3. PERIODIC GENE EXPRESSION DURING THE MITOTIC CELL CYCLE IN FISSION YEAST

This chapter will provide a global overview of the periodic expression profiles during the fission yeast cell cycle. Two different synchronisation methods were used in order to study changes in gene expression as a function of time: centrifugal elutriation and temperature sensitive mutants. The results presented in this chapter were derived from the following eight timecourse experiments: three wild type elutriations, two *cdc25* 'block and release', one *cdc25* experiment done combining elutriation and 'block and release', one *cdc10* experiment again done by combining the two synchronisation methods and one *sep1Δ cdc25* 'block and release'.

3.1 Experimental overview

One aim of this thesis was to identify genes showing a periodic behaviour during the mitotic cell cycle in fission yeast using DNA microarrays. In order to achieve this, the relative abundance of mRNAs was measured as a function of time in cultures synchronised using two different methods, and samples were collected in order to cover two full cell cycles in most experiments. The first method used to obtain a synchronous culture was centrifugal elutriation and the second one was temperature-sensitive mutants. For each timepoint, the gene behaviour in the synchronous population was compared to a reference sample consisting of an asynchronous population of the same strain growing under the same conditions. For each synchronisation method used, at least one experiment has been hybridised with a dye swap. More detailed information on the experimental conditions for all timecourses can be found in Appendix IVa.

When comparing the outcome of several cell cycle microarray studies done in different human cell lines (Cho R.J. *et al.*, 2001; Whitfield M.L. *et al.*, 2002; Iyer V.R. *et al.*, 1999), the overlap between them is poor, and the main reason for this is probably the difference between experimental conditions, especially the different methods used for synchronizing cells. Similarly, the overlap between the two budding yeast studies is relatively small (Cho R.J. *et al.*, 1998; Spellman P.T. *et al.*, 1998). Such differences between studies conducted in similar ways have raised quite a lot of scepticism concerning the validity of the synchronization techniques themselves, claiming that

synchrony cannot be really achieved using traditional whole-culture methods, meaning those experiments where an entire culture of growing cells is used to produce a synchronous cell population (temperature sensitive mutants) (Cooper S. and Shedden K., 2003). From this point of view selective synchronization methods (elutriation) where only a small fraction of a culture is used to obtain a synchronized population are preferable.

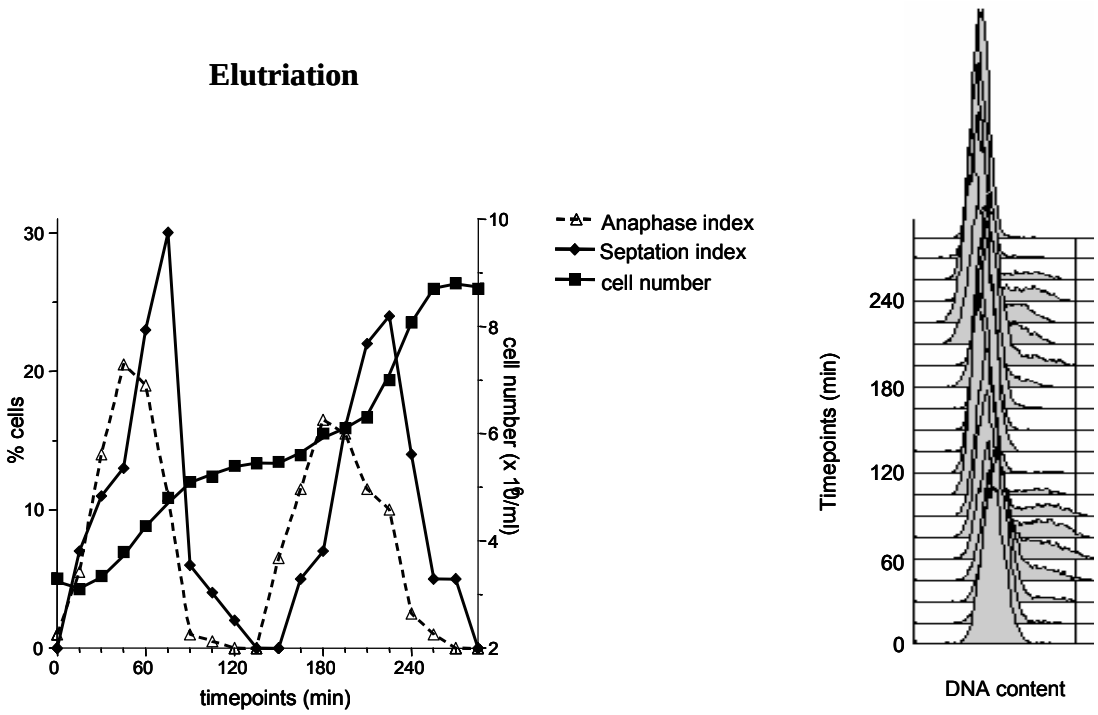
For these reasons, in this thesis both synchronization methods (whole-culture/temperature sensitive mutants and selective/elutriation) were used, and synchrony was assessed by measuring different parameters for each timepoint collected: number of septated cells (named septation index), number of nuclei (DAPI index), DNA content (FACS index) and cell number. The results of these measurements are shown in Figure 3.1A for one elutriation experiment and in Figure 3.1B for one *cdc25* 'block and release' experiment. Additional results for the six remaining timecourses can be found in Appendix V (Fig. V.1-V.3).

In a typical elutriation experiment, nuclear division (Fig. 3.1, DAPI and FACS index) takes place about 1 hour 45 minutes after elutriation (including a 1 hour recovery period for cells). Cell division (Fig. 3.1, septation index) takes place 30 minutes after nuclear division, as also confirmed by the doubling of cell number that can be seen immediately after the peak of septation (Fig. 3.1, cell number). The period between first and second division is normally around 135 minutes in EMM medium at 30°C. In a typical *cdc25* 'block and release' experiment, nuclear division takes place 30 minutes after cells have been shifted back to the permissive temperature, and septation immediately follows at 45 minutes, together with the expected doubling of cell number. The gap between the two divisions is around 120 minutes, in EMM at 25°C. The process of cell division appears to be faster in *cdc25*; because the size of the *cdc25* cell is bigger than the wild type, G2 is therefore shorter because the cell size required to enter mitosis is reached earlier in the cycle.

An apparently higher degree of synchrony was normally achieved in the *cdc25* 'block and release' experiments (60%-70% of septated cells) compared to the elutriations (30% of septated cells). However, much of this difference could be due to extended duration of septation in cdc25 'block and release' experiments; according to the gene expression profiles, there was very little difference in synchrony between the two methods. Despite the fact that temperature sensitive mutants give better synchrony, it is important to consider that a temperature shift likely introduces more artefacts due to the

heat shock involved compared to the moderate mechanical stress experienced during elutriation. For this reason, elutriation is probably the more physiological method for achieving synchrony.

A



B

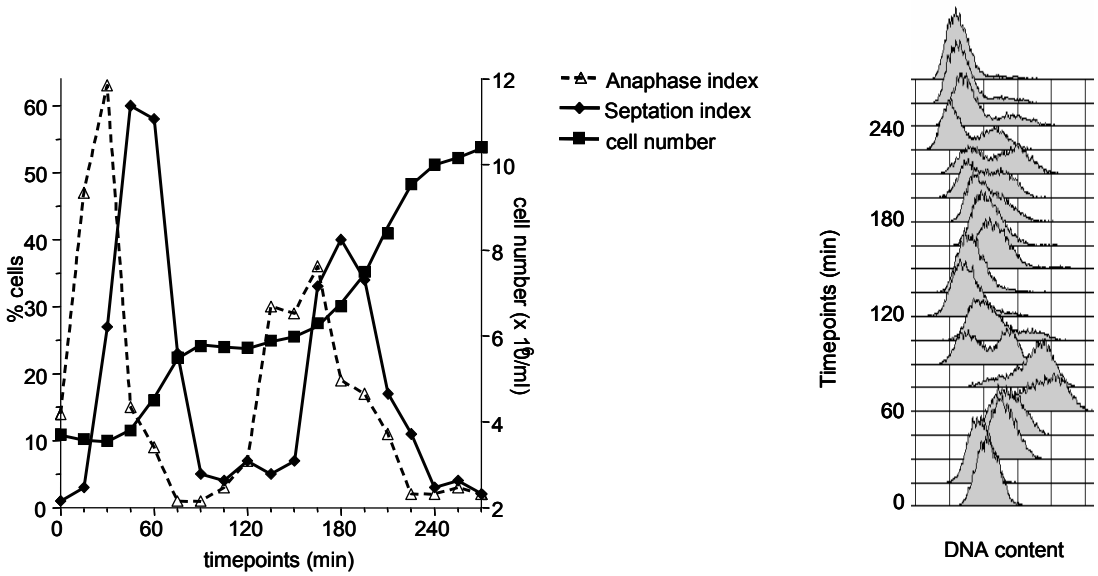


Fig. 3.1 Parameters defining cell cycle synchrony.

Panel A, the left hand side graph shows DAPI index, septation index and cell counting whereas the right hand side graph shows the FACS profile for a typical elutriation experiment. Panel B, shows the same results as panel A for a typical *cdc25* 'block and release' experiment.

3.2 Identification of periodic genes

Genes were ranked according to an average autocorrelation score calculated from the data obtained from three elutriation and three *cdc25-22* 'block and release' experiments (including the *sep1::ura4 cdc25-22* experiment), as described in Materials and Methods. The top 2500 genes were then visually inspected to confirm or reject the periodicity, comparing the profiles with the ones from the *cdc25-22* and *cdc10-129* 'block & release' experiments in which synchronisation was achieved by combining elutriation and temperature shift. In these particular experiments, samples were collected throughout the block and after the release only for one cycle, making it impossible to use these data for the autocorrelation analysis. Around 250 genes were judged to be periodic after this inspection.

These results were compared with the output of the fast Fourier transform and randomisation analysis carried out at the EBI. Following this approach, around 1000 genes had a P value < 0.01 and among these around 800 showed a fold change > 1.5 in all three of the following experiments: two elutriations and one *cdc25* 'block and release'. After visual inspection ~ 400 genes were discarded because their periodicity was not consistent across all experiments or with the duration of the cell cycle. This clearly shows the limitations of an automatic approach for the discovery of periodic genes and underlines again the importance of visually inspecting the dataset to validate the output of any statistical methodology.

From the comparison of the gene lists obtained from these two independent approaches, a total of 407 periodically expressed genes were identified among the 5119 *S. pombe* genes investigated, corresponding to ~ 8% of the total gene number. 136 genes showed a change in expression > 2 fold in the elutriation experiments, and we will refer to them as 'high amplitude' genes. The other 271 showed changes between 1.5 and 2-fold and we will refer to them as 'low amplitude' genes. A complete list of genes can be found in Appendix VI, together with a description of their biological function. The table also

contains information concerning the motifs found in the promoter region of each gene, as will be discussed in chapter 4.

3.3 Clustering of periodic genes

After identification, the 407 genes were classified using a clustering algorithm according to similarities in their expression profiles. Clustering was performed independently using two different methods: ArrayMiner (http://www.optimaldesign.com/Download/ArrayMiner/AM2whitepaper.pdf) and K-means [Materials and Methods – section 2.6.4; (Sherlock G., 2000)]. The ArrayMiner classification resulted in biologically more coherent clusters than the K-means approach in GeneSpring.

Traditional clustering algorithms treat genes as vectors and measure the distance between the two vectors with a distance metric such as Pearson correlation or Euclidean distance (Sherlock G., 2000). This measure reflects the degree of similarity between the two expression profiles and it therefore makes sense to group together those genes that are closest to each other in space. This is an example of *hierarchical clustering*. In the *K-means clustering* method instead, genes are divided from the beginning into an arbitrarily assigned number of clusters. At this stage, the distance between each gene and the centre of the cluster it belongs to are calculated and genes eventually reassigned to a closest cluster. Both methods have their weaknesses. Outliers can represent both meaningful data as well as artefacts created by the experimental procedure. Existing clustering methods will assign them to a cluster, very often altering the structure of the classification. Having to choose from the start the number of clusters is also problematic, especially when little is known about the biological function of each class.

The ArrayMiner non-hierarchical algorithm is based on the assumption that clustering should have the ultimate goal of grouping together genes with distinct biological functions. Therefore the aim is to obtain a distribution of clusters as close as possible to the distribution of the data, modelling the data with a number of Gaussian distributions that best fit the dataset. Outliers are detected and considered as a *uniform distribution* that is competing with the Gaussians. The non-hierarchical approach takes into account the cluster variance and the clusters created remain stable despite the level of detail achieved with the classification. Once a cluster has been identified, increasing the

total number of clusters will result in identifying a subset of smaller clusters within the existing one.

Genes were clustered independently for each of the five experiments into five separate classes. In ambiguous cases, genes have been assigned manually to a cluster when possible or left unclassified if the expression profiles were inconclusive. Ultimately cluster 4 and 5 were merged into one, because separation into two clusters was not consistent across different experiments.

Clusters were assigned to cell cycle phases as follows: cluster 1 corresponding to mitosis, cluster 2 to M/G1 phase (cytokinesis and cell separation), cluster 3 to S phase (DNA replication) and cluster 4 to G2 phase. While this assignment is arbitrary to some degree, especially considering genes at the boundary between two clusters, a good correlation can be found in general between the biological function of the genes and the stage of the cycle when peak of expression occurs. This will be discussed in more detail in the next section. The majority of the 'high amplitude' genes are members of clusters 1, 2 and 3 whereas most cluster 4 genes are only weakly regulated. Fig. 3.2A shows the 407 periodic genes classified into four main clusters for one elutriation experiment and Fig. 3.2B for one *cdc25* 'block and release' experiment. Fig. 3.2C and D show the same clustering for the 136 'high amplitude genes only. Additional figures showing the four clusters in a 3-dimensional view can be found in Appendix V (Fig. V.4-V.5). The purpose of this representation is to show how clusters are separated after projection from the multidimensional expression space into a 2-D space.

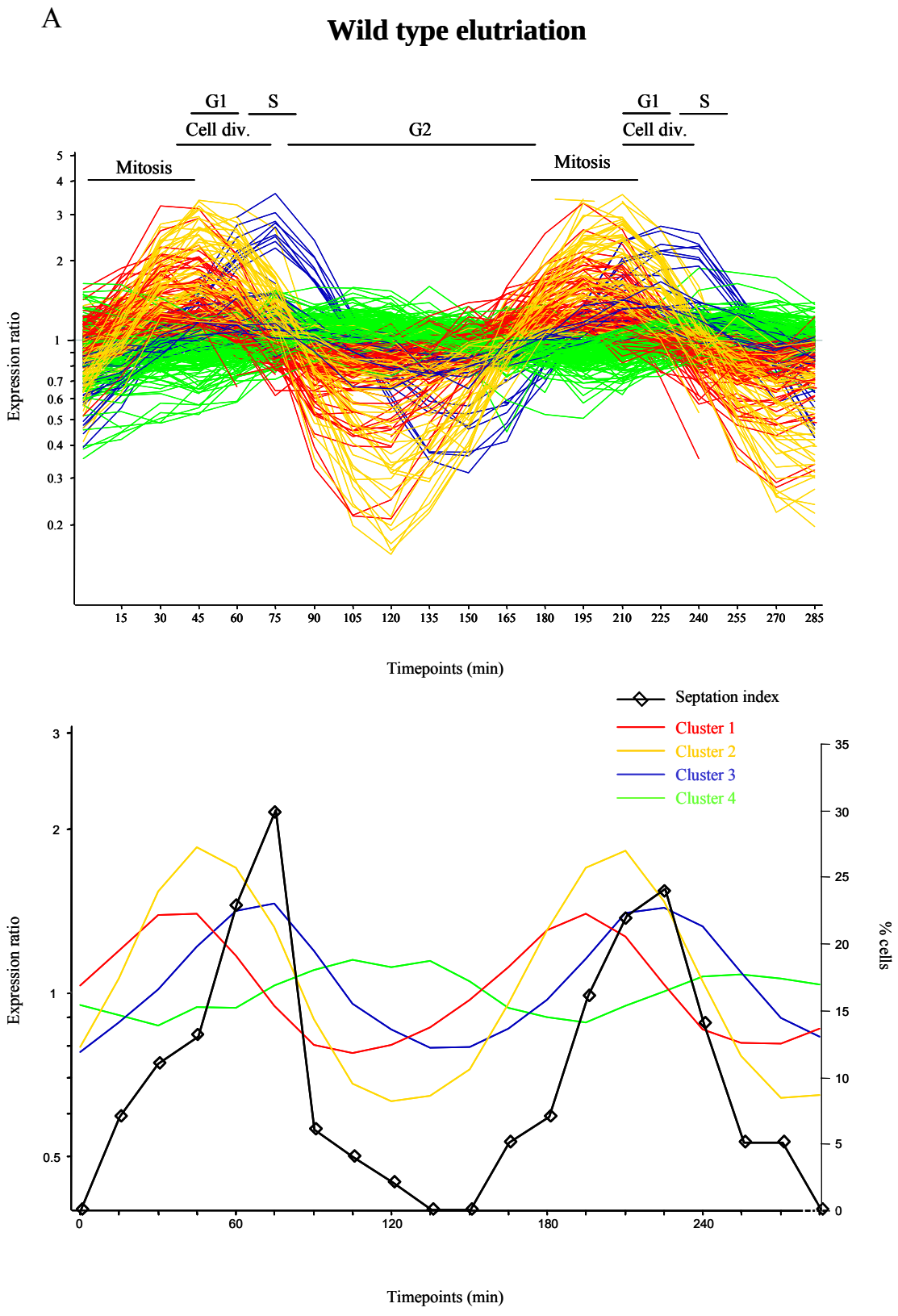


Fig. 3.2 - continues

B *cdc25* 'block and release'

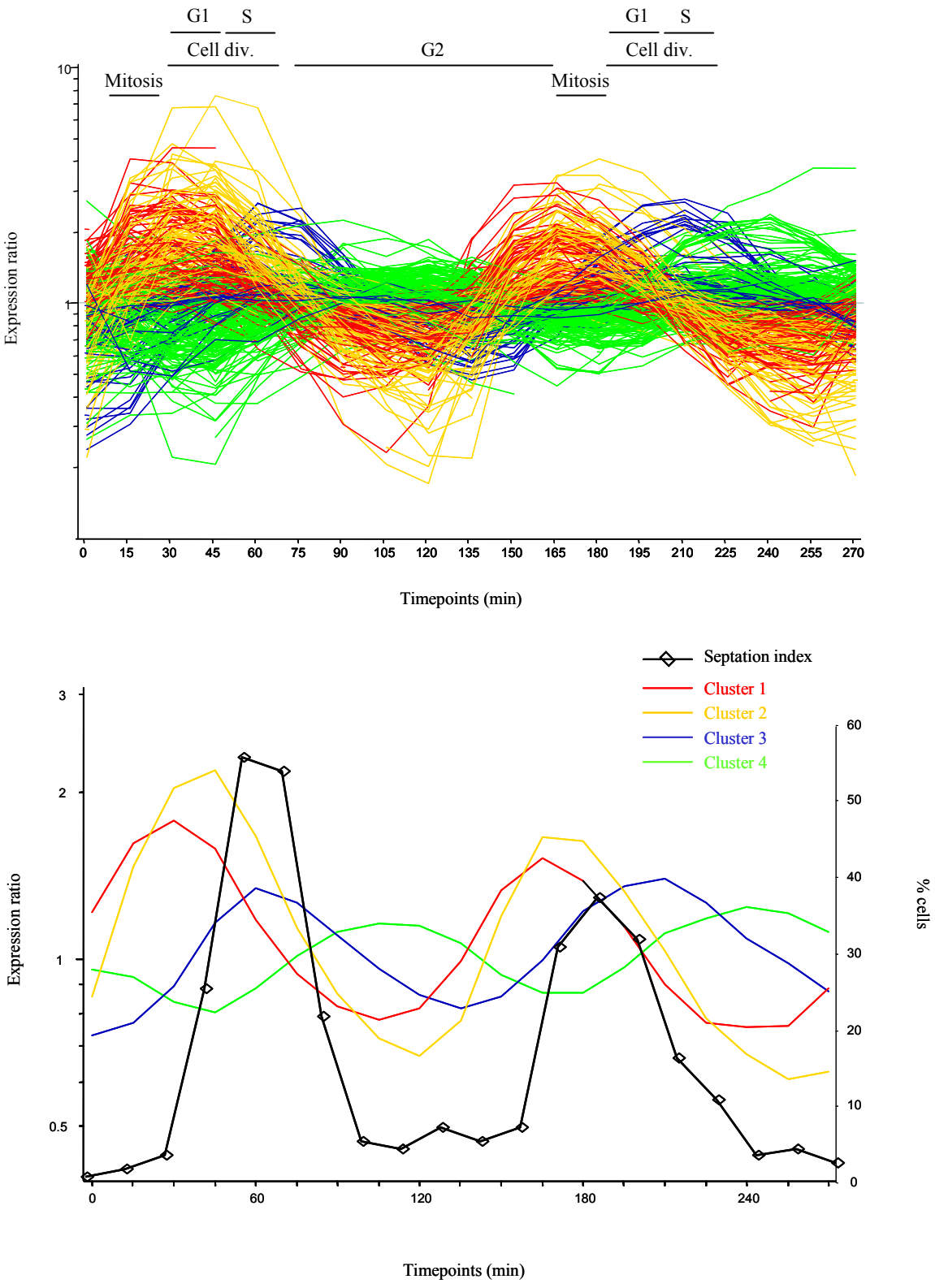


Fig. 3.2 - continues

C

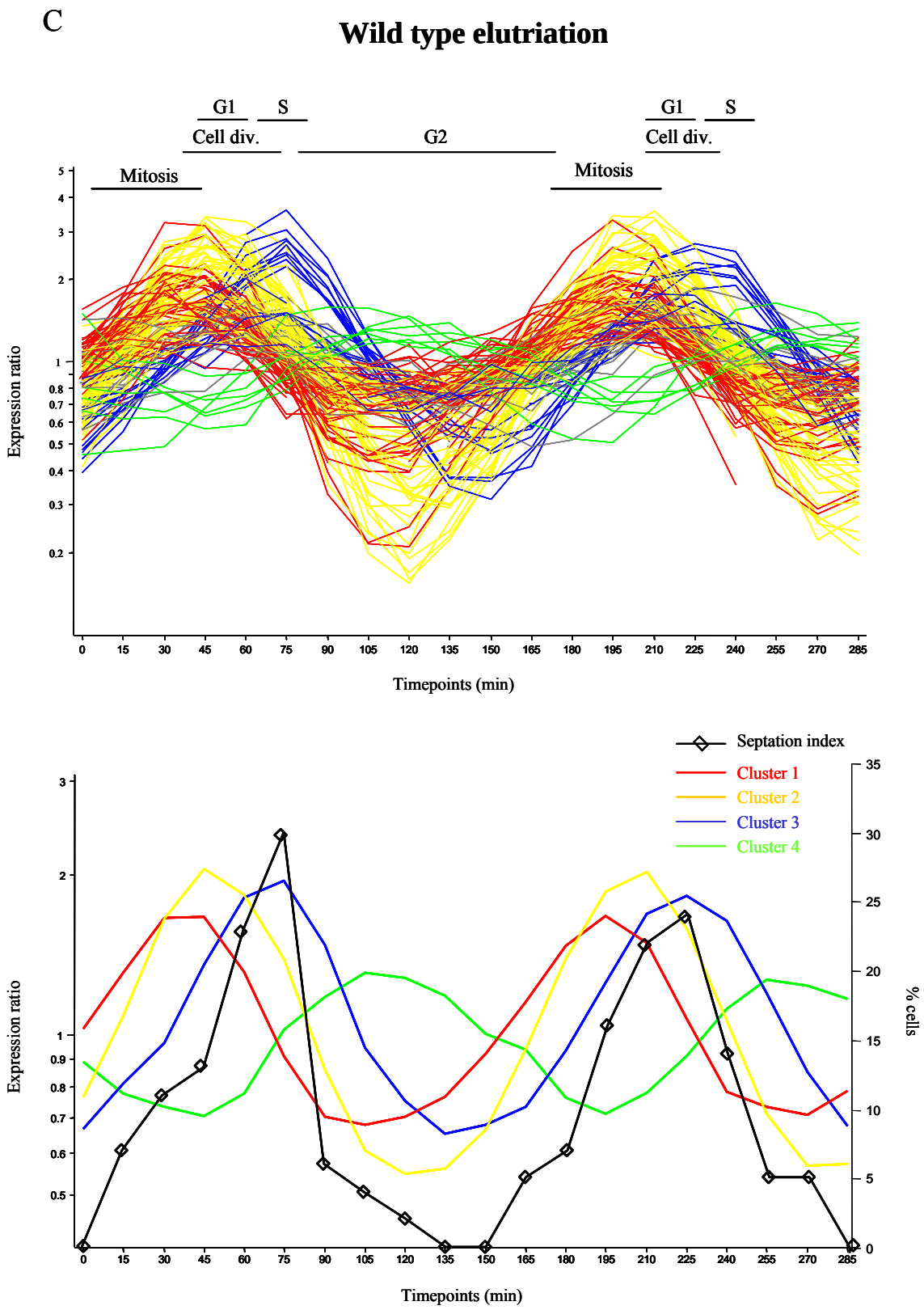


Fig. 3.2 - continues

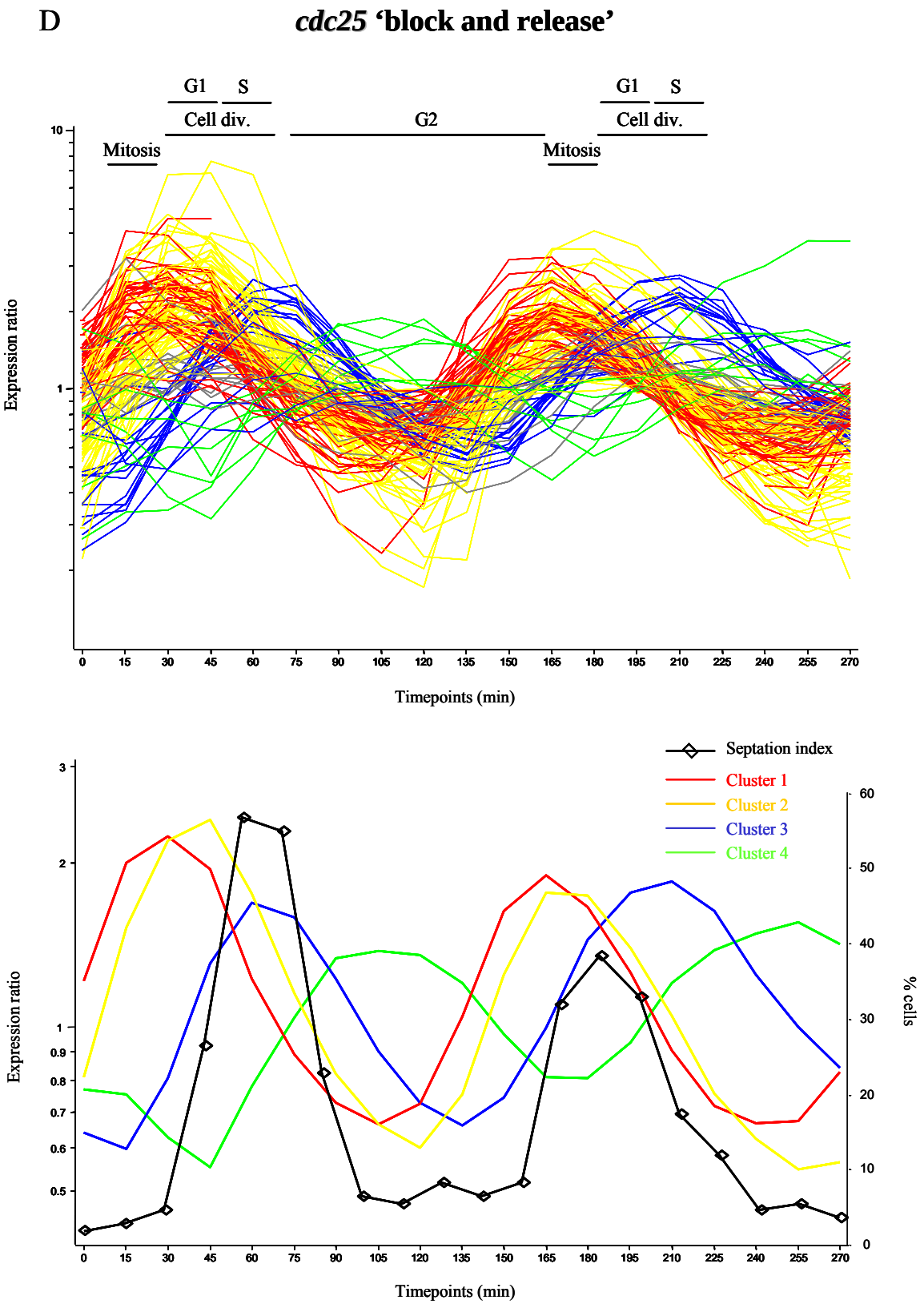


Fig. 3.2 **Clustering of cell cycle regulated genes in *S. pombe*.**

Panel A/C, elutriation experiment (2201). Panel B/D, *cdc25* 'block and release' experiment (2002). In all panels the top graph shows all 407 periodic genes (panel A and B) or the 136 high amplitude genes only (panel C and D) grouped into four main clusters whereas the bottom graph shows the average expression profiles of all genes in each class together with the septation index for the same experiment.

Several fission yeast genes had been previously described as periodic, using traditional molecular genetic approaches. A complete list of the 35 already known cell cycle regulated genes can be found in Table 3.1. 28 of those genes were confirmed to be periodic by our data. Among the seven genes missing, *mid1*, *ppb1*, *res2* and *suc22* showed a weak periodicity, especially in the *cdc25* 'block and release' experiments, but this was insufficient to be included among the periodic genes according to our criteria. It should be mentioned that in two cases (*cdc19* and *ppb1*) results in the original studies concerning the actual periodicity of those genes contradict each other (Anderson M. *et al.*, 2002; Forsburg S.L. and Nurse P., 1994; Plochocka-Zulinska D. *et al.*, 1995). For *res2*, only a marginal periodicity was reported in the original paper (Obara-Ishihara T. and Okayama H., 1994). *Suc22* encodes two separate transcripts: a large one that is weakly expressed and shows a periodic behaviour and a small one, which is much more abundant and continuously expressed through the cycle (Harris P. *et al.*, 1996; Fernandez Sarabia M.J. *et al.*, 1993). In the microarray experiment, the small abundant transcript probably hybridises more efficiently than the large one, explaining why the periodic behaviour of the gene could not be detected. *Cdc19*, *cmk1* and *rrg1* were not periodic under any conditions in study.

All the genes previously reported as periodic are members of the first three clusters, showing a good correlation between the peak of expression in our experiments and the one reported in the literature (see table 3.1 for reference to original publications). None of the genes included in the fourth cluster had been shown before to be periodic.

Table 3.1 Genes previously reported as cell cycle regulated in *S. pombe*

Gene Name	Expression peak	Cluster	References
cdc15	M	1	(Fankhauser C. *et al.*, 1995)
cdc18	G1/S	2	(Kelly T.J. *et al.*, 1993)
cdc19?	M/G1?	ND	(Anderson M. *et al.*, 2002; Forsburg S.L. and Nurse P., 1994)
cdc22	G1/S	2	(Gordon C. and Fantes P., 1986)
cdc25	M	1	(Moreno S. *et al.*, 1990)
cdt1	G1	2	(Hofmann J.F. and Beach D., 1994)
cdt2	G1	2	(Hofmann J.F. and Beach D., 1994)
cig2	G1/S	2	(Connolly T. and Beach D., 1994; Obara-Ishihara T. and Okayama H., 1994)
cmk1	G1/S	ND	(Rasmussen C.D., 2000)
cnp1	G1/S	2	(Takahashi K. *et al.*, 2000)
dfp1	G1/S	2	(Brown G.W. and Kelly T.J., 1999)
fin1	as *cdt1*	2	(Krien M.J. *et al.*, 2002)
eng1	G1/S	2	(Martin-Cuadrado A.B. *et al.*, 2003)
hht1	S	3	(Takahashi K. *et al.*, 2000; Matsumoto S. and Yanagida M., 1985)
hta1	S	3	(Aves S.J. *et al.*, 1985)
htb1	S	3	(Matsumoto S. *et al.*, 1987)
mid1	M/G1	ND	(Anderson M. *et al.*, 2002)
mid2	M	2	(Tasto J.J. *et al.*, 2003)
mik1	G1/S	2	(Christensen P.U. *et al.*, 2000; Ng S.S. *et al.*, 2001; Baber-Furnari B.A. *et al.*, 2000)
mrc1	as *cdc18*	2	(Tanaka K. and Russell P., 2001)
pht1	S	3	(Carr A.M. *et al.*, 1994; Durkacz B.W. *et al.*, 1986)
plo1	M/G1	1	(Anderson M. *et al.*, 2002)
ppb1?	S or M/G1?	ND	(Anderson M. *et al.*, 2002); Plochocka-Zulinska D. *et al.*, 1995)
rad21	G1/S	2	(Birkenbihl R.P. and Subramani S., 1995)
res2?	G1/S?	ND	(Obara-Ishihara T. and Okayama H., 1994)
rhp51	before *cdc22*	1	(Jang Y.K. *et al.*, 1996)
rph1	G1/S	2	(Tanaka H. *et al.*, 2002)
rrg1	G2/M	ND	(Kim M.J. *et al.*, 2002)
rum1	end G2	1	(Benito J. *et al.*, 1998)
sid2	M/G1	1	(Anderson M. *et al.*, 2002)
slp1	M	1	(Yamada H.Y. *et al.*, 2000)
spo12	as *cdc15*	1	(Samuel J.M. *et al.*, 2000)
ssb1	as *cdc22*	2	(Parker A.E. *et al.*, 1997)
ste9	as *cdc18*	2	(Tournier S. and Millar J.B., 2000)
(*suc22*)	G1/S (transcript dep.)	ND	(Harris P. *et al.*, 1996; Fernandez Sarabia M.J. *et al.*, 1993)

3.4 Biological function of genes in four clusters

This section describes each of the four waves of transcription focusing on the biological function of its members. For each cluster, genes are grouped in tables (see below) according to the biological process when their function is performed. It should here be

mentioned that this grouping based on function is not rigorous and was here adopted only to facilitate the presentation of the results. Fig. 3.3 shows the expression profiles of all genes in each cluster.

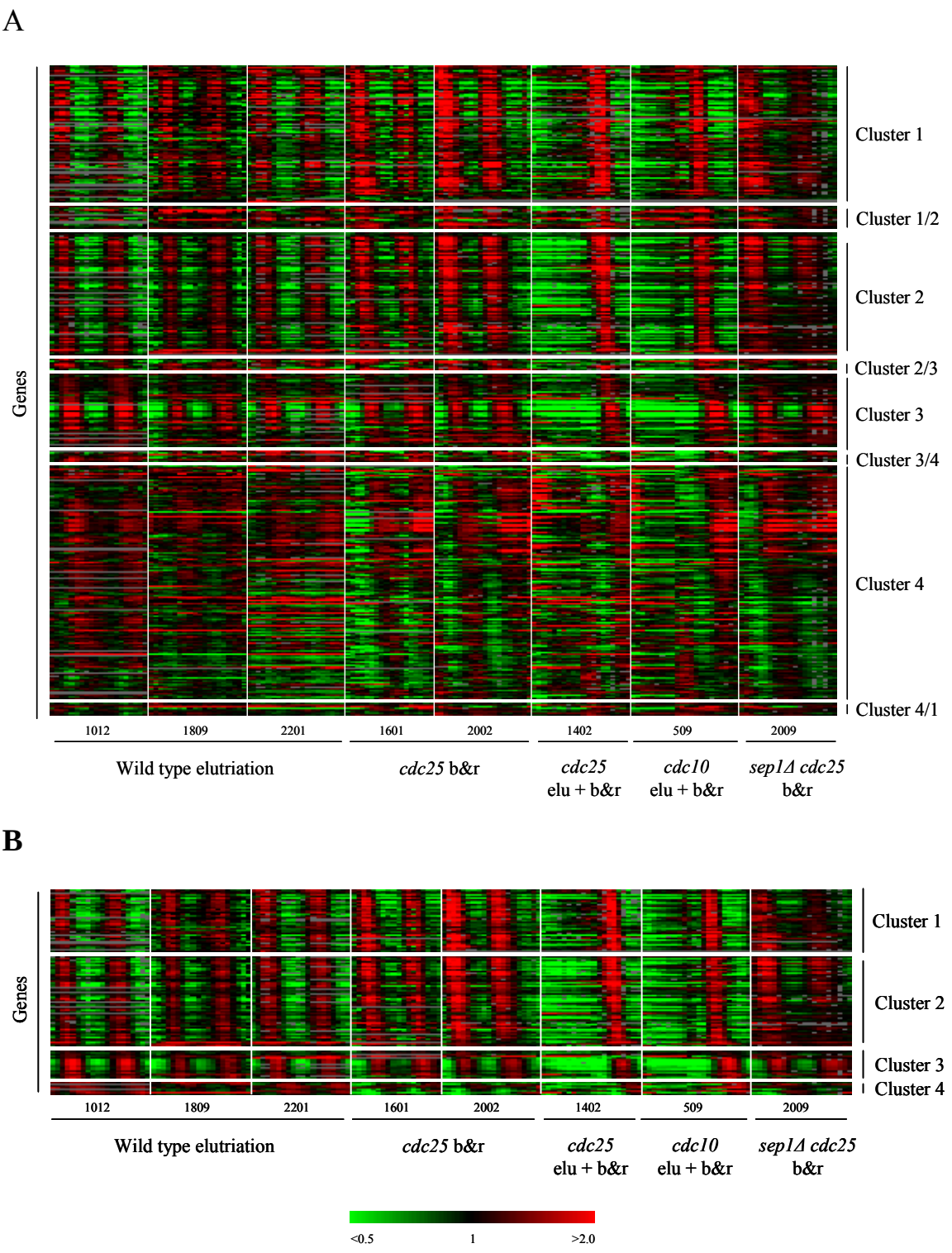


Fig. 3.3 **Cell cycle regulated genes in *S. pombe* and their classification.**

Rows represent the profiles of the 407 periodic genes (panel A) and of the 136 'high amplitude' genes only (panel B) ordered by the time of their peak of expression. Columns represent synchronised experimental samples (8 timecourses of 18-22 timepoints collected at 15 min intervals – 161 timepoints in total). Red: induced expression; green: repressed expression; grey: no data. Classification of genes into 4 major clusters is also shown. Smaller clusters of unclassified genes are also shown (clusters named 1/2, 2/3, 3/4 and 4/1).

Cluster 1

This cluster includes 87 genes, 40 of which are 'high amplitude' genes, showing a peak of expression coincident with mitosis. It is possible to identify subgroups of genes involved in the same biological process.

Many of the events happening during mitosis involve dramatic changes to the cytoskeleton due to rearrangement of the microtubules to form the mitotic spindle as well as nuclear changes like chromosome condensation, segregation and separation (Su S.S.Y. and Yanagida M., 1997). This subset of genes includes: *klp5* and *klp6* (kinesin motors that influence microtubule dynamics), *myo3* (encoding a myosin, which interacts with actin), several genes involved in sister chromatid cohesion (*pds5* – for the establishment of cohesion and *psc3* – encoding a mitotic cohesion subunit) and chromosome segregation (*dis1* which acts in collaboration with klp5p and klp6p) as well as an essential component of the spindle pole body (*sad1*).

Other genes encode proteins involved in cytokinesis and cell separation. In fission yeast the final stage of cell division is characterised by the formation of a septum which will then be digested to allow cell separation (Su S.S.Y. and Yanagida M., 1997). It is not surprising to find expressed at this stage *plo1* (encoding a mitotic regulatory kinase and an inducer of septum formation and cytokinesis), *cdc15* (regulator of septum formation), *imp2* (cdc15p-like protein), *sid2* (protein kinase responsible for triggering septation) and *mob1* (whose gene product binds to sid2p). Many genes involved in metabolic pathways and cell wall biosynthesis are also highly induced at this stage, possibly reflecting the *de novo* synthesis of cell wall and membrane concomitant to cytokinesis.

Several genes involved in cell cycle regulation can also be found in this cluster: *ace2* (transcriptional regulator), *ark1* (encoding the Aurora kinase required for chromosome condensation) as well as genes encoding key regulatory molecules involved

in the regulation of the cyclin-dependent kinase (CDK) cdc2p such as cdc13p (cyclin partner of cdc2p), cdc25p (cdc2p-cdc13p activator), rum1p (cdc2p inhibitor), crk1p and csk1p (cdc2 activators) and cdr1p (indirect activator of cdc2p). *Cdc2* itself is member of cluster 4, as described later.

In addition the gene SPBC16G5.15c, whose function is still uncharacterised, encodes a protein containing a forkhead binding domain typical of a highly conserved class of transcription factors. Two smaller groups, one of meiotic genes [including *mus81* coding for a nuclease and *meu16*, a non coding RNA potentially involved in meiosis regulation (Watanabe T. *et al.*, 2001)] and a second one of genes involved in DNA repair (including *slp1*, component of the spindle pole body checkpoint) are also included. This last group also includes *cdc20*, encoding DNA polymerase epsilon, which had been previously reported as non periodic (Sugino A. *et al.*, 1998).

18 genes are coding for proteins whose functions are still poorly characterised (e. g. domains identified in the protein or function suggested based on similarity with other known proteins).

Table 3.2 Selected cluster 1 members and their biological function

Biological names	Systematic names	Gene description
Cytokinesis and cell separation		
cdc15; rng1	SPAC20G8.05c	Protein involved in cytokinesis
etd1	SPAC1006.08	Protein required for cytokinesis
imp2	SPAC13F4.08c; SPBC11C11.02	Protein required for medial ring disassembly after cytokinesis
mac1	SPAC13G7.04c	Transmembrane protein involved in cell separation
mob1	SPBC428.13c	Protein involved in regulation of cytokinesis
myo3; myp2	SPAC4A8.05c	Myosin-3 isoform, heavy chain (Type II myosin)
plo1	SPAC23C11.16	Polo kinase involved in regulation of mitosis and cytokinesis
rho4	SPAC16A10.04	Rho protein involved in regulation of cytoskeleton, cytokinesis, and cell wall integrity
sid2; pld5	SPAC24B11.11c	Protein kinase involved in regulation of cytokinesis
spn2	SPAC821.06	Septin homolog, involved in cell separation
spn7; mde8	SPBC21.08c; SPBC19F8.01c	Septin homolog, involved in cell separation
Cell cycle control genes		
ace2	SPAC6G10.12c	Zinc finger transcription factor
apc15; apc16	SPBC83.04	Component of APC/cyclosome complex
ark1; sex1	SPCC330.16; SPCC320.13c	Aurora kinase involved in regulation of mitosis
cdc13	SPAC19G10.09C; SPBC582.03	Cyclin that promotes entry into mitosis from G2 phase, forms complex with Cdc2
cdc25; sal2	SPAC24H6.05	Tyrosine phosphatase that activates Cdc2p kinase,

		involved in G2/M transition and DNA damage checkpoints
cdr1; nim1	SPAC644.06c	Protein kinase involved in regulation of mitosis
crk1; mcs6; mop1	SPBC19F8.07	Cyclin-dependent kinase activating kinase (CAK) involved in activating Cdc2p kinase, putative transcription initiation factor TFIIH subunit
csk1	SPAC1D4.06c	Cyclin-dependent kinase activating kinase (CAK) involved in activating Cdc2p (activity partially redundant with Mcs6p-Mcs2p complex)
rum1	SPBC32F12.09	Inhibitor of the Cdc2p cyclin-dependent kinase complex
wis3; spo12	SPAC3F10.15c	Protein likely to play role in regulating cell cycle progression, possibly at G2 to M phase transition
	SPBC16G5.15c	Fork head protein type transcription factor
DNA repair and checkpoints		
cdc20; pol2	SPBC25H2.13c	DNA polymerase epsilon catalytic subunit
msh6	SPCC285.16c	Protein involved in mismatch repair (mutS family)
rhp51; rad51	SPAC644.14C	Required for DNA repair and meiotic recombination
slp1	SPAC821.08c	WD-domain protein of the spindle defect checkpoint and APC activator
Cell wall biosynthesis and maintenance		
chs2	SPBC1734.17; SPBC1709.01	Member of chitin synthase family, involved in cell wall maintenance
sso1; psy1	SPCC825.03c	Syntaxin-like component of the plasma membrane docking/fusion complex
	SPAC23H4.19; SPAC1705.03c	Putative cell wall biogenesis protein
	SPBC1198.07c	Putative glycosylphosphatidylinositol (GPI)-anchored protein involved in cell wall biosynthesis
	SPAC11E3.13c	Member of glycolipid anchored surface protein (GAS1) family, possible involvement in cell wall maintenance
	SPAC19B12.02c	Protein with high similarity to 1,3-beta-glucanosyltransferase, member of glycolipid anchored surface protein (GAS1) family
Chromosome segregation and chromatidin cohesion		
dis1	SPCC736.14	Microtubule-associated protein required for chromosome segregation (functions with Klp5p and Klp6p in kinetochore-spindle attachment)
klp6	SPBC1685.15c; SPBC649.01c	Kinesin motor protein; KIP3 subfamily
klp5	SPBC2F12.13	Kinesin motor protein; KIP3 subfamily
pds5	SPAC110.02	Protein required for maintenance of sister chromatid cohesion
psc3	SPAC17H9.20; SPAC607.01	Cohesin complex component, required for sister chromatid cohesion and normal mitosis
sad1; sta1	SPBC16H5.01c; SPBC12D12.01	Spindle pole body associated protein
top1	SPBC1703.14c	DNA topoisomerase I, involved in chromatin organisation
Meiosis		
mde6	SPAC15A10.10	Protein likely to play a role in meiosis or sporulation, requires Mei4p for transcriptional activation
meu16		Non-coding RNA
mus81	SPCC4G3.05c	Holliday junction resolvase subunit
ste7	SPAC23E2.03c	Protein required for mating and meiosis

Metabolic genes		
cid13	SPAC821.04c	Cytoplasmic poly(A) polymerase involved in regulation of ribonucleotide reductase (suc22) mRNA, TRF family of nucleotidyltransferases
	SPBC646.06c	Member of glycosyl hydrolase family 71, putative glucanase
	SPAC589.09	Protein containing a CRAL-TRIO domain, putative phosphatidylinositol metabolism
	SPCC576.02	Member of aspartate and glutamate racemases family
	SPAC30D11.01c; SPAC56F8.01	Member of glycosyl hydrolases family 31, involved in carbohydrate metabolism
	SPAC13G6.03	Member of type I phosphodiesterase or nucleotide pyrophosphatase family
	SPAC13C5.05c	Member of phosphoglucomutase or phosphomannomutase C-terminal domain containing family
	SPBC27B12.06	Protein with possible role in glycosylphosphatidylinositol biosynthesis
	SPBPB2B2.09c	Member of the ketopantoate reductase PanE or ApbA family, involved in thiamine biosynthesis
	SPCC757.12	Protein containing an alpha amylase N-terminal catalytic domain

Cluster 2

75 genes are members of this cluster, 58 of which are regulated with 'high amplitude' and they are induced around anaphase and cytokinesis, which in *S. pombe* corresponds to G1 phase. Many of them are involved in DNA replication initiation like *cdc18*, *cdt1* and *dfp1* (essential factors interacting with the pre-replicative complex), *cdc22* (coding for the large subunit of ribonucleotide reductase), *cdt2* (potentially involved in the formation of protein complexes required for DNA replication (Yoshida S.H. *et al.*, 2003) and *ssb1* (encoding for Replication Protein A, involved in DNA replication, recombination and repair). Two polymerase subunits are also expressed (*pol1* – alpha subunit and *cdm1* – delta subunit). *Pol1* was previously reported as not cell cycle regulated (Park H. *et al.*, 1993).

In *S. pombe*, septation does not need to be completed for DNA replication to start, and the two processes are partially overlapping due to a very short G1 phase. Therefore, this cluster also includes genes required for cell separation: *eng1* and *exg1* (glucanases involved in septum digestion), *cdc4* and *klp8* (involved in actin and microtubule rearrangements). Similarly, chromosome segregation genes like *fin1* (kinase promoting chromatin condensation), *rad21* (mitotic cohesion subunit), *ams2* (chromatin binding

protein) and *cnp1* (CeNtromere Protein-A-like, histone H3 variant) as well as sister chromatid cohesion genes (*cut2*, *eso1* and *psm3*) are also present.

Some other well characterised genes involved in cell cycle progression are also expressed at this stage: *cdc10* and *rep2* (encoding a component and a regulator of the MBF transcription factor), *ste9* (Anaphase Promoting Complex regulator), *cig2* (cdc2p cyclin partner) and *mik1* (cdc2p inhibitor), and *mrc1* (DNA damage checkpoint protein).

Another non-coding RNA, *meu19*, with a putative meiotic regulatory role, is found in the cluster. Among the members of this cluster with an unknown function, SPBC21B10.13c is worth mentioning, which encodes a protein containing a homeobox domain, a domain frequently found in transcription factors (Gehring W.J. *et al.*, 1994). Its budding yeast homologue, *YOX1*, encodes a homeodomain protein which acts as a transcriptional repressor, restricting expression of a subset of genes to M/G1 (Pramila T. *et al.*, 2002). *YOX1* itself is periodic in *S. cerevisiae*. SPBC21B10.13c might also have a regulatory role in meiosis (Mata J. *et al.*, 2002). 32 genes code for proteins with unknown function.

Table 3.3 Selected cluster 2 members and their biological function

Biological names	Systematic names	Gene description
Cell cycle control genes		
cdc10	SPBC336.12c	Component of MBF transcriptional activation complex involved in control of START
cig2; cyc17	SPAPB2B4.03	Major G1/S-phase cyclin, promotes onset of S phase
mik1	SPBC660.14	Protein kinase that inhibits Cdc2p kinase
	SPBC21B10.13c; SPAC21B10.13c	Homeobox domain (homeodomain) protein, putative transcription factor
rep2	SPBC2F12.11c	Zinc finger transcriptional activator, MBF transcriptional complex
ste9; srw1	SPAC144.13c	Protein required for mating and sporulation, may regulate anaphase promoting complex
DNA replication		
cdc18	SPBC14C8.07c	Protein that couples cell cycle signals to DNA replication machinery and induces replication
cdc22	SPAC1F7.05	Ribonucleoside-diphosphate reductase large chain, likely required for initiation of DNA replication
cdm1	SPBC12D12.02c	DNA polymerase delta subunit
cdt1	SPBC428.18	Protein that coordinates completion of S phase with onset of mitosis
cdt2	SPAC17H9.19c	Protein required for DNA replication
dfp1; him1; rad35	SPCC550.13	Regulatory subunit of the Hsk1p-Dfp1p kinase complex involved in S phase initiation
mrc1; huc1	SPAC694.06c	Protein required for DNA replication checkpoint
pol1; swi7	SPAC3H5.06c	DNA polymerase alpha catalytic subunit

rph1; pfh1; pif1	SPBC887.14c	ATP-dependent DNA helicase involved in telomere maintenance, DNA replication, and DNA repair
ssb1; rad11	SPBC660.13c	Single-stranded DNA-binding protein subunit, required for DNA replication
DNA repair and checkpoint		
mrc1; huc1	SPAC694.06c	Protein required for DNA replication checkpoint
Cell wall biosynthesis		
bgs4; cwg1	SPCC1840.02c	Putative 1,3-beta-glucan synthase component, cell wall synthesis
Meiosis		
meu19		Non-coding RNA
Cytokinesis and cell separation		
cdc4	SPAP8A3.08	EF-hand component of actomyosin contractile ring, required for cytokinesis
chs5	SPAC6G9.12	Protein with fibronectin domain involved in cell surface binding, and BRCT domain found in checkpoint proteins, similar to chitin synthase
eng1	SPAC821.09	Endo-beta-1,3-glucanase required for cell separation
exg1	SPBC1105.05	Putative exo-beta-1,3-glucanase
klp8	SPAC144.14	Protein containing a kinesin motor domain
mid2	SPAPYUG7.03c	Protein required for septin function and stability during cytokinesis
par2; pbp2	SPAC6F12.12	Protein phosphatase PP2A, B' regulatory subunit, required for cytokinesis, morphogenesis, and stress tolerance
pob1	SPBC1289.04c	Protein required for cell polarity and cell separation
	SPBC1289.01c; SPBC1539.11c	Unknown function, putative involvement in chitin biosynthesis
	SPBC3E7.12c	Unknown function, possible role in regulation of chitin synthase
	SPCC1322.10	Unknown function, similar to cell-surface proteins and proteoglycans
	SPAC14C4.09	Unknown function, putative glucanase
Chromosome segregation and chromatid cohesion		
ams2	SPCC4F11.01; SPCC290.04	Protein that binds binds chromatin at centromere and is involved in chromosome segregation
cnp1; sim2	SPBC1105.17	CENP-A-like protein, histone H3 variant specific to inner centromeres and required for chromosome segregation
cut2	SPBC1815.02c; SPBC14C8.01c	Securin; required for sister chromatid separation
eso1; eco1	SPBC16A3.11	DNA polymerase eta, involved in sister chromatid cohesion
fin1	SPAC19E9.02	NimA family kinase; regulates spindle formation and recruitment of Plo1p to SPB, promotes chromatin condensation
psm3; smc3	SPAC10F6.09c	Cohesin complex subunit, involved in sister chromatid cohesion and progression through mitosis
rad21	SPCC338.17c	Cohesin complex subunit, double-strand-break repair protein
Metabolic genes		
	SPCC1322.04	Putative UTP-glucose-1-phosphate uridylyltransferase
	SPBC32F12.10	Protein with phosphoglucomutase or phosphomannomutase C-terminal domain

	SPBPB2B2.13	Protein similar to galactokinase

Cluster 3

This cluster contains 46 genes expressed during DNA replication, which overlaps with septation and cell separation in rapidly growing fission yeast cells. 18 genes are 'high amplitude'.

All histone genes peak during S phase as expected, and represent the tightest subcluster within cluster 3. Other interesting genes are: *rad25*, whose product is responsible for sequestering cdc25p to the cytoplasm causing a G2/M arrest in response to DNA damage and *pas1*, encoding a cyclin partner for the Pef1p kinase complex possibly involved in the regulation of MBF transcription factor (Tanaka K. and Okayama H., 2000).

Proteins with unknown function are encoded by 16 members of this cluster; among those are several proteins containing some well known domains such as zinc fingers, HMG-box and GTPase activation. The non coding RNA *prl36* is also member of this cluster.

Table 3.4 Selected cluster 3 members and their biological function

Biological names	Systematic names	Gene description
Histones		
hht1	SPAC1834.04	Histone H3.1
hht2	SPBC8D2.04	Histone H3.2
hht3; clo5	SPBC1105.11c	Histone H3.3
hhf1; ams1	SPAC1834.03c	Histone H4.1
hhf2; ams3	SPBC8D2.03c	Protein similar to histone H4.1, contains a core histone domain
hhf3; ams4	SPBC1105.12	Histone 4.3
hta1	SPCC622.08c	Histone H2A-alpha
hta2	SPAC19G12.06c	Histone H2A-beta
htb1	SPCC622.09	Histone H2B-alpha
pht1	SPBC11B10.10c	Histone H2A variant
Cell cycle checkpoints		
rad25	SPAC17A2.13c	14-3-3- protein involved in DNA damage checkpoint control
Meiosis		
pas1	SPAC57A10.01; SPAC19E9.03	Cyclin involved in regulation of mating, interacts with Pef1p and Cdc2p kinases
Metabolic genes and others		

	SPBC1348.10c; SPAC1348.10c	Member of lysophospholipase catalytic domain family, putative lysophospholipase precursor
	SPBC21B10.09; SPAC21B10.09	Protein similar to acetyl-CoA transporter
	SPAC977.09c	Member of lysophospholipase catalytic domain containing family, similar to phospholipase B, which deacylates phosphatidylinositol
	SPCC306.08c	Malate dehydrogenase, mitochondrial precursor
	SPCC1906.01	Mannose-1-phosphate guanyltransferase
prl36		Non coding RNA

Cluster 4

This cluster contains 147 genes, peaking at different times during G2, only 7 of which belong to the 'high amplitude' subgroup. It is the largest and most heterogeneous cluster of the four, and only 20% of its members have a characterised function. It contains 68 genes encoding proteins with unknown functions or with only an identified domain.

Cdc2, encoding the kinase responsible for driving cell cycle progression and *cig1*, one of cdc2p cyclin partners, both peak in G2, together with *spd1*, an S-phase inhibitor through association with cdc2p. Many genes encoding membrane transporters or proteins involved in ribosome biogenesis and RNA processing are also transcribed at this stage, reflecting the actively growing state of the cell in G2 phase.

Another subcluster is represented by several tf2-type transposon elements (9 genes). It is also interesting to notice the presence of several stress genes, including the transcription factor *pcr1*, involved in regulating meiosis and stress response. At least 50 genes belonging to this cluster, most of them still uncharacterised, are known to be induced in response to stress (Chen D. *et al.*, 2003).

Table 3.5 Selected cluster 4 members and their biological function

Biological names	**Systematic names**	**Gene description**
Cell cycle control		
cdc2; swo2	SPBC11B10.09	Cyclin-dependent kinase, regulates cell cycle transitions G1/S and G2/M
cig1	SPCC645.01; SPCC4E9.02	B-type cyclin involved in G1 to S phase transition
spd1	SPAC29B12.03	Negative regulator of S phase
Transporters		
mae1	SPAPB8E5.03	Malate transporter
sst1	SPAC521.04c	Member of sodium or calcium exchanger protein family of

		membrane transporters
	SPAP7G5.06	Protein similar to amino acid permease, a proton symport transporter for all naturally-occurring L-amino acids
	SPAC1039.01	Member of amino acid permease family of membrane transporters
	SPCC126.09	Member of ZIP zinc transporter family, possible metal transporter and vacuolar membrane protein
	SPAC139.02c	Probable mitochondrial oxaloacetate transporter
	SPBC16D10.06	Member of ZIP zinc transporter family
	SPBC29A3.01	Member of P-type ATPase, similar to copper-transporting ATPase
	SPAC212.10	Pseudogene; malic acid transport protein; truncated C at terminal
	SPCC548.06c	Protein similar to putative H+-glucose symporter involved in glucose transport
	SPAC9.10	Member of amino acid permease family of membrane transporters
	SPAC869.02c	Member of globin family of oxygen transporters, similar to flavohemoglobin that protects from stress
	SPAPB24D3.09c	Protein with ABC transporter domains, similar to brefeldin A resistance protein involved in multidrug resistance
	SPBC1271.10c	Protein similar to membrane transporter
	SPCC794.03	Member of amino acid permease family of membrane transporters
	SPCC31H12.01; SPCC1183.11	Member of mechanosensitive ion channel family
	SPCC794.11c	Protein with actin binding domain, possible role in formation of clathrin coats at the Golgi and endosomes
Transposons		
TF2-1		Retrotransposable element; tf2-type transposon
TF2-10		Retrotransposable element; tf2-type transposon
TF2-2		Retrotransposable element; tf2-type transposon
TF2-3; TF2-4		Retrotransposable element; tf2-type transposon
TF2-5		Retrotransposable element; tf2-type transposon
TF2-6		Retrotransposable element; tf2-type transposon
TF2-7		Retrotransposable element; tf2-type transposon
TF2-8		Retrotransposable element; tf2-type transposon
TF2-9		Retrotransposable element; tf2-type transposon
Stress genes		
trx1; trx2	SPAC7D4.07c	Putative thioredoxin involved inresponse to heavy metals
pcr1; mts2	SPAC21E11.03c	Transcription factor that plays roles in mating, meiosis and stress response
rds1	SPAC343.12	Stress response protein
ssp1	SPCC297.03	Protein kinase that mediates rapid osmotic stress response at cell surface
uvi15	SPBC649.04	Protein essential for stationary phase survival, induced by stress
Metabolic genes		
arg5	SPBC56F2.09c	Protein similar to amidotransferase small subunit of carbamoylphosphate synthetase
dak1; dak2	SPAC977.16c	Dihydroxyacetone kinase, isoenzyme II
gpd2	SPAC23D3.04c	Glycerol-3-phosphate dehydrogenase
gmh2	SPAC5H10.13c	Protein similar to alpha-1,2-galactosyltransferase
gps2	SPBC365.14c	Putative UDP-glucose 4-epimerase involved in UDP-galactose synthesis and protein glycosylation

	SPBC119.10	Asparagine synthetase
	SPCC1827.06c	Aspartate semialdehyde dehydrogenase
	SPAC5H10.06c	Protein similar to alcohol dehydrogenase IV, which is involved in carbohydrate metabolism
	SPBC8E4.03	Protein with arginase family domain, similar to agmatine ureohydrolase
	SPAC19G12.09	Protein with aldo-keto reductase family domain, similar to aldehyde reductase
RNA processing and ribosome biogenesis		
	SPBC13G1.09	Member of bystin family, possible role in 35S pre-rRNA processing into 18S rRNA
	SPBC17D1.06; SPCC17D1.06	Member of the DEAD or DEAH box ATP-dependent RNA helicase
	SPAC2C4.18; SPAC25G10.01	Protein with RNA recognition motif, possible splicing factor that activates pre-mRNA splicing
csx1	SPAC17A2.09c	Protein containing three RNA recognition motifs, similar to U1 snRNA-associated protein
	SPBP8B7.15c	Protein similar to Polyadenylation Factor I complex component required for mRNA cleavage and polyadenylation
	SPCP1E11.08	Protein similar to nuclear protein involved in ribosome biogenesis
	SPAC1486.09	Protein similar to protein that functions in 20S proteasome maturation and 26S proteasome assembly
	SPAC823.03; SPAC1E11.03	Protein with kinase domain, similar to CDC-like kinase 2, which may regulate mRNA splicing
	SPAC23H4.15	Member of DUF663 protein of unknown function family, possible role in rRNA processing and 40S ribosomal subunit biogenesis
	SPAC31A2.07c	Putative RNA helicase, possible role in ribosome biogenesis
	SPCC1494.06c	Member of the DEAD or DEAH box ATP-dependent RNA helicase, possible role in rRNA processing
	SPCC1682.08c	Protein containing six Pumilio-family RNA binding domains, possible role in mRNA metabolism
	SPAC16C9.03	Possible role in nuclear export of 60S ribosomal subunits
Cell wall biosynthesis and maintenance		
bgl2	SPAC26H5.08c	Protein similar to beta-glucosidase, a cell wall endo-beta-1,3-glucanase
psu1	SPAC1002.13c	Protein required for cell wall integrity, member of SUN protein family
	SPBC11C11.05	Member of yeast cell wall synthesis protein KRE9 or KNH1 family

Unclassified genes

48 genes could not be assigned to a specific cluster (Table 3.5). Regardless of the clustering method used, classification is somewhat arbitrary and this becomes more evident when looking at the genes at the boundary of each cluster where assignment becomes difficult. For all the unclassified genes, Arrayminer gives an estimate of the

closest cluster they could belong to. Taking this into account and also considering the timing of peak of expression of each gene, some of these 48 genes have been assigned to smaller clusters named 1/2, 2/3, 3/4 and 4/1 (Fig. 3.3).

Among the genes peaking in M and G1 (Fig. 3.3 - cluster 1/2) is *hsk1*, encoding a factor responsible for DNA replication initiation in association with its partner dfp1p (gene member of cluster 2). *Mfm2*, *spk1* and *byr2* are all involved in mating, sporulation and the pheromone signalling pathway. Another essential meiotic gene expressed at this boundary is *mei2*, which encodes an RNA binding protein crucial for initiation of premeiotic DNA synthesis and meiosis I.

The genes assigned to the G2/M boundary (Fig. 3.3 - cluster 4/1) include a sulphate transporter family member and several enzymes involved in different metabolic pathways, similar functions to most of the previously characterised cluster 4 members. *Sim4*, involved in chromosome segregation it is also part of this group.

The majority of the genes peaking at either G1/S (Fig. 3.3 - cluster 2/3) or S/G2 (Fig. 3.3 - cluster 3/4) do not have a well characterised function except *fim1*, a fimbrin coding gene, which may be involved in polarised growth, assigned to cluster 3/4.

Table 3.6 Selected unclassified genes and their biological function

Biological names	Systematic names	Gene description
Cluster 1/2 (M-G1 genes)		
vip1	SPAC10F6.06	Protein containing an RNA recognition motif
prl3		Non-coding RNA
hsk1	SPBC776.12c	Protein kinase of the Hsk1p-Dfp1p complex involved in S phase initiation
mfm2	SPAC513.03	Precursor polypeptide for mating pheromone M factor produced by h- cells
spk1	SPAC31G5.09c	MAP kinase (MAPK) acting in the mating and sporulation pathways
byr2; ste8	SPBC2F12.01; SPBC1D7.05	MAP kinase kinase kinase acting upstream of MAPKK Byr1p and MAP kinase Spk1p in pheromone signaling pathway
	SPAC1006.06	Protein with RhoGEF domain, similar to Rho GDP-GTP exchange factor activated by cell wall defects
	SPAC12G12.06c	Probable RNA 3'-terminal phosphate cyclase
	SPBC1683.07	Protein similar to alpha-glucosidase
hri1	SPAC20G4.03c	Translation initiation factor 2 alpha kinase, may play role in negative regulation of eIF2alpha in response to stress
mei2	SPAC27D7.03c	RNA-binding protein involved in meiosis
spm1; pmk1	SPBC119.08	MAP kinase involved in maintenance of cell wall integrity
	SPCC965.06	Protein similar to potassium voltage-gated channel
cnd2	SPCC306.03c	Subunit of condensin complex involved in chromosome condensation

isp6; prb1	SPAC4A8.04	Putative subtilase-type proteinase, role in sexual differentiation+E49
Cluster 4/1 (G2/M genes)		
	SPAC869.05c	Member of sulfate transporter family, similar to sulfate permease
	SPAC1002.17c	Protein with phosphoribosyl transferase domain, possible role in pyrimidine salvage pathway
	SPCC16C4.06c	Protein with tRNA pseudouridine synthase domains
sim4	SPBC18E5.03c	Centromere-associated protein required for chromosome segregation and silencing
	SPBC19G7.07c	Member of PPR repeat containing family
	SPCC330.15c; SPCC320.14	Member of pyridoxal phosphate dependent enzyme family, similar to racemase that catalyzes the racemisation of L-serine to D-serine
Cluster 2/3 (G1/S genes)		
	SPCC553.07c	Member of impB, mucB or samB family, possible role as translesion DNA repair polymerase
	SPBC800.11	Protein with inosine-uridine preferring nucleoside hydrolase domain
	SPBC409.22c; SPBC1306.01c	Protein with elongation factor Tu GTP binding domain, similar to mitochondrial translation elongation factor G
	SPAC17G6.03	Protein with calcineurin-like phosphoesterase domain
	SPBC21B10.07; SPAC21B10.07	Protein with glycosyl hydrolase family 16 domain
	SPAC2E1P3.04	Protein with possible role in detoxifying extracellular amines and nitrogen metabolism
	SPAC29A4.05	Protein similar to calmodulin 1, which regulates the calcium-dependent activity of enzymes including phosphatases
Cluster 3/4 (S/G2 genes)		
	SPAC631.02	Protein with two bromodomains, which interact with acetylated lysine
	SPBC1271.09	Member of sugar (and other) transporter family, possible role in inositol metabolism
	SPCC364.07; SPCC4G3.01	Protein similar to 3-phosphoglycerate dehydrogenase, which catalyzes first step in synthesis of serine
	SPAC3A11.10c	Member of Rnal dipeptidase family, zinc-dependent metalloproteinases that hydrolyze various dipeptides
	SPAC664.03	Member of Paf1 family, components of RNA polymerase II associated complexes
pro1	SPAC821.11	Protein similar to gamma-glutamyl phosphate reductase involved in proline biosynthesis
fim1	SPBC1778.06c	Fimbrin, role in actin organization during medial ring formation and polarized growth
	SPBPB21E7.09; SPAPB21E7.09	Protein similar to L-asparaginase II