Mol Gen Genet (1997) 256: 661–673

ORIGINAL PAPER

G. Leubner-Metzger · B. A. Horwitz · O. C. Yoder
B. G. Turgeon

Transcripts at the mating type locus of *Cochliobolus heterostrophus*

Received: 1 March 1997 / Accepted: 21 June 1997

Abstract The single mating type locus (*MAT*) of the heterothallic ascomycete *Cochliobolus heterostrophus* is composed of a pair of unlike sequences called idiomorphs, each of which encodes one *MAT*-specific gene (*MAT-1* and *MAT-2*). *MAT* transcripts were observed in blots of poly(A)$^+$ RNA isolated from cultures grown in minimal medium, but were not detectable after growth of the fungus in complete medium, suggesting that transcription of *MAT* is tightly regulated. The idiomorphs (*MAT-1* = 1297-bp, *MAT-2* = 1171-bp) encode transcripts of 2.2 kb (*MAT-1*) and 2.1 kb (*MAT-2*), which start 5′ and end 3′ of the idiomorph within sequences common to both mating types. Analyses of *MAT-1* and *MAT-2* cDNAs revealed obligatory splicing of one intron (55-bp in *MAT-1*, 52-bp in *MAT-2*) within each *MAT*-specific ORF and optional splicing of two introns (63 and 79-bp) in the long (approximately 0.55 kb) 5′ untranslated leader sequences; the 3′ untranslated region is 0.46 kb long. Transcription start sites were found 5′ of, and within, the 79-bp intron. Optional splicing of the upstream introns and at least two transcription start sites result in three types of transcript: Type I with both 5′ introns spliced, Type II with only the 63-bp intron spliced, and Type III with neither 5′ intron spliced. The three transcript types are distinguished by various combinations of four short ORFs encoded by the corresponding genomic DNA, in the leader sequences of the *MAT* mRNAs. The transcript structure suggests several mechanisms by which expression of the *MAT* genes might be regulated at the level of translation during sexual development.

Key words Ascomycete · Intron · uORF · UTR · mRNA leader

Communicated by C. A. M. J. J. van den Hondel

G. Leubner-Metzger
Friedrich Miescher Institute, P.O. Box 2543,
CH-4002 Basel, Switzerland

B. A. Horwitz
Technion – Israel Institute of Technology,
Department of Biology, Haifa 32000, Israel

O. C. Yoder · B. G. Turgeon (✉)
Department of Plant Pathology, 334 Plant Science Bldg.,
Cornell University, Ithaca, NY 14853, USA
Fax: +1-607-255-4471; e-mail: bgt1@cornell.edu

Introduction

The single mating type (*MAT*) locus of the heterothallic ascomycete *Cochliobolus heterostrophus* consists of two non-homologous sequences (*MAT-1* and *MAT-2*) called idiomorphs (Metzenberg and Glass 1990), flanked by sequences common to both mating types (Turgeon et al. 1993). Idiomorphs are found at the *MAT* loci of all ascomycetes investigated to date, including *Saccharomyces cerevisiae* (Hicks et al. 1979; Herskowitz 1989), *Schizosaccharomyces pombe* (Kelly et al. 1988; Egel et al. 1990), *Neurospora crassa* (Glass et al. 1990; Staben and Yanofsky 1990), *Podospora anserina* (Picard et al. 1991), and *Magnaporthe grisea* (Kang et al. 1994). Among these genera, it is clear that there are significant differences in morphology of reproductive structures, size of the *MAT* locus, number of genes encoded, location of transcription starts and stops, and type of DNA-binding protein encoded, illustrating that the route to successful mating and reproduction has been achieved in different ways in various taxonomic groups.

In the pyrenomycetes *N. crassa* and *P. anserina*, one idiomorph encodes a single *MAT*-specific transcript, while the other encodes at least three (Debuchy et al. 1993; Chang and Staben 1994; Ferreira et al. 1996). In the hemiascomycete *S. cerevisiae*, each idiomorph (*MATa* or *MATα*) encodes two transcripts; three of the four are required for mating (Astell et al. 1981; Klar et al. 1981; Tatchell et al. 1981; Miller 1984). In *S. pombe* (a hemiascomycete), each idiomorph (*mat1-P* or *mat1-M*), encodes two transcripts; all four are required for sexual reproduction (Kelly et al. 1988; Egel et al. 1990). All four *N. crassa* genes are transcribed, at low levels, during both the vegetative and mating phases, suggesting

involvement in processes other than mating (Ferreira et al. 1996). In *S. cerevisiae, MATa1, α1* and *α2* are expressed during all developmental stages; *MATα1* is down-regulated during the mating process (Esposito and Klapholz 1981; Herskowitz and Oshima 1981; Herskowitz 1989). For *S. pombe*, low levels of two (*mat1-Pc* and *mat1-Mc*) of the four *MAT*-specific transcripts encoded by the idiomorphs are detectable during vegetative growth. Removal of nitrogen from the medium causes an increase in the levels of all four transcripts (Kelly et al. 1988; Egel et al. 1990).

We have undertaken a study of the transcripts of genes encoded by the *MAT* loci in a mating pair of highly inbred *C. heterostrophus* strains, in an effort to determine the number and the regulation of transcripts and to compare these data with those for the other ascomycetes. Our study has revealed that a single *MAT*-specific transcript is encoded by each idiomorph and that expression is tightly regulated by the composition of the growth medium. Transcription starts and stops in common flanking DNA, resulting in mRNA that is almost twice the size of the corresponding ORF. For both *MAT-1* and *MAT-2,* RNA blot and RT-PCR analyses have identified a heterogeneous population of transcripts that differ in start sites, and optional splicing of two introns in the long 5′ untranslated leader sequences. The leader sequences include several short ORFs which may be involved in post-transcriptional regulation.

Materials and methods

Strains, growth media and transformation

The near-isogenic *C. heterostrophus* strains C3 (*MAT-2*), C4 (*MAT-2*), and C5 (*MAT-1*) (Leach et al. 1982) were grown in 500 ml of minimal (MM) or complete (CM) medium with shaking for 3–5 days at 30° C (Turgeon et al. 1985) prior to harvesting for RNA isolation. In some cases, *MAT-1* and *MAT-2* strains were cocultivated (grown together, starting with equal amounts of inoculum). *E. coli* strain DH5α MCR (Gibco BRL) was used to propagate plasmid DNA.

RNA manipulations

Poly(A)$^+$ RNA isolation**.** Total RNA was isolated from fungal strains using guanidine thiocyanate extraction followed by CsCl centrifugation (Sambrook et al. 1989). Lyophilized mycelium was frozen in liquid nitrogen, ground to a powder, and homogenized for 2 min in a Waring blender in 50 ml of extraction buffer (4 M guanidine thiocyanate, 30 mM sodium acetate pH 5.2, 1% (v/v) β-mercaptoethanol) per g dry mycelium. After centrifugation (30 min, 27 000 × *g*) in Corex glass tubes, the supernatant was layered on a cushion (11 ml) of 5.7 M CsCl, 2 mM EDTA in Ultraclear centrifuge tubes (Beckman, 25 × 89 mm) and centrifuged (18–24 h, 120 000 × *g*, 20° C, SW28 rotor). After removal of the supernatant, the RNA pellet was dissolved in TE pH 8.0 (Sambrook et al. 1989), then extracted with chloroform:1-butanol (4:1) and ethanol precipitated. The pellet was dissolved in TE pH 8.0 and stored at −70° C. The RNA concentration was determined by spectrophotometry and by agarose gel electrophoresis using standards. Poly(A)$^+$ RNA was purified on oligo(dT)-cellulose columns (Boehringer-Mannheim) or by using the Oligotex mRNA kit (Qiagen), in each case according to the supplier's directions.

RNA blot analysis. Glyoxal-denatured poly(A)$^+$ RNA (20–30 μg per lane) was electrophoresed in 1.1% (w/v) agarose gels (Sambrook et al. 1989). Capillary transfer to Nytran nylon membranes was carried out according to the supplier's manual (Schleicher and Schuell). Prehybridization and DNA-RNA hybridization (Sambrook et al. 1989) was at 65° C, in 6 × SSC, 5 × Denhardt's solution, 0.1% (w/v) SDS, and 100 μg of denatured, sheared salmon sperm DNA/ml. Following overnight hybridization, filters were washed once at room temperature for 10 min in 2 × SSC, 0.1% SDS, once at 65° C for 20 min in 1 × SSC, 0.1% (w/v) SDS, and once at 65° C for 50 min in 0.2 × SSC, 0.1% (w/v) SDS. RNA-RNA hybridization was at 55° C as above; filters were washed in 2 × SSC, 0.1% (w/v) SDS at 60° C for 5 min. Transcript sizes were determined by comparison with RNA molecular weight marker I (Boehringer Mannheim).

Probes

Sources of probes (Fig. 1) were plasmids pB11 and pdB11, which carry 8.3-kb *Bgl*II and 2.2-kb *Pst*I fragments of the *MAT-1* locus, respectively, and p73HB, which carries a 3.5-kb *Eco*RV fragment of the *MAT-2* locus (Turgeon et al. 1993; Wirsel et al. 1996). Probes were labelled with [α-^{32}P]dCTP using the random primer method (Feinberg and Vogelstein 1983). In addition, PCR was used with C3 genomic DNA as template and primer pairs GL32/GL25i, GL23/GL31, and GL27/GL28 to amplify fragments M1.2, M2.2, and M4.3, respectively (see Fig. 6B). These PCR fragments were cloned into pCR-Script SK(+) (Stratagene), sequenced to verify

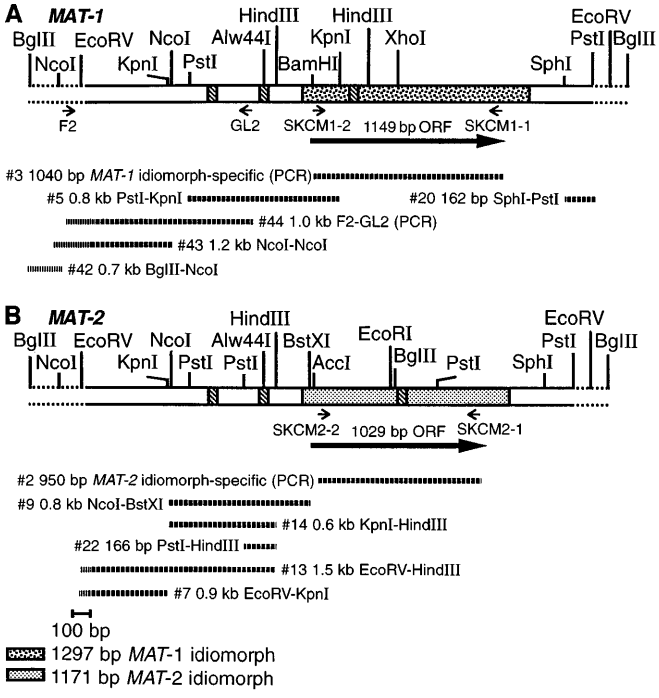


Fig. 1A, B DNA probes used for transcript mapping. At the top of each panel are restriction maps of the chromosomal *MAT-1* (**A**) and *MAT-2* (**B**) loci. The *stippled* (A) and shaded (B) boxes indicate the idiomorphs; the cross-hatched boxes denote putative introns. Below each map the broken bars indicate fragments used as probes in Figs. 2, 3, 4, 9; the large arrows indicate the *MAT*-specific ORFs; and the small arrows indicate the positions of primers listed in Table 1. Most probes were obtained by digesting plasmids carrying *MAT-1* or *MAT-2* DNA (Materials and methods). Others (#2, #3 and #44) were generated by PCR with primers (SKCM2-1/SKCM2-2, SKCM1-1/SKCM1-2, and F2/GL2, respectively). *Dotted lines* designate regions not drawn to scale. Only some restriction enzyme sites are shown

identity, linearized with an appropriate restriction enzyme, purified on an agarose gel, and used as templates for in vitro transcription of antisense RNA labelled with [α-^{32}P]dCTP using T3 or T7 RNA polymerase (Boehringer Mannheim).

PCR amplification of *MAT*-specific cDNAs

Partial *MAT-1* and *MAT-2* cDNAs were amplified by PCR. First-strand cDNA was synthesized using Superscript reverse transcriptase (Gibco BRL), 1 μg of *MAT-1* (strain C5) or *MAT-2* (strain C3) poly(A)$^+$ RNA from fungus grown in MM, and primer SKCM1-1 or SKCM2-1, as appropriate (see Fig. 5 and Tables 1 and 2). Single-stranded cDNAs were purified on Glassmax spin cartidges (Gibco BRL).

PCR reactions using single-stranded cDNAs as templates and primers listed in Table 1 were carried out in a Cetus 9600 thermocycler (Perkin Elmer) in 50-μl reactions containing: 200 μM dNTPs, 2 mM $MgCl_2$, 0.2 μM primer, 1 × AmpliTaq buffer, 0.05 units AmpliTaq DNA polymerase, and single-stranded cDNA corresponding to 0.05–0.1 μg poly(A)$^+$ RNA. Denaturation for 2 min at 95° C preceded 35 cycles of 1 min at 95° C, 1 min at 52–62° C, and 2 min at 72° C. A 7-min extension period and subsequent cooling at 4° C followed the final cycle. A second set of reactions was done using 0.2 μl of the first reaction as template and nested primers (Table 2). PCR products were evaluated by electrophoresis through 3.5% (w/v) MetaPhor-agarose (FMC Bioproducts) stained with ethidium bromide, and by DNA blot analysis (Sambrook et al. 1989) using labelled *MAT*-specific DNA probes.

Following digestion with appropriate restriction enzymes, the PCR-derived cDNAs GLi1/GL4, GLi1/GL5, SKCM1-2/SKCM1-1, and SKCM2-2/SKCM2-1 (Fig. 5) were cloned into pUC18 (Sambrook et al. 1989) and sequenced using the Sequenase 2.0 kit (US Biochemical) with [α-^{35}S]dATP and either gene-specific (Table 1) or common primers provided with the kit. The PCR-derived cDNAs GL13/GL5, GL17/GL5, (Fig. 5), GL23/3, and GL20/3 (Fig. 6B) were cloned into the pCRII vector (Invitrogen). Sequences were determined at the Cornell DNA Sequencing Facility using Taq-Cycle automated sequencing with DyeDeoxy terminators (Applied Biosystems).

The Gibco BRL kit was used for 3′ RACE (Rapid Amplification of cDNA Ends). *MAT-2* first-strand cDNA was obtained as described above using primer GLdTT (Table 1). Two subsequent PCR reactions (annealing temperature 55° C) were performed as above using the primer pairs GL35/GLdT and GL36/GLdT. PCR fragments were cloned into the pCR-Script SK(+) vector and sequenced as described.

Table 1 Sequences of primers used for PCR and for sequencing *MAT* cDNAs

Primers	Sequence (5′ to 3′)
F2	GGCCCGGGTGTGAGTTATCCTCCCTG
GLi1	CCTGTGACTGCCTGTTGAAGCTTGG
GL2	CGTGAGTCGCAGGGAGAGGTTACG
GL3	GTGGAGTCGAAATCTCAGAAACAGG
GL4	GATAGTAGACCAGGCTTTCG
GL5	CTTGTAATTGGGGTGCTGGC
GL9	GCCTTTGTCAAGACTCAGAACAAGAACC
GL10	CTCAAACTCCCCTTGAGTATTAGTGAG
GL11	CCTTACAGACTGCTGCCTCAGACG
GL13	GACAGTGAGTGATGAACTGTGCAC
GL14	CTTCTCGCCAGGCTTCCTTGGAGTG
GL17	CCTTGAGTATTAGTGAGATTTACTC
GL20	CATCTTGCCTGTTTATTCCTAGC
GL21	CCAAGTGATTCCTAGTTAGAGACC
GL23	GTCTCATATATCAAGTCACGGTC
GL25	CCATCCGCGTGTGGCGTCAG
GL25i	CTGACGCCACACGCGGATGG
GL26	GTGCAAGTAAAGCATCAATGGCAC
GL27	AGTGAGGTAAGTAAAGGCG
GL28	AAATCTGGTGATAGCAAACGG
GL31	AAATGTGCATTACTGCGCTGTC
GL32	GCTACAGATGTCTCTTATGCAAGG
GL35	GAGATCTCTACGCGTTGC
GL36	AGCCTACTCGATACGGAG
GLdT	GGCTCTAGAGCTTTTTTT
GLdTT	GGCTCTAGAGCTTTTTTTTTTTTTTTTT
SKCM1-1[a]	GCAGATCTGTCGTCGATGGT
SKCM1-2[a]	GCAGATCTCCGCACTGGAGC
SKCM2-1[b]	GCAAGCTTGTTGCATCTCCG
SKCM2-2[b]	GCAAGCTTGGCTGCAAGGAT

[a] As described previously (Wirsel et al. 1996)
[b] As described previously (Sharon et al. 1996)

Results

Idiomorph-specific transcripts

When poly(A)$^+$ RNA from a *MAT-1* strain grown in MM was probed with a *MAT-1* specific probe (Fig. 1A, probe #3) a 2.2 ± 0.1-kb *MAT-1* specific transcript was evident (Fig. 2A), while a 2.1 ± 0.1-kb signal was observed (Fig. 2B) in RNA from the *MAT-2* strain, probed with the *MAT-2*-specific probe (Fig. 1B, probe #2). No signals were observed when *MAT-1* RNA was probed with the *MAT-2*-specific probe and vice-versa (Figs. 2A, B). When strains of opposite mating type were co-cultivated, both *MAT-1* and *MAT-2* transcripts were detectable (Fig. 2A, B). Both *MAT-1* and *MAT-2* transcripts were present in low abundance compared to the glyceraldehyde 3-phosphate dehydrogenase (*GPD1*) (VanWert and Yoder 1992) signal. About 30 μg of total RNA from fungus grown in MM was required to detect a weak *MAT* signal after 2 weeks exposure; with 20 μg poly(A)$^+$ RNA the transcript was evident after 1 day of exposure. In contrast, the *GPD1* probe with comparable specific activity easily detected a signal with either 30 μg of total RNA (not shown) or 2 μg of poly(A)$^+$ RNA (Fig. 3C) after a few hours of exposure.

Transcription of *MAT* is regulated by composition of the culture medium

Idiomorph-specific transcripts were detected in poly(A)$^+$ RNA from *MAT-1* (strain C5) or *MAT-2* (strain C3 or C4) grown in MM, but not in CM (Figs. 3A, B). In contrast, signals of equal intensity were obtained when poly(A)$^+$ RNA from fungus grown in either MM or CM was probed with the constitutively expressed *GPD1* gene (Fig. 3C).

MAT-specific transcripts start and stop in the common flanking DNA

The observation that *MAT* transcripts are about twice as long as the corresponding ORFs (Turgeon et al. 1993)

Table 2 PCR analysis of intron splicing and transcription start site(s) in *MAT-2* mRNA: templates, primers and products

Template	Primer(s)[a]		PCR product	Size[b]	Intron(s) spliced[c]	cDNA type
	A	B				
First PCR reaction[d]						
RT-PCR/SKCM2-1	GL17	GL14	GL17/14	e		
RT-PCR/SKCM2-1	GL13	GL14	GL13/14	e		
RT-PCR/SKCM2-1	GL25	GL9	GL25/9	e		
RT-PCR/SKCM2-1	GL23	GL9	GL23/9	e		
RT-PCR/SKCM2-1	GL21	GL9	GL21/9	e		
RT-PCR/SKCM2-1	GL26	GL9	GL26/9	e		
RT-PCR/SKCM2-1	GL20	GL9	GL20/9	e		
RT-PCR/SKCM2-1	GL11	GL9	GL11/9	e		
RT-PCR/SKCM2-1	GL17	GL9	GL17/9	e		
RT-PCR/SKCM2-1	GL13	GL9	GL13/9	e		
RT-PCR/SKCM2-1	GLi1	GL9	GLi1/9	e		
Second (nested) PCR reaction[f]						
GL17/14	GL17	GL5	GL17/5	836-bp	55-bp, 63-bp	I or II
Genomic DNA	GL17	GL5	None			
GL13/14	GL13	GL5	GL13/5	863-bp	55-bp	III
Genomic DNA	GL13	GL5	GL13/5	918-bp	–	
GL25/9	GL25	GL3	None			
Genomic DNA	GL25	GL3	GL25/3	623-bp	–	
GL23/9	GL23	GL3	GL23/3	449-bp	79-bp, 63-bp	I
Genomic DNA	GL23	GL3	GL23/3	591-bp	–	
GL21/9	GL21	GL3	GL21/3	343-bp	79-bp, 63-bp	I
Genomic DNA	GL21	GL3	GL21/3	485-bp	–	
GL26/9	GL26	GL3	GL26/3	313-bp	79-bp, 63-bp	I
Genomic DNA	GL26	GL3	None			
GL20/9	GL20	GL3	GL20/3	371-bp	63-bp	II
				434-bp	–	III
Genomic DNA	GL20	GL3	GL20/3	434-bp	–	
GL11/9	GL11	GL3	GL11/3	220-bp	63-bp	I or I
				283-bp	–	III
Genomic DNA	GL11	GL3	GL11/3	283-bp	–	
GL17/9	GL17	GL3	GL17/3	139-bp	63-bp	I or II
Genomic DNA	GL17	GL3	None			
GL13/9	GL13	GL3	GL13/3	166-bp	–	III
Genomic DNA	GL13	GL3	GL13/3	166-bp	–	
GLi1/9	GLi1	GL3	GLi1/3	105-bp	–	I, II or III
Genomic DNA	GLi1	GL3	GLi1/3	105-bp	–	

[a] For primer locations refer to Figs. 5 and 7B; for primer sequences refer to Table 1

[b] Sizes measured by gel electrophoresis are in agreement with sizes calculated from the genomic DNA sequence (Turgeon et al. 1993), and the sequences of partial cDNAs GL23/3 (Type I), GL20/3 (Type II), GL20/3 (Type III), GL17/5, GL13/5, GLi1/GL5, and SKCM2-2/SKCM2-1 (Figs. 5 ,7 and 8)

[c] Note that partial cDNAs were amplified, therefore only introns included in the particular PCR fragment are listed as spliced

[d] First-strand cDNA synthesis was achieved with poly(A)$^+$ RNA from *MAT-2* strain C3 grown in MM using primer SKCM2-1. This served as template (RT-PCR/SKCM2-1) in a set of 'first' PCR reactions (annealing temperature 62°) listed in this Table using a primer from column A in combination with primer GL14 or GL9 (primer B) (Figs. 5 and 7B)

[e] On a gel, products of the first PCR were either invisible or very faint bands (not shown)

[f] Aliquots of 'first' PCR reactions served as templates (primerA/GL14 or primerA/GL9) in a second set of nested PCR amplifications (annealing temperature 62°) using a primer from column A in combination with nested primer GL5 or GL3 (primer B) (Figs. 5 and 7B)

suggested that the transcripts start and/or stop within the sequences flanking the idiomorphs. To test this, poly(A)$^+$ RNAs were probed with 5′ and 3′ flanking DNAs (Fig. 4A–D). For reference, the non-idiomorph sequence 5′ of the 5′ end of the *MAT*-specific ORF is defined as the 5′ flank. Probes containing sequences exclusively from either the 5′ (Fig. 1B, probe #22) or 3′ (Fig. 1A, probe #20) flanks of the idiomorph detected both *MAT-1* and *MAT-2* specific transcripts (Fig. 4A, B). Probes bearing 5′ flank sequences plus a short region of idiomorph (Fig. 1, probes #5 and #9) revealed the same transcripts as probes 22 and 20 and no others (Fig. 4C, D). Thus, flank-specific probes identify the *MAT*-specific transcripts and these transcripts start 5′ and end 3′ in the common regions flanking the idiomorphs.

Since 3′ flank-specific probe #20 (Fig. 4B) detected both *MAT-1* and *MAT-2* specific transcripts, the transcription stop site must be located 3′ of the *Sph*I site in the 3′ flanking region (Fig. 1). Sequencing of four independent *MAT-2* 3′ RACE clones revealed a single transcription stop site (5′...ACGATTTTCAAA...3′) at position 2085 (Turgeon et al. 1993). 3′ RACE sequences were identical to genomic DNA (Genbank Accession Number X68398; Turgeon et al. 1993) except for the addition of a poly(A) tail. Thus, the *MAT*-specific

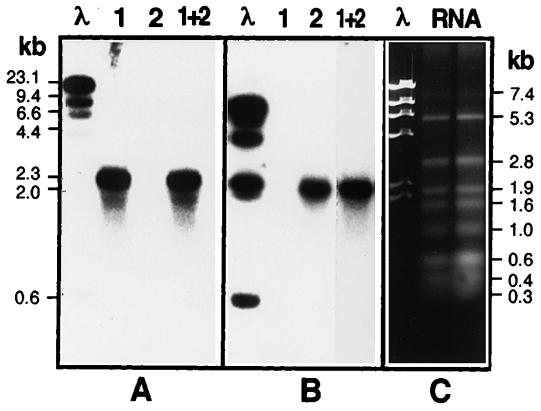


Fig. 2A–C RNA blot analysis of *MAT-1* and *MAT-2* specific transcripts. Poly(A)$^+$ RNAs (20 μg per lane) from *MAT-1* strain C5 (lanes 1), *MAT-2* strain C3 (lanes 2), or strains C5 plus C3 (lanes 1 + 2) grown in MM were analysed with the following probes. **A** *MAT-1* idiomorph-specific PCR fragment #3 (Fig. 1). **B** *MAT-2* idiomorph-specific PCR fragment #2 (Fig. 1). **C** Ethidium bromide-stained gel showing RNA size markers (lanes labeled RNA, sizes indicated on the right) compared with lambda DNA digested with *Hin*dIII (lane labeled λ, sizes indicated on the left). The *MAT-1* probe detects a 2.2 ± 0.1-kb transcript (**A**), the *MAT-2* probe detects a 2.1 ± 0.1-kb transcript (**B**). Both transcripts were detected in poly(A)$^+$ RNA from a co-culture of strains of both mating types (**A**, **B**). No signals were found when *MAT-1* RNA was probed with a *MAT-2* specific probe and vice-versa

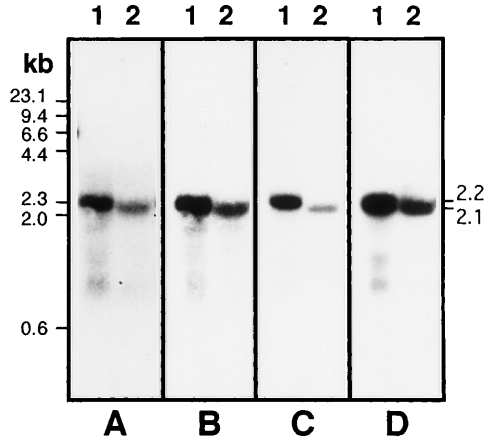


Fig. 4A–D *MAT*-specific transcripts start 5′ and end 3′ of either idiomorph. Poly(A)$^+$ RNA was analysed from *MAT-1* strain C5 (lanes 1) or *MAT-2* strain C3 (lanes 2) using the following probes (see Fig. 1). **A** Fragment #22. **B** Fragment #20. **C** Fragment #5. **D** Fragment #9. Probes unique to either the 5′ flank (A) or the 3′ flank (B) detected both of the *MAT*-specific transcripts, as did probes which overlapped the idiomorph borders (C, D). Sizes of *MAT*-specific RNAs are indicated on the right [based on comparison with lambda and RNA markers (Fig. 2); fragment sizes of lambda DNA digested with *Hin*dIII are indicated on the left)]

transcripts have a common 0.46-kb 3′ untranslated region (3′UTR) following the *MAT*-specific ORF.

To map the approximate transcription start point, additional probes (Fig. 1) were used. 5′ flank-specific DNA probes #13 (Fig. 8A), #22 (Fig. 4A), #14 and #44 (results not shown) detected both *MAT-1* and *MAT-2* transcripts. No *MAT-1* and *MAT-2* signals were obtained with 5′ flank specific DNA probes #7 (Fig. 8B), #43 (Fig. 8C), and #42 (results not shown). Similarly, when poly(A)$^+$ RNA blots were probed with RNA fragment M2.2 (102-bp, Fig. 6B) *MAT*-specific transcripts were detected in *MAT-1* and *MAT-2* RNA, but not when the blots were probed with RNA probe M1.2 (113-bp, Fig. 6B). The DNA probes demonstrated that the *MAT*-specific transcripts start in the region between the 5′ *Kpn*I and *Hin*dIII sites (Figs. 1 and 6B), and the RNA probes localized the start site more precisely to the 110-bp region between the 3′ end of primer GL25 and the 3′ end of primer GL31 (see Fig. 6B). This conclusion is supported by the observation that primer GL23 (Fig. 6B) was the most 5′ primer that resulted in amplification of a cDNA in combination with primer GL3 (Fig. 6B; Table 2). No specific cDNA was amplified with primer GL25. This localizes the transcript start site to the 35-bp region between the 3′ ends of primers GL25 and GL23 (Fig. 6).

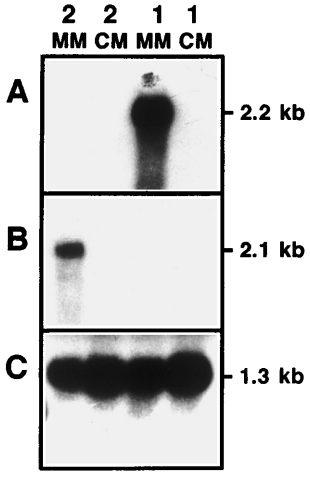


Fig. 3A–C RNA blot analysis showing that transcription of the *MAT* genes is regulated by the composition of the culture medium. Poly(A)$^+$ RNAs from strain C5 (*MAT-1*; lanes 1) or C4 (*MAT-2;* lanes 2) grown in either MM or CM were probed with (**A**) *MAT-1* specific probe #3 (Fig. 1), (**B**) *MAT-2* specific probe #2 (Fig. 1) or (**C**) the constitutively expressed *GPD1* gene (which controls for amounts of RNA loaded). Neither the *MAT-1* nor the *MAT-2* transcript was detectable when cultures were grown in CM

Thus, the *MAT*-specific transcripts contain long (ca. 0.55 kb) 5′ leader sequences preceding the *MAT*-specific ORF sequences (Turgeon et al. 1993). We have not been successful in determining the precise transcription start sites using either 5′ RACE or primer extension, despite many attempts. Preliminary evidence suggests that formation of stable RNA secondary structures interferes with analysis of 5′ end sequences.

An idiomorph-specific intron is spliced within each *MAT*-specific cDNA

Partial *MAT-1* and *MAT-2* cDNAs (GLi1/GL4, SKCM1-2/SKCM1-1, GLi1/GL5, SKCM2-2/SKCM2-1,

GL13/GL5, GL17/5, Fig. 5) were amplified by PCR, cloned and sequenced. The combined sequences represent most of each idiomorph and more than 0.17 kb of the 5′ flank. Comparison of the cDNA and genomic sequences confirmed that the previously proposed *MAT-1* (52-bp) and *MAT-2* (55-bp) specific introns (Turgeon et al. 1993) are spliced. A second putative intron in the *MAT-2* idiomorph (Turgeon et al. 1993) is not spliced. With the exception of the introns, the 0.17-kb 5′ flank plus idiomorph sequences of the cDNAs and the genomic DNAs were identical to each other. RNA blots probed with cDNAs SKCM2-2/SKCM2-1 and SKCM1-2/SKCM1-1 (Fig. 5) showed the 2.1 *MAT-2* and 2.2 kb *MAT-1* transcripts, respectively (not shown).

Differential splicing of introns in the 5′ leader mRNA

Two putative introns, 79 and 63-bp in length, are located 5′ of the idiomorph-specific ORF in the DNA common to both *MAT-1* and *MAT-2* (Figs. 5 and 6B). Two independent strategies, one based on PCR and the other on RNA blot analysis with differential RNA probes, were employed to determine if these introns are spliced.

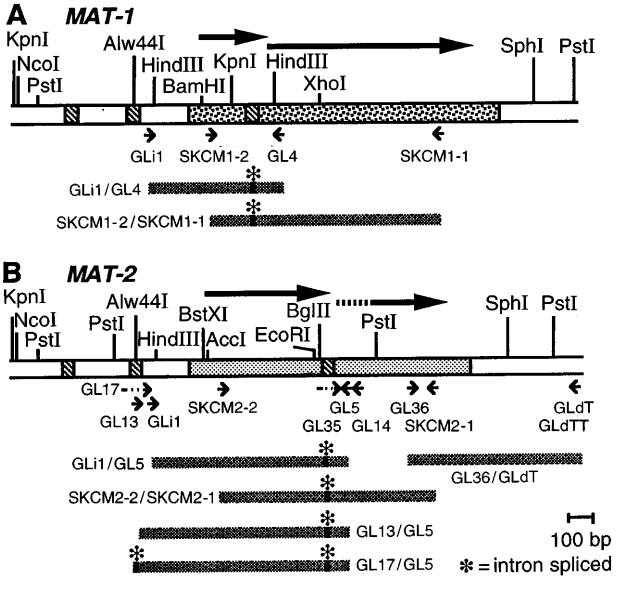


Fig. 5A, B Amplification of partial *MAT-1* (**A**) and *MAT-2* (**B**) cDNAs. The maps at the top of each panel represent the *MAT* loci. Cross-hatched boxes indicate introns; from left, 79-bp, 63-bp and 52-bp (*MAT-1*) or 55-bp (*MAT-2*). The *bars* below represent partial cDNAs with introns (*asterisks*) spliced out, as determined by sequencing. *Large arrows* indicate ORFs previously proposed to be joined via intron splicing (Turgeon et al. 1993). Single-strand cDNAs, synthesized with primers SKCM1-1 (*MAT-1*) or SKCM2-1 (*MAT-2*) or GLdTT for 3′RACE (*MAT-2*), were used as templates to amplify cDNAs by PCR using primers indicated by the *small arrows*. Each cDNA was generated by two sequential PCR reactions using primer pairs and annealing temperatures (°C) as follows: GLi1/GL4 (GLi1 and SKCM1-1, 52°; GLi1 and GL4, 58), GLi1/GL5 (GLi1 and SKCM2-1, 52; GLi1 and GL5, 58), SKCM1-2/SKCM1-1 (GLi1 and SKCM1-1, 55; SKCM1-2 and SKCM1-1, 55), SKCM2-2/SKCM2-1 (GLi1-SKCM2-1, 52; SKCM2-2 and SKCM2-1, 55), GL13/GL5 (G13 and GL14, 62; G13&GL5, 62), GL17/GL5 (G17 and GL14, 62; G17 and GL5, 62), 3′RACE product GL36/GLdT (GL35 and GLdT, 55; GL36 and GLdT, 55)

In the PCR strategy, poly(A)$^+$ RNA was used to synthesize first-strand cDNA by reverse transcription with primer SKCM2-1 (Table 2; Fig. 5B) or SKCM1-1 (Fig. 5A). Controls for this step did not yield products and included reactions (i) without RNA, (ii) with yeast tRNA instead of *MAT* poly(A)$^+$ RNA, or (iii) with *MAT* poly(A)$^+$ RNA and a sense strand primer (e.g., SKCM2-2 for *MAT-2*, SKCM1-2 for *MAT-1*; Fig. 5). With single-strand cDNA as template, one set of PCR reactions was performed using primer pairs A and B (Table 2, First PCR reactions). Aliquots of the first PCR reactions were used as templates in a second set of PCR reactions with primer pairs A and nested B (Table 2, Second PCR reactions). Control PCR reactions without template or with only one primer did not lead to products (not shown).

The 63-bp 5′ UTR intron. Differential splicing of the common 63-bp intron in the 5′ flank of the *MAT-2* idiomorph was demonstrated with cDNAs GL13/GL5 and GL17/GL5 (Table 2; Fig. 5B). Primer GL13 is specific for the 63-bp intron sequence, while primer GL17 is specific for the sequence obtained after splicing of the 63-bp intron (Fig. 5B). The GL13/GL5 cDNA (863-bp) was smaller than the corresponding genomic DNA fragment (918-bp, Table 2). Sequencing of this clone revealed splicing of the idiomorph-specific 55-bp intron, but not the putative 63-bp intron (Table 2; Fig. 5). The idiomorph and 5′ flank sequences of the GL13/GL5 cDNA were identical to the genomic sequence except for the 55-bp intron. In addition, PCR fragments of equal sizes were amplified from cDNA and genomic DNA using primer GL13 (Fig. 6; Table 2) in combination with primer GL3, indicating that the 63-bp intron was not spliced.

Primer pair GL17/GL5 did not amplify a fragment from genomic DNA, but did amplify a fragment from cDNA (Table 2; Fig. 5B), demonstrating the specificity of primer GL17 for the sequence obtained after splicing of the 63-bp intron. Sequencing of this cDNA revealed splicing of the idiomorph-specific 55-bp intron as well as the 63-bp intron in the 5′ flank; the rest of the sequence was identical to that of genomic DNA. Amplification with primer pair GL17/GL3, also resulted in a single cDNA product (Fig. 6; Table 2) with the 63-bp intron spliced. As with GL17/GL5 cDNA (Fig. 5; Table 2), no product was obtained with genomic DNA (Fig. 6; Table 2). Thus, comparison of PCR products obtained with primer pairs GL13 and primer GL5 or GL3 versus GL17 and primer GL5 or GL3, revealed two types of transcript; one with the 63-bp intron spliced out, the other still containing the intron.

The 79-bp 5′ UTR intron. A second series of PCR experiments utilized primer GL9 (Fig. 6B) and various A primers (First PCR Reactions, Table 2) followed by nested reactions using the products of the first reaction as templates with primer GL3 and the same set of A

primers (Second PCR Reactions, Table 2; Fig. 6A). Results for *MAT-2* are summarized below; in some cases, data were also collected for *MAT-1* and found to be identical.

1. GL23 (Fig. 6B) was the most 5′ primer that resulted in amplification of a cDNA in combination with primer GL3 (Fig. 6B; Table 2) as described above.

2. Sequence analysis of cDNA GL23/3 demonstrated that both the 79-bp and the 63-bp intron were spliced out (Figs. 6 and 7; Table 2); the remaining cDNA sequence was identical to that of genomic DNA. Due to intron splicing, cDNAs GL23/3 and GL21/3 (Fig. 6; Table 2) differ in size by 142-bp from the corresponding fragments amplified from genomic DNA. With these primer sets, no cDNAs with only one or no intron spliced out were amplified (Figs. 6 and 7; Table 2), suggesting that whenever the 79-bp intron is spliced out, the 63-bp intron is also removed.

3. Primer GL26 is specific for the sequence obtained when the 79-bp intron is spliced (Fig. 6B). When this primer was used with primer GL3, a single cDNA (GL26/3) with both introns spliced out was obtained. No product was generated using genomic DNA as template (Fig. 6A; Table 2). This finding supports the results with primer pairs GL23/GL3 and GL21/GL3 (point 2).

4. PCR with primers located either inside the 79-bp intron sequence (GL20) or between the introns (GL11), in combination with primer GL3 amplified two cDNA fragments of different sizes (Fig. 6; Table 2). Sequence analysis of both GL20/GL3 cDNAs revealed differential splicing of the 63-bp intron. For these cDNAs, the smaller fragment represents cDNAs in which the 63-bp intron is spliced out. The larger fragment is the same size as the genomic amplification product, and represents cDNAs in which the 63-bp intron has not been spliced (Fig. 6A; Table 2). This result is consistent with the demonstration that GL13/GL5 and GL17/GL5 cDNAs differ only in splicing of the 63-bp intron (Fig. 5A). Contamination of the *MAT* poly(A)$^+$ RNA with genomic DNA cannot account for the amplification of the larger product, since use of the same batch of single-stranded cDNA did not result in fragments of the size expected for amplification from genomic DNA when primers GL25, GL23, or GL21, instead of GL20 or GL11, were used with GL3 (Fig. 6; Table 2), or when primer GL13 was used in combination with GL5 (Fig. 5B; Table 2).

To investigate the possibility that the apparent evidence for differential intron splicing was due to amplification of rare or transient, preprocessed transcripts in the "first PCR reaction" (Table 2), blots of poly(A)$^+$ RNA were probed with RNA fragment M4.3 (72-bp, Fig. 6B) which is specific for the 63-bp intron. Both the *MAT*-specific transcripts were detected (not shown), indicating that transcripts with unspliced 63-bp introns are relatively abundant and not an artefact caused by PCR amplification of rare preprocessed transcripts or by amplification from genomic DNA.

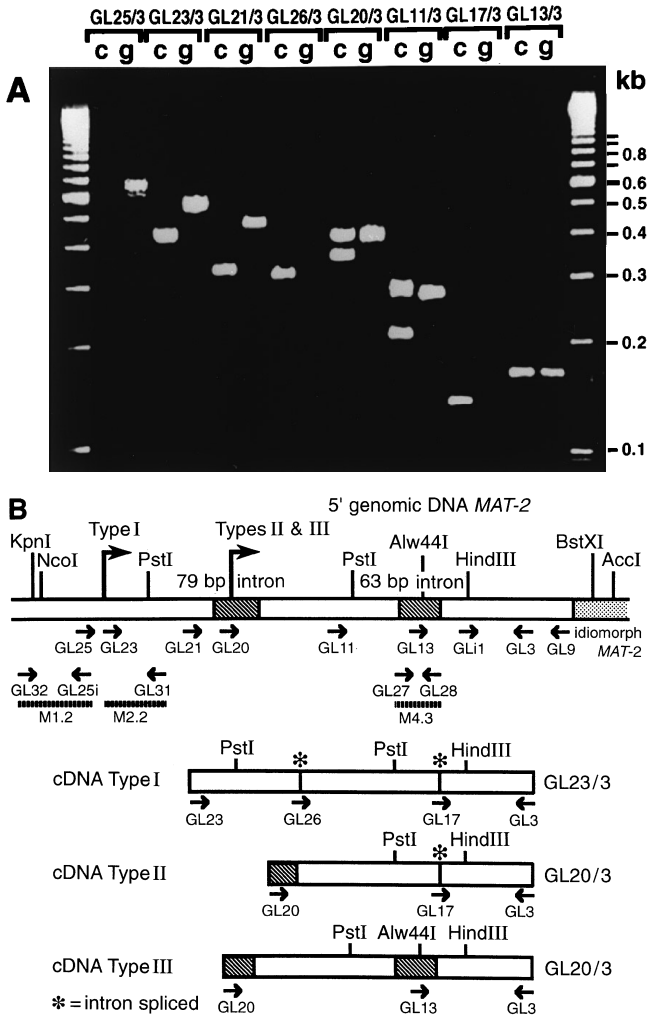


Fig. 6A, B Differential splicing of introns in the 5′ leader of *MAT-2* transcripts. A Mapping of *MAT-2* transcription start sites and differential splicing of the 5′ flank introns. g, genomic DNA; c, cDNA; a 100-bp DNA marker ladder was loaded in each of the outer lanes (fragment sizes indicated on the right). Numbers above the lanes designate primer pairs used (see Figs. 5 and 6B; Table 2). Fragments amplified by PCR from a cDNA template using the 3′ primer GL3 and a series of 5′ primers (GL25, GL23, GL21, GL26, GL20, GL11, GL17, GL13) are compared to the corresponding products amplified from genomic DNA (Table 2). Note that: (i) the most 5′ primer which resulted in *MAT-2*-cDNA amplification is GL23 and no amplification occurred with GL25; (ii) no fragment was amplified from genomic DNA with 5′ primer GL26 or GL17, which are specific for the sequence generated after splicing of the 79-bp or the 63-bp intron, respectively (iii) due to differential splicing of the 63-bp intron, two fragments were amplified from cDNA with 5′ primers located between the 79- and 63-bp introns (GL11) or inside the 79-bp intron sequence (GL20) and (iv) single cDNA fragments were obtained with 5′ primers GL23, GL21, and GL26, indicating that when the 79-bp intron is spliced out, the 63-bp intron is also removed. **B** Classification of *MAT* cDNAs according to the pattern of splicing of the 79- and 63-bp introns in the 5′ flank of the idiomorph. Maps representing the three cDNA types are compared with genomic 5′ flanking DNA of *MAT-2*. Type I, Type II and Type III transcription start sites are indicated *above* the genomic DNA (large arrows). Locations of PCR primers (*small arrows*), non-spliced introns (*hatched boxes*), spliced introns (*asterisks*) and PCR fragments (plus primers used to generate them) used as templates for RNA probes (*thin broken bars*) are indicated. The three cDNA Types are represented by cDNAs GL23/3 and GL20/3

Thus, it is reasonable to conclude that three types of partial *MAT-2* cDNA can be amplified from single strand cDNA generated with idiomorph-specific primer SKMC2-1. All three cDNA types contain idiomorph sequences and 5′ flank sequences and have the idiomorph-specific 55-bp intron spliced out, but differ in splicing of the 63-bp and 79-bp introns located in the 5′ flank (Table 2; Figs. 5B and 6B).

Classification of *MAT* cDNAs

Type I, represented by cDNA GL23/3 (Figs. 6B and 7; Table 2), has both the 79-bp and the 63-bp introns spliced out. Type II, represented by cDNA GL20/3 (Figs. 6B and 7; Table 2), has the 63-bp intron spliced out while the 79-bp intron is still present. Type III, also represented by cDNA GL20/3 (Figs. 6B and 7; Table 2) retains both introns. In all three types, the idiomorph-specific intron is spliced out and a single transcription stop site is evident within the 3′ flanking region of the idiomorph. All Type I transcripts must start upstream of the 79-bp intron sequence. For some (if not all), transcription starts in the 35-bp region between the 3′ ends of primers GL25 and GL23 (Figs. 6B and 7) common to *MAT-1* and *MAT-2*. No transcripts of sequences further 5′ were detectable by PCR or RNA blot analysis (Figs. 6 and 7). No Type II or III cDNAs were obtained that corresponded to a transcription start 5′ of the 79-bp intron sequence. Therefore a start site for these transcripts must lie within the 79-bp intron, most probably within the 3′ part of primer GL20 (Figs. 6 and 7). Thus, at least two *MAT* transcription start sites are evident, leading to production of three types of transcript that differ by as much as 80-bp in their leader sequences.

Non-specific transcripts in the *MAT* region

Probing with DNA 5′ and 3′ of the idiomorphs revealed additional transcripts at the *MAT* locus. Of these a 1.4-kb (Fig. 8A–C) and a 2.4-kb (not shown) transcript were localized to a 3-kb region 5′ of each idiomorph.

The 1.4-kb common transcript was detected at the same levels in poly(A)$^+$ RNA from fungus grown in either MM or CM (Fig. 8C). Since this transcript is detectable with probes (Fig. 1) #13 (Fig. 8A), #7 (Fig. 8B), #43 (Fig. 8C), #14 and #44 (not shown), but not with #5 (Fig. 4C), #9 (Fig. 4D), #22 (Fig. 4A), or #42 (not shown), the gene corresponding to it is located in the 5′*Nco*I fragment (Fig. 1). The 2.4-kb transcript is detectable with probe #42 (not shown) but not with probes #43 (Fig. 8C), #13 (Fig. 8A), #7 (Fig. 8B) and #44 (not shown). Thus, this transcript is located 5′ of the 5′ *Nco*I site (Fig. 1). Cosmid probes containing the *MAT-1* or the *MAT-2* idiomorph, flanked on either side by approximately 16 kb of DNA, detected 3–5 transcripts common to both mating type strains (Fig. 8D).

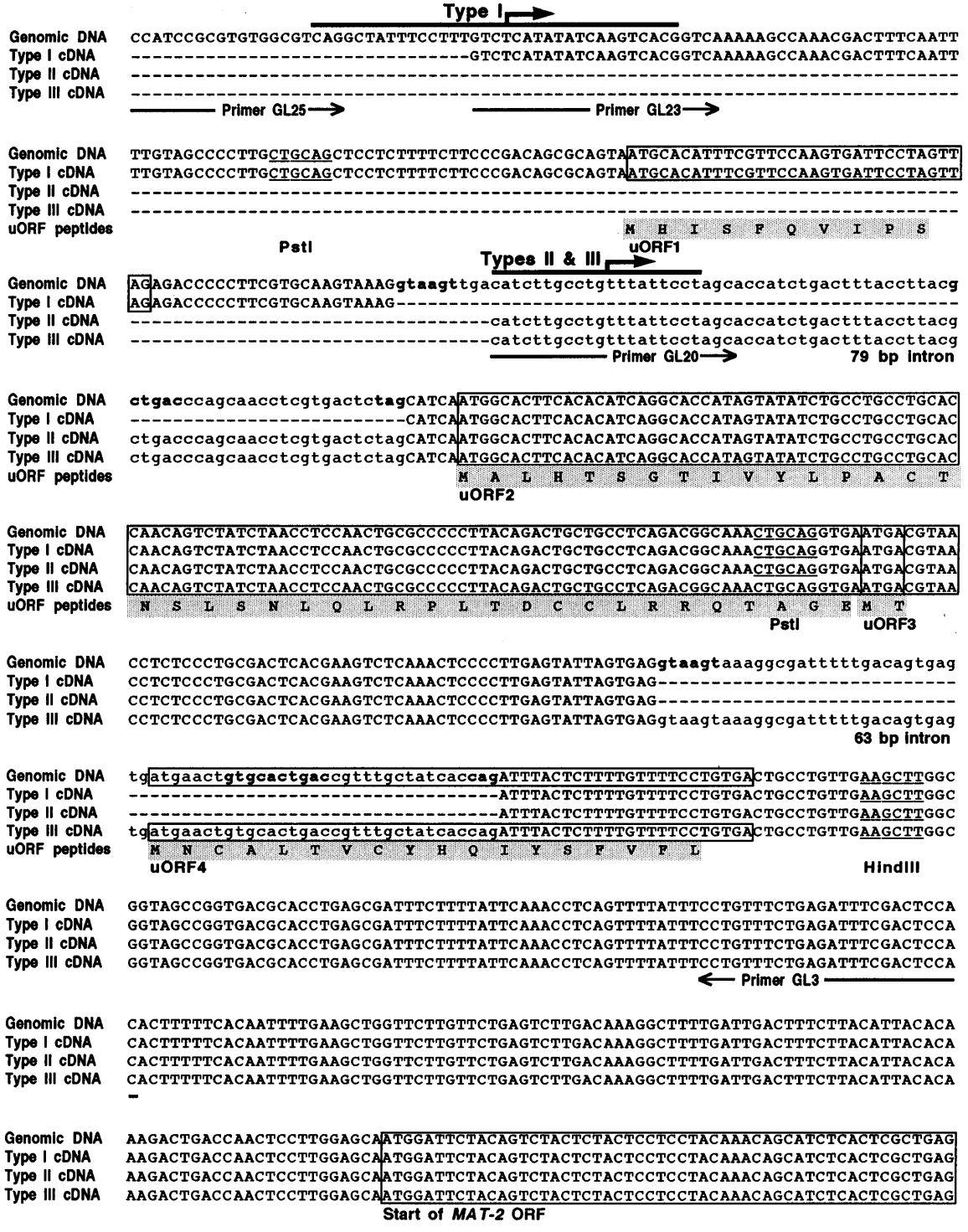
Fig. 7 5′ cDNA sequences aligned with the *MAT-2* genomic DNA sequence. Primer sequences are indicated by *thin arrows*; intron 5′ and 3′ splice signals and the putative branch sequences are in *bold*; *MAT*-specific ORF and uORF nucleotide sequences are *boxed*; uORF amino acid sequences are *shaded*; *Hin*dIII and *Pst*I restriction enzyme sites are *underlined*. Sequences were obtained from partial cDNAs GL23/3 (Type I), GL20/3 (Type II), GL20/3 (Type III), GL17/GL5 (Type I or II), and GL13/GL5 (Type III) (Figs. 5 and 6B); the genomic sequence is from Turgeon et al. (1993) with additional data derived from sequencing of the genomic GL25/3 DNA. Note that, in each of the three types of *MAT-2* cDNAs, the idiomorph-specific 55-bp intron is spliced out. In Type I, represented by cDNA GL23/3, both the 79-bp and the 63-bp introns are spliced out. A proportion of the Type I transcripts start upstream of the *Pst*I site near primer GL23. For Types II and III, represented by GL20/3 cDNAs of different sizes, the 79-bp intron is not spliced out. Some of these transcripts start within the 79-bp intron sequence near primer GL20. Sequencing confirmed that in Type II cDNA the 63-bp intron is spliced out, while it is still present in Type III. Approximate transcription starts are indicated by *horizontal bars* attached to *arrows*

Discussion

RNA analyses have determined that a unique *MAT*-specific transcript is associated with each *MAT* idiomorph, that expression of these transcripts is tightly regulated, and that there is 5′ heterogeneity in the transcript population. We conclude that *C. heterostrophus MAT* expression is regulated at the level of transcription and suggest that *MAT* may also be regulated at the post-transcriptional level, although there is no direct evidence for the latter.

MAT transcripts are regulated by the composition of the culture medium

No transcripts were found in poly(A)$^+$ RNA from cultures grown in liquid complete medium but were readily detectable (although in low abundance) in poly(A)$^+$ RNA from cultures grown in minimal medium. Co-culture of strains of opposite mating type does not appreciably alter the abundance of either *MAT*-specific transcript. The nutrients in minimal medium are the same as those in crossing medium (except for the addition of glucose). Complete medium differs from minimal by the presence of yeast extract and casein hydrolysate (0.2% w/v). In earlier work, the effect of altering the nitrogen concentration, and source, as well as other components of crossing medium on *C. heterostrophus* mating ability, was studied (Leach et al. 1982). While our main objective in that study was to improve mating procedures in the laboratory, we did note that addition of casein hydrolysate to crossing medium reduced or eliminated pseudothecium production, and that trehalose plus casein hydrolysate completely inhibited mating. Nitrogen concentration (8.4 mM nitrate) in minimal and crossing medium vs complete medium (a complex mixture containing 8.4 mM nitrate plus 0.1% yeast extract and 0.1% casein hydrolysate) is likely to be the most



important factor. For *S. pombe,* removal of nitrogen from the medium causes elevation of all four transcripts encoded at *MAT* and induces mating and sporulation (Kelly et al. 1988; Egel et al. 1990). In contrast, nitrogen and carbon starvation are not required for mating of *S. cerevisiae* but are required for initiation of meiosis and for sporulation (Herskowitz 1989). Nitrogen starvation (<10 mM nitrate) is required for mating of *N. crassa* (Nelson and Metzenberg 1992).

We have noted that a cluster of three carbon-repression motifs (5′ C/GT/CGGG/AG3′), found in genes involved in carbon catabolite regulation [e.g., *S. cerevisiae MIG1* (Nehlin et al. 1991; Kulmburg et al. 1993), *A. nidulans CREA* (Cubero and Scazzocchio 1994)], is present in the –1.26 to –1.31 kb region (GCGGAG, GTGGGG, GTGGAG), 5′ of both *C. heterostrophus* idiomorphs. There are also two consensus nitrogen repression motifs (5′GATA/TA) found in genes involved

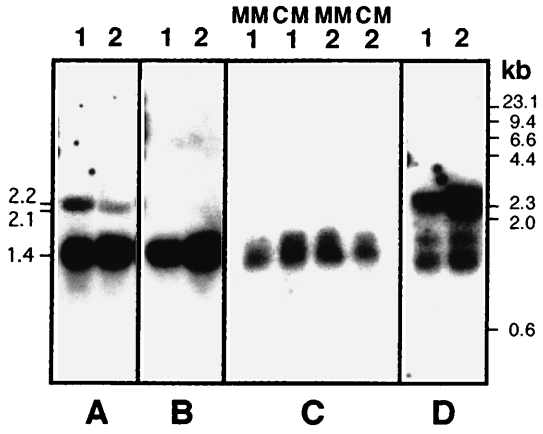


Fig. 8A–D *MAT*-specific and nonspecific transcripts in the *MAT* region. Poly(A)$^+$ RNA was probed with fragments (Fig. 1): #13 (**A**); #7 (**B**); #43 (**C**). The gel in **D** was probed with a cosmid containing *MAT-2* DNA flanked on either side by about 16 kb of DNA. Lanes 1, *MAT-1* strain C5; lanes 2, *MAT-2* strain C3 (**A**, **B**, **D**) or C4 (**C**). Fragment sizes of lambda DNA digested with *Hin*dIII are indicated on the right, RNA marker sizes on the left. MM, minimal medium; CM, complete medium. *MAT*-specific transcripts are evident at about 2 kb in **A**. A 1.4-kb transcript is seen in **A–C**. Several transcripts are evident in **D**

in nitrogen regulation [e.g., *S. cerevisiae GLN3* (Minehart and Magasanik 1991), *A. nidulans AREA* (Peters and Caddick 1994), and *N. crassa NIT2* (Fu and Marzluf 1990)] upstream of both idiomorphs. One of these (GATTA) is at –2.04 kb 5′ of both idiomorphs; the other (GATAA) is at –1.72 kb and is followed by TATCTA, which is an additional DNA binding motif that has been described in *NIT2*. Further experiments designed to directly manipulate these sequences and subsequently examine phenotypic consequence are required to determine if the *C. heterostrophus* sequences are involved in carbon catabolite and nitrogen regulation.

Organization of the *MAT*-specific transcripts

Each *C. heterostrophus MAT*-specific transcript is about twice as long as the corresponding ORF encoding the *MAT*-specific protein (which lies entirely within the idiomorph). For *MAT-1* (1297-bp idiomorph), the ORF is 1149-bp in length; the corresponding transcript is 2.2 ± 0.1 kb long. For *MAT-2* (1171-bp idiomorph), the ORF is 1029-bp in length; the corresponding transcript is 2.1 ± 0.1 kb long. The difference (approx. 0.1 kb) reflects the difference in the size of the two *MAT*-specific ORFs. A combination of DNA-RNA and RNA-RNA hybridization experiments and RT-PCR amplifications has demonstrated that transcription starts in the 5′ and stops in the 3′ common flanking DNA. Since each *MAT*-specific ORF begins within 50-bp of the 5′ end and stops within 50-bp of the 3′ end of the idiomorph, the long 5′ and 3′ untranslated sequences (5′ and 3′ UTRs) are nearly identical for both transcripts. Sequencing of cDNAs has confirmed that a single intron is spliced out of the region which encodes the putative DNA-binding segment of the translated portion of each transcript as proposed previously (Turgeon et al. 1993). The *MAT-1*-specific protein is thus 383 amino acids in length and the *MAT-2*-specific protein is 343 amino acids long. The position of each intron within the sequence coding for the DNA-binding region is conserved in the corresponding *MAT* genes of all filamentous ascomycetes sequenced to date; these introns are not found in the functionally equivalent genes of *S. cerevisiae* (*MATα1*) or *S. pombe* (*mat1-Mc*). The *MAT* transcripts have long 3′ UTRs; 3′ RACE analysis of the *C. heterostrophus MAT-2* transcripts revealed a single transcription stop site within the 3′ flanking region of the idiomorph, resulting in an untranslated region (3′ UTR) of 466 nucleotides. Known polyadenylation signal sequences, such as the AAUAAA motif, are not present around the transcription stop site.

Transcription of genes encoded by the *MAT* loci of all other ascomycetes studied to date begins within the idiomorph and thus their untranslated leader sequences are unique to the particular *MAT* gene. For *S. cerevisiae* and *S. pombe*, each *MAT* idiomorph produces a pair of divergent transcripts. The *S. cerevisiae a1* transcript is encoded entirely within the idiomorph, while the *α1* and *α2* transcripts begin in the idiomorph and end in the common flanking DNA. The *S. pombe mat1-Pm* and *mat1-Mc* transcripts are encoded entirely within the idiomorph, while the *Mm* and *Pc* transcripts start in the idiomorphs and end in the flanking DNA. For *C. heterostrophus*, both *MAT*-specific transcripts start and stop in common flanks and there are only two nucleotide differences in the DNA corresponding to the 5′ untranslated portions of the transcripts and five nucleotide differences in the DNA corresponding to the 3′ untranslated portions of the transcripts, resulting in nearly identical regulatory and termination sequences and suggesting they may be similarly regulated. We speculate that stoichiometric levels of MAT-1 and MAT-2 proteins must be present in diploid cells, perhaps for formation of a MAT-1/MAT-2 heterodimer. The *N. crassa* and *P. anserina* homologs of these *C. heterostrophus MAT* genes are encoded by DNA within the idiomorph. As yet we have no evidence that *C. heterostrophus* encodes or requires the two additional genes found at the *N. crassa mt A* and *P. anserina mat-* loci.

5′ Heterogeneity of *MAT* transcripts

Using two independent strategies, RNA blot analyses with DNA and RNA probes, and RT-PCR with a series of nested primers specific to the common 5′ region, we identified three types of *MAT* transcript distinguished by heterogeneity within the 5′ UTRs. Type I transcripts have both introns of the leader sequence spliced out and start in the 35-bp interval between the 3′ ends of primers GL25 and GL23. Their 5′ UTR leader is thus 553 nucleotides long. A putative CAAAT box is located ~55-bp up-

stream of GL23 and sequences with similarity to the *N. crassa* consensus sequence for transcriptional start sites (TCATCANC; Bruchez et al. 1993), and to one of the *N. crassa mt A-2* gene transcriptional start sites (5′TCATCTTC3′; Ferreira et al. 1996) are present within primer GL23 (5′TCATATATCAAG3′, Fig. 7). Type II and III mRNAs start in the interval between primer GL21, which amplifies only Type I transcripts, and GL20, a primer specific for the 79-bp intron, because no Type II and III transcripts are produced with primer pair GL21/3, while both are when GL20/3 is used (Figs. 6B and 7). The 5′ UTRs of Type II and Type III *MAT* transcripts are about 473 and 536 bases long, respectively. No appropriately positioned consensus sequences for transcription start sites are present. All size classes of transcript migrate together as a broad band in agarose gels. The short RNA probe M4.3, which is specific for the 63-bp intron sequence, detected transcripts in blots of poly(A)$^+$ RNA, thereby demonstrating that differential intron splicing is not an artefact due to amplification of rare unprocessed transcripts.

The *MAT* transcripts have long leader sequences

Long untranslated leader sequences (UTRs) are found in $<10\%$ of fungal and higher eukaryotic transcripts (Kozak 1984, 1987; Ballance 1986; 1991; Gurr et al. 1987); the majority have leaders of <100 bases. These long 5′ UTRs are generally associated with regulatory genes such as the yeast *GCN4* gene, which encodes a major transcriptional activator (Hinnebusch 1988, 1994), and the *CPA1* gene, which encodes carbamoylphosphate synthetase A (Werner et al. 1987). In *N. crassa*, examples include the *CPC1* gene, a cross-pathway regulator of amino acid biosynthesis (Paluh et al. 1988), transcripts of regulatory genes in the quinic acid utilization (*qa*) cluster (Giles et al. 1985; Gurr et al. 1987), and the *N. crassa mt A-2* and *A-3* genes encoded by the *mt A* mating type idiomorph (Ferreira et al. 1996).

Long leader sequences and complexity of *MAT* transcripts may reflect a second mechanism of regulation of *C. heterostrophus MAT* expression. 5′ heterogeneity is due to the use of at least two transcription start sites and to alternative splicing of 5′ UTR introns, creating three transcript types. There are four ATGs and thus four possible short uORFs in the *C. heterostrophus MAT* 5′ UTRs preceding the ATG which opens the reading frame of each *MAT*-specific ORF (Fig. 7). Optional intron splicing would create varying numbers of these in the mature mRNAs (three in Type I and III and two in Type II) which could affect regulation of *MAT* expression as follows. Firstly, different numbers of upstream AUGs may affect the efficiency of translation of the *MAT*-specific ORF, as is observed with transcripts of *S. cerevisiae GCN4* (Hinnebusch 1988, 1994; Abastado et al. 1991; Dever et al. 1992) and *CPA1* (Werner et al. 1990, 1987; Delbecq et al. 1994) where translational control is mediated by upstream AUGs. Secondly, upstream uORFs may be translated, yielding small peptides that function in regulation of *MAT* translation. A 10-amino acid peptide (uORF1) would be produced only from Type I and an 18-amino acid peptide (uORF4) only from Type III transcripts (Fig. 7). All transcript types share uORF2 (which encodes a 39-amino acid peptide that differs by one amino acid residue between *MAT-1* and *MAT-2*). uORF3 encodes two amino acids and is found in all transcript types. For *GCN4,* each of four uORFs consists of only two or three codons**.** Two are involved in translational control and there is evidence for translation of one or more of these (Hinnebusch 1988, 1994; Abastado et al. 1991; Dever et al. 1992; Geballe and Morris 1994). For *CPA1*, regulation by the single uORF is sequence dependent and its translation has been clearly demonstrated (Werner et al. 1987, 1990; Delbecq et al. 1994); the uORF of the *N. crassa* homolog (*arg-2*) is also translated (Luo et al. 1995). Interestingly, a portion of the 25-amino acid *CPA1* peptide (MFSLSNLQ) shares some homology with a portion of the putative *MAT* uORF2 peptide (TNSLSNLQ) and there are also certain similarities between the *C. heterostrophus* uORFs and the uORFs preceding the *N. crassa mt A-2* gene (Ferreira et al. 1996). Whether or not uORFs of *MAT* transcripts are translated and their peptide products involved in regulation is not known. Thirdly, upstream introns may be spliced in a stage- or tissue-specific manner. Further experiments examining *C. heterostrophus MAT* transcripts in specific tissues (protoperithecia, pseudothecia, ascospores) and at different time points in the mating process are required. The *MER2* gene of yeast, for example, is transcribed both during mitosis and meiosis. Splicing of an 80-nucleotide intron within the coding region of the transcript is meiosis-specific and generation of a functional product occurs efficiently only in meiosis (Engebrecht et al. 1991). Lastly, different hairpin loops in the RNA structure of the three types of message may affect ribosome scanning. Application of an RNA fold program (Chan et al. 1991; Zuker and Jacobson 1995) to the 5′ UTRs of the three *C. heterostrophus MAT* transcript types (not shown) revealed several possible stem-loop structures and homologies of up to 88% to yeast tRNA sequences were found, suggesting comparable loop structures (Mazo et al. 1979; Stucka et al. 1987; Hauber et al. 1988). Our unsuccessful attempts to perform 5′ RACE and primer extension are consistent with the existence of stable secondary structures in the 5′ UTR.

Additional transcripts in the *MAT* region

The additional transcripts at the *MAT* locus appear to be constitutively expressed. In separate work (S. Wirsel et al., in preparation) we have identified ORFs corresponding to the 1.4- and 2.4-kb transcripts and are in the process of deleting or truncating them to determine if either is involved in mating.

Acknowledgments We thank Holly Davies for expert technical assistance and Amir Sharon and Sung-Hwan Yun for helpful suggestions and providing some of the RNA samples. We thank Stefan Wirsel for thoughtful discussions and critical reading of the manuscript. A grant to G. Leubner-Metzger by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. This work was supported by a grant to G.T. and O.Y. from the U.S. Department of Agriculture.

References

Abastado J-P, Miller PF, Jackson BM, Hinnebusch AG (1991) Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. Mol Cell Biol 11:486–496

Astell C, Ahlstrom-Jonasson L, Smith M, Tatchell K, Nasmyth K, Hall B, (1981) The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*. Cell 27:15–23

Ballance DJ (1986) Sequences important for gene expression in filamentous fungi. Yeast 2:229–236

Ballance DJ (1991) Transformation systems for filamentous fungi and an overview of fungal gene structure. In: Leong SA, Berka RA (eds) Molecular industrial mycology: systems and applications for filamentous fungi. Marcel Dekker, New York, pp 1–30

Bruchez JJP, Eberle J, Russo VEA (1993) Regulatory sequences in the transcription of *Neurospora crassa* genes: CAAT box, TATA box, introns, poly(a) tail formation sequences. Fungal Genetics Newslett 40:89–96

Chan L, Zuker M, Jacobson AB (1991) A computer method for finding common base paired helices in aligned sequences: application to the analysis of random sequences. Nucl Acids Res 19:353–358

Chang S, Staben C (1994) Directed replacement of *mtA* by *mta-1* effects a mating type switch in *Neurospora crassa*. Genetics 138:75–81

Cubero B, Scazzocchio C (1994) Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. EMBO J 13:407–415

Debuchy R, Arnaise S, Lecellier G (1993) The *mat-* allele of *Podospora anserina* contains three regulatory genes required for the development of fertilized female organs. Mol Gen Genet 241:667–673

Delbecq P, Werner M, Feller A, Filipkowski RK, Messenguy F, Pierard A (1994) A segment of mRNA encoding the leader peptide of the *CPA1* gene confers repression by arginine on a heterologous yeast gene transcript. Mol Cell Biol 14:2378–2390

Dever TE, Feng L, Wek RC, Cigan AM, Donahue TF, Hinnebusch AG (1992) Phosphorylation of initiation factor 2a by protein kinase GCN2 mediates gene-specific translational control of *GCN4* in yeast. Cell 68:585–596

Egel R, Nielsen O, Weilguny D (1990) Sexual differentiation in fission yeast. Trends Genet 6:369–373

Engebrecht J, Voelkel-Meiman K, Roeder GS (1991) Meiosis-specific RNA splicing in yeast. Cell 66:1257–1268

Esposito RE, Klapholz S (1981) Meiosis and ascospore development. In: Strathern JN, Jones EW, Broach JR (eds) The molecular biology of the yeast *Saccharomyces*: life cycle and inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York, pp 211–287

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Ferreira AVB, Saupe S, Glass NL (1996) Transcriptional analysis of the *mt A* idiomorph of *Neurospora crassa* identified genes in addition to *mt A-1*. Mol Gen Genet 250:767–774

Fu YH, Marzluf GA (1990) *Nit-2,* the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. Proc Natl Acad Sci USA 87:5331–5335

Geballe AP, Morris DR (1994) Initiation codons within 5′-leaders of mRNA as regulators of translation. Trends Biochem Sci 19:159–164

Giles NH, Case ME, Baum J, Geever R, Huiet L, Patel V, Tyler B (1985) Gene organization and regulation in the *qa* (quinic acid) gene cluster of *Neurospora crassa*. Microbiol Rev 49:338–358

Glass NL, Grotelueschen J, Metzenberg RL (1990) *Neurospora crassa A* mating-type region. Proc Natl Acad Sci 87:4912–4916

Gurr SJ, Unkles SE, Kinghorn JR (1987) The structure and organization of nuclear genes of filamentous fungi. In: JR Kinghorn (ed) Gene structure in eukaryotic microbes. IRL Press, Oxford, pp 93–139

Hauber J, Stucka R, Krieg R, Feldmann H (1988) Analysis of yeast chromosomal regions carrying members of the glutamate transfer RNA gene family: various transposable elements are associated with them. Nucleic Acids Res 16:10623–10634

Herskowitz I (1989) A regulatory hierarchy for cell specialization in yeast. Nature 342:749–757

Herskowitz I, Oshima Y (1981) Control of cell type in *Saccharomyces cerevisiae*: mating type and mating type inter-conversion. In: Strathern JN, Jones EW, Broach JR (eds) The molecular biology of the yeast *Saccharomyces*: life cycle and inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 181–209

Hicks J, Strathern JN, Klar AJS (1979) Transposable mating type genes in *Saccharomyces cerevisiae*. Nature 282:478–483

Hinnebusch AG (1988) Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev 52:248–273

Hinnebusch AG (1994) Translational control of *GCN4*: An *in vivo* barometer of initiation-factor activity. Trends Biochem Sci 19:409–414

Kang S, Chumley FG, Valent B (1994) Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. Genetics 138:289–296

Kelly M, Burke J, Smith M, Klar A, Beach D (1988) Four mating-type genes control sexual differentiation in the fission yeast. EMBO J 7:1537–1548

Klar AJS, Strathern JN, Broach JR, Hicks JB (1981) Regulation of transcription in expressed and unexpressed mating type cassettes of yeast. Nature 289:239–242

Kozak M (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res 12:857–872

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kulmburg P, Mathieu M, Dowzer C, Kelly J, Felenbok B (1993) Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the *creA* repressor mediating carbon catabolite repression in *Aspergillus nidulans*. Mol Microbiol 7:847–857

Leach J, Lang BR, Yoder OC (1982) Methods for selection of mutants and *in vitro* culture of *Cochliobolus heterostrophus*. J Gen Microbiol 128:1719–1729

Luo Z, Freitag M, Sachs MS (1995) Translational regulation in response to changes in amino acid availability in *Neurospora crassa*. Mol Cell Biol 15:5235–5245

Mazo AM, Mashkova TD, Avdonina TA, Ambartsumyan NS, Kisselev LL (1979) An improved rapid enzymatic method of RNA sequencing using chemical modification. Nucleic Acids Res 7:2469–2482

Metzenberg RL, Glass NL (1990) Mating type and mating strategies in *Neurospora*. Bioessays 12:53–59

Miller AM (1984) The yeast *MATa1* gene contains two introns. EMBO J 3:1061–1065

Minehart PL, Magasanik B (1991) Sequence and expression of *GLN3*, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain. Mol Cell Biol 11:6216–6228

Nehlin JO, Carlberg M, H Ronne (1991) Control of yeast *GAL* genes by *MIG1* repressor: a transcriptional cascade in the glucose response. EMBO J 10:3373–3378

Nelson MA, Metzenberg RL (1992) Sexual development genes of *Neurospora crassa*. Genetics 132:149–162

Paluh JL, Orbach MJ, Legerton TL, Yanofsky C (1988) The cross-pathway control gene of *Neurospora crassa, cpc-1*, encodes a protein similar to *GCN4* of yeast and the DNA-binding domain of the oncogene *v-jun*-encoded protein. Proc Natl Acad Sci USA 85:3728–3732

Peters DG, Caddick MX (1994) Direct analysis of native and chimeric GATA specific DNA binding proteins from *Aspergillus nidulans*. Nucleic Acids Res 22:5164–5172

Picard M, Debuchy R, Coppin E (1991) Cloning the mating types of the heterothallic fungus *Podospora anserina* – Developmental features of haploid transformants carrying both mating types. Genetics 128:539–547

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sharon A, Yamaguchi K, Christiansen S, Horwitz BA, Yoder OC, Turgeon BG (1996) An asexual fungus has the potential for sexual development. Mol Gen Genet 251:60–68

Staben C, Yanofsky C (1990) *Neurospora crassa a* mating type region. Proc Natl Acad Sci USA 87:4917–4921

Stucka R, Hauber J, Feldmann H (1987) One member of the transfer RNA-Glu gene family in yeast codes for a minor GAG transfer RNA-Glu species and is associated with several short transposable elements. Curr Genet 12:323–328

Tatchell K, Nasmyth KA, Hall BD, Astell C, Smith M (1981) In vitro mutation analysis of the mating-type locus in yeast. Cell 27:25–35

Turgeon BG, Garber RC, Yoder OC (1985) Transformation of the fungal maize pathogen *Cochliobolus heterostrophus* using the *Aspergillus nidulans amdS* gene. Mol Gen Genet 201:450–453

Turgeon BG, Bohlmann H, Ciuffetti LM, Christiansen SK, Yang G, Schafer W, Yoder OC (1993) Cloning and analysis of the mating type genes from *Cochliobolus heterostrophus*. Mol Gen Genet 238:270–284

VanWert SL, Yoder OC (1992) Structure of the *Cochliobolus heterostrophus* glyceraldehyde-3-phosphate dehydrogenase gene. Curr Genet 22:29–35

Werner M, Feller A, Messenguy F, Piérard A (1987) The leader peptide of yeast gene *CPA1* is essential for the translational repression of its expression. Cell 49:805–813

Werner M, Feller A, Delbecq P, Pierard A (1990) Translational control by arginine of yeast gene *CPA1*. In: McCarthy JEG, Tuite MF (eds) Post-translational control of gene expression. Springer-Verlag, Berlin, pp 337–346

Wirsel S, Turgeon BG, Yoder OC (1996) Deletion of the *Cochliobolus heterostrophus* mating type (*MAT*) locus promotes function of *MAT* transgenes. Curr Genet 29:241–249

Zuker M, Jacobson AB (1995) 'Well-determined' regions in RNA secondary structure prediction: analysis of small subunit ribosomal RNA. Nucl Acids Res 23:2791–2798