Scientific Research Publishing

American Journal of Molecular Biology, 2022, 12, 203-244
https://www.scirp.org/journal/ajmb
ISSN Online: 2161-6663
ISSN Print: 2161-6620

Cloning and Characterization of Eukaryotic Translation Initiation Factor 4E (eIF4E) Gene Family in *Ipomoea batatas* L. (Lam) for Understanding Hexaploid Sweetpotato-Virus Interactions

Adrianne P. A. Brown[1,2], Marceline Egnin[1], Foaziatu Bukari[3], Osagie Idehen[1], Inocent Ritte[1,2], Desmond Mortley[1], Gregory Bernard[1], Deloris Alexander[2], Conrad Bonsi[1]

[1]Plant Biotech and Genomics Research Lab, Department of Agriculture and Environmental Sciences, College of Agriculture, Environment and Nutrition Sciences (CAENS), Tuskegee University, Tuskegee, USA
[2]Integrative Biosciences PhD Program, Tuskegee University, Tuskegee, USA
[3]Heliae Development, LLC, Gilbert, USA
Email: abrown9633@tuskegee.edu, megnin@tuskegee.edu

How to cite this paper: Brown, A.P.A., Egnin, M., Bukari, F., Idehen, O., Ritte, I., Mortley, D., Bernard, G., Alexander, D. and Bonsi, C. (2022) Cloning and Characterization of Eukaryotic Translation Initiation Factor 4E (eIF4E) Gene Family in *Ipomoea batatas* L. (Lam) for Understanding Hexaploid Sweetpotato-Virus Interactions. *American Journal of Molecular Biology*, **12**, 203-244.
https://doi.org/10.4236/ajmb.2022.124017

Received: July 30, 2022
Accepted: October 28, 2022
Published: October 31, 2022

Abstract

Characterization of genes related to sweetpotato viral disease resistance is critical for understanding plant-pathogen interactions, especially with feathery mottle virus infection. For example, genes encoding eukaryotic translation initiation factor (*eIF*)4*E*, its isoforms, *eIF*(*iso*)4*E*, and the cap-binding protein (*CBP*) in plants, have been implicated in viral infections aside from their importance in protein synthesis. Full-length cDNA encoding these putative eIF targets from susceptible/resistant and unknown hexaploid sweetpotato (*Ipomoea batatas* L. Lam) were amplified based on primers designed from the diploid wild-type relative *Ipomoea trifida* consensus sequences, and designated *Ib*eIF4E, *Ib*eIF(iso)4E and *Ib*CBP. Comparative analyses following direct-sequencing of PCR-amplified cDNAs versus the cloned cDNA sequences identified multiple homeoalleles: one to four *IbeIF4E*, two to three *IbeIF*(*iso*)4*E*, and two *IbCBP* within all cultivars tested. Open reading frames were in the length of 696 bp *Ib*eIF4E, 606 bp *Ib*eIF(iso)4E, and 675 bp *Ib*CBP. The encoded single polypeptide lengths were 232, 202, and 225 amino acids for *Ib*eIF4E, *Ib*eIF(iso)4E, and *Ib*CBP, with a calculated protein molecular mass of 26 kDa, 22.8 kDa, and 25.8 kDa, while their theoretical isoelectric points were 5.1, 5.57, and 6.6, respectively. Although the homeoalleles had similar sequence lengths, single nucleotide polymorphisms and multi-allelic variations were detected within the coding sequences. The multi-sequence

alignment performed revealed a 66.9% - 96.7% sequence similarity between the predicted amino acid sequences obtained from the homeoalleles and closely related species. Furthermore, phylogenetic analysis revealed ancestral relationships between the eIF4E homeoalleles and other species. The outcome herein on the eIF4E superfamily and its correlation in sequence variations suggest opportunities to decipher the role of eIF4E in hexaploid sweetpotato feathery mottle virus infection.

Keywords

Ipomoea batatas, Eukaryotic Translation Initiation Factors, eIF4E, CBP, eIF(iso)4E, Sweetpotato Viral Diseases

1. Introduction

As the only edible crop within the *Convolvulaceae* family, sweetpotato (*Ipomoea batatas* L. Lam) is the seventh most important agricultural crop in terms of production [1]. Grown in over 100 countries, sweetpotato is not only an industrial commodity but also an important staple food (leaves and storage roots) worldwide. It boasts high starch and nutrient contents such as fiber, potassium, vitamin C and harbors the precursor for vitamin A (ß-carotene). As a staple crop, sweetpotato provides a great source of antioxidants and bioactive compounds within its highly variable skin and flesh colors for human consumption. Economically, the global sweetpotato market value reached 42.75 billion U.S. dollars in 2020 with China as its leading producer [2]. In addition, the United States ranks among the top ten producers generating over 30 million hundredweight in 2020 with an estimated value of $726.18 million with most of the production in the U.S. southern states [3]. Although a nutritious and profitable crop, sweetpotato production has been challenging to growers due to the multiple biotic stresses besetting its farming, especially the sweetpotato virus disease (SPVD) [4].

SPVD is one of the most devastating biotic stress constraints to impact sweetpotato production by up to 98% [5] [6] [7] [8]. It is derived from the co-infection of the aphid transmitted sweetpotato feathery mottle virus (SPFMV; Potyvirus) and the white fly-transmitted sweetpotato chlorotic stunt virus (SPCSV; Crinivirus) [7]. Both viruses are (+) sense-single-stranded RNA pathogens that can individually or collectively impact sweetpotato quality and yield in various parts of the world [5] [8] [9] [10]. Globally, sweetpotato viral diseases are becoming more rampant [11] among which, SPFMV has become the most widespread virus to infect sweetpotatoes in comparison to other diseases giving rise to significant yield losses [8] [12]. Furthermore, co-infection with SPCSV generates crop devastation up to 98%. [12]. As a vegetatively propagated crop, sweetpotato is at risk of acquiring systemic pathogens in subsequent transplant materials, therefore, reducing productivity and quality [13]. Currently, the use of virus-free tested

transplanting material is the only method of SPVD prevention. However, gaining access to these materials for subsistence farmers can be a challenge [4] [13]. Recent advancements made in studying the control of RNA viruses in many plant species have identified the eukaryotic translation initiation factor 4E family (eIF4E) as a new class of virus recessive resistance genes [14] [15].

Translation initiation is an important process responsible for regulating the synthesis of many proteins [16]. It is an intricate pathway that involves approximately 10 eukaryotic translation initiation factors (eIFs), mRNAs, rRNAs, and tRNAs that facilitate ribosome-mRNA binding [17] [18] [19]. Of these eIFs, eIF4E is one of the most important factors known for the initial phase of protein synthesis [20]. As a component of the eIF4F pre-initiation complex (along with eIFG and eIFA subunits), its key roles are to bind to the methyl-7-guanosine (m7G) cap on the 5' end of an mRNA sequence in eukaryotes, recruit additional translation initiation factors and deliver the mRNA to the 43S pre-initiation complex; therefore, triggering translation initiation complex assembly, which results in the initiation of protein synthesis [21] [22]. Naturally, at the molecular level, eIFs play a critical role in plant development and response to adverse stimuli [23] [24]. For example, physiologically, eIF4A, eIF3, eIFG, eIF4E, and eIF5A families have been reported to impact vegetative and reproductive growth; while eIF1A, eIF4, eIF5, and eIF2 have been linked with regulating responses to abiotic stress. In conjunction, eIF4G and eIF4E families have been reported to be hijacked by viral pathogens throughout various plant species during infections [24] [25].

In addition to their role in protein synthesis in whole organism homeostasis, eIF4E or its isoforms eIF(iso)4E, and the cap-binding protein (CBP) possess host susceptibility characteristics that make them favorable targets for plant viruses within the *Sobemovirus* genus, *Secoviridae*, and *Potyviridae* virus families, to seize for their replications [26] [27]. These viruses share (+) sense-single-stranded RNA genomes encoding a protein cap known as viral genome-linked protein (VPg), analogous to the m7G cap located at the 5' end of an mRNA sequence in eukaryotes [21] [26] [28]. Like eukaryotic mRNAs, the translation initiation of viral mRNA is dependent upon VPgs interaction with the host's eIF4E proteins or its isoforms for replication and survival [29] [30] [31]. Consequently, translation initiation factors, specifically eIF4E and eIF4G are now considered great genetic resources to improve crop productivity utilizing novel and emerging engineered technologies to mitigate viral impact [30]. For instance, roughly 200 viral resistant genes have been identified in plants and approximately half of them were shown to carry natural recessive resistance against plant viruses [32]. In recent years, an array of recessive viral-resistance genes were cloned and characterized within various plant species such as *Arabidopsis thaliana*, lettuce (*Lactuca sativa*), rice (*Oryza sativa*), mustard (*Brassica rapa*), Barley (*Hordeum-vulgare*) and cassava (*Manihot esculenta*) of which, the majority encode for eIF4E or its isoform eIF(iso)4E [14] [26] [33] [34] [35] [36] [37]. While other

major crops have one gene copy of each like the wild type *I. trifida*, in cassava, and a root crop like sweetpotato, the cloning of the eIF4E protein gene family revealed varying gene copy numbers, one *eIF*4*E*, two *eIF*(*iso*)4*E*, and two of the divergent *nCBP* [26]. In terms of polyploids, the cloning of the allopolyploid sugarcane (*Saccharum officinarum*) eIF4E family identified one single copy for both *CPB* and *eIF*(*iso*)4*E*, and two gene copies for *eIF*4*E* [27]. It has been reported that the numerous eIF4E encoding protein families in conjunction with their homologues, are critical in differentially regulating mRNA translation during normal plant development due to their functional redundancy. The redundancy of the eIF4E family in plant house-keeping functions over other translation initiation factors was exhibited during knockout and down regulation experiments of eIF4E and eIF(iso)4E, where *Arabidopsis* growth tolerated depletion in both alleles with slight symptoms of impairment, which did not affect the health of the host plant, except for phenotypic dwarfism during viral infection [35] [36]. Furthermore, similar results were obtained with individual nonsense mutations in *eIF*4*E* or *eIF*(*iso*)4*E* alleles of tobacco (*Nicotiana tabacum*) and pepper (*Capsicum annum*) showing no significant phenotypic impairment [18] [35] [37] [38] [39] [40]. However, plant species harboring double knockouts resulted in a semi-dwarf phenotype and polysome loading reduction, therefore displaying their role in cell growth regulation [18] [35] [36] [41] [42]. Viral resistance studies performed in tomato (*Solanum lycopersicum*), *Capsicum annuum*, muskmelon (*Cucumis melo*), *Ipomoea batatas*, and *Manihot esculenta* have been used to demonstrate how natural variations of *eIF*4*E* genes can give rise to recessive resistance to potyviruses without impeding on the endogenous function [5] [22] [43]-[49].

SPFMV is a member of the *potyviridae* family encoding a VPg protein within its genome [50]. Its synergistic interaction with SPCSV can cause excessive devastation to sweetpotato production. As a component of SPVD, it is seldom acknowledged that a single infection of SPFMV may decrease storage root production and quality in several cultivars resulting in 20% - 50% yield losses; while in others, symptoms remain indistinguishable [4] [5] [7] [51] [52]. A major concern for SPFMV infection is derived from its synergistic interaction with SPCSV. Several sweetpotato cultivars with periodical resistance to SPFMV exist, especially in commercial varieties (Beauregard, Jewel, and others), which exhibit mild or no symptoms after infection leading to little economic losses. However, resistance is eliminated under dual infection with other viruses [7] [9] [53] [54] [55] [56]. There is limited information on the molecular interactions between SPFMV and sweetpotato host factors or the natural resistance to SPVD [5] [13]. In 2002, Mwanga *et al.* [5] reported that sweetpotato resistance to SPCSV and SPFMV may be mediated by two separate recessive genes. Further research postulated that the VPg of SPFMV may interact with eIF4E or its isoforms for infection, similar to polypoid sugarcane [5] [13] [27] [54]. Therefore, characterizing and elucidating the separate recessive genes to-

ward establishing their interaction with SPFMV VPg and the mechanism(s) responsible for their regulation are now major goals of plant-virus interaction studies.

The hexaploid sweetpotato genome is comprised of many small genomes, while the origin of its polyploid nature (90 chromosomes) has largely been a controversy in the sweetpotato breeding/molecular biology community [1] [56] [57] [58] [59]. Due to its complex genome, it can be challenging to identify single nucleotide polymorphisms (SNPs) because of its duplicative nature and high heterozygosity also seen in other polypoid crops [1] [49] [56] [58] [60] [61] [62]. Hence, this study aims to isolate, clone, and characterize eIF4E protein genes (*eIF4E*, *eIFiso4E*, and *CBP*) initially from four hexaploid sweetpotato cultivars exhibiting susceptible/resistance and unknown phenotypic responses to SPFMV. Knowledge gained from this work will elucidate their potential roles in host susceptibility factors' interaction with VPg-SPFMV for viral replication and subsequently provide a framework for CRISPR-Cas9 engineering of functional resistance against sweetpotato potyvirus infection.

2. Materials and Methods

2.1. Plant Material and Growth Conditions

Sweetpotato cultivars [PI-318846 (D-3; NZ-196), Resisto (Res), Jewel (JWL), and Beauregard (BRG)] utilized in this work have different responses to SPFMV infection. Commercial varieties (JWL and BRG) are known to have periodic resistance to SPFMV due to the mild or absence of symptom development after infection in a growing season [7] [53] [54] [56]. While PI-318846 is unknown and Resisto may show susceptibility at times [5] [7] [15] [56]. In-vitro nodal cuttings from the above cultivars, acquired from the Regional Plant Genetic Resource Conservation Unit of USDA (Griffin, Georgia), were micro-propagated into plantlets on an in-house Multiplication Media (MM; Tuskegee University) based on a modified Murashige and Skoog (MS) media formulation [63] [64]. Briefly, the nodal cuttings were initiated in Magenta-GA-7 vessels (Magenta Corporation) containing sterilized MM media supplemented with 30 g/L sucrose and 5 mg/L gibberellic acid at pH 5.8, then incubated under a photoperiod of 16-h light and 8-h dark at 26˚C ± 2˚C [63] [64]. After eight weeks, leaf tissues from the micro-propagated plantlets were collected for total RNA extraction and subsequent analyses.

2.2. Total RNA Extraction, Reverse-Transcriptase-PCR Amplification of Full-Length *Ib*eIF4E, *Ib*eIF(iso)4E, and *Ib*CBP cDNA

Total RNA Extraction and FirstStrand cDNA Synthesis: Leaf total RNA was isolated from 300 mg of tissue with the Quick-RNA Plant Kit (Zymo Research (Irvine, CA) following the manufacturer's protocol. Total cDNA was synthesized utilizing Maxima H-Minus First-Strand cDNA Synthesis Kit (Thermo

Fisher Scientific, Waltham, MA), following the manufacturer's instructions with modifications. Each starting reaction contained about 5 µg of total RNA and 100 pmol of Oligo(dT)-18 primer in a volume of 20 µl. Due to its GC-rich RNA template, samples were incubated at 65˚C for 5 min and stored on ice where 5X buffer and maxima H-minus reverse transcriptase enzyme mix were added to reach a final volume of 30 µl. Total single-strand cDNA was synthesized with Oligo(dT)-18 primer from poly(A+) RNA for 40 mins at 50˚C followed by a termination step at 85˚C for 5 mins.

cDNA Generation of Full-length Specific cDNA: To amplify the full-length cDNA of *Ipomoea batatas* (*Ib*) *Ib*eIF4E, *Ib*eIF(iso)4E, and *Ib*CBP within each variety, gene-specific primers (Table 1) targeting their respective coding sequences, were designed based on the annotated sequences of *Ipomoea trifida* eIF4E superfamily CDS sequences (http://sweetpotato.plantbiology.msu.edu/) [59]. The eukaryotic translation initiation factors previously identified in *Arabidopsis thaliana* and *Manihot esculenta* (phytozome.jgi.doe.gov) was utilized to perform BLASTp analyses against the two-diploid wild-type relatives (*I. trifida* and *I. triloba*) as reference genomes for the hexaploid varieties. Primer pairs were designed with restriction enzymes (RE) flanking the 5' end of each primer sequence (Table 1). The selected restriction enzyme sites are listed with their name ID underlined/bolded in Table 1. To protect restriction enzymes from degradation, 3 - 5 miscellaneous nucleotide base pairs (bp) were incorporated at the end of each restriction enzyme site located upstream of each primer. Each PCR reaction was performed in a 50 µl reaction volume containing the following:

Table 1. Primers designed to amplify coding sequences of Sweetpotato eIF4E.

A. Primer ID	Sequences 5'-3'	eIF4E Locus
eIF4EFR-HindIII[a]	CGT**AAGCTT**ATGGTGGAAGAAATCGAGAAATCG	itf10g18510.t1
eIF4ERV-BamHI	GCT**GGATCC**TACTGTGTAACGATTCTTGGC	
eIF(iso)4EFR-SalI	GAT**GTCGAC**ATGGCAACCGAGACGGCAG	itf.04g08430.t1
eIF(iso)4ERV-Xbal	GCTCG**TCTAGA**CACACTATAACGGCCCTTAGCTG	
CBPFR-HindIII	GCT**AAGCTT**ATGGAAGAAGCGATAGCAGAG	itf.01g08030
CBPRV-Xbal	GAC**TCTAGA**GCCTCTCAACCAAGTGTTCCG	
B. Primers for Sequencing		
eIF4EFR	ATGGTGGAAGAAATCGAGAAATCG	
eIF4ERV	TACTGTGTAACGATTCTTGGC	
eIF(iso)4EFR	ATGGCAACCGAGACGGCAG	
eIF(iso)4ERV	CACACTATAACGGCCCTTAGCTG	
CBPFR	ATGGAAGAAGCGATAGCAGAG	
CBPRV	GCCTCTCAACCAAGTGTTCCG	

A: Primers utilized to amplify each specific cDNA. B: Primers utilized to sequence the amplicons in both directions.

approximately 300 ng of template cDNA, 1X iProof HiFi Master Mix (BIO-RAD, Hercules, CA), 5 µmol of forward (FR) and reverse (RV) primers. Each reaction was performed under the following thermocycler conditions: initial denaturation 98˚C for 3mins, then 35 cycles-denaturation at 98˚C for 1 min, annealing 60˚C (eif4E/eIF(iso)4E) or 57˚C (CBP) for 1 min, extension 72˚C for 2 mins, and a final extension of 72˚C for 5 mins. About 25 µl of amplified PCR products were resolved on a 0.8% 1X TAE agarose gel. The bands were eluted and purified with PureLink Quick Gel Extraction (Invitrogen, Waltham, MA), and stored at −80˚C for subsequent cloning and sequencing along with the PCR-amplified cDNA amplicons for direct-sequencing. The Sweetpotato ß-actin gene served as a loading control to generate a ~200 bp amplicon using the following: forward 5'-CCGGTATTGCGCATAGAATGAG-3' and reverse 5'-CCACGAGCATCTTTGGATCTT-3' primers. PCR conditions were similar to those utilized with CBP gene specific primers.

2.3. Sequencing Annotation and Phylogenetic Analyses

Direct-sequencing of PCR-amplified cDNA amplicons and the sequencing of the confirmed recombinant clone plasmid DNA were performed by GeneWiz (South Plainfield, NJ). The bioinformatics analyses of all the cloned cDNA sequences and the direct-sequencing PCR-amplified cDNA amplicon data were mined with the DNASTAR-LaserGene software (DNASTAR, Madison, WI). Since sweetpotato is hexaploid, the direct-sequencing of PCR-amplified cDNA amplicon resulted in heterogeneous chromatographic traces with varying multiple picks at certain nucleotide positions. The Mixed Sequence Reader (MSR) software was further utilized for multiple heterozygous base calling to help decipher and annotate the different homeoalles for each of the *IbeIF4E*, *IbeIF*(*iso*)*4E*, and *IbCBP* alleles [65]. Predicted protein sequences of *Ib*eIF4E gene family were examined simultaneously with protein sequences from thirteen plant species collected [*I. trifida*, *I. tribola*, japanese morning glory *I. nil* (XP_019169243.1, XP_019181316.1, XP_019173242.1), *A. thaliana* (NP_193538.1, NP_001332369.1, AAC.172220.1), pepper (*C. annuum*) (XP_016557968.1, NP_001311631.1, XP_016548322.1), cassava (*M. esculenta*) (XP_019169243.1, XP_021606525.1, XP_021620559.1), coffee (*C. canephora*) (CDP14577.1, CDO99285.1, CDP14308.1), rice (*O. sativa*) (NP_001359122.1, XP_015615056.1, XP_0156311510.1), soybean (*G. max*) (KAG4956073.1, NP_001356078.1, NP_001235427.1), pea (*P. sativum*) (AHV79423.1, ABH09880.1, AVH79423.1), tomato (*S. lycopersicum*) (NP_001307578.1, QPB75810.1, XP_004249299.1), potato (*S. tuberosum*) (QQP16443.1, QQP16447.1, XP_006351360.1), and wheat (*T. aestivum*) (M95818, Z12616, CD934979) from the NCBI Genbank database (http://www.ncbi.nlm.nih.gov) to access percentage sequence similarities. The Sequences were inputted into the MEGAX software for phylogenetic analysis using the maximum likelihood, and the cladogram was developed within the same program [66] [67] [68].

3. Results

3.1. Analysis of Sweetpotato *eIF4E* Family Transcripts

To amplify the full-length coding sequences (CDS) of the eIF4E family in *Ipomoea batatas* cultivars, a BLASTp query was conducted using *Arabidopsis thaliana* and *Manihot esculentaeIF4E* genes family (phytozome.jgi.doe.gov) against the sweetpotato genomics resource database (http://sweetpotato.uga.edu/index.shtml), which contains two highly in-bred wild-types diploid *Ipomoe*a species (*Ipomoea trifida* and *Ipomoea triloba*) as reference genomes [59] [69] [70]. The BLASTp screening results confirmed three candidate genes for each species that encode for the conserved eIF4E superfamily protein domains: Itf10g18510.t1(eIF4E), Itf01g08030.t1 (CBP), and Itf04g08430 (eIF(iso)4E) in *I. trifida*; and itb10g18850.t1(eIF4E) itb04g09410.t1 (eIF(iso)4E), and itb01g12380.t1(CBP) in *I. triloba*. Gene-specific primers utilized for the *Ipomea batata* cultivars tested resulted in amplicons of about 700 bp in full-length cDNA of *IbeIF4E*, *IbeIF*(*iso*)*4E*, and *IbCBP* (Figure 1), confirming their approximate theoretical sizes from the query search (Figure 1). Each putative *eIF4E* family gene was present in all four cultivars, which indicates that eIF4E family proteins are expressed in the leaf tissue of all four cultivars. However, the semi-quantitative analysis showed differential expression levels among the cultivars (Figure 1). The specific cDNA amplicons were directly sequenced or cloned and sequenced to screen for homeoalleles/haplotypes, as a result of sweetpotato hexaploid nature.

3.2. Heterozygous Base Calling of Direct-Sequencing Analyses of PCR-Amplified eIF4E cDNA Superfamily in Different Sweetpotato Cultivars

To sequence multiple allelic forms simultaneously, direct-sequencing of

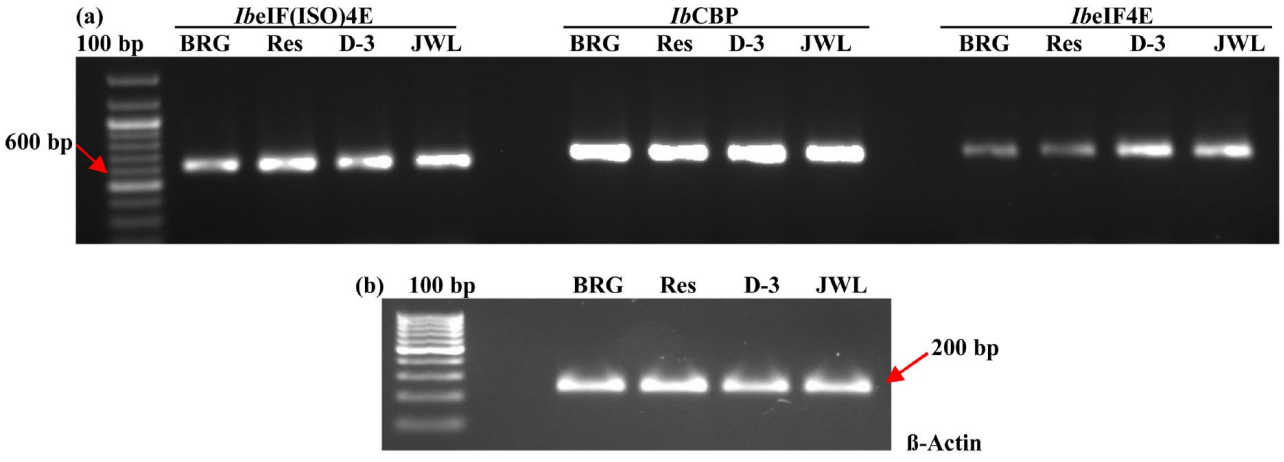


Figure 1. Resolution Patterns of the *IbeIF4E* Superfamily cDNAs in Different Cultivars. Panel (a): End point amplification of leaf tissue *IbeIF*(*iso*)*4E*, *IbCBP* and *IbeIF4E* 700 bp cDNAs from RT-PCR reactions with specific primers on sweetpotato cultivars with SPFMV tolerance/susceptible, and unknown. Panel (b): sweetpotato ß-actin gene, 200 bp amplicon served as a loading control. The Semi-Quantitative analysis showed differential expression levels among the cultivars BRG: Beauregard, Res: Resisto, D-3: PI-318466, JWL: Jewel.

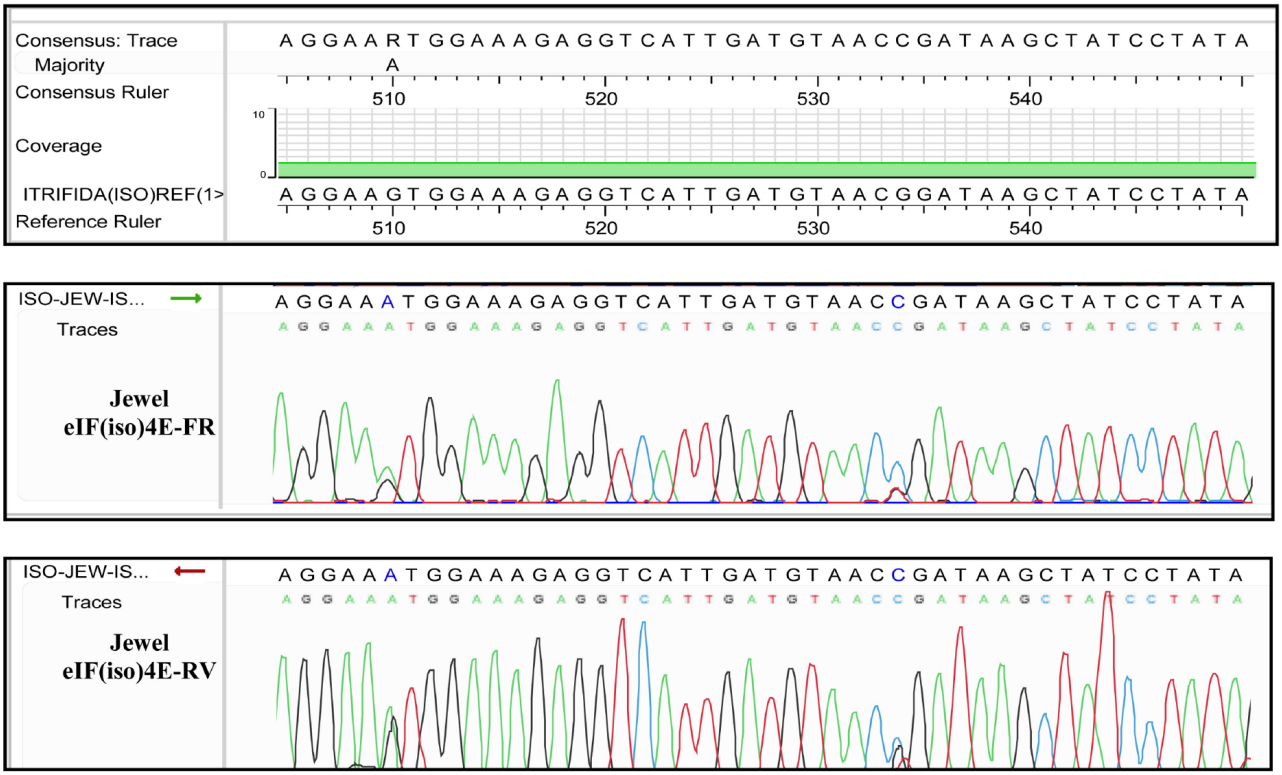
PCR-amplified cDNA amplicons versus sequencing of the cloned cDNAs were performed and mined for heterozygous base calling of the chromatogram traces with the Mixed Sequence Reader software and DNASTAR-LaserGene software to provide a collective mean of polymorphic variations within each targeted gene amongst each cultivar.

Analysis of PCR-amplified IbeIF4E cDNA direct-sequencing chromatograms. The *I. trifida* reference sequence, which contains only one copy of the *eIF4E* gene with the same allele at the itf10g18510.t1 locus served as a comparative reference. From the heterozygous base calling of direct-sequencing of *IbeIF4E*, Resisto results diverged by 10 SNPs from *I. trifida* (T/C at position 168, C/T at position 246, C/A at position 381, A/C at position 384, C/T at position 432, A/G at position 543, C/A at position 552, C/A at position 573, A/G at position 591, and C/T at position 672; where the first letter represents the first nucleotide polymorphism peak and the second letter represents the second nucleotide peak and the wild-type reference sequence). Double peaks with two fluorescent indicators were observed on the chromatogram indicating where the *eIF4E* alleles diverged within the heterozygous sample. These double peaks were represented in all SNP locations except at position 381, which contained a single peak indicating all allele copy numbers contain the same polymorphism. There was no evidence of these SNPs causing alterations to the predicted protein sequence in comparison to the *I. trifida* reference sequence. The North Carolina commercial variety, Jewel, displayed a total of 17 SNPs (C/G at position 87, A/C at position 120, A/G at position 140, T/C at position 141, T/C at position 144, C/T at position 161, C/T at position 168, C/T at position 246, C/A at position 381, A/C at position 384, G/A at position 420, C/T at position 432, A/G at position 543, A/C at position 552, C/A at position 573, A/G at position 591, T/C at position 649), with each SNP position containing double peaks. Meanwhile, the commercial variety Beauregard exhibited 9 SNPs (T/C at position 141, T/C at position 168, C/T at position 246, A/C at position 384, C/T at position 432, A/G at position 543, C/A at position 552, C/A at position 573, and C/T at position 673) and all SNP positions displayed double peaks. In the last analyses, PI-318846 showed 7 SNPs (T/C at position 168, C/T at position 246, C/A at position 381, A/C at position 384, C/T at position 432, A/G at position 543, and C/A at position 552), which All SNP's identified exhibited double peaks when analyzing the chromatogram.

Analysis of PCR-amplified IbeIF(iso)4E cDNA direct-sequencing chromatograms. Based on the single copy reference sequence of the *eIF(iso)4E* gene with the same allele found at locus itf04g08030.t1 of *I. trifida*, Resisto chromatogram mining results revealed a total of 6 SNPs (G/C at position 63, T/C at position 66, A/G at position 231, A/G at position 267, T/C at position 321, A/G at position 510), with position 63 containing a single SNP peak when analyzing the chromatogram. The remaining SNPs identified, contained overlapping peaks. Similarly, Jewel revealed 7 SNPs (T/G at position 60, T/C at position 66, C/T at

position 82, A/G at position 267, G/T at position 376, A/G at position 510, C/G/T at position 534) as shown in Figure 2. Within the trace files, we identified a position containing a triple peak in position 534 which displayed C, G and T nucleotides, which could represent multiple allelic forms. The remaining SNPs identified contained double peaks (Figure 3). Beauregard showed 5 SNPs within the putative *eIF*(*iso*)*4E* allele (G/C at position 63, C/T at position 114, A/G at position 267, A/G at position 510, C/G at position 534) which displayed mixed chromatograms within each SNP location. All mixed chromatogram positions contained a fluorescent signal identical to the reference sequence in conjunction with a single SNP. Lastly, accession PI-318846 was found to contain 7 SNPs (C/T at position 114, A/G at position 231, A/G at position 267, T/C at position 321, A/G at position 510, A/G at position 512 and C/T at position 524), which comprised of multiple fluorescent peaks.

Analysis of PCR-amplified IbCBP cDNA direct-sequencing chromatograms. A single copy of the *I. trifida CBP* gene sequence, with the same allele



(a)

(b)

Jewel eIF(iso)4E F1-A**T**GGAAAGAGGTCATTGATGTAAC**C**GAT
Jewel eIF(iso)4E F2-**G**TGGAAAGAGGTCATTGATGTAAC**G**GAT
Jewel eIF(iso)4E F3-**G**TGGAAAGAGGTCATTGATGTAAC**T**GAT
Jewel eIF(iso)4E F4-A**T**GGAAAGAGGTCATTGATGTAAC**T**GAT

Figure 2. Example of mixed chromatogram trace from direct-sequencing of PCR-Amplified cDNA. (a): The resulting mixed chromatogram was obtained through direct-sequencing of PCR-amplified *IbeIF*(*iso*)*4E* cDNA from Jewel leaf samples. In locations where allelic forms differed, multiple peaks were detected (red arrows). (b): The anticipated base compositions of allelic forms are provided below the chromatogram. SNP combinations are highlighted in red bold.

located on locus Itf.01g08030.t1 served as our reference comparator for the heterozygous base calling analyses. Sequences within each cultivar showed varying polymorphisms. In Resisto, we identified 6 SNPs (C/G at position 50, T/C at position 96, G/A at position 282, G/A at position 438, T/A at position 513, and C/T at position 595) that diverged from the reference sequence. While Jewel was shown to contain 5 SNPs (A/G at position 88, G/A at position 106, G/A at position 276, G/A at position 432, T/C at position 624) that possess mixed peaks within the SNP location. Furthermore, PI-318846 displayed a total of 6 SNPs (A/G at position 88, T/C at position 90, G/A at position 106, G/A at position 432, C/T at position 508, and T/C at position 624) that differed from the reference sequence. Finally, in Beauregard, a total of 5 SNPs (A/G at position 88, T/C at position 90, G/A at position 106, G/A at position 432, and T/C at position 624) were detected.

To confirm SNP variants within the cultivars (**Table 2**), trace files were analyzed with the mixed sequence reader program for heterozygous base calling [65]. Sequencing analyses indicated that the amplified cDNA fragments were exactly 696 nucleotide bp (putative *IbeIF4E*), 606 bp (putative *IbeIF*(*iso*)*4E*), and 675 bp (putative *IbCBP*) in length. To decipher the putative gene structures of the eIF4E subfamily from the direct sequencing results, we compared *I. trifida* reference genomic sequences to each deduced chromatogram sequence of *IbeIF4E*, *IbeIF*(iso)*4E*, and *IbCBP* cDNA. The results showed that both *IbeIF4E*, and *IbeIF*(iso)*4E* each contained five exons and four introns, while *IbCBP* was comprised of six exons and five introns. The deduced amino acid sequences from the direct-sequencing of cDNA amplicons were used to screen for conserved domain regions resulting in the discovery of the IF4E superfamily domains within the sequences representing putative *IbeIF4E*, *IbeIF*(*iso*)*4E* and *IbCBP*, in the following amino acid locations: 57 - 212, 27 - 185, and 46 - 204 similar to Marchler-Bauer *et al.* [71] findings with different species. The genes, *IbeIF4E*, *IbeIF*(*iso*)*4E*, and *IbCBP*, each encoded polypeptides of 232, 202 and 225 amino acid residues with calculated molecular masses of 26kDa, 22.8kDa, and 25.8 kDa with predicted isoelectric points of 5.1, 5.57, and 6.66, respectively. Since data from the direct-sequencing of cDNA exhibited multiple peaks in the chromatogram that could potentially account for homeoalleles or SNPs occurring within the subgenome of the cultivars, similar to previous work [26] [62];

Table 2. Total SNPs identified within the coding regions of IbeIF4E, IbeIF(iso)4E and IbCBP in Sweetpotato cultivars based on direct-sequencing of PCR-Amplified cDNA chromatogram mining.

Cultivars	*IbeIF4E*	*IbeIF*(*iso*)*4E*	*IbCBP*	Total SNPs
PI-318846 (D-3) NZ-196[a]	7	7	4	18
Beauregard (BRG)	9	5	7	21
Jewel (JWL)	17	7	5	29
Resisto (RES)	10	6	6	22

therefore, we sought to further assess and confirm the presence of these SNP variants. Thus, the PCR-amplified cDNAs from all cultivars were cloned and sequenced for comparative bioinformatic analyses.

3.3. Mining of Cloned *Ibe*IF4E Superfamily Sequence Variations within Different Sweetpotato Cultivars

Although direct-sequencing of PCR-amplified cDNA tools provided rapid results as an alternative to utilizing traditional gene cloning, one major constraint with direct PCR product sequencing results in polyploid organisms is the presence of mixed chromatograms containing multiple fluorescent signals in positions where allelic copy numbers may diverge [72]. In solving this daunting challenge of the sequencing read interpretations, we sought to capture individual allelic forms by cloning the cDNA PCR products, sequencing the purified recombinant plasmid DNAs from several positive colonies, and separately characterizing and annotating them to overcome this impediment. In the past, it has been postulated that the closest existing relatives of the commercial sweetpotato are the *I. trifida* and *I. triloba* species [57] [58] [59] [73]. Therefore, for this work, the sequences of the clones were compared against both *I. trifida* and *I. triloba* reference genomes. Sequencing data from this analysis has been submitted to NCBI with the accession numbers OP273667-OP273690.

Haplotype Identification within Cloned IbeIF4E Sequences. The analysis of cloned cDNA sequencing results from the putative *IbeIF4E* gene in Resisto identified four haplotypes with a total of seven SNP locations compared to the reference sequence *I. trifida* (**Table 3** and **Figure 3**). However, when the nucleotide sequences were translated to their predicted proteins, identical amino acid sequences resulting in synonymous mutations (**Figure 4**) were found in all the haplotypes. Following comparative sequence analysis with *I. trifida*, a pairwise sequence alignment was performed with predicted amino acid sequences obtained from the putative eIF4E in Resisto with *I. triloba*, which showed a 98.7%

Table 3. Example of SNP site distribution of mining cloned forms of ibeIF4E coding sequence in Resisto.

Position	246	381	543	552	573	591	672
I. trifida	T	A	G	A	A	G	T
I. triloba	T	A	G	C	A	G	C
Form 1	C	C	A	C	C	A	C
Form 2	C	C	G	A	A	G	T
Form 3	C	C	G	A	A	G	T
Form 4	T	C	A	C	C	A	C
Genotype Combinations	C/T	C	G/A	A/C	A/C	G/A	T/C

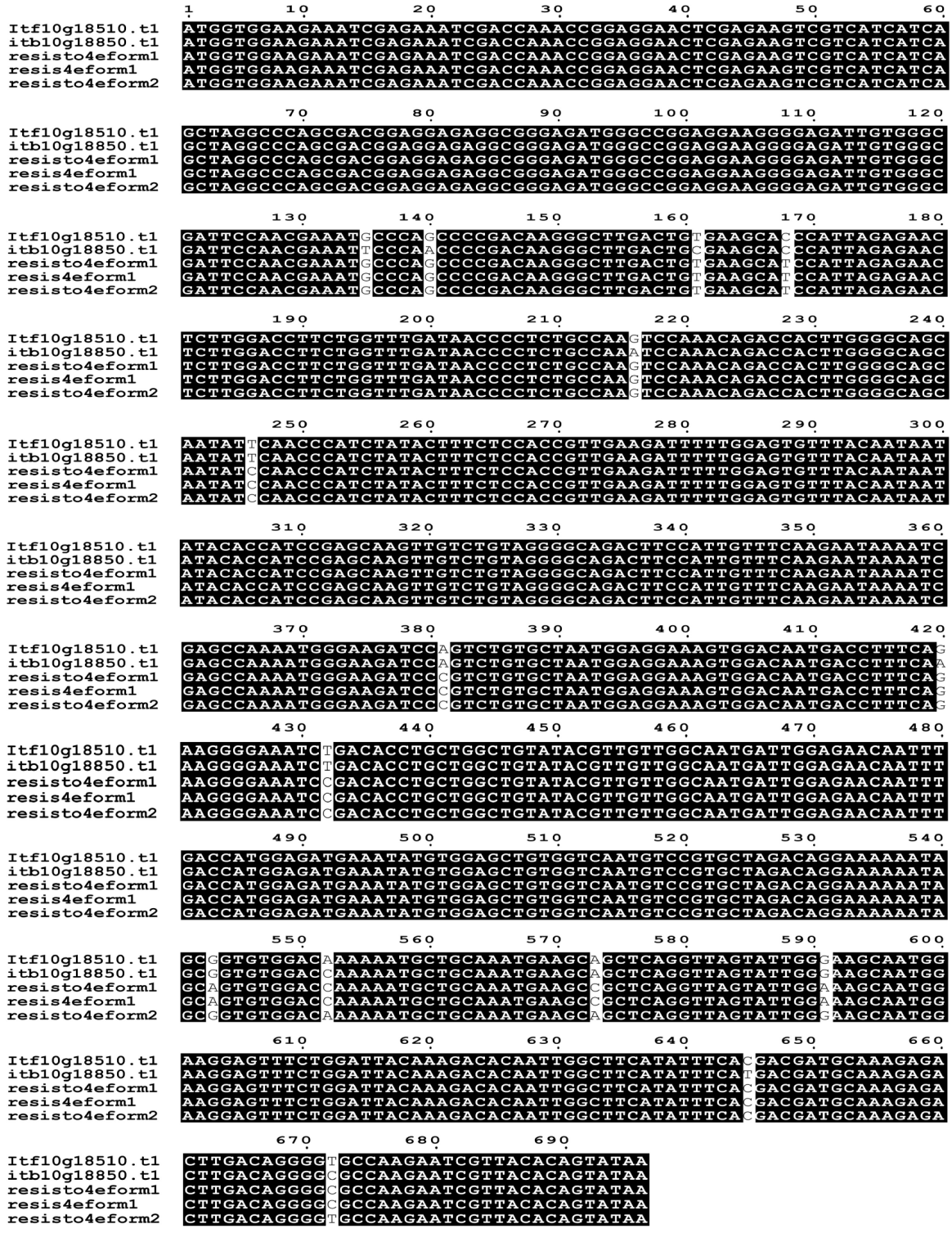


Figure 3. Example of Resisto *IbeIF4E* Allelic Forms Nucleotide Multiple Sequence Alignment. Multiple sequence alignment of selected Resisto *IbeIF4E* allelic forms displaying C/T (246), C/A (381), C/T(432), A/G (543), C/A (552), C/A (573), A/G (591), C/T (672) polymorphisms. The first letter indicates the SNP nucleotide in the full length sweetpotato *IbeIF4E* haplotype sequence, the number gives the position of the nucleotide, and the second letter indicates the nucleotide sequence in the wild-type sequence. Highly conserved regions between allelic forms are shaded in black. Single nucleotide polymorphisms are denoted with a white background. Nucleotide positions are numbered on the above sequences. Note: Multi-sequence alignment also displays SNPs in comparison to *I. triloba*.

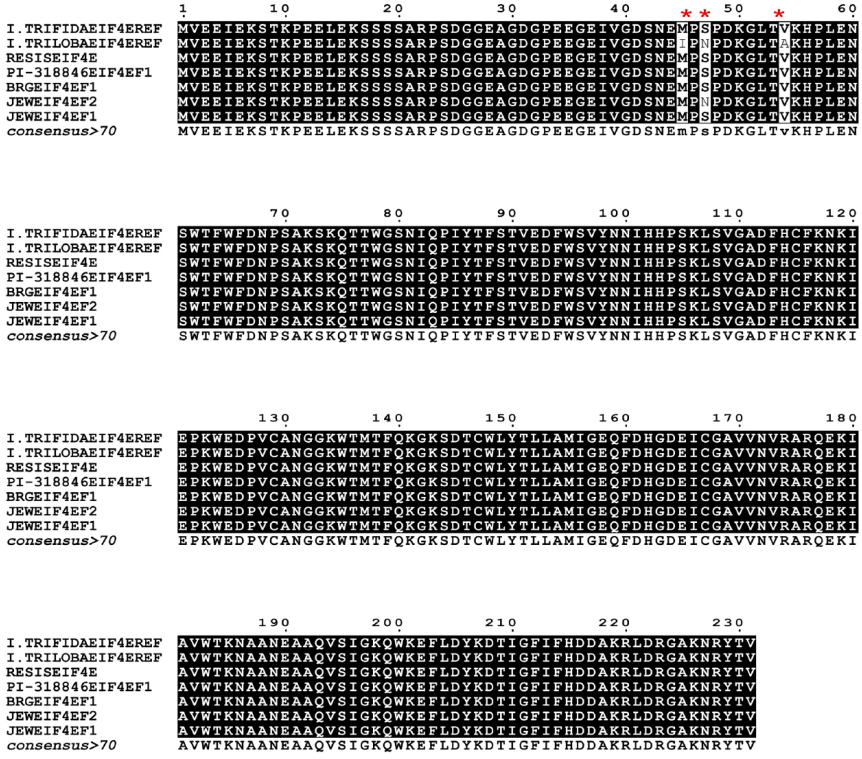


Figure 4. Amino Acid Sequence Alignment of Predicted *Ib*eIF4E Sequences. All the *Ib*eIF4E forms from sweetpotato cultivars Resisto, PI-318846, Jewel and Beauregard were compared with the reference ancestorial lines *I. trifida* (NSP306) and *I. triloba* (NSP323). Red asterisks indicate the location of amino acid variations observed between multiple allelic forms and amongst reference sequences. Identical amino acid sequences are shaded in black. Amino acid sequence positions are numbered above each sequence.

amino acid sequence identity with three of the amino acid variations in locations: M45I, S47N, and V54A (the first letter indicates the residue in the cloned sequence, the number indicates the amino acid position and the second letter indicates the residue present in the diploid reference *I. triloba*) (**Figure 4**).

Of the total colonies screened from PI-318846, three were categorized as haplotype one which displayed identical nucleotide sequences to the reference *I. trifida*. The remaining colonies were identical to each other and were labeled as a second haplotype, with a nucleotide sequence identity, following pairwise analysis, of 99.4% to *I. trifida* (T/C at nt 168), and a 98.7% sequence identity to *I. triloba* (G/T at nt 135, G/A at nt 140, T/C at nt 161, G/A at nt 216, A/C at nt 384, G/A at nt 420, A/C at nt 552, C/T at nt 645, T/C at nt 672). Both haplotype sequences were translated to their predicted amino acid sequences which showed 100% identity to *I. trifida*. However, in comparison to *I. triloba*, there was a 98.7% identity between the amino acid sequences. Polymorphisms in locations 135 and 140 and 161 resulting in amino acid variations M45I, S47N and V54A respectively (**Figure 4**).

All colonies representing Beauregard *Ib*eIF4E sequences showed a 99.4% nucleotide sequence identity to *I. trifida* with four SNPs (T/C at nt 168, C/T at nt

246, C/A at nt 381, and A/C at nt 384), and a 98.3% sequence identity to *I. triloba* (G/T at nt 135; G/A at nt 140; T/C at nt 161; T/C at nt 168; A/G at nt 216; T/C at nt 246; A/C at nt 381; C/A at nt 384; G/A at nt 420; A/C at nt 552; C/T at nt 645 and T/C at nt 672). Translated nucleotide sequences revealed a 100% amino acid sequence identity to *I. trifida* and a 98.7% amino acid sequence homology to *I. triloba*.

We identified two haplotypes from the Jewel *Ib*eIF4E colonies analyzed. Haplotype one exhibited a 99.4% nucleotide sequence identity to *I. trifida* with four SNPs (T/C at nt 168; C/T at nt 246; C/A at nt 381; A/C at nt 384), and a 98.3% nucleotide sequence identity to *I. triloba*, however with 12 SNPs (G/T at nt 135; G/A at nt 140; T/C at nt 161; T/C at nt 168; G/A at nt 216; C/T at nt 246; C/A at nt 381; A/C at nt 384; G/A at nt 420; A/C at nt 552; C/T at nt 645 and T/C at nt 672). Though several SNPs were detected, the translated nucleotide sequences of Jewel *Ib*eIF4E haplotype one displayed a 100% amino acid sequence identity to *I. trifida* and a 99.6% amino acid identity to *I. triloba*. Polymorphisms in nucleotide positions 135, 140 and 161 resulted in amino acid substitutions of M45I, S47N and V54A in the predicted protein sequence, respectively (**Figure 4**).

Haplotype Identification within Cloned IbeIF(iso)4E Sequences. Analysis of sequence variations of Resisto *IbeIF*(*iso*)4*E* clones revealed three different haplotypes. Haplotype one showed a 99.5% nucleotide sequence identity to both *I. trifida*, which harbored three SNPs (G/C at nt 63, T/C at nt 66, and C/G at nt 534), and *I. triloba* with also three SNPs (T/C at nt 66, T/C at nt 171 and G/A at nt 589). Whereas haplotypes two & three showed a 99.2% nucleotide sequence identity to both ancestorial lines *I. trifida and I. triloba*. Specific to *I. trifida*, haplotype two and three displayed a total of five SNPs (haplotype two: G/C and nt 63, T/C at nt 66, A/G at nt 231, A/G at nt 510, C/G at nt 534; haplotype three: G/C at nt 63, T/C at nt 66, T/C at nt 321, A/G at nt 510 and C/G at nt 534). While the comparison of haplotype two and three to *I. triloba* revealed 5 SNPs each, with haplotype 2 consisting of T/C at nt 17, T/C at nt 66, A/G at nt 231, A/G at nt 510, and G/A at nt 589, and haplotype 3 comprised of T/C at nt 66, T/C at nt 171, T/C at nt 321, A/G at nt 510, and G/A at nt 589 (**Figure 5**). Although we identified three haplotypes within Resisto, the predicted amino acid sequences of each haplotype revealed synonymous mutations, that resulted in 100% amino acid sequence identity to both *I. trifida* and *I. triloba* wild-type relatives (**Figure 6**).

Sequencing results of cloned *IbeIF*(*iso*)4*E* from PI-318846 revealed three different allelic forms that resulted in two haplotypes possessing synonymous mutations and one allelic form displaying a non-synonymous mutation. Haplotype one exhibited a 99.2% nucleotide sequence identity (**Figure 4**) to both *I. trifida* (haplotype one: G/C at nt 63; A/G at nt 267; A/G at nt 510; C/G at nt 534 and a A/G at nt 535), and *I. triloba* (haplotype one: T/C at nt 171; A/G at nt 267; A/G at nt 510; A/G at nt 535 and a G/A at nt 589). The analysis from haplotype two revealed a 99.5% nucleotide sequence identity to both wild types *I. trifida*

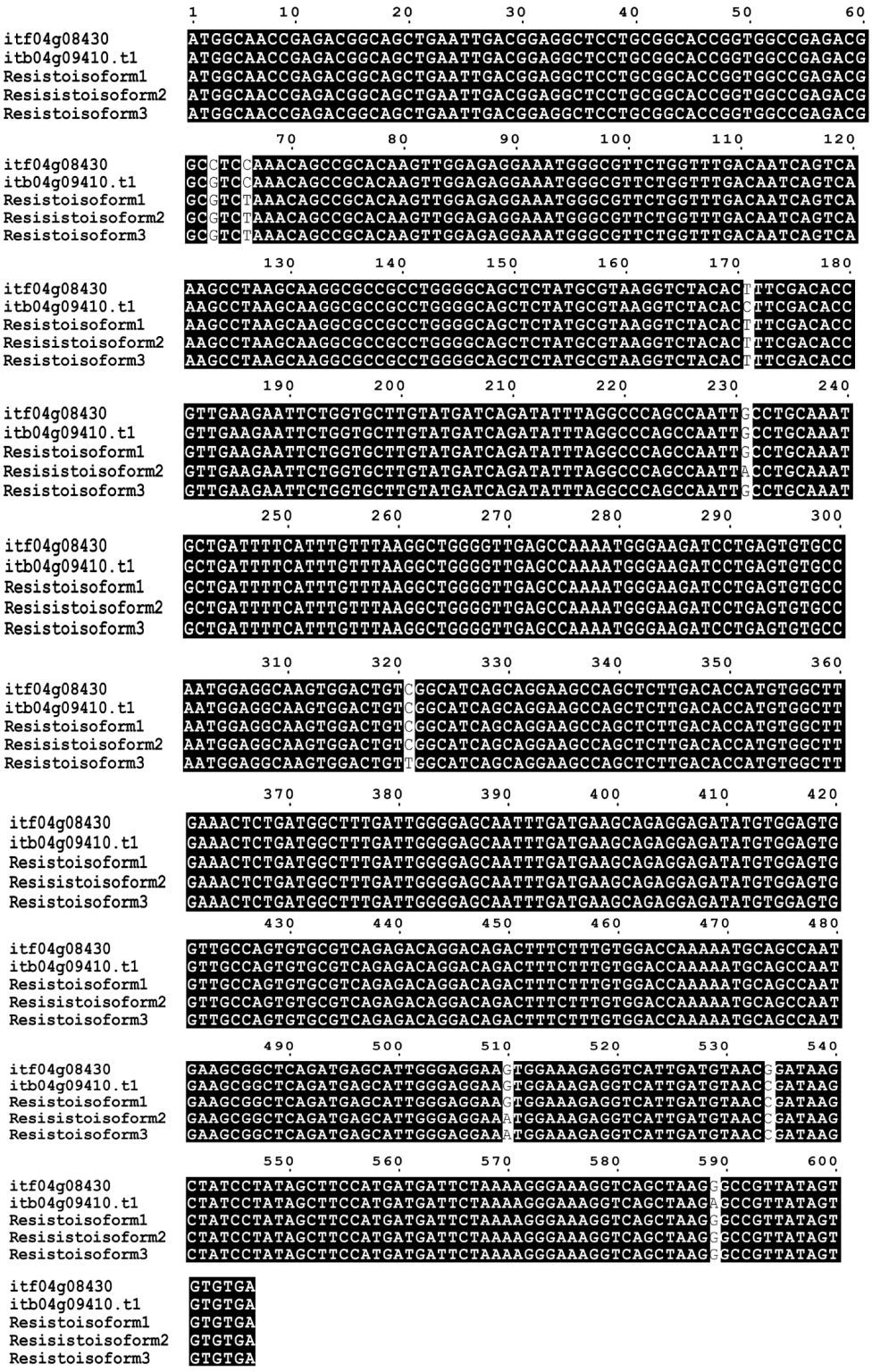


Figure 5. Example of Resisto *IbeIF*(*iso*)4*E* Allelic Forms Nucleotide Multiple Sequence Alignment. Multiple sequence alignment of selected Resisto *Ib*eIF(iso)4E allelic forms are displaying (G/C (63), T/C (66), A/G (231), T/C (321), A/G (510), C/G (534)) polymorphisms in Resisto allelic forms (the first letter indicates the SNP nucleotide in the full length sweetpotato *ibeIF4E* haplotype sequence, the number gives the position of the nucleotide, and the second letter indicates the nucleotide sequence in the wild-type sequence). Highly conserved regions between allelic forms are shaded in black. Single nucleotide polymorphisms are denoted with a white background. Nucleotides are numbered above sequences.

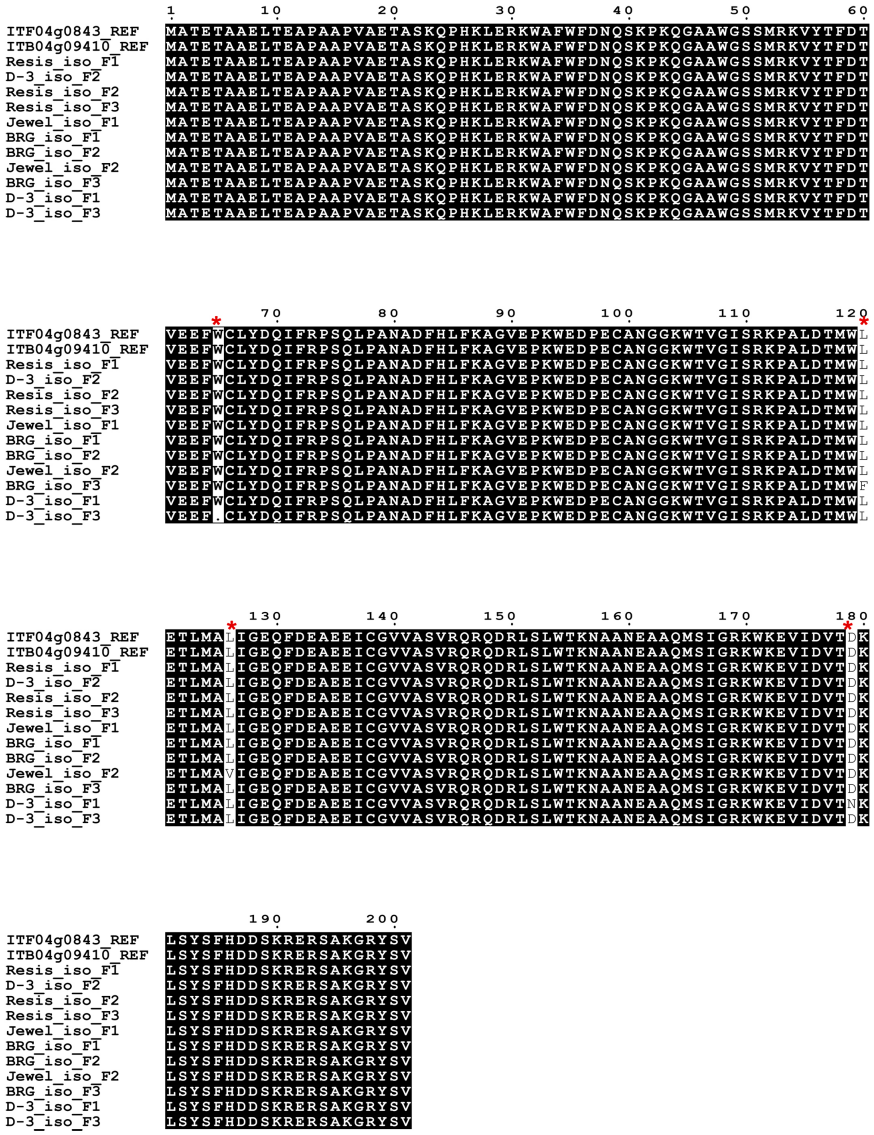


Figure 6. Deduced amino acid sequence alignment of predicted *IbeIF*(*iso*)4*E* Sequences: Alignment of predicted *IbeIF*(*iso*)4*E* amino acids sequences from cloned sweetpotato cultivars Resisto, PI-318846, Jewel and Beauregard with Reference ancestorial Lines *I. trifida* (NSP306) and *I. triloba* (NSP323). Red asterisks indicate amino acid variations observed between multiple allelic forms and amongst reference sequences. Conserved amino acid sequences are shaded in black. Numbers for amino acid residue positions are indicated above the sequences.

(haplotype two: G/C at nt 63; A/G at nt 510; and C/G at nt 534) and *I. triloba* (haplotype two: T/C at nt 171; A/G at nt 510; and G/A at nt 589). The third haplotype identified within PI-318846 revealed a 98.5% nucleotide sequence identity to both diploid wild types *I. trifida* (haplotype three: T/G at nt 42; T/G at nt 60; and G/C at nt 63; C/T at nt 82; nt deletions in regions 193 - 195; A/G at nt 510 and a C/G at nt 534) and *I. triloba* (haplotype three: T/G at nt 42; T/G at nt 60; C/T at nt 82; T/C at nt 170; nt deletions in regions 195 - 197; A/G at nt 510 and a G/A at nt 589). The predicted amino acid sequence of haplotype three contained

a missing codon at nucleotide positions 195 - 197, resulting in an in-frame deletion in exon 1 (**Figure 6**). The in-frame deletion mutation appears to encode a tryptophan (W) residue. In other species, it has been characterized that all known *eIF*4*E* genes consist of 8 conserved tryptophan residues, some of which are involved in mRNA cap binding [74] [75]. Since PI-318846 possesses 3 allelic forms of *IbeIF*(*iso*)4*E* and only one allelic form contained an in-frame mutation that resulted in a missing tryptophan, it suggests that the two forms that do not possess the mutation must dominate the function of *IbeIF*(*iso*)4*E*.

Three haplotypes were identified in Beauregard following the cloning of the putative *IbeIF*(*iso*)4*E* in relationship to both *I. trifida* and *I. triloba*. Compared to *I. trifida*, BRG *Ib*eIF(iso)4E haplotype one and three displayed a 99.2% nucleotide sequence identity, however, they exhibited SNPs at different positions (haplotype one: G/C at nt 63; A/G at nt 267; A/G at nt 510; C/G at nt 534, A/T at nt 567; haplotype three: G/C nt 63; T/C nt 321; T/C nt 171, A/G at nt 267, A/G at nt 510, A/T at nt 567, G/A at nt 589). Haplotype two showed a 99.2% nucleotide sequence identity (C/G at nt 63, C/T at nt 114, T/C at nt 171, A/G at nt 510, and G/A at nt 589). The deduced amino acid sequences of haplotype one and two have a 100% amino acid sequence identity to *I. trifida* and *I. triloba*, inferring that the different SNPs detected resulted in synonymous mutations. On the other hand, haplotype three displayed a 99.5% amino acid sequence identity. However, the SNP resulted in a F120L in comparison to both parental lines (**Figure 4**).

We identified two forms of the cloned *IbeIF*(*iso*)4*E* gene in Jewel in comparison to the wild-type reference sequences. Haplotype one had a 99.2% nucleotide sequence identity to *I. trifida* (T/G at nt 42; G/C at nt 63; C/T at nt 82; A/G at nt 510; C/G at nt 534); resulting in a 100% amino acid sequence identity. The comparison against *I. triloba* showed a 99.2% nucleotide sequence identity (T/G at nt 42; C/T at nt 82; T/C at nt 171; A/G at nt 510; G/A at nt 589); however, it harbored two SNPs in separate regions of the nucleotide sequence in contrast to the pairwise alignment with *I. trifida*. Although we were able to identify SNPs within haplotype one, the SNPs identified did not alter the amino acid sequence; therefore, form one possessed a 100% amino acid sequence identity to both *I. trifida* and *I. triloba*. Haplotype two results revealed a 99.2% nucleotide sequence identity to both reference genes: *I. trifida* (T/G at nt 60; G/C at nt 63; G/T at nt 376; A/G at nt 510; C/G at nt 534), and *I. triloba* (T/G at nt 60; T/C at nt 171; G/T at nt 376; A/G at nt 510; G/A at nt 589). In the predicted amino acid sequence, we observed that the SNP at nucleotide 376 resulted in a V126L amino acid change, which are both non-polar amino acids. This mutation led to a 99.5% amino acid sequence identity (**Figure 4**).

Haplotype Identification within Cloned IbCBP Sequences. The analysis of cloned sequences representing the cap binding protein in Resisto revealed 2 haplotypes, in contrast to the reference sequences. Haplotype one showed a 99.6% nucleotide sequence identity to *I. trifida* with 3 SNPs (C/G at nt 44; T/C at nt 90; G/A at nt 276) (**Figure 7**), while its amino acid sequence identity was

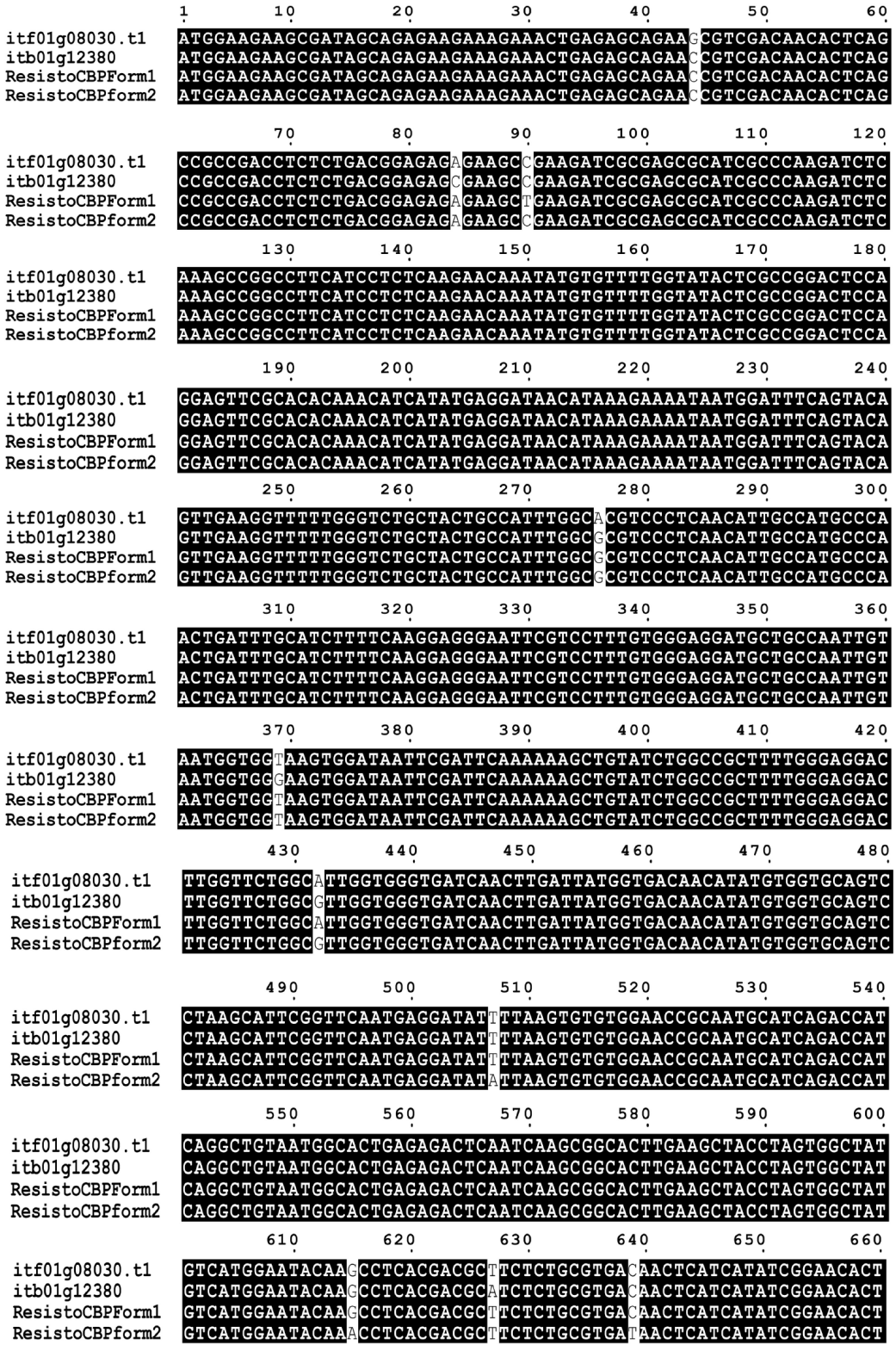


Figure 7. Example of resisto *Ib*CBP allelic forms nucleotide multiple sequence alignment. Multiple sequence alignment of selected Resisto *Ib*eIF(iso)4E allelic forms displaying C/G (44), C/A (84), T/C (90) G/A, G/A (276), G/A (432), A/T(507), A/G (615), T/A (627), T/C (639) polymorphisms in Resisto allelic forms compared to both *I. trifida* and *I. triloba* reference genomes. Highly conserved regions between allelic forms are shaded in black. Single nucleotide polymorphisms are denoted with a white background. Nucleotides are numbered above rows.

99.6% due to the nucleotide substitution in position 44 leading to a T15S amino acid change. In contrast to *I. triloba*, haplotype one displayed a 99.3% nucleotide identity (A/C at nt 84; T/C at nt 90; T/G at nt 369; A/G at nt 432; T/A at nt 627) leading to a synonymous mutation due to the nucleotide substitution in position 84 that altered the amino acid sequence to R28S causing a 99.6% amino acid sequence identity (Figure 8). Haplotype two analyzed in Resisto revealed a 99.1% nucleotide sequence identity in comparison to both reference species. The results against *I. trifida* showed six SNPs (C/G at nt 44; G/A at nt 276; G/A at nt

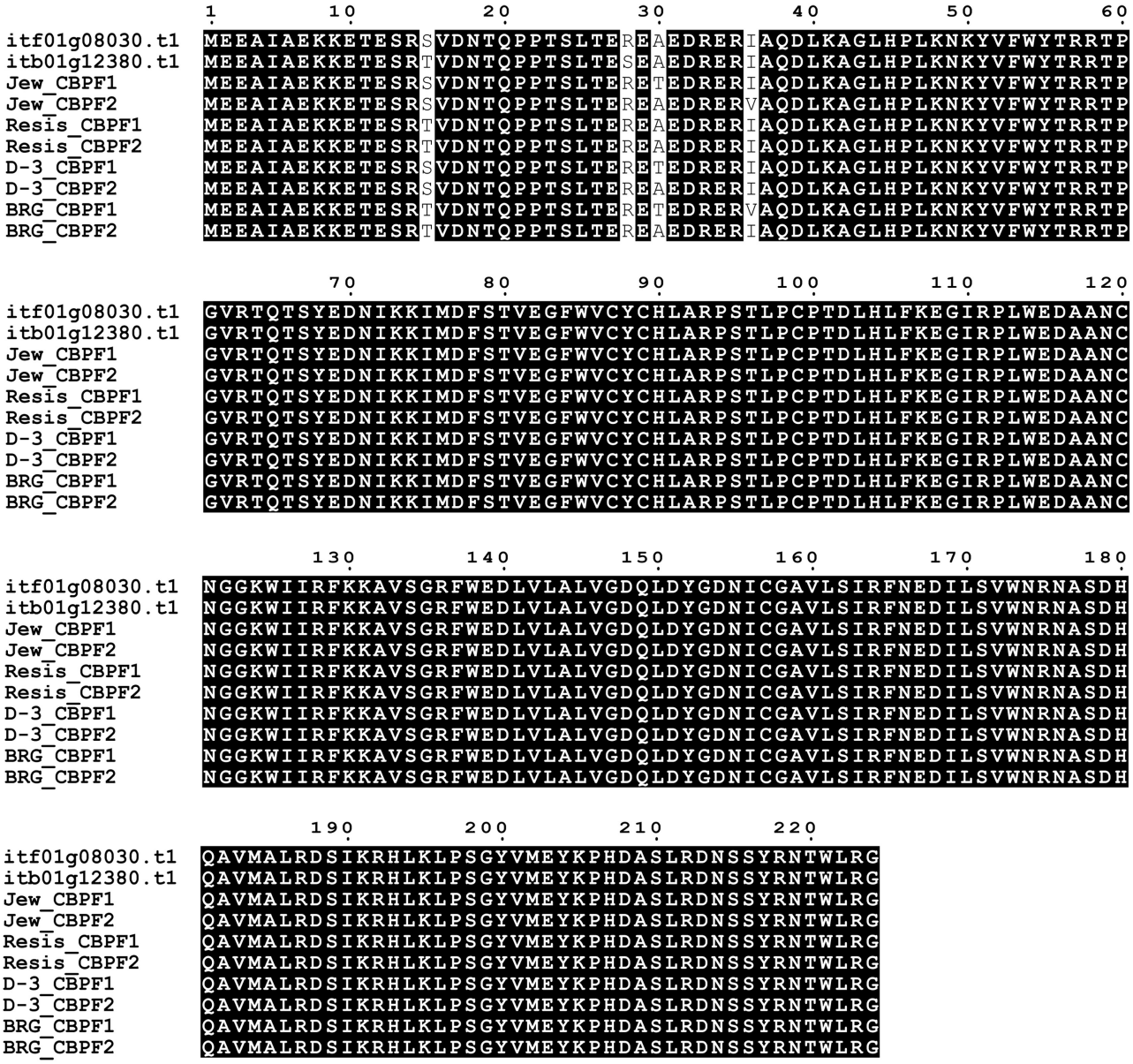


Figure 8. Deduced amino acid sequence alignment of predicted *IbCBP* Sequences. Amino acid sequence alignment of Predicted *Ib*CBP amino acid sequences from the cloned haplotypes compared to the reference ancestorial lines *I. trifida* (NSP306) and *I. triloba* (NSP323). Sweetpotato Cultivars Resisto (Resis), PI-318846 (D-3), Jewel (Jew) and Beauregard (BRG). Red asterisks indicate amino acid variations observed between multiple allelic forms and amongst reference sequences. Conserved amino acid sequences are shaded in black.

430; A/T at nt 507; A/G at nt 615; T/C at nt 639) resulting in the same synonymous mutation as haplotype one, where nucleotide substitution at position 44 changed the amino acid sequence to T15S, which is identical to the amino acid located in the same position in *I. triloba*. The alignment of haplotype two against *I. triloba* identified a 99.1% nucleotide sequence identity (A/C at nt 84; T/G at nt 369; A/T at nt 507; A/G at nt 615; T/A at nt 627; T/C at nt 639) with the same synonymous mutation as haplotype one in contrast to *I. triloba* showing a 99.6% amino acid sequence identity altering R28S, which is also identical to the amino acid located in the same position in *I. trifida*.

In cv. PI-318846, two allelic forms of the cloned *IbCBP* gene were identified in comparison to *I. trifida*. The results from haplotype one revealed a 99.6% nucleotide sequence identity (T/C at nt 90; G/A at nt 276; G/A at nt 432) leading to a 99.6 % amino acid sequence identity. Nucleotide substitution of position 90 resulted in an amino acid change in T30A. Whereas haplotype one in contrast with *I. triloba* shows a 99.3% nucleotide similarity (G/C at nt 44; A/C at nt 84; T/C at nt 90; T/G at nt 369; T/A at nt 627) with a 98.7% amino acid sequence identity. The amino acid changes occurred in the following positions (S15T, R28S, T30A) as a result of the nucleotide substitutions in positions 44, 84, and 90 (**Figure 8**). Sequencing results from haplotype two revealed a 99.1% nucleotide sequence identity with *I. trifida* (A/G at nt 88; T/C at nt 90; G/A at nt 106; G/A at nt 276; C/T at nt 507; T/C at nt 624); which generated a 100% amino acid sequence identity following the translation of sequencing results. Haplotype two sequencing data compared to *I. triloba* revealed a 98.5% nucleotide sequence identity (G/C at nt 44; A/C at nt 84; A/G at nt 88; T/C at nt 90; G/A at nt 106; T/G at nt 369; A/G at nt 432; C/T at nt 507; T/C at nt 624; T/A at nt 627) resulting in a 99.6% amino acid sequence identity displaying amino acid changes in S15T, R28S that also derived from the previously mentioned nucleotide substitutions in positions 44 and 84 (**Figure 8**).

In total, we identified two allelic forms of the *IbCBP* gene in cultivar Beauregard when assessing cloned sequences. In the first haplotype, we performed a pairwise alignment against *I. trifida* revealing a 99% nucleotide sequence identity, showing 7 SNPs (C/G at nt 44; A/G at nt 88; T/C at nt 90; G/A at nt 106; G/A at nt 276; G/A at nt 432; T/C at nt 624). The results from the translated nucleotide sequence revealed a 98.7% amino acid sequence identity showing amino acid changes in T15S, T30A and V36I as a result of nucleotide substitutions in positions 44, 90 and 106 (**Figure 8**). Pairwise alignment results assessed from haplotype one and *I. triloba* revealed the same identity percentage as *I. trifida* displaying a total of 7 SNPs (A/C at nt 84; A/G at nt 88; T/C at nt 90; G/A at nt 106; T/G at nt 369; T/C at nt 624 and T/A at nt 627). However, we were able to identify a different base substitution in position 84 that gave rise to the R28S amino acid change in conjunction with T30A and V36I amino acid changes which exhibited a 98.7% amino acid sequence identity. We then assessed variations in haplotype two, against *I. trifida* and it was revealed that there is a 99.7% nucleo-

tide sequence identity; displaying three SNPs (C/G at nt 44 and G/A at nt 276) that translated into a 99.6% amino acid sequence identity showing a T15S mutation, that was derived from a nucleotide substitution in nucleotide position 44. Next, the pairwise alignment with haplotype 2 and *I. triloba* revealed a 99.4% nucleotide sequence identity (A/C at nt 84; T/G at nt 369; A/G at nt 432 and T/A at nt 627) resulting in a 99.6% amino acid sequence identity that is derived from a nucleotide substitution in position 84 that resulted in a R28S amino acid change (**Figure 8**).

Jewel also displayed two allelic forms of the cloned *IbCBP* gene with a 99.7% nucleotide sequence identity to *I. trifida* (A/G at nt 88; G/A at nt 276). The deduced amino acid sequence revealed a 99.6% amino acid sequence identity with a T30A amino acid change based on the nucleotide substitution in position 88 (**Figure 8**). In comparison to *I. triloba*, haplotype one displayed a 99.1% nucleotide sequence identity (G/C at nt 44; A/C at nt 84; A/G at nt 88; T/G at nt 369; A/G at nt 432; T/A at nt 627). The translated sequence revealed a 98.7% amino acid sequence identity with the following mutations: S15T, R28S, T30A (**Figure 8**). Haplotype two results compared with *I. trifida* showed a 99.1% nucleotide sequence identity (T/C at nt 90; G/A at nt 106; G/A at nt 201; G/A at nt 276; G/A at nt 432; T/C at nt 624) that revealed a single amino acid change V36I from the SNP located in position 106 with a 99.6% amino acid identity. Results inferred from *I. triloba* revealed a 98.8% nucleotide sequence identity (G/C at nt 44; A/C at nt 84; T/C at nt 90; G/A at nt 106; G/A at nt 201; T/G at nt 369; T/C at nt 624; T/A at nt 627). Translated sequences also showed a 98.7% amino acid sequence identity that resulted in the following amino acid changes S15T, R28S, V36I (**Figure 8**).

4. Phylogenetic Analysis of Sweetpotato eIF4E Subclasses

Evolutionary Analysis of Sweetpotato IbeIF4E. To characterize putative *IbeIF4E* genes, a phylogenetic analysis was performed utilizing MUSCLE within the MEGA7 platform to investigate evolutionary relationships among eIF4E orthologues previously identified in other plant species (**Figure 9**). The phylogenetic analysis was developed from predicted amino acid sequences assessed from cloned samples representing allelic forms that were identified. As displayed in **Figure 9**, the dendrogram suggests that the eIF4E family is separated into three distinct classes. Cloned *Ib*eIF4E sequences, eIF4E from *I. nil* (IN_ XP 019169243.1), and the eIF4E sequences obtained from the reference genomes were clustered into the same subgroup, which suggests that the eIF4E protein has a similar function as the eIF4E protein in other *Ipomoea* species such as *I. nil*. The amino acid sequence comparison of *Ib*eIF4E displayed its highest sequence identity with IN_XP 019169243.1 from *Ipomoea nil* (~96% - 97%), and the lowest identity with At_NP_193538.1. from *Arabidopsis thaliana* (~72%) displayed on the genetic distance and identity map shown in supplemental **Figure S1** data image. Interestingly, when assessing the predicted amino acid sequence alignment of *Ib*eIF4E compared with other eIF4E proteins from various

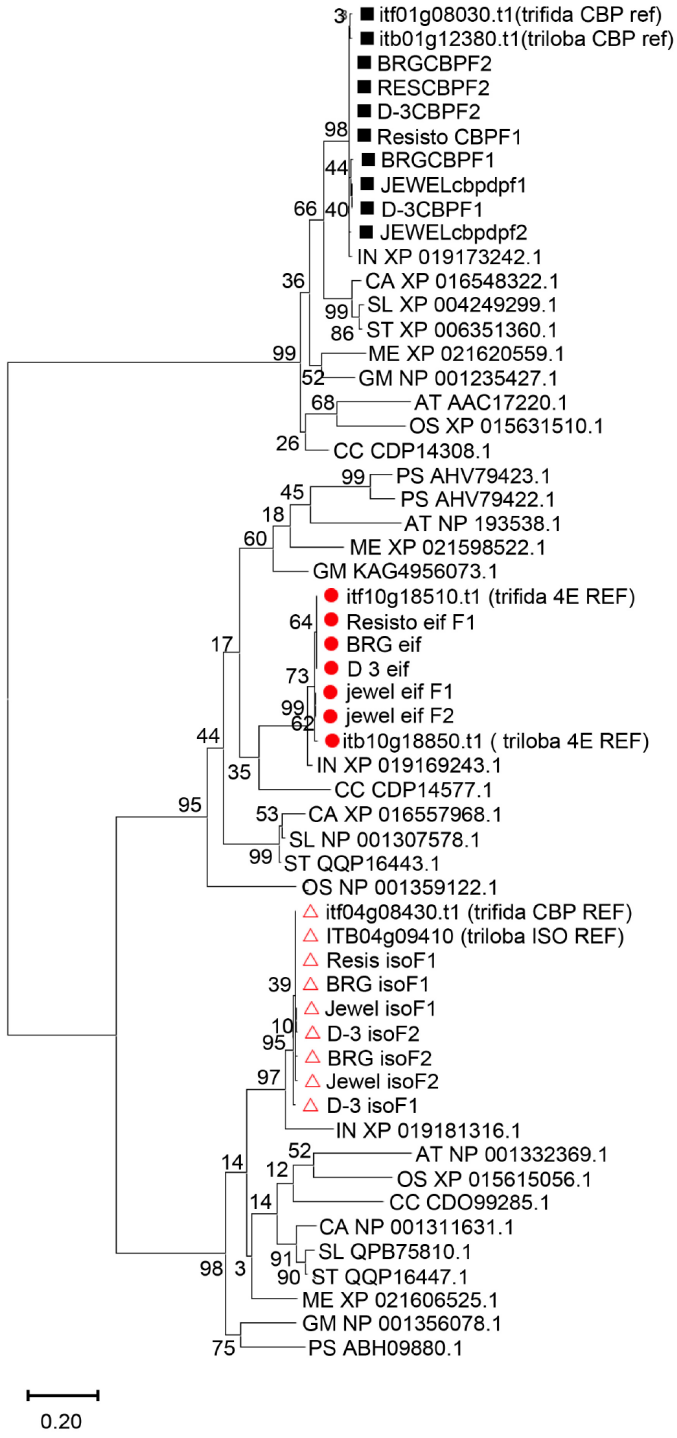


Figure 9. Molecular phylogenetic analysis of the ibeIF4E multigene family by maximum likelihood method against other species. The evolutionary history was inferred using Maximum-Likelihood method based on the JTT matrix-based model [82]. The tree with the highest log likelihood (−4193.37) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7004)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The numbers adjacent to the branches represent the percentage of the replicate trees where the associated taxa clustered within the bootstrap test. The analysis involved 56 amino acid sequences. All positions containing gaps and missing data were eliminated. There is a total of 180 positions in the final dataset. Evolutionary analyses were performed in MEGA7 [66].

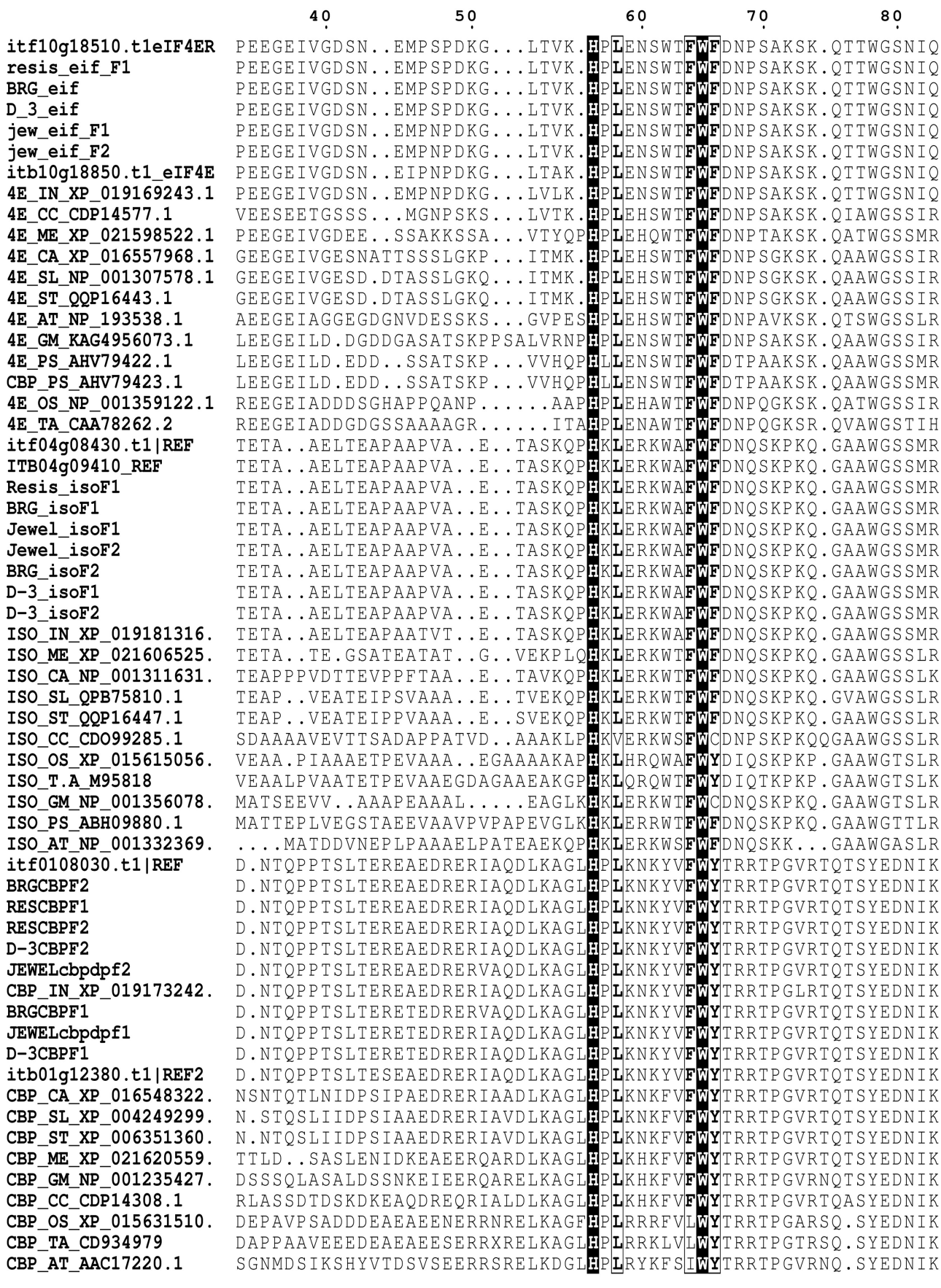
plant species, 109 identical regions were identified amongst all species. Of the 109 conserved regions, eIF4e has been characterized to contain a large consensus of tryptophan (W) residues that have been proven to play a significant role in binding to the 5' mRNA cap in other plant species [21] [26] [74] [75]. In *Ib*eIF4E, our results revealed a total of nine W residues that are identical amongst all species (**Figure 10**). Of the nine residues identified, three W residues (W-78, W-124 and W-183) and a glutamic acid (E-125) were found to be identical in all classified eIF4E family members and have been shown to interact with the mRNA cap binding structure as previously described by Joshi et al. in 2005 and Monzingo *et al*. in 2007 [74] [76] [77] [78].

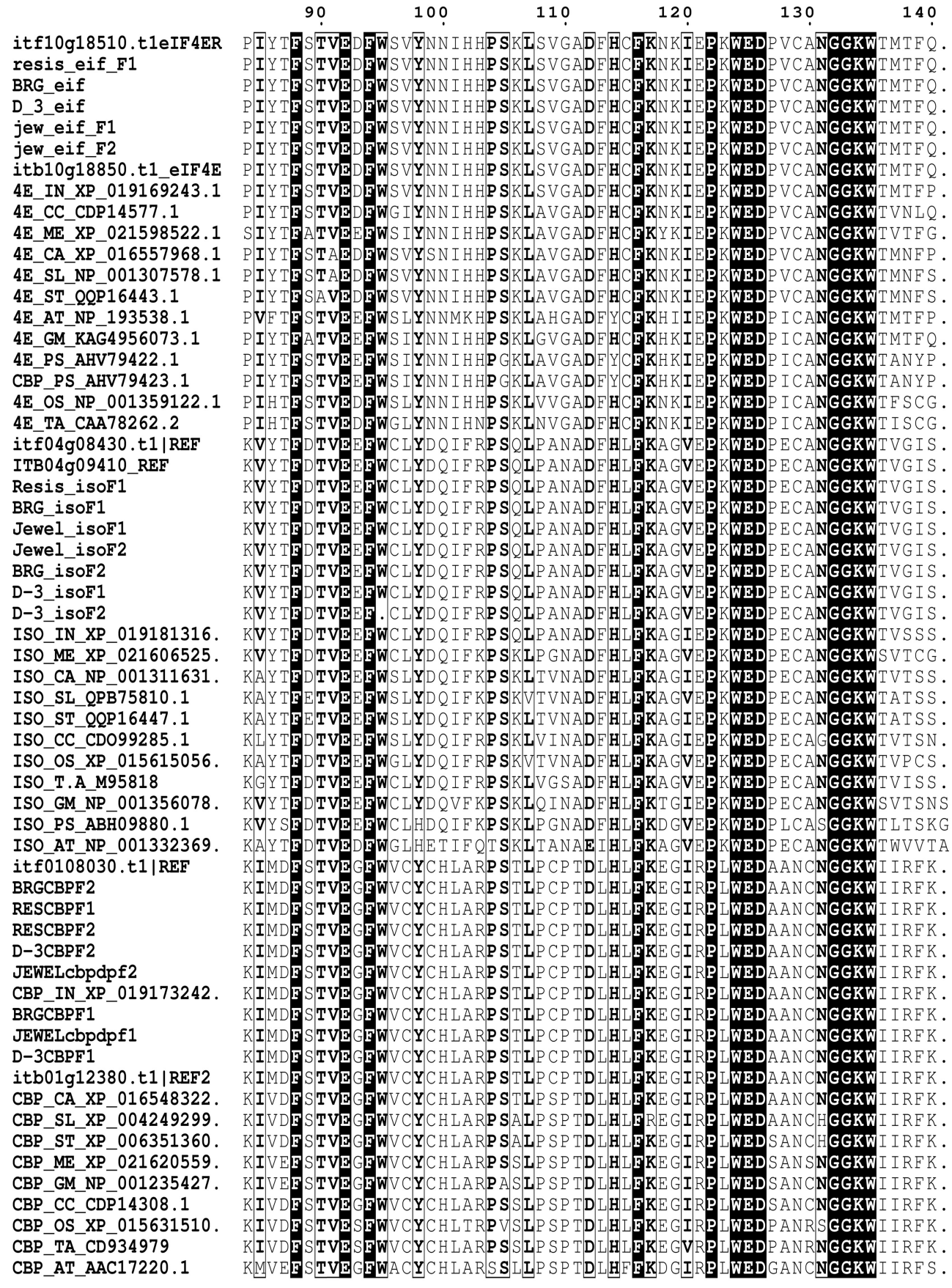
Evolutionary Analysis of Sweetpotato IbeIF(iso)4E. The results generated from the phylogeny revealed that the cloned sequence haplotypes from *Ib*eIF(iso)4E, the *Ib*eIF(iso)4E sequences from the reference genomes and eIF(iso)4E from *I. nil* (IN XP 019181316.1)were clustered into the same subgroup, which also suggests that the *Ib*eIF(iso)4E protein possesses the similar function to eIF(iso)4E proteins in other *Ipomoea* species. The amino acid sequence identity of *Ib*eIF(iso)4E with pre-existing eIF(iso)4Es from other plant species revealed its highest identity with *Ipomoea nil* (96.24%) and the lowest with *Arabidopsis thaliana* (67.20%) AT_NP_001332369.1 (**Figure 6**). The alignment analysis of *Ib*eIF(iso)4E in comparison to other species revealed 82 conserved regions (**Supplemental Figure S2**), nine of which were conserved tryptophan (W) residues as previously observed in the eIF(iso)4E amino acid sequence alignments of other species [76]. Residues W-48, W-94, E-95 and W-194 were identified as positions that were equivalent to residues that are directly associated with *Ib*eIF4E-mRNA-cap binding in other species [22] [23] [79] (**Figure 10**).

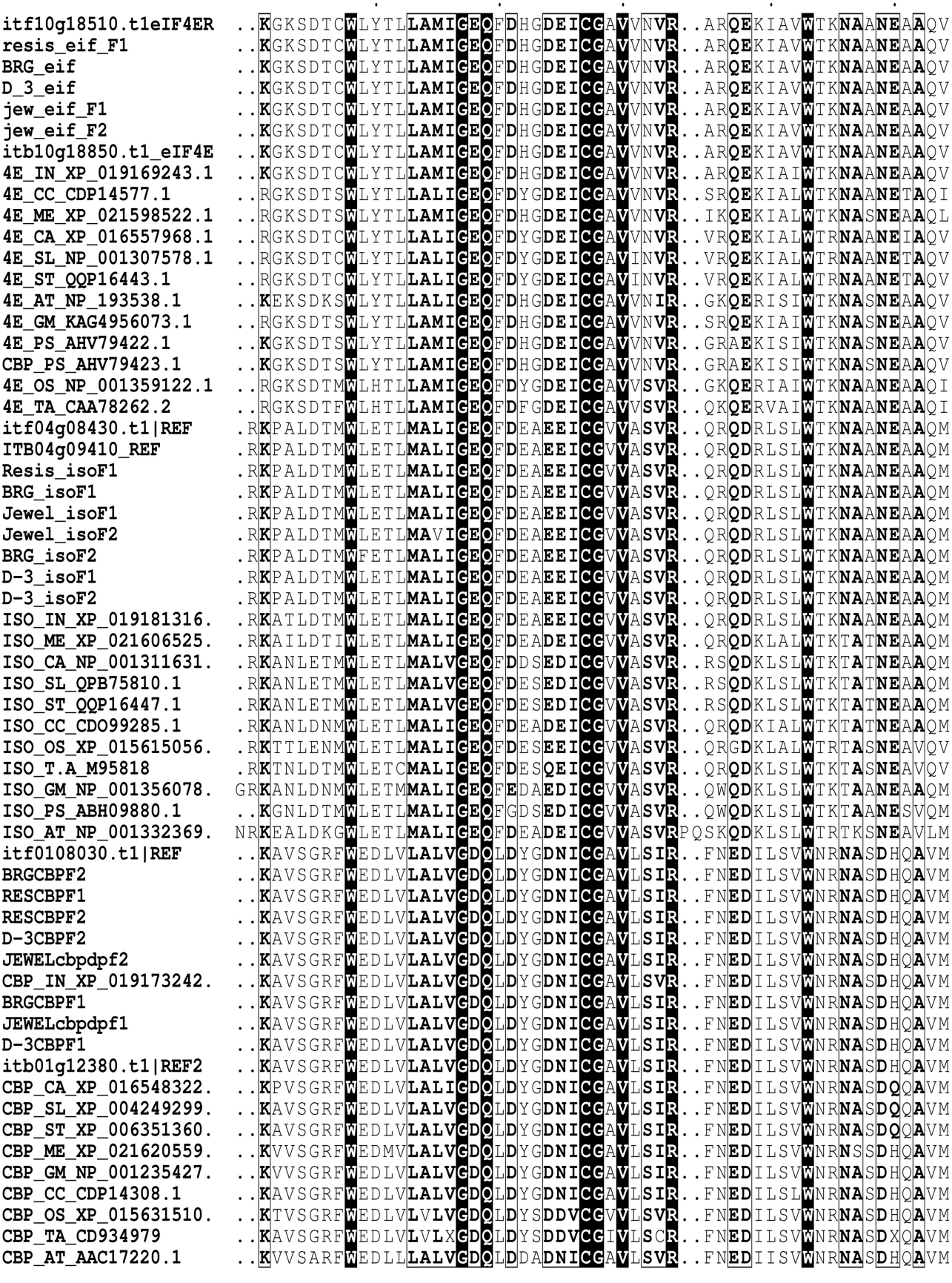
Evolutionary Analysis of Sweetpotato IbCBP. Based on the sequence homology from other plant species, *Ib*CBP haplotypes clustered into a subgroup with *Ib*CBP sequences from the reference genomes and CBP from *I. nil*, therefore, suggesting that *Ib*CBP share similar functions as the CBP proteins in other *Ipomoea* species. The amino acid sequence alignment performed for *Ib*CBP in comparison to other species revealed a high amino acid identity with XP_019173242.1 from *I. nil* (98% - 99%); whereas AAC17220.1 of *Arabidopsis thaliana* is shown as its lowest sequence identity (73% - 74%) based on the distance map generated from the amino acid sequence alignment (**Supplemental Figure S3**). We expected for *I. nil* to possess high sequence homology to *IbCBP* sequences as they share the same genus [80].

The phylogenetic analysis revealed that *Ib*eIF4E, *Ib*eIF(iso)4E and *Ib*CBP were members of three distinctly different evolutionary clades. When assessing the amino acid sequence alignment of *Ib*CBP, a total of 51 conserved regions were identified, of which seven residues accounted for the tryptophan residue. One of the primary observations noted in the amino acid sequence variations of CBP in comparison to eIF4E and eIF(iso)4E was the presence of two amino acid substitutions in CBP in lieu of the first two conserved tryptophan residues in

```
                                                1         10        20        30
itf10g18510.t1eIF4ER  ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
resis_eif_F1          ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
BRG_eif               ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
D_3_eif               ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
jew_eif_F1            ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
jew_eif_F2            ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
itb10g18850.t1_eIF4E  ..........................MVEEIEKSTKPEELEKSSSSARPSDGGEAGDG
4E_IN_XP_019169243.1  ..........................MVEELEKSTKPEEPEKSSSSARPGDGGEAGDG
4E_CC_CDP14577.1      ..........................MVEELEKPMA.EEKEEKIPPVEDDEGPEEGEI
4E_ME_XP_021598522.1  .........................MAAEEPLKSTTEETPNPNLNSNPRAQDDVNDDE
4E_CA_XP_016557968.1  ..........................MADEGNKTTMLGEYKTSSME..........DG
4E_SL_NP_001307578.1  ..........................MADELNKAA.LEEYKSSSVE..........DR
4E_ST_QQP16443.1      ..........................MADELNKAA.SEEYKTSSVE..........DG
4E_AT_NP_193538.1     .........................MAVEDTPKSVVTEEAKPNSIENPIDRYHEEGDD
4E_GM_KAG4956073.1    .........................MVVEDAQKSAITEDQNPSRVANDN....NDDED
4E_PS_AHV79422.1      .........................MVVEDTPKSIITDDQITTNPNRVI....EDDNN
CBP_PS_AHV79423.1     .........................MVVEDTPKSIITDDQITTNPNRVI....EDDNN
4E_OS_NP_001359122.1  ...............................MAEEHETRPPSAGRPPSSGRGRADDADE
4E_TA_CAA78262.2      ...............................MAEDTETRPASAG............AEE
itf04g08430.t1|REF    ..........................................................MA
ITB04g09410_REF       ..........................................................MA
Resis_isoF1           ..........................................................MA
BRG_isoF1             ..........................................................MA
Jewel_isoF1           ..........................................................MA
Jewel_isoF2           ..........................................................MA
BRG_isoF2             ..........................................................MA
D-3_isoF1             ..........................................................MA
D-3_isoF2             ..........................................................MA
ISO_IN_XP_019181316.  ....................MKSIHL.YEPFQERLSPRLQNYLSDSGTRDLY..HEMA
ISO_ME_XP_021606525.  ..........................................................MA
ISO_CA_NP_001311631.  ..........................................................MA
ISO_SL_QPB75810.1     ..........................................................MA
ISO_ST_QQP16447.1     ..........................................................MA
ISO_CC_CDO99285.1     MESSPSGKHQKLLNRPPLVLIGPPRLGYQNLRPKLTSHQNANLRETRGRNLETEREMA
ISO_OS_XP_015615056.  ..........................................................MAE
ISO_T.A_M95818        ..........................................................MAE
ISO_GM_NP_001356078.  ..........................................................
ISO_PS_ABH09880.1     ..........................................................
ISO_AT_NP_001332369.  ..........................................................
itf0108030.t1|REF     ..........................................MEEAIAEK..KETESRSV
BRGCBPF2              ..........................................MEEAIAEK..KETESRTV
RESCBPF1              ..........................................MEEAIAEK..KETESRTV
RESCBPF2              ..........................................MEEAIAEK..KETESRTV
D-3CBPF2              ..........................................MEEAIAEK..KETESRTV
JEWELcbpdpf2          ..........................................MEEAIAEK..KETESRTV
CBP_IN_XP_019173242.  ..........................................MEEAIAEK..KETESRTV
BRGCBPF1              ..........................................MEEAIAEK..KETESRTV
JEWELcbpdpf1          ..........................................MEEAIAEK..KETESRTV
D-3CBPF1              ..........................................MEEAIAEK..KETESRTV
itb01g12380.t1|REF2   ..........................................MEEAIAEK..KETESRTV
CBP_CA_XP_016548322.  ...........................................MEVSSEK..KELDNKNT
CBP_SL_XP_004249299.  ...........................................MEVTSEK..KELDNKNT
CBP_ST_XP_006351360.  ...........................................MEVTPEK..KELDNKNT
CBP_ME_XP_021620559.  .....................................MEIT.EKKDTENNT..NNSNNNAQ
CBP_GM_NP_001235427.  .....................................MDFTAEKKESESNT..ENAHPTTL
CBP_CC_CDP14308.1     .....................................MEVAAAVEKKESET..NNSSSDLS
CBP_OS_XP_015631510.  .....................................MEPAAEKREAEQEE.LQQQH....
CBP_TA_CD934979       .....................................MEPAVERKVPEQEEQLQPSHARAE
CBP_AT_AAC17220.1     ..........................................MEVLDRRDDEIRD
```


40
50
60
70
80
itf10g18510.t1eIF4ER PEEGEIVGDSN..EMPSPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
resis_eif_F1 PEEGEIVGDSN..EMPSPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
BRG_eif PEEGEIVGDSN..EMPSPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
D_3_eif PEEGEIVGDSN..EMPSPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
jew_eif_F1 PEEGEIVGDSN..EMPNPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
jew_eif_F2 PEEGEIVGDSN..EMPNPDKG...LTVK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
itb10g18850.t1_eIF4E PEEGEIVGDSN..EIPNPDKG...LTAK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
4E_IN_XP_019169243.1 PEEGEIVGDSN..EMPNPDKG...LVLK.HPLENSWTFWFDNPSAKSK.QTTWGSNIQ
4E_CC_CDP14577.1 VEESEETGSSS...MGNPSKS...LVTK.HPLEHSWTFWFDNPSAKSK.QIAWGSSIR
4E_ME_XP_021598522.1 PEEGEIVGDEE..SSAKKSSA...VTYQPHPLEHQWTFWFDNPTAKSK.QATWGSSMR
4E_CA_XP_016557968.1 GEEGEIVGESNATTSSSLGKP...ITMK.HPLEHSWTFWFDNPSGKSK.QAAWGSSIR
4E_SL_NP_001307578.1 GEEGEIVGESD.DTASSLGKQ...ITMK.HPLEHSWTFWFDNPSGKSK.QAAWGSSIR
4E_ST_QQP16443.1 GEEGEIVGESD.DTASSLGKQ...ITMK.HPLEHSWTFWFDNPSGKSK.QAAWGSSIR
4E_AT_NP_193538.1 AEEGEIAGGEGDGNVDESSKS...GVPESHPLEHSWTFWFDNPAVKSK.QTSWGSSLR
4E_GM_KAG4956073.1 LEEGEILD.DGDDGASATSKPPSALVRNPHPLENSWTFWFDNPSAKSK.QAAWGSSIR
4E_PS_AHV79422.1 LEEGEILD.EDD..SSATSKP...VVHQPHLLENSWTFWFDTPAAKSK.QAAWGSSMR
CBP_PS_AHV79423.1 LEEGEILD.EDD..SSATSKP...VVHQPHLLENSWTFWFDTPAAKSK.QAAWGSSMR
4E_OS_NP_001359122.1 REEGEIADDDSGHAPPQANP.....AAPHPLEHAWTFWFDNPQGKSK.QATWGSSIR
4E_TA_CAA78262.2 REEGEIADDGDGSSAAAAGR.....ITAHPLENAWTFWFDNPQGKSR.QVAWGSTIH
itf04g08430.t1|REF TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
ITB04g09410_REF TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
Resis_isoF1 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
BRG_isoF1 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
Jewel_isoF1 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
Jewel_isoF2 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
BRG_isoF2 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
D-3_isoF1 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
D-3_isoF2 TETA..AELTEAPAAPVA..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
ISO_IN_XP_019181316. TETA..AELTEAPAATVT..E..TASKQPHKLERKWAFWFDNQSKPKQ.GAAWGSSMR
ISO_ME_XP_021606525. TETA..TE.GSATEATAT..G..VEKPLQHKLERKWTFWFDNQSKPKQ.GAAWGSSLR
ISO_CA_NP_001311631. TEAPPPVDTTEVPPFTAA..E..TAVKQPHKLERKWTFWFDNQSKPKQ.GAAWGSSLK
ISO_SL_QPB75810.1 TEAP..VEATEIPSVAAA..E..TVEKQPHKLERKWTFWFDNQSKPKQ.GVAWGSSLR
ISO_ST_QQP16447.1 TEAP..VEATEIPPVAAA..E..SVEKQPHKLERKWTFWFDNQSKPKQ.GAAWGSSLR
ISO_CC_CDO99285.1 SDAAAAVEVTTSADAPPATVD..AAAKLPHKVERKWSFWCDNPSKPKQQGAAWGSSLR
ISO_OS_XP_015615056. VEAA.PIAAAETPEVAAA..EGAAAAKAPHKLHRQWAFWYDIQSKPKP.GAAWGTSLR
ISO_T.A_M95818 VEAALPVAATETPEVAAEGDAGAAEAKGPHKLQRQWTFWYDIQTKPKP.GAAWGTSLK
ISO_GM_NP_001356078. MATSEEVV..AAAPEAAAL.....EAGLKHKLERKWTFWCDNQSKPKQ.GAAWGTSLR
ISO_PS_ABH09880.1 MATTEPLVEGSTAEEVAAVPVPAPEVGLKHKLERRWTFWFDNQSKPKQ.GAAWGTTLR
ISO_AT_NP_001332369. ....MATDDVNEPLPAAAELPATEAEKQPHKLERKWSFWFDNQSKK...GAAWGASLR
itf0108030.t1|REF D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
BRGCBPF2 D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
RESCBPF1 D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
RESCBPF2 D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
D-3CBPF2 D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
JEWELcbpdpf2 D.NTQPPTSLTEREAEDRERVAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
CBP_IN_XP_019173242. D.NTQPPTSLTEREAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGLRTQTSYEDNIK
BRGCBPF1 D.NTQPPTSLTERETEDRERVAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
JEWELcbpdpf1 D.NTQPPTSLTERETEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
D-3CBPF1 D.NTQPPTSLTERETEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
itb01g12380.t1|REF2 D.NTQPPTSLTESEAEDRERIAQDLKAGLHPLKNKYVFWYTRRTPGVRTQTSYEDNIK
CBP_CA_XP_016548322. NSNTQTLNIDPSIPAEDRERIAADLKAGLHPLKNKFVFWYTRRTPGVRTQTSYEDNIK
CBP_SL_XP_004249299. N.STQSLIIDPSIAAEDRERIAVDLKAGLHPLKNKFVFWYTRRTPGVRTQTSYEDNIK
CBP_ST_XP_006351360. N.NTQSLIIDPSIAAEDRERIAVDLKAGLHPLKNKFVFWYTRRTPGVRTQTSYEDNIK
CBP_ME_XP_021620559. TTLD..SASLENIDKEAEERQARDLKAGLHPLKHKFVFWYTRRTPGVRTQTSYEDNIK
CBP_GM_NP_001235427. DSSSQLASALDSSNKEIEERQARELKAGLHPLKHKFVFWYTRRTPGVRNQTSYEDNIK
CBP_CC_CDP14308.1 RLASSDTDSKDKEAQDREQRIALDLKAGLHPLKHKFVFWYTRRTPGVRTQASYEDNIK
CBP_OS_XP_015631510. DEPAVPSADDDEAEAEENERRNRELKAGFHPLRRRFVLWYTRRTPGARSQ.SYEDNIK
CBP_TA_CD934979 DAPPAAVEEEDEAEAEESERRXRELKAGLHPLRRKLVLWYTRRTPGTRSQ.SYEDNIK
CBP_AT_AAC17220.1 SGNMDSIKSHYVTDSVSEERRSRELKDGLHPLRYKFSIWYTRRTPGVRNQ.SYEDNIK


90
100
110
120
130
140
itf10g18510.t1eIF4ER PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
resis_eif_F1 PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
BRG_eif PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
D_3_eif PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
jew_eif_F1 PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
jew_eif_F2 PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
itb10g18850.t1_eIF4E PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFQ.
4E_IN_XP_019169243.1 PIYTFSTVEDFWSVYNNIHHPSKLSVGADFHCFKNKIEPKWEDPVCANGGKWTMTFP.
4E_CC_CDP14577.1 PIYTFSTVEDFWGIYNNIHHPSKLAVGADFHCFKNKIEPKWEDPVCANGGKWTVNLQ.
4E_ME_XP_021598522.1 SIYTFATVEEFWSIYNNIHHPSKLAVGADFHCFKYKIEPKWEDPVCANGGKWTVTFG.
4E_CA_XP_016557968.1 PIYTFSTAEDFWSVYSNIHHPSKLAVGADFHCFKNKIEPKWEDPVCANGGKWTMNFP.
4E_SL_NP_001307578.1 PIYTFSTAEDFWSVYNNIHHPSKLAVGADFHCFKNKIEPKWEDPVCANGGKWTMNFS.
4E_ST_QQP16443.1 PIYTFSAVEDFWSVYNNIHHPSKLAVGADFHCFKNKIEPKWEDPVCANGGKWTMNFS.
4E_AT_NP_193538.1 PVFTFSTVEEFWSLYNNMKHPSKLAHGADFYCFKHIIEPKWEDPICANGGKWTMTFP.
4E_GM_KAG4956073.1 PIYTFATVEEFWSIYNNIHHPSKLGVGADFHCFKHKIEPKWEDPICANGGKWTMTFQ.
4E_PS_AHV79422.1 PIYTFSTVEEFWSIYNNIHHPGKLAVGADFYCFKHKIEPKWEDPICANGGKWTANYP.
CBP_PS_AHV79423.1 PIYTFSTVEEFWSIYNNIHHPGKLAVGADFYCFKHKIEPKWEDPICANGGKWTANYP.
4E_OS_NP_001359122.1 PIHTFSTVEDFWSLYNNIHHPSKLVVGADFHCFKNKIEPKWEDPICANGGKWTFSCG.
4E_TA_CAA78262.2 PIHTFSTVEDFWGLYNNIHNPSKLNVGADFHCFKNKIEPKWEDPICANGGKWTISCG.
itf04g08430.t1|REF KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
ITB04g09410_REF KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
Resis_isoF1 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
BRG_isoF1 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
Jewel_isoF1 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
Jewel_isoF2 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
BRG_isoF2 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
D-3_isoF1 KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
D-3_isoF2 KVYTFDTVEEF.CLYDQIFRPSQLPANADFHLFKAGVEPKWEDPECANGGKWTVGIS.
ISO_IN_XP_019181316. KVYTFDTVEEFWCLYDQIFRPSQLPANADFHLFKAGIEPKWEDPECANGGKWTVSSS.
ISO_ME_XP_021606525. KVYTFDTVEEFWCLYDQIFKPSKLPGNADFHLFKAGVEPKWEDPECANGGKWSVTCG.
ISO_CA_NP_001311631. KAYTFDTVEEFWSLYDQIFKPSKLTVNADFHLFKAGIEPKWEDPECANGGKWTVTSS.
ISO_SL_QPB75810.1 KAYTFETVEEFWSLYDQIFKPSKVTVNADFHLFKAGVEPKWEDPECANGGKWTATSS.
ISO_ST_QQP16447.1 KAYTFETVEEFWSLYDQIFKPSKLTVNADFHLFKAGIEPKWEDPECANGGKWTATSS.
ISO_CC_CDO99285.1 KLYTFDTVEEFWSLYDQIFRPSKLVINADFHLFKAGIEPKWEDPECAGGGKWTVTSN.
ISO_OS_XP_015615056. KAYTFDTVEEFWGLYDQIFRPSKVTVNADFHLFKAGVEPKWEDPECANGGKWTVPCS.
ISO_T.A_M95818 KGYTFDTVEEFWCLYDQIFRPSKLVGSADFHLFKAGVEPKWEDPECANGGKWTVISS.
ISO_GM_NP_001356078. KVYTFDTVEEFWCLYDQVFKPSKLQINADFHLFKTGIEPKWEDPECANGGKWSVTSNS
ISO_PS_ABH09880.1 KVYSFDTVEEFWCLHDQIFKPSKLPGNADFHLFKDGVEPKWEDPLCASGGKWTLTSKG
ISO_AT_NP_001332369. KAYTFDTVEDFWGLHETIFQTSKLTANAEIHLFKAGVEPKWEDPECANGGKWTWVVTA
itf0108030.t1|REF KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
BRGCBPF2 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
RESCBPF1 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
RESCBPF2 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
D-3CBPF2 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
JEWELcbpdpf2 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
CBP_IN_XP_019173242. KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
BRGCBPF1 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
JEWELcbpdpf1 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
D-3CBPF1 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
itb01g12380.t1|REF2 KIMDFSTVEGFWVCYCHLARPSTLPCPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
CBP_CA_XP_016548322. KIVDFSTVEGFWVCYCHLARPSTLPSPTDLHLFKEGIRPLWEDAANCNGGKWIIRFK.
CBP_SL_XP_004249299. KIVDFSTVEGFWVCYCHLARPSALPSPTDLHLFREGIRPLWEDAANCHGGKWIIRFK.
CBP_ST_XP_006351360. KIVDFSTVEGFWVCYCHLARPSALPSPTDLHLFKEGIRPLWEDSANCHGGKWIIRFK.
CBP_ME_XP_021620559. KIVEFSTVEGFWVCYCHLARPSSLPSPTDLHLFKEGIRPLWEDSANSNGGKWIIRFK.
CBP_GM_NP_001235427. KIVEFSTVEGFWVCYCHLARPASLPSPTDLHLFKEGIRPLWEDSANCNGGKWIIRFK.
CBP_CC_CDP14308.1 KIVDFSTVEGFWVCYCHLARPSSLPSPTDLHLFKEGIRPLWEDSANCNGGKWIIRFK.
CBP_OS_XP_015631510. KIVDFSTVESFWVCYCHLTRPVSLPSPTDLHLFKEGIRPLWEDPANRSGGKWIIRFK.
CBP_TA_CD934979 KIVDFSTVESFWVCYCHLARPSSLPSPTDLHLFKEGVRPLWEDPANRNGGKWIIRFK.
CBP_AT_AAC17220.1 KMVEFSTVEGFWACYCHLARSSLLPSPTDLHFFKDGIRPLWEDGANCNGGKWIIRFS.


150
160
170
180
190
itf10g18510.t1eIF4ER
resis_eif_F1
BRG_eif
D_3_eif
jew_eif_F1
jew_eif_F2
itb10g18850.t1_eIF4E
4E_IN_XP_019169243.1
4E_CC_CDP14577.1
4E_ME_XP_021598522.1
4E_CA_XP_016557968.1
4E_SL_NP_001307578.1
4E_ST_QQP16443.1
4E_AT_NP_193538.1
4E_GM_KAG4956073.1
4E_PS_AHV79422.1
CBP_PS_AHV79423.1
4E_OS_NP_001359122.1
4E_TA_CAA78262.2
itf04g08430.t1|REF
ITB04g09410_REF
Resis_isoF1
BRG_isoF1
Jewel_isoF1
Jewel_isoF2
BRG_isoF2
D-3_isoF1
D-3_isoF2
ISO_IN_XP_019181316.
ISO_ME_XP_021606525.
ISO_CA_NP_001311631.
ISO_SL_QPB75810.1
ISO_ST_QQP16447.1
ISO_CC_CDO99285.1
ISO_OS_XP_015615056.
ISO_T.A_M95818
ISO_GM_NP_001356078.
ISO_PS_ABH09880.1
ISO_AT_NP_001332369.
itf0108030.t1|REF
BRGCBPF2
RESCBPF1
RESCBPF2
D-3CBPF2
JEWELcbpdpf2
CBP_IN_XP_019173242.
BRGCBPF1
JEWELcbpdpf1
D-3CBPF1
itb01g12380.t1|REF2
CBP_CA_XP_016548322.
CBP_SL_XP_004249299.
CBP_ST_XP_006351360.
CBP_ME_XP_021620559.
CBP_GM_NP_001235427.
CBP_CC_CDP14308.1
CBP_OS_XP_015631510.
CBP_TA_CD934979
CBP_AT_AAC17220.1

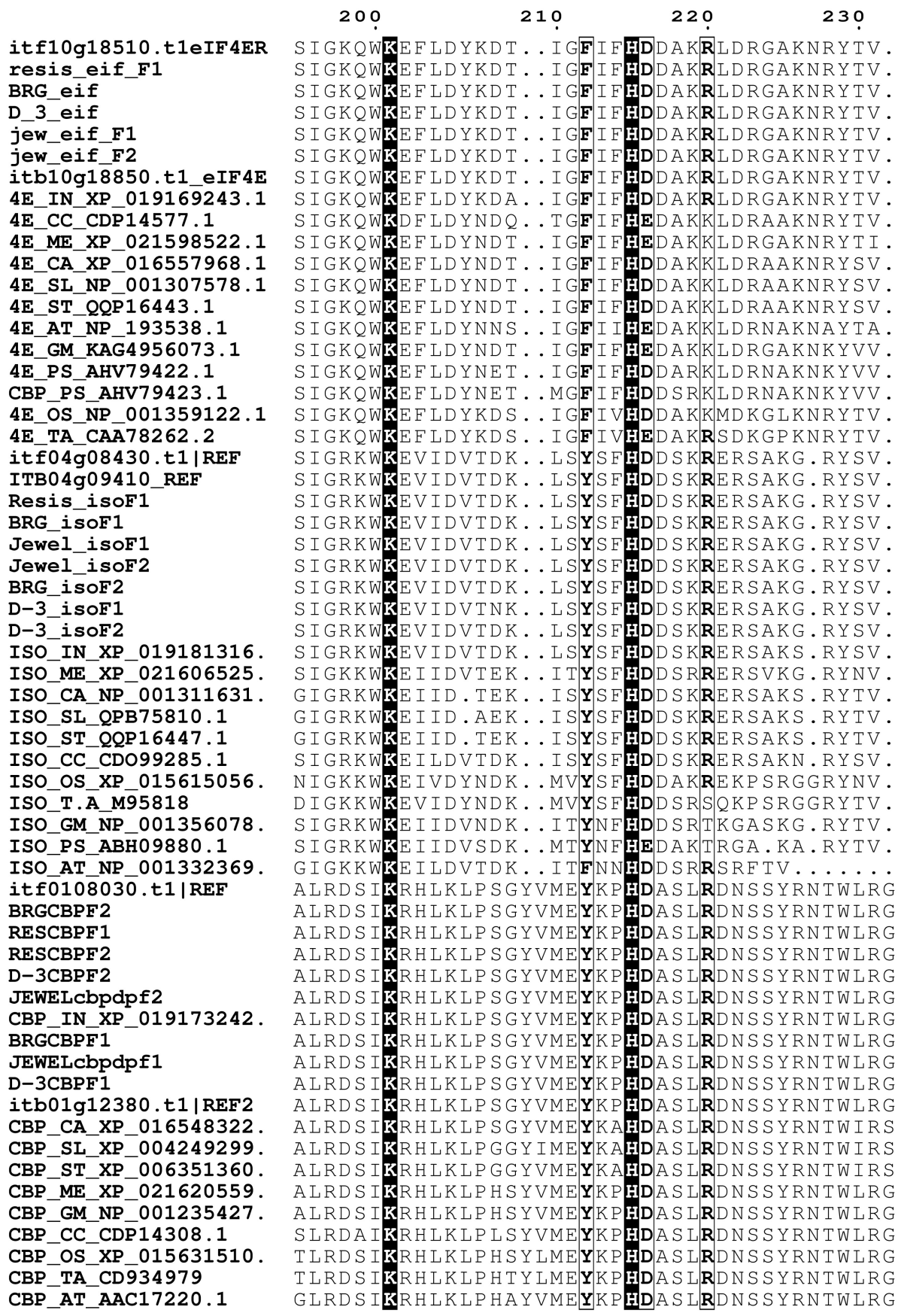

200
210
220
230
itf10g18510.t1eIF4ER SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
resis_eif_F1 SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
BRG_eif SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
D_3_eif SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
jew_eif_F1 SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
jew_eif_F2 SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
itb10g18850.t1_eIF4E SIGKQWKEFLDYKDT..IGFIFHDDAKRLDRGAKNRYTV.
4E_IN_XP_019169243.1 SIGKQWKEFLDYKDA..IGFIFHDDAKRLDRGAKNRYTV.
4E_CC_CDP14577.1 SIGKQWKDFLDYNDQ..TGFIFHEDAKKLDRAAKNRYTV.
4E_ME_XP_021598522.1 SIGKQWKEFLDYNDT..IGFIFHEDAKKLDRGAKNRYTI.
4E_CA_XP_016557968.1 SIGKQWKEFLDYNDT..IGFIFHDDAKKLDRAAKNRYSV.
4E_SL_NP_001307578.1 SIGKQWKEFLDYNDT..IGFIFHDDAKKLDRAAKNRYSV.
4E_ST_QQP16443.1 SIGKQWKEFLDYNDT..IGFIFHDDAKKLDRAAKNRYSV.
4E_AT_NP_193538.1 SIGKQWKEFLDYNNS..IGFIIHEDAKKLDRNAKNAYTA.
4E_GM_KAG4956073.1 SIGKQWKEFLDYNDT..IGFIFHEDAKKLDRGAKNKYVV.
4E_PS_AHV79422.1 SIGKQWKEFLDYNET..IGFIFHDDARKLDRNAKNKYVV.
CBP_PS_AHV79423.1 SIGKQWKEFLDYNET..MGFIFHDDSRKLDRNAKNKYVV.
4E_OS_NP_001359122.1 SIGKQWKEFLDYKDS..IGFIVHDDAKKMDKGLKNRYTV.
4E_TA_CAA78262.2 SIGKQWKEFLDYKDS..IGFIVHEDAKRSDKGPKNRYTV.
itf04g08430.t1|REF SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
ITB04g09410_REF SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
Resis_isoF1 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
BRG_isoF1 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
Jewel_isoF1 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
Jewel_isoF2 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
BRG_isoF2 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
D-3_isoF1 SIGRKWKEVIDVTNK..LSYSFHDDSKRERSAKG.RYSV.
D-3_isoF2 SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKG.RYSV.
ISO_IN_XP_019181316. SIGRKWKEVIDVTDK..LSYSFHDDSKRERSAKS.RYSV.
ISO_ME_XP_021606525. SIGKKWKEIIDVTEK..ITYSFHDDSRRERSVKG.RYNV.
ISO_CA_NP_001311631. GIGRKWKEIID.TEK..ISYSFHDDSKRERSAKS.RYTV.
ISO_SL_QPB75810.1 GIGRKWKEIID.AEK..ISYSFHDDSKRERSAKS.RYTV.
ISO_ST_QQP16447.1 GIGRKWKEIID.TEK..ISYSFHDDSKRERSAKS.RYTV.
ISO_CC_CDO99285.1 SIGRKWKEILDVTDK..ISYSFHDDSKRERSAKN.RYSV.
ISO_OS_XP_015615056. NIGKKWKEIVDYNDK..MVYSFHDDAKREKPSRGGRYNV.
ISO_T.A_M95818 DIGKKWKEVIDYNDK..MVYSFHDDSRSQKPSRGGRYTV.
ISO_GM_NP_001356078. SIGRKWKEIIDVNDK..ITYNFHDDSRTKGASKG.RYTV.
ISO_PS_ABH09880.1 SIGRKWKEIIDVSDK..MTYNFHEDAKTRGA.KA.RYTV.
ISO_AT_NP_001332369. GIGKKWKEILDVTDK..ITFNNHDDSRRSRFTV.......
itf0108030.t1|REF ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
BRGCBPF2 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
RESCBPF1 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
RESCBPF2 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
D-3CBPF2 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
JEWELcbpdpf2 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
CBP_IN_XP_019173242. ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
BRGCBPF1 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
JEWELcbpdpf1 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
D-3CBPF1 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
itb01g12380.t1|REF2 ALRDSIKRHLKLPSGYVMEYKPHDASLRDNSSYRNTWLRG
CBP_CA_XP_016548322. ALRDSIKRHLKLPSGYVMEYKAHDASLRDNSSYRNTWIRS
CBP_SL_XP_004249299. ALRDSIKRHLKLPGGYIMEYKAHDASLRDNSSYRNTWIRS
CBP_ST_XP_006351360. ALRDSIKRHLKLPGGYVMEYKAHDASLRDNSSYRNTWIRS
CBP_ME_XP_021620559. ALRDSIKRHLKLPHSYVMEYKPHDASLRDNSSYRNTWLRG
CBP_GM_NP_001235427. ALRDSIKRHLKLPHSYVMEYKPHDASLRDNSSYRNTWLRG
CBP_CC_CDP14308.1 SLRDAIKRHLKLPLSYVMEYKPHDASLRDNSSYRNTWLRG
CBP_OS_XP_015631510. TLRDSIKRHLKLPHSYLMEYKPHDASLRDNSSYRNTWLRG
CBP_TA_CD934979 TLRDSIKRHLKLPHTYLMEYKPHDASLRDNSSYRNTWLRG
CBP_AT_AAC17220.1 GLRDSIKRHLKLPHAYVMEYKPHDASLRDNSSYRNTWLRG

Figure 10. Amino acid sequence alignment of sweetpotato *IbeIF*4*E*, *IbeIF*(*iso*)4*E* and *IbCBP* Deduced from Cloned Haplotypes against Several Plant species. For *IbeIF*4*E*, sequences were compared with *eIF*4*E* sequences from *A. thaliana* (AT_NP193538.1), pepper (*C. annuum*; CA_XP_016557968.1), cassava (*M. esculenta*; ME_XP_021598522.1), Japanese morning glory (*I. nil*; IN_XP_019169243.1), coffee (*C. canephora*; CC_CDP14577.1), rice (*O. sativa*; OS_NP_001359122.1), soybean (*G. max*; GM_KAG4956073.1), pea (*P. sativum*; PS_AHV79422.1), tomato (*S. lycopersicum*; SL_NP_001307578.1), wheat (*T. aestivum*; CAA78262.2), and potato (*S. tuberosum* ST_QQP16443.1). For *IbeIF*(*iso*)4*E*, sequences were compared with *eIF*(*iso*)4*E* sequences from *A. thaliana* (AT_NP001332369.1), *C. annuum* (CA_XP_001311631.1), *M. esculenta* (ME_XP_021606525.1), *I. nil* (IN_XP_019181316.1), C. *canephora* (CC_CDO99285.1), *O. sativa* (OS_NP_015615056.1), *G. max* (GM_001356078.1), *P. sativum* (PS_ABH09880.1), *S. lycopersicum* (SL_QPB75810.1), *T. aestivum* (TA_M95818) and *S. tuberosum* (ST_QQP16447.1). For *IbCBP*, sequences were compared with *CBP* sequences from *A. thaliana* (AT_AAC17220.1), *C. annuum* (CA_XP_016548322.1), *M. esculenta* (ME_XP_021620559.1), *I. nil* (IN_XP_019173242.1), *C. canephora* (CC_CDP14308.1), *O. sativa* (OS_XP_015631510.1), *G. max* (GM_NP_001235427.1), *P. sativum* (PS_AHV79423.1), *S. lycopersicum* (SL_XP_004249299.1), *T. aestivum* (CD934979) and *S. tuberosum* (ST_XP_006351360.1). The multiple sequence alignment was performed using Blosum [66] and shaded with ESPript 3.0 [68]. Highly conserved tryptophan residues were indicated by numbers. *IbCBP* sequences differentiated by amino acid substitutions in conserved Trp residues are numbered and contain red asterisks. Yellow amino acid residues signify sequence similarities within each group. The principal sequence eIF4E superfamily domain most identified in all eukaryotic translation initiation factors 4Es were underlined in red. Numbers above the sequences indicate amino acid positions from the N-terminal direction.

Arabidopsis thaliana eIF4E and eIF(iso)4E [81]. In exchange, the amino acid residues were replaced with alternative aromatic amino acid residues. In this study, it was observed that the first and the third tryptophan residues in *IbeIF*4*E* and *IbeIF*(*iso*)4*E* were substituted for an alternative aromatic amino acid, tyrosine, in *Ib*CBP. The observations regarding the conservation of tryptophan residues in sweetpotato further confirm the specificity of each gene that was sequenced.

5. Discussion

In this study, hexaploid sweetpotato cultivars Jewel, Resisto, Beauregard, and PI-318846 which exhibit unknown, tolerance or susceptibility to SPFMV served as donors for the cloning and characterization of the eIF4E family because the commercial varieties (Jewel and Beauregard) show periodic resistance to SPFMV due to low viral load during the growing season [5] [6]. Three *eIF*4*E* full-length cDNAs were annotated in these cultivars through two routes: 1) direct-sequencing of PCR-amplified cDNA amplicons and 2) sequencing of the cloned cDNA amplicons that were compared against the sweetpotato genomics resource database (http://sweetpotato.uga.edu/) containing the reference genomes of *I trifida* and *I. triloba* [59]. Additionally, each gene sequence was blasted against the conserved consensus sequences of *M. esculenta* and *A. thaliana* eIF4Es. Results mined from the direct-sequencing of PCR-amplified cDNA and the cloned sequences along with their protein prediction and phylogenetic analyses revealed multiple haplotypes of each gene among the cultivars assessed (**Figures 3-8**) in comparing all sequences to *I. trifida*. We detected a total of four haplotypes in Resisto, two haplotypes in D-3 and Jewel, and one haplotype in Beauregard. While *IbeIF*(*iso*)4*E* sequencing assessment demonstrated the presence of three haplotypes in cultivars Resisto, D-3, and Beauregard, and two haplotypes in Jewel. We also determined that *IbCBP* had two haplotypes within each

of the four cultivars. Similar results were obtained in root the crop cassava and the polyploidy species of sugarcane, where the cloning of the eIF4E protein gene family revealed varying gene copy numbers for eIF(iso)4E, CBP and eIF4E [14] [26] [27] [33] [34] [35] [36] [37]. Screening for the polymorphisms revealed that nucleotide variations against *I. trifida* were identical to the *I. triloba* genome. The outcome from these results further concurs with Austin's [67] postulation that hexaploid sweetpotato species emerged from an allopolyploid origin [67]. However, a selection of the hexaploid sweetpotato predicted translated allelic forms in this work showed silent mutations, therefore encoding the same amino acid sequences resulting in no conformational changes to the protein except for D-3 *ibeIF*(*iso*) haplotype 3, which had an in-frame deletion. The CDS sequences for each gene length were confirmed as 696 bp (*IbeIF4E*), and 606 bp (*IbeIF*(*iso*)*4E*) and 675 bp (*IbCBP*) encoding 232, 202 and 225 amino acids respectively. The multi-sequence alignment analysis revealed that the deduced amino acid sequence for all *Ib*eIF4E classes displayed homology to eIF4E classes from other species, including the conserved 160 amino acid IF4E superfamily domain located in-between the N-terminal and C-terminal ends (**Figure 10**). In conjunction, we identified conserved tryptophan residues within each class that is known to interact with the mRNA-cap structure of eIF4Es from other organisms such as *T. asetivum*, *H. sapiens* (human) and *M. musculus* (mouse) as depicted in **Figure 10** [27] [74] [75] [78].

Furthermore, when comparing the deduced amino acid sequences of each haplotype against both *Ipomoea* reference genomes, amino acid variations were detected within the commercial varieties Jewel (*Ib*eIF4E haplotype two, *IbeIF*(*iso*)*4E* haplotype two, *IbCBP* haplotype one, and haplotype two) and Beauregard (*IbeIF*(*iso*)*4E* haplotype three, *IbCBP* haplotype one, and haplotype two) (**Table 4**). We examined a unique amino acid residue variation within these

Table 4. A consensus of Amino acid variations identified from cloned ibeif4e super-family protein sequences.

Gene/Amino Acid position	*I. trifida*	*I. triloba*	Beauregard	Jewel	Resisto	PI-318846
IbeIF4E47	S	N	S	N	S	S
IbeIF(*iso*)*4E*65	W	W	W	W	W	In-frame deletion
IbeIF(iso)4E120	L	L	F	L	L	L
IbeIF(iso)4E126	V	V	V	L	V	V
IbeIF(iso)4E179	D	D	D	D	D	N
IbCBP15	S	T	T	S	T	T/S
IbCBP30	A	A	T/A	T/A	A	T/A
IbCBP36	I	I	V/I	I/V	I	I

a. A representation of amino acid residue differences in detected haplotype sequences among commercial varieties Jewel (IbeIF4E haplotype two, IbeIF(iso)4E haplotype two, IbCBP haplotype one and haplotype two) and Beauregard (three (IbeIF(iso)4E haplotype three, IbCBP haplotype one and haplotype two) in comparison to susceptible and unknown cultivars (Resisto and D-3), and to the WT diploid species. Different haplotypes are represented by the forward slash /. Columns in gray indicate commercial varieties and amino acid variations among varieties are indicated in red.

commercial varieties in *IbCBP* position 36 in which one of the haplotypes contained a valine (V) in comparison to the reference sequences and other cultivars possessing an isoleucine (I) residue. Both I and V have branched-chain amino acids that have not been documented in CBPs of other plant species. Currently, there is limited information on the structural composition of the divergent initiation factor CBP. However, we predicate that these variations can play a role in protein-protein binding affinity. Further work needs to be performed to establish the role of this mutation among the cultivars. In the unknown and susceptible cultivars an in-frame deletion in PI-318846 *IbeIF*(*iso*)4*E* haplotype three, and single amino acid variations in PI-318846 *IbCBP* haplotype one and Resisto haplotype one were detected (**Table 4**). Similar to our postulation regarding the unique amino acid variation in *IbCBP* within the commercialized varieties, it is suspected that these amino acid variations contribute to protein-protein interactions resulting in virus susceptibility. For example, German-Retana *et al*. [77] reported eIF4E mutations located in the highly conserved tryptophan residues interfered with cap binding capabilities.

Initially, the amino acid variations were compared to the *I. trifida* genome for assessment; however, further comparison of these amino acid changes against the *I. triloba* genome revealed variations between *I. trifida* and *I. triloba* results. Previously, it was indicated that both *I. trifida* and *I. triloba* were the closest relatives of *Ipomoea batatas* according to morphological and taxonomic observations performed by Austin, 1988 [73]. As research progresses, it has been documented that our understanding of the origin and reorganization of the hexaploid sweetpotato genome is restricted [1] [56] [57] [59]. One might question if these variations and the different homeoalles contribute to a degree of yearly/ periodic resistance to SPFMV, based on previous work indicating how homeoalleles may potentially contribute to differentially regulating mRNA translation during development under stress [18] [27] [35]. As published earlier from experiments in several plants, individual knockout, and downregulation of these proteins had no negative effect on *Arabidopsis* and *Nicotiana* growth and development [5] [13] [18] [27] [35]. Traditionally, SNPs have been characterized to give rise to phenotypic traits such as pathogen resistance or differences in allelic forms; however, SNPs that do not translate to altered protein sequences are often overlooked [83]. Based on our nucleotide sequence analysis, we postulate that the SNPs identified amongst the four varieties may contribute to mRNA folding which can result in differences in viral protein and host interaction responses. The findings from the previous work performed in other plant species provided more insight into how the amino acid variations from the allelic forms discovered in this study, against the two reference genomes, compare amongst the commercial (BRG, JWL) varieties and other cultivars (D3 and Resisto) in avoiding SPFMV symptom development post-infection [5] [7] [9] [10] [13] [53] [55] [84] [85].

One suggestion for studying the allelic variations from this work may be to utilize them as an outline for conserved sequence detection to design gRNAs for

CRISPR editing of the eIF4E family in the complex hexaploid sweetpotato, which has been performed in other model organisms [26] [27] [86]. Natural genetic variants have been associated with physiological characteristics, such as disease resistance [32] [87]. In the last decade, eukaryotic translation initiation factors 4E, iso4E, and CBP have been cloned and characterized, which have been documented to play roles in recessive gene-mediated defense in other plant species [14] [22] [30] [44] [45]. For example, natural amino acid substitutions in 4E, iso4E, or CBP are shown to confer resistance to potyviruses in the following plant species: *Capsicum annuum*, *Hordeum vulgare*, *Pisum sativum*, *Oryza sativa*, and *Cucumis melo* [22] [37] [45] [87] [88] [89]. Since most commercial sweetpotato cultivars such as BRG and JWL are described as periodic resistant to SPFMV due to the mild or absent symptom development after infection in addition to previous findings [7] [54]; therefore, the results garnered in **Table 4** may shed light on the correlation between resistance and susceptible of the cultivars tested. While we have surveyed natural variations amongst sweetpotato cultivars with tolerant, susceptible, and unknown responses to SPFMV, it is unclear how these SNPs identified are associated with phenotypic disease responses. Further work will be necessary to comprehend how these variations function at the molecular level.

6. Conclusion

This work reports the isolation and characterization of the *eIF4E* gene subfamily from four hexaploid sweetpotato cultivars (BRG, JWL, Res, and D3) annotated as *IbeIF4E*, *IbeIF*(*iso*)*4E*, and *IbCBP* with multiple allelic forms identified when compared against the diploid reference genome *I. trifida* (http://sweetpotato.uga.edu/). Our data on *IbeIF4E* revealed four haplotypes in Resisto, two haplotypes in D-3 and Jewel, and one haplotype in Beauregard. While *IbeIF*(*iso*)*4E* sequencing assessment demonstrated the presence of three haplotypes in cultivars Resisto, D-3 and Beauregard, and two haplotypes in Jewel. We also determined *IbCBP* had two haplotypes within each of the four cultivars. It could be inferred from the above findings on the various haplotypes, that genetic diversity among resistant, susceptible and unknown sweetpotato cultivars exists in their response to infection. All open reading frames were in the length of 696 bp *IbeIF4E*, 606 bp *IbeIF*(*iso*)*4E*, and 675 bp *IbCBP*, each encoding a single polypeptide chain of 232, 202, and 225 amino acids, respectively. The gene structures inferred from these results showed that both *IbeIF4E* and *IbeIF*(*iso*)*4E* contained five exons and four introns each, while *IbCBP* had six exons and five introns in comparison to *I. trifida* reference genomic sequences. Further sequence mining, protein prediction studies and phylogenetic analysis indicated that *IbeIF4E*, *IbeIF*(*iso*)*4E* and *IbCBP* were highly conserved with other eIF4E proteins previously identified in other plant species, especially the conserved aromatic residues (tryptophan, phenylalanine, and histidine) within each class that is known to interact with the mRNA-cap structure of eIF4Es from other organisms. This research

outcome substantiates the significance of the continued investigation into the function and mechanisms of the eIF4E gene subfamily in hexaploid sweetpotato for the further development of viral-resilient cultivars through genetic and new breeding technologies such as CRISPR. Furthermore, gene editing studies may be performed to reach a conclusion correlating resistance, unknown and susceptible cultivars tested to confirm genetic diversity among host susceptibility factors.

Acknowledgements

This research was supported by USDA-NIFA Grants: 2017-38821-26414-GE and USDA-NIFA-Evans-Allen Grant No. ALX-FVC18 to Tuskegee University Plant Biotech and Genomics Research Laboratory, CAENS, Tuskegee University, AL 36088. The authors wish to acknowledge Dr. Curtis Jolly for the critical edits of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gao, M., Soriano, S.F., Cao, Q., Yang, X. and Lu, G. (2020) Hexaploid Sweetpotato (*Ipomoea batatas* (L.) Lam.) May Not Be a True type to Either Auto- or Allopolyploid. *PLOS ONE*, **15**, e0229624. https://doi.org/10.1371/journal.pone.0229624

[2] Shahbandeh, M. (2022, April 4) Sweet Potato: Worldwide Market Value Forecast 2020-2025. Statista. https://www.statista.com/statistics/1100004/global-sweet-potato-market-size/

[3] USDA-National Agricultural Statistics Service (2021) Vegetables 2020 Summary. https://usda.library.cornell.edu/concern/publications/02870v86p?locale=en

[4] Loebenstein, G. and Katis, N.I. (2015) Advances in Virus Research: Control of Plant Viral Diseases Vegetatively Propagated Crops. *Advances in Virus Research*, **91**, 33-45. https://www.sciencedirect.com/science/article/abs/pii/S0065352714000062 https://doi.org/10.1016/bs.aivir.2014.10.005

[5] Mwanga, R.O.M., Kriegner, A., Cervantes-Flores, J.C., Zhang, D.P., Moyer, J.W. and Yencho, G.C. (2002) Resistance to Sweetpotato Chlorotic Stunt Virus and Sweetpotato Feathery Mottle Virus Is Mediated by Two Separate Recessive Genes in Sweetpotato. *Journal of the American Society for Horticultural Science*, **127**, 798-806. https://doi.org/10.21273/JASHS.127.5.798

[6] Njeru, R.W., Mburu, M.W.K., Cheramgoi, E., Gibson, R.W., Kiburi, Z.M., Obudho, E. and Yobera, D. (2004) Studies on the Physiological Effects of Viruses on Sweetpotato Yield in Kenya. *Annals Applied Biology*, **145**, 71-76. https://doi.org/10.1111/j.1744-7348.2004.tb00360.x

[7] Clark, C.A. and Hoy, M.W. (2006) Effects of Common Viruses on Yield and Quality of Beauregard Sweetpotato in Louisiana. *Plant Disease*, **90**, 83-88. https://doi.org/10.1094/PD-90-0083

[8] Bednarek, R., David, M., Fuentes, S., Kreuze, J. and Fei, Z. (2021) Transcriptome

Analysis Provides Insights into the Responses of Sweetpotato to Sweetpotato Virus Disease (SPVD). *Virus Research*, **295**, Article ID: 198293. https://doi.org/10.1016/j.virusres.2020.198293

[9] Kokkinos, C.D. and Clark, C.A. (2006) Interactionsamong Sweetpotato chlorotic Stunt Virus and Different Potyviruses and Potyvirus Strains Infecting Sweetpotato in the United States. *Plant Disease*, **90**, 1347-1352. https://doi.org/10.1094/PD-90-1347

[10] Karyeija, R.F., Kreuze, J.F., Gibson, R.W. and Valkonen, J.P.T. (2000) Synergistic Interactions of a Potyvirus and a Phloem-Limited Crinivirus in Sweetpotato Plants. *Virology*, **269**, 26-36. https://doi.org/10.1006/viro.1999.0169

[11] Kwak, H.-R., Kim, J., Kim, M.-K., Seo, J.-K., Jung, M.-N., Kim, J.-S., Lee, S., *et al*. (2015) Molecular Characterization of Five Potyviruses Infecting Korean Sweet Potatoes Based on Analyses of Complete Genome Sequences. *The Plant Pathology Journal*, **31**, 388-401. https://doi.org/10.5423/PPJ.OA.04.2015.0072

[12] Wang, H., Wu, Y., Zhang, Y., Yang, J., Fan, W., Zhang, H., Zhao, S., *et al*. (2019) CRISPR/Cas9-Based Mutagenesis of Starch Biosynthetic Genes in Sweet Potato (*Ipomoea batatas*) for the Improvement of Starch Quality. *International Journal of Molecular Sciences*, **20**, Article No. 4702. https://doi.org/10.3390/ijms20194702

[13] Miano, D.W., LaBonte, D.R. and Clark, C.A. (2008) Identification of Molecular Markers Associated with Sweetpotato Resistance to Sweetpotato Virus Disease in Kenya. *Euphytica*, **160**, 15-24. https://doi.org/10.1007/s10681-007-9495-2

[14] Wang, A. and Krishnaswamy, S. (2012) Eukaryotic Translation Initiation Factor 4E-Mediated Recessive Resistance to Plant Viruses and its Utility in Crop Improvement. *Molecular Plant Pathology*, **13**, 795-803. https://doi.org/10.1111/j.1364-3703.2012.00791.x

[15] Maule, A.J., Caranta, C. and Boulton, M.I. (2007) Sources of Natural Resistance to Plant Viruses: Status and Prospects. *Molecular Plant Pathology*, **8**, 223-231. https://doi.org/10.1111/j.1364-3703.2007.00386.x

[16] Studer, S.M. and Joseph, S. (2007) Binding of mRNA to the Bacterial Translation Initiation Complex. *Methods of Enzymology*, **430**, 31-44. https://doi.org/10.1016/S0076-6879(07)30002-5

[17] Prévôt, D., Darlix, J.-L. and Ohlmann, T. (2003) Conducting the Initiation of Protein Synthesis: The Role of eIF4G. *Biology of the Cell*, **95**, 141-156. https://doi.org/10.1016/S0248-4900(03)00031-5

[18] Combe, J.P., Petracek, M.E., van Eldik, G., Meulewaeter, F. and Twell, D. (2005) Translation Initiation Factors eIF4E and eIFiso4E are Required for Polysome Formation and Regulate Plant Growth in Tobacco. *Plant Molecular Biology*, **57**, 749-760. https://doi.org/10.1007/s11103-005-3098-x

[19] Browning, K.S. (1996) The Plant Translational Apparatus. *Plant Molecular Biology*, **32**, 107-144. https://doi.org/10.1007/BF00039380

[20] Browning, K.S. (2004) Plant Translation Initiation Factors: It Is Not Easy to Be Green. *Biochemical Society Transactions*, **32**, 589-591. https://doi.org/10.1042/BST0320589

[21] Coutinho de Oliveira, L., Volpon, L., Rahardjo, A.K., Osborne, M.J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M., Kwok H.B. and Borden L.B.K. (2019) Structural Studies of the eIF4E-VPg Complex Reveal a Direct Competition for Capped RNA: Implications for Translation. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 24056-24065. https://doi.org/10.1073/pnas.1904752116

[22] Charron, C., Nicolaï, M., Gallois, J.-L., Robaglia, C., Moury, B., Palloix, A. and Caranta, C. (2008) Natural Variation and Functional Analyses Provide Evidence for Co-Evolution between Plant eIF4E and Potyviral VPg. *The Plant Journal*, **54**, 56-68. https://doi.org/10.1111/j.1365-313X.2008.03407.x

[23] Rhoads, R.E. (2009) eIF4E: New Family Members, New Binding Partners, New Roles. *Journal of Biological Chemistry*, **284**, 16711-16715. https://doi.org/10.1074/jbc.R900002200

[24] Dutt, S., Parkash, J., Mehra, R., Sharma, N., Singh, B., Raigond, P., Joshi, A., Chopra S. and Singh P.B. (2015) Translation Initiation in Plants: Roles and Implications beyond Protein Synthesis. *Biologia Plantarum Plant*, **59**, 401-412. https://doi.org/10.1007/s10535-015-0517-y

[25] Jackson, R.J., Hellen, C.U.T. and Pestova, T.V. (2010) The Mechanism of Eukaryotic Translation Initiation and Principles of Its Regulation. *Nature Reviews Molecular Cell Biology*, **11**, 113-127. https://doi.org/10.1038/nrm2838

[26] Shi, S., Zhang, X., Mandel, M.A., Zhang, P., Zhang, Y., Ferguson, M., Amuge, T., Rounsley S., Liu Z. and Xiong, Z. (2017) Variations of Five eIF4E Genes across Cassava Accessions Exhibiting Tolerant and Susceptible Responses to Cassava Brown Streak Disease. *PLOS ONE*, **12**, e0181998. https://doi.org/10.1371/journal.pone.0181998

[27] Yang, Z., Dong, M., Cheng, G., Liu, S., Zhang, H., Shang, H., Zhou, Y., Huang, G., Zhang, M., Wang, F. and Xu, J. (2021) Selective Interaction of Sugarcane eIF4E with VPgs from Sugarcane Mosaic Pathogens. *Viruses*, **13**, Article No. 518. https://doi.org/10.3390/v13030518

[28] Jiang, J. and Laliberté, J.-F. (2011) The Genome-Linked Protein VPg of Plant Viruses-A Protein with many Partners. *Current Opinion in Virology*, **1**, 347-354. https://doi.org/10.1016/j.coviro.2011.09.010

[29] Yadav, S. and Chhibbar, A. (2016) Plant-Virus Interactions Molecular Aspects of Plant-Pathogen Interaction. Springer, Cham. https://doi.org/10.1007/978-3-319-25489-0

[30] Bastet, A., Robaglia, C. and Gallois, J.-L. (2017) eIF4E Resistance: Natural Variation Should Guide Gene Editing. *Trends in Plant Science*, **22**, 411-419. https://doi.org/10.1016/j.tplants.2017.01.008

[31] Khan, M.A. and Goss, D.J. (2019) Poly (A) Binding Protein Enhances the Binding Affinity of Potyvirus VPg to Eukaryotic Initiation Factor eIF4F and Activates In-Vitro Translation. *International Journal of Biological Macromolecules*, **121**, 947-955. https://doi.org/10.1016/j.ijbiomac.2018.10.135

[32] Truniger, V. and Aranda, M.A. (2009) Recessive Resistance to Plant Viruses. *Advances in Virus Research*, **75**, 119-159+231. https://doi.org/10.1016/S0065-3527(09)07504-6

[33] Kanyuka, K., Druka, A., Caldwell, D.G., Tymon, A., Mccallum, N., Waugh, R. and Adams, M.J. (2005) Evidence that the Recessive Bymovirus Resistance Locus Rym4 in Barley Corresponds to the Eukaryotic Translation Initiation Factor 4E Gene. *Molecular Plant Pathology*, **6**, 449-458. https://doi.org/10.1111/j.1364-3703.2005.00294.x

[34] Yoshii, M., Shimizu, T., Yamazaki, M., Higashi, T., Miyao, A., Hirochika, H. and Omura, T. (2009) Disruption of a Novel Gene for a NAC-Domain Protein Rice Confers Resistance to Rice Dwarf Virus. *The Plant Journal*, **57**, 615-625. https://doi.org/10.1111/j.1365-313X.2008.03712.x

[35] Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K.S. and Robaglia, C. (2002)

The Arabidopsis Eukaryotic Initiation Factor (iso)4E Is Dispensable for Plant Growth but Required for Susceptibility to Potyviruses. *The Plant Journal*, **32**, 927-934. https://doi.org/10.1046/j.1365-313X.2002.01481.x

[36] Lellis, A.D., Kasschau, K.D., Whitham, S.A. and Carrington, J.C. (2002) Loss-of-Susceptibility Mutants of Arabidopsis Thaliana Reveal an Essential Role for eIF(iso)4E during Potyvirus Infection. *Current Biology*, **12**, 1046-1051. https://doi.org/10.1016/S0960-9822(02)00898-9

[37] Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta C. and LeGall O. (2003) The Eukaryotic Translation Initiation Factor 4E Controls Lettuce Susceptibility to the Potyvirus Lettuce mosaic virus. *Plant Physiology*, **132**, 1272-1282. https://doi.org/10.1104/pp.102.017855

[38] Ruffel, Sandrine, Dussault, M.-H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C. (2002) A Natural Recessive Resistance Gene against Potato Virus Y in Pepper Corresponds to the Eukaryotic Initiation Factor 4E. *The Plant Journal*, **4**, 1067-1075. https://doi.org/10.1046/j.1365-313X.2002.01499.x

[39] Sato, M., Nakahara, K., Yoshii, M., Ishikawa, M. and Uyeda, I. (2005) Selective Involvement of Members of the Eukaryotic Initiation Factor 4E Family in the Infection of Arabidopsis thaliana by Potyviruses. *Federation of European Biochemical Societies (FEBS) Letters*, **579**, 1167-1171. https://doi.org/10.1016/j.febslet.2004.12.086

[40] Yoshii, M., Nishikiori, M., Tomita, K., Yoshioka, N., Kozuka, R., Naito, S. and Ishikawa, M. (2004) The Arabidopsis Cucumovirus Multiplication 1 and 2 Loci Encode Translation Initiation Factors 4E and 4G. *Journal of Virology*, **78**, 6102-6111. https://doi.org/10.1128/JVI.78.12.6102-6111.2004

[41] Zaidi, S.S.-A., Mukhtar, M.S. and Mansoor, S. (2018) Genome Editing: Targeting Susceptibility Genes for Plant Disease Resistance. *Trends in Biotechnology*, **36**, 898-906. https://doi.org/10.1016/j.tibtech.2018.04.005

[42] Sadowy, E., Miłner, M. and Haenni, A.-L. (2001) Proteins Attached to Viral Genomes Are Multifunctional. *Advances in Virus Research*, **57**, 185-262. https://doi.org/10.1016/S0065-3527(01)57004-9

[43] Mazier, M., Flamain, F., Nicolaï, M., Sarnette, V. and Caranta, C. (2011) Knock-Down of Both *eIF4E1* and *eIF4E2* Genes Confers Broad-Spectrum Resistance against Potyviruses in Tomato. *PLOS ONE*, **6**, e29595. https://doi.org/10.1371/journal.pone.0029595

[44] Ruffel, S., Gallois, J.L., Lesage, M.L. and Caranta, C. (2005) The Recessive Potyvirus Resistance Gene Pot-1 Is the Tomato Orthologue of the Pepper pvr2-eIF4E Gene. *Molecular Genetics and Genomics*, **274**, 346-353. https://doi.org/10.1007/s00438-005-0003-x

[45] Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., Puigdomènech, P., Pitrat, M., Caboche, M., Dogimont, C., Garcia-Mas, J., Aranda A.M. and Bendahmane A. (2006) An eIF4E Allele Confers Resistance to an Uncapped and Non-Polyadenylated RNA Virus in Melon. *The Plant Journal*, **48**, 452-462. https://doi.org/10.1111/j.1365-313X.2006.02885.x

[46] Zhang, Y.-Y., Qi, M.-F., Sun, J., Zhang, X.-H., Shi, H.-L., Li, H.-X. and Ye, Z.-B. (2009) Molecular Cloning and Characterization of a Gene Encoding Eukaryotic Initiation Factor iso4E in Tomato (*Solanum lycopersicum*). *Plant Molecular Biology Reporter*, **27**, 400-406. https://doi.org/10.1007/s11105-009-0090-7

[47] Sanfaçon, H. (2015) Plant Translation Factors and Virus Resistance. *Viruses*, **7**, 3392-

3419. https://doi.org/10.3390/v7072778

[48] Poulicard, N., Pacios, L.F., Gallois, J.-L., Piñero, D. and García-Arenal, F. (2016) Human Management of a Wild Plant Modulates the Evolutionary Dynamics of a Gene Determining Recessive Resistance to Virus Infection. *Public Library of Science Genetics*, **12**, Article ID: 1006214. https://doi.org/10.1371/journal.pgen.1006214

[49] Shopan, J., Liu, C., Hu, Z., Zhang, M. and Yang, J. (2020) Identification of Eukaryotic Translation Initiation Factors and the Temperature-Dependent Nature of Turnip mosaic virus Epidemics in Allopolyploid Brassica juncea. 3 *Biotech*, **10**, Article No. 75. https://doi.org/10.1007/s13205-020-2058-0

[50] Tugume, A.K., Cuellar, W.J., Muskasa, S.B. and Valkonen, J.P.T. (2010) Molecular Genetic Analysis of Virus Isolates from Wild and Cultivated Plants Demonstrates that East Africa is a Hotspot for the Evolution and Diversification of Sweetpotato Feathery Mottle Virus. *Molecular Ecology*, **19**, 3139-3156. https://doi.org/10.1111/j.1365-294X.2010.04682.x

[51] Okada, Y., Kobayashi, A., Tabuchi, H. and Kuranouchi, T. (2017) Review of Major Sweetpotato Pests in Japan, with Information on Resistance Breeding Programs. *Breeding Science*, **67**, 73-82. https://doi.org/10.1270/jsbbs.16145

[52] Adikini, S., Mukasa, S.B., Mwanga, R.O.M. and Gibson, R.W. (2016) Effects of Sweetpotato Feathery Mottle Virus and Sweetpotato Chlorotic Stunt Virus on the Yield of Sweet Potato in Uganda. *Journal of Phytopathology*, **164**, 242-254. https://doi.org/10.1111/jph.12451

[53] Karyeija, R.F., Gibson, R.W. and Valkonen, J.P.T. (1998) The Significance of Sweetpotato Feathery Mottle Virus in Subsistence Sweetpotato Production in Africa. *Plant Disease*, **82**, 4-15. https://doi.org/10.1094/PDIS.1998.82.1.4

[54] Gutie'rrez, D.L., Fuentes, S. and Salazar, L.F. (2003) Sweetpotato Virus Disease (SPVD): Distribution, Incidence, and Effect on Sweetpotato Yield in Peru. *Plant Disease*, **87**, 297-302. https://doi.org/10.1094/PDIS.2003.87.3.297

[55] Léonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G. and Laliberté, J.-F. (2000) Complex Formation between Potyvirus VPg and Translation Eukaryotic Initiation Factor 4E Correlates with Virus Infectivity. *Journal of Virology*, **74**, 7730-7737. https://doi.org/10.1128/JVI.74.17.7730-7737.2000

[56] McGregor, C.E., Miano, D.W., LaBonte, D.R., Hoy, M., Clark, C.A. and Rosa, G.J.M. (2009) Differential Gene Expression of Resistant and Susceptible Sweetpotato Plants after Infection with the Causal Agents of Sweetpotato Virus Disease. *Journal of American Society for Horticulture Science*, **134**, 658-666. https://doi.org/10.21273/JASHS.134.6.658

[57] Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V.M., Rossel, G., Tay, D., McKey, D. and Lebot, V. (2013) Disentangling the Origins of Cultivated Sweetpotato (*Ipomoea batatas* (L.) Lam.). *PLOS ONE*, **8**, e62707. https://doi.org/10.1371/journal.pone.0062707

[58] Kyndta, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., Gheysen, G., Kreuze F.J. (2015) The Genome of Cultivated Sweetpotato Contains Agrobacterium T-DNAs with Expressed Genes: An Example of a Naturally Transgenic Food Crop. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5844-5849. https://doi.org/10.1073/pnas.1419685112

[59] Wu, S., Lau, K.H., Cao, Q., Hamilton, J.P., Sun, H., Zhou, C., Eserman, L., Gemenet, C.D., Olukolu, A.B., Wang, H., Crisovan, E., Godden, T.G., Jiao, C., Wang, X., Kitavi, M., Manrique-Carpintero, N., Vaillancourt, B., Wiegert-Rininger, K., Yang, X., Bao K., Schaff, J., Kreuze J., Gruneberg, W., Khan, A., Ghislain, M., Ma, D., Jiang.J.,

Mwanga M.O.R., Leebens-Mack, J., Coin, M.J.L., Yencho G.C., Buell C.R. and Zhangun, F. (2018) Genome Sequences of Two Diploid Wild Relatives of Cultivated Sweetpotato Reveal Targets for Genetic Improvement. *Nature Communications*, **9**, Article No. 4580. https://doi.org/10.1038/s41467-018-06983-8

[60] Comai, L. (2005) The Advantages and Disadvantages of Being Polyploid. *Nature Reviews Genetics*, **6**, 836-846. https://doi.org/10.1038/nrg1711

[61] Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D. and Strömvik, M.V. (2018) Current Strategies of Polyploid Plant Genome Sequence Assembly. *Frontiers in Plant Science*, **9**, Article No. 1660. https://doi.org/10.3389/fpls.2018.01660

[62] Clevenger, J., Chavarro, C., Pearl, S.A., Ozias-Akins, P. and Jackson, S.A. (2015) Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Molecular Plant*, **8**, 831-846. https://doi.org/10.1016/j.molp.2015.02.002

[63] Egnin, Marceline, Mora, A. and Prakash, C.S. (1998) Factors Enhancing Agrobacterium tumefaciens-Mediated Gene Transfer in Peanut (*Arachis hypogaea* L.). *In Vitro Cellular & Developmental Biology-Plant*, **34**, 310-318. https://doi.org/10.1007/BF02822740

[64] Egnin, M., Walker, M., Prakash C.S. and Jaynes J. (2002) Transgenic 'High Protein' Sweetpotatoes (*Ipomoea batatas* L., PI-318846-3) Engineered with an Artificial Storage Protein Gene (asp-1) Alter the Temporal Distribution/ Accumulation of Sporamin and B-amylase. *In Vitro Cellular and Developmental Biology-Plant*, **38**, 56A.

[65] Chang, C.-T., Tsai, C.-N., Tang, C.Y., Chen, C.-H., Lian, J.-H., Hu, C.-Y., Tsai, C.-L., Chao, A., Lai C-H., Wang, T-H. and Lee, Y-S. (2012) Mixed Sequence Reader: A Program for Analyzing DNA Sequences with Heterozygous Base Calling. *The Scientific World Journal*, **2012**, 1-10. https://doi.org/10.1100/2012/365104

[66] Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870-1874. https://doi.org/10.1093/molbev/msw054

[67] Corpet, F. (1988) Multiple Sequence Alignment with Hierarchical Clustering. *Nucleic Acids Research*, **16**, 10881-10890. https://doi.org/10.1093/nar/16.22.10881

[68] Gouet, P., Courcelle, E., Stuart, D. and Metoz, F. (1999) ESPript: Analysis of Multiple Sequence Alignments in PostScript. *Bioinformatics*, **15**, 305-308. https://doi.org/10.1093/bioinformatics/15.4.305

[69] Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., Rokhsar, S.D. and Rounsley, S. (2012) The Cassava Genome: Current Progress, Future Directions. *Tropical Plant Biology*, **5**, 88-94. https://doi.org/10.1007/s12042-011-9088-z

[70] Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, Y.I., Egesi, C., Nauluvula, P., Lebot, V., Ndunguru, J., Makamilo, G., Bart, S.R., Setter, L.T., Gleadow, M.R., Kulakow, P., Ferguson, E.M., Rounsley, S. and Rokhsar, S.D. (2016) Sequencing Wild and Cultivated Cassava and Related Species Reveals Extensive Interspecific Hybridization and Genetic Diversity. *Nature Biotechnology*, **34**, 562-570. https://doi.org/10.1038/nbt.3535

[71] Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire K.M., Geer C.R., Gonzales, R.N., Gwadz, M., Hurwitz, I.D., Lu, F., Marchler, H.G., Song, S.J., Thanki, N., Wang, Z., Yamashita, A.R., Zhang, D., Zheng, C., Geer, Y L. and Bryant H.S. (2017) CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res*, **45**, 200-203.

https://doi.org/10.1093/nar/gkw1129

[72] Kommedal, Ø., Karlsen, B. and Sæbø, Ø. (2008) Analysis of Mixed Sequencing Chromatograms and It's Application in Direct 16S rRNA Gene Sequencing of Polymicrobial Samples. *Journal of Clinical Microbiology*, **46**, 3766-3771. https://doi.org/10.1128/JCM.00213-08

[73] Austin, D.F. (1988) The Taxonomy, Evolution and Genetic Diversity of Sweetpotato and Related Wild Species. In: Gregory, P., Ed., *Exploration, Maintenance and Utilization of Sweetpotato Genetic Resources*, International Potato Center, Lima, 27-60.

[74] Altmann, M., Edery, I., Trachsel, H. and Sonenberg, N. (1988) Site-Directed Mutagenesis of the Tryptophan Residues in Yeast Eukaryotic Initiation Factor 4E Effects on Cap Binding Activity. *Journal of Biological Chemistry*, **263**, 17229-17232. https://doi.org/10.1016/S0021-9258(19)77821-1

[75] Morino, S., Hazama, H., Ozaki, M., Teraoka, Y., Shibata, S., Doi, M., Ueda, H., Ishida T. and Uesugi S. (1996) Analysis of the mRNA Cap-Binding Ability of Human Eukaryotic Initiation Factor-4E by Use of Recombinant Wild-Type and Mutant Forms. *European Journal of Biochemistry*, **239**, 597-601. https://doi.org/10.1111/j.1432-1033.1996.0597u.x

[76] Monzingo, A.F., Dhaliwal, S., Dutt-Chaudhuri, A., Lyon, A., Sadow, J.H., Hoffman, D.W., Robertus, J.D., *et al*. (2007) The Structure of Eukaryotic Translation Initiation Factor-4E from Wheat Reveals a Novel Disulfide Bond. *Plant Physiology*, **143**, 1504-1518. https://doi.org/10.1104/pp.106.093146

[77] German-Retana, S., Walter, J., Doublet, B., Roudet-Tavert, G., Nicaise, V., Lecampion, C., Houvenaghel, M.-C., *et al*. (2008) Mutational Analysis of Plant Cap-Binding Protein eIF4E Reveals Key Amino Acids Involved in Biochemical Functions and Potyvirus Infection. *Journal of Virology*, **82**, 7601-7612. https://doi.org/10.1128/JVI.00209-08

[78] Joshi, B., Lee, K., Maeder, D.L. and Jagus, R. (2005) Phylogenetic Analysis of eIF4E-Family Members. *BioMed Central* (*BMC*) *Ecology and Evolution*, **5**, 1-20. https://doi.org/10.1186/1471-2148-5-48

[79] Mayberry, Laura K., Allen, M.L., Nitka, K.R., Campbell, L., Murphy, P.A. and Browning, K.S. (2011) Plant Cap-Binding Complexes Eukaryotic Initiation Factors eIF4F and eIFISO4F. *Journal of Biological Chemistry*, **286**, 42566-42574. https://doi.org/10.1074/jbc.M111.280099

[80] Hjulsager, C.K., Olsen, B.S., Jensen, D.M.K., Cordea, M.I., Krath, B.N., Johansen, I.E. and Lund, O.S. (2006) Multiple Determinants in the Coding Region of *Pea Seed-Borne Mosaic Virus* P3 Are Involved in Virulence against *sbm*-2 Resistance. *Virology*, **355**, 52-61. https://doi.org/10.1016/j.virol.2006.07.016

[81] Ruud, K.A., Kuhlow, C., Goss, D.J. and Browning, K.S. (1998) Identification and Characterization of a Novel Cap-Binding Protein from Arabidopsis Thaliana. *Journal of Biological Chemistry*, **273**, 10325-10330. https://doi.org/10.1074/jbc.273.17.10325

[82] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The Rapid Generation of Mutation Data Matrices from Protein Sequences. *Bioinformatics*, **8**, 275-282. https://doi.org/10.1093/bioinformatics/8.3.275

[83] Shen, L.X., Basilion, J.P. and Stanton, V.P. (1999) Single-Nucleotide Polymorphisms can Cause Different Structural Folds of mRNA. *Proceedings of the National Academy of Science of the United States of America*, **96**, 7871-7876. https://doi.org/10.1073/pnas.96.14.7871

[84] Galloway, A. and Cowling, V.H. (2019) mRNA Cap Regulation in Mammalian Cell

Function and Fate. *Biochimica et Biophysica Acta* (*BBA*)—*Gene Regulatory Mechanisms*, **1862**, 270-279. https://doi.org/10.1016/j.bbagrm.2018.09.011

[85] Schmitt-Keichinger, C. (2019) Manipulating Cellular Factors to Combat Viruses: A Case Study from the Plant Eukaryotic Translation Initiation Factors eIF4. *Frontiers in Microbiology*, **10**, Article No. 17. https://doi.org/10.3389/fmicb.2019.00017

[86] Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T. and Gal-On, A. (2016) Development of Broad Virus Resistance in Non-Transgenic Cucumber Using CRISPR/Cas9 Technology. *Molecular Plant Pathology*, **17**, 1140-1153. https://doi.org/10.1111/mpp.12375

[87] Richard, M., Chuffart, F., Duplus-Bottin, H., Pouyet, F., Spichty, M., Fulcrand, E., Entrevan, M., Barthelaix, A., Springer, M., Jost, D. and Yvert, G. (2018) Assigning Function to Natural Allelic Variation via Dynamic Modeling of Gene Network Induction. *Molecular Systems Biology*, **14**, e7803. https://doi.org/10.1101/140467

[88] Kang, B.-C., Yeam, I., Frantz, J.D., Murphy, J.F. and Jahn, M.M. (2005) The pvr1 Locus in Capsicum Encodes a Translation Initiation Factor eIF4E That Interacts with Tobacco Etch Virus VPg. *The Plant Journal*, **42**, 392-405. https://doi.org/10.1111/j.1365-313X.2005.02381.x

[89] Stein, N., Perovic, D., Kumlehn, J., Pellio, B., Stracke, S., Streng, S., Ordon, F., *et al*. (2005) The Eukaryotic Translation Initiation Factor 4E Confers Multiallelic Recessive Bymovirus Resistance in *Hordeum vulgare* (L.). *The Plant Journal*, **42**, 912-922. https://doi.org/10.1111/j.1365-313X.2005.02424.x

Supplemental Data Figures

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	itf10g18510.t1eIF4EREF		98.53	100.00	100.00	100.00	99.51	99.51	72.55	82.84	78.92	96.57	81.86	80.88	76.47	74.02	82.35	82.35	74.02
B	itb10g18850.t1_eIF4EREF	0.01		98.53	98.53	98.53	99.02	99.02	72.06	82.35	78.92	96.57	81.86	80.88	76.96	74.02	81.86	81.86	74.02
C	resis_eif_F1	0.00	0.01		100.00	100.00	99.51	99.51	72.55	82.84	78.92	96.57	81.86	80.88	76.47	74.02	82.35	82.35	74.02
D	BRG_eif	0.00	0.01	0.00		100.00	99.51	99.51	72.55	82.84	78.92	96.57	81.86	80.88	76.47	74.02	82.35	82.35	74.02
E	D_3_eif	0.00	0.01	0.00	0.00		99.51	99.51	72.55	82.84	78.92	96.57	81.86	80.88	76.47	74.02	82.35	82.35	74.02
F	jew_eif_F1	0.00	0.01	0.00	0.00	0.00		100.00	72.06	82.35	78.92	97.06	82.35	80.88	76.47	74.02	81.86	81.86	74.02
G	jew_eif_F2	0.00	0.01	0.00	0.00	0.00	0.00		72.06	82.35	78.92	97.06	82.35	80.88	76.47	74.02	81.86	81.86	74.02
H	AT_NP_193538.1	0.27	0.28	0.27	0.27	0.27	0.28	0.28		70.10	74.02	73.04	69.12	75.49	69.61	76.47	69.12	69.12	67.65
I	CA_XP_016557968.1	0.17	0.18	0.17	0.17	0.17	0.18	0.18	0.30		77.94	81.86	82.84	77.94	74.51	74.02	96.08	95.10	70.10
J	ME_XP_021598522.1	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.26	0.22		77.94	79.41	82.35	75.49	76.47	77.45	77.45	70.59
K	IN_XP_019169243.1	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.27	0.18	0.22		82.84	80.39	76.96	74.51	81.37	81.37	74.51
L	CC_CDP14577.1	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.31	0.17	0.21	0.17		79.41	75.00	74.02	83.82	83.82	72.06
M	GM_KAG4956073.1	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.25	0.22	0.18	0.20	0.21		75.98	83.82	77.94	77.94	74.02
N	OS_NP_001359122.1	0.24	0.23	0.24	0.24	0.24	0.24	0.24	0.30	0.25	0.25	0.23	0.25	0.24		71.08	74.51	74.51	84.31
O	PS_AHV79422.1	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.24	0.26	0.24	0.25	0.26	0.16	0.29		73.53	73.53	67.65
P	SL_NP_001307578.1	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.31	0.04	0.23	0.19	0.16	0.22	0.25	0.26		99.02	70.10
Q	ST_QQP16443.1	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.31	0.05	0.23	0.19	0.16	0.22	0.25	0.26	0.01		70.10
R	TA_CAA78262.2	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.32	0.30	0.29	0.25	0.28	0.26	0.16	0.32	0.30	0.30	

Figure S1. Percent identity matrix of IbeIF4E sequences with eIF4E sequences from other plant species.

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	itf04g08430.t1|REF		100.00	100.00	100.00	99.46	100.00	99.46	99.46	100.00	69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
B	ITB04g09410_REF	0.00		100.00	100.00	99.46	100.00	99.46	99.46	100.00	69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
C	Resis_isoF1	0.00	0.00		100.00	99.46	100.00	99.46	99.46	100.00	69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
D	BRG_isoF1	0.00	0.00	0.00		99.46	100.00	99.46	99.46	100.00	69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
E	BRG_isoF2	0.01	0.01	0.01	0.01		99.46	98.92	98.92	99.46	69.35	79.57	81.18	95.16	78.49	73.12	75.27	73.12	77.96	78.49	73.12
F	Jewel_isoF1	0.00	0.00	0.00	0.00	0.01		99.46	99.46	100.00	69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
G	Jewel_isoF2	0.01	0.01	0.01	0.01	0.01	0.01		98.92	99.46	69.35	79.57	81.18	95.16	78.49	73.12	75.27	73.12	77.96	78.49	73.12
H	D-3_isoF1	0.01	0.01	0.01	0.01	0.01	0.01	0.01		99.46	69.35	80.11	81.72	95.16	78.49	73.12	75.27	73.12	78.49	79.03	73.12
I	D-3_isoF2	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01		69.89	80.11	81.72	95.70	79.03	73.66	75.81	73.66	78.49	79.03	73.66
J	AT_NP_001332369.1	0.30	0.30	0.30	0.30	0.31	0.30	0.31	0.31	0.30		69.89	70.97	68.82	70.97	67.20	66.13	64.52	68.82	69.89	64.52
K	CA_NP_001311631.1	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.30		80.65	82.26	79.03	73.66	77.42	74.19	92.47	94.62	72.04
L	ME_XP_021606525.1	0.18	0.18	0.18	0.18	0.19	0.18	0.19	0.18	0.18	0.29	0.19		81.72	77.96	73.12	77.96	73.12	79.03	79.57	74.19
M	IN_XP_019181316.1	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.04	0.31	0.18	0.18		79.57	72.58	76.88	73.12	79.03	80.65	73.12
N	CC_CDO99285.1	0.21	0.21	0.21	0.21	0.22	0.21	0.22	0.22	0.21	0.29	0.21	0.22	0.20		72.58	75.27	68.82	76.34	78.49	71.51
O	OS_XP_015615056.1	0.26	0.26	0.26	0.26	0.27	0.26	0.27	0.27	0.26	0.33	0.26	0.27	0.27	0.27		70.97	67.74	74.73	74.19	77.42
P	GM_NP_001356078.1	0.24	0.24	0.24	0.24	0.25	0.24	0.25	0.25	0.24	0.34	0.23	0.22	0.23	0.25	0.29		77.96	75.81	77.42	70.97
Q	PS_ABH09880.1	0.26	0.26	0.26	0.26	0.27	0.26	0.27	0.27	0.26	0.35	0.26	0.27	0.27	0.31	0.32	0.22		73.12	73.66	68.28
R	SL_QPB75810.1	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.31	0.08	0.21	0.21	0.24	0.25	0.24	0.27		96.77	70.43
S	ST_QQP16447.1	0.21	0.21	0.21	0.21	0.22	0.21	0.22	0.21	0.21	0.30	0.05	0.20	0.19	0.22	0.26	0.23	0.26	0.03		70.97
T	TA_AAA34295.1	0.26	0.26	0.26	0.26	0.27	0.26	0.27	0.27	0.26	0.35	0.28	0.26	0.27	0.28	0.23	0.29	0.32	0.30	0.29	

Figure S2. Percent identity matrix of IbeI(iso)F4E sequences with eIF(iso)4E sequences from other plant species.

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A	itf0108030.t1|		99.08	99.54	99.08	99.54	99.54	99.08	99.54	99.08	99.54	73.85	85.78	80.73	98.62	83.49	75.69	81.19	84.40	84.86
B	itb01g12380.t1|CBP	0.01		99.54	99.08	99.54	99.54	99.08	99.54	99.08	99.54	74.31	86.24	80.73	98.62	83.49	75.69	82.11	84.86	85.32
C	BRGCBPF2	0.00	0.00		99.54	100.00	100.00	99.54	100.00	99.54	100.00	73.85	86.24	80.73	99.08	83.49	75.69	81.65	84.86	85.32
D	BRGCBPF1	0.01	0.01	0.00		99.54	99.54	100.00	99.54	100.00	99.54	73.85	85.78	80.73	98.62	83.03	75.23	81.65	84.40	84.86
E	RESCBPF1	0.00	0.00	0.00	0.00		100.00	99.54	100.00	99.54	100.00	73.85	86.24	80.73	99.08	83.49	75.69	81.65	84.86	85.32
F	RESCBPF2	0.00	0.00	0.00	0.00	0.00		99.54	100.00	99.54	100.00	73.85	86.24	80.73	99.08	83.49	75.69	81.65	84.86	85.32
G	JEWELcbpdpf1	0.01	0.01	0.00	0.00	0.00	0.00		99.54	100.00	99.54	73.85	85.78	80.73	98.62	83.03	75.23	81.65	84.40	84.86
H	JEWELcbpdpf2	0.00	0.00	0.00	0.00	0.00	0.00	0.00		99.54	100.00	73.85	86.24	80.73	99.08	83.49	75.69	81.65	84.86	85.32
I	D-3CBPF1	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00		99.54	73.85	85.78	80.73	98.62	83.03	75.23	81.65	84.40	84.86
J	D-3CBPF2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		73.85	86.24	80.73	99.08	83.49	75.69	81.65	84.86	85.32
K	AT_AAC17220.1	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26		72.94	73.85	73.85	72.94	71.56	76.61	72.02	72.94
L	CA_XP_016548322.1	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.27		80.28	85.32	80.73	72.94	80.73	94.50	94.04
M	ME_XP_021620559.1	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.26	0.20		79.82	84.40	74.77	87.61	79.82	81.19
N	IN_XP_019173242.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.26	0.15	0.20		82.57	76.15	80.73	83.94	84.40
O	CC_CDP14308.1	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.27	0.19	0.16	0.17		76.61	83.94	80.73	82.11
P	OS_XP_015631510.1	0.24	0.24	0.24	0.25	0.24	0.24	0.25	0.24	0.25	0.24	0.28	0.27	0.25	0.24	0.23		76.61	72.94	73.39
Q	GM_NP_001235427.1	0.19	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.23	0.19	0.12	0.19	0.16	0.23		79.36	80.73
R	SL_XP_004249299.1	0.16	0.15	0.15	0.16	0.15	0.15	0.16	0.15	0.16	0.15	0.28	0.06	0.20	0.16	0.19	0.27	0.21		97.71
S	ST_XP_006351360.1	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.27	0.06	0.19	0.16	0.18	0.27	0.19	0.02	

Figure S3. Percent identity matrix of IbCBP sequences with CBP sequences from other plant species.