Genome-wide identification and analysis of the basic leucine zipper (bZIP) transcription factor gene family in *Ustilaginoidea virens*

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0089.R1
Manuscript Type:	Article
Date Submitted by the Author:	01-Jul-2017
Complete List of Authors:	Yin, Weixiao; Huazhong Agriculture University, Department of Plant Protection Cui, Peng; Huazhong Agriculture University, Department of Plant Protection Wei, Wei; Huazhong Agriculture University, Department of Plant Protection Lin, Yang; Huazhong Agriculture University, Department of Plant Protection Luo, Chaoxi; Huazhong Agriculture University
Is the invited manuscript for consideration in a Special Issue? :	This submission is not invited
Keyword:	<i>Ustilaginoidea virens</i>, bZIP, expression analysis, ROS stress

SCHOLARONE™
Manuscripts

1 **Genome-wide identification and analysis of the basic leucine zipper (bZIP) transcription**
2 **factor gene family in *Ustilaginoidea virens***

3 Weixiao Yin, Peng Cui, Wei Wei, Yang Lin and Chaoxi Luo*
4 Department of Plant Protection, College of Plant Science and Technology and the Key Lab of
5 Crop Disease Monitoring & Safety Control in Hubei Province, Huazhong Agricultural
6 University, Wuhan 430070, China;

7
8 wxyin@mail.hzau.edu.cn (Weixiao Yin);
9 1160095043@qq.com (Peng Cui);
10 515292094@qq.com (Wei Wei);
11 yanglin@mail.hzau.edu.cn (Yang Lin);
12 * Correspondence author: cxluo@mail.hzau.edu.cn (Chaoxi Luo); Tel.: +86-27-87281242

Draft

Abstract: The basic leucine zipper (bZIP) transcription factor (TF) family is one of the largest and most diverse TF families widely distributed across the eukaryotes. The bZIP TF family play important roles in growth, development, and response to abiotic or biotic stresses, which have been well characterized in plants, but not in plant pathogenic fungi. In this study, we performed genome-wide and systematic bioinformatics analysis of bZIP genes in the fungus *Ustilaginoidea virens*, the causal agent of rice false smut disease. We identified 28 bZIP family members in the *U. virens* genome by searching for the bZIP domain in predicted genes. The gene structures, motifs and phylogenetic relationships were analyzed for bZIP genes in *U. virens* (*UvbZIPs*). Together with bZIPs from other two fungi, the bZIP genes can be divided into eight groups according to the phylogenetic relationship. Based on RNA-Seq data, the expression profiles of *UvbZIPs* at different infection stages were evaluated. Results showed that 17 of the *UvbZIPs* were up-regulated expression during the infection period. Furthermore, 11 infection-related *UvbZIPs* was investigated under H₂O₂ stress and the expression level of eight genes have changed, which confirmed their roles in stress tolerance and pathogenicity. In summary, our genome wide systematic characterization and expression analysis of *UvbZIPs* provided insights into the molecular functions of these genes in *U. virens* and provides a reference for other pathogens.

Keywords: *Ustilaginoidea virens*; bZIP; expression analysis; ROS stress

65 **Introduction**

66 In multicellular organisms, transcription factors (TFs) play crucial roles in stress responses
 67 and regulate many important developmental processes such as cellular growth and
 68 differentiation. Based on the primary and/or three-dimensional structure similarities in the
 69 DNA-binding and multimerization domains, TFs can be classified into different categories. In
 70 plants, fungi and oomycetes, more than 60 TF families have been identified, e.g., basic
 71 leucine zipper (bZIP) proteins, MADS-box proteins, MYB-like proteins, zinc-finger proteins
 72 and so on (Park et al. 2008; Perez-Rodriguez et al. 2010; Schwechheimer and Bevan 1998).

73 The bZIP family is one of the largest and most diverse TF families, widely distributed
 74 across the eukaryotes (Etxebeste et al. 2008; Fassler et al. 2002; Kong et al. 2015; Tang et al.
 75 2015; Wang et al. 2015b; Ye et al. 2013). The bZIP proteins can be identified based on the
 76 bZIP domain which is 60 to 80 amino acids in length and consist of two functionally distinct
 77 regions, a basic region and a leucine-zipper region (Hurst 1995). The basic region is highly
 78 conserved and consists of approximately 16 amino acid residues with an invariant N-x7-R/K
 79 domain for nuclear localization and sequence-specific DNA binding. The leucine zipper
 80 mediates homo and/or heterodimerization of bZIP proteins and is less conserved, composed
 81 of several heptad repeats of Leu or other bulky hydrophobic amino acids, such as Ile, Val, Phe
 82 or Met, which are located exactly nine amino acid residues toward the C-terminus of the basic
 83 region (Ellenberger et al. 1992; Landschulz et al. 1988; Vinson et al. 1989). In addition to the
 84 bZIP domain, bZIP proteins were found to contain proline-rich, glutamine-rich and acidic
 85 domains, which also are described as being involved in transactivation functions (Liao et al.
 86 2008; Liu et al. 2014; Niu and Gultinan 1994).

87 The bZIP TFs are involved in many critical processes in a diverse range of species. In
 88 *Arabidopsis thaliana*, 75 distinct members of bZIP TFs were identified, which regulate
 89 processes including pathogen defense, light and stress signaling, seed maturation and flower
 90 development (Jakoby et al. 2002). There were 89 bZIP transcription factor-encoding genes in
 91 the rice (*Oryza sativa*) genome (Nijhawan et al. 2008), some of which have been
 92 functionally characterized as involved in seed maturation and germination, floral induction
 93 and development, photomorphogenesis, as well as stress and hormone signaling (Dey et al.
 94 2016; Hossain et al. 2010; Nijhawan et al. 2008; Shimizu et al. 2005; Takahashi et al. 2012;
 95 Zhang et al. 2016). In filamentous fungi or oomycetes, the processes regulated by bZIP
 96 proteins characterized thus can be mainly categorized into development, amino acid
 97 biosynthesis, unfolded protein response, nutrient utilization and various stress responses. In
 98 *Aspergillus* species, the bZIP TF FlbB proteins are required for proper asexual development
 99 and gliotoxin production (Etxebeste et al. 2008; Xiao et al. 2010). In *Fusarium oxysporum*,
 100 the gene *HapX* mediates iron homeostasis that is essential for rhizosphere competence and

virulence (López-Berges et al. 2012), and the gene *MeaB* affects nitrogen metabolite repression (Lopez-Berges et al. 2010). In *Magnaporthe oryzae*, bZIP TFs *MoAp1* and *MoAtf1* are critical in sporulation and pathogenicity (Guo et al. 2011; Guo et al. 2010). Also *MoAtf1* can help the pathogen to overcome reactive oxygen species (ROS)-mediated plant defense by regulating the transcription of laccases and peroxidases (Guo et al. 2010). After these findings, bZIP TFs of *M. oryzae* were functionally characterized for hyphal growth, appressorium formation, nutrient uptake and nitrogen utilization and pathogenicity (Kong et al. 2015; Tang et al. 2015). Some *bZIP* genes have been found to confer resistance to oxidative stress in several fungi, including *Aspergillus* spp. (Asano et al. 2007; Balazs et al. 2010; Hagiwara et al. 2008; Sakamoto et al. 2008), *Ustilago maydis* (Molina and Kahmann 2007), *Botrytis cinerea* (Temme and Tudzynski 2009), and *Neurospora crassa* (Tian et al. 2011) and the oomycete *Phytophthora infestans* (Gamboa-Melendez et al. 2013). Although *bZIP* genes play diverse regulatory roles in fungi or oomycetes, it has been rarely systematically analyzed.

Ustilaginoidea virens (Cooke) Takah (teleomorph: *Villosiclava virens*), a pathogenic ascomycetous fungus that causes rice false smut (RSF), a severe epidemic disease in rice-producing areas worldwide over the past decades (Atia 2004; Bischoff et al. 2004; Rush et al. 2000). Although significant progress has been achieved in research on pathogen morphology (Li et al. 2013; Song et al. 2016; Tang et al. 2013), diversity (Sun et al. 2013; Wang et al. 2014), toxin characterization (Meng et al. 2015) and fungicide resistance (Chen et al. 2013; Wang et al. 2015a) etc., knowledge about the biological functions of *U. virens* proteins is still very limited (Yu et al. 2015). Analysis of the complete genome data of *U. virens* may enhance the systematic identification and analysis for many important proteins such as bZIP TFs. It is anticipated that the characterization of this gene family will provide novel insights into their biological functions.

In this study, we performed a genome-wide systematic characterization of 28 *U. virens* *bZIP* genes (*UvbZIPs*) and found 17 of the *UvbZIP* genes were involved in infection and eight genes involved in ROS stress. This study provided an important starting point to further explore the biological functions of *UvbZIPs* and other TF genes, and increase our understanding for their potential transcriptional regulatory mechanisms in *U. virens*.

Materials and methods

Identification of bZIP transcription factors in *U. virens*

Both local Blast and hidden Markov model (HMM) profile searches were performed against the *U. virens* protein database downloaded from the National Center for Biotechnology Information site (the GeneBank accession code JHTR000000000) (Zhang et al. 2014). The HMM search was performed using an E-value threshold less than or equal to 500.

136 The sequences that appeared multiple times were removed and the most similar ones were
137 kept. The remaining proteins were verified for the presence of the typical bZIP domain by
138 using CD-search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) with an E-value
139 threshold of 0.01. After removing the proteins that do not contain bZIP domain, the remaining
140 sequences were considered as *UvbZIPs* candidates for the further analyses.

141 **Structure of *UvbZIP* genes**

142 To obtain information regarding the intron and exon structures, the positional information
143 of the CDS sequences and the corresponding genomic sequences were loaded into the gene
144 structure display server (GSDS) 2.0 (<http://gsds.cbi.pku.edu.cn/index.php>) (Hu et al. 2015).
145 The status of introns and exons were annotated according to the database. The bZIP domains
146 in each protein were recalculated and given coordinates in CDS sequences, and the
147 corresponding genomic sequences were featured in the gene structure in order to analyze the
148 intron positions and phase within the basic and hinge regions of the bZIP domains.

149 **Phylogenetic analysis**

150 Multiple sequence alignments of the identified bZIP domain amino acids was conducted
151 using the Bioedit program with 1000 bootstrap replicates. Phylogenetic analysis was
152 performed using MEGA5.1 by Maximum Likelihood method (Tamura et al. 2011). The
153 protein sequence databases of *Metarhizium robertsii* were produced by the US Department of
154 Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user
155 community and 31 proteins contains bZIP domain were found. (Gao et al. 2011). The
156 protein database of *M. oryzae* (70-15 MG6) was downloaded from the *M. grisea* Sequencing
157 Project, Broad Institute of Harvard and MIT database (<http://www.broadinstitute.org/>) and 25
158 bZIP protein that have been reported were searched (Tang et al. 2015; Ye et al. 2013).

159 **Inoculation experiments**

160 Inoculation of rice with *U. virens* were done following previous describe (Jia et al. 2015).
161 Briefly, plugs were placed into potato sucrose broth (PSB) and shaken at 160 rpm at 27°C for
162 7 days. The conidial suspension with some hyphal bits was adjusted to 10⁶ conidia per ml.
163 About 2 ml of the conidial suspension in PSB was injected into a single rice panicle from the
164 middle to upper portion at the late booting stage (3-5 days before heading) using a syringe.
165 Inoculated plants were kept in a greenhouse at 27 °C with 90-100% relative humidity.

166 **Analysis of gene expression patterns**

167 Digital gene expression profiling data were obtained from the RNA-Seq data (unpublished
168 data) collected from samples at 3 DPI and 6 DPI. The heat maps were generated using MeV
169 software based on the RPKM (Reads Per Kilo bases per Million reads) data

(www.tm4.org/mev.html). The primers for all genes were designed using Primer Premier 5.0 program and are listed in Table S1. For H₂O₂ treatment, 6-day-old mycelia in potato dextrose broth culture (PDB) were collected and rinsed with sterile water several times, and then were treated with 2 mM H₂O₂ for 0 h, 0.5 h and 1 h. Total RNA was extracted from the mycelia using TaKaRa MiniBEST Plant RNA Extraction Kit (TaKaRa Biotechnology (Dalian) Co., Ltd.). The cDNA was synthesized by using a RevertAid First Strand cDNA Synthesis Kit employing the oligo(dT)₁₈ primer (Thermo Fisher Scientific, Waltham, MA). The reaction mixtures were diluted 20 times with distilled water and used as templates for quantitative real-time PCR. Real-time PCR amplifications of genes were performed in a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories Inc., Hercules, CA) using SYBR Green I fluorescent dye. Amplifications were conducted in a 20 µl volume containing 10 µl of SYBR qPCR Mix (Aidlab, Beijing, China), 1 µl of RT product and 0.2 µM of each primer. All samples were in triplicate with the following profile: an initial preheating at 95 °C for 2 min, followed by 40 cycles at 95 °C for 20 s, 60 °C for 20 s and 72 °C for 20 s. The *α-tubulin* gene served as an internal control for the expression studies, and relative expression was calculated following the comparative Ct(2^{-ΔΔCt}) method (Wong and Medrano 2005).

Results

Identification of the bZIP TF family in *U. virens*

To investigate the bZIP TF family in *U. virens*, the predicted protein set of *U. virens* was downloaded and screened by using the bZIP protein domain sequence as query. In the Pfam database, there are three conserved bZIP domain sequences with domain IDs PF00170 (bZIP_1 domain), PF07716 (bZIP_2) and PF03131 (bZIP_Maf). We obtained 41, 33 and 16 matches, respectively, for each. After combining all the results and removing the duplicates, there remained 47 unique results. To confirm and filter out uncertain bZIP proteins, CD-search (Conserved Domain Database, NCBI) was done, and 25 of them were confirmed to contain the bZIP domain. Based on *U. virens* gene annotations, we also found three additional bZIPs, resulting in 28 bZIP genes were identified from the *U. virens* genome (Table 1), named *UvbZIP1* to *UvbZIP28*. Analysis of the genome of *U. virens* revealed 19 UvbZIP proteins (Zhang et al. 2014), but detailed information about these genes was not supplied. These 19 were represented among the 28, and the additional 9 proteins were successfully identified to contain bZIP domain. The lengths of the predicted bZIP proteins ranged from 121 to 696 aa, with an average of 420 aa.

Conserved domains and motifs in UvbZIP proteins

In the predicted UvbZIP proteins, in addition to bZIP domains, five of these proteins also

204 contained other predicted domains. UvbZIP8 with Aft1 HRA domain, Aft1 OSM domain and
 205 Aft1 HRR domain at the N-terminal of the protein, and UvbZIP11 with PAP1 domain at the
 206 C-terminal. UvbZIP5 contained Rab interacting lysosomal protein-like domain and UvbZIP21
 207 contained Mcm10 replication factor domain, but both domains showed overlap with bZIP
 208 domains. For UvbZIP26, in addition to the carboxypeptidase activation peptide domain,
 209 peptidase M14 carboxypeptidase subfamily A/B-like domain overlapped with the bZIP
 210 domain.

211 The bZIP domain contains a basic region and a leucine-zipper region, and the basic region
 212 of bZIP domain is highly conserved with a N-x7-R/K motif, in which the last amino acid is
 213 Arginine (R) or Lysine (K) (Jakoby et al. 2002). To investigate the characteristics of the
 214 UvbZIP domains, the amino acid sequences were aligned. UvbZIP2 showed less conservation
 215 than other UvbZIPs. UvbZIP22 had the basic region, but without the leucine-zipper region. In
 216 contrast, UvbZIP24 contained the leucine-zipper region without the basic region. Outside of
 217 these exceptions, all of the UvbZIP domains contained the invariant N-x7-R motif, but not the
 218 N-x7-K motif, indicating that Arg was more conserved than Lys in the UvbZIP proteins. For
 219 the Leu zipper, the first heptad repeat Leu was highly conserved, and in three of them, the Leu
 220 was replaced by Met, Ile or Ala. In the other two heptad repeats, variants were frequently
 221 observed (Fig. 1). bZIP proteins also contain some regions, e.g., domains enriched in Pro, Gln
 222 or acidic amino acids. Similarly, the UvbZIP proteins also showed this feature. For example,
 223 the UvbZIP2 contained consecutive Pro or Gln region, while the UvbZIP8 protein, from
 224 residues 218 to 252, more than 70.0% of this region was Pro or Gln.

225 **Intron distribution patterns in *UvbZIPs***

226 The intron/exon arrangement generally carried the imprint of the evolution of some gene
 227 families (Boudet et al. 2001). To gain insights into the structural evolution of the *UvbZIPs*, the
 228 intron distribution in *UvbZIPs* was subjected to alignment analysis using the Gene Structure
 229 Display Server (GSDS), and different intron distribution patterns were observed. As shown in
 230 Fig. 2, all *UvbZIPs* contained introns except for the *UvbZIP20* which contained multiple
 231 repeat sequences. The number of introns ranged from 1 to 4, which varied within the *UvbZIP*
 232 family. The *UvbZIPs* were classified into five classes depending on their intron number from
 233 none to four introns per gene. The most abundant class was single-intron *UvbZIPs* (50.00 %;
 234 14), followed by two introns (32.14 %; 9), four introns (10.71 %; 3), three introns (7.14 %; 2)
 235 and intronless (3.57%; 1). The length of introns was from 24 to 603 base pairs (bp), and more
 236 than half of the introns were between 60 and 150 bp. All introns characteristically contained
 237 the start dinucleotide sequence GT and ended with AG except *UvbZIP9*.

238 The intron positions and phases were diverse and differed from each other. For one gene,
 239 *UvbZIP2*, the intron reading frame or phase was 0, which meant that introns were located

between codons (phase 0). For other genes containing introns, introns were located within codons (phase 1 and 2), but nine of these genes also had phase 0 introns (Fig. 2). The intron patterns within the basic and the hinge region are very important for their functional evolution due to different status of exon splicing in these regions. Therefore, the pattern of intron positions within the basic, hinge and leucine coding sequences regions were analyzed. The results showed that nine *UvbZIPs* contained introns in these regions. All of the introns were inserted in the N-x7-R motif coding regions, and could be divided into five groups according to the position of the introns (Fig. 3).

Phylogenetic analysis of *UvbZIP* proteins

To explore the evolution of the *UvbZIP* genes and relationships with the other fungal bZIPs, the bZIP proteins from other two fungi, *M. oryzae* and *M. robertsii*, were performed and comparing. Together with 25 from *M. oryzae* and 31 from *M. robertsii*, a total 84 bZIP proteins were analyzed and a phylogenetic tree was constructed. All bZIPs were divided into eight clades, and designated as A to H. The *UvbZIPs* distributed to all the clades except clade C which was the minimum and only contained two bZIPs from *M. oryzae* and *M. robertsii*, respectively. None of the bZIPs from *M. oryzae* were contained in clade G, indicating that bZIPs from *U. virens* have closer relationships with the bZIPs from *M. robertsii* than that from *M. oryzae*, which is accord with the current understanding of fungal evolutionary history. There were six clades had bZIPs from *M. oryzae*, *M. robertsii* and *U. virens* which indicating that the evolution of these bZIP genes were conservative between the three fungi. Evolutionary analysis also identified some closely related orthologous bZIPs in three fungi, such as clade C or G etc, indicating that ancestral set of bZIP genes existed prior to the divergence of the three fungi.

Digital expression analysis of *UvbZIPs* at infection stage

In the infection cycle, 3 days post inoculation (DPI) and 6 DPI are important infection stages (Song et al. 2016). To investigate the transcriptional pattern of *UvbZIP* genes in *U. virens* during the process of infection, a RNA-Seq approach was applied to analyze the expression of *UvbZIP* genes at 3 and 6 DPI (Table S2). The results indicated that more fungal genes were involved in the infection at 6 DPI. According to the gene expression profiles, the *UvbZIP* genes can be divided into three groups. Group I exhibited the high expression levels at two time points, e.g., *UvbZIP5*, *UvbZIP6*, *UvbZIP12*, *UvbZIP14*, and *UvbZIP18*. These *UvbZIP* genes might play critical roles at both early and later infection stage. Group II genes exhibited low expression level or no expression at 3 DPI, but higher transcription level at 6 DPI, e.g., *UvbZIP2*, *UvbZIP4*, *UvbZIP7*, *UvbZIP8*, *UvbZIP11*, *UvbZIP13*, *UvbZIP15*,

274 *UvbZIP20*, *UvbZIP21*, *UvbZIP22*, *UvbZIP24* and *UvbZIP25*. These genes might be more
 275 important at the later infection stage. Other genes belonging to Group III exhibited low
 276 expression level or no expression at both time points. These genes might perform functions at
 277 other stages. Overall, expression data implied that 17 *UvbZIPs* may play regulatory roles at
 278 the different stages of the pathogenesis (Fig. 5).

279 ***UvbZIPs* response to H₂O₂ treatment**

280 One of the most rapid plant defense responses against pathogen infection is the oxidative
 281 burst, which constitutes the production of reactive oxygen species (ROS), primarily
 282 superoxide and H₂O₂, at the attempted invasion site of pathogens. To examine *UvbZIPs*
 283 response to ROS, 11 genes expression profiles of infection-induced *UvbZIPs* at 3 and 6 DPI
 284 were examined using real-time PCR. The results showed that compared with 0 h, the
 285 expression level of most selected *UvbZIPs* were up- or down-elevated slightly following H₂O₂
 286 treatment at 0.5 h except *UvbZIP3*, *UvbZIP5*, *UvbZIP6* and *UvbZIP20*. After 1 h H₂O₂
 287 treatment, most of the *UvbZIPs* showed significantly elevated transcript levels except
 288 *UvbZIP3* and *UvbZIP12* which remained the same, *UvbZIP18* showed decreased 0.6 fold
 289 expression. These results indicated that some infection-associated *UvbZIPs* are stimulated by
 290 oxidative stress. Expression level of seven of the *UvbZIPs* at 0.5 h treatment showed weak
 291 stimulation by H₂O₂, while eight of the *UvbZIPs* showed significantly increased expression at
 292 1 h (Fig. 6).

293 **Discussion**

294 With the release of large amounts of genome data, genome-wide identification of
 295 transcription factor families in different species have been carried out, such as *WRKY* (Chen
 296 et al. 2016), *MYB* (Katiyar et al. 2012), *bZIP* (Nijhawan et al. 2008) and so on. The *bZIP*
 297 genes are one of the largest and most diverse TF families across the eukaryotes. There are
 298 many *bZIP* families identified and characterized from various plant species (Hu et al. 2016;
 299 Jakoby et al. 2002; Li et al. 2015a; Li et al. 2015b; Liu and Chu 2015; Nijhawan et al. 2008),
 300 but only a few of them are identified in fungi or oomycetes (Kong et al. 2015; Tang et al.
 301 2015; Ye et al. 2013). Furthermore, some *bZIP* genes may have specific function in different
 302 species, and hence, identification of *bZIP* genes are important for improved understanding of
 303 the infection biology in *U. virens*.

304 In this study, more *bZIP* genes were found from *U. virens* and the detail investage were
 305 performed, which will provide insights for functional research of these genes in *U. virens* and
 306 reference for other pathogens. There were 28 *UvbZIP* genes systematically identified, which
 307 added nine more compared with previous *U. virens* genomic studies (Zhang et al. 2014). The
 308 total number of *UvbZIPs* identified was more than that in *F. graminearum* (25), *M. oryzae* (25)

and *Hyaloperonospora arabidopsidis* (25), but much less than that of the oomycete *P. sojae* (71). The ratio of *UvbZIPs* in the *U. virens* genome was about 0.33%, which was similar to *P. sojae* (0.37%) and more than *M. oryzae* (0.23%), *F. graminearum* (0.18%) and *H. arabidopsidis* (0.17%) (Ye et al. 2013).

Certain bZIP proteins contained other domains in addition to the bZIP domain, which was common in plant species (Li et al. 2015a; Nijhawan et al. 2008). For fungi, few bZIP proteins have been described (Ye et al. 2013). In addition to the bZIP domain, *UvbZIP8* contained Aft1 HRA, Aft1 OSM and Aft1 HRR domains, similar to the Aft1 protein in *Schizosaccharomyces pombe*. In *S. pombe*, Aft1 can form a heterodimer with Pcr1 that also contains a bZIP domain and regulates sexual development, numerous stress responses, and activate meiotic recombination (Gao et al. 2008; Takeda et al. 1995). Aft1 has a modular organization in which distinct regions differentially affect the osmotic stress response (OSA) and meiotic recombination (HRA, HRR). The HRA and HRR regions are necessary and sufficient to activate and repress recombination, respectively (Gao et al. 2008). Thus *UvbZIP8* may play a similar role to regulate development and stress tolerance in *U. virens*. *UvbZIP11* was predicted to contain a PAP1 domain. In *S. pombe*, the Pap1 protein is a transcription factor that activates the downstream gene peroxiredoxin Tpx1 to regulate H₂O₂-sensing (Castillo et al. 2002; Vivancos et al. 2005). YAP1 is the homologue of Pap1 in *S. cerevisiae*, and responsible for transcriptional activation of various genes involved in ROS detoxification. Yap1 homologs were identified in several fungal pathogens, which share the function in stress tolerance (Lessing et al. 2007; Lev et al. 2005). Also Yap1-mediated ROS detoxification was an essential virulence determinant in *M. oryzae*, *U. maydis*, *Alternaria alternata* and *Candida albicans* (Guo et al. 2011; Lev et al. 2005; Lin et al. 2009; Molina and Kahmann 2007). *UvbZIP26* contains the carboxypeptidase activation peptide domain that may be responsible for modulation of folding and activity of the pro-enzyme. The other predicated domain of *UvbZIP5* and *UvbZIP21* showed overlap with bZIP domain indicating that the two bZIPs probably share other functions. The bZIP proteins contained additional conserved domains indicating potential functions. Some regions of *UvbZIP* proteins enriched in Pro, Gln or acidic amino acids also were observed. These special regions probably act as transactivation domains to regulate the expression of downstream genes (Liao et al. 2008; Vettore et al. 1998; Wei et al. 2012).

In eukaryotes, introns might also play important roles to regulate gene expression, since more introns ordinarily mean more complex regulation, which may involve complex biological processes (Callis et al. 1987). In some plant species, up to 10 introns were found in bZIP genes (Hu et al. 2016; Liu et al. 2014; Liu and Chu 2015). Gene structure analysis showed that the number of introns of *UvbZIPs* was varied. Nearly half of *UvbZIP* genes contain two introns and maximum number of introns were four, which was less than plant

species. Previous research showed that the rate of intron loss is faster than intron gain after segmental duplication (Nuruzzaman et al. 2010). Thus, some introns may be lost in *UvbZIPs* compare to plant species. In plant species such as rice, maize and Arabidopsis, the patterns of those motifs exhibited regular conservation and diversity (Nijhawan et al. 2008; Wei et al. 2012). The bZIP domains of nine genes were inserted by introns and all of that were in the N-x7-R motif coding regions but not leucine zipper. The pattern is different from *Oryza sativa* and grape (*Vitis vinifera*) (Liu et al. 2014; Nijhawan et al. 2008), but similar to *Zea mays* (Wei et al. 2012). The intron position in bZIP domain coding regions was different, indicating that some splicing change in the hinge region would change the code of the bZIP domain of the proteins and further affect their function.

Together with bZIPs from *M. oryzae*, *M. robertsii* and *U. virens*, the total bZIPs were divided into eight clades by phylogenetic analysis. All the results indicated the existence of a diversified bZIP family in *U. virens* with diverse functions. Orthology include bZIP TF family often, but not always, have the same function. Comparison of *UvbZIPs* with that in other fungi could shed light on their functions. Compared with other fungi, seven *UvbZIPs* were specific in *U. virens*, suggesting that these genes perform some functions in processes that are different from other fungi. There were three conserved *UvbZIPs*, and these might play core functions across other fungi. *UvbZIP5* is the ortholog of *MobZIP12* which is essential for conidial morphology in *M. oryzae* (Tang et al. 2015). Combining the domain analysis and sequence alignment results, *UvbZIP11* is probably the ortholog of *Yap1* in other fungi, which mediates the oxidative stress response and is required for pathogenicity in *M. oryzae* (Guo et al. 2011). The ortholog of the *UvbZIP13*, *MoMeab* in *M. oryzae*, is involved in the regulation of vegetative growth (Tang et al. 2015). These conserved genes probably derived from an ancestral gene that underwent numerous modifications during the evolution, including horizontal gene transfer and subsequent functional diversification. Furthermore, we also observed that the ortholog of some *UvbZIPs* were present or absent in other fungi according. Some specific bZIPs identified in *U. virens* might perform functions related to the special biological processes.

Some *bZIP* genes in the plant pathogen *M. oryzae* are required for infection, and they show high expression levels at the infection stage (Guo et al. 2010; Tang et al. 2015). Gene expression profiling indicated that 17 *UvbZIPs* showed high transcriptional levels at 3 DPI or 6 DPI, indicating that these genes are involved in the pathogen infection process. Also the *UvbZIPs* that did not show high expression levels, may be involved in other biological processes other than infection.

Oxidative burst is a fundamental defense response to avoid pathogen invasion. There are some *bZIP* genes in plant pathogens reported to be involved in oxidative stress (Gamboa-Melendez et al. 2013; Guo et al. 2011). The 11 infection-associated *UvbZIPs* were

detected under H₂O₂ treatment, and eight of them showed significantly elevated transcript levels at 1 h post treatment compared with no treatment. The expression level of seven *UvbZIPs* showed reduction or no effects at 0.5 h post treatment. For *P. sojae*, many *bZIP* genes showed elevated expression at 0.5 h H₂O₂ treatment (Ye et al. 2013). It is possible that *U. virens* has a distinct mechanism of infection, hence showing a different response to oxidative stress. These results suggested that *UvbZIP* genes may play important roles in plant ROS stress tolerance, which is consistent with previous studies that some *bZIP* genes in plant pathogens are involved in peroxide stress (Gamboa-Melendez et al. 2013; Guo et al. 2010).

Conclusions

The bZIP transcription factors have been shown to play crucial roles in development, physiological processes, metabolism and stress responses in the plant and fungal kingdoms. In this study, we identified a comprehensive set of 28 bZIP TF family members in the *U. virens* genome, which expanded the list from previous studies. We found 17 *UvbZIP* genes involved in infection and eight were involved oxidative stress. In brief, our genome wide systematic characterization and expressional analysis of *UvbZIPs* has provided insights for the molecular functions of these genes in *U. virens* and reference for other plant pathogenic fungi. The next step is to functional studies by knocking out genes.

Acknowledgments: This work was supported by the Fundamental Research Funds for the Central Universities (grant numbers 2662015QC045 and 2662014BQ013). We thank Tom Hsiang from University of Guelph for language modification of the manuscript.

Author Contributions: Weixiao Yin and Chaoxi Luo conceived and designed the study, Weixiao Yin, Peng Cui and Wei Wei analyzed the data, Peng Cui and Yang Lin performed the real-time PCR assay, Weixiao Yin and Chaoxi Luo wrote the paper. All authors have read and approved the final version of the manuscript.

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

References

- Asano, Y., Hagiwara, D., Yamashino, T., and Mizuno, T. 2007. Characterization of the bZip-type transcription factor NapA with reference a to oxidative stress response in *Aspergillus nidulans*. *Biosci Biotech Bioch* **71**(7): 1800-1803. doi: 10.1271/bbb.70133.
- Atia, M.M.M. 2004. Rice false smut (*Ustilaginoidea virens*) in Egypt. *Journal of Plant Diseases and Protection* **111**(1): 71-82. doi: 10.1007/bf03356134.
- Balazs, A., Pocs, I., Hamari, Z., Leiter, E., Emri, T., Miskei, M., Olah, J., Toth, V., Hegedus, N., Prade, R.A., Molnar, M., and Pocs, I. 2010. AtfA bZIP-type transcription factor regulates oxidative and osmotic stress responses in *Aspergillus nidulans*. *Molecular Genetics and Genomics* **283**(3): 289-303. doi: 10.1007/s00438-010-0513-z.

- 418 Bischoff, J.F., Sullivan, R.F., Kjer, K.M., and White, J.F. 2004. Phylogenetic placement of the
419 anamorphic tribe Ustilaginoideae (Hypocreales, Ascomycota). *Mycologia* **96**(5): 1088-1094.
- 420 Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M., and Lecharny, A. 2001. Evolution of
421 intron/exon structure of DEAD helicase family genes in *Arabidopsis*, *Caenorhabditis*, and *Drosophila*.
422 *Genome Research* **11**(12): 2101-2114. doi: 10.1101/gr.200801.
- 423 Callis, J., Fromm, M., and Walbot, V. 1987. Introns increase gene expression in cultured maize cells.
424 *Genes & Development* **1**(10): 1183-1200.
- 425 Castillo, E.A., Ayte, J., Chiva, C., Moldon, A., Carrascal, M., Abian, J., Jones, N., and Hidalgo, E.
426 2002. Diethylmaleate activates the transcription factor Pap1 by covalent modification of critical
427 cysteine residues. *Molecular Microbiology* **45**(1): 243-254.
- 428 Chen, M., Tan, Q., Sun, M., Li, D., Fu, X., Chen, X., Xiao, W., Li, L., and Gao, D. 2016. Genome-wide
429 identification of *WRKY* family genes in peach and analysis of *WRKY* expression during bud dormancy.
430 *Molecular Genetics and Genomics* **291**(3): 1319-1332. doi: 10.1007/s00438-016-1171-6.
- 431 Chen, Y., Zhang, Y., Yao, J., Li, Y.-F., Yang, X., Wang, W.-X., Zhang, A.-F., and Gao, T.-C. 2013.
432 Frequency distribution of sensitivity of *Ustilaginoidea virens* to four EBI fungicides, prochloraz,
433 difenoconazole, propiconazole and tebuconazole, and their efficacy in controlling rice false smut in
434 Anhui Province of China. *Phytoparasitica* **41**(3): 277-284. doi: 10.1007/s12600-013-0288-y.
- 435 Dey, A., Samanta, M.K., Gayen, S., Sen, S.K., and Maiti, M.K. 2016. Enhanced gene expression rather
436 than natural polymorphism in coding sequence of the *OsbZIP23* determines drought tolerance and yield
437 improvement in rice genotypes. *PLoS ONE* **11**(3): e0150763. doi: 10.1371/journal.pone.0150763.
- 438 Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C. 1992. The GCN4 basic region leucine
439 zipper binds DNA as a dimer of uninterrupted α Helices: Crystal structure of the protein-DNA
440 complex. *Cell* **71**(7): 1223-1237. doi: [http://dx.doi.org/10.1016/S0092-8674\(05\)80070-4](http://dx.doi.org/10.1016/S0092-8674(05)80070-4).
- 441 Etxebeste, O., Ni, M., Garzia, A., Kwon, N.J., Fischer, R., Yu, J.H., Espeso, E.A., and Ugalde, U. 2008.
442 Basic-zipper-type transcription factor FlbB controls asexual development in *Aspergillus nidulans*.
443 *Eukaryotic cell* **7**(1): 38-48. doi: 10.1128/EC.00207-07.
- 444 Fassler, J., Landsman, D., Acharya, A., Moll, J.R., Bonovich, M., and Vinson, C. 2002. B-ZIP proteins
445 encoded by the *Drosophila* genome: evaluation of potential dimerization partners. *Genome Research*
446 **12**(8): 1190-1200. doi: 10.1101/gr.67902.
- 447 Gamboa-Melendez, H., Huerta, A.I., and Judelson, H.S. 2013. bZIP transcription factors in the
448 oomycete *Phytophthora infestans* with novel DNA-binding domains are involved in defense against
449 oxidative stress. *Eukaryotic cell* **12**(10): 1403-1412. doi: 10.1128/EC.00141-13.
- 450 Gao, J., Davidson, M.K., and Wahls, W.P. 2008. Distinct regions of ATF/CREB proteins Atf1 and Pcr1
451 control recombination hotspot *ade6-M26* and the osmotic stress response. *Nucleic Acids Research*
452 **36**(9): 2838-2851. doi: 10.1093/nar/gkn037.
- 453 Gao, Q., Jin, K., Ying, S.H., Zhang, Y., Xiao, G., Shang, Y., Duan, Z., Hu, X., Xie, X.Q., Zhou, G.,
454 Peng, G., Luo, Z., Huang, W., Wang, B., Fang, W., Wang, S., Zhong, Y., Ma, L.J., St Leger, R.J., Zhao,
455 G.P., Pei, Y., Feng, M.G., Xia, Y., and Wang, C. 2011. Genome sequencing and comparative
456 transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS*
457 *Genet* **7**(1): e1001264. doi: 10.1371/journal.pgen.1001264.
- 458 Guo, M., Chen, Y., Du, Y., Dong, Y., Guo, W., Zhai, S., Zhang, H., Dong, S., Zhang, Z., Wang, Y.,
459 Wang, P., and Zheng, X. 2011. The bZIP transcription factor MoAP1 mediates the oxidative stress
460 response and is critical for pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog*
461 **7**(2): e1001302. doi: 10.1371/journal.ppat.1001302.

- Guo, M., Guo, W., Chen, Y., Dong, S., Zhang, X., Zhang, H., Song, W., Wang, W., Wang, Q., Lv, R., Zhang, Z., Wang, Y., and Zheng, X. 2010. The basic leucine zipper transcription factor Moatf1 mediates oxidative stress responses and is necessary for full virulence of the rice blast fungus *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions* **23**(8): 1053-1068. doi: 10.1094/MPMI-23-8-1053.
- Hagiwara, D., Asano, Y., Yamashino, T., and Mizuno, T. 2008. Characterization of bZip-type transcription factor AtfA with reference to stress responses of conidia of *Aspergillus nidulans*. *Biosci Biotech Bioch* **72**(10): 2756-2760. doi: 10.1271/bbb.80001.
- Hossain, M.A., Cho, J.-I., Han, M., Ahn, C.-H., Jeon, J.-S., An, G., and Park, P.B. 2010. The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice. *Journal of Plant Physiology* **167**(17): 1512-1520. doi: <http://dx.doi.org/10.1016/j.jplph.2010.05.008>.
- Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**(8): 1296-1297. doi: 10.1093/bioinformatics/btu817.
- Hu, W., Yang, H., Yan, Y., Wei, Y., Tie, W., Ding, Z., Zuo, J., Peng, M., and Li, K. 2016. Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Scientific Reports* **6**: 22783. doi: 10.1038/srep22783.
- Hurst, H.C. 1995. Transcription factors 1: bZIP proteins. *Protein profile* **2**(2): 101-168.
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F. 2002. bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* **7**(3): 106-111. doi: [http://dx.doi.org/10.1016/S1360-1385\(01\)02223-3](http://dx.doi.org/10.1016/S1360-1385(01)02223-3).
- Jia, Q., Lv, B., Guo, M.Y., Luo, C.X., Zheng, L., Hsiang, T., and Huang, J.B. 2015. Effect of rice growth stage, temperature, relative humidity and wetness duration on infection of rice panicles by *Villosiclava virens*. *Eur J Plant Pathol* **141**(1): 15-25. doi: 10.1007/s10658-014-0516-4.
- Katiyar, A., Smita, S., Lenka, S.K., Rajwanshi, R., Chinnusamy, V., and Bansal, K.C. 2012. Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genomics* **13**: 544. doi: 10.1186/1471-2164-13-544.
- Kong, S., Park, S.Y., and Lee, Y.H. 2015. Systematic characterization of the bZIP transcription factor gene family in the rice blast fungus, *Magnaporthe oryzae*. *Environmental microbiology* **17**(4): 1425-1443. doi: 10.1111/1462-2920.12633.
- López-Berges, M.S., Capilla, J., Turrà, D., Schaffner, L., Matthijs, S., Jöchl, C., Cornelis, P., Guarro, J., Haas, H., and Di Pietro, A. 2012. HapX-mediated iron homeostasis is essential for rhizosphere competence and virulence of the soilborne pathogen *Fusarium oxysporum*. *The Plant Cell* **24**(9): 3805-3822. doi: 10.1105/tpc.112.098624.
- Landschulz, W.H., Johnson, P.F., and McKnight, S.L. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**(4860): 1759-1764.
- Lessing, F., Kniemeyer, O., Wozniok, I., Loeffler, J., Kurzai, O., Haertl, A., and Brakhage, A.A. 2007. The *Aspergillus fumigatus* transcriptional regulator AfYap1 represents the major regulator for defense against reactive oxygen intermediates but is dispensable for pathogenicity in an intranasal mouse infection model. *Eukaryotic cell* **6**(12): 2290-2302. doi: 10.1128/EC.00267-07.
- Lev, S., Hadar, R., Amedeo, P., Baker, S.E., Yoder, O.C., and Horwitz, B.A. 2005. Activation of an AP1-like transcription factor of the maize pathogen *Cochliobolus heterostrophus* in response to oxidative stress and plant signals. *Eukaryotic cell* **4**(2): 443-454. doi: 10.1128/EC.4.2.443-454.2005.
- Li, D., Fu, F., Zhang, H., and Song, F. 2015a. Genome-wide systematic characterization of the bZIP

- transcriptional factor family in tomato (*Solanum lycopersicum* L.). BMC Genomics **16**(1): 771. doi: 10.1186/s12864-015-1990-6.
- Li, W., Li, L., Feng, A., Zhu, X., and Li, J. 2013. Rice false smut fungus, *Ustilagoidea virens*, inhibits pollen germination and degrades the integuments of rice ovule. American Journal of Plant Sciences **4**: 2295-2304.
- Li, X., Gao, S., Tang, Y., Li, L., Zhang, F., Feng, B., Fang, Z., Ma, L., and Zhao, C. 2015b. Genome-wide identification and evolutionary analyses of bZIP transcription factors in wheat and its relatives and expression profiles of anther development related *TabZIP* genes. BMC Genomics **16**: 976. doi: 10.1186/s12864-015-2196-7.
- Liao, Y., Zou, H.F., Wei, W., Hao, Y.J., Tian, A.G., Huang, J., Liu, Y.F., Zhang, J.S., and Chen, S.Y. 2008. Soybean *GmbZIP44*, *GmbZIP62* and *GmbZIP78* genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis*. Planta **228**(2): 225-240. doi: 10.1007/s00425-008-0731-3.
- Lin, C.H., Yang, S.L., and Chung, K.R. 2009. The YAP1 homolog-mediated oxidative stress tolerance is crucial for pathogenicity of the necrotrophic fungus *Alternaria alternata* in citrus. Molecular Plant-Microbe Interactions **22**(8): 942-952. doi: 10.1094/Mpmi-22-8-0942.
- Liu, J., Chen, N., Chen, F., Cai, B., Dal Santo, S., Tornielli, G.B., Pezzotti, M., and Cheng, Z.M. 2014. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). BMC Genomics **15**: 281. doi: 10.1186/1471-2164-15-281.
- Liu, X., and Chu, Z. 2015. Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in *Brachypodium distachyon*. BMC Genomics **16**(1): 227. doi: 10.1186/s12864-015-1457-9.
- Lopez-Berges, M.S., Rispail, N., Prados-Rosales, R.C., and Di Pietro, A. 2010. A nitrogen response pathway regulates virulence functions in *Fusarium oxysporum* via the protein kinase TOR and the bZIP protein MeaB. The Plant cell **22**(7): 2459-2475. doi: 10.1105/tpc.110.075937.
- Meng, J., Sun, W., Mao, Z., Xu, D., Wang, X., Lu, S., Lai, D., Liu, Y., Zhou, L., and Zhang, G. 2015. Main Ustilaginoidins and Their Distribution in Rice False Smut Balls. Toxins **7**(10): 4023-4034. doi: 10.3390/toxins7104023.
- Molina, L., and Kahmann, R. 2007. An *Ustilago maydis* gene involved in H₂O₂ detoxification is required for virulence. The Plant cell **19**(7): 2293-2309. doi: 10.1105/tpc.107.052332.
- Nijhawan, A., Jain, M., Tyagi, A.K., and Khurana, J.P. 2008. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiology **146**(2): 333-350. doi: 10.1104/pp.107.112821.
- Niu, X., and Guiltinan, M.J. 1994. DNA binding specificity of the wheat bZIP protein EmBP-1. Nucleic Acids Research **22**(23): 4969-4978.
- Nuruzzaman, M., Manimekalai, R., Sharoni, A.M., Satoh, K., Kondoh, H., Ooka, H., and Kikuchi, S. 2010. Genome-wide analysis of NAC transcription factor family in rice. Gene **465**(1-2): 30-44. doi: 10.1016/j.gene.2010.06.008.
- Park, J., Park, J., Jang, S., Kim, S., Kong, S., Choi, J., Ahn, K., Kim, J., Lee, S., Kim, S., Park, B., Jung, K., Kim, S., Kang, S., and Lee, Y.H. 2008. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. Bioinformatics **24**(7): 1024-1025. doi: 10.1093/bioinformatics/btn058.
- Perez-Rodriguez, P., Riano-Pachon, D.M., Correa, L.G., Rensing, S.A., Kersten, B., and Mueller-Roeber, B. 2010. PlnTFDB: updated content and new features of the plant transcription factor

- 550 database. Nucleic Acids Research **38**(Database issue): D822-827. doi: 10.1093/nar/gkp805.
- 551 Rush, M.C., Shahjahan, A.K.M., Jones, J.P., and Groth, D.E. 2000. Outbreak of false smut of rice in
552 Louisiana. Plant Disease **84**(1): 100-100. doi: 10.1094/PDIS.2000.84.1.100D.
- 553 Sakamoto, K., Arima, T.H., Iwashita, K., Yamada, O., Gomi, K., and Akita, O. 2008. *Aspergillus oryzae*
554 *atfB* encodes a transcription factor required for stress tolerance in conidia. Fungal Genetics and Biology
555 **45**(6): 922-932. doi: 10.1016/j.fgb.2008.03.009.
- 556 Schwechheimer, C., and Bevan, M. 1998. The regulation of transcription factor activity in plants.
557 Trends in Plant Science **3**(10): 378-383. doi: [http://dx.doi.org/10.1016/S1360-1385\(98\)01302-8](http://dx.doi.org/10.1016/S1360-1385(98)01302-8).
- 558 Shimizu, H., Sato, K., Berberich, T., Miyazaki, A., Ozaki, R., Imai, R., and Kusano, T. 2005. LIP19, a
559 basic region leucine zipper protein, is a Fos-like molecular switch in the cold signaling of rice plants.
560 Plant and Cell Physiology **46**(10): 1623-1634. doi: 10.1093/pcp/pci178.
- 561 Song, J.H., Wei, W., Lv, B., Lin, Y., Yin, W.X., Peng, Y.L., Schnabel, G., Huang, J.B., Jiang, D.H., and
562 Luo, C.X. 2016. Rice false smut fungus hijacks the rice nutrients supply by blocking and mimicking
563 the fertilization of rice ovary. Environmental microbiology **18**(11): 3840-3849. doi:
564 10.1111/1462-2920.13343.
- 565 Sun, X., Kang, S., Zhang, Y., Tan, X., Yu, Y., He, H., Zhang, X., Liu, Y., Wang, S., Sun, W., Cai, L., and
566 Li, S. 2013. Genetic diversity and population structure of rice pathogen *Ustilaginoidea virens* in China.
567 PLoS One **8**(9): e76879. doi: 10.1371/journal.pone.0076879.
- 568 Takahashi, H., Kawakatsu, T., Wakasa, Y., Hayashi, S., and Takaiwa, F. 2012. A rice transmembrane
569 bZIP transcription factor, OsbZIP39, regulates the endoplasmic reticulum stress response. Plant and
570 Cell Physiology **53**(1): 144-153. doi: 10.1093/pcp/pcr157.
- 571 Takeda, T., Toda, T., Kominami, K., Kohnosu, A., Yanagida, M., and Jones, N. 1995.
572 *Schizosaccharomyces pombe atf1⁺* encodes a transcription factor required for sexual development and
573 entry into stationary phase. The EMBO Journal **14**(24): 6193-6208.
- 574 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular
575 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum
576 parsimony methods. Molecular Biology and Evolution **28**(10): 2731-2739. doi:
577 10.1093/molbev/msr121.
- 578 Tang, W., Ru, Y., Hong, L., Zhu, Q., Zuo, R., Guo, X., Wang, J., Zhang, H., Zheng, X., Wang, P., and
579 Zhang, Z. 2015. System-wide characterization of bZIP transcription factor proteins involved in
580 infection-related morphogenesis of *Magnaporthe oryzae*. Environmental microbiology **17**(4):
581 1377-1396. doi: 10.1111/1462-2920.12618.
- 582 Tang, Y.X., Jin, J., Hu, D.W., Yong, M.L., Xu, Y., and He, L.P. 2013. Elucidation of the infection
583 process of *Ustilaginoidea virens* (teleomorph: *Villosiclava virens*) in rice spikelets. Plant Pathology
584 **62**(1): 1-8. doi: 10.1111/j.1365-3059.2012.02629.x.
- 585 Temme, N., and Tudzynski, P. 2009. Does *Botrytis cinerea* ignore H₂O₂-induced oxidative stress during
586 infection? characterization of *Botrytis* activator protein 1. Molecular Plant-Microbe Interactions **22**(8):
587 987-998. doi: 10.1094/MPMI-22-8-0987.
- 588 Tian, C., Li, J., and Glass, N.L. 2011. Exploring the bZIP transcription factor regulatory network in
589 *Neurospora crassa*. Microbiology **157**(Pt 3): 747-759. doi: 10.1099/mic.0.045468-0.
- 590 Vettore, A.L., Yunes, J.A., Cord Neto, G., da Silva, M.J., Arruda, P., and Leite, A. 1998. The molecular
591 and functional characterization of an *Opaque2* homologue gene from *Coix* and a new classification of
592 plant bZIP proteins. Plant Mol Biol **36**(2): 249-263.
- 593 Vinson, C.R., Sigler, P.B., and McKnight, S.L. 1989. Scissors-grip model for DNA recognition by a

family of leucine zipper proteins. *Science* **246**(4932): 911-916.

Vivancos, A.P., Castillo, E.A., Biteau, B., Nicot, C., Ayte, J., Toledano, M.B., and Hidalgo, E. 2005. A cysteine-sulfinic acid in peroxiredoxin regulates H₂O₂-sensing by the antioxidant Pap1 pathway. *Proc Natl Acad Sci U S A* **102**(25): 8875-8880. doi: 10.1073/pnas.0503251102.

Wang, F., Lin, Y., Yin, W.X., Peng, Y.L., Schnabel, G., Huang, J.B., and Luo, C.X. 2015a. The Y137H mutation of *VvCYP51* gene confers the reduced sensitivity to tebuconazole in *Villosiclava virens*. *Scientific Reports* **5**: 17575. doi: 10.1038/srep17575.

Wang, F., Zhang, S., Liu, M.-G., Lin, X.-S., Liu, H.-J., Peng, Y.-L., Lin, Y., Huang, J.-B., and Luo, C.-X. 2014. Genetic diversity analysis reveals that geographical environment plays a more important role than rice cultivar in *Villosiclava virens* population selection. *Applied and Environmental Microbiology* **80**(9): 2811-2820. doi: 10.1128/AEM.03936-13.

Wang, X.-L., Zhong, Y., Cheng, Z.-M., and Xiong, J.-S. 2015b. Divergence of the bZIP gene family in strawberry, peach, and apple suggests multiple modes of gene evolution after duplication. *International Journal of Genomics*: 536943. doi: 10.1155/2015/536943.

Wei, K., Chen, J., Wang, Y., Chen, Y., Chen, S., Lin, Y., Pan, S., Zhong, X., and Xie, D. 2012. Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res* **19**(6): 463-476. doi: 10.1093/dnares/dss026.

Wong, M.L., and Medrano, J.F. 2005. Real-time PCR for mRNA quantitation. *Biotechniques* **39**(1): 75-85. doi: 10.2144/05391rv01.

Xiao, P., Shin, K.S., Wang, T.H., and Yu, J.H. 2010. *Aspergillus fumigatus flbB* encodes two basic leucine zipper domain (bZIP) proteins required for proper asexual development and gliotoxin production. *Eukaryotic Cell* **9**(11): 1711-1723. doi: 10.1128/Ec.00198-10.

Ye, W., Wang, Y., Dong, S., Tyler, B.M., and Wang, Y. 2013. Phylogenetic and transcriptional analysis of an expanded bZIP transcription factor family in *Phytophthora sojae*. *BMC Genomics* **14**: 839. doi: 10.1186/1471-2164-14-839

1471-2164-14-839 [pii].

Yu, M., Yu, J., Hu, J., Huang, L., Wang, Y., Yin, X., Nie, Y., Meng, X., Wang, W., and Liu, Y. 2015. Identification of pathogenicity-related genes in the rice pathogen *Ustilaginoidea virens* through random insertional mutagenesis. *Fungal Genetics and Biology* **76**: 10-19. doi: 10.1016/j.fgb.2015.01.004.

Zhang, C., Liu, J., Zhao, T., Gomez, A., Li, C., Yu, C., Li, H., Lin, J., Yang, Y., Liu, B., and Lin, C. 2016. A drought-inducible bZIP transcription factor OsABF1 delays reproductive timing in rice. *Plant Physiology*. doi: 10.1104/pp.16.01691.

Zhang, Y., Zhang, K., Fang, A., Han, Y., Yang, J., Xue, M., Bao, J., Hu, D., Zhou, B., Sun, X., Li, S., Wen, M., Yao, N., Ma, L.J., Liu, Y., Zhang, M., Huang, F., Luo, C., Zhou, L., Li, J., Chen, Z., Miao, J., Wang, S., Lai, J., Xu, J.R., Hsiang, T., Peng, Y.L., and Sun, W. 2014. Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics. *Nature communications* **5**: 3849. doi: 10.1038/ncomms4849.

Fig.1 Sequence alignments of UvbZIP domains. The protein sequence were prepared by Bioedit software. Black or gray shading indicate identical and conversed amino acid residues, asterisk show the conserve amino acide of bZIP domain.

Fig. 2 Intron distribution patterns in the coding sequence of the UvbZIPs. The exons, represented by gray boxes, were drawn to scale. Black lines connecting two exons represented introns.

Fig. 3 The partern of intron insert in the bZIP domain of the UvbZIP proteins. An example of the consensus sequence (UvbZIP1) is given, and the orresponding positions are shown at bar. The red words are the highly conserved residues of bZIP domain, and Leu amino acid are less conserved for UvbZIPs. The black vertical lines in the bar indicate the intron position in the bZIP domain.

Fig. 4 Phylogenetic analysis of bZIP proteins from *M. oryzae*, *M. robertsii* and *U. virens*. A total of 84 bZIP proteins, 25 from *M. oryzae*, 31 from *M. robertsii* and 28 from *U. virens*, were collected. The sequences were collected from organisms as follows: Mo or or MGG, *M. oryzae*; MAA, *M. robertsii*; Uv, *U. virens*. The phylogenetic tree was generated based on the bZIP protein sequences by using the maximum likelihood method in MEGA 5.1. The proteins were clustered into eight clades.

Fig. 5 Expression profiling of UvbZIPs at different infection stage. The heatmap was constructed based on the expression level of each RPKM value of *UvbZIP* genes from RNA-Seq data. Blue and red boxes indicate lower and higher expression levels, respectively. DPI: days post inoculation.

Fig. 6 Expression profiling of UvbZIPs under the H₂O₂ treatment. Relative expression levels of the genes under 0, 0.5 and 1 h of H₂O₂ treatment using RT-PCR analysis. The expression of each *UvbZIP* gene at 0 h was regarded as a reference and normalized to 1.0 using *α-tubulin* gene of *U. virens* as a reference. And the other values represented the expression levels relative to the reference. Mean values and SDs were obtained from three biological replicates.

666 Table 1. List of bZIP transcription factor family identified in the *U. virens* genome
667

Draft

Table 1. List of bZIP transcription factor family identified in the *U. virens* genome

Gene Name	Gene Size(bp)	Protein (aa)	ORF(bp)	Scaffold	Gene ID
<i>UvbZIP1</i>	978	325	1350	Sca_7	KDB17187. 1
<i>UvbZIP2</i>	1320	439	1355	Sca_1	KDB18923. 1
<i>UvbZIP3</i>	1044	347	1125	Sca_1	KDB18718. 1
<i>UvbZIP4</i>	624	207	729	Sca_1	KDB18635. 1
<i>UvbZIP5</i>	1866	621	1943	Sca_20	KDB14776. 1
<i>UvbZIP6</i>	765	254	906	Sca_20	KDB14749. 1
<i>UvbZIP7</i>	2091	696	2229	Sca_24	KDB14356. 1
<i>UvbZIP8</i>	1632	543	2066	Sca_32	KDB13355. 1
<i>UvbZIP9</i>	1290	429	1394	Sca_63	KDB11662. 1
<i>UvbZIP10</i>	882	293	1023	Sca_81	KDB11104. 1
<i>UvbZIP11</i>	1776	591	2057	Sca_1	KDB19098. 1
<i>UvbZIP12</i>	1362	453	1822	Sca_16	KDB15353. 1
<i>UvbZIP13</i>	1155	384	1831	Sca_11	KDB16349. 1
<i>UvbZIP14</i>	1674	557	1737	Sca_1	KDB19084. 1
<i>UvbZIP15</i>	1095	364	1268	Sca_60	KDB11760. 1
<i>UvbZIP16</i>	1206	401	1268	Sca_60	KDB11763. 1
<i>UvbZIP17</i>	1347	448	1969	Sca_25	KDB14168. 1
<i>UvbZIP18</i>	1143	380	1293	Sca_3	KDB18227. 1
<i>UvbZIP19</i>	1062	353	1349	Sca_1	KDB18692. 1
<i>UvbZIP20</i>	1194	397	1194	Sca_1	KDB18880. 1
<i>UvbZIP21</i>	945	314	1052	Sca_1	KDB18696. 1

<i>UvbZIP22</i>	366	121	441	Sca_170	KDB10769. 1
<i>UvbZIP23</i>	924	307	1262	Sca_34	KDB13153. 1
<i>UvbZIP24</i>	2022	673	2142	Sca_28	KDB13840. 1
<i>UvbZIP25</i>	870	289	1473	Sca_2	KDB18401. 1
<i>UvbZIP26</i>	1494	497	1820	Sca_1	KDB18698. 1
<i>UvbZIP27</i>	1374	457	1556	Sca_1	KDB18741. 1
<i>UvbZIP28</i>	1860	619	2542	Sca_76	KDB11243. 1

Draft

Fig.1 Sequence alignments of UvbZIP domains. The protein sequence were prepared by Bioedit software. Black or gray shading indicate identical and conserved amino acid residues, asterisk show the conserve amino acid of bZIP domain.

23x12mm (600 x 600 DPI)

39x26mm (600 x 600 DPI)

Fig. 3 The pattern of intron insert in the bZIP domain of the UvbZIP proteins. An example of the consensus sequence (UvbZIP1) is given, and the corresponding positions are shown at bar. The red words are the highly conserved residues of bZIP domain, and Leu amino acid are less conserved for UvbZIPs. The black vertical lines in the bar indicate the intron position in the bZIP domain.

12x3mm (600 x 600 DPI)

Draft

Fig. 4 Phylogenetic analysis of bZIP proteins from *M. oryzae*, *M. robertsii* and *U. virens*. A total of 84 bZIP proteins, 25 from *M. oryzae*, 31 from *M. robertsii* and 28 from *U. virens*, were collected. The sequences were collected from organisms as follows: Mo or MGG, *M. oryzae*; MAA, *M. robertsii*; Uv, *U. virens*. The phylogenetic tree was generated based on the bZIP protein sequences by using the maximum likelihood method in MEGA 5.1. The proteins were clustered into eight clades.

44x43mm (600 x 600 DPI)

Fig. 5 Expression profiling of *UvbZIPs* at different infection stage. The heatmap was constructed based on the expression level of each RPKM value of *UvbZIP* genes from RNA-Seq data. Blue and red boxes indicate lower and higher expression levels, respectively. DPI: days post inoculation.

121x323mm (600 x 600 DPI)

Fig. 6 Expression profiling of *UvbZIPs* under the H_2O_2 treatment. Relative expression levels of the genes under 0, 0.5 and 1 h of H_2O_2 treatment using RT-PCR analysis. The expression of each *UvbZIP* gene at 0 h was regarded as a reference and normalized to 1.0 using *α -tubulin* gene of *U. vires* as a reference. And the other values represented the expression levels relative to the reference. Mean values and SDs were obtained from three biological replicates.

21x7mm (600 x 600 DPI)

Draft

Table 1. Primers for qRT-PCR assay

Name	Primer sequences
UvbZIP1RTF	ATGAGCGGCTGAAGAAAGAC
UvbZIP1RTR	TAGAAATCCGTCGGGCAGTA
UvbZIP2RTF	CAACGCAAGGAAGGCTACAT
UvbZIP2RTR	GACTGCGACAGGTTGATGTT
UvbZIP3RTF	AGCAAGGAGTCGGTCAACCA
UvbZIP3RTR	CAGACAGCGTGGGACATGA
UvbZIP4RTF	GAGCGTCAGGCAGGAACAA
UvbZIP4RTR	AGCAGCAGTGGTATGGTAG
UvbZIP5RTF	GACGCTTCTGGACAACGTT
UvbZIP5RTR	ATTCATCTGCTCCTGGCTC
UvbZIP6RTF	TACACAGGCGAATCACAA
UvbZIP6RTR	TGCCACTGTCGTTGTCAT
UvbZIP7RTF	AGCCCTTTCACCAATCTCG
UvbZIP7RTR	CCTCGTAGCCTGCACTTCT
UvbZIP12RTF	CATTGATGATCCCAACGAC
UvbZIP12RTR	AGAGCCATCTTCTTCCAGT
UvbZIP14RTF	TGATACTTCCGAGGTTCCC
UvbZIP14RTR	TTCAGCAGGCGTAGAGTCA
UvbZIP18RTF	CATCGCCTTTGAGTTCCC
UvbZIP18RTR	ATAGCGAGGCTGGGTTGA
UvbZIP20RTF	ACGAAGGCTACTTCGTTG
UvbZIP20RTR	TGAGCCAGGCCATGTTGT
α -tubulin-F	GGCGTTTACAATGGCACTTC
α -tubulin-R	CGGAACAGTTGACCAAAAGG

Table S2 The RPKM value of *UvbZIP* genes at different time points in infection

GeneID	3 DPI	6 DPI
<i>UvbZIP1</i>	0.010	12.738
<i>UvbZIP2</i>	0.010	35.158
<i>UvbZIP3</i>	0.010	3.894
<i>UvbZIP4</i>	0.010	42.449
<i>UvbZIP5</i>	49.497	54.603
<i>UvbZIP6</i>	115.741	63.100
<i>UvbZIP7</i>	0.010	63.731
<i>UvbZIP8</i>	0.010	81.153
<i>UvbZIP9</i>	0.010	11.194
<i>UvbZIP10</i>	0.010	2.230
<i>UvbZIP11</i>	0.010	68.076
<i>UvbZIP12</i>	728.300	993.207
<i>UvbZIP13</i>	0.010	133.466
<i>UvbZIP14</i>	331.045	356.105
<i>UvbZIP15</i>	0.010	20.837
<i>UvbZIP16</i>	0.010	0.326
<i>UvbZIP17</i>	0.010	1.155
<i>UvbZIP18</i>	79.967	381.243
<i>UvbZIP19</i>	0.010	3.951
<i>UvbZIP20</i>	157.479	25.376
<i>UvbZIP21</i>	0.010	22.063
<i>UvbZIP22</i>	0.01	43.96280873
<i>UvbZIP23</i>	0.01	5.640019214
<i>UvbZIP24</i>	0.01	42.41311185
<i>UvbZIP25</i>	0.01	248.8463995
<i>UvbZIP26</i>	0.01	14.05963389
<i>UvbZIP27</i>	0.01	5.397621294
<i>UvbZIP28</i>	0.01	5.407441102

Draft

tion stage

Draft