Cell, Vol. 60, 295–306, January 26, 1990, Copyright © 1990 by Cell Press

The *b* Alleles of U. maydis, Whose Combinations Program Pathogenic Development, Code for Polypeptides Containing a Homeodomain-Related Motif

Burkhard Schulz,* Flora Banuett,† Marlis Dahl,* Ramona Schlesinger,* Willi Schäfer,* Thomas Martin,* Ira Herskowitz,† and Regine Kahmann*
* Institut für Genbiologische Forschung Berlin GmbH
Ihnestrasse 63
D-1000 Berlin 33
Federal Republic of Germany
† Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143

Summary

U. maydis is a fungal pathogen of corn with two forms: one is yeast-like and nonpathogenic; the other is filamentous and pathogenic. The *b* locus, with 25 different alleles, regulates this dimorphism: any combination of two different alleles triggers pathogenic development, whereas the presence of identical alleles results in the yeast-like form. We have cloned four *b* alleles (*b1, b2, b3,* and *b4*) and show that the *b* locus contains a single open reading frame (ORF) of 410 amino acids with a variable N-terminal region and a highly conserved C-terminal region (60% and 93% identity, respectively). Mutational analysis confirms that this ORF is responsible for *b* activity. The *b* polypeptides appear to be DNA binding proteins because they contain a motif related to the homeodomain in their constant region. We propose that combinatorial interactions between *b* polypeptides generate regulatory proteins that determine the developmental program of the fungus.

Introduction

Ustilago maydis belongs to the large group of plant pathogenic Basidiomycetes known as the smut fungi and is the causative agent of corn smut disease (for reviews see Christensen, 1963; Banuett and Herskowitz, 1988). Two distinct forms characterize its life cycle: a unicellular haploid form, which divides by budding, produces yeast-like colonies on defined media, and is nonpathogenic; and a filamentous dikaryotic form, whose growth is dependent on the plant and which is pathogenic, causing tumors on leaves, stems, tassels, and ears (Christensen, 1963). This dimorphism is governed by two loci, *a* and *b*, the incompatibility or mating type loci (Rowell and DeVay, 1954; Rowell, 1955; Holliday, 1961). Development of the pathogenic dikaryon is initiated after cell fusion of two haploids that differ at both loci (Rowell and DeVay, 1954; Rowell, 1955; Holliday, 1961; Puhalla, 1970; Day et al., 1971). Formation of the filamentous dikaryon can be assayed on nutrient medium containing charcoal: a mixture of *a1 b1* and *a2 b2* haploids yields a characteristic white fuzziness (the Fuz$^+$ phenotype) due to the formation of dikaryotic filaments, whereas a mixture of haploids containing identical *a* or identical *b* alleles does not yield the filamentous form (Banuett and Herskowitz, 1989).

The *b* locus poses a striking puzzle because there are estimated to be 25 naturally occurring alleles (Rowell and DeVay, 1954; Puhalla, 1968): the filamentous program results if the *b* alleles are different but not if they are the same. The *a* locus has only two alleles (Rowell and DeVay, 1954; Holliday, 1961; Puhalla, 1968), *a1* and *a2*, and different *a* alleles are necessary, together with different *b* alleles, for the development of the filamentous form (Banuett and Herskowitz, 1989). Studies on fruiting body formation by other Basidiomycetes, Schizophyllum commune and Coprinus cinereus, have raised the same question of self-nonself recognition. In these organisms, fruiting body formation requires the presence of different alleles at two different incompatibility loci, both of which have multiple alleles (Raper, 1983; Casselton, 1978).

A variety of molecular explanations have been proposed for the mechanism of recognition of identical versus nonidentical alleles of the Basidiomycete incompatibility loci (Kuhn and Parag, 1972; Ullrich, 1978). According to some schemes, the functional products of the *b* alleles are nucleic acids, either DNA or RNA, and recognition involves formation of heteroduplex structures. According to another class of scheme, the *b* alleles code for polypeptides that combine to form multimeric proteins. A broad variety of speculations have also been made as to the function of such a *b* locus protein. The *b* polypeptides are unlikely to comprise cell surface proteins because monitoring the status of *b* occurs intracellularly (Banuett and Herskowitz, 1989). Thus it has been suggested that the *b* polypeptides may be nuclear membrane or cytoskeletal proteins involved in nuclear behavior (Banuett and Herskowitz, 1989), or they may be proteins involved in regulating gene expression (Banuett and Herskowitz, 1988).

In this paper we show that the *b* locus contains an open reading frame (ORF) of 410 amino acids that is responsible for *b* activity. This finding argues strongly against the nucleic acid models. Comparisons of the ORFs of four different alleles (*b1, b2, b3,* and *b4*) show that each contains a variable domain in the N-terminal 110 amino acids (40% differences) and that the rest of the ORF is highly conserved (93% identity). The amino acid sequence of the constant domain provides a strong clue to the function of the polypeptides encoded by the *b* alleles: it contains a motif related to the homeodomain, a hallmark for a DNA binding protein. Thus, we propose that the *b* alleles of U. maydis code for polypeptides whose association yields a regulatory protein that governs the developmental program and pathogenicity of this organism.

Results

Characterization of *b* Alleles

The classical genetic work of Rowell and DeVay (1954) on the *b* locus demonstrated that this locus is multiallelic and, furthermore, that the presence of different *b* alleles

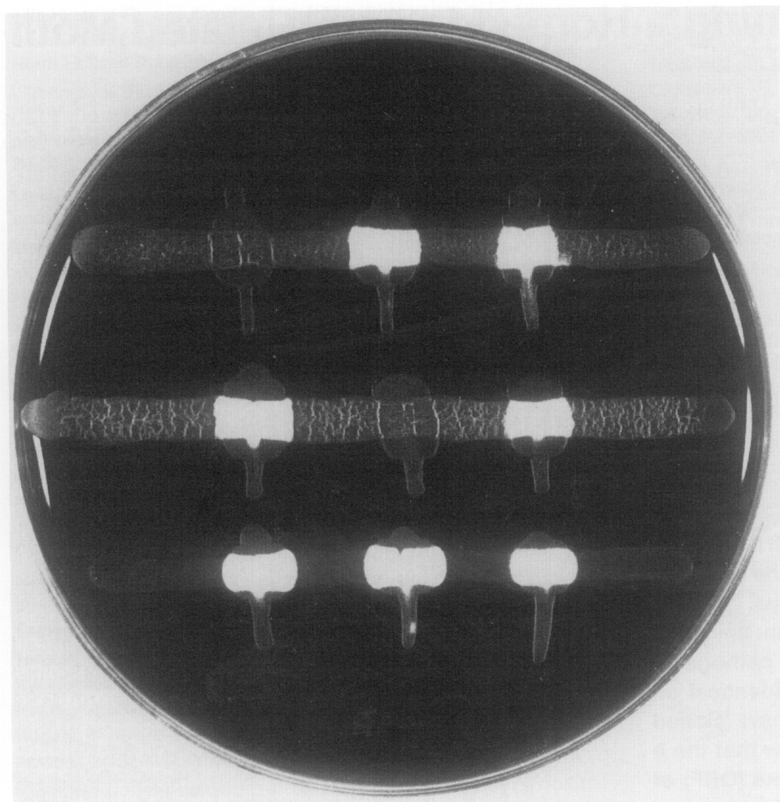


Figure 1. The Presence of Different *b* Alleles Triggers Development of the Filamentous Form of U. maydis

Saturated cultures of haploid strains containing the *b1*, *b2*, or *b4* allele, respectively (horizontal rows), were cross-streaked against haploid strains containing the *b1*, *b2*, or *b3* allele (vertical columns) on charcoal medium and incubated at room temperature for 24 hr. The white fuzziness is due to the formation of dikaryotic filaments after fusion of the haploid strains. Only strains carrying different *b* alleles (and different *a* alleles) exhibit this fuzz reaction (the Fuz^+ phenotype). Strains in the horizontal rows are (from top to bottom) *a2 b1* (FB6a), *a2 b2* (FB2), and *a1 b4* (RK138). Strains in the vertical columns (from left to right) are *a1 b1* (FB1), *a1 b2* (FB6b), and *a2 b3* (RK32).

is necessary for the formation of galls (tumors) and the production of diploid spores (teliospores) in these galls. Other studies (Holliday, 1961; Puhalla, 1970; Day et al., 1971; Banuett and Herskowitz, 1989) have subsequently confirmed these observations. Our studies on the *b* locus use four different alleles: two of which (*b1* and *b2*) have been extensively used by Holliday (1961, 1974), and two of which (*b3* and *b4*) were newly isolated from nature. Figure 1 demonstrates a plate assay used for scoring the *b* alleles and shows that *b1*, *b2*, *b3*, and *b4* are functionally distinct from each other. When strains carrying different *b* alleles are cross-streaked or cospotted on charcoal nutrient medium, a white fuzziness develops due to the formation of dikaryotic filaments (see Banuett and Herskowitz, 1988). Thus, when a strain carrying the *b1* allele is cross-streaked against a strain carrying the *b2* allele (Figure 1, top row, middle column), a white patch is observed at the intersection (the Fuz^+ phenotype). If the strains carry the same *b* allele (for example *b1*), no filaments develop (the Fuz^- phenotype) (Figure 1, top row, left column). Strain RK32, isolated from nature, is capable of inducing filament formation with both *b1* (Figure 1, top row, right column) and *b2* strains (Figure 1, middle row, right column), indicating that it carries a different *b* allele. We designate the *b* allele present in this strain *b3*. Strain RK138 (Figure 1, bottom row) is capable of forming filaments in combination with *b1*, *b2*, and *b3* (left, middle, and right columns, respectively). Thus, we conclude that the *b* allele present in RK138 is distinct from *b1*, *b2*, and *b3* and designate it *b4*. Assay of tumor formation after inoculation of plants with different combinations of these strains confirms these deductions (data not shown).

In the next section we describe the cloning of the *b2* allele by a functional assay and then describe the cloning of the other alleles by nucleic acid hybridization.

Cloning the *b2* Allele by Functional Assay

The strategy for cloning the *b* locus exploits the behavior of diploids homozygous at the *b* locus (Banuett and Herskowitz, 1988). Such diploids (for example, strains that are *a1/a2 b1/b1*) form yeast-like colonies on charcoal nutrient medium (Fuz^-) and are nonpathogenic (Tum^-), in contrast to *a1/a2 b1/b2* diploids, which form mycelial colonies (Fuz^+) and are pathogenic (Tum^+) (Day et al., 1971; Banuett and Herskowitz, 1989). Hence, we transformed an *a1/a2 b1/b1* diploid (FBD12-3) with a genomic library from an *a2 b2* strain (strain 518) and screened for Fuz^+ transformants. A similar strategy was used by Kronstad and Leong (1989) to clone the *b1* allele. Because *b2* resides on an 8 kb BamHI fragment (Kronstad and Leong, 1989), we constructed a genomic library of size-fractionated 8 kb BamHI fragments of U. maydis strain 518 (*a2 b2*) in pHLN (a vector conferring hygromycin B resistance [Hyg^r]). Plasmid DNA from this library was linearized with NotI and used to transform strain FBD12-3. Of 3500 Hyg^r transformants screened on charcoal nutrient medium, five exhibited a Fuz^+ phenotype. Each of these Fuz^+ transformants caused the development of tumors when a pure culture was inoculated into plants, indicating that they contained a functional *b2* allele.

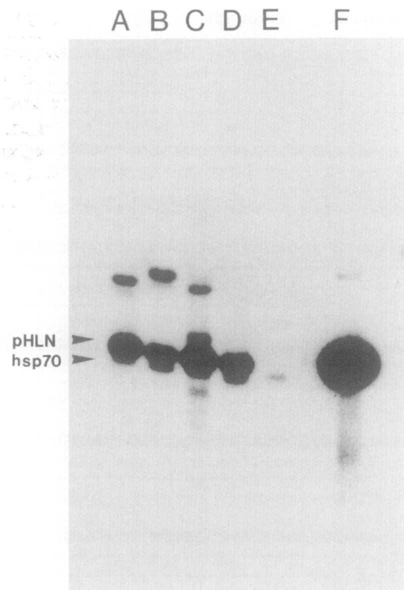


Figure 2. Analysis of Sites of Plasmid Integration in Fuz$^+$ Transformants by Southern Hybridization

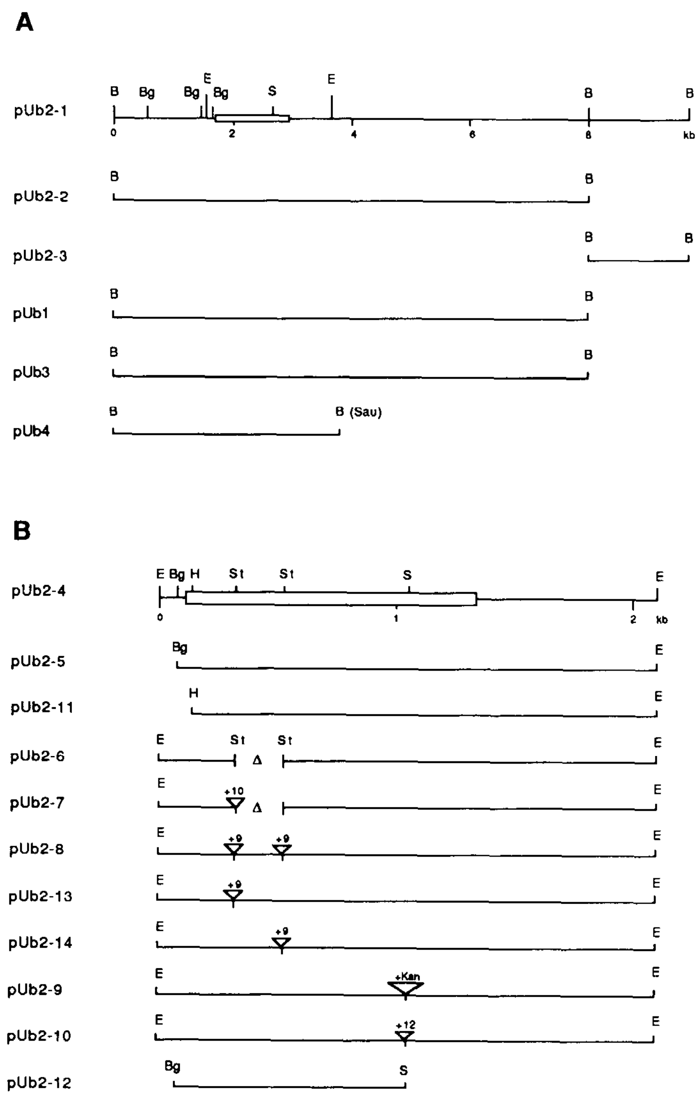
Strains 7-14, 8-15, 8-31, and 6-21 are Fuz$^+$ transformants obtained after introduction of a *b2* library constructed in plasmid pHLN into the Fuz$^-$ recipient strain FBD12-3 (*a1/a2 b1/b1*). Lanes A–D contain BamHI-digested DNA of the four independent transformants 7-14, 8-15, 8-31, and 6-21, respectively. Lane E contains BamHI-digested DNA of U. maydis strain FBD12-3; lane F contains linearized plasmid pHLN. pHLN DNA was used as probe for detection of integrated plasmid. The 6 kb pHLN and the 5 kb *hsp70* fragments are indicated.

We determined by Southern hybridization of transformants whether plasmid integration had occurred at homologous or nonhomologous sites. Chromosomal DNA was prepared from four of the transformants and from the recipient strain, cleaved with BamHI, size fractionated, and hybridized with pHLN DNA after transfer to a nylon filter (Figure 2). All four transformants contain a very prominent signal at 6 kb, the size of the linearized vector pHLN. In DNA of transformants 7-14, 8-15, and 8-31 (Figure 2, lanes A, B, and C, respectively) one or two additional signals that differ in size are observed. We consider these fragments to represent junctions between the vector and U. maydis DNA. These three transformants appear to have been formed by integration of the plasmid at nonhomologous sites. The absence of new junction fragments in transformant 6-21 (Figure 2, lane D) indicates that in this transformant, plasmid integration has occurred by homologous recombination at or near the *b1* locus. In all cases the 6 kb vector signal is overrepresented, suggesting that several copies of the transforming DNA are integrated into the same site. The weak signal at 5 kb seen in all transformants and in the untransformed recipient strain (Figure 2, lane E) can be attributed to *hsp70* sequences (pHLN contains about 2 kb of *hsp70* DNA; Wang et al., 1988).

Recovery of the plasmid responsible for the Fuz$^+$ phenotype was accomplished by partial digestion of chromosomal DNA of transformant 6-21 (Figure 2, lane D) with BamHI followed by ligation and transformation of the ligated DNA into Escherichia coli strain DH5α*mcr* (see Experimental Procedures). While most transformants contained plasmids indistinguishable from pHLN, several transformants carried a plasmid containing two BamHI fragments, one 8 kb and the other approximately 1.7 kb. The NotI site was retained, substantiating the assertion that several copies of the transforming DNA had integrated into the same site. A plasmid indistinguishable from the one recloned from the Fuz$^+$ transformant 6-21 was isolated from our gene bank when we used the 8 kb BamHI fragment as probe in colony hybridization. This plasmid, which never had the opportunity to recombine with the genome of U. maydis, was termed pUb2-1 and was used for all subsequent studies. We have not determined whether the 1.7 kb BamHI fragment of pUb2-1 is adjacent to the 8 kb fragment in the genome or represents a cloning artifact. When pUb2-1 was transformed into a U. maydis *a1/a2 b1/b1* strain (FBD12-3), approximately 70% of the transformants exhibited the Fuz$^+$ phenotype (Figure 3), and all (in this case eight) of the Fuz$^+$ transformants tested were able to induce tumors when inoculated into the plant. All features of pUb2-1 indicate that it carries a functional *b2* allele; in particular, it allows an *a1/a2 b1/b1* strain to exhibit both properties of an *a1/a2 b1/b2* strain: filamentous growth on charcoal nutrient medium and tumor induction in plants.

The finding that multiple copies of *b2* (approximately three to five in the individual original transformants) can be inserted without noticeable effects on the fuzz reaction and pathogenicty suggests, furthermore, that the relative gene dosage of *b2* and *b1* is, within limits, not critical. We have consistently observed that in all instances where an intact *b* allele is introduced into U. maydis, only 70%–80% of the transformants display a Fuz$^+$ phenotype. We have not analyzed this phenomenon but suggest that failure of some transformants to exhibit the Fuz$^+$ phenotype may be due to position effects or to rearrangements occurring during the integration event.

To delimit the boundaries of the *b2* allele on plasmid pUb2-1 (Figure 3), we generated a series of subclones, which were then tested for their properties after introduction into an *a1/a2 b1/b1* strain. These transformants were examined for Fuz$^+$ phenotype and ability to induce tumors after inoculation of plants (Figure 3). The *b2* activity clearly resides on the 8 kb BamHI fragment as pUb2-2 confers a Fuz$^+$ and Tum$^+$ phenotype, whereas pUb2-3 does not. Plasmids containing a 2.1 kb EcoRI as well as a 2.0 kb BglII–EcoRI fragment displayed full *b2* activity (pUb2-4 and pUb2-5, respectively), whereas a plasmid carrying the BglII–SalI fragment (pUb2-12) showed significantly reduced numbers of Fuz$^+$ transformants. In this case only 11% of the transformants were Fuz$^+$ Tum$^+$, in contrast to an average of 80% for transformants carrying a plasmid with an intact *b2*. The HindIII–EcoRI fragment in pUb2-11, on the other hand, yielded only Fuz$^-$ transformants. These results allowed us to place the left border of the functional *b2* allele between the BglII and HindIII sites, and the right border was expected to lie close to the SalI site (Figure 3B).



	No. of TF	Phenotype Fuz (%) +	-	?	Tum
A					
pUb2-1	60	70	12	18	+
pUb2-2	49	76	12	12	+
pUb2-3	56	<2	100	<2	-
pUb1	85	80	20	<1	+
pUb3	5	80	20	<20	+
pUb4	6	83	<17	17	+
B					
pUb2-4	81	82	11	7	+
pUb2-5	79	76	14	10	+
pUb2-11	40	<3	83	17	-
pUb2-6	66	<2	94	6	-
pUb2-7	51	<2	100	<2	-
pUb2-8	55	<2	98	2	-
pUb2-13	37	<3	100	<3	-
pUb2-14	39	<3	100	<3	-
pUb2-9	54	69	31	<2	+
pUb2-10	41	68	32	<2	+
pUb2-12	38	11	84	5	+

Figure 3. Structure of Plasmids Carrying *b2*, *b3*, *b4*, and *b1* and Mutations of *b2* Constructed In Vitro

(A) shows the structure of the original plasmid isolates, and (B) shows the structure of an intact *b2* gene (carried in plasmid pUb2-4) and a variety of mutations constructed in vitro. Only restriction sites used for constructing certain subclones are indicated: B (BamHI), Bg (BglII), E (EcoRI), S (SalI), Sau (Sau3A), H (HindIII), St (StuI). Triangles mark deletions. Insertions have triangles with (+) numbers indicating their length in base pairs. +Kan denotes the insertion of the kanamycin cassette. The ORF of *b2* is shown as an open rectangle. Different scales are used in (A) and (B) as indicated in the top line of each panel.

The right-hand panel summarizes the phenotypes of plasmid constructs after transformation into an *a1/a2 b1/b1* U. maydis strain (FBD12-3), except for pUb1, which was transformed into an *a1/a2 b2/b2* strain (FBD11-21). No. of TF indicates the number of independent transformants that were analyzed. The Fuz phenotype was assayed on charcoal nutrient medium; a plus sign indicates formation of filaments, a minus sign indicates the absence of filaments, and a question mark indicates that the phenotype was ambiguous. Numbers refer to the percent of transformants that exhibited the phenotype. The ability to form tumors when inoculated as a pure culture into plants (six plants were inoculated per culture) was assessed as follows: for transformants displaying a Fuz$^+$ phenotype, cultures of four to six purified colonies derived from independent transformants were tested, and without exception, tumor induction (Tum$^+$) was observed. Fuz$^-$ transformants from the same series were unable to induce tumors, whereas among the Fuz(?) transformants from the same series, we found tumor-inducing as well as nontumor-inducing strains, in a ratio of about 50:50. Tum$^-$ indicates that neither the Fuz$^-$ nor the Fuz(?) transformants obtained with the particular plasmid were able to induce tumors. In these cases all Fuz(?) transformants and four to six Fuz$^-$ transformants were tested in planta.

The *b2* Locus Contains an ORF Essential for *b2* Activity

The nucleotide sequence of *b2* was determined from the leftmost EcoRI site across the SalI site (see Experimental Procedures) (Figure 4). Translation of the sequence revealed a single ORF of 410 amino acids initiating at a methionine codon located between the BglII and HindIII sites, which we had previously inferred to contain the left border of *b2* (Figure 5). The ORF extends for 1230 bp across the SalI site. The deletion in pUb2-11 removes the proposed ATG initiation codon, which would account for its inability to provide *b2* function. The deletion in pUb2-12 has re-

```
      -160                          -130                          -100                          -70
b2  TGATGGAAAACAAAGAACCTGTGCTGAGGCCGAATT*CGTCTGC*AAAATTCC*TTCTTTGTGCACACA**TTCATTT*CTTT***GCGCCGACACCGCT
b3  G**---G--------------C----G---------T----GA-G-------T*--GC--**-------CA----****---CCTC------------**
b4  G**C--G-------------C-------------------G--G-----C-T-*-----**------GCA---------T--****-----T-------A
b1  --G---GG-------------C------C-------T----A--*----A---T---C---CA------**-------*-C--***------------A-
                                                                       +1                             30
b2  CTT**CAACCGCAC*TCCCACCTTTAGCCTCTAACAAAAGATCTTCTATCTCTGCAACATCATCACAATGTCTAACTACCCGAACTTTTCCCTCACAAGC
b3  *--CC--------TT---------CG------T-----*------T-C-----A-------------ATG---GAG--------GC----G---T---AG
b4  *--**---------T---------C-------T-G---*------T-C-----T---T----G---CATG--C-G-G-------GA----GA--------
b1  TC***---TG-T--*-T-AC----CG------T----CGA-*---T-C----A--------------ATG--C-G-G-------T--C--G----T----
                                                                                                     130
b2  TTTGTGGAGTGCCTCAACGAGATCGAGCACGAGTTTCTGCGAGATAAATTGGAAAATCGTCCCGTTCTCGTGCGCAAACTTCAAGAGCTGCGACGAAAGA
b3  --CT--------------------------------------------G-----C-------T----------------------------AG-A-----
b4  --CT------------G------T---G-A------------------GA-A------C---------G---A------------------A--A-----
b1  --CT--------------------------------------------GG---G---TAC--------------T-----C-G--------A--A-----
                                                                                                     230
b2  CACCCAACAACGTTGCAAGCCTATCCTACGATCCGGGAACGATCCATCAAATACACCAGACTACTCACAGGATTAAAGTCGCTGCCAAGGCCTTTATACG
b3  -------AC----A----AA---C--C----G----A--T------G-----------AG--G-A---------G-CA---T----C----A--------
b4  -----------A------A----GA-C----C--C-A---------AA----------------A--------C--T--T--C-T-------------T-
b1  T-------G--A------A----C--CG------C-A---------G-----T-----------A----------G--C--T----C----A--------
                                                                                                     330
b2  TATCGACCAGTCATTCGTCTCTCTGCATTCCGACGCTGTTGAAGACACAAGCAAAGCCTTGAAGAAAGCCGATGCTTCAAGTCCTGTTGTAGGGTGTCGC
b3  -T--------AAG------------TG------GATC---C-C--T--G-C-----T-A--C--G-G---A---T-GTC-------G--A---------A
b4  ----------A---------------G------T--C---------G-G----G----------------A-------------------------C---
b1  -T--------AAG------------TG------G-TC---C-C-G---G-------T-A--C--G-GTT-A---TCGT--------A------------A
                                                                                                     430
b2  GACCTTTCTGAGGATCTTCCAGCGTACCACATGCGGAAACACTTTCTCCATACTCTCGACAACCCTTATCCGACTCAAGAAGAGAAAGAAGGCCTCGTTC
b3  A-------G--A--------T-----------------T--------------------G-----------A---------------------------A
b4  --A-----A--------------------------A--G--------------------A-----A------------------------AC-------A
b1  A-T-----G-----------T-----T--------T-------------T----------------------------------------AC-------A
                                                                                                     530
b2  GACTCACCAACGAATCCACAGCAAGAGTTGGCCTATCCAAAGCAAATCGTCCTCCGCTCGAGGTTCATCAGCTCACGCTATGGTTCATTAACGCACGCCG
b3  --------------------------A--C---C------C----T------------------------------------------C-----------
b4  -GT-G----------------------------AG----TC-T---------------------------------------------------------
b1  -GT-G----------------------------AG----GC-T---------------------------CG----------------------------
                                                                                                     630
b2  TCGTTCTGGATGGTCTCATATTCTGAAGAAGTTCGCACGCGAGGATCGTTCACGGATGAAACATCTGGTACGGGCCAAGCTGTCTTCATCAAACCAGTCT
b3  ------------------------C----------------------CG-----------G-G---C-----------------A--------T------
b4  ------------------------------------------------------------------------------------C--T-----T------
b1  ----------------------------------------------------------------------------------------------------
                                                                                                     730
b2  ACTCCTCCTTCATCGACGTCTGATTCTCTCTCAAACAATCTAGACGACGTTCTATCAGACAACCTCGGACGACCACTCACACCTGTTGACAAGCAGCAGT
b3  --C---------CT---------AAAG-C----G--G--------T----------------T-------------------T--C------A-----A-
b4  ------------C----------A-A--C---------------T---T------------------------------------C------------A-
b1  -------------------A--------------------------------------------------------------------------------
                                                                                                     830
b2  TCGAAGATGACTGGGCAAGCATGATCAGCTGGATCAAGTATGGCGTGAAAGAAAAAGTTGGAGACTGGGTCTATGATCTCTGCGCAGCGAGCAAGAAGAC
b3  ----------------------------------------------------------------------------------------------------
b4  -T--------------------------------------------------------C-----------------------------------------
b1  ----------------------------------------------------------------------------------------------------
                                                                                                     930
b2  TCCGAAGCCTGGCATGCCACGTCCCGTCACCACAGTAGCGAAGCGCCATCCAGCGCGCAAAACTAAACCAGCCGCAAAACCAAAGTCAAGAACAGCGAAC
b3  ----------------------------------------------------------------------------------------------------
b4  ------------------------------T-----------------G-----------G---------------------------------------
b1  ------------------------------------------C--G------------------------------------------------------
                                                                                                    1030
b2  CCGCGAGCAAGCACAACACCTTCCATTGACAGCACCTTGGACAGTTCAAAGCTAGAGTCGACTCCAGAACTCAGCATGTGTTCGACGGCGGACACGAGCT
b3  ----------------------------------------------------------------------------------------------------
b4  ----------------------------------------------------------------------------------------------------
b1  --------------------------------------------------------------------------------------A-------------
                                                                                                    1130
b2  TCAGTACATTCGGGAGCAGTCTCTCGATGTCTCACTACAATCCCTTTCAGGATGGAAACGATATTCTACAGAGTCCGACAGTCAAAGCCAGGGGCAATAG
b3  ----------------------------------------------------------------------------------------------------
b4  --------------------------------------G-----------T-----------------------------T-------------------
b1  -------------------------A--------------------------------------------------------------------------
                                                                                                    1230
b2  GAAAGTCAAAGCGCTTCCGAAGCGGGCTGGAAAGCAGCAGCCTGACGAGGTCGAAAATGGTAAGATTCCTTTCCTTTGCCTTTCCGTTGCGTTCGTTTGA
b3  ------------------------------------------------------C------------------------------------------TGA
b4  -------------------------------------------------------------------------------------------------TGA
b1  ------------------------------------------------------C------------------------------------------TGA
                                                  1284
b2  CTAAACAGACACAATTCACATTCTGCCACACAGACAAACAGGAAAAATTTTC
b3  ----------------------------------------------------
b4  -----------------------------------------A---TT-CGA-
b1  ----------------------------------------------------
```

Figure 4. Nucleotide Sequences of *b2*, *b3*, *b4*, and *b1* Alleles

The nucleotide sequence of the *b2* allele is displayed in the upper line. For alleles *b3*, *b4*, and *b1* dashes are used to denote nucleotide identity with the *b2* allele. In the 5′ noncoding regions, gaps (indicated by asterisks) are introduced to maximize homology among the four *b* alleles. The translational start and stop sequences are underlined.

moved the C-terminal 78 amino acids without destroying *b2* activity completely. This result indicates that this part of the predicted polypeptide is not absolutely essential for *b* function. We have investigated the possibility that the Fuz⁺ Tum⁺ transformants obtained at low frequency with plasmid pUb2-12 could be accounted for by homologous recombination between the transforming DNA and the resident *b1* allele. To this end, DNA of four Fuz⁺ transfor-

```
                     20                  40                  60                  80                 100
b2   MSNYPNFSLTSFVECLNEIEHEFLRDKLENRPVLVRKLQELRRKTPNNVASLSYDPGTIHQIHQTTHRIKVAAKAFIRIDQSFVSLHSDAVEDTSKALKK
b3   --RD-KL--SK-L--------------V-H-----------QQ---KH--K-PHE-EM-Q----AA---DIV-Q----F--K----C-EI-H--T-VMQE
b4   --SD-KI-I---L---S---E------EK-P----S-----QQ-----I-N-DH--K--QK--------N--V----C---T----R------A-R----
b1   --SD-----I--L--------------G--Y-------R--QQ-I--DI-N-PR--E--Q---------RAV-Q----F--K----C-EV-HG---VMQE

                    120                 140                 160                 180                 200
                                                                            ** * *
b2   ADASSPVVGCRDLSEDLPAYHMRKHFLHTLDNPYPTQEEKEGLVRLTNESTARVGLSKANRPPLEVHQLTLWFINARRRSGWSHILKKFAREDRSRMKHL
b3   -NVV--GE---N-----------N------E----------------------IRP-N-I--------------------------------------R-
b4   -N---------E------------------E----------T-------------Q-IV-----------------------------------------
b1   FNVV--D----N---------------L-------------T-------------Q-SV--------HV-------------------------------

                    220                 240                 260                 280                 300
b2   VRAKLSSSNQSTPPSSTSDSLSNNLDDVLSDNLGRPLTPVDKQQFEDDWASMISWIKYGVKEKVGDWVYDLCAASKKTPKPGMPRPVTTVAKRHPARKTK
b3   ---------------L--EKP-DD--V-----------LA------------------------------------------------------------
b4   ---------------P--EYP------F-----------A-----------------------------------------------------Q------
b1   -----------------Y-------------------------------------------------------------------------N--------

                    320                 340                 360                 380                 400
b2   PAAKPKSRTANPRASTTPSIDSTLDSSKLESTPELSMCSTADTSFSTFGSSLSMSHYNPFQDGNDILQSPTVKARGNRKVKALPKRAGKQQPDEVENGKI
b3   -----------------------------------------------------------------------------------------------D----
b4   ---------------------------------------------------------D---Y---------F----------------------------
b1   -----------------------------------------------------------------------------------------------D----

            410
b2   PFLCLSVAFV
b3   ----------
b4   ----------
b1   ----------
```

Figure 5. Alignment of Deduced Amino Acid Sequences of *b2*, *b3*, *b4*, and *b1*

Dashes denote amino acid identity with *b2*. Asterisks mark the four conserved amino acids, WF-N-R, of the homeodomain (see Scott et al., 1989).

mants was cleaved with BamHI and analyzed by Southern hybridization using a probe internal to *b2*. In no case did we observe hybridization signals indicative of a homologous recombination event (data not shown). Other explanations for the Fuz+ transformants obtained with pUb2-12 are considered below.

To verify that the 410 amino acid ORF is responsible for *b2* activity, we created several in vitro mutations and determined their effect on *b2* function by assessing Fuz+ and Tum+ phenotypes after transformation of an *a1/a2 b1/b1* strain with plasmids containing these mutations (see Figure 3B). Deletion of a 200 bp StuI fragment, which in addition to the deletion causes a frameshift mutation (in plasmid pUb2-6), abolished *b2* function. Insertion of an oligonucleotide linker into the StuI deletion to restore the ORF (in plasmid pUb2-7) also inactivated *b2* function. Other more subtle mutations such as in pUb2-8, pUb2-13, and pUb2-14, where three additional amino acids are inserted at both or one of the two StuI sites, also eliminated *b2* activity. These results indicate that integrity of the N-terminal 140 amino acids is essential for *b2* function. The presence of the C-terminal portion of the ORF appears to be less critical. The insertion of a kanamycin resistance cassette into the SalI site (plasmid pUb2-9) or insertion of 3 amino acids at this site (plasmid pUb2-10) did not significantly affect *b2* activity (see Figure 3). In both instances approximately 70% of the transformants display the Fuz+ phenotype. This contrasts with the behavior of pUb2-12 where only 11% of the transformants are Fuz+. The major difference between plasmids pUb2-9 and pUb2-12 is that the latter plasmid lacks sequences downstream of the *b2* coding region, e.g., a putative polyadenylation signal. The presence of such a signal near the integration site for pUb2-12 might determine whether *b2* is expressed or not.

Taken together, the above observations demonstrate that the 410 amino acid ORF is responsible for the Fuz+ Tum+ phenotype and is thus the product of the *b2* gene. The nucleotide sequence does not contain consensus splice sites. Hence, the *b2* gene product appears to be a polypeptide of 410 amino acids with a predicted molecular mass of 45,957 daltons and a predicted isoelectric point of 10.21.

A homology search in the NBRF protein data bank and the EMBL gene bank using the program Wordsearch did not reveal any significant matches between the *b2* coding region and data bank entries.

Cloning and Sequence Analysis of *b1*, *b3*, and *b4*

To determine how different *b* alleles and the products they encode are related to each other, we cloned and determined the nucleotide sequence of three other alleles: *b3*, *b4*, and *b1*.

Segments containing these alleles were isolated from genomic libraries using the 8 kb BamHI fragment that contains *b2* as probe. We first confirmed by Southern hybridization that *b3*, *b4*, and *b1* strains contain fragments with strong homology to *b2*. All three strains exhibit one prominent band, corresponding in position and size (8 kb) to that observed in a *b2* strain (FB2) (data not shown). Kronstad and Leong (1989) had previously shown that a fragment containing *b1* hybridizes to a corresponding fragment containing *b2*.

Libraries of Sau3A fragments of *b3* (strain RK32) and *b4* (strain RK138) DNA in λEMBL3 (see Experimental Procedures) were screened by plaque hybridization with the 8 kb BamHI fragment containing *b2*. BamHI fragments from hybridizing clones were subcloned in pHLN. *b3* was subcloned as an 8 kb BamHI fragment, to yield plasmid pUb3; *b4* was subcloned on a 3.8 kb BamHI fragment yielding plasmid pUb4 (see Figure 3A). The reduced size of the fragment containing *b4* can be explained if joining of the particular Sau3A site in RK138 DNA with the BamHI site of λEMBL3 has recreated a BamHI site. Subsequent analysis indicates that pUb4 lacks sequences downstream of the *b4* coding region (data not shown). Transformation of an *a1/a2 b1/b1* strain (FBD12-3, Figure 3) and an *a1/a2 b2/b2* strain (FBD11-21, data not shown) with pUb3 and pUb4 resulted in Fuz^+ transformants that were able to induce tumors when inoculated into plants. Transformation of pUb4 into an *a2 b3* strain (RK32) also resulted in the appearance of Fuz^+ transformants (data not shown). To identify plasmids containing *b1*, we screened a genomic library of an *a1 b1* strain (521) in pUC18 by colony hybridization with the 8 kb BamHI fragment containing *b2*. Plasmids were identified that contained an 8 kb BamHI fragment, which was then subcloned into pHL1 (Wang et al., 1988) to yield plasmid pUb1. Fuz^+ transformants were obtained after introduction of this plasmid into an *a1/a2 b2/b2* strain (FBD11-21). These Fuz^+ transformants induced tumors when inoculated into plants (see Figure 3A). We hence conclude that plasmids pUb3, pUb4, and pUb1 carry functional *b3*, *b4*, and *b1* alleles, respectively.

The nucleotide sequences of these three alleles were determined (see Experimental Procedures). Figures 4 and 5 show the alignment of their nucleotide and amino acid sequences with that of *b2*. The comparisons reveal several striking features: First, all four alleles have a single ORF for a putative polypeptide of 410 amino acids and can be aligned without the introduction of a single gap. The putative polypeptides have molecular weights of 45,957 (*b2*), 46,322 (*b3*), 46,060 (*b4*), and 46,254 (*b1*). Second, if one discounts conservative changes, the N-terminal 110 amino acid residues are 60% identical in all four *b* alleles, while the degree of identity in the remaining portion of the putative polypeptides is 93%. Of the 410 amino acids, 321 are conserved in all four *b* alleles. The allelic differences are not distributed evenly but rather appear in clusters separated by highly or completely conserved spacers. This is most evident for the highly conserved region (amino acids 111–410), where most differences occur between residues 155 and 160 and between residues 216 and 240. Third, despite the differences in amino acid composition, the predicted isoelectric points of the four polypeptides are extremely similar: 10.21 (*b2*), 10.02 (*b3*), 9.87 (*b4*), and 9.74 (*b1*); thus all four predicted polypeptides are very basic.

Although the coding regions of the four *b* alleles could be aligned without the introduction of a single gap, gaps had to be introduced to maximize homology in the 5′ untranslated regions (Figure 4). Sequences surrounding the first ATG codon (A-T/G-C-A-A-A-ATG-T) fit with the consensus G-C-C-A/G-C-C-ATG-G for a translation initiation site (Kozak, 1987). In the regions up to position −160, we have not detected canonical TATAAA and CAAT motifs that could represent the promoter for the *b* locus, nor have we detected putative polyadenylation sites in the region downstream of the *b* coding regions. The only features that could represent putative promoter elements are five blocks of pyrimidine-rich sequences (position −118 to −109, −97 to −82, −69 to −62, −54 to −43, and −26 to −15). A number of filamentous fungal transcription initiation sites have been shown to lie in or immediately downstream of such CT-rich sequences (see Gurr et al., 1987).

Discussion

The *b* locus of U. maydis governs the switch from a yeast-like form, which is nonpathogenic, to a filamentous form, which is pathogenic on corn plants. This switch involves monitoring identical versus nonidentical *b* alleles, of which there are 25: any combination of different alleles triggers the filamentous, pathogenic program. We have cloned and determined the nucleotide sequence for four of these alleles, *b1*, *b2*, *b3*, and *b4*, and show that each contains an ORF of 410 amino acid residues. Comparison of these ORFs shows that each contains a variable N-terminal region encompassing the first 110 amino acids, where 40% differences are observed, and a more constant region for the remainder of the ORF, where 93% identity is observed. Analysis of the *b2* ORF by constructing mutations in vitro and assaying their function after transformation into U. maydis indicates that this ORF is necessary and sufficient for *b* activity, as determined by the ability to form filaments and to induce tumors in corn plants. Hence, this 410 amino acid ORF is the polypeptide encoded by the *b* locus.

The constant region of the polypeptides encoded by the *b* alleles contains the four highly conserved amino acids WF-N-R (at positions 172, 173, 175, and 177, respectively) found in all multicellular eukaryotic homeodomain proteins (see Scott et al., 1989). These residues are part of a putative DNA recognition helix in the helix-turn-helix motif of homeodomain proteins (Qian et al., 1989). Comparison of this region of the *b* polypeptide with the homeodomain of the Drosophila protein *Antennapedia*, *Antp*, (Figure 6A) shows 17 matches (which includes 6 conservative changes) in a region of 41 amino acids (41% similarity), without the introduction of gaps. These matches include, in addition to the four invariant residues, six of the eight highly conserved (though not invariant) amino acids in homeodomain proteins (Scott et al., 1989). A similar degree of relatedness is observed in comparisons with other Drosophila proteins, including even-skipped (eve) and paired (prd) (not shown). The four highly conserved residues (WF-N-R) are also found in the yeast family of regulatory proteins related to homeodomain proteins (Scott et

A

```
                                                     S
                                                 Q   I
       134       T                       I R P   N V I               H V              179
b2     Y P T Q E E K E G L V R L T N E S T A R V G L S K A N R P P L E V H Q L T L W F I N A R R R
                 :   :     .   .         :   :   * .   * :         :       : .   . : :   :   :   .
Ant    R Y Q T L E L E K E F H F N R Y L T R R R R I E I A H A L C L T E R Q I K I W F Q N R R M K

               Helix I                       Helix II                    Helix III
       -----------------------         ---------------------        -----------------------
```

B

```
                                  T N E S T A R V G L S K A N R

        126                                                                        180
b2      F L H T L D N P Y P T Q E E K E G L V R L P P L E V H Q L T L W F I N A R R R S
MATa1   E Q V F R R K Q S L N S K E K E V A K K C G I T P L Q V R V W F I N K R M R S
MATα2   F A K N I E N P Y L D T K G L E N L M K N T S L S R I Q I K N W V S N R R R K E
Pi      L L L H Y D N P Y P S N S E F Y D L S A A T G L T R T Q L R N W F S N R R R
PHO2    K R K F E I N P T P S L V E R K K I S D L I G M P E K N V R I W F Q N R R A K L
```

Figure 6. The *b* Alleles Contain a Motif Related to the Homeodomain

(A) Amino acid comparison between *b2* (amino acid residues 134–179) and the homeodomain of the Drosophila *Antennapedia* protein, *Antp* (see Scott et al., 1989). Displayed above the *b2* sequence are amino acids found in either *b3*, *b4*, or *b1* at the respective position. Identical residues are indicated by a colon; conservative substitutions are indicated by a single dot. Asterisks mark positions where identity is observed with an amino acid residue of either *b3*, *b4*, or *b1*. Amino acids E (position 139), R (position 153), A (position 159), L (position 164), Q (position 168), and T (position 170) in *b2* occupy positions that are highly conserved in all homeodomain proteins (Scott et al., 1989).

(B) Amino acid alignment of *b2* with the yeast family of homeodomain-related regulatory proteins coded by *MATa1*, *MATα2*, *mat1*-Pi, and *PHO2* genes (see Scott et al., 1989 for references). *b2* is displayed with 15 amino acids deleted to maximize identities. Identical amino acids are boxed.

al., 1989): **a**1 (Miller, 1984) and *PHO2* from Saccharomyces cerevisiae (Bürglin, 1988) and mat1-Pi from Schizosaccharomyces pombe (Kelly et al., 1988). The S. cereviseae α2 protein has three of these residues (Laughon and Scott, 1984; Shepherd et al., 1984). **a**1, α2, and *PHO2* are all known DNA binding proteins (C. Goutte and A. Johnson, personal communication; Johnson and Herskowitz, 1985; Arndt et al., 1987). A high degree of similarity is observed in comparisons of a 40 amino acid region containing WF-N-R between the U. maydis *b* polypeptides and these known yeast DNA binding proteins (Figure 6B). The similarities are particularly evident if 15 amino acid residues (from positions 147 to 161) are deleted from the *b* polypeptides to allow alignment with the yeast sequences. These 15 amino acids correspond in position to the putative helix II of *Antp* (see Figure 6A), and they may indeed comprise the putative helix II of the *b* polypeptides. Alternatively, the residues comprising helix II of the *b* polypeptides may correspond to those in the yeast group. Thus, the homeodomain-related segment of the *b* polypeptides may represent a phylogenetic intermediate between yeast and multicellular eukaryotes: the *b* polypeptides have some characteristics of the yeast family of regulatory proteins (sequences that match DNPYP) and of the *Antennapedia* group of homeodomain proteins (presence of the "additional" 15 amino acid segment).

The substantial similarity between the *b* polypeptides and known yeast DNA binding proteins strongly indicates that the *b* polypeptides function as regulatory proteins. Given the biological observation that the presence of different *b* alleles is necessary for filamentous growth and pathogenicity, we propose that the status of the *b* alleles in the cell is monitored by interaction of monomers, resulting in formation of a multimeric regulatory protein that governs the developmental fate of the organism. This proposal is in some respects analogous to the situation in **a**/α cells of yeast, in which association of the **a**1 and α2 polypeptides—both members of the yeast family of homeodomain-related proteins—creates a novel regulatory species, **a**1-α2 (Goutte and Johnson, 1988; C. Goutte and A. Johnson, personal communication).

In the next section we consider various models for the physical interactions between monomers and briefly discuss the functional consequences of such associations.

Association of *b* Polypeptides: Rules and Functional Consequences

We discuss two broad categories of models to explain how association of the *b* polypeptides might be governed and what the functional consequences of such associations might be (Figure 7). The two classes of models propose different roles for the variable and constant regions and differ primarily in whether the variable regions govern association of monomers or whether they govern activity of the multimeric species. For simplicity, we assume that the functional species is a dimer. For each of these models, we consider the possibilities that the homodimer or the heterodimer is the functional species. The target genes for the regulatory multimer are proposed to be genes for filamentous growth (*fuz* genes) and genes for tumorigenicity (*tum* genes), which are expressed in the dikaryon but not in haploids.

In Model 1 (Figure 7), the variable region is the determinant for association of monomers. If only identical variable regions are capable of association (case 1), then only homodimers are formed, and hence they must be the functional species. If only nonidentical variable regions are capable of association (case 2), then only heterodi-

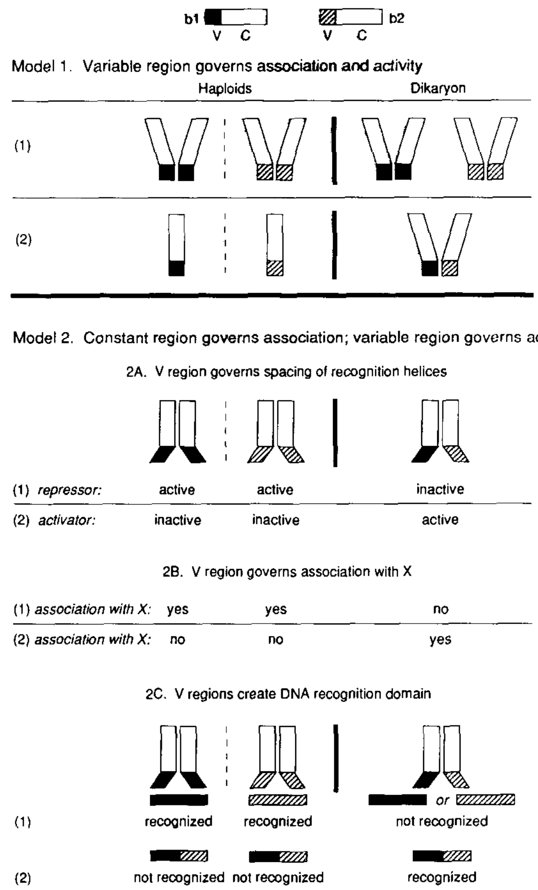


Figure 7. Models for Association of *b* Polypeptides and the Functional Consequences of Their Association

The *b* polypeptide is drawn as a rectangle with an open portion to indicate the constant region and a solid (*b1*) or cross-hatched (*b2*) part indicating the variable region. The active species formed by association of *b* polypeptides is drawn as a dimer. For each of the models, we indicate the species proposed to be present in haploids and in the dikaryon. The two models differ primarily in whether the variable or constant region is the determinant for association of *b* polypeptides.
(Model 1) In case (1), the functional species is a repressor. Haploids have one type of repressor (*b1-b1* or *b2-b2*) and the dikaryon two types of repressor (*b1-b1* and *b2-b2*). In case (2), the functional species, *b1-b2*, is an activator that is present only in the dikaryon.
(Model 2A) In case (1), the functional species is a repressor, and it is active only in haploids. In case (2), the functional species is an activator, and it is active only in the dikaryon.
(Model 2B) In case (1), only identical variable species are capable of interaction with protein X to create the active species. In case (2), only nonidentical variable regions can interact with X to produce the active species. (For both Models 2A and 2B, it is possible to imagine that both the homodimer and the heterodimer are active, one species being active only in haploids and the other only in the dikaryon.)
(Model 2C) In case (1), the functional species is a repressor: *b1-b1* recognizes only the site indicated as a solid rectangle; *b2-b2* recognizes only the site indicated as a cross-hatched rectangle. Each of the 25 homodimeric repressors recognizes its own particular site. In case (2), the functional species is an activator that recognizes a site indicated by a hybrid rectangle. Genes regulated by this activator would require the presence of 300 such sites to allow recognition by each possible heterodimeric combination of *b* polypeptides.

mers are formed, and hence they must be the functional species. Therefore, in this model the variable region governs both association and activity. The production of an active homodimeric species (whether a repressor or an activator) hypothesized by this model can be readily ruled out based on biological observations. For example, if the homodimeric species is a repressor for *fuz* and *tum* genes in the haploid, this repressor will still be present in the dikaryon. We thus conclude that if the variable region determines association, it allows association of nonidentical monomers to yield a functional heterodimer. This is analogous to the interaction of Fos and Jun to produce a functional heterodimer (see for example Nakabeppu et al., 1988; Rauscher et al., 1988; Kouzarides and Ziff, 1988).

In Model 2 (Figure 7), the constant region is the determinant for association. Hence, all polypeptides are capable of association with each other, but the activity of the multimer formed is determined by the variable region. The variable region might govern activity of the associated species in a variety of ways: by determining the spacing of the recognition helix, by providing a site for interaction with another protein, or by creating a DNA recognition domain.

In Model 2A, the variable region determines the position of the putative helix II and helix III regions in the protein so that it is able to recognize a particular target site (or accessory protein; Stern et al., 1989). The *trp* repressor provides an example of how spacing of the recognition helix may be altered, in this case by binding of tryptophan (Zhang et al., 1987; Otwinowski et al., 1988). The functional species could be the homodimer or heterodimer. In either case, there might be preferential association of nonidentical monomers in the dikaryon, to create a novel regulatory species (the heterodimer) or to interfere with formation of homodimers (again by creating a heterodimer, which in this case would be inactive). Precedent for preferential association of different monomers is provided by studies of association between Jun and Fos (O'Shea et al., 1989). In Model 2B, the variable regions of the oligomer allow for binding of the *b* protein to another protein, X, and together this complex constitutes the active species. For example, the heterodimer might bind X, and the homodimer may be unable to do so. In Model 2C, the differences between *b* alleles are responsible for creating recognition domains for different DNA sequences. This model demands that determinants for DNA recognition specificity lie outside helix III (in the variable region and putative helix II) since all *b* alleles have the same proposed helix III. It is worth noting that a patch of variable residues exists in the putative helix II region, which could contribute to such a recognition domain. If only homodimers are functional, they would have to recognize a target site that is unique to each of the alleles; these sites would have to be in or adjacent to the *b* coding region. If only heterodimers are functional, they would have to recognize a site that is not recognized by the *b* homodimers. This seems highly unlikely as it would demand that each target gene have regulatory sites that could be recognized by every possible combination of different *b* polypeptides.

In summary, two broad explanations are most consistent with biological observations: in one, only different monomers associate with each other; in the other, all monomers are capable of associating with each other, and their association allows binding to a common DNA site or to a common protein.

These models do not address the challenging problem of how association of 25 different types of monomers determines protein function. If only homodimers are functional, this would require that 25 combinations of *b* polypeptides be functional and that 300 be nonfunctional. In contrast, if only heterodimers are functional, this would necessitate that 300 combinations of polypeptides work and that 25 do not. In either event, polypeptide chains must monitor whether they are the same or different. One way to accomplish this monitoring would be if corresponding positions of two polypeptide chains are aligned with each other, for example, in parallel α helices or β sheets. Alignment of identical side chains that are charged would be much less favorable than alignment of a charged residue with a residue of opposite charge or one that is polar. It should be possible to identify the determinants in the *b* polypeptides that are responsible for self-nonself recogniton by creating in vitro chimaeric species and assaying their function after transformation into U. maydis. Since functionally identical alleles (for example, *b1*) may actually consist of a family of alleles in which different members exhibit polymorphism without altering allele specificity, a comparison of their sequences should also provide information as to residues that are relevant for self-identity.

The number of *b* alleles, which is cited as 25, is an estimate based on a few independent studies. Eighteen different *b* alleles were identified in a single study by Puhalla (1968), in which he assessed ability to induce tumors. Fifteen different *b* alleles were identified by Rowell and DeVay (1954), but these have not been tested against those of Puhalla (1968) or newly isolated alleles described in this report. Although the number of *b* alleles was estimated by Puhalla (1968) to be 25, it could be even greater. Given that alleles can be cloned and introduced into U. maydis, it is now possible to construct isogenic strains differing only at *b* to test whether all combinations of different *b* alleles behave similarly with respect to filament formation and tumor induction.

It has been suggested, although at present there is neither genetic nor biochemical evidence, that different homeodomain polypeptides of Drosophila could associate to generate heteromultimeric proteins that govern development (Scott et al., 1989). Combinatorial formation of proteins as a mode of generating diversity and specificity is also discussed by Landschulz et al. (1989). We have presented the argument that the U. maydis *b* polypeptides, which contain a motif related to the homeodomain, function in this manner to direct the developmental program during the fungal life cycle. (Two genes that govern filamentous growth, *fuz1* and *fuz2*, have been identified that may be targets of *b* and part of the filamentous, pathogenic program [F. B., unpublished data]). In U. maydis, the multiple forms of the *b* polypeptide are all encoded by a single locus with multiple alleles; in Drosophila, different homeodomain polypeptides are encoded by different loci. The rules by which the *b* polypeptides combine to form regulatory proteins and how the activity of these proteins is governed may provide clues as to how homeodomain polypeptides of Drosophila associate and function.

Experimental Procedures

Strains

Cloning in E. coli was done in DH5α (φ80 Δ*lacZM15*) Δ[*lacZYA-argF*]U196 *recA1 endA1 hsdR17* r^-m^+ *supE44* λ^- *thi1 gyrA relA1*, in DH5α*mcr*, in DH5αF′ (all obtained from Bethesda Research Laboratories), or in HB101 F^- *hsdS20* (rB^-mB^-) *supE44 ara14* λ^- *galK2 lacY1 proA2 rpsL20 xyl5 mtl1 recA13*. λEMBL3 derivatives were propagated on strain P2392 F^- *hsdR514* (rK^- mK^+) *supE44 supF58 lacY1* or Δ(*lacIZY*)6 *galK2 galT22 metB1 trpR55* λ^- (P2). M13 derivatives were propagated on DH5αF′. U. maydis strains 518 (*a2 b2*) and 521 (*a1 b1*) are from the collection of R. Holliday and represent the classical *b1* and *b2* alleles. U. maydis strains FB1 (*a1 b1*), FB2 (*a2 b2*), FB6a (*a2 b1*), FB6b (*a1 b2*), FBD12-3 (*a1/a2 b1/b1*), FBD11-21 (*a1/a2 b2/b2*) have been described (Banuett and Herskowitz, 1989). RK32 and RK138 were isolated as meiotic segregants from teliospores collected in the Bonn area during 1982. Teliospores were kindly provided by H. Hindorf.

Growth Conditions for U. maydis

CM and YEPS were prepared as described (Holliday, 1974; Tsukuda et al., 1988). Strains were grown at 28°C or 32°C. To observe the fuzz reaction, fresh cultures grown to saturation were cospotted or cross-streaked on charcoal nutrient medium (Holliday, 1974; Banuett and Herskowitz, 1989). Plates were sealed with Parafilm and incubated at room temperature. The normal Fuz phenotype could be scored after 24 hr; the reactions were followed for 2–3 days.

Plant Growth and Infection

Corn plants (variety Aztec or B164) were grown in the greenhouse or in a Conviron growth chamber under controlled conditions (14 hr light and 10 hr dark, 28°C day and 20°C night). Plants (8–10 days old) were inoculated as described by Puhalla (1968). Tumors appeared 5–8 days after inoculation.

Isolation of U. maydis DNA

Cultures (100 ml) were grown in YEPS medium (Tsukuda et al., 1988) to early stationary phase at 28°C, washed with SCS (20 mM sodium citrate [pH 5.8], 1 M sorbitol), resuspended in 2 ml of SCS containing Novozyme 234 (Novo Industries) at 20 mg/ml and incubated at 37°C. Protoplast formation was monitored microscopically. Protoplasts were washed in SCS and resuspended in 1 ml of SCS. NDS (6 ml) (0.01 M Tris–HCl [pH 9.5], 0.5 M EDTA, 1% N-laurylsarcosine) containing 0.5 mg/ml proteinase K was added, and after 10 min at 65°C, incubation was continued at 37°C for 1 hr. The DNA was purified by standard CsCl equilibrium centrifugation in the presence of ethidium bromide, extracted with isoamylalcohol, precipitated with ethanol, and resuspended in 10 mM Tris–HCl [pH 7.5], 1 mM EDTA.

Construction of Libraries and Recloning Integrated Plasmid DNA from U. maydis

U. maydis DNA that had been partially digested with Sau3A was size fractionated on NaCl gradients, fragments of 20 to 25 kb were ligated to λEMBL3 arms, packaged in vitro, and used to infect E. coli strain P2392. For plasmid libraries, 8 kb BamHI fragments of U. maydis DNA were isolated after size fractionation on gels. To reclone integrated pHLN-derived plasmid DNA from U. maydis transformants, 1 μg of total DNA was partially digested with BamHI and treated with T4 ligase in a volume of 200 μl. An aliquot of 25 μl was used to transform DH5α*mcr* to ampicillin resistance.

Transformation of U. maydis

The procedure combines protocols described by Wang et al. (1988) and Tsukuda et al. (1988). A 100 ml culture of U. maydis was grown overnight at 28°C to a cell density of 2×10^7/ml. Cells were washed with SCS, resuspended in SCS containing 12.5 mg/ml Novozyme, and kept at room temperature until protoplasts had formed (10–20 min). Then, 12 ml of SCS was added, and the suspension was centrifuged at 2800 rpm at room temperature for 10 min. The pellet was resuspended in 10 ml of SCS and centrifuged again. This step was repeated once with SCS and once with STC (10 mM Tris–HCl [pH 7.5], 100 mM

$CaCl_2$, 1 M sorbitol). Protoplasts were resuspended in STC at a concentration of 2 × 10^8/ml and kept at 0°C. For the transformation, 1–5 µg of DNA in a volume of 10 µl or less was mixed with 1 µl of heparin (15 µg/µl) and 50 µl of protoplasts were added. After 10 min at 0°C, 500 µl of PEG (40% w/v in STC) was added, and incubation at 0°C continued for 15 min. After addition of 5 ml of soft agar (0.7% agar in YEPS, 1 M sorbitol), the mixture was poured on freshly prepared gradient plates (10 ml of 1.5% agar in YEPS, 1 M sorbitol, 400 µg/ml hygromycin overlaid with 5 ml of 1.5% agar in YEPS, 1 M sorbitol). Plates were incubated at 32°C for 3–5 days. As many as 200 transformants were obtained per plate, depending on the recipient strain used. Transformants were colony purified and then tested for their phenotype.

Southern Blotting, Colony and Plaque Hybridizations, and Nucleotide Sequence Determination

Standard procedures for Southern blotting and colony and plaque hybridizations on Gene Screen Plus membranes (DuPont) were employed (Maniatis et al., 1982; Ausubel et al., 1988). Nucleotide sequences were determined using the dideoxy method of Sanger et al. (1977). Templates were either single- or double-stranded; T7 polymerase (Pharmacia) or Sequenase (United States Biochemical Corp.) were used for the elongation. Subclones were generated either in M13mp18 and M13mp19 or in pUC-derived plasmids as appropriate. For regions not accessible for subcloning, specific primers were synthesized at the UCSF Biomolecular Resource Center or at the IGF. All nucleotide sequences were determined for both strands.

Plasmids and Plasmid Constructions

pTZ18R and PT3T718U (Pharmacia) were routinely used for subcloning DNA fragments. pHL1 was kindly provided by S. Leong. It is a pUC12 derivative carrying the prokaryotic hygromycin B resistance gene under control of U. maydis *hsp70* regulatory sequences. The hygB cassette including promoter and 3′ region of *hsp70* can be excised from pHL1 as a HindIII fragment (Wang et al., 1988). In pHLN, the SstI site of the pHL1 polylinker was converted to a NotI site by linker insertion. pUb2-2 and pUb2-3 contain the 8 kb and 1.7 kb BamHI fragments of pUb2-1, respectively, cloned into the BamHI site of pHLN.

For the plasmids described below, most manipulations were first done in pTZ18R. Then BamHI fragments carrying *b2* were excised and inserted into the BamHI site of pHLN. pUb2-4: the 2.1 kb EcoRI fragment of pUb2-1 was ligated to the 115 bp EcoRI polylinker fragment of πVX (Maniatis et al., 1982), which contains a BamHI site, cleaved with BamHI, and inserted into the BamHI site of pHLN. pUb2-5: the 2 kb BglII–BamHI fragment of pUb2-4 was cloned into the BamHI site of pHLN. pUb2-6: the 200 bp StuI fragment was deleted from pUb2-4; this causes a frameshift mutation. pUb2-7: a 10 bp NotI linker (dAGCGGCCGCT) was inserted at the StuI site in pUb2-6, which restores the ORF of *b2*. Translation of the junction reads AKERPLLV. pUb2-8: the 200 bp StuI fragment of pUb2-4 was ligated to 8 bp NotI linkers (dGCGGCCGC), cleaved with NotI, and inserted into the NotI site of pUb2-7 in the original orientation. This creates two 9 bp insertions. Translation across these junctions yields AKERPPF and EGRPLLV, respectively. pUb2-13 and pUb2-14: the 2.1 kb BamHI fragments of pUb2-8 and pUb2-4 were cleaved with SnaBI. The unique SnaBI site is located between the StuI sites. The 0.35 kb fragment of pUb2-8 was ligated to the 1.75 kb fragment of pUb2-4 and inserted into the BamHI site of pHLN to generate pUb2-13. The 0.35 kb fragment of pUb2-4 was ligated to the 1.75 kb fragment of pUb2-8 and inserted into pHLN to generate pUb2-14. pUb2-9: the 1.3 kb kanamycin resistance cassette from pUC-4K (Pharmacia) was excised as a SalI fragment and inserted into the SalI site of pUb2-4. pUb2-10: the kanamycin resistance cassette of pUb2-9 was excised as a PstI fragment, which leaves an insertion of 12 bp that encodes amino acids CRST. pUb2-11: the 1.95 kb HindIII–EcoRI fragment of pUb2-4 was ligated to the HindIII fragment carrying hygB and inserted into pTZ18R cleaved with EcoRI and HindIII. pUb2-12: the 1 kb BglII–SalI fragment of pUb2-4 was ligated to the 275 bp BamHI–SalI fragment of pBR322. The resulting 1275 bp fragment was inserted into the BamHI site of pHLN.

Acknowledgments

We thank Sally Leong for providing plasmid pHL1, K. Hindorf for providing teliospores, Birgit Schroeer for skillful technical assistance, Beate Küsgen for growing the plants, Michael Bölker for assistance in drawing figures and reading the sequencing gels, and Fernando Bazán, Alexander Johnson, Daniel Nathans, Carl Pabo, Steve McKnight, Michael Bölker, and Ulla Bonas for discussion. This research was supported by grants from the German Ministry of Science and Technology to R. K.; F. B. was supported by the Weingart Program in Developmental Genetics at UCSF and by a Research Grant (to I. H.) from the National Institutes of Health (AI 18738).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 21, 1989; revised November 28, 1989.

References

Arndt, K. T., Styles, C., and Fink, G. R. (1987). Multiple global regulators control *HIS4* transcription in yeast. Science *237*, 874–880.

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (1988). Current Protocols in Molecular Biology (New York: Green Publishing Associates and Wiley-Interscience).

Banuett, F., and Herskowitz, I. (1988). *Ustilago maydis*, smut of maize. In Genetics of Plant Pathogenic Fungi, Advances in Plant Pathology, Vol. 6, G. S. Sidhu, ed. (London: Academic Press), pp. 427–455.

Banuett, F., and Herskowitz, I. (1989). Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. Proc. Natl. Acad. Sci. USA *86*, 5878–5882.

Bürglin, T. (1988). The yeast regulatory gene *PHO2* encodes a homeo box. Cell *53*, 339–340.

Casselton, L. A. (1978). Dikaryon formation in higher Basidiomycetes. In The Filamentous Fungi, Vol. 3, J. E. Smith and D. R. Berry, eds. (London: Arnold), pp. 275–297.

Christensen, J. J. (1963). Corn smut caused by *Ustilago maydis*. Am. Phytopathol. Soc. Monogr., No. 2.

Day, P. R., Anagnostakis, S. L., and Puhalla, J. E. (1971). Pathogenicity resulting from mutation at the *b* locus of *Ustilago maydis*. Proc. Natl. Acad. Sci. USA *68*, 533–535.

Goutte, C., and Johnson, A. D. (1988). **a1** protein alters the DNA binding specificity of α2 repressor. Cell *52*, 875–882.

Gurr, S. J., Unkles, S. E., and Kinghorn, J. R. (1987). The structure and organization of nuclear genes of filamentous fungi. In Gene Structure in Eukaryotic Microbes, Special Publications of the Society for General Microbiology, Vol. 22, J. R. Kinghorn, ed. (Oxford: IRL Press), pp. 93–139.

Holliday, R. (1961). The genetics of *Ustilago maydis*. Genet. Res. *2*, 204–230.

Holliday, R. (1974). *Ustilago maydis*. In Handbook of Genetics, Vol. 1, R. C. King, ed. (New York: Plenum), pp. 575–595.

Johnson, A. D., and Herskowitz, I. (1985). A repressor (*MATα2* product) and its operator control expression of a set of cell type specific genes in yeast. Cell *42*, 237–247.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. EMBO J. *7*, 1537–1547.

Kouzarides, T., and Ziff, E. (1988). The role of the leucine zipper in the fos-jun interaction. Nature *336*, 646–651.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15*, 8125–8143.

Kronstad, J. W., and Leong, S. A. (1989). Isolation of two alleles of the *b* locus of *Ustilago maydis*. Proc. Natl. Acad. Sci. USA *86*, 978–982.

Kuhn, J., and Parag, Y. (1972). Protein-subunit aggregation model for self-incompatibility in higher fungi. J. Theor. Biol. *35*, 77–91.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681–1688.

Laughon, A., and Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*, 25–31.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Miller, A. M. (1984). The yeast *MATa1* gene contains two introns. EMBO J. *3*, 1061–1065.

Nakabeppu, Y., Ryder, K., and Nathans, D. (1988). DNA binding activities of three murine Jun proteins: stimulation by Fos. Cell *55*, 907–915.

O'Shea, E. K., Rutkowski, R., Stafford, W. F., III, and Kim, P. S. (1989). Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. Science *245*, 646–648.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, F. B., and Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. Nature *335*, 321–329.

Puhalla, J. E. (1968). Compatibility reactions on solid medium and interstrain inhibition in *Ustilago maydis*. Genetics *60*, 461–474.

Puhalla, J. E. (1970). Genetic studies of the *b* incompatibility locus of *Ustilago maydis*. Genet. Res. *16*, 229–232.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell *59*, 573–580.

Raper, C. A. (1983). Controls for development and differentiation of the dikaryon in Basidiomycetes. In Secondary Metabolism and Differentiation in Fungi, J. W. Bennett and A. Ciegler, eds. (New York: Marcell Dekker, Inc.), pp. 195–238.

Rauscher, F. J., III, Voulalas, P. J., Franza, B. R., Jr., and Curran, T. (1988). Fos and Jun bind cooperatively to the AP-1 site: reconstitution in vitro. Genes Dev. *2*, 1687–1699.

Rowell, J. B. (1955). Functional role of compatibility factors and an *in vitro* test for sexual compatibility with haploid lines of *Ustilago zeae*. Phytopathology *45*, 370–374.

Rowell, J. B., and DeVay, J. E. (1954). Genetics of *Ustilago zeae* in relation to basic problems of its pathogenicity. Phytopathology *44*, 356–362.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W., III (1989). The structure and function of the homeodomain. Biochim. Biophys. Acta *989*, 25–48.

Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M., and Gehring, W. (1984). Fly and frog homeodomains show homologies with yeast mating type regulatory proteins. Nature *310*, 70–71.

Stern, S., Tanaka, M., and Herr, W. (1989). The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature *341*, 624–630.

Tsukuda, T., Carleton, S., Fotheringham, S., and Holloman, W. K. (1988). Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis*. Mol. Cell. Biol. *8*, 3703–3709.

Ullrich, R. C. (1978). On the regulation of gene expression: incompatibility in *Schizophyllum*. Genetics *88*, 709–722.

Wang, J., Holden, D. W., and Leong, S. A. (1988). Gene transfer system for the phytopathogenic fungus *Ustilago maydis*. Proc. Natl. Acad. Sci. USA *85*, 865–869.

Zhang, R.-g., Joachimiak, A., Lawson, C. L., Schevitz, R. W., Otwinowski, Z., and Sigler, P. B. (1987). The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. Nature *327*, 591–597.

GenBank Accession Numbers

The accession numbers for the U. maydis *b1*, *b2*, *b3*, and *b4* sequences reported in this paper are M30648, M30649, M30650, and M30651, respectively.

Note Added in Proof

The sequence of a partial *b4* cDNA clone and additional *b2*, *b3*, and *b4* genomic sequence information suggest that the *b* alleles may contain a 74 bp intron at positions 1192 to 1265 (Figure 4). Translation of the spliced transcript would yield a polypeptide unaltered up to position 397 (Figure 5), containing 76 or 75 additional amino acids at the C-terminus, for a total of 473 (for *b2* and *b3*) or 472 (for *b4*) amino acids. Because the region beyond amino acid position 330 does not contribute to biological activity of the *b* polypeptide, these observations do not affect any of our conclusions, although they raise the possibility of alternative splicing in the two developmentally distinct forms of U. maydis.