# Sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase genes from the basidiomycetes *Schizophyllum commune*, *Phanerochaete chrysosporium* and *Agaricus bisporus*

Martin C. Harmsen<sup>1</sup>, Frank H. J. Schuren<sup>1</sup>, Serge M. Moukha<sup>1,\*</sup>, Carin M. van Zuilen<sup>1,2</sup>, Peter J. Punt<sup>2</sup>, and Joseph G. H. Wessels<sup>1</sup>

<sup>1</sup> Department of Plant Biology, University of Groningen, Kerklaan 30, NL-9751 NN Haren, The Netherlands

<sup>2</sup> Medical Biological Laboratory TNO, P.O. Box 45, NL-2280 AA Rijswijk, The Netherlands

Received March 2, 1992/Accepted May 26, 1992

Abstract. GPD genes encoding glyceraldehyde-3-phosphate dehydrogenase were isolated from the homobasidiomycetes Schizophyllum commune, Phanerochaete chrysosporium and Agaricus bisporus. All three species contain one transcriptionally active GPD gene, but A. bisporus also contains an inactive GPD gene (tandemly linked to the active gene). These genes contain 5-9 introns located at conserved positions, differing (except in one case) from intron positions in ascomycetous GPD genes. The predicted amino-acid sequences of the proteins encoded by the three active GPD genes are highly homologous. A comparison with protein sequences from filamentous ascomycetes shows a clear distinction, whereas the GPD genes from ascomycetous yeasts are quite distinct from both the filamentous ascomycetes and basidiomycetes. Promoter regions of ascomycetous GPD genes do not correspond to those of the GPD genes of basidiomycetes which may (partly) explain poor expression in basidiomycetes of introduced genes driven by an ascomycete GPD promoter.

**Key words:** Glyceraldehyde-3-phosphate dehydrogenase (GPD) – Basiodiomycete – Sequence – Evolution

## Introduction

Glyceraldehyde-3-phosphate dehydrogenase (GPD or GAPDH, E.C. 1.2.1.12) is a key enzyme in glycolysis. The *GPD* gene is highly and constitutively expressed: 2-5% of the poly(A)<sup>+</sup> RNA of yeast may comprise GPD mRNA (Holland and Holland 1978; Edens et al. 1984), while up to 5% of the cellular protein may constitute GPD protein (Krebs 1953). The active holoenzyme is a tetramer composed of identical subunits, each with an approximate M, of 38 000.

The development of transformation systems for homobasidiomycetes, based on selection for hygromycin B resistance and using heterologous expression signals to express the bacterial hygromycin B phosphotransferase gene, has been unsuccessful (Casselton and de la Fuente Herce 1989; Mooibroek et al. 1990; Challen et al. 1991; Royer and Horgen 1991). Only in *Laccaria laccata* have positive results been reported (Barrett et al. 1990). The isolation of strong endogenous promoters to drive the expression of heterologous genes in basidiomycetes would thus seem essential. Tho this end we have cloned the *GPD* genes from *Schizophyllum commune, Phanerochaete chrysosporium* and *Agaricus bisporus*.

Throughout the Kingdoms of organisms amino-acid sequences of GPD proteins are highly conserved (Smith 1989; Michels et al. 1991), which allows for an evolutionary analysis. The nucleotide and predicted amino-acid sequences of the homobasidiomycete *GPDs* are therefore compared with *GPDs* from several ascomycetes, the heterobasidiomycete *Ustilago maydis*, and with some plants and animals.

## Materials and methods

Isolation of GPD genes. (1) S. commune. A genomic library of partial Sau3A genomic fragments of the S. commune monokaryotic strain 4-39 in the BamHI site of  $\lambda$ EMBL4 was screened with the 1.2 kb Scal-Sacl fragment form pAN5-22, containing 267 codons of the Aspergillus nidulans gpdA gene (Punt et al. 1988). Hybridization was in  $6 \times SSC$ , 0.5% SDS,  $5 \times$  Denhardt's solution, 1.5 mg ml<sup>-1</sup> denatured salmon sperm DNA at 58 °C; washings were done twice with 6×SSC, 0.1% SDS, 0.1% NaPPi and twice with 3×SSC, 0.1% SDS, 0.1% NaPPi at 58 °C, each for 30 min. Seven positive clones were isolated, all of which contained a 2.9 kb HindIII fragment that strongly hybridized with the A. nidulans gpdA gene. The 2.9 kb HindIII fragment from  $\lambda 121$  was subcloned (p121-9) and sequenced. (2) P. chrysosporium. A genomic library of the P. chrysosporium heterokaryotic strain ME-446, ATCC34541 in *l*EMBL3 (Schrank et al. 1991) was screened with the 1.4 kb PvuII-HindIII fragment from p121-9, encoding the carboxy-terminal 315 aa of the S. commune GPD gene. Hybridization was in 0.5 M sodium phosphate pH 7.2, 7% SDS and 1% BSA (Church and Gilbert 1984) at 54°C. Non-hybridized probe was removed by washing twice with 0.1 M

<sup>\*</sup> Present address: LBCF/CBAI/INRA, Centre de Grignon, F-78850 Thiverval-Grignon, France

448



Fig. 1. Partial restriction maps of p121-9, pPhc428 and pGBB5 that contain the GPD gene(s) of S. commune, P. chrysosporium and A. bisporus, respectively. Exons and introns are indicated under each map by black and white boxes respectively, whereas the tran-

scription direction is indicated by an arrow. The GPD<sup>Ag1</sup> gene, for which no transcripts have yet been found, could be transcribed in the same direction as the  $GPD^{Ag2}$  gene

			•-1	50			~~~~	+ - + -				***	*~~~	*-*	t		•-	100				_									•-5	0								
aar	grg	cağı		ayyı	.cyc	c y	cuya	rara		alle	cyrt	lly	icay	1.81	cyat	cyya		licy	yycy	r 99	gtat	8999	gcg	rgre	gee	cyci	gici	66 6	teti		cag	cact	ccca	tcc	agag	cac	1100	ctet	CC	
cat	cgc	atco	ca	tcad	aca	a t	•1 a At M	G gt	gago	accg	cgc	gcga	cgg	cgag	ctcg	ca g	ga <u>cg</u>	ctga	<u>c</u> at	ctgc	•50 tcgc	gat	ag G	ICC G A	TG A V	AG 0 K	ITT G V	GA A G	IT A	N C	igg 1 G	TC G F	i gtg	icgta	icgg	• ccca	100 Igati	tgg c	gc	10
aga	tct	c <u>9c</u> 0	: ct	gact	:ccc	tc	cgca	g GT G	CGC R	ATT	GGC G	CGT R	•15 ATC I	0 GTT V	CTC L	CGT R	AAT N	GCC A	CTT L	CAG Q	CTG L	GGC G	AAC N	ATC	GAG E	GTC V	GTC V	GCC A	ATO	200 : AAC N	GA D	gtga	igtgo	ag c	gago	tcag	ıt ti	tccat	cttc	34
aag	tct	gacg	) cga	•2 atct	250 ag	C C 	CG T P	TC A F	TC G I	ICC C A	TT G	AC T D	AC A Y	TG G M	TCT. V	AC A Y I	TG T M 1	TC A F	AG T K	AC G Y	AC A	300 CC G T	TC C V	AC G H	GC C G	GC T R	AC A Y	AG G K	GC ≠ G	ICC G	TC G V	AG G E	ITG A V	AG G K	AC G D	igc A G	JAG ( K	•35 2TC G	0 TC V	67
GTC V	GA D	C 60 0	ic c/	AT G H	icc . A	ATC I	ACC T	GTC V	TTC F	GCT A	GAG E	AAG K	AAC N	CCC P	GCC A	•4 GAC D	DO ATC I	AAG K	TGG ₩	GGC G	TCT S	GCT A	GGC G	GCC A	GAC D	TAC Y	ATC I	GTC V	GAG	i tcc s	ACC	gta	•450 icgto	ata	tatg	cgct	.cc (	cattt	cccgc	98
act	<u>tcc</u>	taac	cto	ctct	cgt	ag I	•5 GGT G	00 GTC V	TTC F	ACC T	ACC T	GTC V	GAG	AAG K	GCG A	TCC ( S	CTG ( L	CAC H	TTG L	CAG N	GGC G	GGC I G	GCG A	•55 AAG K	0 AAGI K	GTC V	GTC V	ATC I	TCG S	GCG A	CCC P	TCC S	GCC A	GAC D	GCC A	CCG P	ATG M	TTC F	GTC V	131
GTC V	•6 GG G	00 C GT V	CA/	AC C N	TC I L	GAC D	AAG K	TAC Y	GAC D	TCG S	AAG K	TAC Y	CAG N	GTC V	ATC I	TCG S	AAC N	GCG A	650 TCC S	TGC C	ACG T	ACC T	AAC N	TGC C	CTG L	GCG A	CCG P	CTC L	GCG A	i AAG K	GTC V	ATC	CAC H	GAC D	•7 AAG K	00 TAC Y	: GGC G	C ATC	GCG A	170
GAG E	GG G	r ct L	C A1	TG A M	CC J	ACC T	GTG V	CAC H	GCC A	ACG T	ACC T	GCG A	•75 ACC T	CAG N	AAG K	ACC T	GTC V	GAC D	GGC G	CCG P	TCG S	CAC H	AAG K	GAC D	TGG W	CGC R	GGT G	GGC G	CGC R	800 TCG S	GTC V	AAC N	AAC N	AAC N	ATC	ATC I	CCC P	C TCG S	TCG S	209
ACT T	GG G	r go A	C GC	CT A	AG I K	GCC A	•8! GTC V	50 GGC G	AAG K	GTC V	ATC 1	CCC P	TCG S	стс L	AAC N	666 6	CGC R	CTC L	AC T	gtac	stct	ga ti	•9 ctgt	00 ccag	a tc	tcct	tagc	ggc	c <u>tcc</u>	<u>taa</u>	tete	ttc	ссса	cag	C GG G	iC CT i L	C G(	•950 3C TT & F	C CGT R	233
GTC V	CC P	G AC	C C1	rc G -	AC ( D	STC V	TCC S	GTC V	GTC V	GAC	CTC L	GTC V	GTC V	CGC R	•10 ctc L	DOO GAG E	AAG K	GAG E	GCG A	TCG S	TAC Y	GAC D	GAG E	ATC I	GTC V	GCG A	ACC	GTC V	AAG K	GAG E	GCG A	•10 TCC S	50 GAG E	GGC G	CCG P	CTC L	AAC K	3 GGC G	ATC I	272
CTC L	GG G	דז : F	C AC	CC G	AC ( D	GAG E	TCC S	GTC V	GTC V	1100 TCG S	ACC T	GAC D	TTC F	ACC T	GGC G	GCG A	AAC N	GAG E	TCG S	TCC S	ATC I	TTC F	GAC D	TCC S	AAG K	•1 GCC A	150 GGC G	ATC I	GCG A	ATC I	AGC S	AAG K	i AGC S	TTC F	GTG V	AAG K	CTC L	: ATT I	GCC A	311
tgg W	TAI Y	GA D	•1 C A4	1200 AC G	AG '	rgg W	GGC G	TAC Y	TCC S	CGC R	CGT R	GTC V	TGC C	GAC D	CTG L	CTC L	GTC V	TAC Y	GCG A	GCG A	AAG K	CAG N	GAC D	GGC G	GCG A	CTC L	TAA	gcg	gcga	agc	aaga	cggc	gc g	ctct	aagt	1300 g cc	tgat	cctt		337
taa.	acq	:tta	tac	ttt	acci	t c1	tacat	taga	c ta	aacc	ite agag	350 agti	aaaa	aa i	aaa	aaat	ta ta	acd	• ttat	1400 a ca	acat	atct	aat	aatt	ata (	cgat	ccca	аа												

Fig. 2. Nucleotide sequence and derived amino-acid sequence of the GPD gene of S. commune. Exon sequences are capitalized, conserved internal intron sequences are underlined. Numbering of nucleotides is with respect to the start of the coding sequence. Putative TATA and CAAT boxes are single and double overlined, respectively.

The putative polyadenylation signal of GPD<sup>Sc</sup> has been double underlined. The major tsp is marked with an asterisk. The last base of the transcript is marked with an arrow. A 27-nt repeat in the 3' end of the gene has been double overlined

gcg	cgat	tga	a tạ	ggag	gcg:	aa t	ctcc	gata	t cca	aatco	gatg	acga	8880	ct t	taaa	tacc	g to	gccs	gccgt	; cCi	gtcc	ctcc	tco	ccat	aac	tttc	aatc	ct c	cctc	caco	:C C1	acac	tact	t tg	Jac .	ATG M	CCG P	gto	:agt	a	2
сас	caca	acaş	<b>;</b> c	ccga	accg	cg a	cgac	cacg	t gc	tgac	•50 ttcg	ctti	ccag	GTC V	K AAA	GCA A	GGA G	ATC I	C AAC N	5 GG G	gta	agcc	gcc	gatg	atgo	•100 gc g	tttt	agga	ic ga	cga	acto	acc	gag	1999	t <u>tt</u>	<u>ctga</u>	i <u>c</u> tc	gtç	jcag		9
ŢŢ	•150 TC ( F	) G g1	tga	gtad	ccc i	gcgc	atac	at go	cacci	aaaga	a cco	ccg <u>ta</u>	ictg	agca	tcca	200 Ica e	ag (	IT CO	GT A1 R 1	TG	GC C	GTA R	TC ( I	STC C V	TC C	GT A R	AT G N	ACT C	TC C	ττι L	CAC ( H	GA G	250 AC 1 D	I I	GAC D	GTC V	GTC V	GCG A	3 GT V	c	32
AAC N	GA D	gta	acg	tcci	tc t	caag	tccc	c cc'	•3 ttac	00 tcag	gga	cgcg <u>i</u>	tc	tgacg	00001	a aa	iccti	tacc	t ca	) C I	P	TTC F	ATT I	•35 GAC D	O CTC L	GAG E	TAC Y	ATG M	gtat	gtc	:ag '	ttta	tac	gc g	tege	tacc	:c c	acag	•4 gcga	00 tg	42
gcg	atti	gca:	a	gcaa	aagc	ccg	aca <u>c</u>	gctg	<u>ac</u> 9	cgtg	tgct	t cca	ag G'	•450 ГСТ/ / \	) NCA1 Y P	1 G T1 1 I	IC A/	AG T/	AC GJ Y î	AC T	CC G S	TT C V	AC ( H	GGT ( G	IGC 1 R	TC A F	AG G K	igt t g	cc o s	NTC I	•500 GAG ( E	) GCC / A	AG I K	GAC D	GGC G	AAG K	CTC L	TA1 Y	t gt V	C	68
GAG E	GGI G	C AV	AG (	CCT P	ATC I	CAC H	GTC V	TTC F	•55 GCT A	GAG E	AAG K	GAC D	CCG P	GCC A	AAC N	ATC I	CCC P	TGG W	GGC G	TCT S	GTC V	GGC G	GC A	T GAG	TAC Y	600 : ATC I	сто V	GAG E	S TCC	i aci	G GG G	1 G T C V	i TTI F	C AC T	C AC T	C AC 1	:C G	iag ∕ E	AAG K	GCC A	107
TCT S	GCI A	C C/	•65 AC H	0 TTG L	AAG K	GGC G	GGT G	GCC A	AAG K	AAG K	GTC V	ATC I	ATC 1	TCG S	GCC A	CCC P	TCT S	GCT A	GAT D	•70 GCC A	0 CCC P	ATG M	5 TT F	T G10 V	: TGC C	: GG1 G	GTC V	C AAC N	: ст( L	C GA	C GCI A	G TAC Y	GAI D	C TC S	CAA K	.G TA	-750 NC A	AG I K	GTC V	ATC 1	146
TCG S	AA( N	C G(	CG (	tcc s	TGC C	ACC T	ACC T	AAC N	TGC C	TTG L	GCG A	CCC P	CTT L	•800 GCC A	) AAG K	стс V	ATC I	CAC H	GAC D	AAG K	TTC F	GGC G	AT I	т GT0 V	G CAG	G GG1 G	CTC L	: ATG M	ACC T	C AC	850 GTI V	C CAC H	GC A	T AC	C AC T	c GC	:Т А 4	CC ( T	CAG Q	AAG K	185
ACC T	GTI V	C G/ [	AT I	GGC G	CCC P	TCG S	i AAC N	•9 AAG K	00 GAC D	TGG W	CGT R	GGT G	GGC G	CGT R	TCC S	GTC V	GGC G	AAC N	AAC N	ATC I	ATC I	CCC P	: TC S	G TCO S	•9 <u>5</u> i AC1 T	60 1 GG1 G	GCC	GCC A	: AAG K	G GC A	с GTi V	000 C G	: AA K	G GT V	C AT I	c c(	;тт	CG ( S	CTC L	AAC N	224
GGC G	AAI K	•100 G C'	00 TG 4	AAC N	gtg	acga	ttt	acga	cctc	tc a	tgtg	attt	gg;	acgad	g <u>tg</u>	ctaa	1050 aggad	) ctc (	ccga	ag	GGT G	стс	GCC A	TTC F	CGT R	GTC V	CCC P	ACC T	GTC V	GAT D	GTC V	TCC S	•11 GTC V	00 GTC V	GAC D	CTC L	: GT V	'C GT	TC C V	GT R	247
CTC L	GAI E	G A/	AG (	CCC P	GCT A	TCT S	TAC Y	GAC D	GAG E	ATC I	1150 AAG K	CAG Q	GCC A	ATC I	AAG K	GAG E	GCG A	TCG S	GAG E	ACC T	ACC T	CAC H	: AA K	G GG( G	: ATC I	сто L	•1 GGC G	1200 : TAC Y	C ACC	C GA E	G GA	G AAO K	GT V	C GT V	C TC S	C A(	:С G	AC ' D	TTC F	ACC	286
000 0	AAI N	C G/ I	AC J	AAC N	•12 TCG S	50 TCG S	ATC I	TTC F	GAT D	CGT R	GAC D	GCG A	GGT G	ATC I	GCG A	CTC L	AAC N	AAG K	ACA T	TTC F	• ото V	1300 AAG K	) 5 CTI L	C ATO I	: TCC S	: TGC W	S TAC Y	GAC D	: AAG N	GA E	G TGI W	G GG( G	TA Y	C TC S	C CC	C C(	атт ₹	GC ' C	•13 TGC C	GAC D	325
CTC L	CTO L	C G(	GT (	TAC Y	GCC A	GCG A	i AAG K	GTC V	GAC D	GGT G	GCC A	CTC L	TAA	cgto	gcto	1400 ac (	) cgaco	cgtg	ct g	tac	ggtc	c to	399	cggad	ata	igaaa	itta	agaa	agci	145 tgt	) acag	tgtal	c t	gcgg	ttgt	a co	cat:	taa	tc g	locod	337
atg	tat	ttaa	a a:	•15 acca	500 aaaa	ta g	atgc	atggi	a gg	attg	taca	aata	icaci	ttg 1	ctc	jacgo	a ti	•15 ttta	50 tgcti	tt:	gcgc	gatg	) ca	tcgaş	lcac	aaco	tcgg	jcg t	tcc	gtc	ga t	•160 gacci	)0 Igca	9							

Fig. 3. Nucleotide sequence and derived amino-acid sequence of the GPD gene of P. chrysosporium. For details, see legend to Fig. 1

sodium phosphate pH7.2, 0.2% SDS at 54 °C for 30 min. Four positive phages were isolated. Of these  $\lambda$ Phc42 and  $\lambda$ Phc52 contained strongly hybridizing *Eco*RI fragments of 8 and 7 kb, respectively. The 8 kb fragment from  $\lambda$ Phc42 was subcloned (pPhc428) and relevant parts sequenced. (3) *A. bisporus*. A genomic library of partial *Sau*3A fragments of *A. bisporus* (strain Horst<sup>®</sup> U3), constructed in the *XhoI* site of  $\lambda$ FIX II, was screened with the 1.2 kb *ScaI-SacI* fragment form pAN5-22, containing the *A. nidulans gpd*A sequence (Harmsen et al. 1991). Hybridizations were done as described for the isolation of *P. chrysosporium* clones. One positive clone ( $\lambda$ ABU3-412) contained a strongly hybridizing 5 kb *Bam*HI fragment that was subcloned (pGBB5) and sequenced.

A-100 .

Sequence analyses. After restriction mapping a series of overlapping subclones of p121-9, pPhc428 and pGBB5 were constructed in pUC-derived vectors. Both strands from the derived subclones were sequenced at least twice by the dideoxy chain-termination method using T7 DNA polymerase (Tabor and Richardson 1987).

Transcript mapping. The transcription start points (*tsp*) of the *GPD* genes were determined using primer extension with AMV reverse transcriptase. Total ssRNA (20 µg) in 120 µl 250 mM KCl, 10 mM Tris-HCl pH 8.3, 1 mM EDTA containing 66 ng of the <sup>32</sup>P-labelled oligonucleotide was heated at 80 °C for 3 min after which annealing took place at 45 °C for 1 h. The mixture was then brought to 160 µl 188 mM KCl, 70 mM Tris-HCl pH 8.3, 7.5 mM MgCl<sub>2</sub>, 6 mM DTT, 1.25 mM of each dATP, dTTP, dCTP, dGTP and 20 U of AMV reverse transcriptase. Reverse transcription was done at 42 °C for 30 min. The products were precipitated and analyzed on a 6% sequence gel alongside with the products of sequencing reactions to determine the *tsp*. The following were used as primers: 5'-ATTCC-

CGGGCATTATTGTGTGATGGGATG-3' for S. commune, 5'-GAGGACGATACGGCC-3' for P. chrysosporium and 5'-CAT-TACGGAGGACAATG-3' for A. bisporus. The 3' end of the S. commune GPD mRNA was determined by S1 nuclease protection mapping, performed as previously described (Schuren and Wessels 1990), using the 511 nt XhoI-HindIII fragment cloned into M13mp18.

Computer analyses. The complete nucleotide and derived aminoacid sequences were analyzed using the PC/GENE programs package (version 6.01). The amino-acid sequence and the position of introns was determined by comparison with known GPD aminoacid sequences and by comparison with intron consensus sequences. Multiple sequence alignments were done with CLUSTAL (Higgins and Sharp 1988) and the PALIGN program of PC/GENE which both use the algorithm of Myers and Miller (1988). Sequences from the GPD genes of U. maydis (Smith and Leong 1990), A. nidulans (Punt et al. 1988), Aspergillus niger (R. F. M. van Gorcom unpublished results) Curvularia lunata (Osiewacz and Ridder 1991), Cryphonectria parasitica (Choi and Nuss 1990), Saccharomyces cerevisiae 1 (Holland and Holland 1980), S. cerevisiae 2 (Holland and Holland 1979), S. cerevisiae 3 (Holland et al. 1983), Kluyveromyces lactis (Shuster 1990), Zygosaccharomyces rouxii (Imura et al. 1987), tobacco (Shih et al. 1986), hamster (Vincent and Fort 1990), human liver (Arcari et al. 1984), chicken (Stone et al. 1985), Drosophila melanogaster (Tso et al. 1985), Homarus americanus (Davidson et al. 1967), Schistosoma mansoni (Goudot-Crozel et al. 1989), Caenorhabditis elegans (Huang et al. 1989), maize (Brinkmann et al. 1987) and mustard (Martin and Cerff 1986) were obtained from the EMBL (Release 27) and SWPROT (Release 18) data libraries or manually entered using PC/GENE.

449

•1

	•-	250														•-	200										-	_		•-	150							
attga	gatg	g acq	atgg	ttt a	itgag	actt	g ac	gago	catta	gcg	agtg	act	gcat	caca	ta g	ccgt	gcto	c go	tttg	ttcc	tca	tgta	tcc	tgta	tatg	aa t	cacc	aatg	a ga	caac	aact	tac	gcaa	agg	atga	acgta	gt	
				_	100.												-50																					
acctt	ggtg	c aad	ctcg	tat g	tgag	atct	a ca	aaga	aaga	tat	atag	Itgg	agaa	agta	ca a	ccgt	gtaa	ig ga	ccto	cttc	att	cacg	tcc	agag	acac	ac a	gaga	aata	g ta	tcat	aatt	cac	•1 gAT M	G gt	aatt	ctct		1
									•50															•100														
ttctt	ttgg	g ata	agtt	gaa a	cccg	laacg	la gg	l <u>aact</u>	<u>taat</u> c	ttt	cact	cgg	tgta	g GT V	T AA N	T GT V	TGG	A AT	C AA	T GG G	gta	cgtt	ctt	gaga	aatg	tt t	cata	actt	t tt	catc	cttt	tat	atgc	ctc	gtac	gata	g	8
•15 G TTT	) Gg	taagt	ttgt	ggaa	ctgc	ct g	ittgc	atca	aa to	gc <u>ca</u>	ctaa	<u>a</u> ta	aaaa	200 aaa	tatg	aag	GG A	GG A	tc G	GA C	GT C	TC G	rc c	TCA	GA A	AT G	CA T	TA C	AA A1	250 IG C	AA A'	TT C	TC A	CT G	ITT C	STA G	CT	
•																	6	ĸ	1	6 1	ĸ	L '	v	L	к	N .	A I	- '		1	-	1 :	L	T	v	ν.	A	30
GTT AJ V I	AT GA	Agt )	tgag	catc	cttt	cacc	•30 aa c	0 aatg	gccat	a aa	gggt	gtg <u>c</u>	gct	<u>tac</u> c	jag :	gcta	tatc	ta g	тс	CT T' P I	rc c	•350 TT G/ L I	AC G	TTG, V I	AA T	AC A' Y I	TG g M	taaa	gcgc1	t tc	tatc	tact	ggt	ctca	ctt	e. gctt	400 <u>c</u>	41
actca	aggt	t tta	tag	GCG T A	AT C Y	TG T L	TCA F	AG T K	TAT G	AT T D	CC G	TT C V	AT G H	•45 GAC G	GA T	AT C Y	AA G Q	GA A G	AA G K	TCG/ VE	AA A	CC A/	AG G	AC GI	GG A. G I	AA T K	TG A	IC A	TT GA	AT G	500 GA C/	AT AJ	AA A	TC G	icg g A	ICT T	TC F	75
									-55	0																												
GCAG/	A CI	GT GA R E	A CCI P	GCA	AAT N	ATT I	AAA K	TGG W	GCC A	GAT D	TGC C	GGC G	GCT A	GAG E	TAC Y	ATC I	GTT V	GAA E	TCT S	ACT T	gtg	agtag	gat :	gtgci	ecgg	tt c	ttg	tac	c cta	act	ttg <u>a</u>	gcta	aga	99 <u>9</u>	cata	tag (	GGC G	98
	650														•700	0														•	750							
GTG T		KA AC	A GA/ E	GAA E	TT L	gtgt	gtaa	tg a	tatg	ttgg	a at	ctgt	ttcg	gat	gcga	aat	tg <u>at</u>	<u>cgta</u>	<u>t</u> ct	tgggi	tag /	A GCO A	G AAG K	G GAO E	G CA' H	L	G AAG K	G GG1 G	r GGG G	GCC A	: AA/ K	A AA/ K	A GT V	r GT V	C AT I	C ACO T	C GCT A	121
CCT C	· .		60 CT		C AC	а та	с ст <b>и</b>	•8 • • •	100 100			T CT												•8	50													
P (	iu /	S	G	A CC / P	τ Τ	Y	V	V	G	V	N	L	D	I AA/ K	Ŷ	D	P	K	E	V	V	i gta	agta	атаа	tati	atato	et d	tato	agcg	it ad	gegg	jcaaa	a aag	jaa <u>a</u>	tctg	<u>ac</u> ga	38	144
•9( ataaca	00 ag A1	TT TC	A AAT N	GCT A	TCG S	TGC C	ACT T	ACC T	AAT N	TGC C	CTA L	GCA A	GTC V	CTG L	GCG A	•9 AAG K	50 GTC V	ATC I	AAT N	GAC D	AAA K	TTT F	GGA G	ATT I	GTG V	GAA E	GGC G	TTG L	ATG M	ACG T	ACA T	GTG V	•100 CAT H	GCC A	ACC T	ACA T	GCC A	181
										•	1050																•11	00										
ACG CA	IG A/	G AC	T GTO V	GAT D	GCT A	CCT P	GCA A	AAG K	i AAG K	GAT D	TGG W	CGT R	TCT S	GGA G	AGG R	AGT S	GTT	ACA T	AAT N	AAC N	ATC I	ATT I	CCA P	GCA A	TCT S	ACG T	GGT G	GCC A	GCT A	AAA K	GCT A	GTT V	ACA T	AAG K	GCG A	ATT I	CCT P	220
					•11	50												•	1200																•12	50		
GAT TI D L	'G G/	G GG	A AAA K	L	AC : T	gtac	gtcag	ga c	tcgt1	tatt	t tc	caaag	gacc	ttca	igc <u>c</u>	<u>ic t</u>	<u>gac</u> a;	gcga	c tg	tggca	ġТ	GGA G	CTG L	GCA A	TTC F	CGA R	GTC V	P	ACA T	CTC L	GAC D	GTA V	TCG S	GTT V	GTT V	GAC D	L	243
												•	1300																•13	50								
	r co r F		E	K AAG	GAA E	ACC T	AGT S	Y	D	GAC D	GTC V	AAA K	AAA K	GCC A	M	AGG R	GAC D	GCA A	GCC A	GAC D	GGT G	AAA K	CAC H	P	GGC G	ATC I	GAG E	AAA K	GGC G	I	GTC V	GAC D	TAT Y	ACG T	GAA E	GAA E	GAC D	282
GTT G1 V V	тт / 9	C AC	C GA1 D	TTC F	GTT V	GGG G	•140 AGC S	00 AAC N	TAT Y	TCG S	ATG M	ATC I	TTT F	GAC D	GCA A	AAA K	GCC A	GGG G	ATC I	GCG A	TTG L	AAC N	•1 TCG S	450 CGT R	TTT F	ATG M	AAG K	TTG L	GTT V	GCA A	tgg W	TAT Y	GAT D	AAT N	GAG E	TGG W	GGA G	321
		1500															•	1550													•16	00						
TAT GO Y A	G CO	T AG	A GTO V	TGC C	GAT D	GAG E	GTT V	GTG V	TAT Y	GTA V	GCG A	AAG K	AAG K	AAT N	TAA -	gag	gtcc	jca i	agta	gattg	a aa	gttc	agta	i cgt	tttt	aac	aata	gago	at t	ctcş	aggo	t tg	cgto	atto	:tg	tgtca	ggc	337
Fig	4 4	R	Nu	cleo	tide	s sei	ane	nce	, an	d d	eriv	ed	ami	no-	acio	d se	an	enc	e of	? the	e tu	o t	and	eml	lv li	inke	ed (	7PI	) ge	enes	i of	A	hisi	nor	us:	GP	$D^{Ag}$	<sup>1</sup> (A)

Fig. 4A, B. Nucleotide sequence and derived amino-acid sequence of the two tandemly linked *GPD* genes of *A. bisporus; GPD*<sup>Ag1</sup> (A) and  $GPD^{Ag2}$  (B). A hexameric repeat of aa in the predicted protein of  $GPD^{Ag2}$  has been *underlined* (B). For other details, see legend to Fig. 1

## Results

## GPD sequences

The GPD gene of S. commune was isolated first, using the gpdA gene from A. nidulans as a probe. Southern blots of P. chrvsosporium DNA hybridized more strongly with the GPD gene from S. commune than with that of A. *nidulans*. The former gene was, therefore, used to isolate the GPD gene from P. chrysosporium. Both probes hybridized to the same fragments on Southern blots of A. bisporus DNA, but the S. commune GPD gene also hybridized faintly to other fragments. Consequently, the A. nidulans GPD gene was used to recover the A. bisporus gene. Unexpectedly, sequencing of the 5 kb BamHI fragment from clone pGBB5 revealed the presence of two GPD genes on this fragment (Fig. 1). The complete sequences of the GPD genes of S. commune (GPD<sup>Sc</sup>, Fig. 2), P, chrysosporium (GPD<sup>Pc</sup>, Fig. 3) and A. bisporus (GPD<sup>Ag1</sup> and GPD<sup>Ag2</sup>, Fig. 4A, B) were determined from overlapping subclones of p121-9, pPhc428 and pG-BB5, respectively. The predicted GPD proteins (shown in

Figs. 2–5) all contain 337 aa residues, except for  $GPD^{Ag2}$  which encodes 338 aa. Of the two tandemly linked GPD genes of *A. bisporus* only  $GPD^{Ag2}$  appears highly expressed in both mycelium and fruit bodies (Harmsen et al. 1991). The mRNA lengths for  $GPD^{Sc}$ ,  $GPD^{Pc}$  and  $GPD^{Ag2}$  were determined to be approximately 1.25 kb in all three cases (data not shown).

For all four GPD genes TATA box-like elements can be discerned (overlined in Figs. 2-4). GPD Sc, GPD Pc, GPD<sup>Ag1</sup> and GPD<sup>Ag2</sup> contain five, six, nine and nine short introns, respectively, with an average size of 56 nt. The initiator codon for Met is immediately followed by an intron, except in GPD<sup>Pc</sup> where a Pro is present between Met and the first intron. In GPD<sup>Pc</sup>, GPD<sup>Ag1</sup> and  $GPD^{Ag2}$  another small exon of 5 nt (encoding Phe) is present. The average G + C contents of  $GPD^{Sc}$ ,  $GPD^{Pc}$ and GPD<sup>Ag1+Ag2</sup> are 60%, 58% and 47%, respectively, similar to reported values (58%, 59% and 43-45%, respectively) for total genomic G+C composition (Dons et al. 1979; Raeder and Broda 1984; Horgen et al. 1984). The coding parts of the genes have a 5-14%higher G + C fraction than the introns and other noncoding regions.

tag	icag	ttt	a t	aag	cgtt	ga (	ggato	taga	ng ct	•- gctg	150 Itttc	cgc	gtct	cga i	atgt	tctc	gg t	gttt	aggg	jg tt	agca	atct	gat	•-1 atga	00 taa	taat	ttgt	ga t	gaca	tcga	it ag	taca	aaaa	- a cca	ccaa	ttcc	ggt	•-5 caca	0 tcc	
асс	* atc	tcc	g t	ttt	ctcc	ca	tctad	acad	a ac	aagc	ttat	cgc	•1 c at M	G gt	ttgt	ctct	cgc	ttgc	atac	cat	ccag	c <u>ag</u>	ctca	<u>c</u> tga	tg t	cgac	•5 ttgt	0 ag	GTT V	AAA K	GTT V	GGA G	ATC I	AAC N	GG G	gta	agtg	tt t	ttgt	8
cgt	cgc	gct	g t	•1 ggt	00 tccg	iga 1	tcatg	tcag	<u>a c</u> t	ttgg	gtgt	ctt	gcag	T T	TC G F	gtg	agtg	acc	•150 acco	) :tgca	itt c	tggc	tata	t gc	gtga	<u>tact</u>	gac	catc	gct	caag	I GT G	CGT R	•20 ATC I	00 GGC -G	CGC R	ATT I	GTC V	CTC L	CGT R	18
AAT N	GC A	t c	TC L	CAA Q	TTC F	CAC Q	G GAC D	: ATC I	GAA E	•25 GTT V	GTC V	GCC A	GTG V	AAC N	GA D	gtgg	gtga	ct t	atgt	gtcc	c at	atct	atcg	at <u>a</u>	gcta	300 <u>ac</u> a	attc	atgg	ca g	сс	CG T P	TC A F	ITT O	GAC ( D	CTC L	GAA E	TAC Y	ATG : M	gtacg	41
atg	• atc	350 tag	ia g	ttt	atca	ca a	attcg	atag	ıt ga	tc <u>ac</u>	cgta	<u>t</u> at	gcag	GCA A	TAC Y	ATG M	400 TTC F	AAG K	TAC Y	GAC	TCC S	GTC V	CAC H	GGT G	CGC R	TTC F	AAG K	GGT G	ACC T	GTT V	GAG	GTC V	•4 AAG K	50 6 AAC N	C GG G	C AG	C TT F	T GT V	C GTT V	67
GAC D	GG G	CA	GG R	сст Р	ATG M	i AA/ K	A GTO V	: TTT F	GCT A	GAA E	•50 CGC R	0 GAT D	ccc P	GCT A	GCC A	ATC I	CCT P	TGG W	GGT G	TCA S	GTC V	GGC G	GCG A	GAC D	TAC Y	GTC V	стс V	550 GAA E	TCC S	ACA T	gtg	cgtc	ctg	acto	ctga	ctt	ggta	ttga	tc	97
tta	tct	eat	• <u>c t</u>	600 ctt	<u>at</u> ct	ac g	gtcga	iccta	Ig GG G	T GT	A TT	C AC T	T AC T	T ATO I	C GA	CAA K	gtg	cgtt	•650 atc	) gatg	cgag	ca a	gcaa	tcat	t ca	tatc	ttct	gat	gttt	ctg	cag	G GC A	•7 T TC	'00 :G G( ; /	CT C	AC T' H I	TG A L	AG G K	GG GGC G G	113
GCC A	AA. K	A A	AA K	GTC V	GTT V	ATO	с тсс s	GCT	ССТ	TCG S	750 GCC A	GAT D	GCG A	CCG P	ATG M	TAT Y	GTC V	TGC C	GGT G	GTT V	AAC N	CTT L	GAC D	AAG K	TAC Y	AAT N	•8 CCC P	00 AAG K	GAC D	ACA T	ATT I	gta	cgto	gca	tta	catc	gtt	gttt	ttgtat	144
tac	agg	•85  ttg	0 at	ttte	sgtc	gt g	ggtag	I ATC	TCG S	AAC N	GCT A	TCT S	TGC C	ACA T	ACC T	AAT N	•9 TGC C	00 TTG	GCT A	ACT	CTT L	GCT A	AAA K	GTC V	ATT I	CAC H	GAT D	AAC N	TTT F	GGT G	ATC I	GTT V	GAG	•95 GG1 G	50 ГСТ L	G ATI M	G AC	C AC	T GTT V	176
CAC H	•1000     •1050 AC GCC ACC ACC GCT ACT CAA AAG ACT GTG GAT GGT CCT TCT CAC AAG GAC TGG CGT GGT GGC CGT GGT GTC GGC AAT AAC ATC ATT CCT TCC TCT ACT GGC GCC GCC AAG GCC GTC H A T T A T Q K T V D G P S H K D W R G G R G V G N N I I P S S T G A A K A V 215     •1100     •115																																							
GGA G	+1000 C GCC ACC GCT ACT CAA AAG ACT GTG GAT GGT CCT TCT CAC AAG GAC TGG CGT GGT GGC GGT GGT GGT GGC GAT AAC ATC ATT CCT TCC TCT ACT GGC GCC AAG GCC GTC A T T A T Q K T V D G P S H K D W R G G R G V G N N I I P S S T G A A K A V 215 +1100 61150 61 AAG GTT ATC CCT TCA CTG ACG GGC AAG CTC AC gtatgtttga ttgtgtggct gtctagccct tgtact <u>cact aat</u> tctctgt catgcatag T GGT CTC TCG ATG CGT GTT CCC ACT CAG GAC G K V I P S L N G K L T G L S M R V P T Q D 237																																							
GTT V	TCI S	C G1 \	TT I	GTC V	GAT D	CTT L	GTT V	GTT V	CGT R	CTT L	GAG E	AAG K	CCC P	•12 GCT A	250 TCC S	TAT Y	GAA E	CAG		AAG K	GAG E	GTC V	ATG M	CGC R	AAG K	GCC A	GCT A	GAA E	GGC	GAA E	•13 TAC Y	00 AAG K	GGA G	ATT	TAT	C GC/	₹ TA; γ	C AC	C GAC	276
GAG E	GA( D	: G1 \	TG I	GTT V	TCC S	ACT T	GAC	• TTC F	1350 ATT I	AGT S	GAT D	AAC N	AAT N	TCT S	TGT C	GTC V	TTC F	GAT D	GCG A	AAG K	GCC A	GGA G	ATT	CAG	•14 CTT	400 AGC S	CCG	AAC	TTT	GTC	AAG K	CTG	ATT	GCT	TG	G TAC	GA D	T AAI	C GAA	315
TGG ₩	66/ 6	•1 \ 74 \	145) AC Y	D TCG S	CGC R	CGT R	GTT V	TGC C	AAC N	СТС	CTC	CAA	TAC	GTT V	GCA A	AAG K	GAG	GAC	GCC	1500 AAG K	GCT	GGC	ATT	TAG	atag	gttg	ctt g	jaati	gcgc	cg c	tcgt	caaa	a aa	•1 gaaa	550 atc:	g aad	ettt <sup>.</sup>	ttat	-	778
agt	gtaa	ntgg	g ti	atca	agti	tta	gaat	atgc	g ct	•10 gttc	600 tgtg	atti	cati	tg t	gttt	agaa	- sg tg	gcgt	taag	g ga	•16 tgata	50 atat	tgat	tacat	tg a	atgga	atgta	a ga	aatg	catg	a tc	acgt	gctt	ttt	•1 gate	700 99cg	tgai	agtt	tcc	550
Fiş	g. 4	B	(0	on	tin	uec	1)																																	

# Protein and codons

The amino-acid sequences of the derived GPD proteins are highly conserved (Fig. 5). The codon usage of  $GPD^{sc}$ and  $GPD^{Pc}$  is heavily biased with 63% and 60% C at the third position of the codons, respectively. This is less pronounced for  $GPD^{Ag2}$ , which has only 40% codons that end with C.  $GPD^{Ag1}$ , for which no transcript was found, shows no clear codon bias.

## Discussion

# Gene structure and expression

A consensus TATA box (TATAAAA) is present only in  $GPD^{Sc}$  (overlined in Fig 2). In  $GPD^{Pc}$  and  $GPD^{Ag2}$  this element may be represented by TTAAAT and TACAAAAA, respectively, because these sequences are located at the expected position, i.e., at -31 to -36 nt from the major *tsp* (overlined in Figs. 3 and 4B). In  $GPD^{Ag1}$ , homologs of a TATA box, TACAAA and TACAAA, are found at -96 nt and -68 nt from the start

codon (overlined in Fig. 4A). In the three active GPD genes these putative TATA elements are followed by pyrimidine-rich stretches which generally precede the tsp in fungal genes (Gurr et al. 1987). Sequences conserved between A. nidulans and A. niger GPD promoter regions, such as a gpd box, pgk box, qut box and qa box (Punt et al. 1990), are not present in the promoter regions of the four basidiomycetous GPD genes. These features, in addition to the absence of clear TATA boxes in these genes may (partly) explain why bacterial genes driven by the A. nidulans GPD promoter are poorly expressed in the basidiomycetes S. commune (Mooibroek et al. 1990) and Coprinus cinereus (Casselton and de la Fuente Herce 1989). Remarkably, the reverse does not seem to hold because the isocitrate lyase gene of the basidiomycete C. cinereus has been expressed in A. nidulans (Hynes 1989); similarly, the ADE2 gene from the adenine biosynthetic pathway of S. commune also functions in Neurospora crassa (Alic et al. 1990).

The  $GPD^{Ag2}$  gene is strongly expressed in both mycelium and fruit bodies of *A. bisporus* (Harmsen et al. 1991). Only 223 bp separate the *tsp* of this gene from the last putative codon of the inactive  $GPD^{Ag1}$  gene. Neither in

<u>S. commune</u> <u>P. chrysosporium</u> <u>A. bisporus 2</u> <u>A. bisporus 1</u>	MAVKYGINGFÖRIGRIVERNALQLGNIE VVAINDPFIALDYNYMFKYDTVRGRYKGTVEVKDGKLVVDGHAITVFAEKNPADIKU NPVKAGINGFGRIGRIVERNALLIGO IDVVAVNDPFIDLEYNAYNFKYDSVHGRFKGSVEAKDGKLYVDGKPINYFAKDPANIPU H VKVGINGFGRIGRIVERNALGPGDIEVAVNDPFIDLEYNAYNFKYDSVHGRFKGTVEVKNGSFVYDGRPHKYFAERDPANIPU H -VNVGINGFGRIGRIVERNALGPGDIEVAVNDPFIDLEYNAYNFKYDSVHGRFGGTVETKGGKLIIDGHKKAAFAEREPANIKU
<u>S. commune</u> P. chrysosporium <u>A. bisporus 2</u> <u>A. bisporus 1</u>	GSAGADY I VESTGVFTTVEKASL HUGGAKKVVI SAPSADAPHFVVGVALDKYDSKYDVI SNASCTINCLAPLAKVI HDKYGIAEG BSVDAEY I VESTGVFTTTEKASAHLKGGAKKVVI SAPSADAPHFVCGVALDAYDSKYDVI SNASCTINCLAPLAKVI HDKFGI VGG GSVGADYVVESTGVFTTI DKASAHLKGGAKKVVI SAPSADAPHYCGVALDKYDPKTI SIAASCTINCLATLAKVI HDKFGI VGG ADCGAEY I VESTGVFKTEELAKEHLKGGAKKVVI TAPGSGVPTYYVGVALDKYDPKTVI SNASCTINCLAVLAKVI HDKFGI VEG
<u>S. commune</u> <u>P. chrysosporium</u> <u>A. bisporus 2</u> <u>A. bisporus 1</u>	LMTTVHATTATOKTVDOPSHKDURGORSVNINI I IPSSTDAAKAVGKVI PSI NGRE I GLAFRVPTLDVSVVDLVVRLEKERSYDE IV LNTTVHATTATOKTVDOPSNKDURGORSVGNI I IPSSTDAAKAVGKVI PSI NGRE I GLAFRVPTDVSVVDLVVRLEKPSYDE IV LNTTVHATTATOKTVDOPSNKDURGORSVGNI I IPSSTDAAKAVGKVI PSI NGRE I GLAFRVPTDVSVVDLVVRLEKPSYDE I LNTTVHATTATOKTVDAPAKKDURSGRSVTNI I IPASTGAAKAVGKVI PSI NGRE I GLAFRVPTLDVSVVDLVVRLEKETSYDDVK
<u>S. commune</u> <u>P. chrysosporium</u> <u>A. bisporus 2</u> <u>A. bisporus 1</u>	ATVKERSEGPLKGILGFIDESVVSTDFTGANESSIFDSKAGIAISKSFVKLIAVVDNENGYSRRVCDLEVYAAKODCAL GAIKEASETTHKGILGYTEEKVVSTDFTGNDNSSIFDRDAGIALNKTFWKLISVVDNENGYSRRVCDLEVGAAKVDCAL EVMRKARGEYKGIINTDEDVSTDFTGNDNSCVPAKAGIGISNFVKLIAVVDNENGYSRRVCDLEVYVAKKN KAMRDAADGKHPGIEKGIVDYTEEDVVSTDFVGSNYSMIFDAKAGIALNSREMKLVANVDNENGYARRVCDEVVYVAKKN





Fig. 6. Comparison of intron positions in *GPD* genes; *horizontal lines* represent the coding region (336–338 codons) of each *GPD* gene. The *vertical bars* mark the position of introns

mycelium nor in fruit bodies did we detect transcripts of GPD<sup>Ag1</sup>. This suggests that all sequences necessary for efficient transcription of the GPD<sup>Ag2</sup> gene are located within this relatively short stretch of DNA. The predicted protein of GPD<sup>Ag1</sup> is homologous to the highly expressed GPDs of all three basidiomycetes (Fig. 5), although GPD<sup>Ag1</sup> and GPD<sup>Ag2</sup> do not cross-hybridize due to a large number of differences in the third base of codons which often conserve the encoded amino acid. Also the main residue defining the binding site of glyceraldehyde-3-phosphate, Cys-150 (marked with an arrow in Fig. 5), is present in the putative gene product of GPD<sup>Ag1</sup>, suggesting that the gene does encode a functional GPD protein. In addition, all nucleotide sequences required for transcription and subsequent splicing appear to be present. Therefore, we suggest that the  $\tilde{GPD}^{Ag1}$  gene is active at some other phase of the life cycle or under particular environmental conditions, for example during anaerobiosis occurring in the lower layers of compost during the colonization phase. In maize, of three GPD genes coding for cytosolic proteins, one is only expressed during anaerobiosis (Russell and Sachs 1989). However, the possibility that GPD<sup>Ag1</sup> is a pseudogene cannot be excluded.

Although the number of introns in the four homobasidiomycetous GPD genes differ, their positions are strongly conserved (Fig. 6); even to the nucleotide where a codon is interrupted (cf. Figs. 2-4). The position of the single intron in the GPD gene of the heterobasidiomycete U. maydis is also conserved among the three investigated homobasidiomycetes. Within the ascomycetous species that have been studied (Fig. 6), intron positions in the GPD genes are also conserved; contrary to an earlier report (Osiewacz and Ridder 1991) all intron positions in the Curvularia lunata GPD gene are also present in the Aspergillus GPD genes. On the other hand, only the position of one intron is found to be conserved between basidiomycetes and ascomycetes (Fig. 6). An intron is also found at this position in chicken but the other introns in metazoan GPD genes are located differently from those in fungi. This comparison suggests that there is a clear boundary between ascomycetes and basidiomycetes with respect to intron positions within the GPD genes.

## Amino-acid sequence comparisons

Table I shows the large extent of homology between GPD proteins in various organisms. Except for GPD<sup>Ag1</sup>, the similarity of GPD proteins within the filamentous fungi (ascomycetes and basidiomycetes) is higher than between filamentous fungi and yeasts. The unique position of the yeasts is further illustrated by the construction of a tree representing amino-acid sequence divergence between GPDs predicted with the CLUSTAL program (Fig. 7). The tree is essentially similar to the phylogenetic trees constructed by Smith (1989), using an evolutionary parsimony method (Lake 1987) on GPD nucleotide sequences, and Michels et al. (1991), who examined GPD aminoacid sequences using a distance matrix method (Fitch and Margoliash 1967). However, we have expanded the number of fungal GPDs. Our analysis reveals that basidiomycetes and ascomycetes are two diverged groups of fungi, which agrees well with the analysis of intron positions (Fig. 6). This is in accordance with the classification



Fig. 7. Phylogenetic tree of similarities in amino-acid sequences of GPD genes from various organisms. The  $GPD^{Ag1}$  gene product differed greatly from all other known sequences

Table 1. Comparison of GPD amino-acid sequences with the Myers and Miller (1988) algorithm: percentage amino-acid identitiy between all analyzed fungal species, a plant and a mammalian species

Sp	ecies	Basie	diomy	cetes			Filar asco	nento mycet	us es		Yeas asco:	t-like mycet	es			Othe	r
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	A. bisporus 1	100	70	69	70	64	63	64	59	62	62	61	61	64	59	62	63
2	A. bisporus 2		100	76	77	70	69	69	- 67	70	62	63	63	66	63	67	70
3	P. chrysosporium			100	81	75	71	71	70	69	60	62	62	62	62	68	71
4	S. commune				100	72	69	69	69	67	64	64	64	67	62	69	70
5	Ustilago maydis					100	72	72	73	69	63	65	63	64	67	68	73
6	Aspergillus nidulans						100	90	81	79	63	66	65	64	64	67	70
7	Aspergillus niger							100	80	77	64	67	66	65	67	69	72
8	Curvularia lunata								100	82	62	64	63	64	65	69	69
9	Cryphonectria parasitica									100	64	65	64	66	65	68	69
10	Saccharomyces cerevisiae 1										100	95	87	80	79	64	65
11	Saccharomyces cerevisiae 2											100	87	82	80	67	66
12	Saccharomyces cerevisiae 3												100	81	78	66	64
13	Kluvveromyces lactis													100	78	68	65
14	Zygosaccharomyces rouxii														100	68	64
15 16	Nicotiana tabacum (tobacco) Cricetulus griseus (hamster)															100	68 100

based on morphological characteristics. However, the classification of the analyzed ascomycetous yeasts within the ascomycetes is not reflected in GPD similarities. It is conceivable that this divergence from the filamentous ascomycetes reflects the unicellular mode of growth of yeasts and/or their facultative fermentative metabolism. With respect to unicellular growth, a mutation in a single nucleus could more readily be fixed in a population than in a multicellular filamentous fungus, thus allowing for a more rapid rate of evolution. Another conclusion from Fig. 7 is that the *GPDs* from filamentous fungi appear more closely related to the *GPDs* from metazoa than to those from plants.

cus bisporus stock cultures and fruit bodies. E. Wolbert, J. Scheer and T. A. Schuurs are acknowledged for sequencing large parts of the *GPD* genes. Dr. R. F. M. Van Gorcom is thanked for providing the sequence of the *A. niger GPD* gene prior to publication. Dr. T. Boekhout from CBS, Delft (yeast division), is thanked for his critical reading of the manuscript. This research was supported by the Netherlands Foundation for Technical Research (STW), by the Foundation for Biological Research (BION) and by the European Community (EC). The nucleotide-sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Databases under the accession numbers M81724, M81754, M81727 and M81728 for *GPD*<sup>se</sup>, *GPD*<sup>Pe</sup>, *GPD*<sup>Ag1</sup> and *GPD*<sup>Ag2</sup>, respectively.

# References

Acknowledgements. We thank Dr. L. J. L. D. van Griensven, director of the Mushroom Experimental Station, for the supply of Agari-

Alic M, Clark EK, Kornegay JR, Gold MH (1990) Curr Genet 17:305-311

- Arcari P, Martinelli R, Salvatore F (1984) Nucleic Acids Res 12:9179-9189
- Barrett V, Dixon RK, Lemke PA (1990) Appl Microbiol Biotechnol 33:313-316
- Brinkmann H, Martinez P, Quigley F, Martin WF, Cerff R (1987) J Mol Evol 26: 320-328
- Casselton LA, de la Fuente Herce A (1989) Curr Genet 16:35-40
- Challen MP, Rao BG, Elliott TJ (1991) In: Van Griensven LJLD (ed) Genetics and breeding of *Agaricus*. Pudoc, Wageningen, pp 129-135
- Choi GH, Nuss DL (1990) Nucleic Acids Res 18:5566
- Church GM, Gilbert W (1984) Proc Natl Acad Sci USA 81: 1991– 1995
- Davidson BE, Sajgo M, Noller HF, Harris JI (1967) Nature 216:1181-1185
- Dons JJM, De Vries OMH, Wessels JGH (1979) Biochim Biophys Acta 563:100-112
- Edens LE, Bom J, Ledeboer AM, Maat J, Toonen MY, Visser C, Verrips CT (1984) Cell 37:629–633
- Fitch WM, Margoliash E (1967) Science 155:279-284
- Goudot-Crozel V, Caillol D, Djabali M, Dessein AJ (1989) J Exp Med 170: 2065-2080
- Gurr SJ, Unkles SE, Kinghorn JR (1987) In: Kinghorn JR (ed) Gene structure in eukaryotic microbes. IRL Press, Oxford, pp 93-139
- Harmsen MC, Scheer J, Schuurs TA, Wessels JGH (1991) In: Van Griensven LJLD (ed) Genetics and breeding of Agaricus. Pudoc, Wageningen, pp 140–145
- Higgins DG, Sharp Pm (1988) Gene 73:237-244
- Holland MJ, Holland JP (1978) Biochemistry 17:4900-4907
- Holland JP, Holland MJ (1979) J Biol Chem 254:9839-9845
- Holland MJ, Holland JP (1980) J Biol Chem 255:2596-2605
- Holland JP, Labieniec L, Swimmer C, Holland MJ (1983) J Biol Chem 258: 5291-5299
- Horgen PA, Arthur R, Davy O, Moum A, Herr F, Strauss N, Anderson J (1984) Can J Microbiol 30: 587-593

- Huang XY, Barrios LAM, Vonkhorporn P, Honda S, Albertson DG, Hecht RM (1989) J Mol Biol 206:411-424
- Hynes MJ (1989) Exp Mycol 13:196-198
- Imura T, Utatsu I, Toh-E A (1987) Agric Biol Chem 51:1641-1647
- Krebs EG (1952) J Biol Chem 200:471-478
- Lake JA (1987) Mol Biol Evol 4:167-191
- Martin WF, Cerff R (1986) Eur J Biochem 159: 323-331
- Michels PAM, Marchand M, Kohl L, Allert S, Wierenga K, Opperdoes FR (1991) Eur J Biochem 198:421–428
- Mooibroek H, Kuipers AGJ, Sietsma JH, Punt PJ, Wessels JGH (1990) Mol Gen Genet 222:41-48
- Myers EW, Miller W (1988) Cabios 4:11-17
- Osiewacz HD, Ridder R (1991) Curr Genet 20:151-155
- Punt PJ, Dingemanse MA, Jabobs-Meijsing BJM, Pouwels PH, van den Hondel CAMJJ (1988) Gene 69:49-57
- Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RDM, Pouwels PM, van den Hondel CAMJJ (1990) Gene 93:101-109
- Raeder U, Broda P (1984) Curr Genet 8:499-506
- Royer JC, Horgen PA (1991) IN: Van Griensven LJLD (ed) Genetics and breeding of Agaricus. Pudoc, Wageningen, pp 135–140
- Russell DA, Sachs MM (1989) The Plant Cell 1:793-803
- Schrank A, Tempelaars C, Sims PFG, Oliver SG, Broda P (1991) Mol Microbiol 5:467-476
- Schuren FHJ, Wessels JGH (1990) Gene 90:199-205
- Shih MC, Lazar G, Goodman HM (1986) Cell 47:73-80
- Shuster JR (1990) Nucleic Acids Res 18:4271
- Smith TL (1989) Proc Natl Acad Sci USA 86:7063-7066
- Smith TL, Leong SA (1990) Gene 93:111-117
- Stone EM, Rothblum KN, Alevy MC, Kuo TM, Schwartz RJ (1985) Proc Natl Acad Sci USA 82:1628-1632
- Tabor S, Richardson C (1987) Proc Natl Acad Sci USA 84:4767– 4771
- Tso JY, Sun XH, Wu R (1985) J Biol Chem 260:8220-8228
- Vincent S, Fort P (1990) Nucleic Acids Res 18:3054

Communicated by K. Esser