Insight into the bZIP Gene Family in Solanum tuberosum: Genome and Transcriptome Analysis to Understand the Roles of Gene Diversification in Spatiotemporal Gene Expression and Function

Venura Herath 1,2 and Jeanmarie Verchot 1,*

1 Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX 77802, USA; venura.herath@tamu.edu
2 Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya 20400, Sri Lanka
* Correspondence: jm.verchot@tamu.edu; Tel.: +1-979-845-1788

Abstract: The basic region-leucine zipper (bZIP) transcription factors (TFs) form homodimers and heterodimers via the coil–coil region. The bZIP dimerization network influences gene expression across plant development and in response to a range of environmental stresses. The recent release of the most comprehensive potato reference genome was used to identify 80 StbZIP genes and to characterize their gene structure, phylogenetic relationships, and gene expression profiles. The StbZIP genes have undergone 22 segmental and one tandem duplication events. Ka/Ks analysis suggested that most duplications experienced purifying selection. Amino acid sequence alignments and phylogenetic comparisons made with the Arabidopsis bZIP family were used to assign the StbZIP genes to functional groups based on the Arabidopsis orthologs. The patterns of introns and exons were conserved within the assigned functional groups which are supportive of the phylogeny and evidence of a common progenitor. Inspection of the leucine repeat heptads within the bZIP domains identified a pattern of attractive pairs favoring homodimerization, and repulsive pairs favoring heterodimerization. These patterns of attractive and repulsive heptads were similar within each functional group for Arabidopsis and S. tuberosum orthologs. High-throughput RNA-seq data indicated the most highly expressed and repressed genes that might play significant roles in tissue growth and development, abiotic stress response, and response to pathogens including Potato virus X. These data provide useful information for further functional analysis of the StbZIP gene family and their potential applications in crop improvement.

Keywords: bZIP transcription factor family; Solanum tuberosum; drought stress response; heat stress response; virus stress response; gene regulatory networks; RNA sequence analysis; stress physiology and genetics; potato genome; potato transcriptome; genomic survey; tuber genome

1. Introduction

Broad networks of transcription factors (TFs) exert basal control of gene expression, acting at the core promoter and engaging with RNA polymerase to initiate transcription. Regulatory TFs bind proximal and distal promoter regions to stimulate gene expression in a spatiotemporal and tissue-specific manner. They often recognize consensus sequences in promoters and can act on multiple genes. Such regulatory TFs are the master controllers of transcription networks and are fundamental for plant growth, development, and responses to environmental factors [1,2].

The basic leucine zipper (bZIP) domain TFs exist in all eukaryotes but remains one of the largest groups of TFs in plants [3]. The bZIP family is central to the regulation of developmental and physiological processes as well as abiotic and biotic stress responses [4].
The basic region of 16 amino acid residues contains a nuclear localization signal and an N-x7-R/K motif that binds DNA. The leucine zipper resides toward the C-terminus within an alpha-helical domain and drives protein dimerization [5]. The dimerization of two bZIP factors enables DNA binding [6,7].

Jakoby et al. (2002) were the first to report the Arabidopsis bZIP family contains 75 unique members that classify into ten phylogeny groups [8]. Droge-Laser et al. (2018) updated the classification and described 78 members of the bZIP-family in Arabidopsis and identified 13 groups [4]. The functional groups participate in gene regulation in response to abiotic stress, systemic acquired resistance to a broad spectrum of pathogens, energy metabolism, hypocotyl development, endoplasmic reticulum stress response, abscisic acid response, and virus infection. Plant bZIPs bind to an ACGT core sequence within an A-box (TACGTA), C-box (GACGTC), and G-box (CACGTG) [9].

The bZIP TF family is conserved across eukaryotes and has undergone intensive gene expansion in angiosperms [10]. Diploid angiosperms have undergone at least one duplication event and, polyploids have undergone more duplication events [11,12], making the number of bZIP family members in angiosperms larger than in Drosophila (27 bZIP members) or homo sapiens (53 bZIP members). The number of bZIP family member genes in Arabidopsis is 78, barley is 89, maize is 125, polpar is 99, rapeseed is 247, rice is 89, sorghum is 92, and soybean is 131 [4,10,13–19]. This expansion is attributed to the whole genome, segmental, and tandem duplications. Being sessile, plants have a robust need for expanded stress adaptation. While expanding the number of bZIP family members is one way to broader adaptation, another way to expand stress adaptation can be achieved through complex regulation of gene expression.

Potato, Solanum tuberosum, ranks in the top four most important crops in the world [20]. Potatoes are grown in all environments, latitudes, and hemispheres of the world. With climate change, there is a need to breed new potato varieties that can handle changing environments and warmer temperatures. There is also a need to understand the landscape of gene families that influence growth, development, and adaptive stress responses. The current available complete genome sequence and genome annotation [21] known as DM v6.1 allowed us to carry out a comprehensive identification and analysis of the StbZIP gene family and to conduct comparisons with the Arabidopsis bZIP gene family. In this study, we screened the most recent reference genome assembly and identified 80 StbZIP genes. We provide functional and regulatory classification groups based on the framework for classification of Arabidopsis bZIP family members. This study reports on the analysis of StbZIP gene expression profiles to understand the general involvement of the bZIP TF family in biological processes in Solanum tuberosum.

2. Results and Discussion

2.1. Identification and Phylogenetic Analysis of bZIP Family Members in S. tuberosum

We identified 67 bZIP genes in the S. tuberosum genome (EnsemblPlants SolTub_3.0 assembly) by restricting our query to the predicted domain model (IPR004827). We also identified thirteen additional bZIP genes in the SpudDB (DM1-3 R44 v6.1) which has the newest genome assembly (released September 2020) with higher coverage and resolution (Table 1 and Table S1). We performed reciprocal blast, phylogenetic comparisons, and sequence comparisons between the putative StbZIP family and the well-studied, 78 unique members of the AtbZIP family [4,8,10]. We designated the S. tuberosum genes as StbZIP1 to StbZIP80 according to the corresponding AtbZIP homolog or their locus ID. Table S1 is a compilation of the current locus IDs referenced in the 2020 updated DM v6.1 assembly and SolTub_3.0. We analyzed the amino acid (aa) length, molecular weight (MW), and protein isoelectric (PI) points of the StbZIPs. The StbZIP lengths vary from 91 aa to 876 aa, the MWs range from 10.11 to 96.59 kDa, and the PIs ranged from 4.56 to 10.15 (Table S1).
Figure 1. Phylogenetic tree constructed using PhyML method (v1.5) and Seaview (v4.7) contains the Arabidopsis and potato basic region-leucine zipper (bZIP) proteins. iTOL (v4) was used to visualize the output, and the diagrams were compiled, labeled, and color-coded using Adobe Photoshop CC (2017). Branch support was assessed with 1000 ultrafast bootstrap approximation and SH-aLRT branch test with 1000 replicates. The legend contains the functional groups assigned according to [8,10]. The bootstrap scores are provided at each node.

Table 1. AtbZIP and StbZIP family members organized by functional.

| Group          | Gene Name | Locus ID | Synonym     | GenBank   |
|----------------|-----------|----------|-------------|-----------|
| A              | AtbZIP12  | AT2G44070| DPBF4       | AF334209  |
|                | AtbZIP13  | AT5G44080|            |           |
|                | AtbZIP14  | AT4G39900| FD          | NM119756.5|
|                | AtbZIP15  | AT5G42910| MBD2        | NM123656.2|
|                | AtbZIP27  | AT2G17770| FD paralog  | NM127313.1|
| ABA Responsive | AtbZIP35  | AT1G49720| ABF1        | NM001198254.2|
| Flowering      | AtbZIP36  | AT1G45249| ABF2, ATAREB1| NM001332526.1|
| time           | AtbZIP37  | AT4G34000| ABF3        | NM001342246.1|
| Seed germination| AtbZIP38  | AT3G19290| ABF4, AREB2 | NM001203005.2|
| Abiotic Stress | AtbZIP41  | AT4G36730| GBF1        | NM179179.3|
|                | AtbZIP54  | AT4G01120| GBF2        | NM116342.3|
|                | AtbZIP55  | AT2G46270| GBF3        | NM001337182.1|
|                | AtbZIP68  | AT1G32150|            | NM102948.4|
| G              | AtbZIP16  | AT2G35530| DPBF4       | AF334209  |
|                | AtbZIP41  | AT4G36730| GBF1        | NM179179.3|
|                | AtbZIP54  | AT4G01120| GBF2        | NM116342.3|
|                | AtbZIP55  | AT2G46270| GBF3        | NM001337182.1|
|                | AtbZIP68  | AT1G32150|            | NM102948.4|
|                | AtbZIP12  | AT2G44070| DPBF4       | AF334209  |
|                | AtbZIP13  | AT5G44080|            |           |
|                | AtbZIP14  | AT4G39900| FD          | NM119756.5|
|                | AtbZIP15  | AT5G42910| MBD2        | NM123656.2|
|                | AtbZIP27  | AT2G17770| FD paralog  | NM127313.1|
|                | AtbZIP35  | AT1G49720| ABF1        | NM001198254.2|
|                | AtbZIP36  | AT1G45249| ABF2, ATAREB1| NM001332526.1|
|                | AtbZIP37  | AT4G34000| ABF3        | NM001342246.1|
|                | AtbZIP38  | AT3G19290| ABF4, AREB2 | NM001203005.2|
|                | StbZIP16  | Soltu.DM.08G003670 | | |
|                | StbZIP24  | Soltu.DM.02G006750 | | |
|                | StbZIP41  | Soltu.DM.02G025470 | | |
|                | StbZIP54  | Soltu.DM.05G019830 | | |
| Group          | Gene Name | Locus ID     | Synonym | GenBank   | Group          | Gene Name | Locus ID     | Synonym | GenBank   |
|---------------|-----------|--------------|---------|-----------|---------------|-----------|--------------|---------|-----------|
| AtbZIP39      | AT2G36270 | ABI5         | NM129185.4 |           | SbZIP53      | Soltu.DM.01G3034570 | SbZIP64 |
| AtbZIP40      | AT1G03970 | CRF4         | NM100278.3 |           | SbZIP56      | Soltu.DM.08G021830 |       |
| AtbZIP66      | AT2G3870  | ABE3         | NM115544.3 |           | H            | AtbZIP56   | AT5G11260 HY5 | NM001343175.1 |
| AtbZIP67      | AT3G44460 | DPBF2        | NM114314.5 |           | AtbZIP64     | AT3G17609 HYH | NM001084700.2 |
| SbZIP2        | Soltu.DM.01G005870   |   |       |           | SbZIP56      | Soltu.DM.08G0111730 |       |
| SbZIP3        | Soltu.DM.01G006940   |   |       |           | I            | AtbZIP18   | AT2G40620 | NM129624.5 |
| SbZIP12       | Soltu.DM.07G030340   | ABL2        |       |           | AtbZIP29    | AT4G38900 | NM0011036733.4 |
| SbZIP13       | Soltu.DM.04G021710   |   |       |           | AtbZIP30    | AT2G21230 | NM0011335735.1 |
| SbZIP14       | Soltu.DM.02G022460   |   |       |           | AtbZIP31    | AT2G13150 | NM126912.1   |
| SbZIP27       | Soltu.DM.02G005680   |   |       |           | AtbZIP32    | AT2G12940 UNE4 | NM126904.6 |
| SbZIP35       | Soltu.DM.11G016910   | ABE4        |       |           | AtbZIP33    | AT2G12900 | NM126901.1   |
| SbZIP36       | Soltu.DM.04G003390   | ABE2        |       |           | AtbZIP34    | AT1G3700  | NM100495.4   |
| SbZIP37       | Soltu.DM.01G015000   | ABE3        |       |           | AtbZIP35    | AT1G3700  | NM100495.4   |
| SbZIP38       | Soltu.DM.01G0047570  | ABE1        |       |           | AtbZIP36    | AT1G3700  | NM100495.4   |
| SbZIP39       | Soltu.DM.09G0023620  | ABI5        |       |           | AtbZIP37    | AT1G3700  | NM100495.4   |
| SbZIP40       | Soltu.DM.01G0043800  |   |       |           | AtbZIP38    | AT1G3700  | NM100495.4   |
| SbZIP66       | Soltu.DM.01G0025900  | ABL1        |       |           | AtbZIP40    | AT1G3700  | NM100495.4   |
| B             | AtbZIP17    | AT2G40950    | NM129659.3 |           | SbZIP74    | AT2G21235 | NM179679.2   |
| ER Stress     | AtbZIP28    | AT3G10800    | NM111917.5 |           | SbZIP18    | Soltu.DM.06G011870 |       |
| Responsive     | AtbZIP49    | AT3G56660    | NM115525.2 |           | SbZIP29    | Soltu.DM.01G050330 |       |
|                | SbZIP17     | Soltu.DM.01G0029910 |       |           | SbZIP30    | Soltu.DM.04G026300 |       |
|                | SbZIP28     | Soltu.DM.04G021020 |       |           | SbZIP40    | Soltu.DM.04G026570 |       |
|                 | SbZIP33     | Soltu.DM.04G020920 |       |           | SbZIP41    | Soltu.DM.06G017610 |       |
|                 | SbZIP67     | Soltu.DM.08G001940 |       |           | SbZIP51    | Soltu.DM.04G026570 |       |
|                 | SbZIP70     | Soltu.DM.08G0019530 |       |           | SbZIP52    | Soltu.DM.06G011430 |       |
|                 | SbZIP71     | Soltu.DM.08G0019590 |       |           | SbZIP53    | Soltu.DM.06G014230 |       |
| C             | AtbZIP9     | AT5G24800    | NM122389.4 |           | SbZIP62    | AT5G19490 | NM101806.3   |
| C/S1          | AtbZIP10    | AT4G02640    | NM001340389.1 |           | SbZIP62    | Soltu.DM.10G024960 |       |
| Regulatory     | AtbZIP25    | AT3G54620    | NM115319.4 |           | SbZIP72    | Soltu.DM.09G023820 |       |
| Network,       | AtbZIP63    | AT5G28770    | NM122760.4 |           | K           | AtbZIP60    | AT5G42990 | NM103458.3 |
| Energy        | AtbZIP9     | Soltu.DM.08G002700 |       |           | SbZIP60    | Soltu.DM.04G038150 |       |
| Starvation,    | AtbZIP10    | Soltu.DM.08G0028360 |       |           | M           | SbZIP72    | AT5G07160 | NM120798.2 |
| Seed          | AtbZIP63    | Soltu.DM.01G0036510 |       |           | N           | SbZIP25    | Soltu.DM.03G005150 |       |
| Development    | SbZIP63     | Soltu.DM.08G0008360 |       |           | SbZIP26    | Soltu.DM.03G025200 |       |
| D             | AtbZIP20    | AT3G06950    | NM001036768.2 |       | SbZIP59    | Soltu.DM.06G0232460 |       |
| SA            | AtbZIP21    | ATIC08320    | TGA9    | NM001331769.1 | SbZIP75    | Soltu.DM.11G017940 |       |
| Responsive or  | AtbZIP22    | ATIG22070    | TGA3    | NM102057.4 | S           | AtbZIP1    | AT5G49450 | NM124322.3 |
| Early Flowering| AtbZIP26    | AT3G06960    | CBF5;   | NM203016.2 | K           | AtbZIP3    | AT5G15830 | NM121588.3 |
|                | AtbZIP45    | AT3G12250    | TGA6    | NM202564.2 | S           | AtbZIP4    | AT5G39530 | NM104646 |
|                | AtbZIP46    | ATIG06640    | PAN     | NM105536  | S           | AtbZIP5    | AT3G49760 | NM114836.3 |
|                | AtbZIP47    | AT3G5210     | TGA1    | NM125919.3 |   |           |       |           |
Table 1. Cont.

| Group | Gene Name | Locus ID | Synonym | GenBank  | Group | Gene Name | Locus ID | Synonym | GenBank  |
|-------|-----------|----------|---------|---------|-------|-----------|----------|---------|---------|
| AtbZIP50 | AT1G77920 | TGA7     | NM0013443041.4 | AtbZIP6 | AT2G22850 | TGA7     | NM0013443041.4 |
| AtbZIP57 | AT5G10030 | TGA4     | NM0013443041.4 | AtbZIP7 | AT4G37730 | TGA3     | NM0013443041.4 |
| StbZIP1 | Soluto.DM.01G | 005540 | AtbZIP8 | AT1G68800 | TGA3     | NM0013443041.4 |
| StbZIP20 | Soluto.DM.11G | 019010 | AtbZIP11 | AT4G34590 | TGA3     | NM0013443041.4 |
| StbZIP21 | Soluto.DM.05G | 029750 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP43 | Soluto.DM.05G | 022450 | AtbZIP42 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP45 | Soluto.DM.10G | 026600 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP46 | Soluto.DM.09G | 007000 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP47 | Soluto.DM.12G | 007200 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP50 | Soluto.DM.04G | 025840 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP57 | Soluto.DM.10G | 029320 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP65 | Soluto.DM.10G | 029320 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP74 | Soluto.DM.10G | 027000 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| StbZIP80 | Soluto.DM.11G | 021800 | AtbZIP4 | AT2G35030 | TGA3     | NM0013443041.4 |
| AtbZIP34 | AT2G2380 | TGA7     | NM0013443041.4 | AtbZIP7 | AT4G73270 | TGA7     | NM0013443041.4 |
| AtbZIP61 | AT3G58120 | TGA4     | NM0013443041.4 | AtbZIP8 | AT3G68800 | TGA4     | NM0013443041.4 |
| AtbZIP76 | AT1G58110 | TGA4     | NM0013443041.4 | AtbZIP9 | AT1G68800 | TGA4     | NM0013443041.4 |
| AtbZIP77 | AT1G58110 | TGA4     | NM0013443041.4 | AtbZIP10 | AT1G68800 | TGA4     | NM0013443041.4 |
| AtbZIP79 | AT1G58110 | TGA4     | NM0013443041.4 | AtbZIP11 | AT1G68800 | TGA4     | NM0013443041.4 |
| StbZIP34 | Soluto.DM.12G | 028390 | StbZIP6 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP61 | Soluto.DM.07G | 020910 | StbZIP7 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP64 | Soluto.DM.07G | 020910 | StbZIP8 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP76 | Soluto.DM.10G | 011180 | StbZIP9 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP77 | Soluto.DM.11G | 012470 | StbZIP10 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP78 | Soluto.DM.03G | 042790 | StbZIP11 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| StbZIP79 | Soluto.DM.11G | 007430 | StbZIP12 | Soltu.DM.01G024850 | TGA4     | NM0013443041.4 |
| F | AtbZIP19 | AT1G35040 | TGA7     | NM0013443041.4 | StbZIP13 | Soltu.DM.10G | 016860 | TGA7     | NM0013443041.4 |
| AtbZIP23 | AT2G16770 | TGA7     | NM0013443041.4 | StbZIP14 | Soltu.DM.10G | 016860 | TGA7     | NM0013443041.4 |
| AtbZIP24 | AT3G51960 | TGA7     | NM0013443041.4 | StbZIP15 | Soltu.DM.10G | 016860 | TGA7     | NM0013443041.4 |
| StbZIP19 | Soluto.DM.10G | 017560 | StbZIP16 | Soltu.DM.10G | 016860 | TGA7     | NM0013443041.4 |
| StbZIP23 | Soluto.DM.10G | 017560 | StbZIP17 | Soltu.DM.10G | 016860 | TGA7     | NM0013443041.4 |

Functional Groups, Gene ID, Locus ID, Synonyms and NCBI accessions are provided for all genes used in this study. The color coding for the functional groups correspond to the colors in the legends in Figure 1.

The available literature across angiosperms using neighbor-joining (NJ) and maximum likelihood (ML) phylogenetic analyses has revealed bootstrap support for thirteen groups of unique bZIP transcription factors [8,10]. According to this convention, we performed ML analysis to examine the congruences between the AtbZIP and StbZIP factors and provided the classification of the StbZIP-family within the prescribed 13 functional groups A through M, and S (Figure 1). We identified an additional functional group N that contains...
four StbZIP proteins. The phylogeny shows three deeply rooted branches. The first deep branch bifurcates with group F separating from all other groups. The second bifurcation separates group D, and the third branch expands to all remaining groups. The third branch also bifurcates. One branch extends to groups A, B, E, H, I, K, and N. The other branch extends to groups C, G, J, and S. Each group shows a pattern of AtbZIPs and StbZIPs sharing a common shallow node suggesting these gene pairs are likely orthologs. Other internal nodes seem to represent two or more paralogs suggestive of gene duplication events (Figure 1).

Table 1 shows the gene names and identifiers for the AtbZIPs and StbZIPs used in this study. The biological functions of the AtbZIP protein groups have been extensively studied [4] and used here as the framework for our categorization and discussion of StbZIPs. There are thirteen AtbZIP and thirteen StbZIP members of group A [22]. The Arabidopsis group A includes AtbZIP14/FD, and the paralog AtbZIP27/FD is mainly expressed in the shoot apical meristem and is involved in flowering time. The abscisic acid (ABA)-responsive element (ABRE) binding protein or ABRE binding factor (ABF), and the ABA insensitive/DC3 promoter-binding factors (ABI/DPBF) (Table 1) are essential for stress responses under conditions of dehydration, salinity, or osmotic stress. Mutations in these genes alter plant resiliency to drought, salinity, and osmotic stress. The ABI/DPBF factors regulate seed germination [4,8,23,24]. The S. tuberosum orthologues, which were assigned synonyms in a recent study [22], include StbZIP12, StbZIP35, StbZIP36, StbZIP37, StbZIP38, StbZIP39, and StbZIP66. Group B and group K are closely related and include AtbZIP17, AtbZIP28, AtbZIP49, and AtbZIP60 and have in common the presence of a transmembrane domain [25]. The Arabidopsis TFs form homodimers and heteromers and regulate endoplasmic reticulum (ER) stress. Their transmembrane domains facilitate interactions with other cellular proteins such as AtBAG-7 and BiP [26–28]. The AtbZIP28 also interacts with NF-Y subunits to form a transcription complex suggesting some promiscuity among these group members [29–31]. S. tuberosum has a single ortholog of the AtbZIP60 in group K and six group B factors StbZIP17, StbZIP28, StbZIP33, StbZIP67, StbZIP70, and StbZIP71 (Figure 1). Expansion of StbZIP genes in group B likely occurred by gene duplication [11,12].

The Arabidopsis literature [4,31] describes group C and S as sister clades that cooperate for some biological functions (Figure 1). For Arabidopsis, group S subdivides into subgroups S1, S2, and S3 [31]. The C/S1 network regulates genes that respond to energy starvation, express primarily in sink tissues, are involved in seed and pollen development, and control the expression of seed storage proteins (SSPs) [32–34]. For Arabidopsis, the C and S1 factors do not generally form homodimers, are not promiscuous and, have specificity in their interactions with each other, which is why this is known as the C/S1 gene regulatory network [35].

In the phylogenetic analysis of StbZIPs reported in Figure 1, the group C StbZIP9, StbZIP10, and StbZIP63 share a node with two AtbZIP factors; AtbZIP9 and AtbZIP63. There does not appear to be a direct orthologue in S. tuberosum for AtbZIP25. The Arabidopsis subgroup S1 consists of AtbZIP1, AtbZIP2, AtbZIP11, AtbZIP44, and AtbZIP53, and S. tuberosum has eight S1 members identified as StbZIP4, StbZIP8, StbZIP11, StbZIP15, StbZIP22, StbZIP44, StbZIP53, and StbZIP73. In total, group S has 17 Arabidopsis bZIPs and 15 StbZIPs, which suggests that some genes were lost in potatoes during angiosperm evolution. This loss may relate to different genetic programming in potatoes needed to produce tubers with buds for asexual reproduction. Group D has twelve StbZIP members and includes the Arabidopsis TGACG-binding (TGA) factors, which contribute to cellular defenses [36]. Groups E, F, and G have seven, two, and six StbZIP members, respectively. StbZIP61 is in group E and functions in salicylic acid signaling and defense against Phytophthora infestans [37]. Group H has factors involved in light regulation and anthocyanin accumulation and has one StbZIP member. Groups I, J, M, and N contain eight, one, zero, and four StbZIP members, respectively. Group G and H consist of Arabidopsis members that primarily form homodimers for their group [38].
2.2. Chromosomal Distribution of StbZIPs and Analysis of Gene Duplication Events

We mapped these 80 StbZIP genes onto the 12 chromosomes (Figure 2) [22]. Thirteen StbZIPs map to Chr 1, eight StbZIPs map to Chr 2, four StbZIPs map to Chr 3, fourteen StbZIPs map to Chr 4, three StbZIPs map to Chr 5, seven StbZIPs map to Chr 6, two StbZIPs map to Chr 7 and Chr 12, eight StbZIPs map to Chr 8, two StbZIPs map to Chr 9, eleven StbZIPs map to Chr 10, and six StbZIPs map to Chr 11. We did not see an apparent pattern of members belonging to specific functional groups clustering on chromosomes.

Gene duplication analysis was carried out using MCScanX. We identified 22 StbZIP paralogs (27.5%) that arose by segmental duplication. We calculated the nonsynonymous mutation rate (K_A), synonymous mutation rate (K_S), and K_A/K_S values to study pairs of StbZIP paralogs and to understand the selection pressures affecting sequence divergence. A K_A/K_S > 1.0 indicates positive selection, a K_A/K_S = 1.0 indicates neutral selection, and a K_A/K_S < 1.0 indicates negative or purifying selection. Among the StbZIP pairs tested, the K_A/K_S values (<1) were between 0.14 and 0.48, suggesting purifying selection. Only the StbZIP28 and StbZIP33 paralogs appear to be the result of tandem duplication (Table 2).

Table 2. The nonsynonymous (Ka) and synonymous (Ks) substitution ratio (Ka/Ks) test for 22 gene pairs.

| Duplicated Gene Pair | Ka   | Ks   | Ka/Ks |
|----------------------|------|------|-------|
| StbZIP55/StbZIP54    | 0.18 | 0.63 | 0.29  |
| StbZIP38/StbZIP37    | 0.19 | 0.67 | 0.29  |
| StbZIP38/StbZIP36    | 0.39 | 2.86 | 0.14  |
| StbZIP29/StbZIP30    | 0.36 | 1.88 | 0.19  |
| StbZIP23/StbZIP19    | 0.17 | 0.67 | 0.26  |
| StbZIP27/StbZIP14    | 0.4  | 1.33 | 0.30  |
| StbZIP14/StbZIP37    | 0.69 | 1.43 | 0.48  |
Table 2. Cont.

| Duplicated Gene Pair       | Ka  | Ks   | Ka/Ks |
|----------------------------|-----|------|-------|
| StbZIP11/StbZIP44          | 0.26| 1.17 | 0.23  |
| StbZIP6/StbZIP7            | 0.44| 1.39 | 0.32  |
| StbZIP6/StbZIP5            | 0.26| 0.87 | 0.30  |
| StbZIP5/StbZIP6            | 0.26| 0.87 | 0.30  |
| StbZIP7/StbZIP6            | 0.44| 1.39 | 0.32  |
| StbZIP36/StbZIP38          | 0.39| 2.86 | 0.14  |
| StbZIP44/StbZIP11          | 0.26| 1.17 | 0.23  |
| StbZIP30/StbZIP29          | 0.36| 1.88 | 0.19  |
| StbZIP54/StbZIP55          | 0.18| 0.63 | 0.29  |
| StbZIP61/StbZIP34          | 0.25| 1.1  | 0.23  |
| StbZIP37/StbZIP38          | 0.19| 0.67 | 0.29  |
| StbZIP19/StbZIP23          | 0.17| 0.67 | 0.26  |
| StbZIP61/StbZIP12          | 0.15| 0.57 | 0.26  |
| StbZIP12/StbZIP66          | 0.15| 0.57 | 0.26  |
| StbZIP34/StbZIP61          | 0.25| 1.1  | 0.23  |

2.3. Gene Structure Analysis of StbZIP Genes

The position of introns within a codon phase 0, 1, or 2, were mapped for each StbZIP gene (Figure 3) because the intron positions and the frequency of intron phase combinations in related genes can be evidence of a common progenitor. A non-random pattern of introns indicates that they were acquired from a progenitor and stabilized through evolution [39]. Random phase distribution of introns suggests exon shuffling through evolution, which is typically evidence of new functional elements of protein gene products.

The representative coding sequences (CDS) of the 80 StbZIPs vary from 274 bp to 2631 bases and were used to map the boundaries of genes for structural analysis. The group A genes range from zero to five symmetric exons (0,0), except StbZIP3, which has one asymmetric (1,0) exon. Group B has zero, one, or two phase 1 introns. Group C has three genes with the same pattern of three asymmetric exons (0,1), (1,2), and (2,0) and a 3’ symmetric exon (0,0). StbZIP10 has an added phase 1 intron at the 3’ end. Group D has two intron phase patterns. First is StbZIP1, StbZIP20, StbZIP43, StbZIP47, StbZIP50, StbZIP57 whose 5’ ends have two symmetric exons (0,0) and the 3’ ends have one symmetric exon (0,0) and a central asymmetric exon (0,1). The second pattern is represented by StbZIP14, StbZIP30, and StbZIP44, which have added asymmetric exons at the 5’ end (2,0). The StbZIP80 is an asymmetric exon at the 3’ end (0,2). Group E has two asymmetric exons (2,0) and (0,1), except for StbZIP77, which has an additional 5’ asymmetric exon (0,2). Group F has one phase 1 intron. Group G genes have four central asymmetric exons (0,2), (2,0), (0,1), (1,0) and adjoining one or two 3’ symmetric (0,0) exons. The 5’ has between two and four symmetric (0,0) exons. Group H has two asymmetric exons (1,2) and (2,0). Group I has two asymmetric exons (2,0) and (0,1) except for StbZIP69, which has (0,1). Group J has two genes with different 5’ exon patterns and three symmetric 3’ exons (0,0). Group K has a simple asymmetric (2,1) exon pattern. Group N has asymmetric (2,0) and (0,1) exons. The Arabidopsis and potato group S members are unique intronless genes [35,40], except for StbZIP31, which has a single phase 0 intron. Overall, the common intron patterns and phases provide further support for segmental duplication as the major mechanism for gene expansion (Figure 3).
Figure 3. Exon/intron organization of StbZIP genes depicted for each group. The legend in Figure 2 explains the color scheme highlighting gene names. Boxes represent exons, and lines represent introns. The numbers “0”, “1”, and “2” identify the splicing phases.
2.4. Analysis of the Amino Acid Composition Among Leucine-Rich Repeats belonging to S. tuberosum bZIP TFs

To further validate the genes that we identified in this study encoding bZIP factors, we analyzed the amino acid composition of the basic and leucine-rich repeat domains which define the bZIP family. The leucine zipper governs dimerization and has a structural repeat of heptads. It is standard to designate the amino acid positions using letters a, b, c, d, e, f, and g [6,7]. For Arabidopsis, the bZIP-TFs vary between three heptads and ten or more heptads (Figure 4A) [7,33]. A region of basic amino acids (N X (7) R/K) defines the N-terminal boundary of the leucine zipper. The C-terminal limit can be identified by the presence of proline or other amino acids that disrupt the alpha-helical structure and follow a heptad with the leucine in the d position [6,7]. Identifying the C-terminal disruptor can be difficult based on the amino acid sequence alone. Prior structural studies identified important amino acid features that are essential to forming a productive dimerization interface. For example, the g and e positions typically contain charged residues D, E, K, or R, which provide attractive or repulsive pairing in each heptad. Leucine typically occupies the d position, although other combinations of aliphatic residues can occupy the a and d positions to create the dimerization interface needed to stabilize the leucine zipper [6,33,41].

![Figure 4](image-url)  
Figure 4. Analysis of amino acid composition for heptads. (A) Depiction of bZIP transcription factor binding DNA and dimeric heptads. (B) Pie chart presenting the frequency (%) of amino acids in the g, a, d, e position of heptads from the 80 SthZIP proteins.

We assessed the amino acid content for the bZIP factors of S. tuberosum at the a, d, e, and g positions within the heptads L0 through L4 (Figure 4B). Sixty-five percent of heptads contained L in the d position, consistent with reports for humans and Arabidopsis [6,7]. Considering the g ⇔ e pairs, the presence of D, E, K, and R, which most often drive attractive and repulsive interactions were 43% for the g position and 28% for the e position. Finally, the a ⇔ a’ pairs also provide an energetic contribution to leucine zipper stability.
Homotypic valine, isoleucine, or arginine interactions are favorable, and charged amino acids can be destabilizing. In the ‘a’ position, N⇔V pairs favor heterodimerization, whereas N⇔I pairs are destabilizing. For potato, N, I, V, M occupy the ‘a’ position in 30% of heptads, and charged residues (K, R) are infrequent (Figure 4B). These data are within the realm of expectations for bZIP-family members reported for humans and Arabidopsis [6,7].

We aligned the amino acid sequences of the Arabidopsis and S. tuberosum bZIP factors. Within the bZIP domains, heptads were labeled using the standard designations of a, b, c, d, e, f, and g, counting from the first leucine in the d position of the L0 heptad (Table S2) [4,6,13,22]. The L1 and L2 heptads consist of attractive pairs favoring homodimerization, repulsive pairs favoring heterodimerization, or a combination of attractive and repulsive heptads. It is interesting to note that the patterns of attractive and repulsive heptads were similar in each group for Arabidopsis and S. tuberosum orthologs. Furthermore, the N residues in the ‘a’ position of L2 heptads were often conserved between Arabidopsis and S. tuberosum bZIPS in the same group. These observations suggest that the experimental data documenting interactions among Arabidopsis bZIP proteins may be useful to predict the interactions among S. tuberosum bZIP proteins. There was a broad pattern of early heptads that were similarly attractive or repulsive across each family, although this pattern was not strictly maintained throughout adjoining heptads within each family (Table S2).

Importantly the AREB/ABF/ABI5 transcription factors in group A have four characteristic phosphorylation sites (R-X-X-S/T) across the protein, including one known as the C4 site after the leucine-rich domain [8,22]. Also notable in the alignment, is the StbZIP14 in group A, StbZIP67 in group B, StbZIP1 in group D, and StbZIP26 in group N showed only two or three complete heptads and leucines were lacking in the heptads defining the leucine zipper. For example, the basic regions of StbZIP14 were highly conserved with group A members, but the three heptads did not have leucine in the d position, and the g⇔e’ pairs were not well conserved with other members of the family.

The StbZIP67 in group B has conserved basic region and L0 heptad but the L1 and L2 heptads lacked leucine in the d position. The StbZIP1 in group D lacked L in the d position of L2 and L3. The StbZIP26 in group N did not have leucine in the d position of L2 heptad (Table S2). Possibly, the sequences recovered by database mining produced candidates that show a high degree of similarity but require further refinement built on more extensive functional analysis in future studies. We expect that further refinement of the StbZIP family will be necessary considering that several studies surveying the Arabidopsis genome reported 74, 75, and 78 bZIP-family members [4,8,16]. The most recent report of 78 AtbZIP-family members derives from an enormous amount of functional data used to refine the list of family members [4].

AtbZIPS belonging to group C and subgroup S1 have eight hydrophobic repeats, higher than other groups [7,33]. These AtbZIPS form specific heteromers that cooperate in regulating seed maturation genes and nutrient allocation. Notably, the heptad patterns are conserved among group C and subgroup S1 AtbZIP and StbZIP factors (Table S2).

2.5. Analysis of Conserved Protein Motifs

Grouping of bZIP family members is based on homology of the basic region and other conserved motifs. The heptad analysis identified the StbZIP14 in group A, StbZIP67 in group B, StbZIP1 in group D, and StbZIP26 were lacking heptads defining the leucine zipper. Therefore, we conducted a motif search using MEME Suite software to identify fixed pattern motifs that are present in proteins within these four functional groups (Figure 5). We identified the bZIP domains and labeled the conserved motifs lying toward the N- or C-terminus (Figure 5 and Figure S1). Notably, all groups have the same fixed bZIP segment consisting of a basic region (N-X(7)-R/K) and the L0 heptad, followed by longer leucine-rich motifs that were specific to each functional group. Group A members segregate into two clusters of conserved bZIP motifs. The first cluster is StbZIP2, StbZIP12, StbZIP13, StbZIP27, StbZIP40, and StbZIP66. The second cluster consists of StbZIP35, StbZIP36, StbZIP37, StbZIP38, and StbZIP39. These two clusters also occur in the functional group A.
of the phylogeny. We propose that the conserved motifs outside the bZIP domain account for the functional specificity of each major group (Figure 5).

**Figure 5.** Conserved motifs identified using MEME software suite. All factors, except six, were predicted to have one bZIP domain ($p < 0.001$). The conserved basic N-X (7)-R/K motif is colored red, and adjoining leucine-rich motifs are shaded blue. Figure S1 provides the details for all motifs in this figure.

Each group had at least one member that lacks a fixed basic and leucine-rich motif representative, although containing other conserved motifs. Specifically, the StbZIP14 of group A, StbZIP67 of group B, StbZIP1 in group D, and StbZIP26 of group N lacked the fixed basic and leucine-rich pattern motif. In addition, StbZIP3 in group A and StbZIP71 in group...
B have the conserved basic region but lack the leucine-rich sequences. Importantly, these genes were identified by InterPro as bZIP domain-containing proteins. Their phylogeny and gene structure shows significant conservation with other bZIP family members despite the MEME analysis suggesting that the leucine-rich sequences are not synonymous with these fixed motifs of related factors. Considering the results of motif analysis, six candidate StbZIP TFs may not have the same dimerization properties of other bZIP proteins.

2.6. StbZIP Promoters Enriched with Developmental, Hormone-Response, and Stress-Related TF Binding Sites

To identify the basis of differential gene expression, we selected 2000 bp upstream of the predicted transcription start site for the StbZIP promoters and derived the predicted cis-regulatory elements (CREs). We identified 18 transcription factor binding sites. The CAAT-box is a ubiquitous core element of eukaryotic promoters and is abundant in the StbZIP promoters. The number of CAAT-box sequences ranges from 25 to less than 60 among all StbZIP promoters (Figure 6A,B). Other elements such as the A-box, CAT-box, or HD-ZIP1 core elements occur in 10 and 27 promoters. The CAT-box and HD-ZIP elements are meristems and leaf regulatory elements. The abscisic acid-responsive elements (ABRE) are consistently represented in the majority of bZIP promoters, while only seven promoters have auxin-responsive elements (AuxRR-core). Most promoters contain between one and four copies of the MeJA-responsive and light response elements CGTCA-, G-box, and/or giberillic acid responsive element (GARE)- motifs (Figure 6A). Fourteen promoters have one or two TGA-elements (Figure 6A). There appear to be more stress-responsive elements than hormone regulating elements (Figure 6A,B). There are between 1 and 10 copies of the AE-box, ARE, and/or Box 4 elements across all StbZIPs. More than half the genes in each functional group have between one and four copies of the low temperature responsive element (LTR), MYB binding sequence (MBS), and TC-rich elements.

2.7. Differential Gene Expression in Plant Tissues

The close relationship between the S. tuberosum and Arabidopsis bZIP proteins within each group suggests that their biological roles may be shaped by their evolutionary history [11]. Therefore, we undertook a series of studies to determine whether the StbZIP gene expression profiles can be useful to predict their functions. First, we utilized large transcriptome datasets from publicly available repositories (detailed in Table S3) to produce hierarchical clustering of gene expression and to understand which StbZIPs express to higher levels in particular tissues (Figure 7). We obtained the gene expression profiles of the StbZIP genes in whole RH89-039-16 (RH) genotype plants and fourteen specific tissues: stem, leaf, petioles, roots, flowers, stamens, stolon, young and mature tubers, as well as tuber peel, pith, and cortex. S. tuberosum plants that were grown in the greenhouse, and tissues were harvested at the 12-leaf stage. Stamens were collected from open flowers, tubers were collected from plants after senescence, and sprouts were collected from harvested tubers. The data were compiled from five plants. The expression patterns differed among StbZIP gene families indicating members of multiple functional groups were mutually expressed in various tissues, as expected, given their role in a wide range of physiological processes. Notably, in many experimental studies, bZIP factors a one- or two-fold induction is significant for increased activity and, therefore greater levels of expression or repression point to important affiliations [42,43]. It is also notable that some gene pairs that arose by segmental duplication do not show the same expression patterns indicating functional divergence.
Figure 6. Distribution and frequency of cis-regulatory elements (CREs) on *StbZIP* promoters. Promoter sequences representing 2000 bp from the transcription start site of *StbZIP* genes were extracted from the DM1-3 516 R44 v 6.1 assembly at SpudDB as described in Materials and Methods Section 3.4. (A) The grid provides the number of sites that contain the CREs involved in the development, hormone signaling, and stress responses. Shades of green, red, and blue point to CREs that are highly represented (dark-colored), moderately represented (medium-colored), once or twice represented (light color), or not present (white color). (B) Bar graph demonstrates the distribution frequency of CREs involved in the development, hormone signaling, and stress responses across the 2000 bp sequences analyzed for each *StbZIP*. 
Figure 7. Heat map representation and hierarchical clustering of StbZIP transcript levels across different tissues/organs/developmental stages. The independent gene expression datasets used for this study are provided in Supplementary Table S3. The expression values map to a color gradient from low (blue) to high (red) expression. Values are present using the log2 scale.
On a log2 scale of zero to 10, 34 bZIP genes show moderate expression levels (4.0–6.0) in one specific tissue but the trend toward lower expression (≤2.0) across most other tissues (Figure 7). Examples of moderate expression in one or more specific tissues include: (a) StbZIP54, StbZIP56, and StbZIP62 in tuber sprouts; (b) StbZIP37, StbZIP47, and StbZIP43 in roots and petioles, and (c) StbZIP27 in roots and stolons. Interestingly, StbZIP64 is moderately expressed in the shoot apex, young tubers, and mature tubers. StbZIP77 and StbZIP66 highly (6.0–10.0) expressed in flowers and stamens and low (<2.0) in all other tissues (Figure 7). Group D contains many of the Arabidopsis TGA transcription factors (Table 1) which are essential for salicylic acid signaling, disease resistance, stress mitigation, and flower development. Many of the S. tuberosum orthologues in group D (StbZIP47, StbZIP21, StbZIP65, StbZIP43, StbZIP46, and StbZIP74) are moderately expressed in roots and vegetative tissues but show low expression in the majority of tissues. In group E, the AtbZIP34 is linked to pollen germination [4] and the data show that StbZIP77 which is also in group E is highly expressed in the flower and stamen which might indicate a similar role (Figure 7).

Twenty-one factors were moderately expressed in the majority of tissues (>2.0 to 6.0). These included group A factors StbZIP36, StbZIP40, and StbZIP66; group C StbZIP19; group D StbZIP20, StbZIP50, and StbZIP57; in addition to the S1 group StbZIP53, and StbZIP73. Group I has eight StbZIP proteins, and four of these are moderate to highly expressed across a range of tissues: StbZIP18, StbZIP30, StbZIP48, and StbZIP51.

Seven bZIP factors were moderate or highly expressed (4.0–10.0) across a wide range of tissues and whole plants (Figure 7). By comparing the gene expression profiles with the comparative phylogeny of AtbZIP and StbZIP factors, we can make inferences about the functions of many StbZIP factors. Interestingly, group A StbZIP3 which lacks a full bZIP domain is highly expressed across tissues. The S1 subgroup members StbZIP4, StbZIP15, StbZIP44, are highly expressed (6.0–10.0) in the shoot apex, flower, and tubers, whereas another S1 subgroup member, StbZIP8 is only expressed in the flower and stamen. The StbZIP60 and StbZIP17 genes belonging to groups B and K are moderate to highly expressed in all tissues. Since these genes contribute to ER stress and protein quality control regulation, it makes sense that they are expressed across tissues. Group D contains many of the Arabidopsis TGA transcription factors (Table 1), which are essential for salicylic acid signaling, disease resistance, stress mitigation, and flower development [36]. The group I StbZIP49 is also in this category (Figure 7).

2.8. Expression of StbZIP Family Members Varies from Repression to Activation in Response to Five Hormones

Seven important plant phytohormones regulate essential plant processes. To identify differentially expressed bZIP genes, we utilized a large transcriptome dataset obtained from a publicly available repository (detailed in Table S3). In vitro grown double monoploid potato plants were treated with 10 µM 6-benzyl amino purine (6-BAP), 50 µM ABA, 10 µM indole-3-acetic acid IAA, or 50 µM gibberellic acid (GA3) (Figure 8A). The substance 6-BAP aids in the regulation of cell division and growth and is used in plant tissue culture media to support tissue regeneration. Following treatment with 10 µM 6-BAP, only one gene, StbZIP65 showed a mild increase in transcript accumulation, while the remaining StbZIP genes were repressed. StbZIP65 was also mildly activated by ABA, IAA, and GA3. Looking at the phylogeny StbZIP65 is an orthologue of AtbZIP65/TGA10, a factor that contributes to plant growth and development, hydrogen peroxide-induced responses, and responses to bacterial infection [44,45].
Figure 8. Heat map and hierarchical clustering of StbZIP transcript levels in response to the hormone, abiotic stress, and biotic stress. The independent gene expression datasets are detailed in Supplementary Table S3. The relative changes in bZIP gene expression levels were obtained following treatment with: (A) benzyl amino purine (BAP), abscisic acid (ABA), indole-3-acetic acid (IAA), or gibberellic acid (GA3); (B) heat, water stress, salt, mannitol, and wounding; (C) P. infestans, D,L-β-aminobutyric acid (BABA) or Benzo[1–3]thiadiazole-7-carbothionic acid (BTH). The log2 fold change values map to a color gradient from low (blue) to high (red) expression.

As mentioned previously, ABA is also known as a “stress hormone” involved in adaptive responses to a wide range of abiotic and biotic stresses. Twelve factors were...
were repressed in ABA treated leaves, eight that were unaffected, and the remaining factors were induced. Most notable, are four factors that are significantly repressed (−4.0 to −2.0) StbZIP64, StbZIP61, StbZIP34, and StbZIP11. The most highly induced genes (2.0 to 4.0) are also noteworthy, StbZIP6, StbZIP8, StbZIP19, StbZIP35/AREB4, StbZIP36/AREB2, StbZIP54, StbZIP55, StbZIP62, StbZIP65, and StbZIP70. Following IAA treatment StbZIP8, StbZIP32, StbZIP65, and StbZIP70 were notably induced (1.0 to 4.0) while the majority of bZIP genes were mildly induced. After GA3 treatment there were eleven induced genes (1.0 to 4.0) StbZIP6, StbZIP9, StbZIP11, StbZIP29, StbZIP32, StbZIP36/AREB2, StbZIP54, StbZIP55, StbZIP58, and StbZIP65 (Figure 8A).

2.9. Gene Expression Profiles Show Differential Responses to Various Abiotic and Biotic Stresses

We investigated the StbZIP family genes that are responsive to environmental assaults using high-quality RNA-seq datasets from publicly available repositories (detailed in Table S3). We calculated the relative StbZIPs expression levels in tissues subjected to heat, water, salt, and osmotic (mannitol) stress. Many of the genes that are induced by abiotic stress are also involved in ABA responses, or oxidative stress and defense responses to pathogens. While we expected some similar patterns of gene induction, the transcriptome datasets also identified highly induced and highly repressed genes that were differentially affected by the various abiotic stresses (Figure 8B). For example, six genes were highly repressed following the heat treatment (≤−0.2): StbZIP6, StbZIP19, StbZIP34, StbZIP42, and StbZIP61. Genes that are highly induced by heat stress include StbZIP32, StbZIP11, StbZIP18, StbZIP21, StbZIP43, StbZIP58, StbZIP65, and StbZIP74 (≥2.0), suggesting these are factors contribute to heat response signaling cascades in potato.

We examined the transcriptome profile derived from leaves following wilting (2 days without water) (Figure 8B). Seven factors were marginally repressed including StbZIP11, StbZIP18, StbZIP19, StbZIP42, StbZIP43, StbZIP50, and StbZIP79. At least eighteen factors were induced in response to water stress (≥2.0) including StbZIP2, StbZIP4, StbZIP6, StbZIP8, StbZIP15, StbZIP35/AREB4, StbZIP36/AREB2, StbZIP38/AREB1, StbZIP40, StbZIP41, StbZIP47, StbZIP48, StbZIP52, StbZIP51, StbZIP54, StbZIP55, StbZIP57 and StbZIP60.

Salt and mannitol stress-induced similar genes to different levels (Figure 8B). For example, mannitol and salt stress led to high levels of StbZIP12/ABL2 and StbZIP27. In addition, salt stress led to repression of StbZIP70 and mannitol repressed StbZIP58. The final treatment was wounding. The underside of a primary leaf was wounded, and then the same primary (P) leaf and an upper secondary (S) leaf were harvested after 24 hrs. Seventeen similar bZIP genes were induced in the P and S leaves. Unique to P leaves was the induction of StbZIP7, StbZIP8, StbZIP11, StbZIP22, StbZIP36/AREB2, StbZIP44, StbZIP58, and StbZIP70. There were two uniquely induced genes in S leaves, StbZIP47 and StbZIP48. Eleven StbZIPs were repressed in both tissues.

For biotic stress, detached leaves were treated with P. infestans or D,L-β-aminobutyric acid (BABA) or benzo[1-3]thiadiazole-7-carbothionic acid (BTH) (Figure 8C). Following P. infestans treatment, only StbZIP77 was induced while twelve factors were repressed: StbZIP15, StbZIP19, StbZIP27, StbZIP36, StbZIP43, StbZIP42, StbZIP54, StbZIP55, StbZIP62, StbZIP70, StbZIP73, and StbZIP74. Following treatment with BABA or BTH similar genes were repressed (Figure 8C). Only StbZIP5 and StbZIP63 were induced by BABA, while StbZIP34, StbZIP39, StbZIP64, and StbZIP77 were induced by BTH.

To be comprehensive, we examined four additional datasets (Table S3) although these did not show significant changes in gene expression among bZIP family members. First, four cultivars of potato were subjected to 60 days of drought stress in the greenhouse and field plots, and there were no significant changes in bZIP gene expression (Table S4). In another dataset, plants were subjected to 0, 1, or 5 mg/kg cadmium, and then leaves or roots were harvested. There was no indication that cadmium treatment altered the expression levels of the StbZIP genes (Table S5). Goyer et al. (2015) inoculated Premier Russet (Potato virus Y strain O (PVYO)-resistant) and Russet Burbank (PVY susceptible) potatoes with PVYO or PVYNTN in a greenhouse and harvested leaves at 4 and 10 hrs
Only StbZIP34 was induced in both potatoes inoculated with PVY\textsuperscript{O} at 4 h. In another experiment, potato cyst nematode (Globodera rostochiensis) infective-juveniles were used to infect potato roots (Table S7), and there were no significant changes in bZIP gene expression. In the final dataset, potato tubers were treated with \textit{P. infestans} (Table S8), and only StbZIP5 and StbZIP11 showed significant repression. Considering that these two genes were significantly induced in the leaves following abiotic stress or BABA treatment, it is interesting to see the opposite effect on StbZIP5 expression in the tubers treated with \textit{P. infestans}.

2.10. Gene Expression Profiles Show Differential Responses to Potato Virus X (PVX) Infection

Extensive investigations of ER stress responses to potexvirus infection in Arabidopsis determined that the viral encoded activates the unfolded protein response (UPR), leading to transiently increased accumulation of \textit{AtbZIP60} and \textit{AtbZIP17} transcripts between 2- and 5-days post-inoculation (dpi). Arabidopsis is a host for a related potexvirus, \textit{Plantago asiatica mosaic virus}, and in \textit{atbzip60}, \textit{atbzip17}, and \textit{atbzip60atbzip17} knock out mutant lines, virus accumulation is higher than in wild-type Arabidopsis plants [27]. In \textit{N. benthamiana} plants that were silenced for \textit{bZIP60}, the PVX and PVY infection levels were elevated compared to control leaves [47]. These combined data demonstrate a role for these two bZIP factors in suppressing virus accumulation in plants. In this study, we examined transcript profiles derived from leaves that were inoculated with PVX and then harvested at 2 and 3 dpi to identify the StbZIP factors whose expression is altered early in virus infection.

RNA-seq analysis yielded ~453 million reads, and approximately 407 million reads were mapped to the reference potato genome with an average of 89.75% alignment rate (Table S9). We estimated the gene-level transcript abundance using RNA-Seq read counts. Among the differentially upregulated genes, six genes belonged to the potato bZIP family (Figure 9). We were surprised to see that StbZIP72 was suppressed at 3 dpi, and five other StbZIP factors were elevated as an early response to infection, StbZIP37/AREB3, StbZIP42, StbZIP46, StbZIP58, and StbZIP61. These factors were not identified as suppressed or induced in the datasets involving biotic stressors BTH, BABA, or \textit{P. infestans} suggesting these are virus-specific response factors. The StbZIP72 belongs to group J, and the only Arabidopsis member belongs to this group, \textit{AtbZIP62} was involved in oxidative stress and drought stress signaling [48]. The group A StbZIP37 was assigned the synonym of AREB3 in a recent study [22]. However, our analysis in this study identifies it as a putative ortholog of AtbZIP37/ABF3, a TF involved in the ABA-mediated signaling pathway that helps plants to acquire tolerance to drought, salt, cold head, and oxidative stress. The StbZIP37 likely has overlapping roles with other ABF/AREB family members [49, 50].
3. Materials and Methods

3.1. Sequence Retrieval and Domain Characterization

We retrieved the bZIPs in the earlier assembly of double monoploid *S. tuberosum* group Phureja DM1-3 (genome assembly SolTub_3.0) and the Arabidopsis bZIPs (genome assembly TAIR10) from Ensemble Plants database release 44 (http://plants.ensembl.org/) based on the presence of a bZIP domain (InterPro ID #IPR004827) [20,21,51–53]. We also retrieved the StbZIPs in the latest genome assembly (ver 6.1) of potato (DM 1-3 516 R44) which we downloaded from the SpudDB (http://solanaceae.plantbiology.msu.edu/, last accessed 06.10.2020) [21]. We discovered bZIPs in the early assembly and recent assembly of potato and compared them. We determined that the new assembly has better coverage and quality and therefore gave the focus of our work to the latest assembly for further analysis of the bZIP family members.

Representative protein models were used to identify conserved protein domains obtained from the databases Pfam, ProSite Profiles, and SMART using InterPro v5.45–80.0 [54–57]. Only proteins with bZIP domains (InterPro ID #IPR004827) were further analyzed (Table S1). Candidate sequences were chosen based on an E-value of $\leq1 \times 10^{-25}$.

3.2. Phylogeny, Chromosomal Locations, Gene Duplications, and Intron/Exon Gene Structure Analysis

Multiple sequence analysis was carried out using MAFFT server v7 (https://mafft.cbrc.jp/alignment/server/, last accessed 07.10.2020) using the E-INS-i Iterative refinement method. TrimAl v1.2 was used to remove the unambiguous alignments using gappyout trimming mode. The Maximum likelihood (ML) phylogenetic tree was generated using IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/; last accessed 10.10.2020) using JTT+G4.
amino acid substitution model with 1000 ultrafast bootstraps and SH-aLRT branch test with 1000 replicates [58–61]. The ML phylogenetic tree was visualized using iTOL (v4) [62]. Potential gene duplication events were identified using MCSnax tools embedded in TBTools v1.051 [63,64]. This analysis used a tabulated BLASTP input file generated using legacy_blast.pl script of NCBI blast 2.90.+ using the following settings: blastp -e 1e-10 -b 5 -v 5 -m 8, and a gene finding format (GFF) file retrieved from the SpudDB (http://solanaceae.plantbiology.msu.edu/; last accessed 06.10.2020). The divergence analysis of paralogues StbZIP genes was carried out using the Ka/Ks calculator (Nei and Gojobori (NG) method) included in the TBTools ver 1.051.

The intron–exon structures and the intron phases of bZIP transcription factor genes were organized by aligning each CDS with their corresponding genomic sequences and visualized using Gene Structure Display Server (GSDS) 2.0 (http://gsds.gao-lab.org/, last accessed 08.10.2020) [65].

3.3. Amino acid Sequence Alignments and Analysis of the Conserved basic and Leucine-rich Domain

CLC Genomics Workbench 8.0 was used to visualize multiple protein sequence alignments. The basic region and C-terminal leucine-rich heptads were extracted into EXCEL spreadsheets (Supplementary Table S2) and conserved residues were manually analyzed according to [6,7]. The transmembrane domains were predicted using Protter v1.0, TMHMM Server v2.0 (http://www.cbs.dtu.dk/services/TMHMM/; last accessed 10.10.2020), and TOPCONS server [66,67]. The molecular weight (MW) of each StbZIP protein was calculated using COPid server (http://crdd.osdd.net/raghava/copid/; last accessed 10.10.2020) [68] and the isolectric points were calculated using IPC calculator v1.0 using EMBoss pKa set [69]. Conserved protein motifs among StbZIP families were analyzed using MEME Suite v5.2.0 (http://meme-suite.org/tools/meme; last accessed on 23.10.2020) [70]. The classic motif discovery mode was used to obtain the distribution pattern of zero to one occurrence per sequence (zoops). The maximum motif parameter was set to 50, and the motif width between 6 to 70 amino acids. Data were organized using Adobe Illustrator CC (2020).

3.4. Ab Initio Promoter Analysis

Promoter sequences representing 2000 bp from the transcription start site of StbZIP genes were extracted from the DM 1-3 516 R44-v6.1 assembly available at SpudDB (http://solanaceae.plantbiology.msu.edu/data/DM_1-3_516_R44_potato_genome_assembly.v6.1.fa.gz; last accessed on 26.10.2020). The locus annotations were provided (http://solanaceae.plantbiology.msu.edu/data/DM_1-3_516_R44_potato_genome_assembly.v6.1.refGenemodels.gff3.gz; last accessed on 26.10.2020). CREs were analyzed against 417 CREs derived from monocotyledonous species (150), dicotyledonous species (263), and conifers (4) using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [71]. CREs were categorized based on their involvement in plant growth, hormonal regulation, and stress responses.

3.5. In Silico bZIP Gene Expression Analysis

Multiple independent gene expression datasets from the potato genotypes DM1-3 516 R44 genotype and RH89-039-16 reported in [20,72–75] and available in publicly accessible repositories were used in this study (Supplementary Table S3). Three datasets derived from RH genotype potato plants (Accession E-MTAB-552) were used to analyze bZIP expression profiles related to various tissues and developmental stages and were obtained from the Expression Atlas database (https://www.ebi.ac.uk/gxa/; last accessed 14.09.2020). The first dataset was obtained from RH plants that were grown in soil-filled pots in the greenhouse and at the 12th leaf stage, various tissues were sampled from five plants including fully expanded leaves, shoot apex, petioles, stem, mature tubers, and roots. Stamens were collected from fully open flowers. Non-tuberizing stolons were harvested and then young tubers that were ≤1 cm in size were collected for gene expression profiles at one week after the first swellings. Mature tubers were collected from senescing plants and then the peel, cortex, and
pith were sampled. Tuber sprouts were obtained after storage in the dark for 3–4 months. The second dataset was obtained from RH plants that were grown in vitro on media. The third dataset was from water-stressed RH plants that were grown in the greenhouse and denied water for two days. Then the 2nd–4th fully expanded, wilted leaves were harvested. The bZIP transcriptomic profiles of DM potato (Accession E-MTAB-552, −553, −554, −555) following treatment with various hormones or abiotic stressors (Supplementary Table S3) were downloaded from SpudDB (http://solanaceae.plantbiology.msu.edu/index.shtml; DM_RH_RNA-Seq_FPKM_expression_matrix_for_DM_v4.03_132dec2013_desc.xlsx; last accessed date 14.09.2020) [20,73]. These gene expression datasets were generated using in vitro grown DM plants maintained at 22 °C day/18 °C night with a 16 h photoperiod. Plants were treated for 24 h and then roots and shoots were harvested together for RNA extraction. Transcriptome datasets were obtained following treatment with abscisic acid (ABA; 50 µM), indole-3-acetic acid (IAA; 10 µM), gibberellic acid (GA3; 50 µM), and 6-benzyl amino purine (6-BAP; 10 µM). Other abiotic stress conditions included heat (35 °C), salt (150 mM NaCl), and mannitol (260 µM) [20]. For wounding, the bottom two leaflets were mechanically injured, and then the primary leaflets and secondary non-wounded leaflets were harvested at 24 h.

The bZIP transcriptomic profiles of DM potato under biotic stress (Supplementary Table S3) were also retrieved from SpudDB (E-MTAB-552, EMTAB-4301, and E-MTAB-5215) [20,73,74]. In these experiments, six leaves were detached from greenhouse-grown DM plants and then spray-inoculated with *P. infestans* (Fi isolate US8: Pi02-007) using 0.5 mL of inoculum concentration of 30,000 sporangia/mL, acibenzolar-S-methyl (BTH; 100 mg/mL) or DL-B-amino-n-butyric acid (BABA; 2 mg/mL), and a mock inoculation was also conducted. Inoculated leaves were kept in the dark at room temperature for 8–10 h and then under lights for seven days. Infection experiments were repeated three times. Tissues from inoculated and mock-inoculated leaves were collected at 24, 48, and 72 hrs. for RNA isolation. RNAs were pooled.

One independent dataset (E-GEOID-77826) belongs to a study of greenhouse and field-grown plants belonging to four commercial cultivars ‘Alegria’, ‘Milva’, ‘Desiree’, and ‘Saturna’ [75]. Plants were subjected to drought stress at 12 days after transfer to pots in the greenhouse by withholding water. After 20 days plants received 30% of the water given to control plants. RNA was extracted from leaves of two to four replicates per cultivar. Four experimental datasets (E-MTAB-771, E-MTAB-4301, E-MTAB-5215, SRP058212, and SRP058230) were analyzed to obtain gene expression patterns following treatment with cadmium, *P. infestans*, *G. rostochiensis*, and potato virus Y (Supplementary Table S2).

Expression analysis was performed using each independent dataset. For hierarchical clustering, we relied on TBTools ver 1.0532 to perform complete linkage clustering and Euclidean distance measures. Then the data were visualized using the HeatMap tool built into TBTools ver 1.0532 [63].

3.6. *PVX* Inoculation of Potato Leaves for Transcriptomic Analysis

*S. tuberosum* cultivar “Russet Norkota” was multiplied by cuttings or in vitro on Murashige and Skoog medium. Rooted cuttings were grown in a growth room with a 12 h photoperiod at 20 °C for four weeks. The PVX-GFP infectious clone is maintained in *Agrobacterium tumefaciens* strain GV3101 [76]. A fresh culture originated from a single colony was used for to agro-infiltrate plants using standard methods [27]. Inoculated *Nicotiana benthamiana* plants grown under 12 h light at 20 °C. Then upper leaves were harvested after the appearance of symptoms (2 weeks), ground 1:10 (w/v) in 0.01 M phosphate buffer (pH 7.0), centrifuged at 6000 rpm for 5 min, and then the soluble phase (sap extract) was stored at −80 °C. Standard infectivity assays were carried out to estimate the amount of infectious virus in the sap preparation used, to ensure future reproducibility [77]. *Chenopodium quinoa* leaves (n = 6) were rub-inoculated with 20 µL of sap and the numbers of chlorotic foci were counted after 7–12 days. The average number of foci across three leaves was 44. After determination of the infectivity, three potato leaves (cultivar ‘Russett Norkota’)
were mechanically inoculated with 20 µL of sap after dusting with carborundum. Mock treatment was carried out using only the phosphate buffer (three biological replicates for PVX infected and mock treated plants). To study the transcriptomic changes in early infected leaves at 2 and 3 dpi, inoculated leaf samples were collected and immediately frozen in liquid nitrogen, and stored at −80 °C freezer for use in transcriptomic studies.

3.7. Transcriptomic Analysis

To investigate early changes in gene expression, we used the inoculated russet potato leaves that were harvested at 2 dpi and 3 dpi following PVX-GFP inoculation and mock treatment, and were stored at −80 °C. Three frozen leaves from each PVX-GFP inoculated or mock treated russet potato plants were combined and homogenized for RNA extraction (3 biological replicates). The RNeasy Mini Kit (Qiagen Co., Hilden, Germany) was used to extract total RNA. RNA purity was assessed using Epoch 2 Microplate Spectrophotometer (BioTek Instruments Inc., VT, USA). All samples produced A260/A280 ratios ranging between 1.9–2.1. RNA integrity was assessed using Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and all samples had an RNA integrity number (RIN number) >7.3.

The mRNA purification, fragmentation, cDNA synthesis, second-strand synthesis, adapter ligation, cDNA library purification, and transcriptomic sequencing were performed at the Beijing Genomics Institute (BGI, Shenzhen, China) using the BGISEQ-500 platform. BGI performed PE150 strand-specific library preparation, generated raw data, and provided clean reads as follows. First, the poly-A-containing mRNA was purified using oligo(dT)-coupled magnetic beads. Then mRNA fragmentation was carried out using divalent cations under elevated temperature. The cleaved fragments were converted into the first-strand cDNA using reverse transcriptase and random primers. Then second-strand cDNA synthesis was used by applying DNA polymerase I and incorporating dUTP (2′-deoxyuridine 5′-triphosphate) in place of dTTP (2′-deoxyguanosine 5′-triphosphate) to generate double stranded cDNA. The final cDNA library was generated by purifying and PCR enriching the product from the earlier step. Using a rolling-circle replication mechanism, single-stranded DNA circles containing DNA nanoballs were generated. Then the DNA nanoballs were then loaded into patterned nanoarrays, and paired-end reads of 150 bp were generated with the BGISEQ-500. The raw data with adapter sequences or low-quality sequences were filtered using SOAPnuke (v2.1.0) [78]. FASTQC was used to assess read qualities (version 0.11.9). The subsequent analysis returned clean reads.

Reference guided mapping was carried out using the latest genome assembly (DM v6.1) of potato [21]. Reads from PVX-infected and mock-treated datasets at 2 and 3 dpi were aligned to the potato reference genome (DM v6.1) using HISAT2 (v2.2.0). The SAM files were converted to BAM files and indexed using SAMtools (v1.9) [79]. Transcripts assembly and abundance were determined using StringTie (v2.1.4) [80] and using the annotations obtained from the reference genome (DM v6.1). Then the results were converted to DESEQ2 format using prepDE.py python scripts available with the program. Differential sequence analysis was carried out using DESEQ2 (v1.28.1) in RStudio (v1.3.959) [81]. Differentially regulated bZIP genes with ≤−1.2 or ≥1.2 log2-fold difference with an adjusted emph:p-value of ≤0.05 at each 2 dpi and 3 dpi were used for the visualization using the HeatMap tool built into TBTools v1.0532 [63].

4. Conclusions

In this study, we identified 80 bZIP genes in *S. tuberosum*. The results revealed the structural and functional diversification of bZIPS in potato. It is evident that gene duplications have contributed towards the expansion of the bZIP family. During its evolutionary trajectory, the bZIP gene family expanded to many groups and these groups were mostly conserved between the potato and Arabidopsis, indicating the functional importance of the family in growth, development, and stress responses. We also identified a novel bZIP gene group (N) consisting of four genes, while group M of Arabidopsis has been lost in
potato. The presence of multiple members in each group (except for group K) indicates the functional redundancy and differential expression patterns observed in our study. The identification of the bZIP gene family in potato will act as the first step towards structural and functional characterization of the bZIP family. Together with the recent reports describing quantitative trait loci associated with the key developmental and stress-related traits [82–86], these findings will help breeders to develop future-proof potato varieties with enhanced yield potential.

Supplementary Materials: The following are available online at https://www.mdpi.com/1422-0067/22/1/253/s1; Figure S1. Motif sequences identified by MEME Tools; Table S1. Compilation of SthbZIP genes, their identifiers, and protein properties; Table S2. Brief alignment of the basic region and leucine-rich domain of Arabidopsis and S. tuberosum bZIP factors; Table S3. Compiled independent transcriptome datasets used for in silico bZIP gene expression analysis; Table S4. Relative expression levels of greenhouse (GH) and field-grown (FG) plants treated with 60 days of drought stress; Table S5. Relative gene expression profile of potato plants treated with 0, 1, and 5 mg/kg cadmium; Table S6. Relative gene expression profiles of early stages of potato virus Y infection in resistant and susceptible potato varieties; Table S7. Relative gene expression profile of potato plants treated with G. rostochiensis; Table S8. Relative gene expression profiles of P. infestans inoculated potatoes and RB transgenic potatoes; Table S9. Summary of the RNA-sequencing read data.

Author Contributions: V.H. compiled lists of genes, identifiers, protein properties. He analyzed the phylogenies, intron/exon patterns, Ka/Ks, and chromosome distribution. He compiled and analyzed the independent transcriptome datasets. V.H. purified RNA from infected potato plants and sent it to Beijing Genome Institute (BGI) for transcriptomics, and performed analysis on retrieved datasets. J.V. performed protein alignments, bZIP protein domain analyses to identify the amino acid composition, heptad characterizations, cross-referencing proteins with the literature to cross-reference synonyms, worked on MEME analysis, and intron/exon phase analysis. Both authors compiled and edited figures. J.V. carried out infectivity assays to standardize inoculum used on potato plants for RNA-seq. J.V. and V.H. wrote and edited the manuscript. Both authors have read and agreed to publish this version of the manuscript.

Funding: This work was supported by a grant from NSF (IOS #1759034). Portions of this research were conducted with high-performance research computing resources provided by Texas A&M University (https://hprc.tamu.edu).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Transcriptomic data are available on the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA: https://www.ncbi.nlm.nih.gov/sra) under the bioproject PRJNA679820.

Acknowledgments: We appreciate the support of the Texas A&M AgriLife Genomics and Bioinformatics Service (TxGen).

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

References
1. Rushton, P.J.; Bokowiec, M.T.; Han, S.; Zhang, H.; Brannock, J.F.; Chen, X.; Laudeman, T.W.; Timko, M.P. Tobacco Transcription Factors: Novel insights into transcriptional regulation in the Solanaceae. Plant Physiol. 2008, 147, 280–295. [CrossRef] [PubMed]
2. Shiu, S.-H.; Shih, M.-C.; Li, W.-H. Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol. 2005, 139, 18–26. [CrossRef]
3. Jindrich, K.; Degnan, B.M. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. BMC Evol. Biol. 2016, 16, 28. [CrossRef]
4. Dröge-Laser, W.; Snoek, B.L.; Snel, B.; Weiste, C. The Arabidopsis bZIP transcription factor family—An update. Curr. Opin. Plant Biol. 2018, 45, 36–49. [CrossRef] [PubMed]
5. Li, Q.; Xiong, L.; Gao, J.; Zhang, H.-Y. Molecular mechanisms of the protein-protein interaction–regulated binding specificity of basic-region leucine zipper transcription factors. J. Mol. Model. 2019, 25, 246. [CrossRef] [PubMed]
6. Vinson, C.; Acharya, A.; Taparovsky, E.J. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochim. Biophys. Acta Gene Struct. Expr. 2006, 1759, 4–12. [CrossRef] [PubMed]
6. Deppmann, C.D. Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: A comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res.* 2004, 32, 3435–3445. [CrossRef]

7. Jakoby, M.; Weisshaar, B.; Dröge-Laser, W.; Vicente-Carbajosa, J.; Tiedemann, J.; Kroy, T.; Parcy, F. bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* 2002, 7, 106–111. [CrossRef]

8. Foster, R.; Izawa, T.; Chua, N. Plant bZIP proteins gather at ACGT elements. *FASEB J.* 1994, 8, 192–200. [CrossRef]

9. Corrêa, L.G.G.; Riaño-Pachón, D.M.; Sbagh, C.G.; Vicentini dos Santos, R.; Mueller-Roelder, B.; Vincentz, M. The Role of bZIP transcription factors in green plant evolution: Adaptive features emerging from four founder genes. *PLoS ONE* 2008, 3, e2944. [CrossRef]

10. Panchy, N.; Letti-Shiu, M.; Shi, S.-H. Evolution of gene duplication in plants. *Plant Physiol.* 2016, 171, 2294–2316. [CrossRef] [PubMed]

11. Afrin, T.; Diwan, D.; Sahawneh, K.; Pajerowska-Mukhtar, K. Multilevel regulation of endoplasmic reticulum stress responses in Arabidopsis. *J. Integr. Plant Biol.* 2020, 8, 189–195. [CrossRef] [PubMed]

12. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 2019, 20, 38. [CrossRef] [PubMed]

13. Nijhawan, A.; Jain, M.; Tyagi, A.K.; Khurana, J.P. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* 2008, 146, 333–350. [CrossRef] [PubMed]

14. Wei, K.; Chen, J.; Wang, Y.; Chen, Y.; Chen, S.; Lin, Y.; Pan, S.; Zhong, X.; Xie, D. Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res.* 2012, 19, 463–476. [CrossRef] [PubMed]

15. Pourabed, E.; Ghane Golmohamadi, F.; Soleymani Monfared, P.; Razavi, S.M.; Shobbar, Z.-S. Basic leucine zipper family in barley: Genome-wide characterization of members and expression analysis. *Mol. Biotechnol.* 2015, 57, 12–26. [CrossRef] [PubMed]

16. Ji, Q.; Zhang, L.; Wang, Y.; Wang, J. Genome-wide analysis of basic leucine zipper transcription factor family in Arabidopsis thaliana, Oryza sativa and Populus trichocarpa. *J. Shanghai Univ.* 2009, 13, 174–182. [CrossRef]

17. Zhou, Y.; Xu, D.; Jia, L.; Huang, X.; Ma, G.; Wang, S.; Zhu, M.; Zhang, A.; Guan, M.; Lu, K.; et al. Genome-wide identification and structural analysis of bZIP transcription factor genes in Brassica napus. *Genes* 2017, 8, 288. [CrossRef]

18. Wang, J.; Zhou, J.; Zhang, B.; Vanitha, J.; Ramachandran, S.; Jiang, S.-Y. Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on Sorghum. *J. Integr. Plant Biol.* 2011, 53, 212–231. [CrossRef] [PubMed]

19. Zhang, M.; Liu, Y.; Shi, H.; Guo, M.; Chai, M.; He, Q.; Yan, M.; Cao, D.; Zhao, L.; Cai, H.; et al. Evolutionary and expression analyses of soybean basic Leucine zipper transcription factor family. *BMC Genom.* 2018, 19, 159. [CrossRef] [PubMed]

20. Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S.; Li, R.; Wang, J.; et al. Genome sequence and analysis of the tuber crop potato. *Nature* 2011, 475, 189–195. [CrossRef] [PubMed]

21. Pham, G.M.; Hamilton, J.P.; Wood, J.C.; Burke, J.T.; Zhao, H.; Vaillancourt, B.; Ou, S.; Jiang, J.; Robin Buell, C. Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* 2020, 9, 1–11. [CrossRef] [PubMed]

22. Liu, T.; Zhou, T.; Lian, M.; Liu, T.; Hou, J.; Ijaz, R.; Song, B. Genome-wide identification and characterization of the AREB/ABF/ABI5 subfamily members from *Solanum tuberosum*. *Int. J. Mol. Sci.* 2019, 20, 311. [CrossRef] [PubMed]

23. Yoshida, T.; Fujita, Y.; Maruyama, K.; Mogami, J.; Todaka, D.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. *Plant Cell Environ.* 2015, 38, 35–49. [CrossRef] [PubMed]

24. Banerjee, A.; Roychoudhury, A. Abscisic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress. *Protoplasma* 2017, 254, 3–16. [CrossRef]

25. Srivastava, R.; Li, Z.; Russo, G.; Tang, J.; Bi, R.; Muppirala, U.; Chudalayandi, S.; Severin, A.; He, M.; Vaitkevicius, S.I.; et al. Response to persistent er stress in plants: A multiphasic process that transitions cells from prosurvival activities to cell death. *Plant Cell* 2018, 30, 1220–1242. [CrossRef]

26. Afrin, T.; Diwan, D.; Sahawneh, K.; Pajerowska-Mukhtar, K. Multilevel regulation of endoplasmic reticulum stress responses in plants: Where old roads and new paths meet. *J. Exp. Bot.* 2020, 71, 1659–1667. [CrossRef]

27. Gayral, M.; Arias Gaguncela, O.; Vasquez, E.; Herath, V.; Flores, F.J.; Dickman, M.B.; Verchot, J. Multiple ER-to-nucleus stress signaling pathways are activated during Plantago asiatica mosaic virus and Turnip mosaic virus infection in Arabidopsis thaliana. *Plant J.* 2020, 103, 1233–1245. [CrossRef] [PubMed]

28. Tajima, H.; Iwata, Y.; Ivano, M.; Takayama, S.; Koizumi, N. Identification of an Arabidopsis transmembrane bZIP transcription factor involved in the endoplasmic reticulum stress response. *Biochem. Biophys. Res. Commun.* 2008, 374, 242–247. [CrossRef]

29. Liu, J.X.; Howell, S.H. bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress responses in Arabidopsis. *Plant Cell* 2010, 22, 782–796. [CrossRef]

30. Williams, B.; Kabbage, M.; Brit, R.; Dickman, M.B. AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc. Natl. Acad. Sci. USA* 2010, 107, 6088–6093. [CrossRef]

31. Srivastava, R.; Deng, Y.; Shah, S.; Rao, A.G.; Howell, S.H. Binding protein is a master regulator of the endoplasmic reticulum stress sensor/transducer bZIP28 in Arabidopsis. *Plant Cell* 2013, 25, 1416–1429. [CrossRef]

32. Alonso, R.; Oríate-Sánehez, L.; Weltmeier, F.; Ehler, A.; Diaz, I.; Dietrich, K.; Vicente-Carbajosa, J.; Dröge-Laser, W. A Pivotal Role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* 2009, 21, 1747–1761. [CrossRef] [PubMed]

33. Ehler, A.; Weltmeier, F.; Wang, X.; Mayer, C.S.; Smeekens, S.; Vicente-Carbajosa, J.; Dröge-Laser, W. Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: Establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J.* 2006, 46, 890–900. [CrossRef] [PubMed]
