Article

Analysis of Heat Shock Proteins Based on Amino Acids for the Tomato Genome

Meshal M. Almutairi * and Hany M. Almotairy

National Center of Agricultural Technology, Sustainability and Environment, King Abdulaziz City for Science and Technology KACST, Box 6086, Riyadh 11442, Saudi Arabia
* Correspondence: malmetairy@kacst.edu.sa

Abstract: This research aimed to investigate heat shock proteins in the tomato genome through the analysis of amino acids. The highest length among sequences was found in seq19 with 3534 base pairs. This seq19 was reported and contained a family of proteins known as HsfA that have a domain of transcriptional activation for tolerance to heat and other abiotic stresses. The values of the codon adaptation index (CAI) ranged from 0.80 in Seq19 to 0.65 in Seq10, based on the mRNA of heat shock proteins for tomatoes. Asparagine (AAT, AAC), aspartic acid (GAT, GAC), phenylalanine (TTT, TTC), and tyrosine (TAT, TAC) have relative synonymous codon usage (RSCU) values bigger than 0.5. In modified relative codon bias (MRCBS), the high gene expressions of the amino acids under heat stress were histidine, tryptophan, asparagine, aspartic acid, lysine, phenylalanine, isoleucine, cysteine, and threonine. RSCU values that were less than 0.5 were considered rare codons that affected the rate of translation, and thus selection could be effective by reducing the frequency of expressed genes under heat stress. The normal distribution of RSCU shows about 68% of the values drawn from the standard normal distribution were within 0.22 and −0.22 standard deviations that tend to cluster around the mean. The most critical component based on principal component analysis (PCA) was the RSCU. These findings would help plant breeders in the development of growth habits for tomatoes during breeding programs.

Keywords: amino acids; codon adaptation index; relative synonymous codon usage; modified relative codon bias; principal component analysis

1. Introduction

Temperature is the primary environmental variable playing a significant role in plant growth and survival. Increasing the temperature to about 12 °C above the threshold point reduces plant productivity by affecting embryogenesis, germination, and fruit development. Tomato is sensitive to heat stress; increasing a few degrees above the average day temperature (25 °C) can impair reproductive organs, resulting in the failure of fruit setting. Heat stress reduced tomato production in both open field and non-controlled growing conditions. The limitations in understanding the genetic basis of heat tolerance traits limits the breeding of tomatoes, which requires a depth of knowledge about their genetic architecture [1]. Tomatoes respond to heat stress using a group of polypeptides during cellular activity known as heat stress transcription factors (HSFs). The factors that are molecular chaperones mediate the rapid accumulation of heat shock proteins during biotic and abiotic stresses, which help plants to protect themselves from damage and degradation of proteins and maintain optimal growth [2]. A small gene family code called Hsfs exists as inactive proteins with up to five members in plants. Hsfs contain two clusters of basic amino acid residues considered as K/R motifs for localization. Moreover, these factors are classified into different families localized to cellular compartments such as the nucleus, endoplasmic reticulum mitochondria, and chloroplasts. These HSFs prevent aggregation and translocation by binding to denatured proteins [2,3]. Stress causes activation with oligomerization when HsfA1 binds with its target sequences, where present, in the promoter of
the genes. Some studies showed that the heat shock proteins’ chloroplast is involved in plant thermotolerance, which protects the photosystem II. A few of the chaperones such as Hsp17–CII, Hsp70, and Hsp90 are involved in the second part of the Hsf cycle that leads to the restoration of the inactive state. More than 365-trained deep learning models of all tomatoes’ genomes were applied for better predictions [4–7].

Tomatoes that belong to the family Solanaceae have a unique flavor and high nutritional value. Tomatoes are diploid, with a genome size of about 950 Mb and 12 chromosomes. During the domestication and breeding processes, tomatoes dramatically change fruit shape, color, and plant tolerance according to a biotic and/or an abiotic stress [8]. More than 1000 tomato varieties and wild species have been sequenced [9]. Tomatoes are extremely sensitive to heat stress, which causes abortion of the male gametophyte, reflecting a reduction in fruit set and yield when there is an increase in daytime temperature above 26 °C. High temperatures (44 °C) cause cell death due to released cytochrome and induced caspase-like enzymes [10]. This extremely high temperature affects membrane fluidity and metabolic mechanisms, which cause the overproduction of reactive oxygen species and oxidative stress [11,12]. Heat shock proteins are produced in cells due to exposure to stress. It was documented that these proteins protect cells from destructive effects during the synthesis of proteins. Some studies have evaluated heat tolerance in tomatoes using different parameters such as the phenotypic index, while others have suggested the investigation of codon usage patterns. Thus, codon usage patterns provide one insight into the impact of high temperatures on tomato proteins. The arrangement of genetic code responds to the synthesis of all protein molecules, which are revealed by using sixty-one codons to encode 20 amino acids. These codon usage patterns vary among genes. However, it is still challenging to decide the most common dominant codon bias due to compositional constraint and codon–anticodon interaction [12].

Amino acids are a fundamental measurement for the plant growth stage due to their connection to protein structures, functions, and much metabolic flux [13]. Thus, these amino acids are essential for establishing the basics of genetic information. In addition, a genetic code uses 61 codons, corresponding to 20 amino acids encoded to more than one codon, that can be revealed as a synonymous codon except for both methionine and tryptophan, which are encoded by only one codon [14,15]. Bias in synonymous codons may arise because of natural selection or mutation [16,17]. It is well known that genetic codes direct all proteins in the DNA genome. However, codon usage pattern varies among organisms and genes. Thus, this variation suggests different factors that influence the codon usage pattern. Several reports explained the codon distribution in the proteins among different synonymous codons. These factors are nucleotide composition, tRNA abundance, the structure of the protein, and translation processes. To explain more about the codon usage pattern, the codon adaption index (CAI) is one of the best methods for calculating the frequency of the synonymous codon for more understanding of gene expression, new gene discovery, molecular mechanism, and genetic engineering. It was found that heat shock proteins were accumulated in the mRNA. However, there is a lack of understanding of the genetic processes involved in heat stress during tomato growth. Therefore, this research aimed to understand the mechanisms of heat shock proteins in tomatoes by investigating heat shock proteins through analysis of amino acids [18–22].

2. Materials and Methods

The heat shock protein sequences for the *Solanum lycopersicum* cultivar Heinz 1706 were retrieved from the National Center for Biotechnology NCBI (https://www.ncbi.nlm.nih.gov/nuccore/?term= (accessed on 7 June 2021)). These 19 proteins, with their accession numbers, were provided (Table 1) [23–28].
Table 1. The basic of *Solanum lycopersicum* sequence attributes under heat shock.

| Sequences | Accession No. | CAI  | GC%  | Length (bp) |
|-----------|---------------|------|------|-------------|
| 1         | XM_010327263.3 | 0.77 | 41.99 | 2023        |
| 2         | NM_001309248.1 | 0.66 | 44.15 | 1973        |
| 3         | NM_001321563.1 | 0.68 | 41.71 | 2539        |
| 4         | NM_001246851.2 | 0.78 | 42.14 | 2589        |
| 5         | NM_001247296.2 | 0.77 | 40.39 | 822         |
| 6         | NM_001320262.1 | 0.76 | 39.32 | 819         |
| 7         | NM_001321450.1 | 0.75 | 39.28 | 1647        |
| 8         | XM_026032687.1 | 0.65 | 38.75 | 2080        |
| 9         | XM_010326728.3 | 0.76 | 38.83 | 2127        |
| 10        | XM_026032686.1 | 0.65 | 39.08 | 2165        |
| 11        | XM_026032685.1 | 0.76 | 37.92 | 2379        |
| 12        | NM_001247047.2 | 0.78 | 33.29 | 1673        |
| 13        | NM_001247134.1 | 0.66 | 37.72 | 3982        |
| 14        | XM_015230190.2 | 0.70 | 44.01 | 1977        |
| 15        | XM_027919128.1 | 0.68 | 40.06 | 1343        |
| 16        | XM_027919127.1 | 0.67 | 38.49 | 1642        |
| 17        | XM_015208371.2 | 0.69 | 40.13 | 1687        |
| 18        | XM_027913778.1 | 0.67 | 43.82 | 1376        |
| 19        | NM_001247342.2 | 0.80 | 40.95 | 3534        |

CAI stands for the codon adaptation index, GC% stands for guanine-cytosine content, and bp is base pair.

These proteins were derived from a whole-genome shotgun sequence supported by EST evidence. In short, tomato genomes are an international consortium and have been extensively characterized to better understand genetic diversity [29,30]. All proteins were checked based on the following considerations, such as the presence of START and STOP codons, and removed duplicated sequences [18].

A variety of analysis tools were utilized, such as the codon adaptation index (CAI) given by Sharp and Li [31], and RSCU provided by Yu et al. [22]. In addition, guanine-cytosine content (GC) content, relative codon bias strength (RCBS), and MRCBS were applied in this study [32].

The following formula represents the codon adaptation index: $\text{CAI} = \left(\prod_{i=1}^{N} w_i\right)^{1/n}$, where $N$ is the number of a codon in the gene and relativeness $w_i$ is described as $w_i = f_i/f(aa, \text{max})$. $F_i$ is the frequency of the $i$th codon, and $f(aa, \text{max})$ is the maximum frequency of the codon most often used to encode amino acid $aa$ [31].

A variety of synonymous codon usage can be estimated using the following formula: $\text{RSCU} = \frac{X_{ij}}{\sum_{j=1}^{ni} X_{ij}}$, where $x_{ij}$ shows the frequency of codon $j$ coding for the $i$th amino acid and $n_i$ represents the number of the synonymous codon encoding the $i$th amino acids. The interpretation for the RSCU is that an RSCU value greater than 0.5 has more frequency, and an RSCU value smaller than 0.5 indicates less frequency in a gene for a particular codon. Whereas RSCU > 1 indicates an overrepresented codon for the corresponding amino acid [33].

Relative synonymous codon usage can be estimated using the following formula: $\text{RSCU} = \frac{X_{ij}}{\sum_{j=1}^{ni} X_{ij}}$, where $x_{ij}$ shows the frequency of codon $j$ coding for the $i$th amino acid and $n_i$ represents the number of the synonymous codon encoding the $i$th amino acids. The interpretation for the RSCU is that an RSCU value greater than 0.5 has more frequency, and an RSCU value smaller than 0.5 indicates less frequency in a gene for a particular codon. Whereas RSCU > 1 indicates an overrepresented codon for the corresponding amino acid [33].

Modified relative codon bias measures the expression level of a gene [32]. MRCBS shows in the formula $MRCBS = \left[ \prod_{i=1}^{N} (MRCBS_{xyz}) \right]$. And $(MRCBS_{xyz}) = \frac{(RCBS_{xyz})}{(RCBS_{aa, \text{max}})}$. And $(RCBS_{xyz}) = \frac{f_{xyz}}{(f_{1}(x) f_{2}(y) f_{3}(z))}$, $f_{xyz}$ is the normalized codon frequency of a codon $xyz$ and $f_{n}(m)$ is the normalized frequency of base $m$ at codon position $n$ in a gene. The $(RCB aa, \text{max})$ reflects the maximum value of the RCBS of a codon encoding the same amino acid $aa$. The score of MRCBS ranges from zero to one.

The data from RSCU, GC content, RCBS, and MRCBS were computed integrally using Microsoft Excel. Two-way analysis of variance was computed using the PROC GLM procedure at a significance level of the $p$-value (0.001) in SAS computer packages version 9.2 for Windows (SAS Institute Inc., Cary, NC, USA). In addition, analysis of principal components helps to re-set data into a few data sets computed using PROC PRINCOMP in SAS to analyze total variations.
3. Results

The ANOVA test for the MRCBS data reflects the significant difference among 21 amino acids, 64 codons, and 19 sequences with a p-value < 0.0001 (Table 2). For example, the maximum means $\pm$SD across all sequences in MRCBS was 0.67 $\pm$ 0.36 for seq14. In contrast, the maximum value for RSCU was 0.28 $\pm$ 0.22 in both sequences 1 and 19. In addition, the maximum means $\pm$SD for the amino acids asparagine, tyrosine, and tryptophan were 0.87 $\pm$ 0.17, 0.85 $\pm$ 0.24, and 0.80 $\pm$ 0.44, respectively. Thus, it was found that all parameters for testing a normal distribution were significant, following a uniform distribution.

Table 2. The ANOVA for MRCBS, amino acids and codons for the Solanum lycopersicum cultivar Heinz 1706.

| Source    | DF | Anova SS | Mean Square | F Value | Pr > F |
|-----------|----|----------|-------------|---------|--------|
| Sequences | 18 | 1.69     | 0.09        | 3.39    | <0.0001|
| Amino acids | 20 | 23.09    | 1.15        | 41.58   | <0.0001|
| Codons    | 63 | 82.47    | 1.31        | 47.14   | <0.0001|

3.1. Genome Sequences Based on the Codon Adaptation Index and GC%

The highest length among sequences was found in seq13 and seq19, with 3982 and 3534 base pairs, respectively (Table 2). The CAI ranges from zero to one, and the higher values of CAI were found in seq19 (0.80) while the lower ones were found in seq10 (0.65). It was reported that sequence19 contained a family of proteins known as HsfA that have the domain of transcriptional activation for tolerance to heat and other abiotic stress [24]. Under stress, hetero-oligomers were formed by the interaction between HsfA1 and HsfB1. It was found that HsfA1 plays an important role in protecting cellular homeostasis under heat stress. This illustrates that the HsfA1 group is a master regulator of heat-stress response in tomatoes, as reported by Mishra et al. [34]. In addition, the HsfB1 expressed significantly with increased heat stress, as known by Kumar et al. [35]. These heat stress transcription factors (Hsfs) are the terminal components for the activation of genes responsive to heat and other stresses. The HsfA1, HsfA2, and HsfB1 are located between the L1 and L2 positions and on C4 at the DNA-binding domain in the tomato genome. However, the HsfA1 is different from all others by an insertion of 21 amino acid residues in the linker. Moreover, protein resulted in higher GC% in seq2 (44.15), and the length of sequences demonstrates the impact of amino acid composition on synonymous codons. This impact is essential to illustrate the limitation of the expected value of CAI based on the sequence. Contrary to previous studies, the length of sequences did not affect either GC% or CAI. The lowest GC% content was 33.29% in Seq12, while the highest was 44.15% in Seq2. These proteins ranged from the most increased length, about 3982 base pairs in Seq13, to the smallest length (819 base pairs) in Seq6. A few reports illustrated that the GC content of rRNA and tRNA showed a strong positive correlation with their optimal growth temperature.

The GC content is an important qualitative aspect in the genomic nucleotide reflected genomic architecture. The importance of this GC content is in conferring stability to higher order structure of DNA and RNA transcripts. Here, it was observed that GC% for heat shock protein (HsfA1) was 37.72%, reflecting the moderated stability of DNA compared to the AU base pairs that were less stable. The second higher of GC% was Seq14, with accession no XM_015230190.2, known also as heat shock protein (Hsf8). The variation in GC% within the identical genomes contributed to mapping genes and biased recombination associated with DNA repair. It was reported that some repeated sequences are indeed GC rich, include the tandemly repeated and interspersed rDNA. Sometimes, the repairing of mismatched base pairs is hypothesized to be biased to certain nucleotides during DNA repair or gene transcription [36]. Carels and Bernardi [37] reported that proteins produced by GC-rich and GC-poor genes showed functional differences. Both of these GCs were different in amino acid composition. In addition, it was found in grasses that fewer introns
control GC-rich genes. GC content and genome size showed a quadratic relationship due to the biochemical costs of GC base synthesis [38].

3.2. The Codons of Amino Acids Based on RSCU Analysis

The synonymous codon usage was analyzed using the above, given by Yu et al. [22]. RSCU is calculated as the observed frequency of the codon divided by its expected frequency. For more illustration, an RSCU value bigger than 0.5 indicates a more frequently used codon. While an RSCU value less than 0.5 indicates a less frequently used codon (Figure 1). It can be seen that histidine, asparagine, aspartic acid, cysteine, phenylalanine, and tyrosine have RSCU values bigger than 0.5. Asparagine (AAT, AAC), Aspartic acid (GAT, GAC), Phenylalanine (TTT, TTC), and Tyrosine (TAT, TAC) have second positions as A and T, while the third position in these amino acids has either T or C. This result reveals that the high RSCU values were rich with these two codons, A (adenine) and T (thymine). These second nucleotide positions should emphasize the nature of the encoded amino acids that determine the transmission of information from DNA to mRNA to protein.

Moreover, RSCU values less than 0.5 were considered rare codons such as lysine, glycine, threonine, isoleucine, proline, stop codon, valine, leucine, glutamic acid, arginine, serine, and glutamine, which affect the translation [28]. Thus, selection could be effective by reducing the frequency of expressed genes under heat stress. The normal distribution of RSCU is important to reflect the population distribution (Figure 2). This distribution is known as symmetric around the means. The means and standard deviation of the RSCU values were 0.27 and 0.22, respectively. Therefore, 68% of values drawn from the standard normal distribution are within 0.22 and ±0.22 standard deviations, which tends to cluster around the mean.

3.3. The MRCBS Analysis of Codons for Heat Shock Proteins

The MRCBS values were calculated based on all tomato proteins under heat shock to obtain the most gene expression for several amino acids that varied from one to 0.14 (Figure 3). The threshold of the MRCB values was 0.62 greater than this threshold and was taken as the standard for determination of the high gene expression. Therefore, the increased gene expression of the amino acids under heat stress was histidine, tryptophan, asparagine, aspartic acid, lysine, phenylalanine, isoleucine, cysteine, threonine, valine, glutamic acid, and glycine. Other amino acids that less than the threshold (0.62) were considered as less gene expression. Some significant scores were found in some codons that contribute to gene expressions such as (CAT) As (Ala), TGC (Arg), ACA (Arg), TAC (Asp), AGA (Cys), GGC (Glu), CTG (Gly), CAA (Ile), TAA (Ile), GGG (Leu), CCG (Lys), GAC (Pro), GTC (Ser), CAG (Ser), CGA, CGT (Stop), TAG, GCG (Thr), GGA (Tyr), and CTA (Val) < 0.05.
These significant codons may affect translational efficiency [39,40]. Most of the high gene expressions were unknown to their characterized hypothetical genes. Thus, genes with high-predicted expressions may have important functions that were unknown. These proteins with their codons may provide targets for identifying critical features of tomato genomes under heat stress [41].

**Figure 2.** The normal distribution of the RSCU values in heat shock protein sequences based on mRNA for the *Solanum lycopersicum* cultivar Heinz 1706.

**Figure 3.** The MRCBS mean in heat shock protein sequences for the *Solanum lycopersicum* cultivar Heinz 1706, excluding Methionine and Tryptophan, with standard error shown by the red line and label with standard deviation (±).
3.4. Principal Component Analysis

Principal component analysis is useful for identifying unknown trends in tomato genomes and simplifying the description of the tomato genome by analyzing the structure of the observations and variables. The two dominations in Figure 4 show the total variability for 19 mRNA sequences related to heat shock proteins. The first domination accounts for 80.5% of the total variance, while the second component accounts for 6.2% of the total variance. The eigenvalue greater than 1 was the RSCU with a value of 1.94, which correlated to PC1 with 64.7% of the total variance. While the eigenvalues that were less than one were in both MRCBS with 0.80 and RCBS with 0.26, which together account for 35.2% of the total variance that correlated to PC2 and PC3, respectively. In addition, the most important component of the tomato genome was the RSCU and MRCBS. Thus, the most variation was explained in these first and second components. While the RCBS had less importance in revealing the variation and its eigenvalue was 0.26. The higher contribution of variations among 19 mRNA sequences in the tomato genomes are from the following sequences: Seq14, Seq2, Seq3, Seq16, and Seq15, while the lower contributions of variations are from Seq18, Seq4, Seq19, Seq8, and Seq6. Other sequences were between the higher and lower contributions to total variation.

Figure 4. The percentage of two dominations explains variability through principal components for 19 mRNA sequences related to heat shock proteins for the *Solanum lycopersicum* cultivar Heinz 1706, the right legend shows the contributions of variables to PCs. Dim1 expresses dimension 1 while Dim2 expresses dimension 2.
4. Discussion

A sustainable crop system such as tomato is important for its food and economic value. Global temperature shows an increasing trend, making heat stress a critical issue; meanwhile genetic and physiological responses to this stress should be considered. The tomato is known to be extremely sensitive to heat stress, reflecting a reduction in fruit set and yield when there is an increase in daytime temperature above 26 °C [11]. The tomato responds to a high temperature by inducing the synthesis of a group of polypeptides known as heat shock proteins ranging from about 15 to 42 kD. Out of 20 amino acids, 18 are encoded with more than two synonymous codons at the level of the genetic code. Some codons are translated more efficiently and accurately due to these codons being favored over others by natural selection [15].

Predicting gene expression from nucleotide sequences is considered fundamental knowledge for modern bioinformatics. This gene could be turned on for desired genes or edited in genes for more specificity. This gene expression is controlled by several factors, such as protein biosynthesis, mutation, and natural selection. These nucleotides and gene expressions would increase the translation efficiency of the gene population. Due to several factors, the CAI values may indicate some nuclear singles in codon bias. One example of these factors is the neutral mutation press for shaping these codons’ usage, while another example is the translation levels [4–6].

The GC content helps breeders to predict conserved amino acid patterns [1]. Several studies have argued that GC content may reflect the primary influence on codon usage patterns that change the expression profile. The GC content was reported as one of the principal influences of codon usage patterns that changed the gene expression. Thirty optimal codons ended with A, T, or G with 13, 15, or 2, respectively. It was stated that the GC index is more sensitive to rare amino acids, and thus it is poorly informed and could not provide details for codon usage; however, a small gene that is less than 100 codons could be used in the GC index more informatively [16].

5. Conclusions

This paper showed that RSCU, RCBS, and MRCBS might be useful tools for predicting the amino acid frequency and gene expression during the growth stages of tomatoes. The idea of supporting our method and finding is based primarily on the responsible codon usage pattern for the regulation of gene translations that occur during protein translation. This objective would help plant breeders in the development of growth habits for tomatoes. It also helps to demonstrate significant heterogeneity in codon usage among genes in tomato genomes [26,27].

The CAI and GC% were used to measure the synonymous codon usage bias for the mRNA of heat shock proteins in tomatoes. The values of CAI ranged from 0.80 in Seq19 to 0.65 in Seq10. Protein resulted in higher GC% in Seq2 (44.15), and the length demonstrates the impact of amino acid composition on synonymous codons. An RSCU value bigger than 0.5 indicated a more frequently used codon, while an RSCU value less than 0.5 indicated a less frequently used codon. RSCU values bigger than 0.5 were observed in histidine, asparagine, aspartic acid, cysteine, phenylalanine, and tyrosine have asparagine (AAT, AAC), aspartic acid (GAT, GAC), phenylalanine (TTT, TTC), and tyrosine (TAT, TAC). In addition, this result reveals that the high RSCU values were rich with these two codons, As (adenine) and Ts (thymine). The overall results here indicate that the highest genes expressed of the amino acids were histidine, asparagine, aspartic acid, cysteine, phenylalanine, tyrosine, and asparagine. In addition, the result of the codon adaption index can help for more confidence in predictions. More studies are required to understand the amino acid network metabolism, such as biochemical, molecular, and genomics [34,42].

Gene function can be determined through natural selection and could be developed for adaptability to the environment. It was found that arginine and proline metabolism, alanine, aspartate, and glutamate were significantly disturbed by drought stress [43–45].
Our results provide the first understanding of the tomato genome to facilitate the study of genetic mechanisms in tomato biology.

Author Contributions: Conceptualization, M.M.A. methodology, M.M.A. and H.M.A.; software, M.M.A.; validation, M.M.A. and H.M.A.; formal analysis, M.M.A.; investigation, M.M.A. and H.M.A.; resources, M.M.A.; data curation, M.M.A. and H.M.A.; writing—original draft preparation, M.M.A.; writing—review and editing, M.M.A. and H.M.A.; visualization, M.M.A.; supervision, M.M.A.; project administration, M.M.A. and H.M.A.; funding acquisition, M.M.A. and H.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank KACST for the facilities and materials provided, which helped to set up and finish this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Araus, J.; Slafer, G.A.; Royo, C.; Serret, M.D. Breeding for Yield Potential and Stress Adaptation in Cereals. *Crit. Rev. Plant Sci.* 2008, 27, 377–412. [CrossRef] [PubMed]

2. Hong, J.C. General Aspects of Plant Transcription Factor Families. In *Plant Transcription Factors*; Gonzalez, D.H., Ed.; Academic Press: Cambridge, MA, USA, 2016; pp. 35–56. [CrossRef] [PubMed]

3. Waters, E.R. The Evolution, Function, Structure, and Expression of the Plant sHSPs. *J. Exp. Bot.* 2013, 64, 391–403. [CrossRef] [PubMed]

4. Liu, Y.; Yang, Q.; Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annu. Rev. Biochem.* 2021, 90, 375–401. [CrossRef] [PubMed]

5. Akagi, T.; Masuda, K.; Kuwada, E.; Takeshita, K.; Kawakatsu, T.; Ariizumi, T.; Kubo, Y.; Ushijima, K.; Uchida, S. Genome-wide cis-decoding for Expression Design in Tomato Using Cistrome Data and Explainable Deep Learning. *Plant Cell* 2022, 34, 2174–2187. [CrossRef] [PubMed]

6. Guo, Y.; Guettouche, T.; Fenna, M.; Boellmann, F.; Pratt, W.; Toft, D.; Smith, D.; Voellmy, R. Evidence for a Mechanism of Repression of Heat Shock Factor 1 Transcriptional Activity by a Multichaperone Complex. *J. Biol. Chem.* 2001, 276, 45791–45799. [CrossRef] [PubMed]

7. Kim, B.; Schöffl, F. Interaction between Arabidopsis Heat Shock Transcription Factor 1 and 70 kDa Heat Shock Proteins. *J. Exp. Bot.* 2002, 53, 371–375. [CrossRef] [PubMed]

8. Port, M.; Tripp, J.; Zielinski, D.; Weber, C.; Heerklotz, D.; Winkelhaus, S.; BublaK, D.; Scharf, K. Role of Hsp17.4-CII as Coregulatory and Cytoplasmic Retention Factor of Tomato Heat Stress Transcription Factor HsfA2. *Plant Physiol.* 2004, 135, 1457–1470. [CrossRef]

9. Dabbaghizadeh, A.; Tanguay, M. Structural and Functional Properties of Proteins Interacting with Small Heat Shock Proteins. *Cell Stress Chaperones* 2020, 25, 629–637. [CrossRef]

10. Aflitos, S.; Schijlen, E.; Jong, H.; Ridder, D.; Smit, S.; Finkers, R.; Wang, J.; Zhang, G.; Li, N.; Mao, L.; et al. Exploring Genetic Variation in The Tomato (Solanum Section Lycopersicon) Clade by Whole-Genome Sequencing. *Plant Physiol.* 2014, 162, 136–148. [CrossRef]

11. Qu, G.; Liu, X.; Zhang, Y.; Yao, D.; Ma, Q.; Yang, M.; Zhu, W.; Yu, S.; Luo, Y. Evidence for Programmed Cell Death and Activation of Specific Caspase-Like Enzymes in The Tomato Fruit Heat Stress Response. *Planta* 2009, 229, 1269–1279. [CrossRef]

12. Wehmeyer, N.; Hernandez, L.; Finkelstein, R.; Vierling, E. Synthesis of Small Heat-Shock Proteins is Part of The Developmental Program of Late Seed Maturation. *Plant Physiol.* 1996, 112, 747–757. [CrossRef] [PubMed]

13. Sun, W.; Bernard, C.; Cotte, B.; Montagu, M.; Verbruggen, N. At-HSP17.6A, Encoding A Small Heat-Shock Protein in Arabidopsis, Can Enhance Osmotolerance Upon Overexpression. *Plant J.* 2001, 27, 407–415. [CrossRef]

14. Tian, R.; Yang, Y.; Chen, M. Genome-Wide Survey of the Amino Acid Transporter Gene Family in Wheat (*Triticum Aestivum* L.): Identification, Expression Analysis and Response to Abiotic Stress. *Int. J. Biol. Macromol.* 2020, 162, 1372–1387. [CrossRef] [PubMed]

15. Hershberg, R. Codon Usage and Translational Selection. *Encycl. Evol. Biol.* 2016, 293–298. [CrossRef]

16. Duret, L. Evolution of Synonymous Codon Usage in Metazoans. *Curr. Opin. Genet. Dev.* 2002, 12, 640–649. [CrossRef]

17. Song, H.S.; Liu, J.; Chen, T.; Nan, Z. Synonymous Codon Usage Pattern in Model Legume *Medicago truncatula*. *J. Integr. Agric.* 2018, 17, 2074–2081. [CrossRef]

18. Shafat, Z.; Ahmed, A.; Parvez, K.; Parveen, S. Decoding the Codon Usage Patterns in Y-Domain Region of Hepatitis E Viruses. *J. Genet. Eng. Biotechnol.* 2022, 20, 56. [CrossRef] [PubMed]
19. Brandis, G.; Hughes, D. The Selective Advantage of Synonymous Codon Usage Bias in Salmonella. *PLoS Genet.* 2016, 12, e1005926. [CrossRef]

20. Carbone, A.; Zinovyev, A.; Képès, F. Codon Adaptation Index as a Measure of Dominating Codon Bias. *Bioinformatics* 2003, 19, 2005–2015. [CrossRef]

21. Lithwick, G.; Margalit, H. Relative Predicted Protein Levels of Functionally Associated Proteins are Conserved across Organisms. *Nucleic Acids Res.* 2005, 33, 1051–1057. [CrossRef]

22. Yu, C.; Dang, Y.; Zhou, Z.; Wu, C.; Zhao, F.; Sachs, M.; Liu, Y. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-Translational Protein Folding. *Mol. Cell.* 2015, 59, 744–754. [CrossRef] [PubMed]

23. Yamaguchi, H.; Ohnishi, J.; Saito, A.; Ohyama, A.; Nunome, T.; Miyatake, K.; Fukouka, H. An NB-LRR gene, TYNS1, Is Responsible for Resistance Mediated by The Ty-2 Begomovirus Resistance Locus of Tomato. *Theor. Appl. Genet.* 2018, 131, 1345–1362. [CrossRef] [PubMed]

24. Zhou, J.; Xu, X.; Cao, J.; Yin, L.; Xia, X.; Shi, K.; Zhou, Y.; Yu, J. Heat Shock Factor Hsfa1a is Essential for R Gene-Mediated Nematode Resistance and Triggers H2O2 Production. *J. Plant Physiol.* 2018, 176, 2456–2471. [CrossRef] [PubMed]

25. Vu, N.; Kamiya, K.; Fukushima, A.; Hao, S.; Ning, W.; Arizumi, T.; Ezura, H.; Kusano, M. Comparative Co-Expression Network Analysis Extracts the Slhsp70 Gene Affecting to Shoot Elongation of Tomato. *J. Plant Biotechnol.* 2019, 36, 143–153. [CrossRef] [PubMed]

26. Shukla, V.; Upadhyay, R.; Tucker, M.; Varnavoni, J.; Rudrabhatla, S.; Mattao, A. Transient Regulation of Three Clustered Tomato Class-I Small Heat-Shock Chaperone Genes by Ethylene is Mediated by Smds-RIN Transcription Factor. *Sci. Rep. Nat.* 2017, 7, 6474. [CrossRef]

27. Fragkostefanakis, S.; Simm, S.; El-Shershaby, A.; Hu, Y.; Bublak, D.; Meselhuvic, A.; Darm, K.; Mishra, S.; Tschiersch, B.; Theres, K.; et al. The Repressor and Co-activator Hsf1b Regulates the Major Heat Stress Transcription Factors in Tomato. *Plant Cell Environ.* 2019, 42, 874–890. [CrossRef]

28. Bolger, A.; Scossa, F.; Bolger, M.; Lanz, C.; Maumus, F.; Tohge, T.; Quesneville, H.; Sørensen, I.; Lichtenstein, G.; et al. The Genome of The Stress-Tolerant Wild Tomato Species Solanum Pennellii. *Nat. Genet.* 2014, 46, 1034–1038. [CrossRef]

29. Wang, Y.; Hoeven, R.S.; Nielsen, R.; Mueller, L.A.; Tanksley, S.D. Characteristics of the Tomato Nuclear Genome as Determined by Sequencing Undermethylated EcoRI Digested Fragments. *Theor. Appl. Genet.* 2005, 112, 72–84. [CrossRef]

30. Van Rens, W.M.; Schmidt, M.H.; Eijfgen, S.; Wang, Y.; Zaidan, M.W.A.M.; Huettel, B.; Schouten, J.; Usadel, B.; Underwood, C. A Chromosome Scale Tomato Genome Built from Complementary Pacbio and Nanopore Sequences Alone Reveals Extensive Linkage Drag During Breeding. *Plant J.* 2022, 110, 572–588. [CrossRef]

31. Sharp, P.M.; Li, W.H. The Codon Adaptation Index a measure of Directional Synonymous Codon Usage Bias and Its Potential Applications. *Nucleic Acids Res.* 1987, 15, 1281–1295. [CrossRef]

32. Shoo, S.; Das, S.S.; Rakshit, R. Codon Usage Pattern and Predicted Gene Expression in *Arabidopsis Thaliana*. *Gene* 2019, 721, 10012. [CrossRef]

33. Wong, E.H.; Smith, D.K.; Rabadan, R.; Peiris, M.; Poon, L.L. Codon Usage Bias and the Evolution of Influenza a Viruses. Codon Usage Biases of Influenza Virus. *BMC Ecol. Evol.* 2010, 10, 253. [CrossRef]

34. Mishra, S.K.; Tripp, J.; Winkelhaus, S.; Tschiersch, B.; Theres, K.; Nover, L.; Scharf, K.D. In The Complex Family of Heat Stress Transcription Factors, Hsf1a Has a Unique Role as Master Regulator of Thermostolerance in Tomato. *Genes Dev.* 2002, 16, 1555–1567. [CrossRef]

35. Kumar, M.; Busch, W.; Birke, H.; Kemmerling, B.; Nurnberger, T.; Schoffl, F. Heat Shock Factors Hsf1 and Hsf2b Are Involved in the Regulation of Pdi1.2 Expression and Pathogen Resistance in Arabidopsis. *Mol. Plant* 2009, 2, 152–165. [CrossRef]

36. Brown, T.C.; Jiricny, J. Different Base/Base Mispairs are Corrected with Different Efficiencies and Specificities in Monkey Kidney Cells. *Cell* 1988, 54, 705–711. [CrossRef]

37. Caires, N.; Bernardi, G. Two Classes of Genes in Plants. *Genetics* 2000, 154, 1819–1825. [CrossRef]

38. Shi, X.; Wang, X.; Li, Z.; Zhu, Q.; Yang, J.; Ge, S.; Luo, J. Evidence That Natural Selection is the Primary Cause of the Guanine-Cytosine Content Variation in Rice Genes. *J. Integr. Plant Biol.* 2007, 49, 1393–1399. [CrossRef]

39. Fazakerley, G.V.; Quignard, E.; Wiosar, A.; Guschlbauer, W.; van der Marel, G.A.; van Boom, J.H.; Jones, M.; Radman, M. Structures of Mismatched Base Pairs in DNA and Their Recognition by the Escherichia Coli Mismatch Repair system. *EMBO J.* 1986, 5, 3679–3703. [CrossRef]

40. Roymondal, U.; Das, S.; Sahoo, S. Predicting Gene Expression Level from Relative Codon Usage Bias: An Application to Escherichia Coli Genome. *DNA Res.* 2009, 16, 13–30. [CrossRef]

41. Das, S.; Paul, S.; Dutta, C. Synonymous Codon Usage in Adenoviruses: Influence of Mutation, Selection and Protein Hydropathy. *Virus Res.* 2006, 117, 227–236. [CrossRef]

42. Zhang, W.; Zhou, J.; Li, Z.; Wang, L.; Gu, X.; Zhong, Y. Comparative Analysis of Codon Usage Patterns Among Mitochondrion, Chloroplast and Nuclear Genes in *Triticum Aestivum* L. *J. Integr. Plant Biol.* 2007, 49, 246–254. [CrossRef]

43. Galili, S.; Amir, R.; Galili, G. Genetic Engineering of Amino Acid Metabolism in Plants. *Adv. Plant Biochem. Mol. Biol.* 2008, 1, 49–80. [CrossRef]

44. You, J.; Zhang, Y.; Lui, A.; Li, D.; Wang, X.; Dossa, K.; Zhou, R.; Yu, Y.; Zhang, Y.; Wang, L.; et al. Transcriptomic and Metabolomics Profiling of Drought-Tolerant and Susceptible Sesame Genotypes in Response to Drought Stress. *MBC Plant Biol.* 2019, 19, 267. [CrossRef] [PubMed]

45. Sanders, A.; Collier, R.; Trethewy, A.; Gould, G.; Sieker, R.; Tegeder, M. AAP1 Regulates Import of Amino Acids into Developing Arabidopsis Embryos. *Plant J.* 2009, 59, 540–552. [CrossRef] [PubMed]