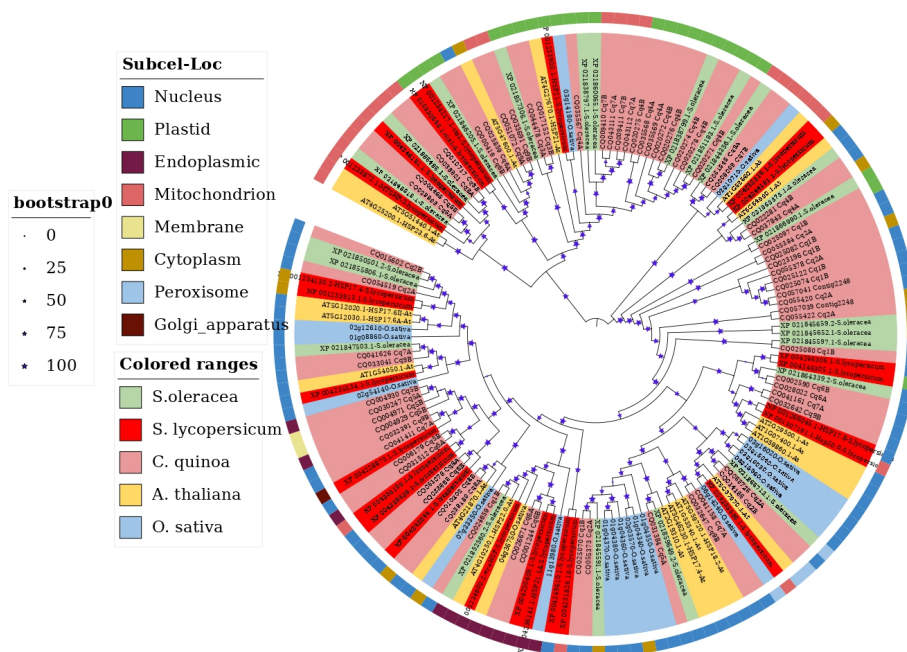
Most CqHsp20 genes showed a simple structure, being intronless or containing a single intron, as expected for small Hsp20 proteins; only nine genes contained two or more introns. Exon-intron organization mapped onto the phylogeny revealed a conserved gene structure within and across subclades (Figure S1, Supplementary Material). Intronless genes were mainly nuclear and cytoplasmic Hsp20s, whereas plastid- and mitochondrial-localized members generally contained one or two introns. The scheme shows the coordinates and hierarchical data for each gene retrieved from GFF-QQ74\_V2 so that it may contain annotation errors, such as the absence of UTR regions (e.g., CQ055420-Cq2A, CQ055373-Cq2A, CQ032642-Cq9B) or tandem disposition (e.g., CQ041626-Cq7A/CQ033041-Cq9, CQ033695-Cq9B/CQ051076-Cq9A, and putative homologous; CQ035569-Cq4A). However, the pattern showed similar structures among putative homologous pairs primarily differing in intron length. The analysis of conserved motifs in protein sequences also revealed a conserved organization among members that share a subcellular localization (Figure 2S). Nuclear Hsp20s predominantly contained five motifs, whereas most cytosolic members exhibited four. A closely related basal nuclear subclade displayed only three motifs with minor divergence. Endoplasmic reticulum-localized and membrane-associated Hsp20s also contained three motifs, while plastid and mitochondrial Hsp20s generally exhibited four motifs. Analysis of the 1 kb upstream regions of CqHsp20 genes identified putative cis-acting element motifs (TFBS), providing information on functional responses and their potential involvement in abiotic stress regulation (Table S8). A total of 621 TFBS were predicted (p-value < 0.0001), with NAC (ion transport) and bZIP elements being the most abundant; thirty-six CqHsp20 genes contained at least 3 NAC motifs and particularly nine genes (CQ010717, CQ028869, CQ033691, CQ017152, CQ022281, CQ024659, CQ025122, CQ053725, CQ025097) presented 5 to 8 motifs. Twenty-seven CqHsp20 genes presented at least three motifs for bZIP. Although the class of bHLH-related elements was moderately represented, the three types of cis-elements are related to TFs linked to salinity and drought tolerance. In contrast, HSE and WRKY elements were less frequent. However, 15 genes harboured 3-5 HSE motifs, suggesting potential responsiveness to heat stress.

### 3.4. Phylogenetic Analysis

A maximum-likelihood (ML) phylogenetic tree of Hsp20 protein sequences was constructed under the JTT+G4 substitution model, which was selected as the best-fitting model by ModelFinder based on the lowest AIC and BIC values (Figure 1). The resulting topology resolved two major, strongly supported clusters (bootstrap = 100) that grouped 67 CqHsp20 proteins largely according to their predicted subcellular localization, together with two more divergent members, localized to plastids (CP) (CQ000416-Cq6B) and the nucleus (N) (CQ0025859-Cq6A). The first, more divergent cluster comprised 21 CP and mitochondrial (M) Hsp20s (organelle-localized), organized into seven homeologous gene pairs involving chromosomes 4A/4B, 5A/5B, 7A/7B, 8A/8B, and 9A/9B, along with four more distantly related members located on chromosomes 3A, 3B, 7A, and 7B, and three closely related proteins encoded on chromosome 4B. The second central cluster contained 46 proteins and included several well-supported pairwise groupings of M, N, and cytosolic (C) Hsp20s, corresponding to genes located on chromosomes 9A/7B and 4A/4B (likely homologous). Within this cluster, 11 proteins predicted to localize to the N, endoplasmic reticulum (ER), or membranes formed a supported subgroup (bootstrap > 80.75), comprising three homeologous pairs on chromosomes 5A/5B and 6A/6B, two closely related members on chromosomes 7A and 9B, and two proteins with divergent signal peptides localized to the ER and membranes despite being encoded on chromosome 5B. The remaining 31 CqHsp20 proteins, predominantly N with three C members and single representatives localized to the peroxisome and plastid, clustered together. They included six homeologous gene pairs located on chromosomes 2A/2B, 6A/6B, 8A/8B, and 9A/9B. The final subcluster showed a close phylogenetic relationship among three nuclear-localized proteins encoded on chromosomes 7A and 9B. Another ML phylogenetic tree including Hsp20 proteins from *A. thaliana* (At; 19), *O. sativa* (Os; 19), *S. lycopersicum* (Sl; 24), and *S. oleracea* (So; 22) was constructed under JTT+F+R6, the best-fitting according to AIC (Figure 2). The model uses an exchangeability matrix (JTT) based on amino-acid frequencies estimated from the alignment (+F) and models among-site rate heterogeneity with six freely estimated rate categories (R6). Although another substitution model (VT+R6) resulted in the best-fitting model under BIC, the phylogenies inferred under both models were largely congruent, preserving the major split between organelle-localized Hsp20s and the remaining N, ER, C and membrane members, with the only difference being the alternative placement of a predominantly ER-localized cluster. The most strongly supported and phylogenetically distant cluster comprised organelle-localized Hsp20s, in which SoHsp20 members showed the closest relationships with CqHsp20 proteins in at least six homologous pairs. In contrast, Hsp20s from *A. thaliana*, *O. sativa*, and *S. lycopersicum* displayed more divergent relationships. Notably, a pair of CqHsp20 genes (CQ051646-Cq9A and CQ008069-Cq7B) clustered with one representative from each model species, supporting the definition of this subcluster. The remaining CqHsp20 proteins are grouped with Hsp20 proteins from different model species based on subcellular localization, reflecting distinct organelle-targeting peptide signals.



**Figure 1.** Phylogenetic relationships among CqHsp20 proteins. An unrooted phylogenetic tree was generated with Maximum Likelihood method with the matrix-based JTT model and gamma distribution with four categories to rate heterogeneity among-site. The colors in the right stripe indicate the seven subcellular localizations predicted with MULocDeep. The dark filled circles indicate permutation values.



**Figure 2.** Phylogenetic relationships among Hsp20 proteins from *Arabidopsis thaliana* (light yellow), *Spinacia oleracea* (light green), *Solanum lycopersicum* (red), *Oryza sativa* (light blue) and *Chenopodium quinoa* (pink). An unrooted phylogenetic tree was generated with Maximum Likelihood method with the matrix-based JTT model and empirical amino acid frequencies for heterogeneity estimation under a FreeRate model with six categories. The colors in the external stripe indicate the eight subcellular localizations predicted with MULocDeep meanwhile the color used for highlighting indicates plant species. The blue stars indicate permutation values.

### 3.5. Chromosomal Organization, Gene Duplications and Evolutionary Analysis

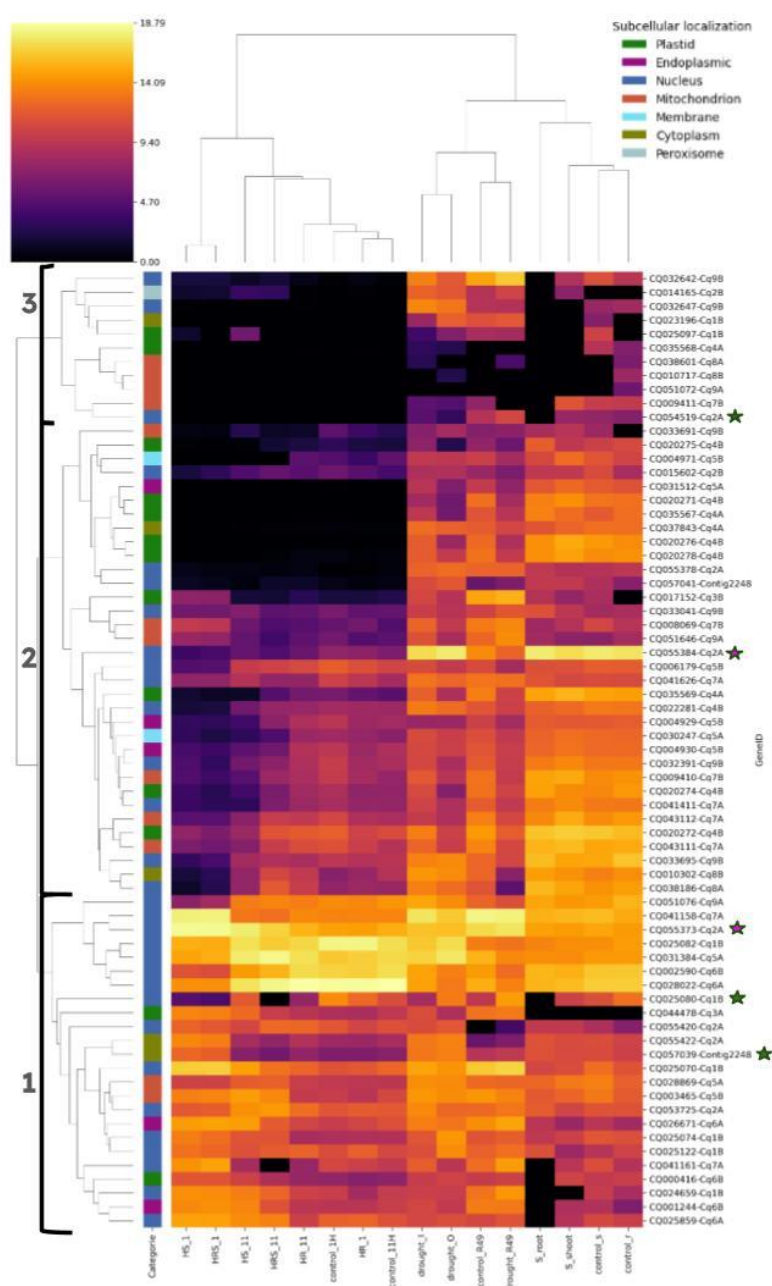
The chromosomal distribution and intergenic distances of CqHsp20 genes revealed several tandem gene arrays on chromosomes 1B, 2A, 4A, 4B, 5B, 7A, 7B, and Contig2248, suggesting their origin through local duplication events. The gene pairs CQ009410/CQ009411 on chromosome 7B and CQ043111/CQ043112 on chromosome 7A exhibited the highest sequence identity and similarity (84%) and were located contiguously. In contrast, the genes CQ055373, CQ055378, CQ055384, CQ055420, and CQ055422 on chromosome 2A, although not strictly contiguous, were positioned in proximity and displayed pairwise sequence identities ranging from 66.8% to 83.7%. Additional tandem-like pairs were identified on chromosome 1B, including CQ025070/CQ025074 (70.2% identity) and CQ025080/CQ025082 (approximately 65% identity). On chromosome 5B, the contiguous gene pair CQ006178/CQ006179 showed a relatively low sequence identity (27%), and the substantially longer genomic length of the latter suggests that it may represent an alternatively spliced isoform (Table S5). Notably, chromosome 4B harbors a tandem cluster of six CqHsp20 genes for which only the pair CQ020272/CQ020276 exhibited sequence identity and similarity above 80%. Overall, several tandemly arranged gene pairs displayed sequence identities below 80%. Mean Ka/Ks ratios for both tandem arrays and homeologous pairs were consistently < 1. Tandem arrays on chromosomes 5B, 4A, and 4B, each composed of more than two genes, exhibited comparatively higher Ka/Ks values (0.39–0.52), whereas the remaining tandem arrays showed lower ratios (0.14–0.18) (Table S6). Among the 19 putative CqHsp20 homeologous pairs analyzed (Table S7), only seven displayed Ka/Ks values above 0.50 (range: 0.52–0.84).

### 3.6. Expression Profiling of the CqHsp20 Genes in Different Abiotic Stresses

To assess CqHsp20 transcriptional activity under heat, drought, and salinity stress, raw reads from four RNA-seq experiments were mapped to the QQ74\_V2 genome and quantified using the described pipeline in the Section 2. The heat and drought datasets had 65–85% of reads assigned to exonic features, whereas salinity had 20–45%, with missing feature annotations as the primary source of unassigned reads (Table S10). The read assignment rate indicates the proportion of mapped reads that overlap (as set by `miOverlap`) with annotated exons used for gene-level counts (Figure 3S). Although some bias

is expected due to differences in read assignment, and despite TPM quantification being suitable for comparing samples of different origins and compositions [75,76], the analysis focuses on describing patterns of quantification, interpreted as relative transcriptional levels under stress response. The pattern may encompass different levels of omic information about CqHsp20, including differences in expression across tissues and protein target localization, as well as the specific responses of genes to a particular abiotic stress (Figure 3). At the top of the figure, hierarchical clustering discriminates between samples, separating them into heat (HS, HR, HRS and controls) and drought+salinity at the top level showing greater chaperone activity under salinity and drought than under heat. However, the last level (more similar samples) grouped samples from the same experiment. The hierarchical clustering on the left groups CqHsp20 genes into clusters 1, 2 and 3, which are then split into sub-clusters. Cluster 1 groups 17 genes (at bottom) with transcriptional activity under all conditions. Regarding heat stress, similar leaf transcriptional patterns were observed between the controls and treatments in which only the root was heated (HR\_1 and HR\_11); meanwhile, noticeable changes of transcriptional activity occurred when the shoot (HS) or both organs simultaneously (HRS) were exposed to heat. On day 1 of heating, the transcriptional profiles seem homogeneous; in contrast, more significant differences emerged by the end of the heating period (HS\_11 and HRS\_11). The lowest transcriptional activity is observed for CQ025080 (N), with a lower number of transcripts (4.3 - 4.7) when the shoot is heating compared to controls (11.1 - 14.1) and to the treatment in which only the root is heated (HR1, 12.6). However, when heat exposure is prolonged in the shoot (HS\_11) or root (HR\_11), transcription increases (11.7 and 7.5, respectively) but results in null activity when heat exposure is prolonged in both organs (HRS\_11) simultaneously (Figure 3S). In the same condition the CQ041161 (N) also showed null transcription; however, transcription remains active when only the shoot (HS\_11 = 8.4) or the root (HR\_11 = 7.6) is exposed to prolonged high temperature and it exhibits a pronounced transcriptional increase (18) when the shoot (HS\_1) or the shoot and root (HSR\_1) are exposed to heat for only one day. These genes also show null transcription in root tissue under salinity (vs control\_r = 12.8) but high transcriptional activity under drought (drought\_R49 = 14.8 vs control\_R49 = 9.5; drought\_O = 13.3 and drought\_I = 8.4). The ER CqHsp20, CQ001244 and CQ026671 show greater transcriptional activity in aerial parts (shoots and leaves) under heat and drought stress than under salinity, meanwhile CQ044478 (CP) shows null transcription under salinity, whereas it shows high transcriptional activity (13) under heat and drought. The CQ055422 (C) and CQ057039 (C) (which are likely a homologous pair), increase the transcription in the leaf tissue after the first day of heat exposure (between 12 and 13.8). In contrast, the levels in control samples or those exposed for prolonged periods decreased by around a half (between 6.0 and 8.9). A superior sub-cluster of Cluster 1, composed of 7 genes encoding N CqHsp20, shows high transcriptional activity across stresses, such as the genes CQ041158 and CQ055373 (18) in all tissues under heat and drought, and slightly decreases under salinity (15). In both, only CQ41158 shows a decrease in activity with prolonged heat exposure. On the contrary, the remaining genes, CQ025082, CQ031384, CQ002590, and CQ028022, show intense transcriptional activity in all tissues, with some differences between drought samples and a homogeneous profile across salinity samples, and, above all, reflect transcriptional activity changes after the first heat exposure. CqHsp20 genes belonging to this subcluster are likely good candidates for biological validation since they seem to play a central role in proteostasis across stresses. While cluster 1 appears to group ubiquitous CqHsp20s, clusters 2 and 3 show differences in transcriptional activity between stresses. Cluster 2 grouped a diverse set of CqHsp20, splitting into two sub-clusters. The one below, composed of 18 genes, shows higher transcriptional activity under salinity (without noticeable differences between tissues exposed or not), less activity in root and shoot tissues under drought, and the least activity in leaves after one day of heat exposure (HS\_1 and HRS\_1). Among this sub-cluster, two ER CqHsp20 (CQ004929 and CQ004930, also part of a tandem) show a different behaviour with respect to the two previously described, reflecting a greater chaperone activity under salinity and drought than under heat. In particular, CQ055384 (N) showed the highest transcriptional activity under salinity (17), a mild response to drought, and lower activity under heat (5). The middle sub-cluster, composed of both M CQ008069 and CQ51646, CQ033041 (N)

and CQ017152 (CP), shows greater activity under drought conditions than the others, with a slight increase as the first response to heat exposure. The third sub-cluster, composed of 12 CqHsp20, shows a low or null transcription in leaf tissue under heat but a response against drought and salinity. The nuclear CQ055378 (2A), CQ057041 (Contig2248) (putatively homologous) and CQ015602 (2B) barely respond under heat exposure; meanwhile, three CqHsp20 members (CP, M and membrane members) showed expression in controls or when only the root is heated. The transcriptional activity for tandemly duplicated CQ020278, CQ020276, CQ020271 (in chromosome 4B) and their homologous prevail in root and shoot tissue under drought and salinity rather than heat. The CqHsp20 genes of 4B tandem are good candidates for validating function and specialization. Cluster 3 contains 11 genes, none of which are transcribed in leaf or root tissue under heat or salinity conditions. The cluster is composed of four mitochondrial, three nuclear, two plastidial, one cytoplasmic and the only CqHsp20, whose peptide localization is the peroxisoma. Among them, significant transcriptional activity is observed under drought; however, this group appears to be the least responsive to stress.



**Figure 3.** Heatmap and clustering of CqHsp20 quantification profile across abiotic stress condition. Heat experiment quantifies transcription in leaf tissue when the root (HR), shoot (HS), or both (HRS) are exposed to

high temperature (30 °C for the root and 35 °C for the shoot) for 1 day (1) or 11 days (11) during the flowering stage. Each moment has the appropriate control at 23 °C (control\_1H; control\_11H). Drought experiment (D1) contained the quantification of transcription in root tissue for two quinoa varieties: Ollague (drought\_O) and Ingapirca (drought\_I) after exposure to drought stress, meanwhile the other drought dataset (D2) shows the quantification of transcription in root and shoot tissues measured on R49 cultivar (drought\_R49) with the respectively control (control\_R49). Salinity experiment quantifies transcription in root (S\_root) and shoot (S\_shoot) tissues exposed to salinity conditions, along with their respective controls (control\_s; control\_r). Green stars indicate CqHsp20 genes validated according to [5]; particularly pink stars indicate genes CQ055373 and CQ055384, identified both under the same ID: XP\_021732378.1.

#### 4. Discussion

Plants face several abiotic stresses by synthesis of a wide variety of HSPs (from 20 to 100 kDa) resulting in a concerted defence response strategy. HSPs are essential to confer tolerance by stabilizing cellular components, preventing protein aggregation and misfolding, and maintaining cellular homeostasis [77,78]. Given the role of the Hsp20 subfamily in plant species [79], a large number and diversity of genes are expected, as observed in *Chenopodium quinoa*. A total of 69 genomic sequences were identified as members of the Hsp20 subfamily, distributed across the nine chromosomes of each subgenome (A and B). These sequences are organized primarily in homoeologous pairs, as expected due to allotetraploidization and paralogy, with the majority of the arrangements located in the B subgenome. The presence of more tandem CqHsp20 genes in sub-genome B is consistent with the genome expansion described for quinoa [32]. Carried distinct subgenomes, making quinoa an interesting model for studying subgenome evolution, interactions, and divergence. Besides, whole-genome duplication (WGD) can lead to gene loss (pseudogenization), subfunctionalization, or neofunctionalization, while tandemly arranged genes often exhibit divergent expression patterns [16,80]. Anyway, duplicated genes improve resilience to environmental changes by providing a buffer against stress. The observed gene distribution suggests a well-conserved evolutionary pattern within the species, with some clusters including paralogs, indicating that gene duplication contributes to the expansion of the CqHsp20 family, as was also the case for CqHsp70 and CqWRKY in quinoa [18,48] and other model plants [17,22,81]. As shown in Figure 3, tandemly duplicated CQ009410 (cluster 2) and CQ009411 (cluster 3) on chromosome 7B, show differences in transcriptional quantification and response pattern. On the other hand, the paralogs and homologs located on chromosomes 4A and 4B (grouped in cluster 2) showed differences in transcriptional activity across stresses (Figure 3S), moreover the tandemly duplicated CQ020272 showed high transcriptional activity under heat. Despite divergent transcriptional responses, duplicated genes exhibit Ka/Ks ratios below 1 (Tables S6 and S7), indicating purifying selection as was observed for *Phaseolus vulgaris* [19]. This suggests that functional diversification is driven mainly by regulatory divergence rather than amino acid sequence evolution. The phylogenetic analysis clustered the CqHsp20 protein sequences according to their subcellular target sites, indicating the presence of appropriate organelle-targeting signals [3], even when we included the hsp20 from other model plants (Figure 2). The intron-exon structure and pattern of conserved motifs in protein sequences (Figure 2S) supported the CqHsp20 clustering, as demonstrated in more recent genomic studies of Hsp20 gene family in pulse crops [19,20,82]. As the chaperone role in stress-responsiveness is intermediate among other HSPs, it is expected that their function is not compartment-limited, as we observed in the hierarchical clustering of 3. Although nuclear Hsp20 dominates cluster 1, it also contains a smaller number of CqHsp20 peptides predicted to localize to CP, C, M, and ER, indicating heterogeneity. Genes within this cluster show transcriptional responsiveness to heat, salinity, and drought across a variety of tissues, reflecting their ubiquitous role in maintaining cellular proteostasis. As we used genomic data from the annotated Hsp20 in the QQ74\_V2 reference, the structures of some members may require curation to distinguish artefacts from biological data, as we mentioned previously regarding the absence of UTRs. Nevertheless, a simple genic structure, as observed in intronless or at least in the presence of only one intron, as in the majority of genes, was in accordance

with the function of Hsp20. In response to abiotic stress, the accumulation of Hsp20 protein is an early response, given its simple gene structure [9]. Currently, other reference genomes of *Chenopodium* from ATGC complex, including diploids, exhibit high-quality assembly [83–86] and can be appropriate resources for curating genomic information. Quinoa is well known for its intrinsic adaptation to diverse abiotic stresses, such as salinity and drought [36,37], but high temperatures negatively impact yield [42,87]. Little research has been conducted at the genetic and genomic levels to understand the abiotic stress response [88], particularly to elucidate the mechanisms underlying intrinsic quinoa tolerance. The integration of transcriptomic data enables us to analyse the transcriptional patterns of CqHsp20 genes that exhibit differential responses across abiotic stress conditions in different tissues. A higher number of CqHsp20 genes present transcriptional activity across drought and salinity than heat stress in coherence with prediction of TFBS in promoter region; a higher quantity and diversity of NAC, bZIP and then bHLH motifs than HSE and WRKY showing a potentially greater possibility of transcription triggered by the response to osmotic stress (reflected in drought and salinity) rather than to oxidative stress [9]. The promoter architecture provides a mechanistic link between external stress perception and transcriptional activation by defining candidate transcription factor binding sites [89], however the presence of a *cis*-element does not guarantee transcription factor binding, nor does it ensure transcriptional activation. A more comprehensive analysis of co-expression regulatory networks is necessary to complement this analysis. The hierarchical clustering of CqHsp20 genes in 3 reveals a structured, stress-dependent organization of transcriptional responses, indicating functional diversification within the gene family. As we mentioned before, cluster 1 represents a core of CqHsp20 reflecting a behaviour as general stress-responsive chaperones, as they are transcriptionally active across multiple tissues and under different abiotic stress conditions. However, the heterogeneous organ-specific expression patterns observed under heat and salinity conditions indicate that even within this broadly responsive cluster, individual genes are subject to fine regulatory control. For example, the suppression or loss of transcription of specific nuclear CqHsp20s, such as CQ025080 and CQ041161, in root tissue under prolonged heat exposure or salinity conditions suggests stress and tissue-specific functional specialization. Clusters 2 and 3 exhibit, in general, more specialized transcriptional profiles; furthermore, three sub-clusters in cluster 2 showed differences among stresses (as fully explained in 3.6). The pattern showed an enrichment in genes preferentially responsive to salinity and drought stress, consistent with roles in ionic or osmotic stress adaptation. Within these conditions, there are no apparent differences among tissues. Meanwhile, cluster 3 genes show limited heat and salinity responsiveness but remain active under drought. The genes grouped into clusters 2 and 3 indicate condition-specific deployment of Hsp20 functions, including the most organelle-specific CqHsp20s. Twelve genes, grouped into clusters 2 and 3, belonging to the Plastid and Mitochondrial Hsp20 families, showed reduced transcription under heat stress compared to drought and salinity. Nonetheless, it has been noted that an increase in Hsp20 within organelles serves as a thermotolerance response in other plant species [10]. In order to add new information regarding CqHsp20 reported in other studies, we contrasted the transcriptional pattern between CQ055373 and CQ055384, both identified in chromosome 2A and aligned with high confidence to the 161 amino acids length protein (XP\_021732378.1: 18.3 kDa class I heat shock protein [*Chenopodium quinoa*]) identified by [5] in the *C. quinoa* proteome, suggests that both genes may exist. The gene CQ055373 (cluster 1) shows high transcriptional activity across all conditions, whereas CQ055384 (cluster 2) shows a differential pattern, with a different response to salinity and drought compared to heat. Schmöckel et al. [43] used transcriptomic and genomic data in an integrated manner identifying the genes CQ025074 and CQ025122 (cluster 1, Figure 3) located in chromosome 1B (putative paralogous), and CQ004971 and CQ035569 genes, located on chromosomes 5B and 4A, respectively (cluster 2, 3) as unique for *C. quinoa*, therefore proposed as salinity tolerance candidate. The results of this study showed similar transcriptional activity across all conditions for those in chromosome 1B, in contrast to the others, which showed high transcriptional activity mainly under salinity. Omic studies of *C. quinoa* have emerged, enabling progress in understanding stress resistance through data integration. Considering the limitations of an

*in silico* descriptive analysis, a group of 24 CqHsp20 genes that showed transcriptional activity across all conditions might be considered ubiquitous Hsp20 (cluster 1), whereas the remaining members showed differential activity, with a strong capacity to respond to osmotic stresses caused by drought and salinity. Exposure of aerial parts to high temperatures results in greater conditioning of Hsp20 expression. The transcriptome derived from the high-temperature exposure experiment yielded interesting results regarding plant yield. Plants with heated shoots lost 60–85% yield compared with control plants, reflected in lower fruit production and fewer seeds per plant. Besides, plants with heated shoots had delayed maturity and greater non-reproductive shoot biomass. In contrast, plants with both heated roots and shoots produced higher yields from panicles that had escaped the heat than did the control plants, suggesting that quinoa uses an avoidance strategy to survive heat [42]. In cross-observation of the transcriptional pattern, treatments with shoot and root prolonged exposure to high temperatures, or only root together with controls, showed sustained transcriptional activity of CqHsp20, unlike treatments where only the aerial parts were exposed to high temperatures for a single day. The adverse effects of high temperatures on cultivated *C. quinoa* limit the crop's expansion to the most arid countries and increase its vulnerability to climate change globally. The most current approaches to improve this aspect are based on the use of wild relatives as a source of introgression of genetic bases that confer heat tolerance [90]. Therefore, a phylogenetic study of genes encoding Hsp20 in wild relatives could provide insights into the evolution of the gene family and its theoretically differential response to high temperatures.

## 5. Conclusions

Given the increasing relevance of quinoa as a climate-resilient crop, the characterization of the CqHsp20 gene family using an integrative omic approach provides a valuable resource for applied breeding and biotechnological strategies. Phylogenetic relationships, together with conserved motif architecture, indicate strong structural conservation within subfamilies, supporting functional constraints across evolutionary lineages. The expansion of the family is associated with the allopolyploid origin of the quinoa genome, as evidenced by well-defined homeologous pairs and multiple tandemly duplicated genes, which might contribute to genomic variation and to the adaptation to extreme environmental conditions—key attributes of quinoa. Stress-responsive CqHsp20 genes, particularly those showing stable or condition-specific transcriptional patterns across heat, drought, and salinity, represent promising candidates for the development of molecular markers associated with abiotic stress tolerance, as was done in another species. This study's data lay the groundwork for characterizing CqHsp20 family members, deepen our understanding of quinoa's resilience to stress, and provide valuable guidance for molecular marker design as breeding tools or for exploring and developing biotechnological approaches to improve heat-stress tolerance.

**Supplementary Materials:** The following supporting information can be downloaded at: [https://drive.google.com/drive/folders/1v\\_zEkmgvBrmSfwSxpSqLNINwskedI52Y?usp=drive\\_link](https://drive.google.com/drive/folders/1v_zEkmgvBrmSfwSxpSqLNINwskedI52Y?usp=drive_link).

**Author Contributions:** SCT: wrote the original draft, made the data acquisition and interpretation. DPA: co-wrote the original draft, contributed to the interpretation of formal analysis and conceptualization. GT: co-work in bioinformatic approach, manuscript review and editing. GRP: manuscript review and editing, conceptualization, final approval of the version submitted.

**Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author, SM Costa-Tártara, upon request. Repository of genomic and transcriptomic data integration: [https://github.com/YellowSabri/Quinoa\\_sHSPs/](https://github.com/YellowSabri/Quinoa_sHSPs/) (available once the paper is accepted for publication).

**Acknowledgments:** We thank CIDETIC (Center for Research, Teaching and Extension in Information and Communication Technologies) an individual unit of the National University of Luján (<http://cidetic.unlu.edu.ar/>) for providing computing resources.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Lichtenthaler, H.K. The Stress Concept in Plants: An Introduction. *Annals NY Acad Sci* **1998**, *851*, 187–198. <https://doi.org/10.1111/j.1749-6632.1998.tb08993.x>.
2. Shao, H.B.; Guo, Q.J.; Chu, L.Y.; Zhao, X.N.; Su, Z.L.; Hu, Y.C.; Cheng, J.F. Understanding molecular mechanism of higher plant plasticity under abiotic stress. *Colloids and Surfaces B: Biointerfaces* **2007**, *54*, 37–45. <https://doi.org/10.1016/j.colsurfb.2006.07.002>.
3. Waters, E.R. The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany* **2013**, *64*, 391–403. <https://doi.org/10.1093/jxb/ers355>.
4. Carra, S.; Alberti, S.; Benesch, J.L.; Boelens, W.; Buchner, J.; Carver, J.A.; Cecconi, C.; Ecroyd, H.; Gusev, N.; Hightower, L.E.; et al. Small heat shock proteins: multifaceted proteins with important implications for life. *Cell Stress and Chaperones* **2019**, *24*, 295–308. <https://doi.org/10.1007/s12192-019-00979-z>.
5. Rizzo, A.J.; Palacios, M.B.; Vale, E.M.; Zelada, A.M.; Silveira, V.; Burrieza, H.P. Snapshot of four mature quinoa (*Chenopodium quinoa*) seeds: a shotgun proteomics analysis with emphasis on seed maturation, reserves and early germination. *Physiol Mol Biol Plants* **2023**, *29*, 319–334. <https://doi.org/10.1007/s12298-023-01295-8>.
6. Arce, D.P.; De Las Rivas, J.; Pratta, G.R. Interactomic analysis of the sHSP family during tomato fruit ripening. *Plant Gene* **2020**, *21*. <https://doi.org/10.1016/j.plgene.2019.100208>.
7. Cacchiarelli, P.; Arce, D.P.; Tapia, E.; Pratta, G. Structural and functional analysis of two sHSP subfamilies in tomato ripening. *Plant Gene* **2021**, *27*, 100297. <https://doi.org/10.1016/J.PLGENE.2021.100297>.
8. Goytia Bertero, V.; Cacchiarelli, P.; Pratta, G.R.; Arce, D.P. Comparative and integrative omic analysis focused on chaperones and interactors in a cultivated and an exotic tomato at different fruit ripening stages. *Plant Gene* **2024**, *37*, 100448. <https://doi.org/10.1016/j.plgene.2024.100448>.
9. Sun, W.; Van Montagu, M.; Verbruggen, N. Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **2002**, *1577*, 1–9. [https://doi.org/10.1016/S0167-4781\(02\)00417-7](https://doi.org/10.1016/S0167-4781(02)00417-7).
10. Al-Whaibi, M.H. Plant heat-shock proteins: A mini review. *Journal of King Saud University - Science* **2011**, *23*, 139–150. <https://doi.org/10.1016/j.jksus.2010.06.022>.
11. Scharf, K.D.; Siddique, M.; Vierling, E. The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing alfa-crystallin domains (Acid proteins). *Cell Stress Society International* **2001**, *6*, 225 – 237. [https://doi.org/10.1379/1466-1268\(2001\)006<0225:tefoat>2.0.co;2](https://doi.org/10.1379/1466-1268(2001)006<0225:tefoat>2.0.co;2).
12. Waters, E.R. The molecular evolution of the small heat-shock proteins in plants. *Genetics* **1995**, *141*, 785–795. <https://doi.org/10.1093/genetics/141.2.785>.
13. Waters, E.R.; Aebermann, B.D.; Sanders-Reed, Z. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress and Chaperones* **2008**, *13*, 127–142. <https://doi.org/10.1007/s12192-008-0023-7>.
14. Haslbeck, M.; Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *Journal of Molecular Biology* **2015**, *427*, 1537–1548. <https://doi.org/10.1016/j.jmb.2015.02.002>.
15. Lallemand, T.; Leduc, M.; Landès, C.; Rizzon, C.; Lerat, E. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. *Genes* **2020**, *11*, 1046. <https://doi.org/10.3390/genes11091046>.
16. Panchy, N.; Lehti-Shiu, M.; Shiu, S.H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. <https://doi.org/10.1104/pp.16.00523>.
17. Krsticevic, F.J.; Arce, D.P.; Ezpeleta, J.; Tapia, E. Tandem Duplication Events in the Expansion of the Small Heat Shock Protein Gene Family in *Solanum lycopersicum* (cv. Heinz 1706). *G3 Genes | Genomes | Genetics* **2016**, *6*, 3027–3034. <https://doi.org/10.1534/g3.116.032045>.
18. Liu, J.; Wang, R.; Liu, W.; Zhang, H.; Guo, Y.; Wen, R. Genome-Wide Characterization of Heat-Shock Protein 70s from *Chenopodium quinoa* and Expression Analyses of Cqhs70s in Response to Drought Stress. *Genes* **2018**, *9*, 15. ISBN: 8613935230446, <https://doi.org/10.3390/genes9020035>.
19. De Souza Resende, J.S.; Dos Santos, T.B.; Souza, S.G.H.D. Small heat shock protein (Hsp20) gene family in *Phaseolus vulgaris* L.: Genome-wide identification, evolutionary and expression analysis. *Plant Gene* **2022**, *31*, 100370. <https://doi.org/10.1016/j.plgene.2022.100370>.
20. Ramakrishna, G.; Kaur, P.; Singh, A.; Yadav, S.S.; Sharma, S.; Singh, N.K.; Gaikwad, K. Comparative transcriptome analyses revealed different heat stress responses in pigeonpea (*Cajanus cajan*) and its crop wild relatives. *Plant Cell Rep* **2021**, *40*, 881–898. <https://doi.org/10.1007/s00299-021-02686-5>.

21. Do, J.M.; Kim, H.J.; Shin, S.Y.; Park, S.I.; Kim, J.J.; Yoon, H.S. OsHSP 17.9, a Small Heat Shock Protein, Confers Improved Productivity and Tolerance to High Temperature and Salinity in a Natural Paddy Field in Transgenic Rice Plants. *Agriculture* **2023**, *13*, 931. <https://doi.org/10.3390/agriculture13050931>.
22. Guo, M.; Liu, J.H.; Lu, J.P.; Zhai, Y.F.; Wang, H.; Gong, Z.H.; Wang, S.B.; Lu, M.H. Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Front. Plant Sci.* **2015**, *6*. <https://doi.org/10.3389/fpls.2015.00806>.
23. Biondi, E.; Ruiz, K.B.; Martinez, E.A.; Zurita-Silva, A.; Orsini, F.; Antognoni, F.; Dinelli, G.; Marotti, I.; Giaquinto, G.; Maldonado, S.; et al. Tolerance to saline conditions. In *Estado del arte de la quinua en el mundo en 2013*; FAO and CIRAD: Santiago de Chile, Chile; Montpellier, Francia, 2014.
24. Ruiz, K.B.; Biondi, S.; Osés, R.; Acuña-Rodríguez, I.S.; Antognoni, F.; Martínez-Mosqueira, E.A.; Coulibaly, A.; Canahua-Murillo, A.; Pinto, M.; Zurita-Silva, A.; et al. Quinoa biodiversity and sustainability for food security under climate change. A review. *Agronomy for Sustainable Development* **2014**, *34*, 349–359. <https://doi.org/10.1007/s13593-013-0195-0>.
25. Pickersgill, B. Domestication of plants in the Americas: Insights from Mendelian and molecular genetics. *Annals of Botany* **2007**, *100*, 925–940. <https://doi.org/10.1093/aob/mcm193>.
26. Mu, H.; Xue, S.; Sun, Q.; Shi, J.; Zhang, D.; Wang, D.; Wei, J. Research Progress of Quinoa Seeds (Chenopodium quinoa Willd.): Nutritional Components, Technological Treatment, and Application. *Foods* **2023**, *12*, 2087. <https://doi.org/10.3390/foods12102087>.
27. Pereira, E.; Encina-Zelada, C.; Barros, L.; Gonzales-Barron, U.; Cadavez, V.; C.F.R. Ferreira, I. Chemical and nutritional characterization of Chenopodium quinoa Willd (quinoa) grains: A good alternative to nutritious food. *Food Chemistry* **2019**, *280*, 110–114. <https://doi.org/10.1016/j.foodchem.2018.12.068>.
28. Jellen, E.N.; Jarvis, D.E.; Benet-Pierce, N.; Maughan, P.J. *Botanical Context for Domestication in North America*; Springer International Publishing, 2021. [https://doi.org/10.1007/978-3-030-65237-1\\_3](https://doi.org/10.1007/978-3-030-65237-1_3).
29. Wilson, H.; Manhart, J. Crop/weed gene flow:Chenopodium quinoa Willd. andC. berlandieri Moq. *Theoretical and Applied Genetics* **1993**, *86*, 642–648. <https://doi.org/10.1007/BF00838721>.
30. Jarvis, D.E.; Ho, Y.S.; Lightfoot, D.J.; Schmöckel, S.M.; Li, B.; Borm, T.J.A.; Ohyanagi, H.; Mineta, K.; Michell, C.T.; Saber, N.; et al. The genome of Chenopodium quinoa. *Nature* **2017**, *542*, 307–312. <https://doi.org/10.1038/nature21370>.
31. Maughan, P.J.; Chaney, L.; Lightfoot, D.J.; Cox, B.J.; Tester, M.; Jellen, E.N.; Jarvis, D.E. Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa Willd.). *Scientific Reports* **2019**, *9*, 1–11. <https://doi.org/10.1038/s41598-018-36693-6>.
32. Rey, E.; Maughan, P.J.; Maumus, F.; Lewis, D.; Wilson, L.; Fuller, J.; Schmöckel, S.M.; Jellen, E.N.; Tester, M.; Jarvis, D.E. A chromosome-scale assembly of the quinoa genome provides insights into the structure and dynamics of its subgenomes. *Commun Biol* **2023**, *6*, 1263. <https://doi.org/10.1038/s42003-023-05613-4>.
33. Young, L.A.; Maughan, P.J.; Jarvis, D.E.; Hunt, S.P.; Warner, H.C.; Durrant, K.K.; Kohlert, T.; Curti, R.N.; Bertero, D.; Filippi, G.A.; et al. A chromosome-scale reference of *Chenopodium watsonii* helps elucidate relationships within the North American A-genome *Chenopodium* species and with quinoa. *The Plant Genome* **2023**, *16*, e20349. <https://doi.org/10.1002/tpg2.20349>.
34. Alandia, G.; Rodriguez, J.P.; Jacobsen, S.E.E.; Bazile, D.; Condori, B. Global expansion of quinoa and challenges for the Andean region. *Global Food Security* **2020**, *26*, 100429. <https://doi.org/10.1016/j.gfs.2020.100429>.
35. Bazile, D.; Jacobsen, S.E.; Verniau, A. The Global Expansion of Quinoa: Trends and Limits. *Front. Plant Sci.* **2016**, *7*. <https://doi.org/10.3389/fpls.2016.00622>.
36. Hinojosa, L.; González, J.A.; Barrios-Masias, F.H.; Fuentes, F.; Murphy, K.M. Quinoa abiotic stress responses: A review. *Plants* **2018**, *7*. <https://doi.org/10.3390/plants7040106>.
37. Schmöckel, S.M., Ed. *The Quinoa Genome*; Compendium of Plant Genomes, Springer International Publishing: Cham, 2021. <https://doi.org/10.1007/978-3-030-65237-1>.
38. García-Parra, M.; Zurita-silva, A.; Stechauner-rohringer, R.; Roa-acosta, D.; Jacobsen, S.E.E.; García-Parra, M.; Zurita-silva, A.; Stechauner-rohringer, R.; Roa-acosta, D.; Jacobsen, S.E.E. Quinoa (Chenopodium quinoa Willd.) and its relationship with agroclimatic characteristics: A Colombian perspective. *Chilean Journal of Agricultural Research* **2020**, *80*, 290–302. <https://doi.org/10.4067/S0718-58392020000200290>.
39. Maughan, P.J.; Turner, T.B.; Coleman, C.E.; Elzinga, D.B.; Jellen, E.N.; Morales, J.A.; Udall, J.A.; Fairbanks, D.J.; Bonifacio, A. Characterization of Salt Overly Sensitive 1 (SOS1) gene homoeologs in quinoa (Chenopodium quinoa Willd.). *Genome* **2009**, *52*, 647–657. <https://doi.org/10.1139/G09-041>.

40. Ruiz-Carrasco, K.; Antognoni, F.; Coulibaly, A.K.; Lizardi, S.; Covarrubias, A.; Martínez, E.A.; Molina-Montenegro, M.A.; Biondi, S.; Zurita-Silva, A. Variation in salinity tolerance of four lowland genotypes of quinoa (*Chenopodium quinoa* Willd.) as assessed by growth, physiological traits, and sodium transporter gene expression. *Plant Physiology and Biochemistry* **2011**, *49*, 1333–1341. <https://doi.org/10.1016/j.plaphy.2011.08.005>.
41. Morales, A.; Zurita-Silva, A.; Maldonado, J.; Silva, H. Transcriptional Responses of Chilean Quinoa (*Chenopodium quinoa* Willd.) Under Water Deficit Conditions Uncover ABA-Independent Expression Patterns. *Frontiers in Plant Science* **2017**, *8*, 1–13. <https://doi.org/10.3389/fpls.2017.00216>.
42. Tovar, J.C.; Quillatupa, C.; Callen, S.T.; Castillo, S.E.; Pearson, P.; Shamin, A.; Schuhl, H.; Fahlgren, N.; Gehan, M.A. Heating quinoa shoots results in yield loss by inhibiting fruit production and delaying maturity. *The Plant Journal* **2020**, *102*, 1058–1073. <https://doi.org/10.1111/tpj.14699>.
43. Schmöckel, S.M.; Lightfoot, D.J.; Razali, R.; Tester, M.; Jarvis, D.E. Identification of Putative Transmembrane Proteins Involved in Salinity Tolerance in *Chenopodium quinoa* by Integrating Physiological Data, RNAseq, and SNP Analyses. *Front. Plant Sci.* **2017**, *8*, 1023. <https://doi.org/10.3389/fpls.2017.01023>.
44. Li, F.; Guo, X.; Liu, J.; Zhou, F.; Liu, W.; Wu, J.; Zhang, H.; Cao, H.; Su, H.; Wen, R. Genome-Wide Identification, Characterization, and Expression Analysis of the NAC Transcription Factor in *Chenopodium quinoa*. *Genes* **2019**, *10*, 500. <https://doi.org/10.3390/genes10070500>.
45. Li, F.; Liu, J.; Guo, X.; Yin, L.; Zhang, H.; Wen, R. Genome-wide survey, characterization, and expression analysis of bZIP transcription factors in *Chenopodium quinoa*. *BMC plant biology* **2020**, *20*, 405. <https://doi.org/10.1186/s12870-020-02620-z>.
46. Tashi, G.; Zhan, H.; Xing, G.; Chang, X.; Zhang, H.; Nie, X.; Ji, W. Genome-Wide Identification and Expression Analysis of Heat Shock Transcription Factor Family in *Chenopodium quinoa* Willd. *Agronomy* **2018**, *8*, 103. <https://doi.org/10.3390/agronomy8070103>.
47. Wu, Q.; Bai, X.; Zhao, W.; Shi, X.; Xiang, D.; Wan, Y.; Wu, X.; Sun, Y.; Zhao, J.; Peng, L.; et al. Investigation into the underlying regulatory mechanisms shaping inflorescence architecture in *Chenopodium quinoa*. *BMC Genomics* **2019**, *20*, 1–20. <https://doi.org/10.1186/s12864-019-6027-0>.
48. Yue, H.; Chang, X.; Zhi, Y.; Wang, L.; Xing, G.; Song, W.; Nie, X. Evolution and identification of the WRKY gene family in Quinoa (*Chenopodium quinoa*). *Genes* **2019**, *10*. <https://doi.org/10.3390/genes10020131>.
49. Golicz, A.A.; Steinfort, U.; Arya, H.; Singh, M.B.; Bhalla, P.L. Analysis of the quinoa genome reveals conservation and divergence of the flowering pathways. *Funct Integr Genomics* **2020**, *20*, 245–258. <https://doi.org/10.1007/s10142-019-00711-1>.
50. Patiranage, D.S.; Asare, E.; Maldonado-Taípe, N.; Rey, E.; Emrani, N.; Tester, M.; Jung, C. Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions. *Plant Cell and Environment* **2021**, *44*, 1–15. <https://doi.org/10.1111/pce.14071>.
51. Patiranage, D.S.R.; Rey, E.; Emrani, N.; Wellman, G.; Schmid, K.; Schmöckel, S.M.; Tester, M.; Jung, C. Genome-wide association study in the pseudocereal quinoa reveals selection pattern typical for crops with a short breeding history. *eLife* **2022**, pp. 1–22. <https://doi.org/10.1101/2020.12.03.410050>.
52. Paysan-Lafosse, T.; Andreeva, A.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.; Salazar, G.; Bileschi, M.; Llinares-López, F.; Meng-Papaxanthos, L.; et al. The Pfam protein families database: embracing AI/ML. *Nucleic Acids Research* **2025**, *53*, D523–D534, [<https://academic.oup.com/nar/article-pdf/53/D1/D523/60667756/gkae997.pdf>]. <https://doi.org/10.1093/nar/gkae997>.
53. The UniProt Consortium.; Bateman, A.; Martin, M.J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; Bye-A-Jee, H.; et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2023**, *51*, D523–D531. <https://doi.org/10.1093/nar/gkac1052>.
54. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, 2005; pp. 571–607. <https://doi.org/10.1385/1-59259-890-0:571>.
55. Jiang, Y.; Wang, D.; Yao, Y.; Eubel, H.; Künzler, P.; Møller, I.M.; Xu, D. MULocDeep: A deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. *Computational and Structural Biotechnology Journal* **2021**, *19*, 4825–4839. <https://doi.org/10.1016/j.csbj.2021.08.027>.
56. Jiang, Y.; Jiang, L.; Akhil, C.S.; Wang, D.; Zhang, Z.; Zhang, W.; Xu, D. MULocDeep web service for protein localization prediction and visualization at subcellular and suborganellar levels. *Nucleic Acids Res* **2023**, *51*, W343–W349. <https://doi.org/10.1093/nar/gkad374>.

57. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. <https://doi.org/10.1093/bioinformatics/btu817>.
58. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Research* **2015**, *43*, W39–W49. <https://doi.org/10.1093/nar/gkv416>.
59. O'Malley, R.; Huang, S.s.; Song, L.; Lewsey, M.; Bartlett, A.; Nery, J.; Galli, M.; Gallavotti, A.; Ecker, J. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **2016**, *165*, 1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038>.
60. Xue, G.; Fan, Y.; Zheng, C.; Yang, H.; Feng, L.; Chen, X.; Yang, Y.; Yao, X.; Weng, W.; Kong, L.; et al. bHLH transcription factor family identification, phylogeny, and its response to abiotic stress in *Chenopodium quinoa*. *Front Plant Sci* **2023**, *14*, 1171518. Place: Switzerland, <https://doi.org/10.3389/fpls.2023.1171518>.
61. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004**, *32*, 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
62. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **2020**, *37*, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
63. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **2017**, *14*, 587–589. <https://doi.org/10.1038/nmeth.4285>.
64. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **2018**, *35*, 518–522. <https://doi.org/10.1093/molbev/msx281>.
65. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **2000**, *16*, 276–277. Place: England, [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2).
66. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **2010**, *8*, 77–80. Place: England, [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3).
67. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **2006**, *34*, W609–612. Place: England, <https://doi.org/10.1093/nar/gkl315>.
68. Li, W.H.; Wu, C.I.; Luo, C.C. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol* **1984**, *21*, 58–71. Place: Germany, <https://doi.org/10.1007/BF02100628>.
69. Raney, J.A.; Reynolds, D.J.; Elzinga, D.B.; Page, J.; A. Udall, J.; Jellen, E.N.; Bonfacio, A.; Fairbanks, D.J.; Maughan, P.J. Transcriptome Analysis of Drought Induced Stress in *Chenopodium quinoa*. *AJPS* **2014**, *05*, 338–357. <https://doi.org/10.4236/ajps.2014.53047>.
70. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. Place: England, <https://doi.org/10.1093/bioinformatics/btu170>.
71. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **2019**, *37*, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
72. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCftools. *Gigascience* **2021**, *10*. Place: United States, <https://doi.org/10.1093/gigascience/giab008>.
73. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
74. Liao, Y.; Smyth, G.K.; Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **2019**, *47*, e47. Place: England, <https://doi.org/10.1093/nar/gkz114>.
75. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* **2016**, *17*, 13. <https://doi.org/10.1186/s13059-016-0881-8>.
76. Zhao, Y.; Li, M.C.; Konaté, M.M.; Chen, L.; Das, B.; Karlovich, C.; Williams, P.M.; Evrard, Y.A.; Doroshow, J.H.; McShane, L.M. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *J Transl Med* **2021**, *19*, 269. Place: England, <https://doi.org/10.1186/s12967-021-02936-w>.

77. Kumar, P.; Paul, D.; Jhahriya, S.; Kumar, R.; Dutta, S.; Siwach, P.; Das, S. Understanding heat-shock proteins' abundance and pivotal function under multiple abiotic stresses. *Journal of Plant Biochemistry and Biotechnology* **2024**, *33*, 492–513. <https://doi.org/10.1007/s13562-024-00932-x>.
78. Mukhopadhyay, R.; Boro, P.; Karmakar, K.; Pradhan, P.; Saha Chowdhury, R.; Das, B.; Mandal, R.; Kumar, D. Advances in the understanding of heat shock proteins and their functions in reducing abiotic stress in plants. *Journal of Plant Biochemistry and Biotechnology* **2024**, *33*, 474–491. <https://doi.org/10.1007/s13562-024-00895-z>.
79. Waters, E.R.; Vierling, E. Plant small heat shock proteins - evolutionary and functional diversity. *New Phytol* **2020**, *227*, 24–37. Place: England, <https://doi.org/10.1111/nph.16536>.
80. Glover, N.M.; Redestig, H.; Dessimoz, C. Homoeologs: What Are They and How Do We Infer Them? *Trends in Plant Science* **2016**, *21*, 609–621. <https://doi.org/10.1016/j.tplants.2016.02.005>.
81. Yu, J.; Cheng, Y.; Feng, K.; Ruan, M.; Ye, Q.; Wang, R.; Li, Z.; Zhou, G.; Yao, Z.; Yang, Y.; et al. Genome-Wide Identification and Expression Profiling of Tomato Hsp20 Gene Family in Response to Biotic and Abiotic Stresses. *Front. Plant Sci.* **2016**, *7*. <https://doi.org/10.3389/fpls.2016.01215>.
82. Huang, J.; Hai, Z.; Wang, R.; Yu, Y.; Chen, X.; Liang, W.; Wang, H. Genome-wide analysis of HSP20 gene family and expression patterns under heat stress in cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* **2022**, *13*, 968418. <https://doi.org/10.3389/fpls.2022.968418>.
83. Jaggi, K.E.; Krak, K.; Štorchová, H.; Mandák, B.; Marcheschi, A.; Belyayev, A.; Jellen, E.N.; Sproul, J.; Jarvis, D.; Maughan, P.J. A pangenome reveals LTR repeat dynamics as a major driver of genome evolution in *Chenopodium*. *The Plant Genome* **2025**, *18*, e70010. <https://doi.org/10.1002/tpg2.70010>.
84. Maughan, P.J.; Jarvis, D.E.; De La Cruz-Torres, E.; Jaggi, K.E.; Warner, H.C.; Marcheschi, A.K.; Bertero, H.D.; Gomez-Pando, L.; Fuentes, F.; Mayta-Anco, M.E.; et al. North American pitseed goosefoot (*Chenopodium berlandieri*) is a genetic resource to improve Andean quinoa (*C. quinoa*). *Sci Rep* **2024**, *14*, 12345. <https://doi.org/10.1038/s41598-024-63106-8>.
85. Rey, E.; Abrouk, M.; Dufau, I.; Rodde, N.; Saber, N.; Cizkova, J.; Fiene, G.; Stanschewski, C.; Jarvis, D.E.; Jellen, E.N.; et al. Genome assembly of a diversity panel of *Chenopodium quinoa*. *Sci Data* **2024**, *11*, 1366. <https://doi.org/10.1038/s41597-024-04200-4>.
86. Stephensen, K.B.; Costa-Tártara, S.M.; Roser, R.L.; Jarvis, D.E.; Maughan, P.J.; Jellen, E.N. Germplasm Pools for Quinoa Improvement. *Crops* **2025**, *6*, 4. <https://doi.org/10.3390/crops6010004>.
87. Matías, J.; Cruz, V.; Reguera, M. Heat Stress Impact on Yield and Composition of Quinoa Straw under Mediterranean Field Conditions. *Plants (Basel)* **2021**, *10*. Place: Switzerland, <https://doi.org/10.3390/plants10050955>.
88. Grenfell-Shaw, L.; Tester, M. Abiotic Stress Tolerance in Quinoa. In *The Quinoa Genome*; Schmöckel, S.M., Ed.; Springer International Publishing: Cham, 2021; pp. 139–167. [https://doi.org/10.1007/978-3-030-65237-1\\_9](https://doi.org/10.1007/978-3-030-65237-1_9).
89. Ohama, N.; Sato, H.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science* **2017**, *22*, 53–65. <https://doi.org/10.1016/j.tplants.2016.08.015>.
90. Xu, J.; Farooq, H.U.; Hashim, M.; Rey, E.; Curti, R.; Morris, A.; Maughan, P.J.; Jellen, E.N.; Jarvis, D.E.; Bertero, D.; et al. Wild relatives to improve heat tolerance of cultivated quinoa (*Chenopodium quinoa*): pollen viability and grain number. *Journal of Experimental Botany* **2025**, *76*, 5117–5128. <https://doi.org/10.1093/jxb/eraf235>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.