Gene 453 (2010) 1–8

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Myb transcription factors in the oomycete *Phytophthora* with novel diversified DNA-binding domains and developmental stage-specific expression

Qijun Xiang, Howard S. Judelson *

Department of Plant Pathology and Microbiology, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 4 August 2009
Received in revised form 14 December 2009
Accepted 15 December 2009
Available online 7 January 2010

Received by I.K. Jordan

Keywords:
Gene evolution
DNA-binding domain
Developmental regulation
Transcriptional regulation

ABSTRACT

Transcription factors containing two or three imperfect tandem repeats of the Myb DNA-binding domain (named R2R3 and R1R2R3, respectively) regulate important processes in growth and development. This study characterizes the structure, evolution, and expression of these proteins in the potato pathogen *Phytophthora infestans* and other oomycetes. *P. infestans* was found to encode five R2R3 and nine R1R2R3 transcription factor-like proteins, plus several with additional configurations of Myb domains. Sets of R2R3 and R1R2R3 orthologs are well-conserved in three *Phytophthora* species. Analyses of sites that bind DNA in canonical Myb transcription factors, such as mammalian c-Myb, revealed unusual diversification in the DNA recognition helices of the oomycete proteins. While oomycete R2R3 proteins contain c-Myb-like helices, R1R2R3 proteins exhibit either c-Myb-like or novel sequences. This suggests divergence in their DNA-binding specificities, which was confirmed by electrophoretic mobility shift assays. Eight of the *P. infestans* R2R3 and R1R2R3 genes are up-regulated during sporulation and three during zoospore release, which suggests their involvement in spore development. This is supported by the observation that an oomycete that does not form zoospores, *Hyaloperonospora arabidopsidis*, contains one-third fewer of these genes than *Phytophthora*.

1. Introduction

Myb proteins represent a diverse and widely distributed class of eukaryotic DNA-binding factors, many of which are sequence-specific transcriptional regulators. Most contain a tandem array of two or three variants of the Myb DNA-binding domain, which is named after features first defined in the *v-myb* oncogene of Avian Myeloblastosis Virus (Beug et al., 1979). The cellular counterpart of *v-myb*, the mammalian *c-myb* gene, contains three Myb domains called R1, R2, and R3 and is now the most-studied family member (Gonda et al., 1985). Animals usually express one to three such R1R2R3 proteins, which regulate cellular proliferation and differentiation (Ramsay and Gonda, 2008). Plants also make a limited number of R1R2R3 proteins, plus 100 or more that contain only the R2 and R3 Myb variants. These R2R3 transcription factors control plant-specific metabolic, morphogenic, or stress pathways (Jin and Martin, 1999; Ito, 2005). Species-specific processes regulated by R2R3 or R1R2R3 proteins in other taxa include conidiophore formation in *Aspergillus nidulans* (Wieser and Adams, 1995), development in *Dictyostelium discoideum* (Otsuka and Van Haastert, 1998), encystment in *Giardia lamblia* (Sun et al., 2002) and iron response in *Trichomonas vaginalis* (Ong et al., 2007). Some plants and lower eukaryotes also express proteins with a single Myb domain (Rose et al., 1999; Ehrenkaufer et al., 2009).

In R2R3 and R1R2R3 proteins, each Myb repeat spans about 50 amino acids, and forms three α-helices in which three tryptophans and eight other residues form a stable hydrophobic core (Anton and Frampton, 1988; Saikumar et al., 1990; Ogata et al., 1994). Their second and third helices fold into helix–turn–helix structures which bind the major groove of DNA, like the helix–turn–helix regions of homeodomain proteins. In R1R2R3 proteins the specificity of DNA binding is conferred by the third α-helices of R2 and R3, and not by R1 (Howe et al., 1990). While Myb domains are fairly well-conserved between species, some plant proteins contain unusual amino acid substitutions that may influence the specificity of DNA binding (Solano et al., 1997). Investigations of Myb proteins in more diverse eukaryotes should help reveal the extent of variation in the domain and the evolution of the family.

Myb transcription factors have not been studied previously in *Phytophthora*, a member of the oomycete class of the Kingdom Stramenopila. This is a major but under-studied eukaryotic lineage

Abbreviations: R2R3, protein with Myb domains R2 and R3; R1R2R3, protein with Myb domains R1, R2 and R3; DNA, deoxyribonucleic acid; d, days; g, gravity; min, minutes; rpm, revolutions per secton; mM, millimolar; C, celsius; hr, hours; RT-PCR, reverse transcription polymerase chain reaction; cDNA, complementary DNA; MBP, maltose binding protein; ng, nanogram; A, adenine; C, cytidine; G, guanidine; T, thymidine; µg, microgram; dI:dC, polymer of deoxyinosine and deoxycytidine; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; KCl, potassium chloride; nt, nucleotide; Y, C or T; Asn, asparagine; Glu, glutamic acid; Arg, argnine; His, histidine; Thr, threonine; Phe, phenylalanine; Trp, tryptophan; EMSA, electrophoretic mobility shift assay; Lys, lysine.

* Corresponding author. Tel.: +1 951 827 4199; fax: +1 951 827 4294.
E-mail address: howard.judelson@ucr.edu (H.S. Judelson).

0378-1119/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2009.12.006

that also includes brown algae and diatoms (Adl et al., 2005). *Phytophthora infestans* is typical of many oomycetes in being an important pathogen (potato and tomato late blight) in which disease is integrated with asexual development (Judelson and Blanco, 2005). Central to the disease cycle is the formation from mycelia of multinucleate sporangia. These typically germinate by releasing six or more biflagellated zoospores that lack cell walls, which swim to infection sites. There they transform into walled cysts, which make germ tubes that penetrate plant surfaces using appressoria. Specific sets of genes are transcribed at each of these stages, and a few promoter motifs responsible for stage-specific transcription are known (Tani and Judelson, 2006; Ah Fong et al., 2007; Xiang et al., 2009). Some of the motifs resemble Myb binding sites, raising the possibility that Myb proteins regulate the stage-specific genes.

In this paper, we describe the Myb proteins of *P. infestans* and several other oomycetes to better explore the function and evolution of this important family. Although *P. infestans* is also predicted to produce 34 proteins with single Myb-like helix–turn–helix domains, the emphasis of the analysis was on its nine R1R2R3 and five R2R3 transcription factor-like proteins since these provide more insight into the molecular evolution of the family. Similar numbers of proteins were predicted from the draft genomes of *P. infestans*, *Phytophthora ramorum* and *Phytophthora sojae* but less in an oomycete that lacks the zoospore life stage, *Hyaloperonospora arabidopsidis*. Interestingly, either c-Myb-like or oomycete-specific DNA-binding domains were detected in the R1R2R3 proteins, and several other novel features were also present. Most R1R2R3 and R2R3 genes are expressed specifically during sporulation or germination.

2. Materials and methods

2.1. Sequence analysis

The 104 amino acids of human c-Myb (NCBI accession No. AAB49039) representing its R2R3 domains were used to search the protein and nucleotide databases of the genomes of *P. infestans* (Broad Institute; www.broad.mit.edu); *P. ramorum*, *P. sojae*, and the diatom *Thalassiosira pseudonana* (v. 1.1, 1.1, and 3.0 respectively at the Joint Genome Institute; genome.jgi-psf.org); *H. arabidopsidis* (v. 8.3 at the Virginia Bioinformatics Institute; annuminas.vbi.vt.edu); and the red algae *Cyanidioschyzon merolae* (merolae.biol.s.u-tokyo.ac.jp). Other sequences, including single-Myb genes, were also identified by keyword searches. Sequences having BLAST matches with $E<10^{-3}$ were examined with MOTIFSCAN (Hulo et al., 2008) to assess if they contained Myb-like helix–turn–helix domains. Recursive searches within each species were used to check for additional family members. Supplementary Table 1 lists the Broad Institute gene model numbers of each *P. infestans* gene, the more descriptive names used in Results and discussion, and their orthologs from other oomycetes. *H. arabidopsidis* genes are named in this paper with the prefix "Ha_," followed by the gene number in the Virginia Bioinformatics Institute database.

Alignments of proteins from each species were performed using CLUSTALW (align.genome.jp) and clusters aligned with MULTALIN (bioinfo.genotoul.fr/multalin/ multalin.html). Neighbor-joining trees based on CLUSTALW were used to group Myb domains with R1, R2 or R3 of human c-Myb. Other phylogenetic analyses employed MEGA v.4, using neighbor-joining trees made with Poisson correction and 1000 bootstrap replicates (Tamura et al., 2007).

2.2. Developmental stages of *P. infestans*

Isolate 1306 was maintained on rye-sucrose agar. Vegetative (nonsporulating) mycelia was obtained from clarified rye-sucrose broth cultures initiated by placing sporangia from 9 to 10 day agar plates into 35 ml of clarified broth, incubating 3 days at 18 °C, and harvesting by vacuum filtration. Sporangia were harvested with room temperature water from 9 to 10 day agar cultures. Cleaving sporangia were obtained by collecting sporangia from cultures using ice-cold water, which were incubated 1 h at 8 °C and harvested by centrifugation at 1000×*g*. Swimming zoospores were isolated from sporangia held at 8 °C for 135 min to stimulate zoospore release, and harvested by centrifugation. Germinated cysts were obtained by adding 1 mM $CaCl_2$ to zoospores, which were encysted by shaking for 3 min at 100 rpm on a platform shaker, incubated 6 h at 18 °C to allow germination, and harvested by centrifugation. Directly germinated sporangia were prepared by collecting sporangia from agar plates with room temperature water, which were incubated in clarified rye-sucrose broth for 6 h at 18 °C and harvested by centrifugation.

2.3. Reverse transcription-polymerase chain reaction (RT-PCR)

Tissues were frozen and ground in liquid nitrogen. RNA was extracted with the Qiagen RNAeasy Plant RNA kit, and then 2.5 μg of RNA was treated with 5 units of RNAse-free DNAse for 30 min at 37 °C. cDNA was made using a first-strand synthesis kit from Invitrogen, and subjected to PCR using the gene-specific primers listed in Supplemental Table 2. Reactions were performed with a 55 °C annealing temperature and normalized against the constitutively expressed gene for ribosomal protein S3a.

2.4. Myb protein expression and DNA-binding assays

Recombinant R2R3 domains were expressed as fusions with maltose binding protein (MBP). This involved amplifying the regions from cDNA by PCR, cloning them into pMAL-c2x, expressing the proteins in *E. coli* strain BL21 using 4 h of IPTG induction at 28 °C, and purifying the proteins with amylose beads. The R2R3 domains were amplified from PiMyb3R1 using primers 3R1U and 3R1L, and from PiMyb2R1 using primers 2R1U and 2R1L (Supplementary Table 2). These correspond to amino acids 480–590 and 15–132 of those proteins, respectively. Electrophoretic mobility shift assays (EMSA) were performed essentially as described (Xiang et al., 2009), using the DNAs listed in Supplementary Table 2, in reactions containing 0.35 μg of protein, 1 μg bovine serum albumin, and 1 μg poly(dI:dC), in a buffer of 25 mM $MgCl_2$, 1 mM DTT, 0.15 M KCl, 10 mM Tris–HCl (pH 8.0), 0.5 mM EDTA, and 12.5% glycerol. Cold competitors were incubated with PiMyb2R1 in binding solution for 15 min at room temperature and then incubated with the radiolabeled DNA target for 15 min before electrophoresis.

3. Results and discussion

3.1. Phytophthora species encode several types of Myb proteins

The genomes of *P. infestans*, *P. ramorum, and P. sojae* were searched for sequences encoding proteins with similarity to the R2R3 repeats of human c-Myb, as these domains define the canonical sequence-specific transcription factors in the Myb superfamily. Fifteen genes were found in each genome, including well-conserved ortholog sets for each gene (Supplemental Table 1). Nine of the proteins possess three Myb domains in a R1R2R3 array similar to that of c-Myb, as illustrated in Fig. 1B; these are named PiMyb3R1 to PiMyb3R9, with 3R referring to the three domains. Five contain only R2R3 configurations of Myb domains; these are named PiMyb2R1 to PiMyb2R5. One protein, PiMyb5R, contained a R1R2R3 array plus an atypical combination of other Myb domains.

Like R2R3 and R1R2R3 proteins from other taxonomic groups, the *Phytophthora* proteins exhibited little similarity outside of their DNA-binding domains. Among the *P. infestans* Myb proteins, PiMyb3R5 and PiMyb3R6 are most similar with 75% identity and 85% similarity in the two Myb domains, but are only 37% identical and 46% similar elsewhere. These genes likely were generated by a duplication event

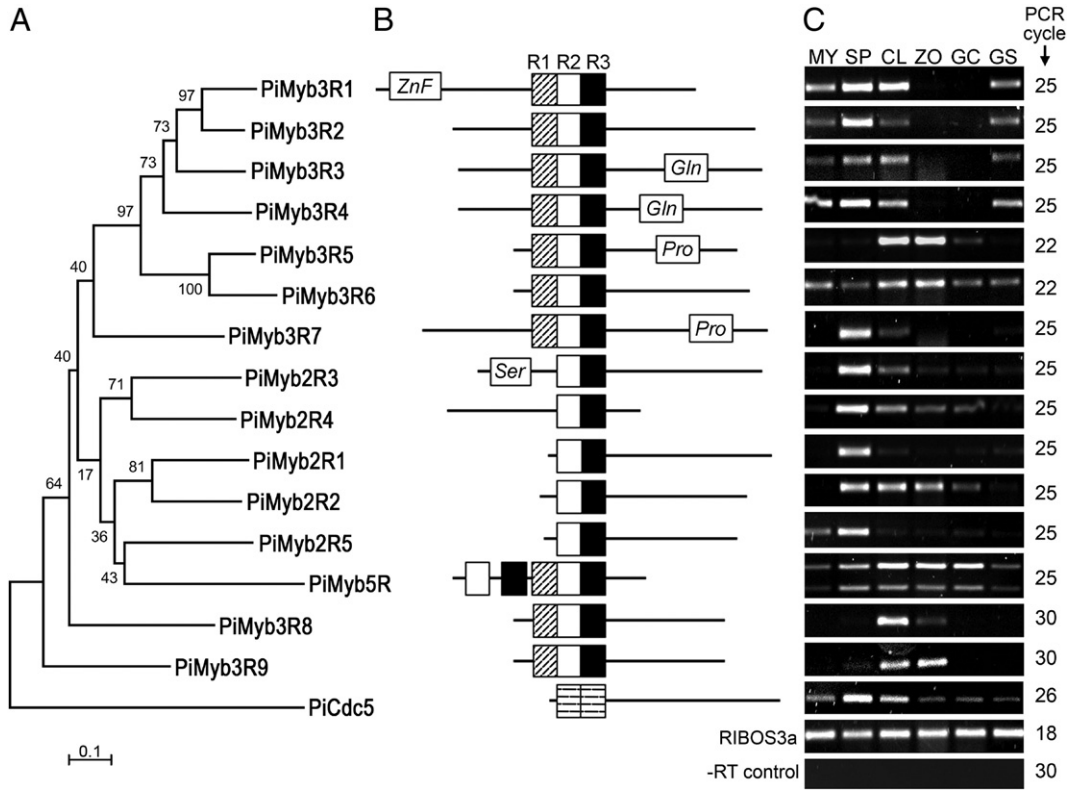

Fig. 1. Classification, structure, and expression patterns of R2R3 and R1R2R3 transcription factors from *P. infestans*. (A) Neighbor-joining tree based on R2R3 domains. Shown is the consensus tree from 1000 bootstrap replicates, with the percentage of node occurrences shown. (B) Relative structures of the proteins. Domains matching R1, R2 and R3 are indicated by grey, white, and black boxes. Other features include a C2H2 type zinc finger (ZnF), and domains that are rich in Gln, Pro, or Ser. (C) Gene expression determination by semi-quantitative RT-PCR. RNA are from mycelia, MY; sporangia, SP; sporangia undergoing cleavage, i.e. zoosporogenesis, CL; swimming zoospores, ZO; germinated zoospore cysts, GC; and sporangia germinated in rye-sucrose media, GS. Indicated in the right margin is the number of cycles used for each series of reactions.

since they are only 194 nt apart from each other. Such tandem duplications are common within *Phytophthora*, but the diversification of sequences between the two genes is greater than average. The other R2R3 and R1R2R3 genes are all unlinked and highly dissimilar outside of their Myb domains, although a few contain features typical of transactivation domains (Latchman, 2003). These include glutamine-rich tracts C-terminal to the Myb domains of PiMyb3R3 and PiMyb3R4, proline-rich C-terminal regions in PiMyb3R5 and PiMyb3R7, and a serine-rich region upstream of the Myb domains in PiMyb2R3 (Fig. 1B).

Besides canonical R2R3 and R1R2R3 proteins, other multidomain Myb proteins were also identified (Supplementary Table 1). For example, PiCdc5 falls into the atypical class of multidomain Myb proteins, as neither of its two tandem Myb domains resemble closely the R2 or R3 regions of c-Myb. Such proteins are found in many eukaryotes, are named after CDC5 of *S. pombe*, and are involved in mRNA splicing but may also exhibit transcription factor activity (McDonald et al., 1999; Nakazawa et al., 2008). Several proteins with novel combinations of Myb domains were also observed. For example, six predicted *P. infestans* proteins contain two or more Myb domains that are not tandemly repeated. Another contains five Myb domains, including a R1R2R3-like tandem array near the C-terminus and unlinked R2 and R3-like regions.

To study the relationships between the R2R3 and R1R2R3 proteins, phylogenetic analyses were performed based on their R2R3 regions, using PiCdc5 as an outgroup. This revealed that all five R2R3 proteins form a coherent clade albeit with a low bootstrap value, and that six of the nine R1R2R3 proteins comprise a distinct and strongly supported group (Fig. 1A). The three other R1R2R3 proteins (PiMyb3R7, PiMyb3R8, PiMyb3R9) are not highly similar to members of either clade, and their position in the tree suggests that PiMyb3R8 or PiMyb3R9 may represent an evolutionary antecedent of the other two and three-domain proteins. As described later, this clustering pattern results largely from oomycete-specific substitutions in the canonical c-Myb sequence that likely confer novel DNA-binding specificities.

P. infestans is also predicted to encode 44 proteins that bear a single Myb-like primary sequence of amino acids, of which 34 form the helix–turn–helix configuration characteristic of Myb DNA-binding domains. Thirty-one of these belong to the SHAQKY family based on the presence of that sequence near their C-termini (Rose et al., 1999). Studies in plants and protists have shown that some SHAQKY proteins are sequence-specific transcription factors (Rubio-Somoza et al., 2006; Ehrenkaufer et al., 2009). However, the SHAQKY Myb domains do not cluster with the R1, R2 or R3 motifs in phylogenetic analyses and have distinct DNA-binding specificities (Ehrenkaufer et al., 2009). Another single-Myb protein (PiMyb1R32) resembles a family of telomeric DNA-binding proteins found in animals and plants, including the presence of a signature VDLKDKWR motif (Karamysheva et al., 2004). Two other single-Myb domain proteins (PiMyb1R33, PiMyb1R34) do not resemble single-domain proteins from other taxa.

3.2. Most Myb genes are differentially expressed during development

Expression profiling using microarrays showed that nearly half of the *P. infestans* transcriptome is differentially expressed during asexual sporulation or germination (Judelson et al., 2008, 2009). However, good data were not obtained for most *Myb* genes due to their low mRNA levels and incomplete coverage of the genome on those microarrays. Semi-quantitative RT-PCR was therefore performed against the 15 R2R3 and R1R2R3 genes (Fig. 1C). All were expressed in one or more stages of asexual development, in several patterns which are described below.

Eight genes are up-regulated during sporulation, to varying degrees. Five have little or extremely low expression in mycelia but are strongly induced during sporulation. These include four R2R3 genes (*PiMyb2R1*, *PiMyb2R2*, *PiMyb2R3*, *PiMyb2R4*) and one R1R2R3 gene (*PiMyb3R7*). Their transcripts persist in sporangia undergoing zoosporogenesis (i.e. cleaving sporangia) and swimming zoospores, but decline by the germinated cyst stage. The mRNA also disappears when sporangia are placed in broth media, which stimulates the production of mycelia from sporangia; this is known as direct germination, which is an alternative to the zoospore mode of germination. Three genes contained moderate levels of mRNA in mycelia, which increased during sporulation. These included the R2R3-encoding gene *PiMyb2R5*, *PiMyb5R* which encodes a protein with an atypical combination of Myb domains, and PiCdc5. In these analyses, two RT-PCR products differing in size by 133 nt were observed for PiMyb5R. When these were cloned and sequenced, the upper band was found to result from incompletely spliced transcripts in which 61 and 72-nt introns were still present.

Three genes display little or no expression in mycelia or sporangia, but are induced during the cleavage stage. These were *PiMyb3R5*, *PiMyb3R8* and *PiMyb3R9*. mRNAs for these genes are then down-regulated in germinated cysts.

Transcripts for *PiMyb3R1*, *PiMyb3R2*, *PiMyb3R3* were detected in mycelia and sporangia, but disappeared in the zoospore and germinated cyst stages. One interpretation of this pattern is that they may be required for filamentous growth. Since sporangia are capable of germinating directly, such mRNAs may need to be maintained in that stage.

The fact that many Myb genes are up-regulated during sporulation or zoosporogenesis suggests that they play important roles in development. Interestingly, two promoter motifs known to activate transcription during sporulation resemble the preferred mammalian c-Myb binding site, which is defined as (C/T)AAC(g/t)G (Deng et al., 1996; Ording et al., 1996). One is a CAACGG responsible for inducing *PiPks1* during intermediate sporulation and which is in one-third of similarly expressed promoters (Xiang et al., 2009). In addition, a 24-nt region that activates early sporulation-induced genes such as *PiCdc14* contains two sequences that match the c-Myb target (Ah Fong et al., 2007). It should be noted that this region of the *PiCdc14* promoter was described originally as including three closely spaced CTYAAC motifs. However, the region can also be read as three (C/T) AACNG sequences, which match the c-Myb binding site. As will be described later, a *P. infestans* R2R3 protein will be shown later to specifically bind this site in the *PiCdc14* promoter. Possibly, a transcription factor cascade involving Myb proteins is active during the spore cycle, in which genes induced at one stage activate genes required for the next.

3.3. Canonical and oomycete-specific sequences within Myb domains R2 and R3

DNA binding by Myb proteins is mediated mostly by the R2 and R3 motifs, with the third α-helix of each motif determining specificity. Within R2, structural studies of c-Myb complexed with DNA indicate that residues Lys128 and Asn136 make specific contacts with DNA bases (Ogata et al., 1994). Within R3, residues Lys182 and Asn183 make specific contacts with DNA bases, Asn179 and Asn186 probably also bind DNA, and Glu178 forms a salt bridge with Arg131 that stabilizes the interaction between R2 and R3 (Ogata et al., 1994). Due to such functions, residues 128, 136, 178, 179, 182, and 183 are nearly invariant in R2R3 and R1R2R3 transcription factors studied to date and explain why most R2R3 and R1R2R3 transcription factors bind a similar sequence in DNA. As shown in Fig. 2, these amino acids are also conserved in all *P. infestans* R2R3 proteins such as PiMyb2R1, and three R1R2R3 proteins such as PiMyb3R7. As discussed in the following paragraphs, however, modest differences from canonical c-Myb sequences were noted within domain R2 of several *P. infestans* proteins, and major changes within domain R3 of six R1R2R3 proteins.

Deviations from c-Myb within some *P. infestans* R2 domains may be predicted to alter the structure of that domain but not necessarily its DNA-binding specificity. For example, the linker between the first and second helices of R2 is one amino acid smaller in most *P. infestans* R1R2R3 proteins, and two residues larger in PiMyb3R7. In addition, residue 123 (near the junction of helices 2 and 3) is Lys in most metazoan c-Myb proteins but Pro in *P. infestans* except for PiMyb3R8 which retains the canonical Lys. This change is also observed in plant Myb proteins (Jiang et al., 2004). Some residues in helix 3 of domain R2 known to influence DNA binding are also sometimes altered. For example, instead of Glu at position 132 as seen in c-Myb, PiMyb2R1 and PiMyb2R2 contain Thr which likely disrupts the interaction of this residue with Arg131, which stabilizes the R2-R3 interaction (Ogata et al., 1994). Based on mutation studies of a plant R2R3 protein, the substitutions in these *P. infestans* proteins may also alter their DNA binding (Williams and Grotewold, 1997).

In contrast to the generally minor changes observed within R2, strong deviations from the canonical c-Myb pattern were observed in the R3 domain of six R1R2R3 proteins, such as PiMyb3R1. Of the five residues conserved in Myb proteins from other taxa (Asp178, Asn179, Lys182, Asn183, Asn186), only Lys182 is preserved within these *P. infestans* proteins. Elsewhere, Asp178 is replaced by Glu, Asn179 by Asp, and Asn183 by Ile or Leu. Residue 186, which is His, Tyr, Arg, Thr or a hydrophobic amino acid in non-oomycete R2R3 or R1R2R3 proteins, is changed to Lys in the proteins that group with PiMyb3R1. Identical changes are seen in orthologs from the other species of *Phytophthora*. Since these substitutions are nonconservative, they likely alter the DNA specificity of these proteins compared to canonical Myb proteins. This is borne out by studies demonstrating that a Glu to Leu change in a Myb domain of a plant R2R3 protein shifted binding from one DNA target to another (Solano et al., 1997). Therefore, the sequences of PiMyb3R1 and relatives define an oomycete-specific Myb-like domain that presumably binds a novel DNA sequence.

Only a few examples exist of proteins from other taxa that deviate from the c-Myb consensus at these important sites within their R3 DNA recognition helices. These include rs-2 of maize and Myb2 of *T. vaginalis*. As would be expected, these bind DNA sequences that differ from the mammalian c-Myb target (Timmermans et al., 1999; Tsiantis et al., 1999; Ong et al., 2007).

An additional notable feature of the R3 domains of the six proteins with the oomycete-specific Myb domain, such as PiMyb3R1, was the substitution of the Trp147 seen in animal Myb proteins by Tyr or Phe (Fig. 2). Substitutions at this site, including Phe, have also been detected in plant R2R3 proteins. While mutation studies in c-Myb suggest that it may not affect DNA binding, Phe147 is considered to be one of the evolutionary hallmarks of plant R2R3 proteins (Saikumar et al., 1990; Dias et al., 2003). However, this feature is apparently shared by oomycete proteins. In contrast, the Tyr substitution is apparently oomycete-specific and is the first reported deviation from Trp in any R1R2R3 protein.

3.4. The novel R3 domain also exists in distantly related oomycetes

Insight into the origin of the unique R3 DNA recognition helix of *Phytophthora* was obtained by examining R2R3 and R1R2R3 proteins from another oomycete, the downy mildew *H. arabidopsidis*. This species is in the family Peronosporaceae, compared to *Phytophthora* which is in the Pythiaceae. *H. arabidopsidis* also has a more simplified life cycle, as it lacks the zoospore stage.

This analysis revealed that *H. arabidopsidis* encodes one R2R3 and eight R1R2R3 proteins, compared to five and nine in *P. infestans*. Seven of the *H. arabidopsidis* R1R2R3 proteins had orthologs in the group from *P. infestans* with the oomycete-specific R3 domain (PiMyb3R1

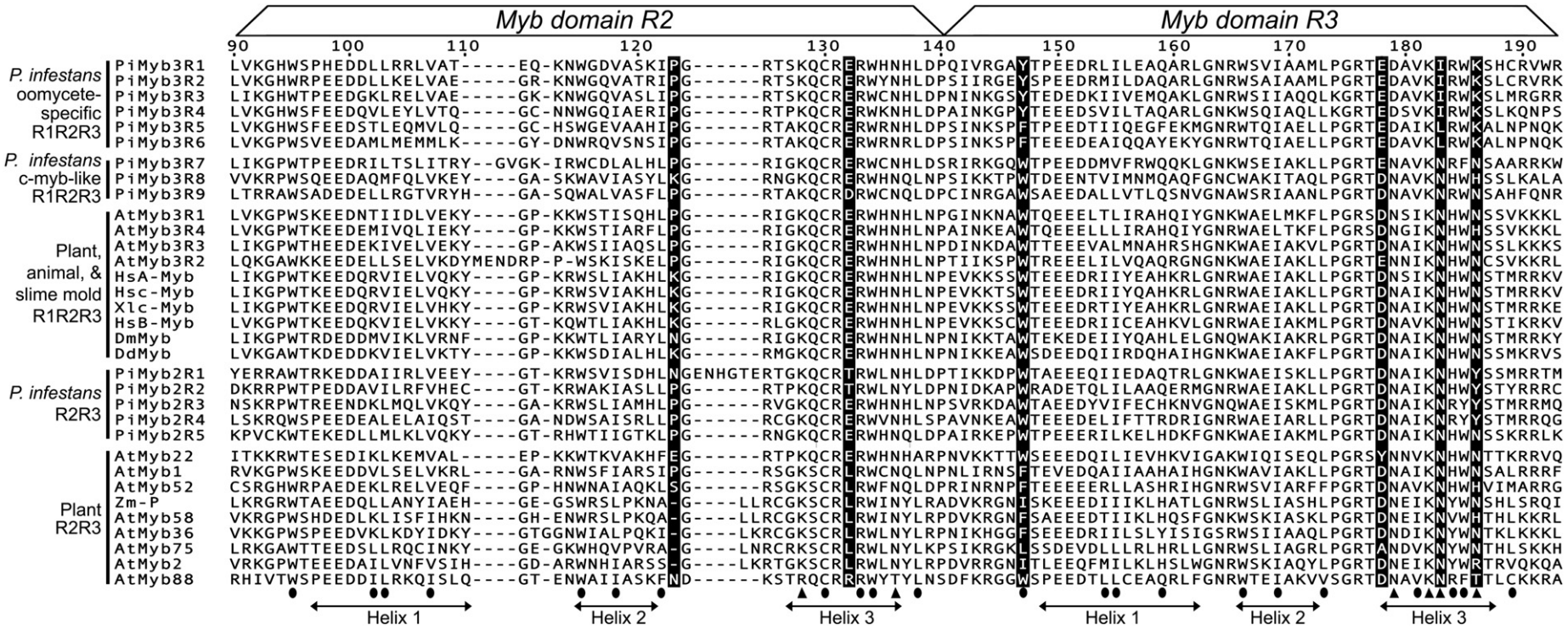

Fig. 2. Alignment of R2 and R3 domains of Myb transcription factors. Numbering is based on human c-Myb. Indicated at the base of the figure are the three helices within each Myb domain, residues that form their hydrophobic cores (ovals), and residues shown to interact with DNA in mammalian c-Myb proteins (triangles; (Ogata et al., 1994). Vertical black blocks indicate sites where many of the *P. infestans* proteins contain notable differences when compared to canonical c-Myb-like proteins. *A. thaliana* sequences are named as described (Stracke et al., 2001). Others are *H. sapiens* proteins Hsc-Myb (GenBank accession AAB49039), HsA-Myb (P10243), and HsB-Myb (P10244); c-Myb relatives from *Drosophila* (DmMyb, AAO25019), *Xenopus* (Xlc-Myb, L22741), and *Dictyostelium discoideum* (DdMyb, DDB0219911), and the maize R2R3 protein Zm-P (P27898).

and relatives). These *H. arabidopsidis* proteins had the same unusual amino acid substitutions seen in helix 3 of that domain in *P. infestans*, except that Asn183 is substituted by either Ile, Leu, or Met. The slightly greater size of the oomycete-specific R1R2R3 group in the downy mildew (seven vs. six in *P. infestans*) is due to a duplication of the *PiMyb3R2* ortholog.

Only one *H. arabidopsidis* R1R2R3 protein, Ha_705789, belonged to the c-Myb-like class. Ha_705789 and PiMyb3R7 of *P. infestans* appear to be the closest orthologous pair. PiMyb3R7 may therefore be the closest existing relative of the ancestral oomycete R1R2R3, which diverged to form the oomycete-specific group.

The largest difference between *P. infestans* and *H. arabidopsidis* was observed in the R2R3 family. While the former encodes five R2R3 proteins, the latter only encodes one (Ha_710573, orthologous to PiMyb2R3). It is notable that most of the missing genes are sporulation-induced in *P. infestans*. This is consistent with the inability of *H. arabidopsidis* to produce zoospores, since a prior microarray study showed that most zoospore-specific genes in *P. infestans* are induced during sporulation (Judelson et al., 2009).

Interestingly, orthologs of many other *P. infestans* sporulation-induced genes are also absent from *H. arabidopsidis*. Of 341 *P. infestans* genes induced >10-fold in sporangia compared to mycelia, only 23% had orthologs in *H. arabidopsidis*. In contrast, orthologous pairs were identified for 74% of genes from a control set. The implication is that some Myb transcription factors and their target genes were both lost in the downy mildew, along with the ability to produce zoospores.

To place the evolution of the Myb family in an even broader context, R2R3 and R1R2R3 proteins were identified from the diatom *T. pseudonana*, which represents another class in the stramenopile kingdom. Eight R2R3 and three R1R2R3 proteins were identified (Supplemental Table 4). These are too diverged from the oomycete proteins to accurately establish orthologous pairs. However, their R3 DNA recognition helices resemble the canonical c-Myb sequence. Therefore, the novel R3 sequences seen in *Phytophthora* and *Hyaloperonospora* likely evolved after the separation of oomycetes from other stramenopiles.

R2R3 and R1R2R3 proteins from the red algae *C. merolae* were also checked for the novel R3 sequences, since many stramenopile genes are believed to have been acquired from such a symbiont (Robertson and Tartar, 2006). Three and one members of those two groups were detected, respectively (Supplemental Table 4). However, all contain DNA recognition helices that resemble the canonical c-Myb protein. Therefore, the novel sequences observed in the oomycetes do not have a red algal origin.

3.5. DNA-binding specificity of canonical and oomycete-specific Myb domains

EMSA was used to assess if the *P. infestans* proteins with DNA recognition helices that match c-Myb bind the same target DNA as c-Myb, and if the oomycete-specific domains show altered specificity. This involved testing recombinant proteins engineered to contain only the R2 and R3 domains of *PiMyb2R1*, which resembles c-Myb, and *PiMyb3R1*, which has the oomycete-specific substitutions. These proteins were then tested for their ability to bind double-stranded oligonucleotides containing the c-Myb-like targets in the *PiCdc14* promoter, (C/T)AACNG. Human c-Myb as well as v-Myb has been shown to bind this sequence (Biedenkapp et al., 1988; Ording et al., 1996).

The results indicated that the R2R3 domains from the c-Myb-like *PiMyb2R1* protein bind to that sequence, while the oomycete-specific R2R3 domains from *PiMyb3R1* do not (Fig. 3A, left panel). As a control, binding was not observed against a double-stranded oligonucleotide lacking the c-Myb-like target (Fig. 3A, right panel). Moreover, EMSA using cold competitors showed that the binding of *PiMyb2R1* to that sequence was specific (Fig. 3B). This is because the wild-type cold target oligonucleotide was an effective competitor for binding, unlike the same oligonucleotide with a mutated target and an unrelated oligonucleotide.

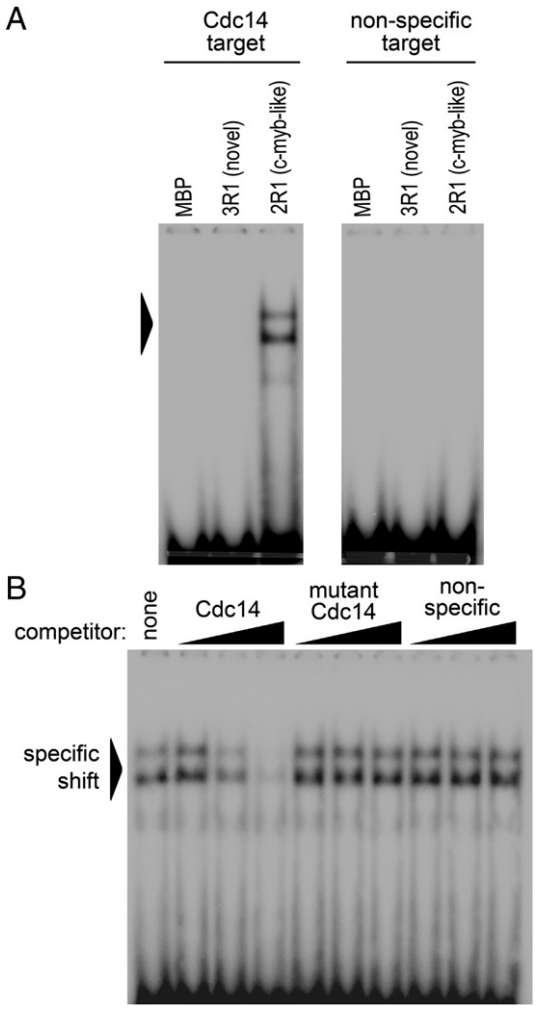

Fig. 3. Binding of Myb proteins to DNA targets. EMSA was performed as described in Materials and methods, using double-stranded oligonucleotides as listed in Supplementary Table 2. (A) EMSA employing maltose binding protein (MBP), MBP fused to the R2R3 region of PiMyb3R1 which contains the novel oomycete-specific domains (3R1-novel), and MBP fused to the R2R3 region of PiMyb2R1 which resembles c-Myb (2R1-c-myb-like). The DNA probe used in the left panel is an oligonucleotide containing three tandem c-Myb-like DNA-binding sites (CAACNG) from the *Cdc14* gene of *P. infestans*. The non-specific probe in the right panel is an unrelated oligonucleotide. (B) EMSA competition assay. This used the PiMyb2R1 fusion protein described above, 1.6 ng of the radiolabeled *Cdc14* probe described above, and 1.6, 16, and 160 ng of unlabeled competitors (left to right). The competitors were either the same as the radiolabeled *Cdc14* probe sequence; the *Cdc14* sequence containing mutations in the c-Myb-like DNA-binding motifs (mutant); and an unrelated oligonucleotide as a non-specific probe.

3.6. Oomycete and plant R2R3 proteins do not share a recent ancestor

Studies of Myb transcription factors in plants suggest that their R2R3 proteins evolved from the R1R2R3 group through the loss of domain R1 (Lipsick, 1996). This is supported by the fact that animals do not produce R2R3 proteins, although their three-Myb proteins may have evolved by duplicating one repeat in a now-extinct R2R3 ancestor (Jiang et al., 2004). Our discovery that oomycetes also encode R2R3 proteins consequently provides an opportunity to better understand the evolution of this group. Phylogenetic approaches were therefore used to test if the R2R3 proteins of *Phytophthora* evolved from an oomycete R1R2R3 protein or a plant-like R2R3 protein.

Neighbor-joining, minimum evolution, and maximum parsimony tree-building strategies based on the R2R3 domains indicate that the

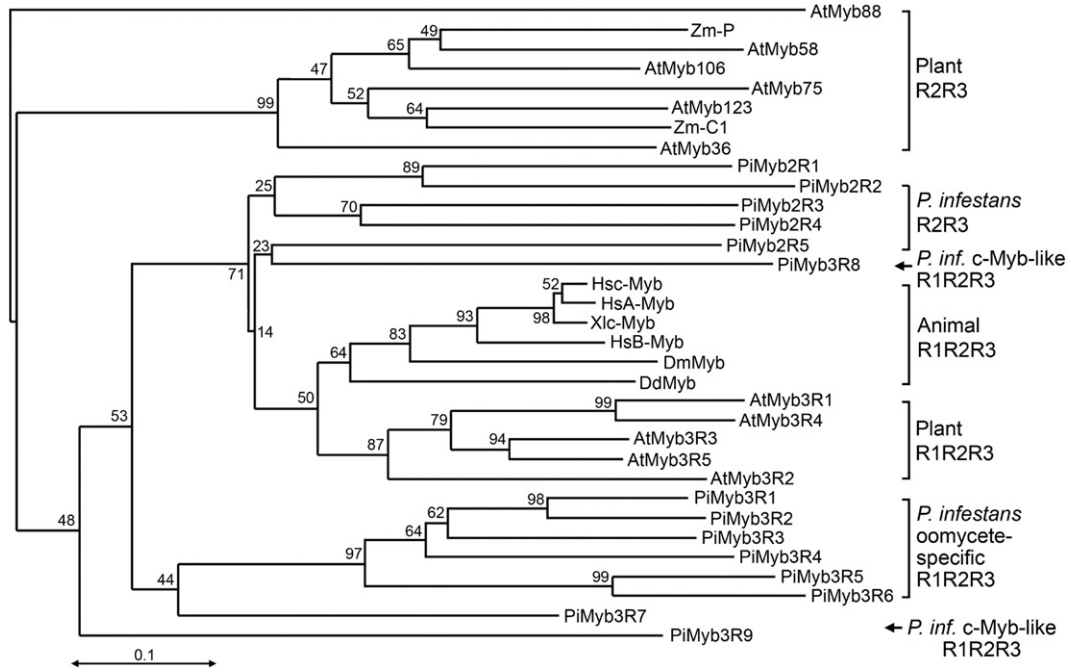

Fig. 4. Neighbor-joining tree of R2R3 and R1R2R3 proteins. This is based on the R2R3 domains of the protein. Values at nodes indicate their percent occurrence in 1000 bootstrap replicates. The tree was generated using the proteins shown in Fig. 2, supplemented with additional *A. thaliana* proteins and Zm-C1 from maize (Genbank accession M62879). Similar topologies were observed in maximum parsimony and minimum evolution trees. PiMybR5, the atypical *P. infestans* protein with five Myb-like domains, is not shown in this figure. However, when included it fell into a clade adjacent to PiMyb3R8, similar to the relationship portrayed in Fig. 1.

P. infestans R2R3 proteins lack strong affinity with plant R2R3 proteins; this is illustrated by the neighbor-joining tree in Fig. 4. Instead, the *P. infestans* R2R3 proteins cluster with R1R2R3 lineages. This suggests that the evolution of the *P. infestans* R2R3 proteins involved a loss of the R1 domain that occurred independently of its elimination from the plant lineage. It is not obvious which current R1R2R3 protein from *Phytophthora* is their closest ancestor. While in each of the three tree-building strategies PiMyb3R8 clustered with *P. infestans* R2R3 protein PiMyb2R5, the bootstrap values were low. It is possible that the R1R2R3 lineage that gave rise to the *P. infestans* R2R3 proteins may have been lost from the genome.

One surprising observation was that the three *P. infestans* R1R2R3 proteins in the c-Myb-like group did not cluster with each other in Fig. 4. In contrast, R1R2R3 proteins from most other taxonomic groups fall into tight clades. This reflects greater diversity in *P. infestans* within most of the helices in the R2 and R3 domains. For example, while the three human R1R2R3 proteins are identical at 13 of 14 positions within helix 1 of domain R2, the three *P. infestans* proteins are conserved at only 2 of the 14 sites. This presumably means that the c-Myb-like proteins of *P. infestans* may show greater diversification in their DNA-binding specificities.

4. Concluding remarks

This study has shown that compared to animals, Myb transcription factors are highly diversified in oomycetes. Oomycetes contain R2R3 proteins in addition to the standard complement of R1R2R3 proteins. The oomycete R1R2R3 family also comprises two groups, one with DNA-binding helices that resemble c-Myb plus a novel group. In contrast, R1R2R3 proteins in plants or animals are usually monophyletic with DNA-binding domains that are similar to c-Myb.

The diversification of the DNA-binding domains in oomycete Myb proteins should alter their DNA specificities, based on mutation studies in other species and EMSA data for selected *P. infestans* proteins. Besides differences in the four amino acids shown to play key roles in the binding of mammalian c-Myb to DNA, variation also exists in other R2R3 sites that are well-conserved in other taxa, and in helix spacing. Most oomycete R2R3 and R1R2R3 proteins are also highly dissimilar outside their DNA-binding domains. By comparison, the three human R1R2R3 proteins have nearly identical Myb domains but distinct target specificities, due to differences elsewhere in the proteins and post-translational modifications (Ness, 2003; Lei et al., 2004). It follows that the diversification existing within and outside the DNA-binding regions of the *P. infestans* Myb proteins would cause them to regulate discrete classes of genes during the life cycle.

Roles of Myb proteins in the life cycle are also suggested by the finding that most R2R3 and R1R2R3 proteins of *P. infestans* are differentially expressed during spore formation or germination. Nearly all R2R3 proteins are predicted to be specific to those stages, which resembles the situation in plants where R2R3 factors control many plant-specific processes (Jin and Martin, 1999; Ito, 2005). Functions of Myb proteins in spores are supported further by the finding that *H. arabidopsidis*, which lacks the zoospore stage, contains fewer of these transcription factors than *Phytophthora.* Several classes of sporulation-induced promoters also contain c-Myb-like recognition sequences and bind at least some of the Myb transcription factors (Ah Fong et al., 2007; Xiang et al., 2009). Signaling pathways involving Myb proteins may therefore be useful targets for chemical-based or other strategies to control these devastating pathogens.

Acknowledgments

This work was supported by a grant from the U.S. Department of Agriculture-Cooperative State Research, Education, and Extension Service to H.S.J. We thank the Department of Energy's Joint Genome Institute, the Virginia Bioinformatics Institute, and the Broad Institute of MIT and Harvard for providing access to their sequence databases.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2009.12.006.

References

Adl, S.M., et al., 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399–451.

Ah Fong, A., Xlang, Q., Judelson, H.S., 2007. Architecture of the sporulation-specific Cdc14 promoter from the oomycete *Phytophthora infestans*. Eukaryot. Cell 6, 2222–2230.

Anton, I.A., Frampton, J., 1988. Tryptophans in myb proteins. Nature 336, 719.

Beug, H., von Kirchbach, A., Doderlein, G., Conscience, J.F., Graf, T., 1979. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. Cell 18, 375–390.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E., Klempnauer, K.H., 1988. Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature 335, 835–837.

Deng, Q.L., Ishii, S., Sarai, A., 1996. Binding site analysis of c-Myb: screening of potential binding sites by using the mutation matrix derived from systematic binding affinity measurements. Nucleic Acids Res. 24, 766–774.

Dias, A.P., Braun, E.L., McMullen, M.D., Grotewold, E., 2003. Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. Plant Physiol. 131, 610–620.

Ehrenkaufer, G.M., Hackney, J.A., Singh, U., 2009. A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in *Entamoeba histolytica*. Cell. Microbiol. 11, 898–910.

Gonda, T.J., Gough, N.M., Dunn, A.R., de Blaquiere, J., 1985. Nucleotide sequence of cDNA clones of the murine myb proto-oncogene. EMBO J. 4, 2003–2008.

Howe, K.M., Reakes, C.F., Watson, R.J., 1990. Characterization of the sequence-specific interaction of mouse c-myb protein with DNA. EMBO J. 9, 161–169.

Hulo, N., et al., 2008. The 20 years of PROSITE. Nucleic Acids Res. 36, D245–249.

Ito, M., 2005. Conservation and diversification of three-repeat Myb transcription factors in plants. J. Plant Res. 118, 61–69.

Jiang, C., Gu, J., Chopra, S., Gu, X., Peterson, T., 2004. Ordered origin of the typical two- and three-repeat Myb genes. Gene 326, 13–22.

Jin, H., Martin, C., 1999. Multifunctionality and diversity within the plant MYB-gene family. Plant Mol. Biol. 41, 577–585.

Judelson, H.S., Blanco, F.A., 2005. The spores of *Phytophthora:* weapons of the plant destroyer. Nat. Microbiol. Rev. 3, 47–58.

Judelson, H.S., et al., 2008. Gene expression profiling during asexual development of the late blight pathogen *Phytophthora infestans* reveals a highly dynamic transcriptome. Mol. Plant-Microbe Interact. 21, 433–447.

Judelson, H.S., Narayan, R.D., Ah-Fong, A.M., Kim, K.S., 2009. Gene expression changes during asexual sporulation by the late blight agent *Phytophthora infestans* occur in discrete temporal stages. Mol. Genet. Genome 281, 193–206.

Karamysheva, Z.N., Surovtseva, Y.V., Vespa, L., Shakirov, E.V., Shippen, D.E., 2004. A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in Arabidopsis. J. Biol. Chem. 279, 47799–47807.

Latchman, D.S., 2003. Eukaryotic Transcription Factors. Academic Press, San Diego.

Lei, W., Rushton, J.J., Davis, L.M., Liu, F., Ness, S.A., 2004. Positive and negative determinants of target gene specificity in myb transcription factors. J. Biol. Chem. 279, 29519–29527.

Lipsick, J.S., 1996. One billion years of Myb. Oncogene 13, 223–235.

McDonald, W.H., Ohi, R., Smelkova, N., Frendewey, D., Gould, K.L., 1999. Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. Mol. Cell. Biol. 19, 5352–5362.

Nakazawa, T., et al., 2008. Basidiomycete *Lentinula* edodes CDC5 and a novel interacting protein CIPB bind to a newly isolated target gene in an unusual manner. Fungal Genet. Biol. 45, 818–828.

Ness, S.A., 2003. Myb protein specificity: evidence of a context-specific transcription factor code. Blood Cells Mol. Dis. 31, 192–200.

Ogata, K., et al., 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell 79, 639–648.

Ong, S.J., Hsu, H.M., Liu, H.W., Chu, C.H., Tai, J.H., 2007. Activation of multifarious transcription of an adhesion protein ap65-1 gene by a novel Myb2 protein in the protozoan parasite *Trichomonas vaginalis*. J. Biol. Chem. 282, 6716–6725.

Ording, E., Bergholtz, S., Brendeford, E.M., Jamin, N., Gabrielsen, O.S., 1996. Flexibility in the second half-site sequence recognised by the c-Myb R2 domain in vitro and in vivo analysis. Oncogene 13, 1043–1051.

Otsuka, H., Van Haastert, P.J., 1998. A novel Myb homolog initiates *Dictyostelium* development by induction of adenylyl cyclase expression. Genes Dev. 12, 1738–1748.

Ramsay, R.G., Gonda, T.J., 2008. MYB function in normal and cancer cells. Nat. Rev. Cancer 8, 523–534.

Robertson, D.L., Tartar, A., 2006. Evolution of glutamine synthetase in heterokonts: evidence for endosymbiotic gene transfer and the early evolution of photosynthesis. Mol. Biol. Evol. 23, 1048–1055.

Rose, A., Meier, I., Wienand, U., 1999. The tomato I-box binding factor LeMYBI is a member of a novel class of myb-like proteins. Plant J. 20, 641–652.

Rubio-Somoza, I., Martinez, M., Diaz, I., Carbonero, P., 2006. HvMCB1, a R1MYB transcription factor from barley with antagonistic regulatory functions during seed development and germination. Plant J. 45, 17–30.

Saikumar, P., Murali, R., Reddy, E.P., 1990. Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. Proc. Natl. Acad. Sci. U. S. A. 87, 8452–8456.

Solano, R., Fuertes, A., Sanchez-Pulido, L., Valencia, A., Paz-Ares, J., 1997. A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor MYB.Ph3 to the animal c-MYB specificity. J. Biol. Chem. 272, 2889–2895.

Stracke, R., Werber, M., Weisshaar, B., 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. 4, 447–456.

Sun, C.H., Palm, D., McArthur, A.G., Svard, S.G., Gillin, F.D., 2002. A novel Myb-related protein involved in transcriptional activation of encystation genes in *Giardia lamblia*. Mol. Microbiol. 46, 971–984.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Tani, S., Judelson, H.S., 2006. Activation of zoosporogenesis-specific genes in *Phytophthora infestans* involves a 7-nucleotide promoter motif and cold-induced membrane rigidity. Eukaryot. Cell 5, 745–752.

Timmermans, M.C., Hudson, A., Becraft, P.W., Nelson, T., 1999. ROUGH SHEATH2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia. Science 284, 151–153.

Tsiantis, M., Schneeberger, R., Golz, J.F., Freeling, M., Langdale, J.A., 1999. The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. Science 284, 154–156.

Wieser, J., Adams, T.H., 1995. *flbD* encodes a Myb-like DNA-binding protein that coordinates initiation of *Aspergillus nidulans* conidiophore development. Genes Dev. 9, 491–502.

Williams, C.E., Grotewold, E., 1997. Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities. J. Biol. Chem. 272, 563–571.

Xiang, Q., Roy, S., Kim, K.S., Judelson, H.S., 2009. A motif within a complex promoter from the oomycete *Phytophthora infestans* determines transcription during an intermediate stage of sporulation. Fungal Genet. Biol. 46, 400–409.