The mating-type region of Schizosaccharomyces pombe h−S 972: sequencing and analysis of 69 kb including the expressed mat1 locus

ARTICLE *in* YEAST · JULY 2000

Impact Factor: 1.63 · DOI: 10.1002/1097-0061(200008)16:11<1061::AID-YEA603>3.0.CO;2-6

7 AUTHORS, INCLUDING:

Valerie Wood

University of Cambridge

57 PUBLICATIONS **6,269** CITATIONS

SEE PROFILE

Karen Moore

University of Exeter

42 PUBLICATIONS **1,434** CITATIONS

SEE PROFILE

Cherryl Hunt

University of Exeter

26 PUBLICATIONS **1,394** CITATIONS

SEE PROFILE

Stephen Aves

University of Exeter

44 PUBLICATIONS **2,204** CITATIONS

SEE PROFILE

Yeast
Yeast 2000; **16**: 1061–1067.

Yeast Sequencing Report

The mating-type region of *Schizosaccharomyces pombe* h^{-S} 972: sequencing and analysis of 69 kb including the expressed *mat1* locus

Zheng Xiang[1]†, Valerie Wood[2], Maire-Adèle Rajandream[2], Barclay G. Barrell[2], Karen Moore[1], Cherryl Hunt[1] and Stephen J. Aves[1]*

[1] *School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK*
[2] *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK*

*Correspondence to:
S. J. Aves, School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK.
E-mail: S.J.Aves@exeter.ac.uk

†Present address:
National Institute for Medical Research, Division of Virology, Mill Hill, London NW7 1AA, UK.

Received: 5 January 2000
Accepted: 29 March 2000

Abstract

The sequence has been determined of 68 897 bp of genomic DNA including the expressed *mat1* mating-type locus from *Schizosaccharomyces pombe* h^{-S} strain 972. The DNA sequence, located on the long arm of fission yeast chromosome II and contained in two cosmid clones, was analysed to reveal one autonomously replicating sequence, two retrotransposon long terminal repeats (LTRs), one $tRNA^{Gly}$ gene and 33 open reading frames (ORFs), of which 15 contain introns. Nine of these ORFs code for previously described genes (*trt1*, *rpl10*, *rps21*, *nif1*, *sui1* (*psu1*), *matMi*, *matMc*, *let1* and *rpa4*), one of which (*trt1*) contains 15 introns, the highest number yet recorded in a gene of *S. pombe*. Of the remaining 24 ORFs, sequence similarity suggests that the function of 13 of the encoded proteins may be predicted and these include four mitochondrial proteins, two transport proteins, two signalling molecules, a component of serine palmitolytransferase, a homologue of 3-methyladenine DNA glycosylase, a multifunctional alcohol dehydrogenase, a killer toxin sensitivity factor and an acetyl transferase. Six deduced sequences appear to be related to proteins of unknown function in *Saccharomyces cerevisiae* or *S. pombe* and the remaining five are hypothetical proteins. This sequence has been submitted to the EMBL database under the following entries: SPBC23G7 (Accession No. AL035065), SPBC18E5 (AL035077) and SPBC29A3 (part) (AL022299).

Keywords: 26S protease; aminotransferase; ARS element; cell cycle inhibitor; copper transporter; cytochrome *c*1; DNA-3-methyladenine glycosidase; GAP; GEF; haem protein precursor; helix–loop–helix; L-spacer; mitochondrial processing peptidase; mitotic inhibitor; NADH dehydrogenase; NADH-dependent flavin oxidoreductase; protein translation factor; ribosomal protein; telomerase reverse transcriptase; urea transporter

Introduction

Mating type in *Schizosaccharomyces pombe* is determined by DNA sequence present at the expressed *mat1* locus. This locus can contain either mating-type plus (P) or minus (M) information and switching can occur by copy transposition from the silent *mat2-P* and *mat3-M* loci in the wild-type homothallic h^{90} strain (Beach, 1983). For genetic analysis, heterothallic strains unable to switch mating type are preferred, such as h^{-S} and h^{+N} isolated by Leupold (1950). In the case of the h^{-S} 972 strain, the inability to switch is due to a deletion of *mat2-P* sequences, the result of a loopout which has given a *mat2:3-M* fusion (Beach and Klar, 1984).

As part of the *S. pombe* genome sequencing project, we have sequenced 68.9 kb of h^{-S} 972 genomic DNA, including the *mat1* locus and most of the 'L-spacer' separating this from the silent *mat2:3-M* locus. We present here an analysis of this sequence, which includes nine known genes,

two retrotransposon long terminal repeats (LTRs) and genes encoding a further 24 predicted proteins and one tRNA.

Methods and materials

Strains, cosmids and vectors

Cosmids c18E5 and c23G7 were obtained from Rhian Gwilliam, the DNA coordinator for the *S. pombe* genome sequencing project. These cosmids contain partial *Sau*3A-digested genomic DNA fragments from *S. pombe* h^{-S} strain 972, cloned into the unique *Bam*HI site of the cosmid vector Lawrist4 (Hoheisel *et al.*, 1993), and their inserts overlap by 3.3 kb. Bacteriophage vector M13mp18 (Messing and Vieira, 1982) was used for subcloning in host *Escherichia coli* strain TG1 (Stratagene) *supE thi-1* Δ*(lac-proAB)* Δ*(mcrB-hsdSM)5* (r_k^- m_k^-) [F′ *traD36 proAB* $lacI^qZ\Delta M15$].

Sequencing strategy and analysis

Sequencing was by a shotgun strategy in M13 followed by gap-filling by primer walking to obtain complete coverage on both strands. Single-stranded phage DNAs were prepared and sequenced using BigDye dye terminator kits (PE Applied Biosystems) as previously described (Xiang *et al.*, 1999). Products were analysed on an ABI377 automated sequencer (PE Applied Biosystems) and read lengths of 500–700 bases (<1% N residues) were obtained for 90% of sequence. Primer walking was performed on existing templates using custom primers (MWG Biotech GmbH) to ensure that all the sequence had been obtained on both strands and to check ambiguities. Sequence quality was checked by visual inspection of gel trace files. 1254 random and 40 directed sequencing reactions were performed in total for cosmids c23G7 and c18E5: the final average number of reads per base pair was 7.0 and all base pairs were sequenced on both strands.

Sequence assembly was performed as previously described (Xiang *et al.*, 1999). DNA sequence was compared with the *S. pombe* sequencing project finished and unfinished data and with the EMBL Nucleotide Sequence Database using BLAST 2.0 (Altschul *et al.*, 1997), against a non-redundant protein database (SWISSPROT + TrEMBL + TrEMBLNew) using BLASTX, and against *S. pombe* ESTs using Est2genome. Gene prediction was performed as previously described (Xiang *et al.*, 1999). Sequences of predicted proteins were compared against the non-redundant protein sequence database using FASTA2 (Pearson and Lipman, 1988) and scanned for Pfam motifs (Sonnhammer *et al.*, 1998).

Results and discussion

DNA sequencing

The overlapping *S. pombe* inserts in cosmid clones c18E5 and c23G7 were sequenced on both strands to give a total 68 897 bp of contiguous genomic sequence, which has been submitted to the EMBL database under entries: SPBC23G7 (Accession No. AL035065), SPBC18E5 (AL035077), and SPBC29A3 (part) (AL022299). Cosmid c18E5 overlaps cosmid c29A3 (sequenced by G. Volckaert, Katholieke Universiteit, Leuven) by 24 242 bp and the two sequences are 100% identical. *S. pombe* genomic sequencing results can be viewed at: www.sanger.ac.uk/Projects/S_pombe/genomic_sequence.shtml.

Sequence analysis

The sequenced region (Figure 1) has a GC content of 50.2% and was analysed as described in Materials and methods. This region contains 33 known genes or open reading frames (ORFs) longer than 100 nucleotides that are potential protein coding regions, giving a density of one gene per 2.1 kb. Fifteen of these ORFs (45%) incorporate one or more introns. The ORFs have been numbered according to *S. pombe* genome project conventions, e.g. for SPBC23G7.01c, SP = *S. pombe*, B = chromosome 2, c23G7 = cosmid name, .01 = first ORF, c = complementary strand. Table 1 gives details of all 33 ORFs and their predicted protein products, 17 of which contain motifs from the Pfam database which are indicative of function.

Of the 33 ORFs, nine are formerly identified *S. pombe* genes (*trt1*, *rpl10*, *rps21*, *nif1*, *sui1 (psu1)*, *matMi*, *matMc*, *let1*, *rpa4*), and three novel genes show greatest similarity to *S. pombe* genes, although one of these is of unknown function. Eleven of the 21 other novel genes have products which show homology with proteins of known function (FASTA scores >100) from other species, five have homology with proteins of unknown function from other species, and five show no marked homologies.

In addition to the protein coding genes the region harbours a gene encoding a $tRNA^{Gly}$, two single

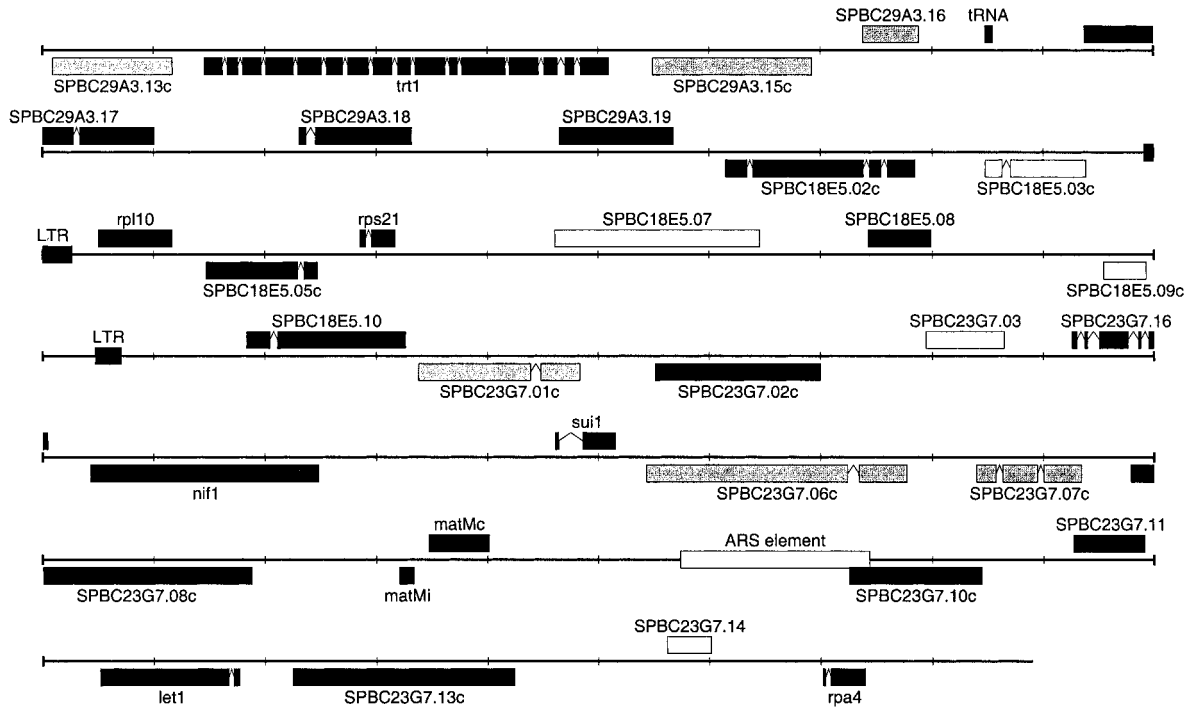

Figure 1. Map of the sequenced region showing previously identified genes and ORFs with similarity to known genes (bars in solid black), ORFs with similarity to genes of unknown function (grey bars) and ORFs with no significant similarity to any sequences in the databases (white bars). ORFs and the tRNA gene reading on the top strand are shown above the line, those on the bottom strand below the line. Introns are signified by bridged gaps in the bars

LTRs and an autonomously replicating sequence (ARS).

Previously described genes

Nine of the 33 ORFs within the sequenced c18E5 and c23G7 cosmids are previously described *S. pombe* genes (Table 1), of which four encode translation-related proteins: 60S ribosomal protein L10 (Masson *et al.*, 1996; EMBL Q09127) (*rpl10*), 40S ribosomal protein S21 (Itoh *et al.*, 1985; PROSITE PO5764) (*rps21*), 60S acidic ribosomal protein P2-beta (Beltrame and Bianchi, 1990; TrEMBL P17478) (*rpa4*), and protein translation factor sui1 (psu1) (TrEMBL P79060); two encode mating-type M-specific polypeptides Mi (*matMi*) and Mc (*matMc*) (Kelly *et al.*, 1988; TrEMBL P10839 and P10840); and the three remaining genes *trt1*, *nif1*, *let1* encode, respectively: telomerase reverse transcriptase 1 (Nakamura *et al.*, 1997; TrEMBL O13339), nif1, an inhibitor of mitosis (Wu and Russell, 1997; TrEMBL P87159), and 26S protease regulatory subunit 8 (Michael *et al.*, 1994; P41836).

The two polypeptides encoded by the *mat1* locus are both mating-type minus (M)-specific in the heterothallic h^{-S} 972 strain (Beach and Klar, 1984). The Mi polypeptide (also termed Mm) is only 42 amino acids in length—it should be noted that this is much smaller than the arbitrary cut-off of 100 amino acids used in deciding whether or not to annotate novel ORFs in the *S. pombe* genome sequencing project. The region of DNA we have sequenced does not include the silent *mat2:3-M* locus present in the h^{-S} 972 strain, but 14.9 kb of DNA separating *mat1* from *mat2:3-M* has been sequenced representing most of this 'L-spacer' (estimated at 18 kb; our unpublished data). The L-spacer sequence is essential for viability and contains the *let1* and *rpa4* genes, each of which has been shown to be essential (Michael *et al.*, 1994; Beltrame and Bianchi, 1990). We have identified a further four ORFs in this region.

Located between *mat1-M* and *let1* and overlapping part of the 3′ end of SPBC23G7.10c is

Table 1. Characteristics of open reading frames (ORFs) identified in cosmids c18E5 and c23G7

ORF name[1]	Intron	Total ORF length (aa)	M_R	pI	Highest FASTA score (opt)	Best homology[2]	Pfam motifs
SPBC29A3.13c	0	359	42.4	9.4	245	*S. pombe* SPBC215.07C hypothetical protein O94312	PF00855 PWWP domain
SPBC29A3.14c	15	988	116.3	10.3		*trt1*	PF00078 reverse transcriptase (RNA-dependent DNA polymerase)
SPBC29A3.15c	0	476	55.2	9.7	211	*S. cerevisiae* YGL129C, hypothetical protein Q01163	PS00038 Myc-type, helix–loop–helix dimerization domain signature
SPBC29A3.16	0	166	19.8	10.3	387	*S. cerevisiae* YOR294W hypothetical protein Q08746	
SPBC29A3.17	1	525	62.5	6.8	280	*S. pombe* SPAC24H6.09 hypothetical protein Q09763	PF00621 RhoGEF domain
SPBC29A3.18	1	307	35.6	9.4	1118	*Neurospora crassa* cytochrome c1 haem protein precursor P07142	PS00190 cytochrome *c* family haem-binding site signature
SPBC29A3.19	0	342	39.7	5.6	257	*Lactococcus lactis orfb* O86281	
SPBC18E5.02c	3	509	59.5	6.0	1130	*Cricetulus griseus* Lbc1 serine palmitoyltransferase subunit1 O54695	PF00222 aminotransferase class II
SPBC18E5.03c	1	277	32.7	6.5			
SPBC18E5.04	0	221	26.1	10.8		*rpl10*	PF00826 ribosomal L10e
SPBC18E5.05c	1	314	36.9	5.0	288	*S. cerevisiae* Ikl1 P38874	
SPBC18E5.06	1	87	9.7	5.0		*rps21*	PF01249 ribosomal S21e
SPBC18E5.07	0	615	68.0	4.8			
SPBC18E5.08	0	186	21.9	9.2	142	*Streptomyces coelicolor* SCJ9A.01 putative acetyl transferase	
SPBC18E5.09c	0	128	15.2	6.9			
SPBC18E5.10	1	452	51.8	8.6	1029	*C. elegans* NADH dehydrogenase complex subunit Q17880	
SPBC23G7.01c	0	454	50.9	8.8	159	*S. cerevisiae* Yeb5 YEL015W hypothetical protein P39998	
SPBC23G7.02c	1	494	56.7	6.5	1050	*Blastocladiella emersonii* mitochondrial processing peptidase alpha subunit P97997	PF00675 insulinase P50025 mitochondrial carrier
SPBC23G7.03	0	233	28.0	9.9			
SPBC23G7.16	4	148	17.2	10.0	254	*S. cerevisiae* Ctr2 copper transport protein P38865	
SPBC23G7.04c	0	681	75.5	6.4		*nif1*	
SPBC23G7.05	1	109	12.9	8.9		*sui1* (*psu1*)	PF01253 *sui1* translation initiation factor
SPBC23G7.06c	1	745	86.3	9.3	573	*S. cerevisiae* YPR091C hypothetical protein Q06833	
SPBC23G7.07c	2	277	32.6	8.41	162	*S. cerevisiae* YLR003C hypothetical protein Q07897	
SPBC23G7.08c	0	695	79.7	8.6	418	*Cochliobolus heterostrophus* putative GTPase activating protein O13384	PF00620 RhoGAP domain
SPBC23G7.17c	0	42	4.8	9.4		*matMi*	
SPBC23G7.09	0	181	22.0	9.7		*matMc*	PF00505 high mobility group (HMG) box
SPBC23G7.10c	0	395	45.2	6.0	343	*Bacillus subtilis* NADH-dependent flavin oxidoreductase (yqjm) P54550	PF00724 FMN oxidoreductase
SPBC23G7.11	0	213	25.0	7.0	586	*S. pombe* mag1 DNA-3-methyladenine glycosidase Q92383	PF01332 alkylb DNA glycoside AlkA family
SPBC23G7.12c	1	403	46.8	8.3		*let1*	PF00004 ATPase associated with various cellular activities (AAA)
SPBC23G7.13c	0	664	74.5	8.3	1998	*S. cerevisiae* Dur3 urea active transporter P33413	PS00017 ATP/GTP-binding site motif A (P-loop)
SPBC23G7.14	0	131	16.3	5.2			
SPBC23G7.15c	1	113	11.6	4.0		*rpa4*	PF00428 60S acidic ribosomal protein

[1]ORFs SPBC18E5.11c and SPBC18E5.12c in overlapping regions are not listed as they are wholly included within SPBC23G7.01c and SPBC23G7.02c, respectively.

[2]SWISSPROT/TrEMBL database Accession Nos are given. Best homologies are not listed for FASTA scores <100. *S. cerevisiae* and *C. elegans* sequences are from the respective genome sequencing projects (Goffeau *et al.*, 1997; the *C. elegans* sequencing consortium 1998).

ars756 (Maundrell *et al.*, 1988). The position of this origin of DNA replication in relation to the *mat1* locus is significant in relation to the recent proposal that the imprinting of mating-type switching is determined by the orientation of DNA synthesis, such that *mat1* is preferentially replicated by a centromere-distal origin(s), so that the strand-specific imprint occurs only during lagging-strand synthesis (Dalgaard and Klar, 1999).

Interestingly, with around 90% of the genome submitted to EMBL, the telomerase reverse transcriptase gene has the highest number of introns yet found in any *S. pombe* gene. Only 44% of fission yeast genes possess introns and of those, 97% have five or less while only 0.9% have more than seven.

The sequence determined during this project agrees completely with the database entries for *trt1*, *rps21*, *nif1*, *matMi*, *matMc*, *let1* and *rpa4*, but *sui1* and *rpl10* show discrepancies with the original database entries. The nucleotide sequence of *sui1* is identical to that found in SWISS-PROT (Accession No. P79060), where the annotated reading frame began with GTG and the start codon was not identified. However, the recognition of an intron has allowed the identification of an ATG start codon and indicates that the N-terminal region encodes amino acids MSAIQQNFNTVDPF rather than (M)LDPF. In nine sequence reads from independent clones we have found a guanine insertion in *rpl10* when compared to the SWISS-PROT entry Q09127: the insertion occurs between nucleotides 681 and 682 and results in a frame shift within the ORF. This has the effect of shortening the C-terminal region of the predicted protein from LLRLTCNNEQCVN to LA.

Novel genes whose products show homology to proteins of known function

The predicted proteins of 13 of the 24 novel genes identified in the sequenced region have similarity to known proteins (FASTA scores >100) and their function may therefore be predicted. Four may code for mitochondrial proteins, and the structures of another five suggest their association with membrane-related functions.

Of the deduced proteins with mitochondrial-related functions, three are involved in redox reactions and the fourth is a structural protein. The product of SPBC18E5.10 is most similar to a *C. elegans* protein which is one of around 25 different subunits together forming the NADH dehydrogenase complex, responsible for oxidizing NADH as the first step in electron transport. The complex transfers electrons from the reduced flavin to the respiratory carrier co-enzyme Q. The predicted protein of SPBC23G7.10c has some similarity with NADH-dependent flavin oxidoreductase from *Bacillus subtilis (ygjm)*, that catalyses the reduction of free flavins by NADPH or NADH. SPBC29A3.18 encodes a protein that is very similar to cytochrome *c*1 haem protein precursor from other fungi, including *Neurospora crassa*, *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (Gbelská *et al.*, 1996). This enzyme is a component of the cytochrome *b–c*1 complex that accepts electrons from Rieske protein and transfers electrons to cytochrome *c* in the mitochondrial respiratory chain. SPBC23G7.02c encodes a protein with some similarity to an alpha subunit of mitochondrial processing peptidase, Mmm1, that is required for establishing and maintaining the structure of budding yeast mitochondria (Burgess *et al.*, 1994).

SPBC18E5.02c appears to encode a biosynthetic enzyme belonging to class II of pyridoxal phosphate-dependent aminotransferases. This membrane-associated enzyme encodes a component of serine palmitoyltransferase, the enzyme catalysing the first step in sphingolipid synthesis. Two other ORFs encode polypeptides which contain alternating hydrophobic and hydrophilic regions characteristic of integral membrane proteins, but these are involved in transport processes: a putative copper transport protein (from SPBC23G7.16) with 39% identity to Ctr2 from *S. cerevisiae* (Kampfenkel *et al.*, 1995) and a protein similar to that encoded by *DUR3* (SPBC23G7.13c), which is involved in active transport of urea in *S. cerevisiae* (ElBerry *et al.*, 1993). The presence of Rho domains in two ORFs suggests that their polypeptides are involved in signal transduction pathways. The Rho family of G proteins activate effectors involved in the regulation of cytoskeleton formation, cell proliferation and the JNK signalling pathway. G proteins generally have a low intrinsic GTPase hydrolytic activity but there are family-specific groups of GTPase-activating proteins (GAPs) that enhance the rate of GTP hydrolysis. Small G proteins transduce signals from plasma membrane receptors to control a wide range of cellular functions. These proteins are clustered into distinct families but all act as molecular switches, active in their GTP-

bound form but inactive when GDP-bound. SPBC23G7.08c encodes a putative GAP, whereas the hypothetical protein encoded by SPBC29A3.17 ORF may be a guanidine nucleotide exchange factor (GEF).

The product of SPBC23G7.11 shows 45% overall identity to another *S. pombe* enzyme, 3-methyladenine DNA glycosylase (MAG), that is involved in hydrolysis of the deoxyribose N-glycosidic bond to excise 3-methyladenine or 7-methyladenine from DNA damaged by alkylation lesions (Memisoglu and Samson, 1996). 3-methyladenine DNA glycosylases can initiate the base excision repair (BER) of an extraordinarily wide range of substrate bases and most organisms possess families of alkylbase DNA glycosidases (AlkA). The advantage of such broad substrate recognition is that these enzymes provide resistance to a wide variety of DNA-damaging agents (Wyatt *et al.*, 1999). The occurrence of a second MAG-like gene may indicate the presence of an AlkA family in *S. pombe*.

The ORF SPBC18E5.08 has homology with prokaryotic acetyl transferases. SPBC23G7.03 shows very weak similarity to some bacterial translational initiation factors, but at this stage it must be regarded as a hypothetical protein.

SPBC18E5.05c has a low level (26%) of identity to *S. cerevisiae IKI1*: *iki1* mutants show insensitivity to a linear double-stranded plasmid, pGKL1, that encodes a killer toxin consisting of 97 K, 31 K and 28 K subunits that is secreted into the culture medium (Tokunaga *et al.*, 1989). *S. cerevisiae IKI1* and *IKI3* genes show no significant identity with other known genes and they do not belong to any homology domain group, gene family, or superfamily (Yajima *et al.*, 1997).

SPBC29A3.19 has 38% overall identity to *Lactococcus lactis* orfb (Arnau *et al.*, 1998) a multifunctional alcohol dehydrogenase that catalyses the reduction of acetaldehyde to ethanol during fermentation. Five other alcohol dehydrogenases have been identified in *S. pombe*.

Other novel genes

The protein products of five ORFs, SPBC29A3.15c, SPBC29A3.16, SPBC23G7.01c, SPBC23G7.06c and SPBC23G7.07c, each show homology to different hypothetical proteins from *S. cerevisiae*. It is thus likely that these are real proteins, but for now nothing certain can be said concerning their functions. The predicted product of SPBCA3.13c is similar to an *S. pombe* hypothetical protein.

The previously mentioned SPBC23G7.03 and the remaining four ORFs (SPBC18E5.03c, SPBC18E5.07, SPBC18E5.09c and SPBC23G7.14) have no significant similarity to any sequences in the databases and must therefore be regarded as hypothetical proteins. At least one of these is expressed, since the proline-serine-rich protein encoded by SPBC18E5.07 is almost identical between amino acid residues 304 and 548 to a partial cDNA clone (Yoshioka *et al.*, 1997).

Sequences not encoding proteins

In addition to *ars756*, previously mentioned, there is a possibility of further ARS elements within the sequence, since *S. pombe* ARS elements, unlike those in budding yeast, cannot be identified by sequence alone but require functional assays. Two lone LTRs are located in the sequenced region. Solo LTRs represent ancient transposon insertions that have been eliminated from the genome: when the 5′ and the 3′ LTRs of a retroviral element recombine, a circular sequence consisting of a single LTR and the internal coding sequence is often excised from the genome, leaving a solo LTR behind. In *S. pombe*, LTRs have been categorized as Tf1- or Tf2-type (Levin *et al.*, 1990) and the fission yeast genome is littered with numerous lone LTRs from this process. LTRs are widely used as molecular markers for linkage mapping and biodiversity studies. One of the LTRs sequenced here has become degraded, suggesting that it may represent a more ancient insertion than the other, which is clearly a Tf2-type.

The sequenced region harbours a gene encoding $tRNA^{Gly}$ with the anticodon CCC. This is one of nine genes encoding $tRNA^{Gly}$ out of 141 tRNA genes identified so far in *S. pombe*. Many genes encoding tRNAs are clustered in the repeated sequences adjacent to the centromeres (Kuhn *et al.*, 1991), although in many cases the multiple genes coding for the family of tRNAs carrying a given amino acid are dispersed throughout the genome.

Acknowledgements

This work was funded as part of the European Union *Schizosaccharomyces* genome sequencing project (BIO4-960159). We thank Rhian Gwilliam for supplying cosmid

DNA. We also thank Sîan Ellard, Michael Bulman and Andrew Hattersley of the University of Exeter School of Postgraduate Medicine and Health Sciences for assistance with ABI 377 usage.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Arnau J, Jorgensen F, Madsen SM, Vrang A, Israelsen H. 1998. Cloning of the *Lactococcus lactis adhE* gene, encoding a multifunctional alcohol dehydrogenase, by complementation of a fermentative mutant of *Escherichia coli*. *J Bacteriol* **180**: 3049–3055.

Beach DH. 1983. Cell type switching by DNA transposition in fission yeast. *Nature* **305**: 682–688.

Beach DH, Klar AJS. 1984. Rearrangements of the transposable mating-type cassettes of fission yeast. *EMBO J* **3**: 603–610.

Beltrame M, Bianchi ME. 1990. A gene family for acidic ribosomal proteins in *Schizosaccharomyces pombe*: two essential and two non-essential genes. *Mol Cell Biol* **10**: 2341–2348.

Burgess SM, Delannoy M, Jensen RE. 1994. *MMM1* encodes a mitochondrial outer membrane protein essential for establishing and maintaining the structure of yeast mitochondria. *J Cell Biol* **126**: 1375–1391.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *Caenorhabditis elegans*: a platform for investigating biology. *Science* **282**: 2012–2018.

Dalgaard JZ, Klar AJS. 1999. Orientation of DNA replication establishes mating-type switching pattern in *S. pombe*. *Nature* **400**: 181–184.

ElBerry HM, Majumdar ML, Cunningham TS, Sumrada RA, Cooper TG. 1993. Regulation of the urea active transporter gene (*DUR3*) in *Saccharomyces cerevisiae*. *J Bacteriol* **175**: 4688–4698.

Gbelská Y, Horváthová K, van der Aart QJM, Zonneveld BJM, Steensma HY, Subik J. 1996. Isolation and molecular analysis of the gene for cytochrome c_1 from *Kluyveromyces lactis*. *Curr Genet* **30**: 145–150.

Goffeau A *et al.* 1997. The Yeast Genome Directory. *Nature* **387** (suppl): 1–105.

Hoheisel JD, Maier E, Mott R, McCarthy L, Grigoriev AV, Schalkwyk LC, Nizetic D, Francis F, Lehrach H. 1993. High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73**: 109–120.

Itoh T, Otaka E, Matsui KA. 1985. Primary structures of ribosomal protein YS25 from *Saccharomyces cerevisiae* and its counterparts from *Schizosaccharomyces pombe* and rat liver. *Biochemistry* **24**: 7418–7423.

Kampfenkel K, Kushnir S, Babiychuk E, Inzé D, Van Montagu M. 1995. Molecular characterization of a putative *Arabidopsis thaliana* copper transporter and its yeast homologue. *J Biol Chem* **270**: 28479–28486.

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**: 1537–1547.

Kuhn RM, Clarke L, Carbon J. 1991. Clustered tRNA genes in *Schizosaccharomyces pombe* centromeric DNA sequence repeats. *Proc Natl Acad Sci USA* **88**: 1306–1310.

Leupold U. 1950. Die Vererbung von Homothallie und Heterothallie bei *Schizosaccharomyces pombe*. *C R Lab Carlsberg Sér Physiol* **24**: 381–480.

Levin HL, Weaver DC, Boeke JD. 1990. Two related families of retrotransposons from *Schizosaccharomyces pombe*. *Mol Cell Biol* **10**: 6791–6798.

Masson JY, Vadnais J, Ramotar D. 1996. The *Schizosaccharomyces pombe spqM* gene is a new member of the Qm transcription factor family. *Gene* **170**: 153–154.

Maundrell K, Hutchison A, Shall S. 1988. Sequence analysis of ARS elements in fission yeast. *EMBO J* **7**: 2203–2209.

Memisoglu A, Samson L. 1996. Cloning and characterization of a cDNA encoding a 3-methyladenine DNA glycosylase from the fission yeast *Schizosaccharomyces pombe*. *Gene* **177**: 229–235.

Messing J, Vieira J. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**: 269–276.

Michael H, Schmidt H, Fleck O, Gutz H, Liedtke C, Lorentz A, Ostermann K. 1994. The mating-type region of *Schizosaccharomyces pombe* contains an essential gene encoding a protein homologous to human modulators of HIV transactivation. *Gene* **145**: 205–210.

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR. 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**: 955–959.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence analysis. *Proc Natl Acad Sci USA* **85**: 2444–2448.

Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* **26**: 320–322.

Tokunaga M, Kawamura A, Hishinuma F. 1989. Expression of pGKL killer 28K subunit in *Saccharomyces cerevisiae*: identification of 28K subunit as a killer protein. *Nucleic Acids Res* **17**: 3435–3446.

Wu L, Russell P. 1997. Nif1, a novel mitotic inhibitor in *Schizosaccharomyces pombe*. *EMBO J* **16**: 1342–1350.

Wyatt MD, Allan JM, Lau AY, Ellenberger TE, Samson LD. 1999. 3-methyladenine DNA glycosylases: structure, function, and biological importance. *Bioessays* **8**: 668–676.

Xiang Z, Lyne MH, Wood V, Rajandream M-A, Barrell BG, Aves SJ. 1999. DNA sequencing and analysis of a 67.4 kb region from the right arm of *Schizosaccharomyces pombe* chromosome II reveals 28 open reading frames including the genes *his5*, *pol5*, *ppa2*, *rip1*, *rpb8* and *skb1*. *Yeast* **15**: 893–901.

Yajima H, Tokunaga M, Nakayama-Murayama A, Hishinuma F. 1997. Characterization of *IKI1* and *IKI3* genes conferring pGKL killer sensitivity on *Saccharomyces cerevisiae*. *Biosci Biotechnol Biochem* **61**: 704–709.

Yoshioka S, Kato K, Nakai K, Okayama H, Nojima H. 1997. Identification of open reading frames in *Schizosaccharomyces pombe* cDNAs. *DNA Res* **4**: 363–369.

