

The origin and significance of the *THI5* gene family in *Saccharomyces cerevisiae*.

**Thesis submitted for the Degree of
Doctor of Philosophy at the
University of Leicester**

By

Raymond Wightman

Department of Genetics
University of Leicester

December 2001

Abstract.

The *Saccharomyces cerevisiae* *THI5* gene family consists of four highly conserved members; *THI5*, *THI11*, *THI12* and *THI13*. Each member is located within the subtelomeric region of a different chromosome and they share these regions with members of other gene families. A detailed analysis of the genome environment of each *THI5* family member has shown all four to exist with their neighbouring genes in a highly conserved arrangement.

A survey of *THI5* copy number among a selection of hemiascomycetes suggests that the existence of *THI5* as a gene family is exclusive to those yeasts of the *Saccharomyces sensu stricto* complex. Furthermore, the association and organisation with respect to neighbouring members of other subtelomeric gene families appears to be conserved between *S. cerevisiae* and other *sensu stricto* yeasts. This conservation of local gene order does not appear to extend to yeasts outside of this subgroup.

S. cerevisiae strains have been constructed which contain deletions of *THI5*, *THI11*, *THI12* and *THI13* in all possible combinations to yield the single, double, triple and quadruple mutants. Phenotypic analysis of the quadruple deletion mutant has found it to exhibit thiamin (vitamin B1) auxotrophy on medium containing glucose as the carbon source. This auxotrophy can be remedied by the exogenous addition of thiamin or one of its precursors hydroxymethyl-pyrimidine (HMP). Analysis of other mutant strains has shown that the *THI5* gene family members are functionally redundant, with each encoded isozyme having an apparent equal role in HMP formation from pyridoxine (vitamin B6). Physiological studies of these mutant strains have examined the regulation of each of the four genes by thiamin and its precursors, as well as investigating a recently proposed anaerobic HMP biosynthetic pathway.

Abbreviations

AAD	Aryl alcohol dehydrogenase
ADH	Alcohol dehydrogenase
ADP	Adenosine diphosphate
AHAS	Acetohydroxyacid synthase
AIR	5-Aminoimidazole ribotide
ALS	δ -aminolevulinate synthase
Amp	Ampicillin
ARS	Autonomously replicating sequence
ATP	Adenosine triphosphate
bp	Base pairs
CBS	Centraalbureau voor Schimmelcultures
dATP	Deoxyadenosine 5'-triphosphate
dCTP	Deoxycytidine 5'-triphosphate
DET	Derepressed expression on thiamin
dGTP	Deoxyguanosine 5'-triphosphate
dITP	Deoxyinosine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide 5'-triphosphate
dTTP	Deoxythymidine 5'-triphosphate
EDTA	Ethylenediamine tetra acetic acid
GCG	Genetics computer group
GFP	Green fluorescent protein
Gln	Glutamine
HET	4-Methyl-5- β -hydroxyethylthiazole
HET-p	HET phosphate
HMP	2-Methyl-4-amino-5-hydroxymethylpyrimidine
HMP-p	HMP phosphate
HMP-pp	HMP diphosphate
IMS	Industrial methylated spirits
Kb	Kilobases
KGD	α -Ketoglutarate dehydrogenase
LB	Luria Bertani
mRNA	Messenger ribonucleic acid
NAD	Nicotinic adenine dinucleotide
NCYC	National collection of yeast cultures
NMR	Nuclear magnetic resonance
nmt	No message in thiamin
nt	Nucleotides
OD	Optical density

ONPG	O-Nitrophenyl-D-galactopyranoside
ORF	Open reading frame
PCR	Polymerase chain reaction
PDC	Pyruvate decarboxylase
PDH	Pyruvate dehydrogenase
PEG	Polyethylene glycol
PLP	Pyridoxal-5'-phosphate
PRA	Phosphoribosyl-1-amine
PRPP	5-phosphoribosyl-1-pyrophosphate
Pyr	Pyridoxine
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RTM	Resistance to toxicity of molasses
SCM	Synthetic complete media
SD	Synthetic dextrose
SDS	Sodium dodecyl sulphate
SGD	<i>Saccharomyces</i> genome database
SUC	Sucrose utilisation
TCA	Tricarboxylic acid
Thi	Thiamin
TK	Transketolase
ThMP	Thiamin monophosphate
TMP-PP	Thiamin phosphate pyrophosphorylase
ThP	Thiamin phosphate
ThDP	Thiamin diphosphate
TKL	Transketolase
TPS	Thiamin phosphate synthase
UV	Ultraviolet
W/V	Weight / volume
YPD	Yeast peptone dextrose
YLE	Yeast lytic enzyme
YGSC	Yeast Genetic Stock Centre

Contents.

Chapter 1 General introduction

1.1 The hemiascomycetes.....1

1.1.1 Taxonomy.....1

1.1.2 The *Saccharomyces* genus.....2

1.1.3 The *Saccharomyces sensu stricto* complex.....2

1.2 *Saccharomyces cerevisiae*.....3

1.2.1 Anaerobic growth.....4

1.2.2 The *S. cerevisiae* genome and evidence of a whole-genome duplication event.....5

1.2.3 Gene families.....7

1.2.4 Subtelomeric regions.....8

1.3 The cofactor thiamin diphosphate.....11

1.4 Thiamin diphosphate biosynthesis.....13

1.4.1 Fungal vs. bacterial thiamin biosynthesis.....13

1.4.2 Synthesis of active cofactor from thiazole and pyrimidine precursors in *S. cerevisiae*.....14

1.4.3 Synthesis of the thiazole precursor.....15

1.4.4 Uptake of thiamin and the precursors.....16

1.4.5 Regulation of the biosynthetic enzymes.....18

1.4.6 *S. cerevisiae* mutants that exhibit derepression in the presence of thiamin.....22

1.5 Biosynthesis of the pyrimidine precursor, HMP-PP.....22

1.5.1 Precursor labelling studies.....23

1.5.2 The *S. cerevisiae* *THI5* gene family.....24

1.5.3 The *THI5* homologue in *S. pombe*, *nmt1*.....25

1.5.4 The *S. cerevisiae* HMP-P kinase family.....26

1.5.5 Evidence of an anaerobic HMP biosynthetic pathway.....27

1.6 Pyridoxine.....28

1.6.1 Biosynthesis.....28

1.6.2 Regulation of *SNO* and *SNZ* by pyridoxine and thiamin.....31

1.6.3 The effect of thiamin in pyridoxine-free medium.....31

1.7 Aims of the project.....32

Chapter 2 A bioinformatic analysis of thiamin biosynthesis in various organisms and of the *S. cerevisiae* *THI5* gene family.

2.1 Introduction.....34

2.2 BLASTP searches.....35

2.2.1 Assignment of pathways to thiamin precursors within the domains of bacteria, archaea, plants and fungi.35

2.2.2 Fungal Thi5p homologues.37

2.3 Analysis of the genome environment of the *S. cerevisiae* *THI5* gene family.....38

2.3.1 Comparison of the *THI5*, *THI11*, *THI12* and *THI13* promoters in *S. cerevisiae*.....38

2.3.2 Organisation of the *THI5* genes in *S. cerevisiae*.39

2.5 Discussion.....41

Chapter 3 A phylogenetic analysis of *THI5* copy number among the hemiascomycetes.

3.1 Introduction.....44

3.2 Survey of *THI5* copy number.45

3.2.1 The parental strains of *S. cerevisiae*. S288C.45

3.2.2 Laboratory and Commercial strains of *S. cerevisiae*.....46

3.2.3 Other yeasts of the *Saccharomyces sensu stricto* complex.48

3.2.4 The *Saccharomyces sensu lato*.....49

3.2.5 The *Kluyveromyces* genus.....50

3.3 Further analysis of hemiascomycetes that lack a *THI5* gene.....50

3.4 *S. cariocanus* is a thiamin auxotroph.51

3.5 Discussion.....52

Chapter 4 Determination of *THI5* gene order conservation between the *Saccharomyces sensu stricto*, *Saccharomyces kluyveri* and *Kluyveromyces lactis*.

4.1 Introduction.....55

4.2 The *Saccharomyces sensu stricto*.....56

4.3 *Saccharomyces kluyveri*.56

4.3.1 Isolation of rescuing clones.56

4.3.2 DNA sequencing and contig construction of *S. kluyveri* rescuing clone, pRW3a.57

4.3.3 The ORF adjacent to *Sk.THI5* resembles a *S. cerevisiae* DNA binding transcription factor involved in pyrimidine nucleotide biosynthesis.58

4.4 *Kluyveromyces lactis*.....59

4.4.1 DNA sequencing and contig construction of *K. lactis* clones pDW5a and pDW5b.....59

4.4.2 Identification of adjacent ORFs to the <i>THI5</i> homologue, <i>Kl.THI5</i>	59
4.4.3 Determining the role of ORF3.	60
4.4.4 Analysis of <i>Kl.COS</i> copy number in <i>K. lactis</i>	61
4.5 Discussion.....	62
 Chapter 5 A functional analysis of the <i>THI5</i> gene family in <i>S. cerevisiae</i>.	
5.1 Introduction.....	65
5.2 Creation of mutant strains containing total deletions of the <i>THI5</i> loci.....	65
5.2.1 Deletion strategy.....	65
5.2.2 Creation and confirmation of all four single gene deletions, strains RWY1-RWY4.....	66
5.2.3 Creation and confirmation of the six double mutants and the four triple mutants, strains RWY5-RWY14.....	67
5.2.4 The construction of the quadruple mutant strains, RWY15 and RWY16.....	68
5.3 Phenotypic analysis of the RWY mutant <i>S. cerevisiae</i> strains.	68
5.3.1 Phenotypic analysis with regards to thiamine biosynthesis.	68
5.3.2 Effect of alternative carbon sources on thiamin auxotrophy of RWY15.....	69
5.3.3 Growth on blackstrap molasses.	70
5.4 Discussion.....	71
 Chapter 6 Physiological characterisation of <i>THI5</i>, <i>THI11</i>, <i>THI12</i> and <i>THI13</i>.	
6.1 Introduction.....	75
6.2 Phenotypic comparison of <i>S. cerevisiae</i> strain RWY15 and the <i>K.lactis klthi5</i> mutant.	75
6.2.1 Phenotypic analysis of the <i>K. lactis klthi5</i> mutant.....	75
6.2.2 The slow growth phenotype of RWY15.....	76
6.3 Aerobic vs. anaerobic thiamin biosynthesis in <i>S. cerevisiae</i>	77
6.3.1 The growth of RWY15 compared to parent strains.....	77
6.3.2 Monitoring growth in aerobic and anaerobic batch cultures.....	78
6.4 Regulation of the <i>THI5</i> gene family.....	80
6.4.1 Northern blot hybridisation of the triple mutants.....	80
6.4.2 Expression of <i>THI5</i> in <i>S. cerevisiae</i> during glucose batch culture.....	82
6.4.3 Slot blot hybridisation analysis.	82
6.4.4 Construction of promoter-reporter gene fusions.	84
6.4.5 Assays of the reporter plasmids.....	85
6.5 Discussion.....	86

Chapter 7 General discussion.

7.1 The HMP gene clusters of *Saccharomyces cerevisiae*.....96

7.2 Mechanism of subtelomeric
amplification.....97

Chapter 8 Materials and Methods.

8.1 Growth media and conditions.....99

8.1.1 Luria broth (LB).....99

8.1.2 Yeast peptone (YP) medium.99

8.1.3 Synthetic dextrose (SD) medium.....99

8.1.4 Wickerham’s minimal medium..... 100

8.1.5 Blackstrap molasses agar medium..... 100

8.1.6 The synthetic complete powder mix (SCM)..... 100

8.2 Bacterial and Yeast strains. 101

8.2.1 *E.coli* strain storage..... 101

8.2.2 *S. cerevisiae* strain storage 101

8.3 Plasmids, vectors and genomic libraries..... 101

8.3.1 The *S. kluyveri* genomic library..... 101

8.4 Isolation of DNA and RNA..... 103

8.4.1 Small scale *E. coli* plasmid preparations..... 103

8.4.2 Medium scale, highly purified *E. coli* plasmid preparations..... 103

8.4.3 Large scale, highly purified *E. coli* plasmid preparations. 103

8.4.4 Small scale, rapid *S. cerevisiae* plasmid preparations..... 103

8.4.5 Rapid preparation of chromosomal DNA from *S. cerevisiae*. 104

8.4.6 Large scale preparation of highly purified chromosomal DNA from the
hemiascomycetes. 104

8.4.7 Preparation of total RNA from *S. cerevisiae*. 105

8.5 DNA transformation of bacteria and yeast. 106

8.5.1 *E.coli* transformation..... 106

8.5.2 High efficiency transformation of *S. cerevisiae*..... 106

8.6 Nucleic acid electrophoresis. 107

8.6.1 DNA electrophoresis..... 107

8.6.2 RNA electrophoresis..... 107

8.7 Southern-, northern-, and slot-blot hybridisations..... 107

8.7.1 Southern blot hybridisation..... 107

8.7.2 Northern blot hybridisation..... 108

8.7.3 Slot blot hybridisation..... 108

8.7.4 Preparation of high specific activity probes for northern- and slot-blot hybridisation.. 109

8.7.5 Pre-hybridisation and hybridisation of RNA-bound membranes. 109

8.7.6 DNA probes used in this study. 110

8.8 DNA manipulations..... 110

8.8.1 DNA restriction endonuclease digestion. 110

8.8.2 DNA elution for agarose gels. 110

8.8.3 DNA ligation..... 111

8.9 Methods using the Polymerase Chain Reaction (PCR)..... 111

8.9.1 Reaction conditions for American Yeast Foam (Chapter 3)..... 111

8.9.2 Reaction conditions for degenerate PCR analysis (Chapter 4)..... 111

8.9.3 Reaction conditions for synthesis of disruption cassettes..... 112

8.9.4 Reaction conditions for amplification of promoters. 112

8.9.5 Oligonucleotide primers. 113

8.10 DNA sequencing..... 113

8.11 Assay of β-galactosidase activity in yeast..... 113

8.12 Large-scale aerobic and anaerobic batch fermentations..... 114

8.13 Methods using computer-based analyses..... 115

8.13.1 Computer-generated alignments..... 115

8.13.2 Contig construction of DNA sequence data..... 115

8.13.3 Measuring intensities of bands on RNA northern- and slot- blots. 115

References.....116

Chapter 1.

General introduction.

1.1 The hemiascomycetes.

1.1.1 Taxonomy.

The yeasts are a unique group of fungi which grow vegetatively as predominantly unicellular organisms. They are distributed between the phyla basidiomycota and ascomycota. Yeasts of the latter phylum are distinguished from other ascomycete fungi by the formation of sexual asci that are not enclosed by a fruiting body.

Modern classification of the fungi is largely based upon the use of molecular comparisons especially with regards to similarities in ribosomal RNA sequences (Kurtzman, 1994). This has resulted in a revised phylogeny of the ascomycota which branches into three classes (Souciet *et al.*, 2000). The first class is the Archiascomycetes and includes the fission yeast *Schizosaccharomyces pombe*. It is thought that *S. pombe* represents the lineage that diverged the earliest from the common ancestor of the ascomycota and basidiomycota (Nishida & Sugiyama, 1993). The second is the Euascomycetes and contains the filamentous fungi such as *Neurospora* and *Aspergillus* species whilst the third class, hemiascomycetes, represents the budding yeasts. Many of the ascomycete yeasts exhibit a very high degree of divergence from each other. For example, rRNA sequence analysis shows that *S. pombe* is phylogenetically as distant from budding yeasts as it is from humans (Komiya *et al.*, 1981).

Figure 1.1 shows the phylogenetic relationships of a selection of hemiascomycetes based on comparisons of the nucleotide sequences of their 18S rRNA subunits. These yeasts are of the genera *Saccharomyces*, *Kluyveromyces*, *Torulaspora* and *Zygosaccharomyces*. Species were originally placed in one of these domains based entirely on phenotypic characteristics including morphology, nutritional

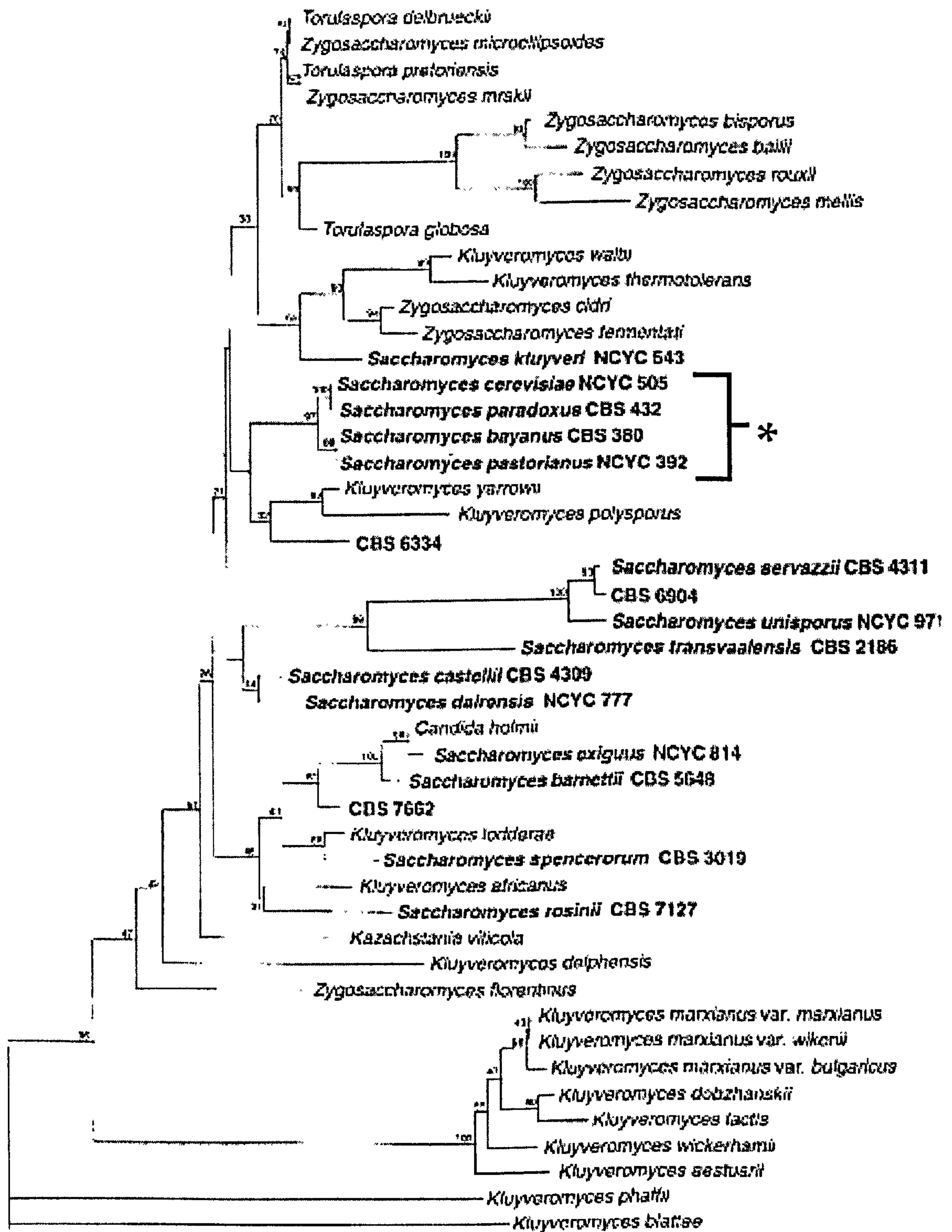


Figure 1.1 Phylogenetic tree based on 18S rRNA sequence data from species of the genera *Kluyveromyces*, *Saccharomyces*, *Torulaspora* and *Zygosaccharomyces*.

* indicates those yeasts of the *Saccharomyces sensu stricto* complex.

Tree taken from James *et al.* Int. J. Syst. Bacteriol. 47 (1997) 453-460. The authors determined the 18S rRNA sequences of those organisms shown in bold type.

requirements and substrate assimilation. However, molecular classification often sees these species to be phylogenetically intermixed with those of other genera (Figure 1.1; James *et al.*, 1997).

1.1.2 The *Saccharomyces* genus.

As shown in figure 1.1, many species of the *Saccharomyces* genus are taxonomically distributed between other genera. There is a noticeable exception which can be seen in figure 1.1. Four *Saccharomyces* yeasts are grouped very close together, displaying 18S rRNA sequence similarity of greater than 99% (James, *et al.*, 1997). These are *Saccharomyces cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces bayanus* and *Saccharomyces pastorianus* and are representative species of the closely related subgroup of the *Saccharomyces* termed the *sensu stricto* complex. These yeasts exhibit high overall phenotypic resemblance and are only distinguishable by a few phenotypic traits such as growth temperature and assimilation of D-mannitol (Vaughan-Martini & Martini, 1993).

In contrast to the *sensu stricto* complex, other *Saccharomyces* yeasts are only loosely related both in terms of rRNA sequence data and physical characteristics (Vaughan-Martini and Martini, 1993). For this reason they are further classified as the *Saccharomyces sensu lato* subgroup. Unlike the *sensu stricto* species, which possess very similar numbers of chromosomes, members of the *sensu lato* have varied numbers ranging from seven (*Saccharomyces kluyveri*) to sixteen (*Saccharomyces exiguus*) (Keogh *et al.*, 1998; Petersen *et al.*, 1999).

1.1.3 The *Saccharomyces sensu stricto* complex.

Although they exhibit very similar properties, members of the *sensu stricto* subgroup are discrete species as shown by DNA-DNA reassociation studies (Martini & Kurtzman, 1985). Haploid cells of *S. cerevisiae*, *S. bayanus* and *S. paradoxus* are able to mate with each other and form zygotes that can sporulate. The spores of these

interspecific hybrids are all inviable which is probably due to sequence divergence between these species which prevents meiotic pairing and recombination (Mortimer, 2000). Hybrids of *S. cerevisiae* and *S. paradoxus* which possess deletions of the mismatch repair genes *PMS1* or *MSH2* show an increase in meiotic recombination and an increase in spore viability (Hunter *et al.*, 1996). The commercial bottom-fermenting yeast, *S. pastorianus*, is thought to be an interspecific hybrid as it forms spores that are inviable (Rainieri *et al.*, 1999). Southern blot hybridisation of whole chromosomal separations has been carried out on this yeast and a variety of known *S. cerevisiae* genes have been used as probes (Tamai *et al.*, 1998). When compared to *S. cerevisiae* and *S. bayanus*, many of the gene probes produce two positive hybridising chromosomes, one of similar size to that of the former yeast and the other of similar size to that of the latter. This suggests that two types of chromosomes exist in *S. pastorianus*, one from each parent. Therefore *S. pastorianus* is considered to be a hybrid of a *S. cerevisiae* strain and a *S. bayanus* strain (Vaughan Martini and Martini, 1987; Tamai *et al.*, 1998).

Recently, three more yeasts have been included within the *sensu stricto*. All yield sterile hybrids when crossed with one another and with other members of the subgroup. These new members were obtained from geographically isolated areas in Brazil and Japan (Naumov *et al.*, 1995a; Naumov *et al.*, 1995b). They have been formally described under the names *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae* (Naumov *et al.*, 2000).

1.2 *Saccharomyces cerevisiae*.

Most of the laboratory strains currently used in studies on the genetics and molecular biology of *S. cerevisiae* are derived from strain S288C. S288C was constructed from several progenitor strains called EM93, EM126, NRRL-210 and the commercial baking yeasts FLD, LK and American Yeast Foam. It has been calculated that EM93 has contributed 88% of the S288C genome whilst the rest make up the balance (Mortimer & Johnston, 1986). EM93 was the first strain used by Lindegren,

one of the pioneers of yeast genetics. This particular strain was isolated in 1938 from rotting figs in California and was probably a wine yeast carried there by insects (Mortimer, 2000; Mortimer and Johnston, 1986). Lindegren was fortunate that this yeast was heterothallic, meaning that the mating type of haploid cells are stable which therefore facilitates the construction of genetic crosses.

Another principal laboratory strain is Σ 1278b. S288C cannot grow pseudohyphally when starved for nitrogen whereas Σ 1278b can. This is due to a nonsense mutation in the *FLO8* gene in S288C (Liu *et al.*, 1996). Therefore, Σ 1278b is the strain of choice for studies into this dimorphic switch. It is derived from two parents, one of which is American Yeast Foam (Mortimer and Johnston, 1986).

Unlike laboratory strains which are either haploid or diploid, many commercial strains of *S. cerevisiae* are either aneuploid or polyploid. In addition, they exhibit extremely low spore viabilities and possess subtle phenotypic differences such as the ability to assimilate certain substrates. The diversity of strains among this species has prompted much work to discriminate them based on molecular polymorphisms (Azumi and Goto-Yamamoto, 2001; de Barros Lopes *et al.*, 1998; Wightman *et al.*, 1996). This is considered important in the brewing and wine industries where various strains of *S. cerevisiae* have been implicated in spoilage of fermented products (Baleiras Couto *et al.*, 1996).

1.2.1 Anaerobic growth.

S. cerevisiae is a facultative anaerobe. In the absence of oxygen, catabolism of sugars such as glucose is directed completely to fermentation products. During anaerobiosis, this yeast needs an exogenous supply of ergosterol and oleic acid due to the absolute requirement for oxygen during their biosynthesis (Walker, 1998) (pp236-238). However, for other metabolic pathways which require a redox step, *S. cerevisiae* has evolved alternative mechanisms to allow the pathways to function under anaerobic conditions. For example, in the aerobic yeast *S. pombe*, the single

redox step in the production of pyrimidine nucleotides is catalysed by dihydroorotate dehydrogenase. This enzyme is localised in the mitochondria where the respiratory chain provides the route to disposal of the reducing equivalents of the reaction. The gene encoding this protein can functionally rescue the mutant *S. cerevisiae* homologue but only under aerobic conditions. This is because the *S. cerevisiae* protein lacks a mitochondrial targeting sequence and instead this yeast uses a dicarboxylic acid as a reducing agent (Nagy *et al.*, 1992).

The ability to grow anaerobically is uncommon among the hemiascomycetes in general (Piskur, 2001). *S. cerevisiae* is a 'petite-positive' yeast, that is, it can produce respiration-deficient mutants which are characterised by a reduced colony size. *S. kluyveri*, like *S. cerevisiae*, can grow anaerobically but cannot generate true petite mutants and *K. lactis* can neither grow anaerobically nor exhibit the petite phenotype (Moller *et al.*, 2001). Based on this evidence, it appears that the ancestral yeast of *S. cerevisiae* and *K. lactis* was an obligate aerobe. Upon separation of these lineages *S. cerevisiae* developed the ability to grow anaerobically which, after further separation from *S. kluyveri*, could happen independently of the respiratory chain. Presented in the next section will be a hypothesis that a whole-genome duplication event occurred somewhere along the *S. cerevisiae* lineage. It is suggested that this duplication allowed the specialisation of a number of genes which enabled their optimal function even in the absence of oxygen and mitochondrial function (Piskur, 2001; Wolfe & Shields, 1997) .

1.2.2 The *S. cerevisiae* genome and evidence of a whole-genome duplication event.

S. cerevisiae was the first eukaryote to have its whole genome sequence elucidated with a predicted gene number of just under 6000. It quickly became apparent that a significant number of genes are duplicated within the genome (Goffeau *et al.*, 1996). Many of these homologues are themselves part of larger duplicated blocks of genes, called cluster homology regions, that are interspersed with single-copy genes. A total of 55 of these regions (excluding the subtelomeres) have been identified and

these comprise 50% of the whole genome (Wolfe and Shields, 1997). A model has been put forward by Wolfe and Shields to explain these observations. They suggest that *S. cerevisiae* is a degenerate tetraploid formed by the fusion of two diploid ancestors, each containing 5000 genes. It is thought that approximately 85% of the duplicated copies were deleted, followed by a series of reciprocal translocations between duplicated chromosomes, to yield the present *S. cerevisiae* genome encompassing 6000 genes. The time of the proposed genome duplication has been placed after the *Kluyveromyces* genus diverged from *Saccharomyces* (Wolfe and Shields, 1997). This has since been narrowed down to some time after the divergence of *Saccharomyces sensu stricto* from *S. kluyveri* (Keogh, *et al.*, 1998).

Another hypothesis has been put forward by workers involved in the recent Génolevures project which has partially sequenced the genomes of thirteen hemiascomycetes ranging from *S. bayanus* to *Yarrowia lipolytica* (Feldmann, 2000). The Genolevures team have identified all the gene couples in their sequence data and have compared their positions on the *S. cerevisiae* chromosome maps. They have found evidence of single gene and chromosome segmental inversions that have occurred in the lineages between *S. cerevisiae* and the thirteen other yeasts. The fact that gene orientation with respect to the centromeres is not always conserved between species indicates that the duplications were probably not due to a whole genome duplication. The team have put forward a hypothesis of yeast genome evolution occurring by duplication of segments that encompass a few genes rather than by the duplication of entire chromosomes. They propose that the duplicate segment is inserted at a different location in a random orientation relative to the centromeres and is followed by numerous single gene deletions within these segments (Llorente *et al.*, 2000). Both theories of *S. cerevisiae* genome evolution agree that duplication is followed by many small deletions.

1.2.3 Gene families.

The proposed genome duplication of a *S. cerevisiae* ancestor has resulted in a large number of duplicated genes that are still present within the existing genome. In addition, there are a significant number of gene families containing more than two members; the largest is the seripauperin (*PAU*) gene family containing 23 members in total (Viswanathan *et al.*, 1994). Such families of paralogous genes account for about 40% of the total number of predicted open reading frames in *S. cerevisiae* (Blandin *et al.*, 2000).

Members of some gene families have different functions whilst others appear to be functionally redundant. An example of functional divergence is seen within the alcohol dehydrogenase family which encompasses five isozymes. The enzyme encoded by *ADH1* is involved in the production of ethanol from acetaldehyde whilst *ADH2* encodes the enzyme which catalyses the reverse reaction used in gluconeogenesis. Unlike *ADH1*, *ADH2* expression is glucose-repressed. This glucose regulation is only seen within the *sensu stricto* complex of the *Saccharomyces* genus (Young *et al.*, 2000). Both *ADH1* and *ADH2* encode cytosolic isozymes whereas *Adh3p* is targeted to the mitochondria. This enzyme is important for regenerating NAD through the reduction of acetaldehyde in the mitochondria during anaerobic growth (Bakker *et al.*, 2000). The roles of *Adh4p* and *Adh5p* are unclear.

An example of functional redundancy is seen in the hydroxymethylpyrimidine-phosphate (HMP-P) kinase family and will be described in more detail in section 1.5.4. There are three members of this gene family; *THI20* and *THI21* encode proteins that are functionally redundant in terms of HMP-P phosphorylation, whilst a third, more diverged member, *THI22*, appears not to function in this pathway (Llorente *et al.*, 1999). The reason for retaining multiple genes in yeast has been partly addressed by Seoighe and Wolfe, who have examined the large number of duplicates within the genome (Seoighe & Wolfe, 1999). They suggest two evolutionary reasons for retaining both copies; selection for increased levels of gene

expression or divergence of gene function. This may explain the role of the three genes encoding HMP-P kinases. *THI20* and *THI21* may serve a requirement for high levels of mRNA whilst *THI22* may encode an enzyme with a different but as yet unidentified function.

1.2.4 Subtelomeric regions.

The subtelomeric regions from an increasing variety of organisms have been characterised as part of their respective genome projects and are found to contain a combination of various repetitive sequences. In many organisms these regions are dynamic and exhibit frequent exchanges. Studies on a number of parasites and pathogens have shown the subtelomeric region to be used for the generation of antigenic diversity. Examples include the var gene family of *Plasmodium falciparum* and the genes encoding the Variant Surface Glycoprotein of *Trypanosoma brucei* (Freitas-Junior *et al.*, 2000; Rudenko *et al.*, 1996). Consistent with this, the subtelomeric domains of the human pathogen *Plasmodium vivax* have been recently examined and a novel gene family has been identified. This gene family, called *vir*, is described as being 'immunovariant' during infection where the high subtelomeric recombination frequency results in the rapid generation of *P. vivax* variants with novel antigenic phenotypes (del Portillo *et al.*, 2001).

In *S. cerevisiae*, the region adjacent to the TG₁₋₃ repeats that define the telomere, consists of two elements called Y' and X (Louis, 1995). Y' elements are only found at some of the chromosome ends and are not generally found within all *Saccharomyces* genomes, being restricted to the *sensu stricto* members *S. paradoxus* and *S. cerevisiae*. These elements are highly conserved and their precise function is unclear. In comparison, X elements are considerably less conserved but are found at all chromosome ends, indicating a possible role in chromosome stability.

The completion of the *S. cerevisiae* genome project revealed the presence of a large number of gene families concentrated within the subtelomeric regions of its chromosomes (Goffeau, *et al.*, 1996). Most of these families have members that are

found centromere-proximal to the X element whilst the *RTM* (Resistance to Toxicity of Molasses) and *SUC* (sucrose utilisation) genes are found between X and the Y' element on the telomere-proximal side (Louis, 1995).

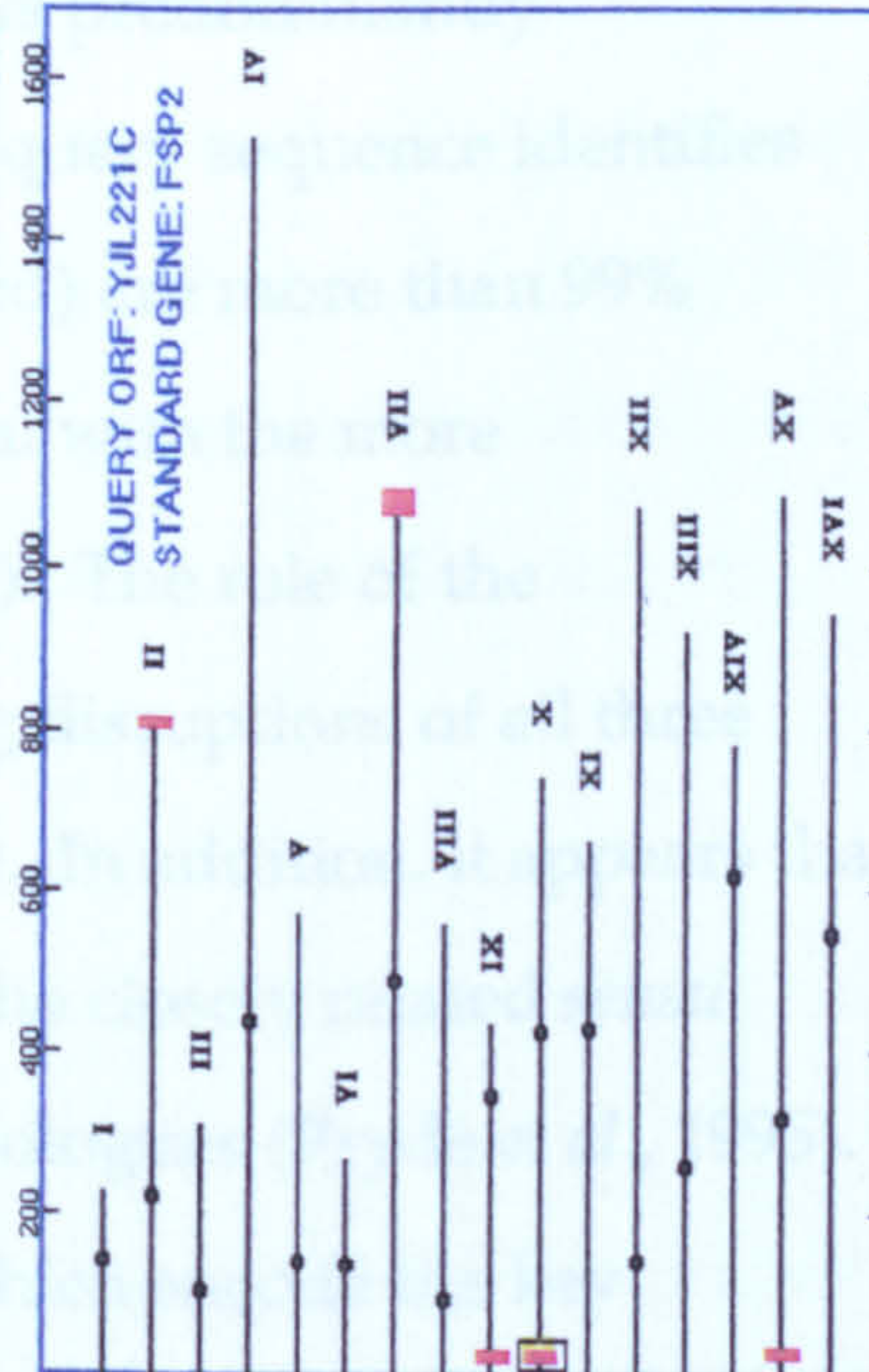
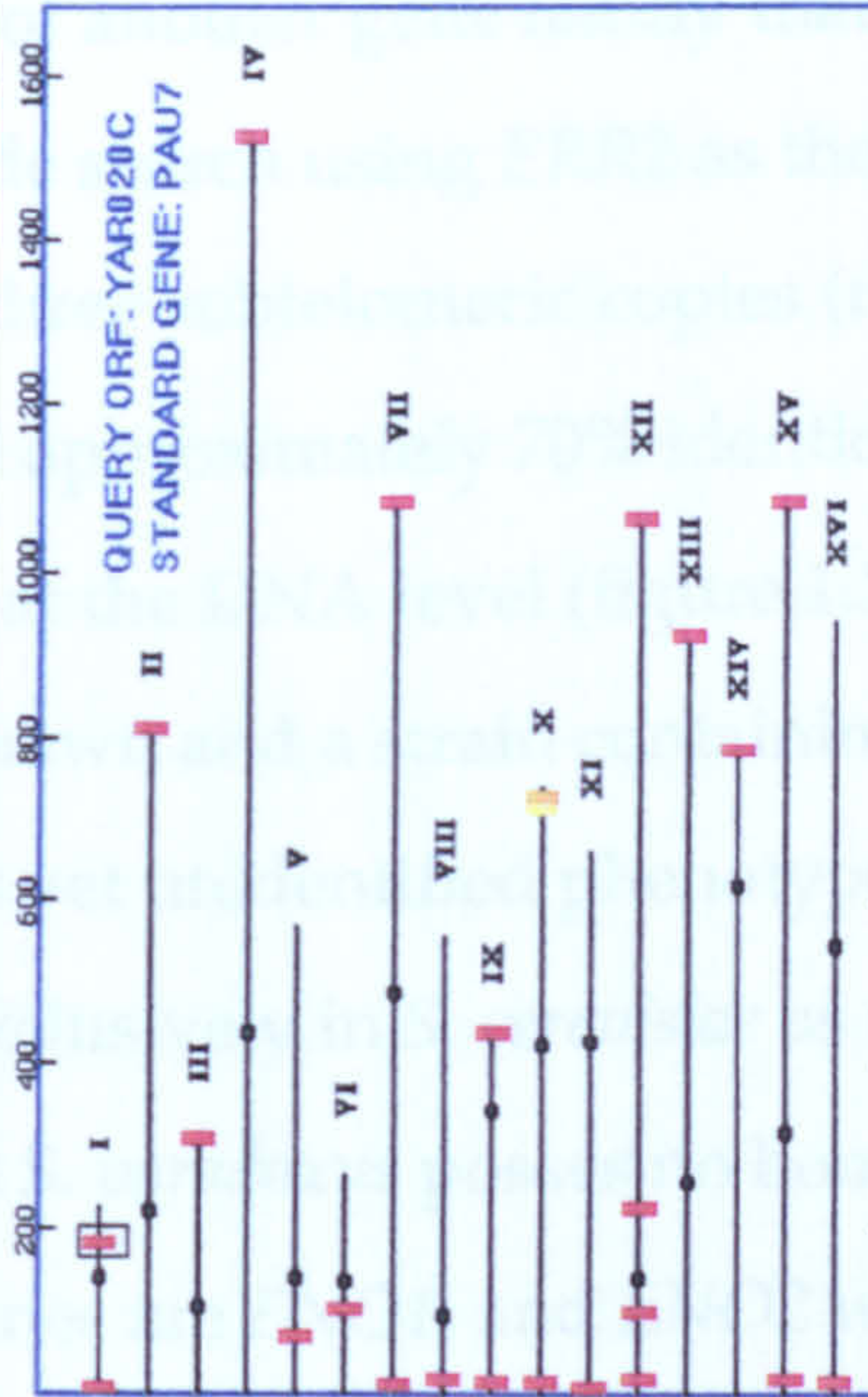
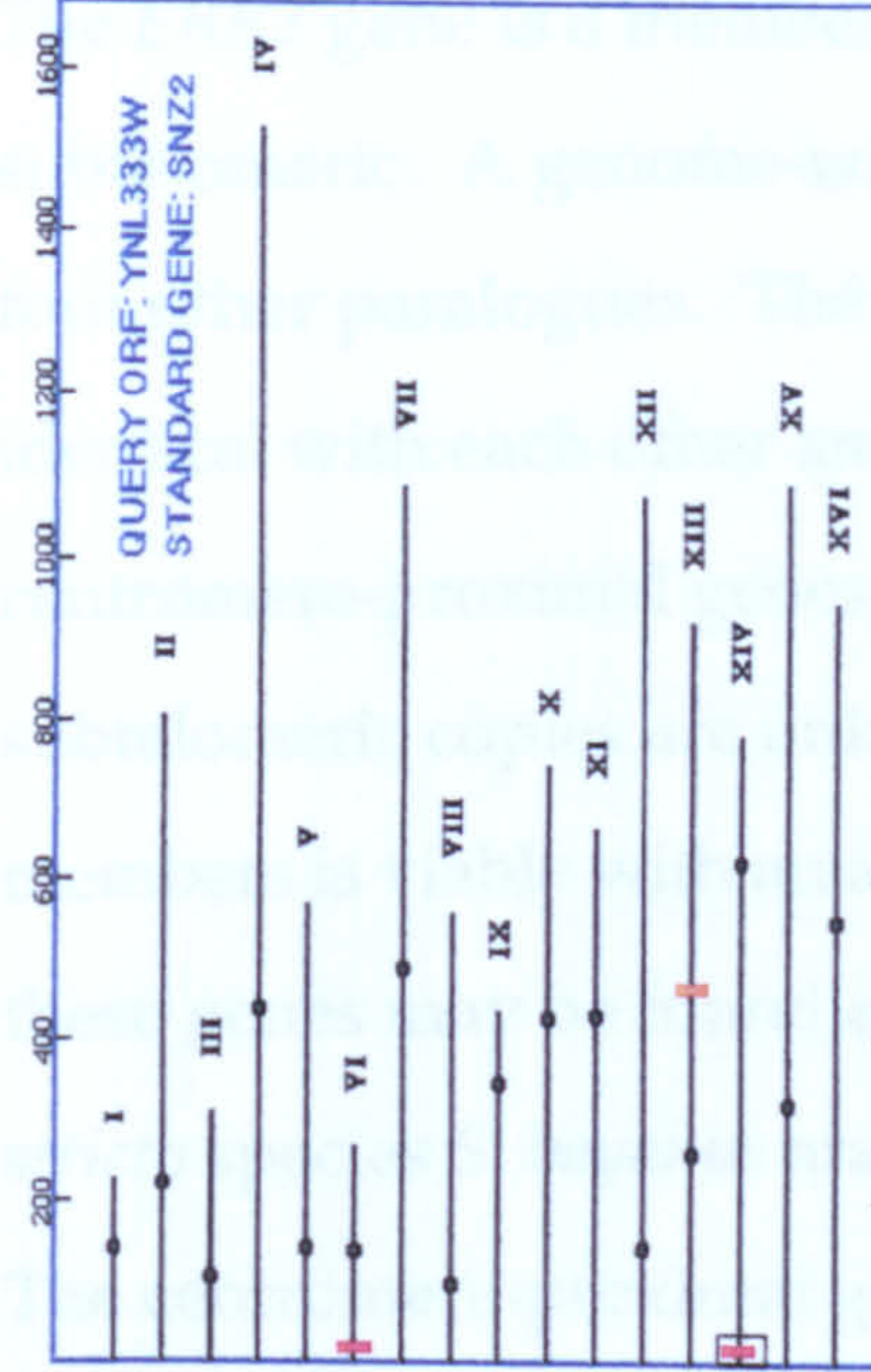
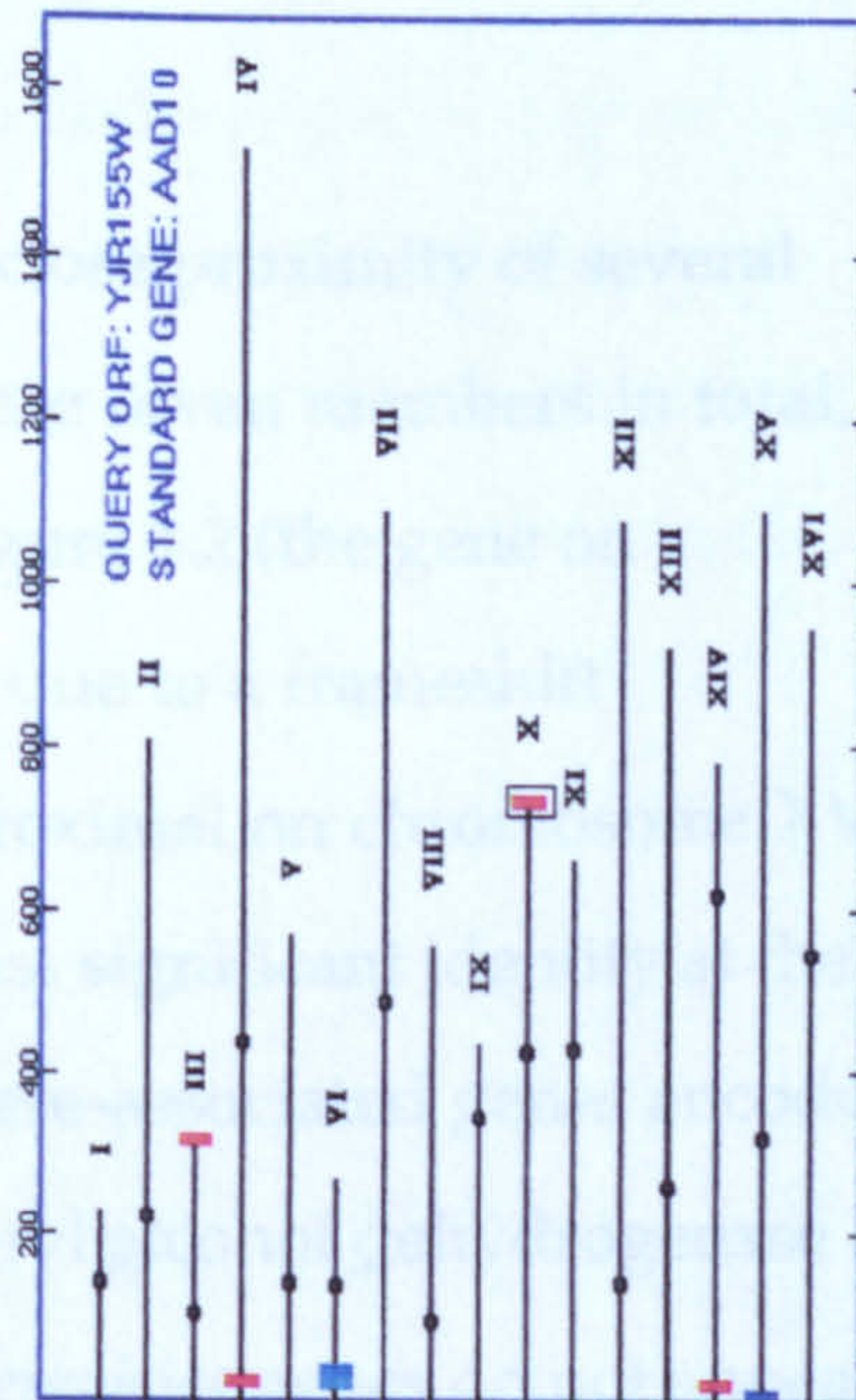
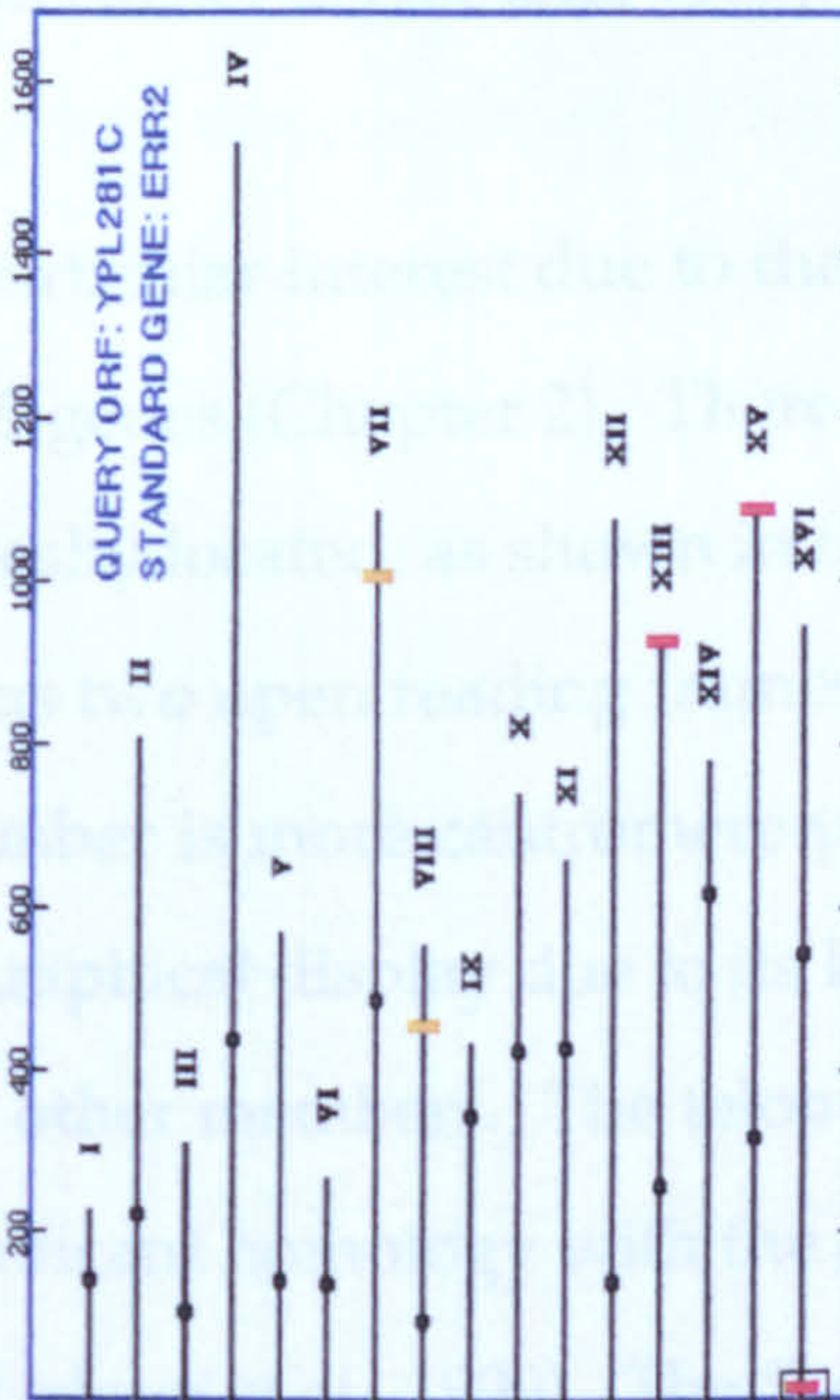
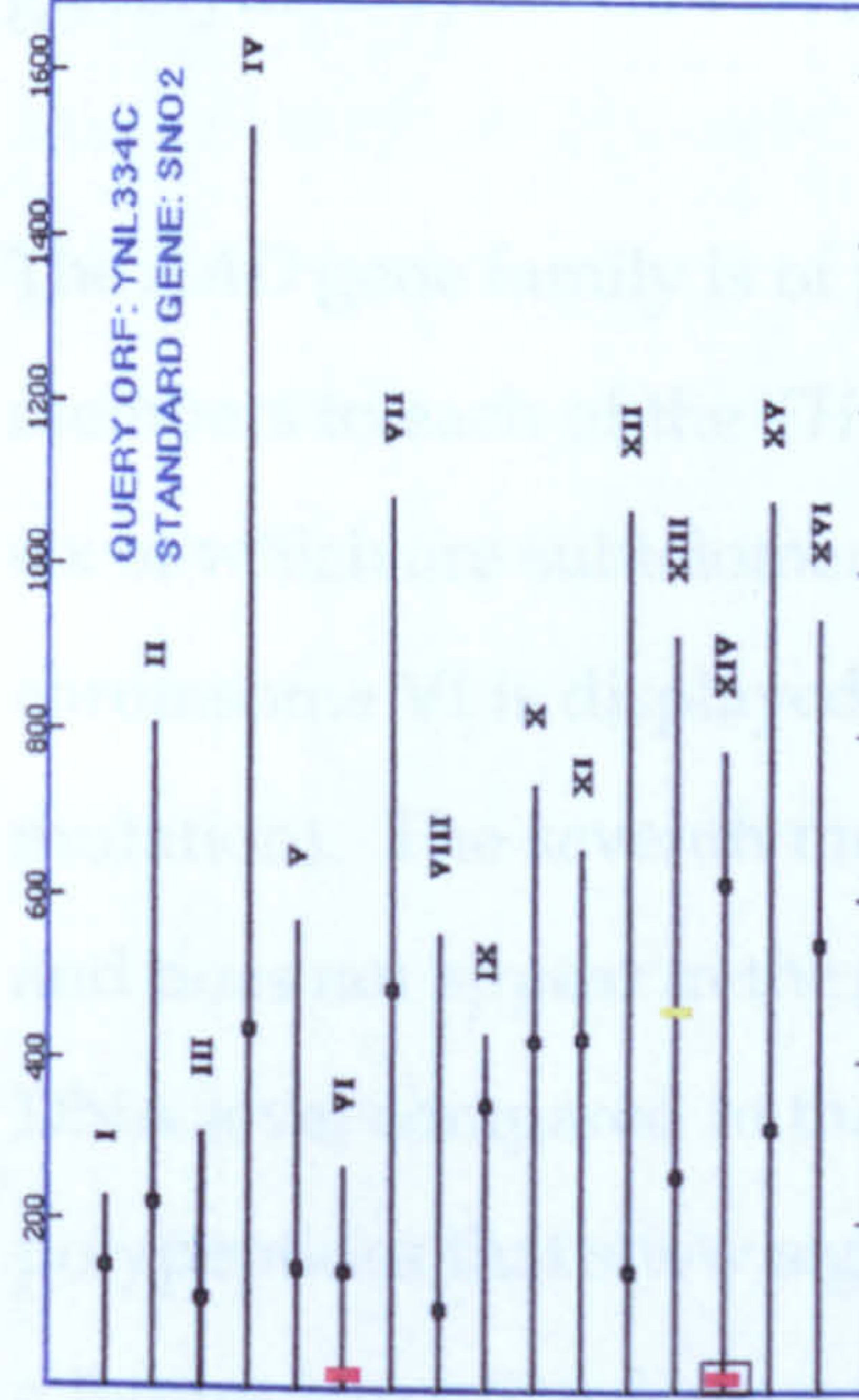
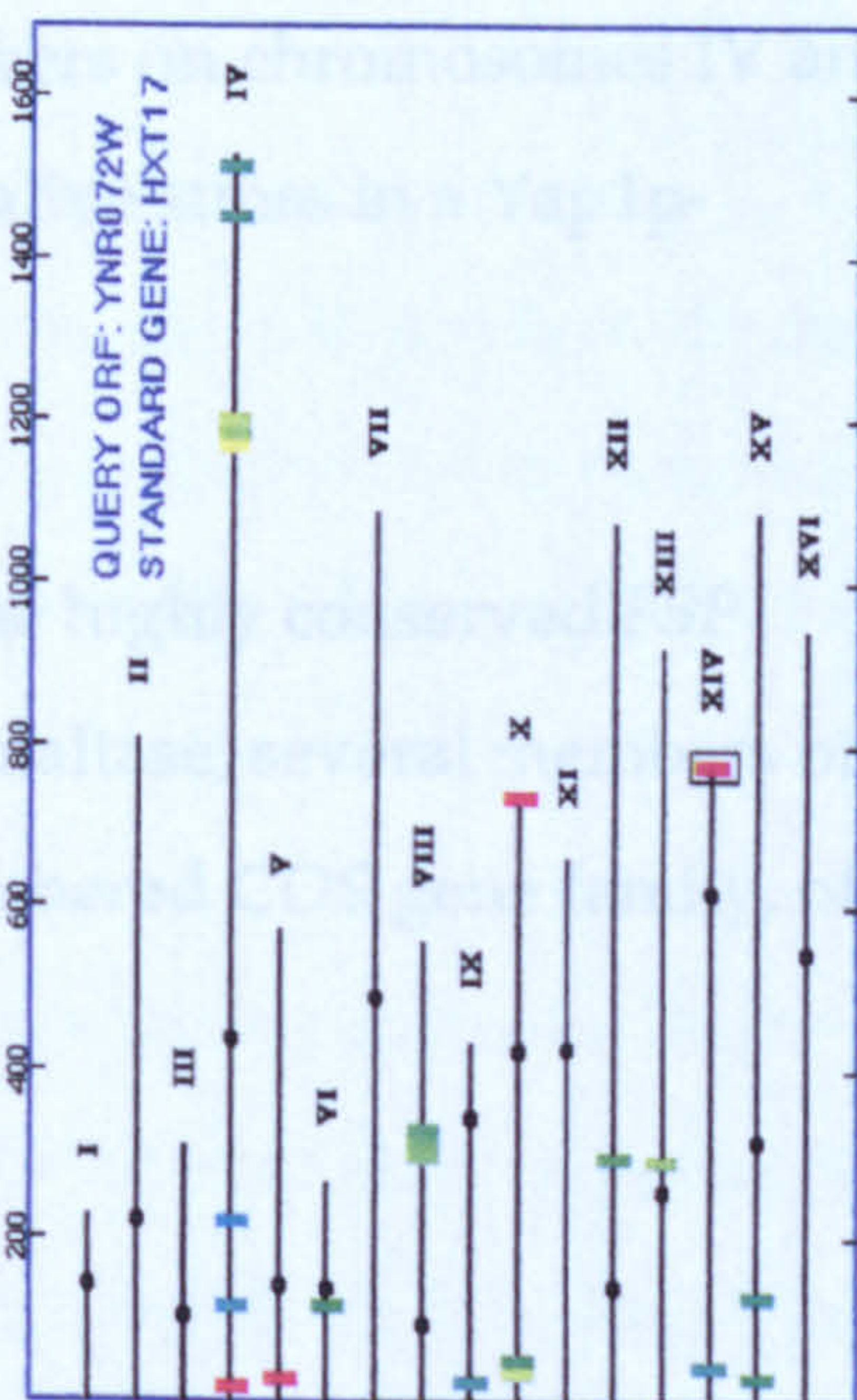
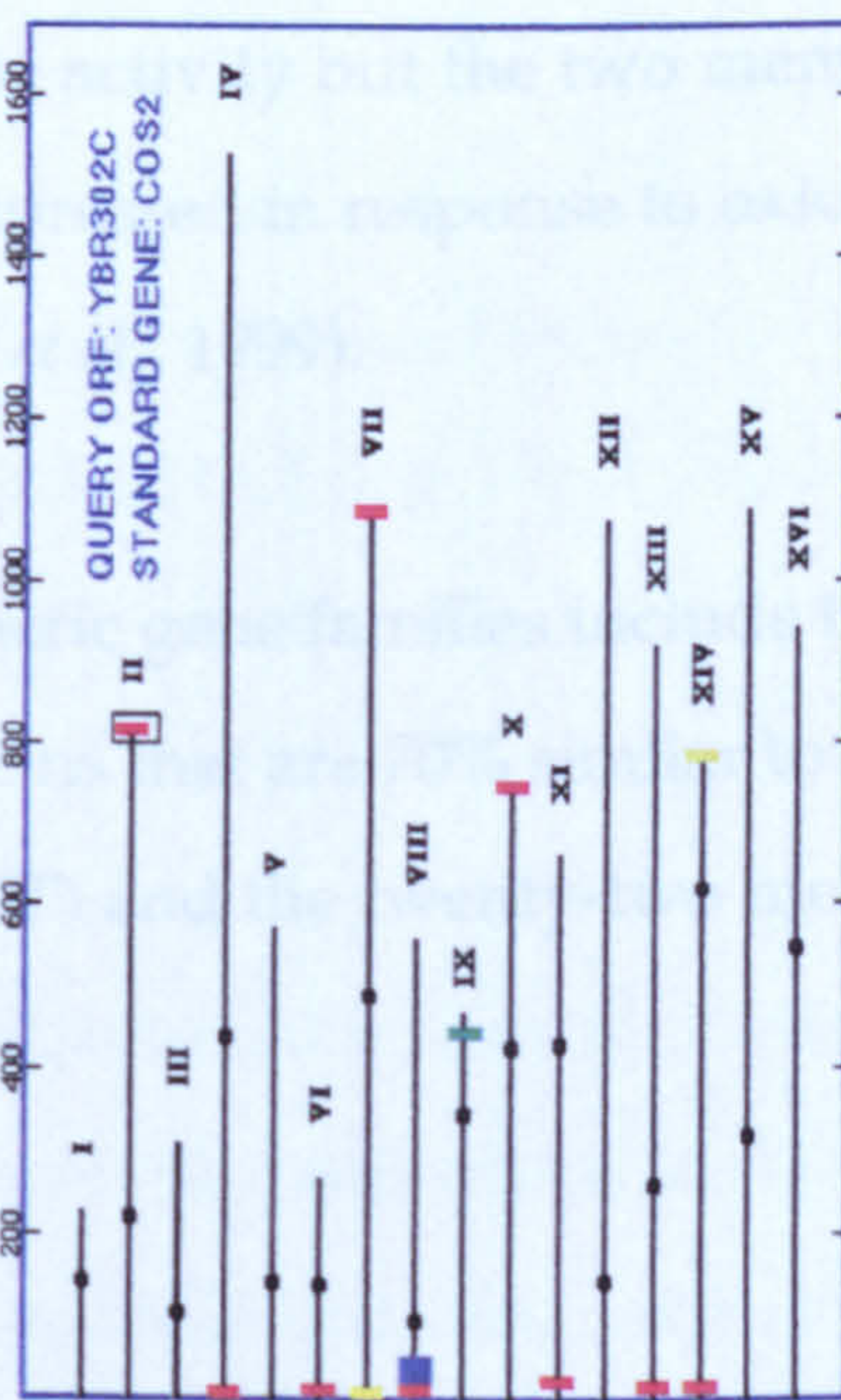
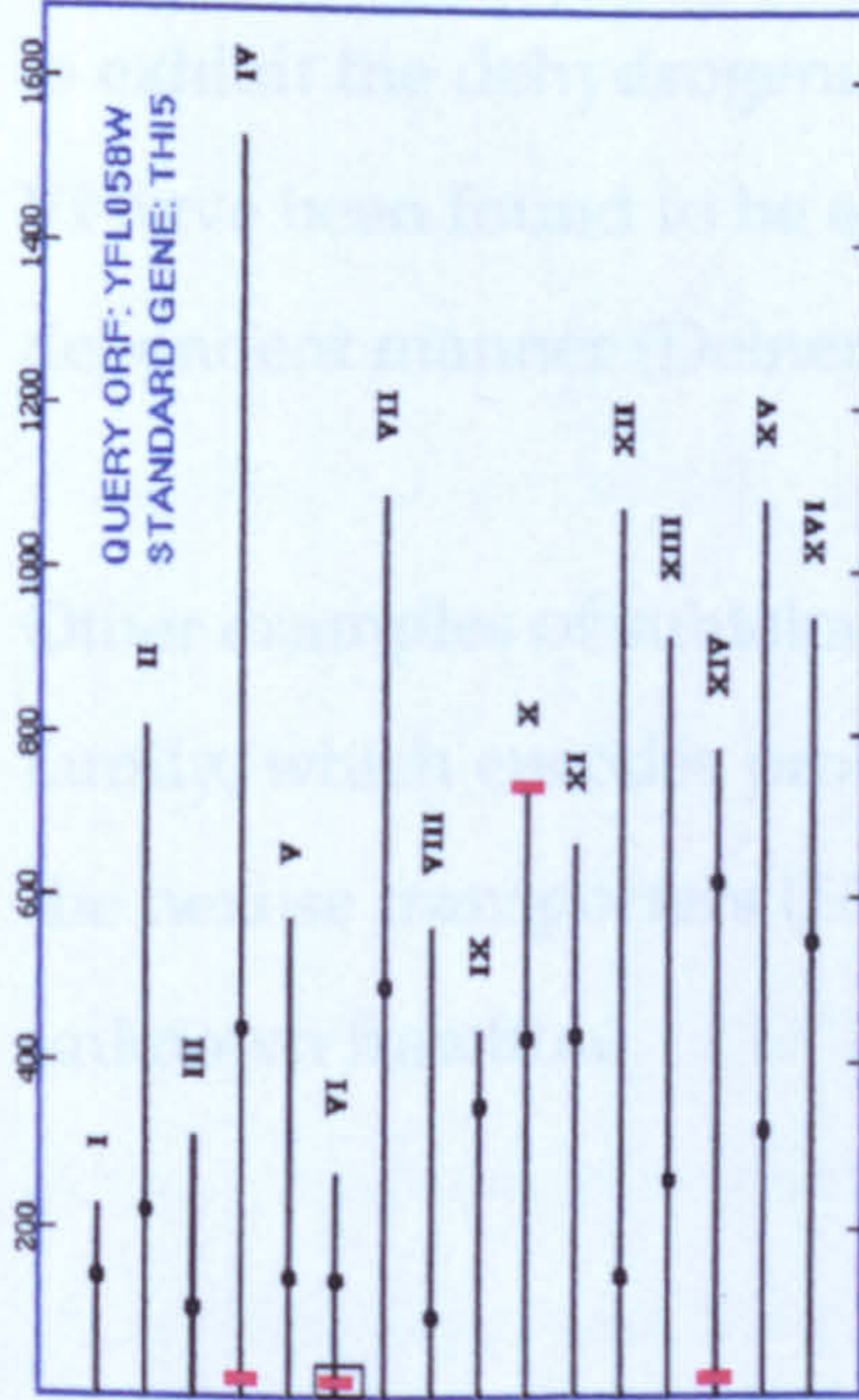
There is a striking observation when comparing the subtelomeric members of such families, in that they are highly conserved. The identity between these genes is often considerably higher than with their more centromere-proximal counterparts. This suggests that a great deal of ectopic recombination may have occurred over evolutionary time. These events appear to have served two purposes; firstly, as a route to gene amplification at the chromosome ends and, secondly, to maintain a high level of sequence identity between these amplified regions. The idea of frequent exchange within the subtelomeres has been put forward by Louis, however, experimental observations show that, at least in meiosis, there is a notably reduced level of homologous recombination in these regions compared with the rest of the genome (Louis, 1995) .

Figure 1.2 shows a graphical representation of a selection of gene families where all or most members are subtelomerically located. The individual genes are colour-coded as a function of their identity to each other at the DNA level; those that exhibit identity of greater than 75% are coloured red. Such genes are only generally found at the subtelomeres, although not all subtelomeric genes exhibit this high degree of conservation. The largest of the *S. cerevisiae* gene families, *PAU*, comprises highly conserved members that are distributed at the subtelomeres and elsewhere in the genome. A thorough functional analysis of this family has not yet been carried out but *PAU* mRNA is transiently expressed at the beginning of stationary phase during wine-making alcoholic fermentation on synthetic 'must' media. A characteristic of this type of fermentation is strong anaerobiosis. The promoters of several members of the gene family have been found to induce expression of a reporter gene under this condition on various media, with maximal activity at the end of the exponential growth phase (Rachidi *et al.*, 2000).

Figure 1.2 A genome-wide DNA similarity view showing the position of members of gene families found at the subtelomeres of the sequenced genome of *S. cerevisiae* S288Ca. Shown here are a selection of gene families which have all or several members at subtelomeric locations. The displays were created using an ORF-DNA alignment tool with graphic output on the Saccharomyces Genome Database (URL: <http://genome-www.stanford.edu/cgi-bin/SGD/SWA/swaEntryForm.pl>). The display parameters were set as follows (default unless otherwise stated); colour criterion max 75% identity and min 30% identity, min 35% aligned. In each case the query open reading frame is a subtelomeric gene. Bars represent the location of similar genes and are coloured according to the degree of identity with the query open reading frame as follows;



Identity to query	>75%	70%	60%	50%	40%	>35%
-------------------	------	-----	-----	-----	-----	------



The *ERR2* gene is a member of another gene family that is predominantly subtelomeric. A genome-wide search using *ERR2* as the query sequence identifies four other paralogues. The three subtelomeric copies (red) are more than 99% identical with each other and approximately 70% identical with the more centromere-proximal genes, at the DNA level (figure 1.2). The role of the subtelomeric copies are unknown and a strain containing disruptions of all three members is viable with an as yet unidentified phenotype. In addition, it appears that these genes may be found exclusively in *S. cerevisiae* as the closely related *sensu stricto* species *S. bayanus* and *S. paradoxus* possess no homologues (Pryde *et al.*, 1995). The centromere-proximal genes are *ENO1*, and *ENO2* which encode the key glycolytic enzyme enolase (McAlister & Holland, 1982).

The *AAD* gene family is of particular interest due to the close proximity of several members to each of the *THI5* genes (Chapter 2). There are seven members in total, six of which are subtelomerically located, as shown in figure 1.2 (the gene on chromosome VI is displayed as two open reading frames due to a frameshift mutation). The seventh member is more centromere-proximal on chromosome XVI and does not appear in the graphical display due to its less significant identity at the DNA level compared to the other members. The telomere-associated genes encode polypeptides that show significant homology with the aryl alcohol dehydrogenase of a lignin-degrading fungus (Delneri *et al.*, 1999). The *S. cerevisiae* genes do not appear to exhibit the dehydrogenase activity but the two members on chromosomes IV and VI have been found to be expressed in response to oxidative stress in a Yap1p-dependent manner (Delneri *et al.*, 1999).

Other examples of subtelomeric gene families include the highly conserved *FSP* family, which encodes proteins that are 70% similar to maltase, several members of the hexose transporters (*HXT*) and the twenty-two membered *COS* gene family, of unknown function.

The *THI5* gene family is the focus of this study and, based on the protein sequence similarities to a *S. pombe* homologue, is thought to be involved in the biosynthesis of thiamin (vitamin B1). As shown in figure 1.2, there are four members which are all subtelomeric and highly conserved at the DNA level. This gene family and its role in thiamin biosynthesis will be discussed in detail in the next sections along with the functionally related *SNO* and *SNZ* gene families.

1.3 The cofactor thiamin diphosphate.

The active form of thiamin, thiamin diphosphate (ThDP), is a cofactor for a number of key metabolic enzymes (Table 1.1). It is composed of two moieties, a pyrimidine and a thiazolium (thiazole), which are connected by a methylene bridge (figure 1.3a). ThDP serves as coenzyme for enzymatic systems that involve cleavage of carbon-carbon bonds. More accurately, ThDP attacks the carbonyl group of mainly α -keto acids or ketoses so liberating carbon dioxide or an aldose, respectively. This results in a ThDP-bound intermediate called an 'active aldehyde' ('active glycoaldehyde' when a ketose is the substrate). For an α -keto acid the general formula for the reaction, called a decarboxylation, is:



(Louloudi & Hadjiliadis, 1994).

The 'active' intermediate is bound to ThDP and, depending on the attached apoenzyme, then reacts with one of several species to form the end product. In the case of pyruvate dehydrogenase (PDH), this intermediate is formed from pyruvate and is transferred to coenzyme A (CoA) which yields Acetyl CoA for entry into the TCA cycle.

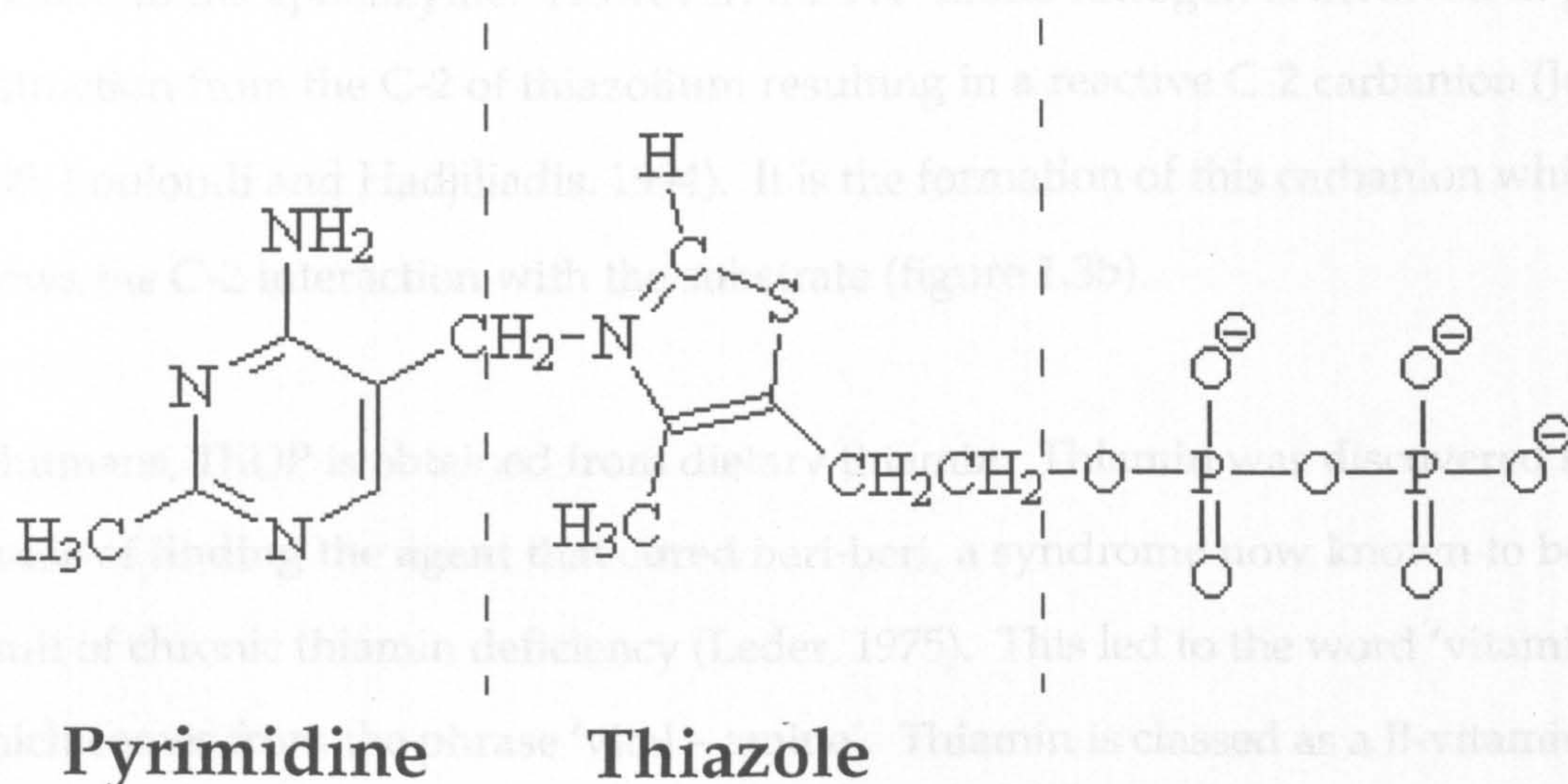
It is the C-2 carbon of the thiazolium which provides the initial interaction with the carbonyl group of the keto-molecule and which binds the intermediate. Figure 1.3b shows the important atoms of ThDP and their respective roles in the function of the

Gene	Enzyme	Metabolic role	Catalytic step
<i>PDC1</i>	Pyruvate decarboxylase	Pyruvate catabolism*	Pyruvate ⇒ Acetaldehyde
<i>PDC5</i>	Pyruvate decarboxylase	Pyruvate catabolism*	
<i>PDC6</i>	Pyruvate decarboxylase	Unknown*	Unknown
<i>ILV2</i>	Acetolactate synthase	Biosynthesis of isoleucine and valine.	Pyruvate ⇌ 2-Acetolactate
<i>THI3</i>	α-Ketoisocaproate decarboxylase	Leucine catabolism*	α-Ketoisocaproate ⇒ Isoamyl alcohol
<i>YDR380w</i>	35% identical to Pdc1p	Unknown*	Unknown
<i>YEL020c</i>	23% identical to Pdc1p	Unknown	Unknown
<i>TKL1</i>	Transketolase	Pentose phosphate interconversions	G-3-P + C6/C7-P ⇌ C5-P + C4/C5-P
<i>TKL2</i>	Transketolase	Pentose phosphate inter.	
<i>PDA1</i>	Pyruvate dehydrogenase E1α	Oxidative pyruvate metabolism	Pyruvate ⇒ Acetyl CoA
<i>KGD1</i>	α-Ketoglutarate dehydrogenase	TCA cycle	α-Ketoglutarate ⇒ Succinyl CoA

Table 1.1 List of thiamin-diphosphate dependent enzymes in *S. cerevisiae*. Also shown are the known metabolic steps catalysed by the enzyme. The transketolase reactions involve the interconversion of polysaccharide phosphates with either 4, 5, 6 or 7 carbon atoms. This Table was adapted from Hohmann and Meacock (1998).

* These enzymes have recently been found to have an additional activity in the catabolism of isoleucine to 2-methylbutanol. This is the only function assigned to the gene product of *YDR380w* so far (Dickinson *et al.* 2000).

a.)



b.)

4'-imino group in close proximity to the C-2 of thiazole resulting in proton abstraction (grey line)

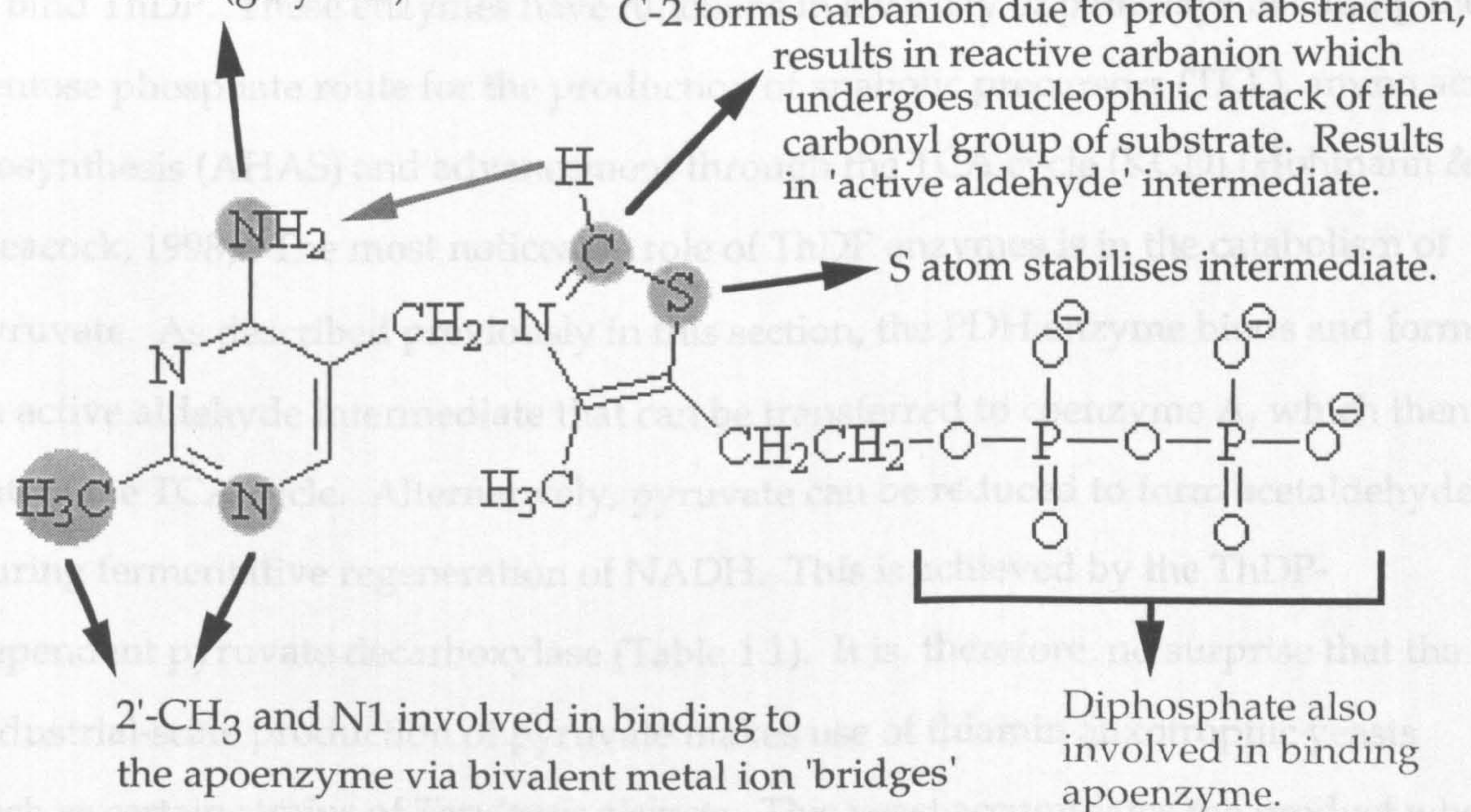


Figure 1.3 Diagram showing a.) the structure of thiamin diphosphate and b.) important atoms and their role in catalysis or protein binding.

Information compiled from Louloudi and Hadjiliadis (1994), Jordan (2000).

cofactor. The pyrimidine part of the molecule mostly has a role in the tight binding of ThDP to the apoenzyme. However, the N4'-imino nitrogen is involved in proton abstraction from the C-2 of thiazolium resulting in a reactive C-2 carbanion (Jordan, 2000; Louloudi and Hadjiliadis, 1994). It is the formation of this carbanion which allows the C-2 interaction with the substrate (figure 1.3b).

In humans, ThDP is obtained from dietary thiamin. Thiamin was discovered in the course of finding the agent that cured beri-beri, a syndrome now known to be a result of chronic thiamin deficiency (Leder, 1975). This led to the word 'vitamin' which comes from the phrase 'vital – amine'. Thiamin is classed as a B-vitamin and has been given the designation B1. Other B-vitamins include riboflavin (B2), niacin (B3) and pyridoxine (B6).

As mentioned earlier in this section, ThDP-containing enzymes have central roles in metabolism. The vast majority of these functions are shared among living organisms from bacteria to humans. Table 1.1 lists the yeast proteins that are known to bind ThDP. These enzymes have functions in a variety of pathways including the pentose phosphate route for the production of anabolic precursors (TKL), amino acid biosynthesis (AHAS) and advancement through the TCA cycle (KGD) (Hohmann & Meacock, 1998). The most noticeable role of ThDP enzymes is in the catabolism of pyruvate. As described previously in this section, the PDH enzyme binds and forms an active aldehyde intermediate that can be transferred to coenzyme A, which then enters the TCA cycle. Alternatively, pyruvate can be reduced to form acetaldehyde during fermentative regeneration of NADH. This is achieved by the ThDP-dependent pyruvate decarboxylase (Table 1.1). It is, therefore, no surprise that the industrial-scale production of pyruvate makes use of thiamin auxotrophic yeasts such as certain strains of *Torulopsis glabrata*. This yeast accumulates the product when supplied with the appropriate limiting quantities of the vitamin (Hua *et al.*, 2001).

1.4 Thiamin diphosphate biosynthesis.

Unlike higher eukaryotes, many of the bacteria, lower eukaryotes and plants have the ability to synthesise ThDP *de novo*. This section will present the current understanding of the metabolic pathways responsible, with a particular focus on the scheme used by *S. cerevisiae*.

1.4.1 Fungal vs. bacterial thiamin biosynthesis.

There are a small number of key organisms where the chemistry and genetics of ThDP biosynthesis has been extensively studied. Within the bacteria, these are *Escherichia coli*, *Salmonella typhimurium* and, to a lesser extent, *Bacillus subtilis*. Studies within the fungal kingdom have almost exclusively focused on the yeasts *S. cerevisiae* and *S. pombe*. Each of the biosynthetic pathways of these organisms has been reviewed and compared by Begley (1996). All the organisms are known to form the cofactor through separate synthesis of both the thiazole and pyrimidine moieties, which are then brought together in a condensation reaction. The pyrimidine used in this reaction is 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate, often shortened to hydroxymethylpyrimidine-PP (HMP-PP). The thiazole takes the form of 4-methyl-5(β -hydroxyethyl)thiazole phosphate, shortened to hydroxyethylthiazole-P (HET-P). The structures of both these precursors are shown in figure 1.4 along with the reaction scheme (described by Begley (1996)). The product of the condensation of HMP-PP and HET-P is thiamin phosphate.

Similarities between the bacteria and yeast end here. Substantial differences are seen in the biosynthesis of the precursors and the formation of the ThDP from thiamin phosphate. Bacteria synthesise the HET-P precursor from tyrosine (glycine in *B. subtilis*), deoxy-D-xylulose and cysteine. Bacterial HMP-PP is formed from 5-aminoimidazole ribotide (AIR) which is itself an intermediate of *de novo* purine nucleotide biosynthesis (reviewed in Begley *et al.*, 1999). Further studies in *S. typhimurium* show that this organism can still make HMP-PP in a mutant defective in

purine biosynthesis, and much work has since been performed to characterise this alternative pyrimidine pathway (Downs & Roth, 1991).

The origin of the precursors in *S. cerevisiae* has been reviewed by Begley (Begley, 1996). HET-P is formed from a 2-pentulose-5-phosphate in addition to glycine and cysteine. HMP-PP is synthesised from glutamine and 5-phosphoribosyl-1-amine via pyridoxine (described in detail in section 1.5). Apart from the broad similarities between kinases, the enzymes involved in precursor formation are very different between bacteria and yeast.

Further evidence of the differences between *E. coli* and *S. cerevisiae* thiamin biosynthetic pathways has been supplied by NMR spectroscopy studies (Himmeldirk *et al.*, 1998). These experiments directly compared the incorporation of various ¹³C-labelled substrates such as glucose and glycine into the thiamin molecule and found substantial dissimilarities in the patterns of incorporation.

1.4.2 Synthesis of active cofactor from thiazole and pyrimidine precursors in *S. cerevisiae*.

As outlined above, both the HET-P and HMP-PP precursors are brought together in a condensation reaction to yield thiamin phosphate (figure 1.4). The enzyme responsible for this step of the pathway is thiamin phosphate pyrophosphorylase (TMP-PP) which was isolated through characterisation of a mutant yeast strain resistant to the antimetabolite 2-aminohydroxyethylthiazole. It was found that this strain lacked both TMP-PP and HET-kinase enzyme activities which were subsequently found to copurify (Kawasaki, 1993). A clone was obtained containing a gene that rescued the mutant phenotype. The gene, *THI6*, encodes a bifunctional enzyme where the two activities are located in different domains of the same protein. Disruption of *THI6* results in thiamin auxotrophy (Nosaka *et al.*, 1994).

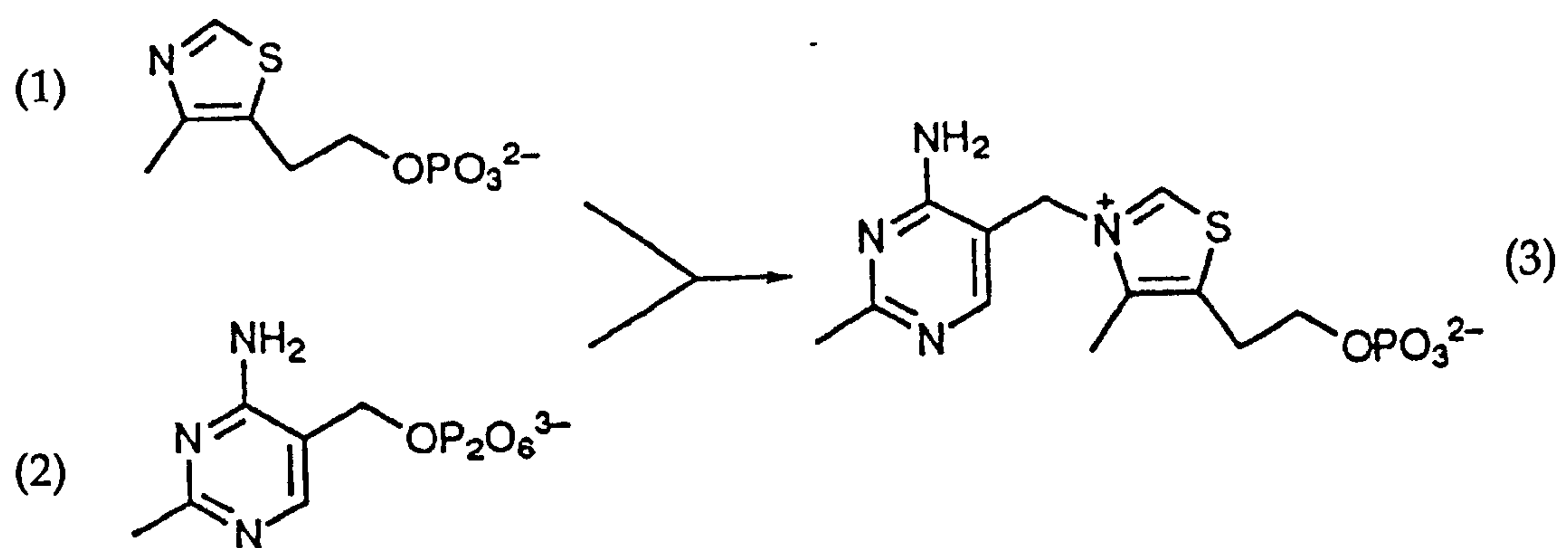


Figure 1.4 Biosynthesis of thiamin phosphate (3) by the condensation of 4-methyl-5(β-hydroxyethyl)thiazole phosphate (1) and 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate (2). Taken from Begley, Nat. Prod. Rep. 13 (1996) 177-185.

The steps involved in the conversion of thiamin phosphate to the active cofactor, ThDP, have been elucidated based on early experiments by Kaziro who purified a protein which enabled the *in vitro* synthesis of the active cofactor from thiamin. He called this protein thiaminokinase. The substrates for this enzyme included thiamin and ATP but not thiamin phosphate or ADP (Kaziro, 1959). This suggested that after the condensation of the HET-P and HMP-PP precursors, thiamin phosphate must be dephosphorylated to free thiamin before being acted upon by thiaminokinase. This enzyme, now known as thiamin pyrophosphokinase, has since been found to be encoded by the *THI80* gene. Disruption of *THI80* is lethal regardless of the presence of exogenous thiamin (Nosaka *et al.*, 1993). To date, no enzyme has been purified which dephosphorylates intracellular thiamin phosphate. The positions of both Thi6p and Thi80p in the overall scheme of thiamin biosynthesis are shown in figure 1.5.

1.4.3 Synthesis of the thiazole precursor.

The identification of substrates for synthesis of the thiazole precursor, HET-P, has been achieved through radiotracer experiments using mainly ¹⁴C-labelled compounds. These substrates are glycine (White & Spenser, 1979), a 2-pentulose-5-phosphate from the pentose phosphate pathway (White & Spenser, 1982) and cysteine, which provides the sulphur atom (Tazuya *et al.*, 1987).

Apart from one of the enzyme activities of Thi6p, only one other enzyme has been found to be involved in HET-P formation. This protein is encoded by the *THI4* gene (figure 1.5). *THI4* was originally isolated on a cDNA clone obtained through a screen for genes expressed during entry into stationary phase, when *S. cerevisiae* is grown on industrial molasses medium (Praekelt & Meacock, 1992). Several genes were identified, one of which was named *MOL1* (MOLasses-inducible). *MOL1* was later found to be expressed as a result of the exogenous depletion of thiamin and was renamed *THI4*. Disruption of *THI4* results in thiamine auxotrophy which is rescuable by the supplementation of HET, but not HMP, in the growth medium (Praekelt *et al.*,

1994). Curiously, the *THI4* gene product has been demonstrated to have another role in mitochondrial DNA damage tolerance. The *thi4* disrupted strain exhibits a far higher than average frequency of respiratory-deficient mutants when treated with the DNA damaging agents methyl methanesulfonate and ultra violet-light (Machado *et al.*, 1997).

An attempt to identify other genes encoding thiazole biosynthetic enzymes has been unsuccessful. In a screen to search for thiazole auxotrophic mutants, only *thi4* mutations were isolated and assigned to this pathway (Byrne, 1998).

1.4.4 Uptake of thiamin and the precursors.

Despite the ability of yeast to synthesise thiamin *de novo*, the active cofactor is preferentially formed from an exogenous source, if present, in the growth medium. This uptake results in a significant increase in the growth rate of yeast, which can dispense with the energy-costly route of biosynthesis (Suomalainen & Oura, 1971).

Studies using ^{14}C -labelled thiamin have been carried out in order to investigate the transport of this vitamin in *S. cerevisiae* (Iwashima *et al.*, 1973). The labelled thiamin is seen to be taken up from the external medium in a linear fashion for the first twenty minutes after addition, reaching a maximum intracellular concentration of 10 nmoles/mg dry weight cells at thirty minutes and 37°C. This intracellular thiamin concentration is calculated to be 10,000 times the external concentration and no uptake is seen at 0°C. In addition, yeast exhibits significantly less internalisation of ^{14}C -thiamin in the absence of glucose or in the presence of metabolic inhibitors like cyanide. Taken together, the experimental evidence shows that thiamin transport is active. In addition, another study has compared uptake of ^{14}C -thiamin, ^{14}C -ThMP and ^{14}C -ThDP. The results indicate that only free, unphosphorylated thiamin, crosses the plasma membrane (Nishimura *et al.*, 1982). Thiamin phosphates must therefore be dephosphorylated prior to transport across the cell membrane. This is achieved by a periplasmic thiamin-repressible acid phosphatase encoded by the *PHO3* gene

(Nosaka, 1990). A *S. cerevisiae* strain containing a mutation in *PHO3*, called *pho3-81*, exhibits normal uptake activity of thiamin but considerably less uptake of phosphorylated forms compared to wild type (Nosaka *et al.*, 1989).

Transport of the thiamin precursors has also been investigated. Using ³H-labelled hydroxymethylpyrimidine, it has been found that linear uptake of HMP occurs for the first minute after addition to a maximum of 1.5 nmol/mg dry weight cells after five minutes. This corresponds to an intracellular HMP concentration of 610 times that of the external concentration. This transport is metabolically and temperature dependent and so, like thiamin, uptake is by an active process. HMP transport into the cell uses a common system to thiamin since HMP uptake is competitively inhibited by the free vitamin. In addition, *S. cerevisiae* mutants that have been shown to be defective in thiamin uptake are also defective in their HMP uptake (Iwashima *et al.*, 1990a).

In contrast to thiamin and HMP, thiazole is not concentrated within the cell and is instead acquired by passive diffusion across the membrane. The thiazole accumulates only in its phosphorylated form, HET-P, due to HET kinase activity (Iwashima *et al.*, 1986).

The evidence of an active transport system has prompted the identification of genes encoding thiamin transporters. The cloning of one such gene used an ingenious approach by Singleton (1997). His strategy relied on the observation that a thiamin analog, called pyrithiamin, is lethal to yeast. A collection of *S. cerevisiae* strains that are resistant to pyrithiamin were isolated and were found to be severely deficient in thiamin transport. These strains therefore rely on the *de novo* biosynthesis of the cofactor. A *thi4* disruption created in this laboratory, along with a plasmid-borne *THI4* wild type allele, was introduced into the existing transport mutants thus making survival dependent upon biosynthesis using the extrachromosomal *THI4*. Singleton then concurrently transformed a genomic library into these new strains whilst selectively removing the plasmid-borne *THI4*. Only those transformants that

contained a clone conferring thiamin transport survived. These same transformants were found to be restored for pyrithiamin sensitivity. The rescuing clones were subsequently found to contain an open reading frame, corresponding to YLR237w, that has been named *THI7*. A *THI7* disruption results in pyrithiamin resistance and a *thi7, thi4* double mutant results in lethality since the cell has no other route to acquire or produce thiamin.

The same thiamin transporter has been identified by another group using a similar approach to Singleton. Enjo *et al.* (Enjo *et al.*, 1997) isolated a genomic clone that rescued the thiamin transport defect of a pyrithiamin-resistant mutant. The gene responsible for the restoration of transport activity was named *THI10*, and is identical to *THI7*.

Since HMP and thiamin uptake have been shown to involve a common system, it is therefore very likely that *THI7* encodes the common transporter. Thi7p-dependent HMP uptake has not been determined experimentally. Uptake by Thi7p has been included in the overall scheme of figure 1.5.

1.4.5 Regulation of the biosynthetic enzymes.

The uptake of thiamin by *S. cerevisiae* does not occur for phosphorylated forms of the vitamin. As described in the previous section, an external source of the active cofactor (ThDP) is dephosphorylated in the periplasmic space prior to transport across the plasma membrane. Biosynthesis of ThDP from the HMP-PP and HET-P precursors proceeds via unphosphorylated thiamin as an intermediate (section 1.4.2). Therefore the final pyrophosphorylation step yielding the active cofactor is essential regardless of the thiamin availability outside of the cell. This step is catalysed by the product of the *THI80* gene (figure 1.5). Although a strain disrupted for *THI80* is lethal, a mutant strain, called *thi80-1*, has been isolated which has significantly lower Thi80p enzyme activity. A *thi80-1* strain, when grown in medium containing thiamin, accumulates much more unphosphorylated thiamin than wild type but

accumulates much less of the phosphorylated form, ThDP. This is consistent with a lower pyrophosphokinase activity in the mutant strain. Along with the decrease in ThDP levels, the *thi80-1* mutant shows constitutive activity of biosynthetic enzymes and transport proteins in the presence of exogenous thiamin. Under the same conditions, wild type yeast are observed to repress such activities. This shows that thiamin biosynthesis and transport in *S. cerevisiae* is negatively controlled by the intracellular ThDP levels (Nishimura *et al.*, 1991). Another mutant strain, called *thi81*, exhibits similar properties to *thi80-1*. Further characterisation of *thi81* shows that repression of enzyme activity by ThDP occurs at the level of gene expression (Nishimura *et al.*, 1997). Based on the experimental evidence, the proteins affected by both the *thi80-1* and *thi81* mutations represent negative regulators of thiamin-dependent gene expression.

Thiamin regulation at the level of gene expression has been examined in much more detail by Burrows (1997). Burrows examined enzyme activity of the *lacZ* reporter gene product when fused to the *THI4*, *THI5* or *THI12* promoter. As described earlier, *THI4* encodes an enzyme involved in HET-P biosynthesis whilst *THI5* and *THI12* are members of a larger gene family possibly involved in HMP-PP formation and are described further in section 1.5.

The *THI4:lacZ* reporter promoter fusion, when introduced into wild type *S. cerevisiae*, was found to be expressed at high levels in the absence of exogenous thiamin and fully repressed at concentrations of 0.74 μ M or higher (Praekelt, *et al.*, 1994).

Burrows found that this repression was also seen for the constructs containing the *THI5* and *THI12* promoters. When introduced into the *thi80-1* mutant strain, all three constructs conferred appreciable β -galactosidase activity in the presence of thiamine, which correlated to derepression of the promoters. This reinforces the evidence that gene repression occurs by ThDP and not the unphosphorylated form. In addition, Burrows examined the regulation of these constructs in response to the precursors HMP and HET. He found that expression from all three promoters was completely repressed in the presence of 1.5 μ M thiamin in the growth medium, however,

repression of the *THI4* promoter construct was not seen in thiamin-deficient media containing either HET or HMP. The *THI5* and *THI12* constructs show partial repression by the precursors. Complete repression, as seen in the presence of thiamin, is observed for *THI4*, *THI5* and *THI12* promoters in the presence of both precursors together. This regulation by the precursors will be examined further in Chapter 6 of this study.

Three positive regulators of thiamin-dependent expression have been identified. *PHO6* was originally identified as a gene encoding a protein that is required for the synthesis of the thiamin-repressible acid phosphatase, Pho3p (Toh-E *et al.*, 1975a). A *pho6* mutant strain is a thiamin auxotroph. This *pho6* strain can grow if supplied with a limited amount of thiamin which does not normally result in repression of biosynthetic activity. There is no activity of the various enzymes involved in thiamin biosynthesis in the mutant, neither is there significant expression of *THI4*, *THI5* and *THI12* (Burrows, 1997; Kawasaki *et al.*, 1990). Thiamin transport, which is also repressed by thiamin, is not governed by *PHO6* (Nishimura *et al.*, 1992a). Due to its role in the regulation of gene expression by thiamin, *PHO6* has been renamed *THI2*. *THI2* encodes a protein containing a classic zinc-finger motif similar to the transcription factor Gal4p, which suggests a role in DNA-binding.

In a screen to find more regulators of thiamin metabolism, a mutant strain of *S. cerevisiae* was isolated which exhibits almost no activity of thiamin biosynthetic enzymes as seen for the *thi2* mutant described above (Nishimura *et al.*, 1992b). However, unlike *thi2*, this novel strain also showed a significantly reduced rate of thiamin transport compared to wild type. A clone that rescues the mutant phenotype was obtained and the gene responsible for the complementation has been named *THI3*. A *thi3* mutant exhibits thiamin auxotrophy and, like the *thi2* strain, displays greatly reduced expression of the *THI4*-, *THI5*- and *THI12-lacZ* constructs (Burrows, 1997; Nishimura, *et al.*, 1992b). Thi3p has a strong similarity to the ThDP-dependent enzyme pyruvate decarboxylase (Pdc). Like Pdc, Thi3p possesses a motif involved in binding of the ThDP cofactor but the role of this

protein is unclear. Unlike Thi2p, Thi3p has no similarity to a transcriptional activator. A hypothesis has been put forward that this protein may sense the intracellular ThDP levels by binding the cofactor and it may exert transcriptional control through interaction with Thi2p (Hohmann and Meacock, 1998). The role of Thi3p has been further complicated by a proposed function as the α -ketoisocaproate decarboxylase in leucine catabolism (Dickinson *et al.*, 1997).

The third positive regulatory gene was isolated by the functional complementation of two thiamin auxotrophic strains of *S. cerevisiae*. These strains, UV2 and UV3, were isolated as thiamin auxotrophs after mutagenesis of a prototrophic strain with ultraviolet light (Byrne, 1998). Both strains contained recessive mutations in the same gene. However, these strains had different precursor requirements. UV2, in the absence of exogenous thiamin, required both the HET and HMP precursors for growth whilst UV3 only required the addition of HMP. Consistent with this observation, in UV2 both the *THI4* (HET biosynthesis) and *THI5* (HMP) genes were not expressed in derepressing (i.e. thiamin limiting) conditions whilst *THI4*, but not *THI5*, showed expression in UV3 (Richards, 1996). The single gene that functionally complemented the thiamin auxotrophy and which restored gene expression of *THI4* and *THI5* was identified as *PDC2*. The molecular basis of the different precursor requirements of UV2 and UV3 await definition. It therefore appears that Pdc2p is a positive regulator of thiamin biosynthetic gene expression. Pdc2p has also been identified as a positive regulator of the *PDC1* and *PDC5* genes which are the structural genes of pyruvate decarboxylase (Hohmann, 1993). Consistent with a putative role in transcriptional activation, a fusion protein containing the DNA-binding domain of Gal4p and the C-terminal portion of Pdc2p is able to activate transcription of a *GAL1-lacZ* reporter construct (Raghuram *et al.*, 1994). In addition, a Pdc2p:GFP fusion protein has been seen to have a specific subcellular localisation that appears to be the nucleus (S. Savill, personal communication). A strain deleted for *PDC2* is a thiamin auxotroph rescuable only with the supplementation of thiamin or both precursors together (Richards, 1996).

1.4.6 *S. cerevisiae* mutants that exhibit derepression in the presence of thiamin.

In an attempt to identify negative regulators of thiamin biosynthetic genes, a genetic approach was taken to isolate mutant strains which exhibit constitutive expression of these genes in the presence of normally repressing levels of thiamin (Burrows *et al.*, 2000). Four mutants, when grown in thiamin supplemented medium, were found to display activity from the *THI4-lacZ* construct described in the previous section. This derepression was also found for other biosynthetic genes including *THI5* and *THI12*. By examining the meiotic segregation patterns of the four mutants, called *DET* (Derepressed Expression on Thiamin), it was found that the mutations define three different genes.

Three of the *det* mutations have been mapped to two genes: the positive regulator *THI3*; and the pyruvate decarboxylase structural gene *PDC1*. The mutations give rise to amino acid substitutions which possibly result in compromised binding of the active cofactor ThDP. As both Thi3p and Pdc1p have enzymatic roles in leucine and pyruvate metabolism respectively, any activity altered by the *det* mutations may be transmitted to the nucleus whereby the thiamine genes are derepressed (Byrne and Meacock, in preparation).

1.5 Biosynthesis of the pyrimidine precursor, HMP-PP.

Presented here is a detailed review of the current knowledge about the formation of the pyrimidine moiety of thiamin. At the commencement of these studies, no genes had been unambiguously assigned a role in HMP-PP biosynthesis. During the course of this investigation, the gene products that confer the kinase activity were identified (Llorente, *et al.*, 1999). In addition, as a result of work performed in this laboratory, several genes encoding enzymes involved in pyridoxine (vitamin B6) biosynthesis have been identified. As described in the following section, pyridoxine is itself a precursor for HMP-PP formation. Further work has led to the discovery of a regulatory link between the thiamin and pyridoxine pathways. This work will be

presented in section 1.6. There is also mounting evidence of an alternative route to the formation of HMP which is discussed in section 1.5.5. This is the subject of a further and more detailed examination as part of this thesis (Chapter 6).

1.5.1 Precursor labelling studies.

Similar to the HET-P biosynthetic pathway, identification of the starting substrates in HMP-PP formation has made use of the data provided by radiotracer experiments. These experiments measure the incorporation of specific radioisotopes into the HMP-PP molecule from a variety of labelled compounds. A scheme representing our current understanding of the origins of the various atoms of HMP is displayed in figure 1.6. Part of the HMP moiety of thiamin is itself derived from another B-vitamin, called pyridoxine (vitamin B1). The earliest evidence showing pyridoxine to be a precursor of HMP formation came from an analysis of the incorporation of $^{15}\text{NH}_4\text{Cl}$ into the pyrimidine ring. The nitrogen isotope (^{15}N) of this molecule was found to be incorporated into the N-1 position of HMP (Figure 1.6). Addition of unlabelled pyridoxine was found to competitively inhibit ^{15}N incorporation at this position (Tazuya *et al.*, 1993). Further experiments were carried out where the pyridoxine molecule was labelled with radioisotopes at the C-6 and 5'-H atoms. The data showed the efficient incorporation of both isotopes into the same positions (C-6 and 5'-H) of the HMP molecule (Tazuya *et al.*, 1995). Tazuya and colleagues also identified the exact part of the HMP molecule that originates from pyridoxine. This was found to be the region consisting of the C-2', C-2, N-1, C-6, C-5, C-5' atoms and is illustrated in figure 1.6. Subsequent experiments have found the N-1 atom of HMP and pyridoxine to come from the amide nitrogen atom of the amino acid glutamine (Tazuya *et al.*, 1995). A hypothesis has been made that a substantial part of the incorporated pyridoxine fragment comes from 5-phosphoribosyl-1-amine (Figure 1.6). This molecule is a very good candidate since it is itself formed using an amidotransferase (Ade4p) from 5-phosphoribosyl-1-pyrophosphate (PRPP) (Mantsala & Zalkin, 1984). Glutamine is the nitrogen donor in this reaction. PRPP is a very important substrate for biosynthesis of both pyrimidine and purine

Figure 1.6 Biosynthesis of the HMP-PP precursor in *S. cerevisiae*.

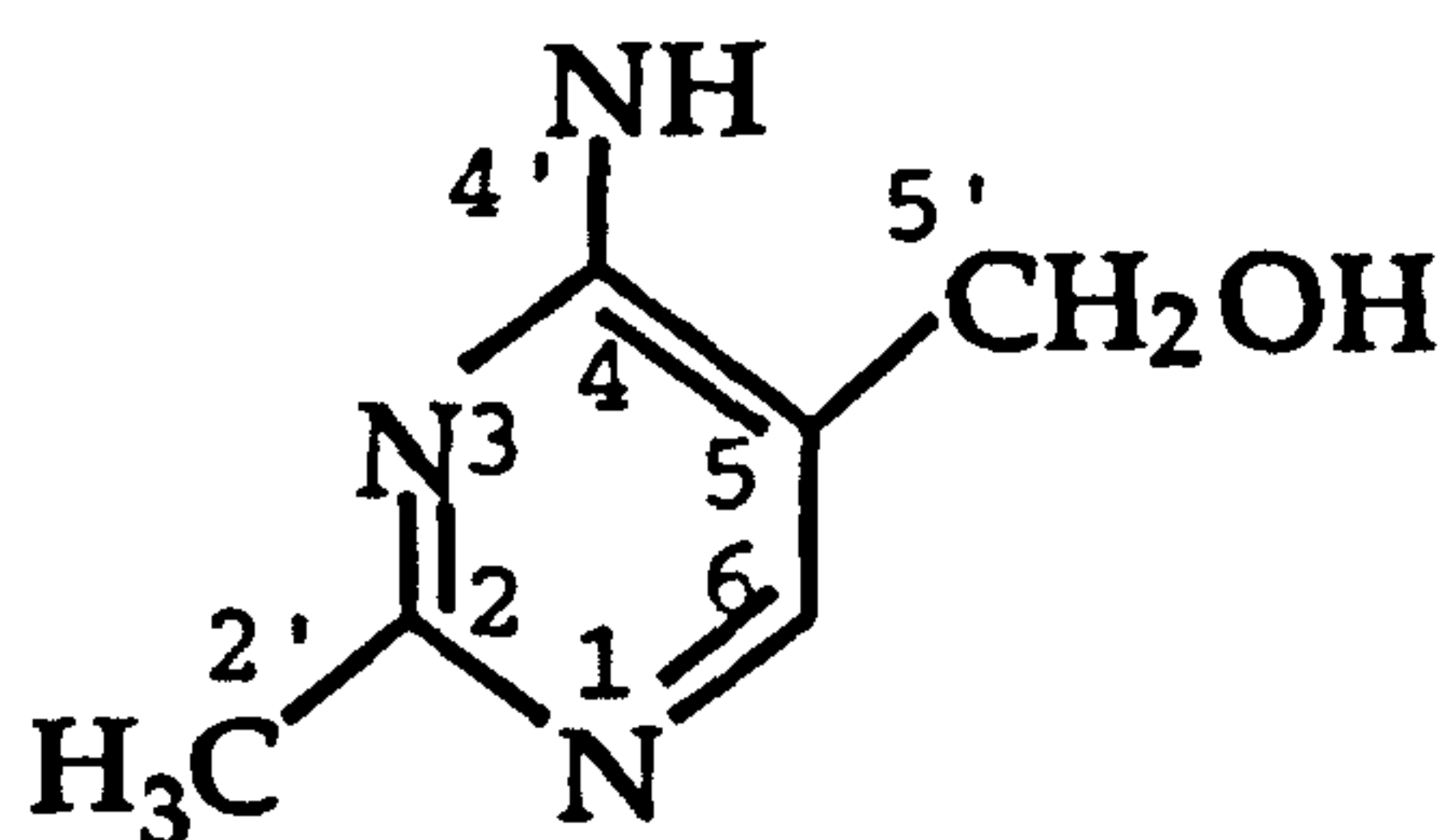
a.) Anabolic pathway leading to the formation of the HMP-PP precursor of thiamin.

Genes whose products are known to be involved at any step are shown. This diagram only displays those intermediates that have been currently identified in HMP-PP formation.

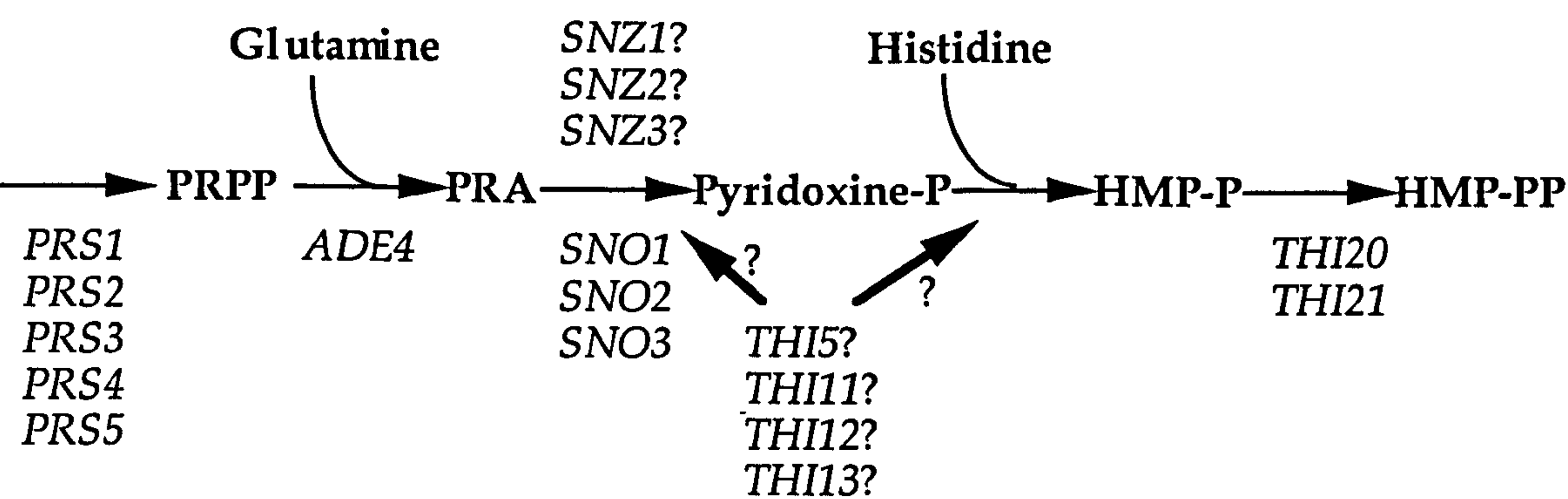
PRPP (5-phosphoribosyl-1-pyrophosphate), PRA (5-phosphoribosyl-1-amine), HMP (2-methyl-4 amino-5-hydroxymethylpyrimidine), P (phosphate), PP(pyro- or di- phosphate).

b.) Scheme showing the likely origins of the atoms of the HMP molecule as taken from Tazuya *et al.* (Biochem. Mol. Biol. Int. 36 (1995) 883-888). Gln (glutamine).

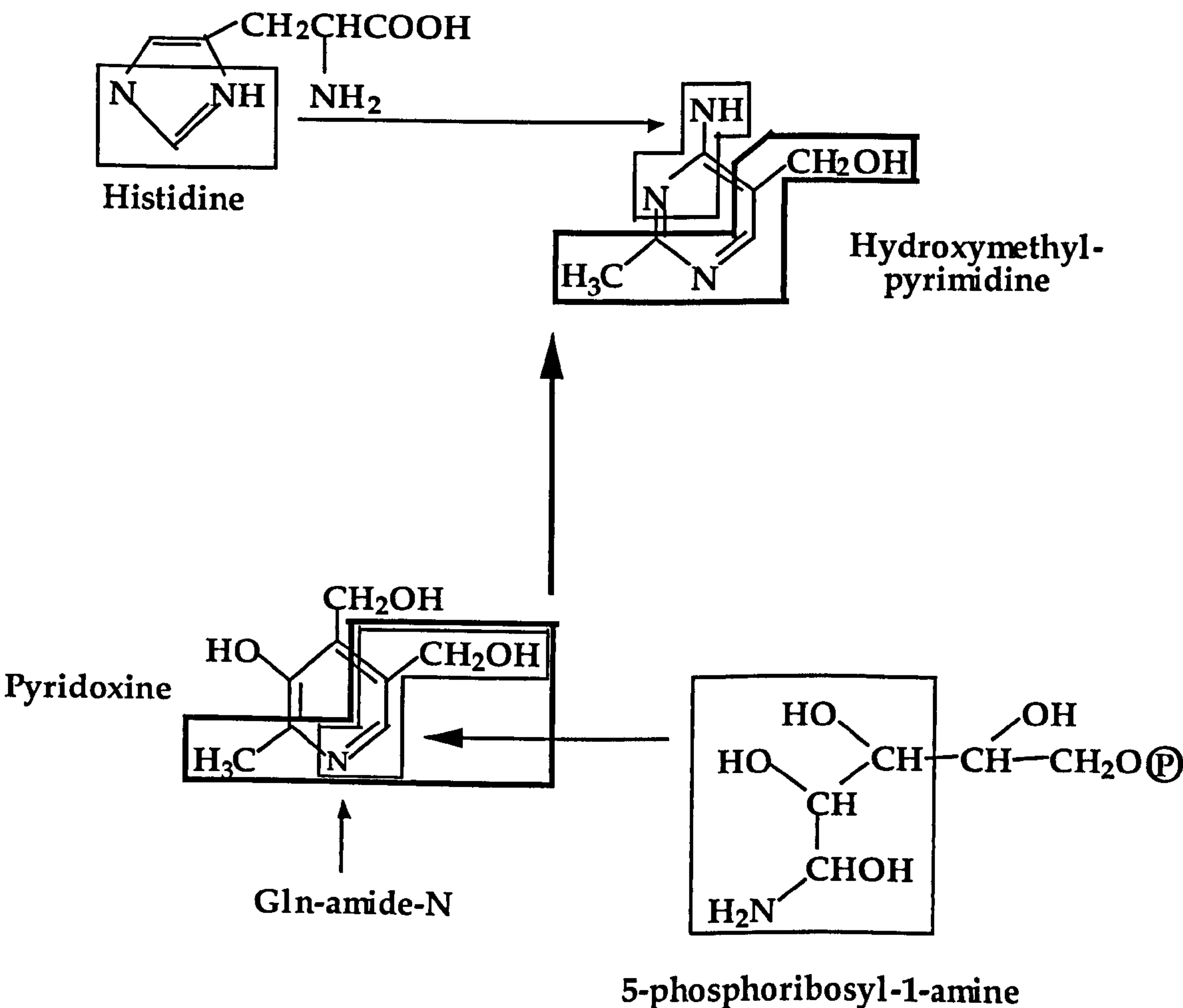
The numbering system for the principal atoms of HMP are as follows;



a.)



b.)



nucleotides, which proceed via very different pathways to that of pyridoxine and subsequently HMP formation. A key *S. cerevisiae* enzyme, called phosphoribosylpyrophosphate synthetase which is involved in the biosynthesis of PRPP, has been found to be encoded by the *PRS* gene family (Carter *et al.*, 1997).

The remainder of the HMP molecule, represented by the N-3, C-4, N-4' backbone, has been found to originate from part of the imidazole ring of the amino acid histidine (Figure 1.6, (Tazuya *et al.*, 1989)).

1.5.2 The *S. cerevisiae* *THI5* gene family.

THI5 was originally isolated as a cDNA clone obtained in the same screen as *THI4*, for genes that are expressed at entry into stationary phase. This gene was initially designated *MOL2* (Praekelt and Meacock, 1992). The *THI5* cDNA, when used as probe for Southern blot hybridisation, hybridised to several fragments for a variety of genomic restriction digests (Hather, 1996). This observation led to an early hypothesis that *THI5* was a part of a gene family comprising three members. Elucidation of the complete genome sequence of *S. cerevisiae* revealed the presence of a fourth gene. Each gene is 1023 bp in length and found within the subtelomeric region of four different chromosomes. The designations (with systematic codes in brackets) are *THI5* (*YFL058w*, chromosome VI), *THI11* (*YJR156c*, chromosome X), *THI12* (*YNL332w*, chromosome XIV) and *THI13* (*YDL244w*, chromosome IV). Collectively, these genes are referred to as the *THI5* gene family. Each of these members share the subtelomeric regions with members of other gene families. As discussed earlier in this chapter, there is a concentration of genes within the subtelomeres that are members of gene families.

The four genes, *THI5*, *THI11*, *THI12* and *THI13*, and their products, show a remarkable degree of sequence conservation at the DNA and protein levels. There are twenty seven nucleotide differences between the four open reading frames. Of

these, only three result in amino acid changes in the protein (Figure 1.7). Because of these subtle differences, each gene encodes an apparent isozyme.

There is a variety of experimental evidence to support a role for one or more of the isozymes in the formation of HMP-PP and therefore thiamin biosynthesis. Firstly, the *S. pombe* homologue, called *nmt1p*, has been functionally assigned to this pathway and is the focus of the next section. Secondly, the pool of transcripts derived from the *THI5* gene family is abundant as shown by northern blot hybridisation of total RNA isolated from *S. cerevisiae* which has been cultivated in minimal medium lacking thiamin. No detectable transcript is present in yeast grown in the same medium supplemented with greater than 1.5 μ M thiamin (Burrows, 1997; Hather, 1996). In addition, at least three of the isozymes have been detected by two-dimensional electrophoresis of whole cell extracts of *S. cerevisiae* grown in thiamin-limited medium; these proteins are undetectable when the yeast are grown in the presence of thiamin (Muller *et al.*, 1999).

A *S. cerevisiae* strain disrupted for both *THI5* and *THI12* has been constructed (Hather, 1996). This *thi5 thi12* mutant is not a thiamin, HET or HMP auxotroph indicating that Thi11p and/or Thi13p are involved in thiamin biosynthesis. These data and the protein analysis suggest that there is a degree of functional redundancy within the *THI5* family. This apparent redundancy can only be determined with the construction of strains containing all combinations of single, double and the triple mutants once the phenotype of the quadruple mutant has been determined. The creation and phenotypic analysis of these mutants is one of the main aims of this study and will be discussed in Chapter 5.

1.5.3 The *THI5* homologue in *S. pombe*, *nmt1*.

The proposed role for one or more of the *S. cerevisiae* Thi5p isozymes in HMP-PP biosynthesis is entirely based on the studies of the *S. pombe* homologue, *nmt1p*. Each of the four isozymes are approximately 62% identical to the *S. pombe* enzyme. The

			*	
Thi5p	1	MSTDKITFLLNWQPTPYHIPIFLAQT	KGYFKEQGLDMAILEPTNPSDVTE	LIGSGKVDMG
Thi12p	1	MSTDKITFLLNWQPTPYHIPIFLAQT	KGYFKEQGLDMAILEPTNPSDVTE	LIGSGKVDMG
Thi13p	1	MSTDKITFLLNWQPTPYHIPIFLAQT	KGYFKEQGLDMAILEPTNPSDVTE	LIGSGKVDMG
Thi11p	1	MSTDKITFLLNWQPTPYHIPIFLAQT	KGYFKEQGLDMAILEPTNPSDVTE	LIGSGKVDMG
Thi5p	61	LKAMIHTLAAKARGFPVTSVASLLDEP	FTGVLYLKSGGITEDFQSLKGKKIGY	VGEFGKI
Thi12p	61	LKAMIHTLAAKARGFPVTSVASLLDEP	FTGVLYLKSGGITEDFQSLKGKKIGY	VGEFGKI
Thi13p	61	LKAMIHTLAAKARGFPVTSVASLLDEP	FTGVLYLKSGGITEDFQSLKGKKIGY	VGEFGKI
Thi11p	61	LKAMIHTLAAKARGFPVTSVASLLDEP	FTGVLYLKSGGITEDFQSLKGKKIGY	VGEFGKI
			*	
Thi5p	121	QIDELTKHYGMKPEDYTAVRCGMNVA	KYIIIEGKIDAGIGIECMQQVELEEY	LAKQGRPAS
Thi12p	121	QIDELTKHYGMKPEDYTAVRCGMNVA	KYIIIEGKIDAGIGIECMQQVELEEY	LAKQGRPAS
Thi13p	121	QIDELTKHYGMKPEDYTAVRCGMNVA	KYIIIEGKIDAGIGIECMQQVELEEY	LAKQGRPAS
Thi11p	121	QIDELTKHYGMKPEDYTAVRCGMNVA	KYIIIEGKIDAGIGIECMQQVELEEY	LAKQGRPAS
Thi5p	181	DAKMLRIDKLACLGCCCFCTVLYICN	DEFLKKNPEKVRKFLKA	IKKATDYVLADPVKAWK
Thi12p	181	DAKMLRIDKLACLGCCCFCTVLYICN	DEFLKKNPEKVRKFLKA	IKKATDYVLADPVKAWK
Thi13p	181	DAKMLRIDKLACLGCCCFCTVLYICN	DEFLKKNPEKVRKFLKA	IKKATDYVLADPVKAWK
Thi11p	181	DAKMLRIDKLACLGCCCFCTVLYICN	DEFLKKNPEKVRKFLKA	IKKATDYVLADPVKAWK
			*	
Thi5p	241	EYIDFKPQLNNDLSYKQYQRCYAYF	SSSLYNVHRDWKKVTGYGKRLAIL	PPDYVSNYTNE
Thi12p	241	EYIDFKPQLNNDLSYKQYQRCYAYF	SSSLYNVHRDWKKVTGYGKRLAIL	PPDYVSNYTNE
Thi13p	241	EYIDFKPQLNNDLSYKQYQRCYAYF	SSSLYNVHRDWKKVTGYGKRLAIL	PPDYVSNYTNE
Thi11p	241	EYIDFKPQLNNDLSYKQYQRCYAYF	SSSLYNVHRDWKKVTGYGKRLAIL	PPDYVSNYTNE
Thi5p	301	YLSWPEPEEVSDPLEAQRLMAIHQEK	CRQEGTFKRLALPA~	
Thi12p	301	YLSWPEPEEVSDPLEAQRLMAIHQEK	CRQEGTFKRLALPA~	
Thi13p	301	YLSWPEPEEVSDPLEAQRLMAIHQEK	CRQEGTFKRLALPA~	
Thi11p	301	YLSWPEPEEVSDPLEAQRLMAIHQEK	CRQEGTFKRLALPA~	

Figure 1.7 Alignment of the Thi5p, Thi11p, Thi12p and Thi13p isozymes.

Alignments were created using default settings on PILEUP (see materials and methods).

Differences between any of the proteins are displayed as an unhighlighted or grey highlighted character. The positions where differences are found are marked by an asterix(*).

nmt1 gene was isolated in a screen for genes whose expression is sensitive to thiamin (Maundrell, 1990). Messenger RNA from two *S. pombe* cultures, one grown in minimal medium and one grown in minimal medium supplemented with thiamin, was used to create the corresponding labelled cDNA probes. These probes were used in a differential screen of a *S. pombe* genomic library. One of the clones obtained in this way was further characterised and the gene within the insert was named *nmt1* (*n*o *m*essage in *t*hiamin). Northern blot hybridisation of *S. pombe* mRNA grown in varying quantities of thiamin shows that *nmt1* is fully repressed at concentrations greater than 0.5 μ M thiamin. This repression is due to inhibition of transcription initiation rather than an increase in mRNA turnover (Maundrell, 1990). The *nmt1* gene was the first fully repressible gene identified in this yeast and its promoter is extensively used as a tool for the inducible expression of cloned genes (Maundrell, 1993). The thiamin regulatory element has been identified by examining the inducible expression of a reporter gene containing internal deletions in the *nmt1* promoter. This element has been localised to between 51 bp and 61 bp upstream of the ATG translational start codon (Zurlinden & Schweingruber, 1997).

A *S. pombe* strain containing disrupted *nmt1* is a thiamin auxotroph (Maundrell, 1990). This auxotrophy can be rescued by the addition of thiamin or HMP, but not HET, to the minimal growth medium, confirming its role in HMP-PP biosynthesis (Schweingruber *et al.*, 1991). The position of Nmt1p activity in the HMP-PP anabolic pathway relative to pyridoxine has not yet been determined.

1.5.4 The *S. cerevisiae* HMP-P kinase family.

During the course of this study, the first *S. cerevisiae* genes to encode a biosynthetic enzyme involved in the formation of HMP-PP were characterised (Llorente, *et al.*, 1999). The investigation by Llorente *et al.* examined a three-membered gene family with the systematic ORF codes YOL055c, YPL258c and YPR121w. The amino-terminal parts of these genes share approximately 32% identity with the known HMP-P kinase of *S. typhimurum*. For this reason, *S. cerevisiae* strains covering all

combinations of the single, double and triple deletion mutants have been constructed and their phenotypes examined with respect to their thiamin requirement. Thiamin auxotrophy is only seen for those strains that contain a deletion in both the *YOL055c* and *YPL258c* ORFs; these genes are therefore functionally redundant. On the other hand, *YPR121w* is not required for thiamin prototrophy. Despite these observations in the mutant strains, all three genes are expressed in response to thiamin depletion and expression is dependent upon the regulatory factors *THI2* and *THI3* in each case. Therefore, *YPR121w* appears to have an as yet unidentified role in thiamin biosynthesis. One possible role is in the anaerobic HMP pathway described below. The *YOL055c* and *YPL258c* ORFs have been given the designations of *THI20* and *THI21*, respectively and have been included in the scheme of Figure 1.6.

1.5.5 Evidence of an anaerobic HMP biosynthetic pathway.

It has long been known that there is an alternative route to HMP synthesis in *S. typhimurum*. Recently, evidence has been published by Tanaka *et al.* describing an alternative but unrelated HMP biosynthetic pathway that functions when *S. cerevisiae* is grown under anaerobic conditions (Tanaka *et al.*, 2000). The HMP pathway discussed previously in this chapter has been investigated in aerobic cultures of *S. cerevisiae*. Under aerobiosis, radiolabelled formate is incorporated into the pyrimidine ring via the amino acid histidine (Tazuya *et al.*, 1989). Radiolabel from pyridoxine is also efficiently incorporated (Tazuya *et al.*, 1994). Experiments have been performed examining the incorporation of [¹³C]-labelled formate and [5'-²H₂]-labelled pyridoxine during anaerobiosis. In these conditions, *S. cerevisiae* has been found not to incorporate either of the labelled compounds into the HMP molecule (Tanaka *et al.*, 2000). These data confirm that the pathway used for biosynthesis of the pyrimidine moiety of thiamin in *S. cerevisiae* under anaerobic conditions differs from that used under aerobic conditions.

Further evidence supporting the existence of an alternative HMP pathway comes from work carried out by Grue-Sorensen and colleagues. They carefully examined

the incorporation of [^{14}C]-formate into the pyrimidine ring under aerobic conditions (Grue-Sorensen *et al.*, 1986). The results found incorporation at two sites within the HMP molecule, the C-2 carbon and the C-4 carbon (see Figure 1.6 legend for number designations). Moreover, the radiolabel was not uniformly distributed between the two atoms; 70% was located at C-4 whilst only 20% at C-2. This observation can only be explained by the presence of two independent pathways where [^{14}C]-formate enters C-4 by a more efficient route and C-2 by a less efficient route (Figure 1.8). The C-4 route represents the recognised pathway described in section 1.5.1.

Both Grue-Sorensen and Tanaka agree that there are two different pathways to HMP formation. However, in Tanaka's proposed pathway, formate is not incorporated into the pyrimidine. If the results from both teams are correct then *S. cerevisiae* could have as many as three different routes to HMP formation. Although possible, this appears unlikely. Grue-Sorensen's C-4 principal route is that which proceeds via pyridoxine and histidine. As this has been shown to be absent under anaerobiosis, then a good candidate for the alternative route is the minor C-2 route. Had Tanaka and colleagues grown *S. cerevisiae* under appropriate anaerobic conditions then they may have noticed a small incorporation of the labeled formate. They failed to add two important growth requirements to allow the successful propagation of *S. cerevisiae* during anaerobiosis, oleic acid and ergosterol. As reported in section 1.2.1, both of these must be supplied in the growth medium.

1.6 Pyridoxine.

1.6.1 Biosynthesis.

As well as being an important precursor for HMP-PP biosynthesis, pyridoxine (vitamin B6) is also an essential cofactor for a number of enzymes. The active form is pyridoxal 5' phosphate (PLP) that functions in a diverse range of reactions, the most common of which is in the metabolism of amino acids (reviewed in (Mittenhuber, 2001)). Whereas the ThDP-containing enzymes in *S. cerevisiae* have been identified

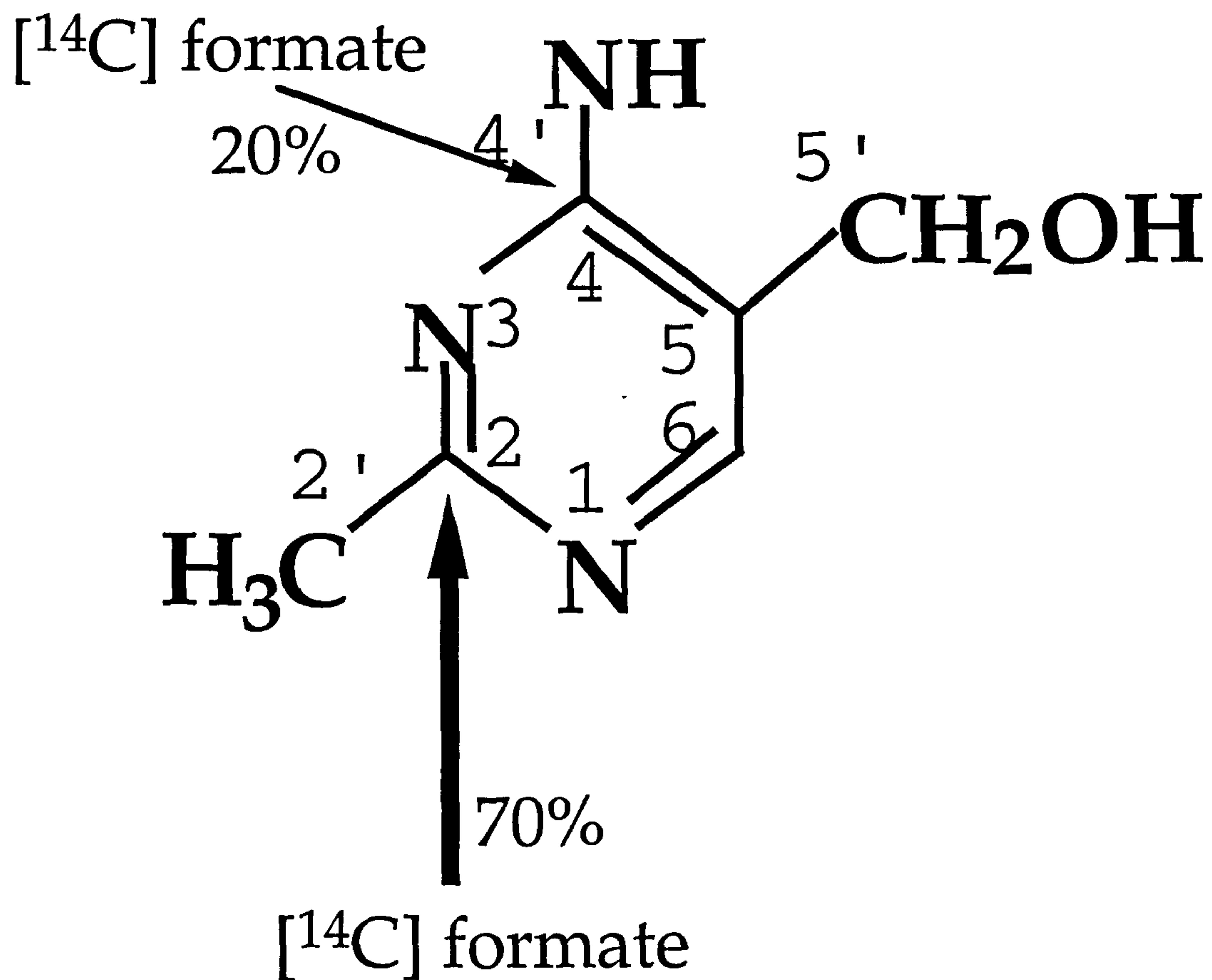


Figure 1.8 Incorporation of $[^{14}\text{C}]$ -formate into the HMP molecule. The labelled carbon can enter at one of two sites. Figures printed as percentages represent the radioactivity emitted at that site as a proportion of the total present in the HMP molecule.

based on the conservation of binding sites (Hohmann and Meacock, 1998), a thorough search for genes encoding PLP-dependent enzymes has not been carried out.

Like ThDP synthesis, PLP formation occurs by different pathways in *E. coli* and *S. cerevisiae*. As described above, pyridoxine is synthesised from PRA in *S. cerevisiae* and HMP-PP synthesis is just a further extension of this pathway. In *E. coli*, pyridoxine is synthesised from 1-deoxy-D-xylulose-5-phosphate which is also a substrate for the formation of the thiazole moiety (Begley, *et al.*, 1999). Although several genes have been identified which encode PLP biosynthetic enzymes in *E. coli*, no such enzymes or the genes that encode them had been identified in *S. cerevisiae* prior to the commencement of this study.

A number of genes encoding pyridoxine biosynthetic activities have now been identified as a result of work carried out in this laboratory. Two enzymes have been found to be encoded by gene families called *SNO* and *SNZ* (Marsh, 2000). These have been included in the scheme of Figure 1.6. The *SNZ* and *SNO* (*SNZ*-proximal *QRF*) genes have been the focus of much previous work, in an attempt to identify their functional roles. The *SNZ* genes were originally identified due to their expression at stationary phase. As all *SNZ* family members in *S. cerevisiae* are adjacent to a *SNO* gene, it was presumed that they are functionally related. This is reinforced by the observation that their homologues in other organisms are also adjacent. Much of the earlier studies of these genes were performed in the laboratory of Werner-Washburne. By creating strains containing various disruptions of the *SNZ* or *SNO* genes, Werner-Washburne was able to perform a series of phenotypic tests and found a noticeable sensitivity of several strains to 6-azouracil, an inhibitor of both purine and pyrimidine synthesis. Further analysis suggested some role of *SNO* and *SNZ* in pyrimidine biosynthesis (Padilla *et al.*, 1998). These findings were of immediate relevance to this thesis for two reasons; firstly, *SNO* and *SNZ* may have a role in HMP-PP formation rather than in the synthesis of pyrimidine nucleotides; secondly, each of the two gene families comprises three

members. Two members of the *SNO* and *SNZ* families are very highly conserved, exhibiting greater than 99% identity at the DNA level (Marsh, 2000). These highly conserved members are subtelomerically located (Figure 1.2) and are found immediately telomere-proximal to *THI5* and *THI12*.

A suggested functional role for both gene families came after the single copy *SNZ* homologues in the fungi *Aspergillus nidulans* (*pyroA*) and *Cercospora nicotinae* (*SOR1*) were found to encode enzymes involved in pyridoxine biosynthesis (Ehrenshaft *et al.*, 1999; Osmani *et al.*, 1999).

Marsh focused on the *SNO* gene family and carried out a comprehensive functional analysis of the three members. *SNO2* and *SNO3* are the genes which are positioned close to *THI5* and *THI12*, respectively (Chapter 2, Figure 2.5). *SNO1* is not subtelomerically located and exhibits significant sequence divergence at both the DNA and protein (67% identity) levels compared to *SNO2* and *SNO3*. Marsh constructed strains comprising all combinations of the single, double and triple deletion mutants. Phenotypic analysis with respect to thiamin and pyridoxine prototrophy has yielded very interesting results. On medium lacking pyridoxine, a strain deleted for all three *SNO* genes exhibits no growth. This pyridoxine auxotrophy is not affected by the presence of thiamin in the growth medium. Strains containing a deletion for *SNO1*, either alone or in combination with a deletion of *SNO2* or *SNO3*, all show severely retarded growth on medium lacking pyridoxine but containing thiamin. The strains grow normally if both pyridoxine and thiamin are lacking from the medium, however, on this same growth medium, a strain deleted for both *SNO2* and *SNO3* shows severely retarded growth (Marsh, 2000).

Taken together, these data suggest that *SNO1* serves to synthesise pyridoxine for use as a cofactor when the subtelomeric genes (*SNO2* and *SNO3*) are dispensable i.e. when there is an exogenous supply of thiamin. However, in conditions of thiamin limitation, both *SNO2* and *SNO3* gene products are required to supply pyridoxine for HMP-PP biosynthesis (Figure 1.6).

1.6.2 Regulation of *SNO* and *SNZ* by pyridoxine and thiamin.

Preliminary analysis of the regulation of the *SNO* gene family has been carried out in this laboratory by assaying the activity of promoter:*lacZ* fusions (D McKissock, personal communication). Data concerning the regulation of the subtelomeric copies are currently available and show that the *SNO2/3* promoter:reporter construct has enhanced activity in the absence of thiamin (Figure 1.9). In the absence of this vitamin, activity of the *lacZ* product gives approximately 750 β -galactosidase units compared with below 200 units when thiamin is present. Curiously, maximal activity is only observed in the presence of pyridoxine. Based upon the functional analysis by Marsh, it is expected that *SNO1* will be regulated by pyridoxine and not thiamin.

1.6.3 The effect of thiamin in pyridoxine-free medium.

Before the elucidation of the origins of thiamin and its precursors, a number of experiments in the early 1980's found that thiamin seriously affected growth of *S. carlsbergensis* (now *S. pastorianus*) and other yeasts which were grouped as 'type 1' and which we now know as the *Saccharomyces sensu stricto*. It was observed that the addition of thiamin to a pyridoxine-free medium caused this growth inhibition (reviewed in Kamihara & Nakamura, 1982). When pyridoxine is added to the growth medium, either on its own or together with thiamin, normal growth was observed. Coupled with the growth inhibition, the addition of thiamin to pyridoxine-free medium also causes a marked reduction in the intracellular pyridoxine content, respiratory deficiency and a reduction in the activity of δ -aminolevulinate synthase (ALS) which is essential for heme biosynthesis. ALS is known to be a PLP-dependent enzyme. Both Kamihara and Nakamura concluded that the addition of thiamin repressed pyridoxine biosynthesis which in turn caused a decrease in ALS activity, a decrease in the intracellular heme content followed by a decrease in respiratory activity which caused the growth inhibition.

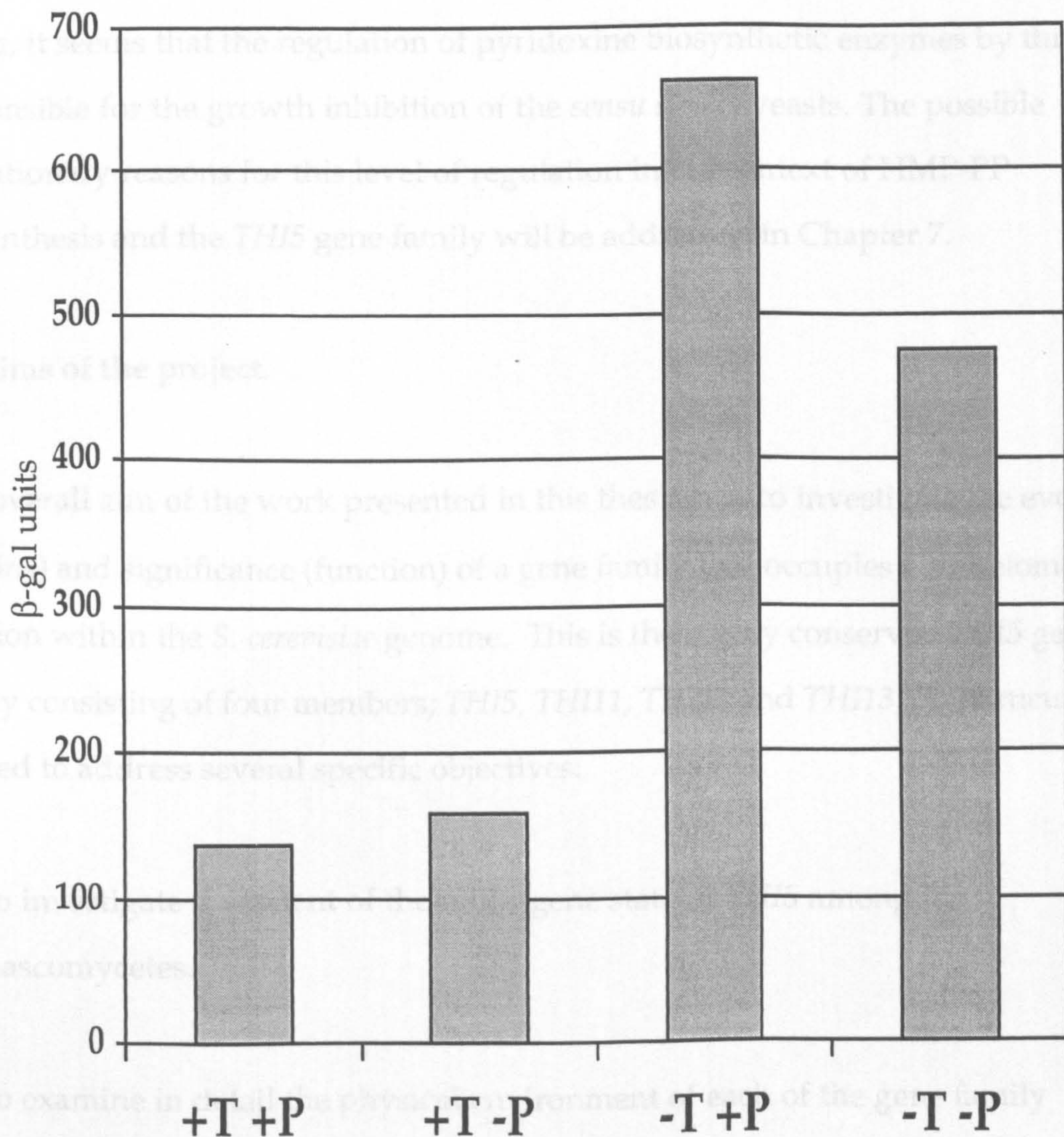


Figure 1.9 Activity of a *lacZ*-reporter construct containing the *SNO2/3* promoter. Activity was measured in yeast grown in medium supplemented with (+) or deficient for (-) the vitamins thiamin (T) or pyridoxine (P) as shown in the graph. The data for this graph were obtained from D. McKissock and P.A. Meacock, University of Leicester.

It is interesting to note that Kamihura and Nakamura came to these conclusions a decade before it was known that pyridoxine is itself a precursor for the biosynthesis of the pyrimidine moiety of thiamin. Together with the SNO information described above, it seems that the regulation of pyridoxine biosynthetic enzymes by thiamin is responsible for the growth inhibition of the *sensu stricto* yeasts. The possible evolutionary reasons for this level of regulation in the context of HMP-PP biosynthesis and the *THI5* gene family will be addressed in Chapter 7.

1.7 Aims of the project.

The overall aim of the work presented in this thesis was to investigate the evolution (origins) and significance (function) of a gene family that occupies a subtelomeric location within the *S. cerevisiae* genome. This is the highly conserved *THI5* gene family consisting of four members; *THI5*, *THI11*, *THI12* and *THI13*. In particular I wished to address several specific objectives:

- (1) To investigate the extent of the multi-gene state of *THI5* among the hemiascomycetes.
- (2) To examine in detail the physical environment of each of the gene family members within the *S. cerevisiae* genome in order to find clues about the molecular processes that generated the four copies. Extending this analysis to *THI5* homologues and their flanking regions in related yeasts might provide information about the ancestral state.
- (3) To deduce the functional role of the isozymes encoded by these genes, and the extent of redundancy amongst the four family members. Based on their very high degree of similarity to each other it seemed possible that this gene family may be fully functionally redundant in terms of HMP biosynthesis for ThDP formation. I also wished to determine if the expression of all four genes is regulated in a similar

manner, to establish whether there were phenotypic differences at this level. Analysis of other related gene families, eg. *SNZ* and *SNO*, has revealed subtle regulatory differences between the members (Chapter 1, section 1.6.2).

(4) To investigate reports of an alternative pathway for HMP biosynthesis as proposed in section 1.5.5.

Chapter 2.

A bioinformatic analysis of thiamin biosynthesis in various organisms and of the *S. cerevisiae* *THI5* gene family.

2.1 Introduction.

The differences between thiamin biosynthesis in bacteria and fungi are predominantly within the pathways leading to the two precursor molecules, hydroxyethylthiazole (HET) and hydroxymethylpyrimidine (HMP). The fate of radio-labelled substrates into the thiamin molecule has been compared between *E. coli* and *S. cerevisiae* using high-resolution NMR spectroscopy (Himmeldirk, *et al.*, 1998). This study showed that there are different routes to precursor biosynthesis between the two organisms (Chapter 1). Differences are also seen at the genetic level. HET biosynthesis has been found to involve at least six gene products in bacteria whereas only a single, unrelated gene, *THI4*, has so far been implicated in yeast. Similarly, HMP formation requires the product of *ThiC* in bacteria (Begley, *et al.*, 1999; Praekelt, *et al.*, 1994). The role of the *S. cerevisiae* Thi5 protein and its isozymes in thiamin biosynthesis is the focus of this study and, based upon characterisation of the *S. pombe* homologue, it is expected that one or more function in HMP formation. There is no ThiC homologue in *S. cerevisiae*, neither is there a Thi5p homologue in *E. coli*.

The literature that describes thiamin biosynthesis in *E. coli*, *S. typhimurum* and in yeast presumes that the deduced pathways from these organisms are representative of their entire phylogenetic kingdoms. Also, it has been incorrectly documented (e.g. Himmeldirk *et al.*, 1998) that the yeast biosynthetic route represents the pathway found in all eukaryotes that make thiamin. It is believed that plants synthesise HET-P from the same substrates as *E. coli* and *S. typhimurum*, but may use the fungal enzyme Thi4p (Juilliard & Douce, 1991; Machado, *et al.*, 1997). It is unknown what route is taken by plants for synthesis of the HMP precursor.

Included in this chapter are the results of a search of the genomes of various organisms of the plant, fungal and prokaryotic kingdoms for homologues of *E. coli* and *S. cerevisiae* thiamin biosynthetic genes. By assigning the correct pathway based upon the genes present, it will be possible to determine the extent of the various biosynthetic routes that are present within the three kingdoms. Particular attention is paid to the fungi, especially with regards to the organisation of the *THI5* family members within the *S. cerevisiae* genome which, to date, is the only organism known to possess multiple copies of the gene.

2.2 BLASTP searches.

2.2.1 Assignment of pathways to thiamin precursors within the domains of bacteria, archaea, plants and fungi.

An analysis of the large amounts of available DNA sequence data for various organisms was carried out to look for genes that encode putative thiamin biosynthetic enzymes. Table 2.1 displays the results of BLASTP searches of computer databases to look for homologues of *E. coli* and *S. cerevisiae* thiamin biosynthetic proteins. These are the pyrimidine biosynthetic enzymes ThiCp and Thi5p and the thiazole enzymes ThiGp and Thi4p of *E. coli* and *S. cerevisiae*, respectively. The homologues are arranged according to the phylogenetic domain of the parental genome. Their presence or absence provide an indication of the route(s) these organisms use for the synthesis of the two principal thiamin precursors.

The BLASTP analysis shows that homologues of *E. coli* ThiCp, involved in the synthesis of the pyrimidine moiety from a purine intermediate (5-aminoimidazole ribotide), are found in three of the four domains. Fungi are the exception; all possessing a Thi5p homologue and therefore, like *S. cerevisiae*, probably synthesise HMP as an extension of pyridoxine biosynthesis.

Table 2.1 The genome resource tools at NCBI were used to display all the homologues of the hydroxymethylpyrimidine biosynthesis enzymes ThiC (*E.coli*) and Thi5 (*S. cerevisiae*), and the homologues of the hydroxyethylthiazole enzymes ThiG (*E. coli*) and Thi4 (*S. cerevisiae*). The BLASTP results, with a default cut-off score of 95, are grouped according to the phylogenetic domain of the organism. Those organisms for which the genome has been completely sequenced were checked individually for the presence of the query genes.

A ✓ indicates the organism contains a homolog, X indicates that a homologue is not present within the completed genome.

Escherichia coli;

(<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/altvik?gi=115&db=g&gene=thiC>)

(<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/altvik?gi=115&db=g&gene=thiG>)

Saccharomyces cerevisiae;

([http://www.ncbi.nlm.nih.gov/cgi-](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/sgeenik?gene=THI5&db=Genome&org=Saccharomyces+cerevisiae)

[bin/Entrez/sgeenik?gene=THI5&db=Genome&org=Saccharomyces+cerevisiae](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/sgeenik?gene=THI5&db=Genome&org=Saccharomyces+cerevisiae))

([http://www.ncbi.nlm.nih.gov/cgi-](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/sgeenik?gene=THI4&db=Genome&org=Saccharomyces+cerevisiae)

[bin/Entrez/sgeenik?gene=THI4&db=Genome&org=Saccharomyces+cerevisiae](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/sgeenik?gene=THI4&db=Genome&org=Saccharomyces+cerevisiae)).

For BLASTP searches of completed genomes;

(<http://www.ncbi.nlm.nih.gov/BLAST/>)

- Select 'Standard protein-protein BLAST [blastp]'
- Query sequences were the ThiC, ThiG, Thi5 and Thi4 peptides.
- The name of the subject organism was entered under 'Limit by entrez query' .
- All other options were at their default settings.

	Pyrimidine		Thiazole	
	ThiC	Thi5p	ThiG	Thi4p
Fungi (F)				
<i>*Saccharomyces cerevisiae</i>	X	✓	X	✓
<i>Schizosaccharomyces pombe</i>		✓		✓
<i>Uromyces fabae</i>		✓		✓
<i>Aspergillus parasiticus</i>		✓		
<i>Fusarium oxysporum</i>				✓
Plants (P)				
<i>*Arabidopsis thaliana</i>	✓	X	X	✓
<i>Zea mays</i>				✓
<i>Poa secunda</i>	✓			
<i>Oryza sativa</i>	✓			
<i>Citrus sinensis</i>	.			✓
Archaea (A)				
<i>*Archaeoglobus fulgidus</i>	✓	✓	X	✓
<i>Methanothermobacter thermautotrophicus</i>	✓			✓
<i>*Methanococcus jannaschii</i>	✓	X	X	✓
<i>Pyrococcus abyssi</i>	✓			✓
<i>Halobacterium</i> sp. NRC-1	✓			✓
Bacteria (B)				
<i>*Thermotoga maritima</i>	✓	✓		✓
<i>Pseudomonas aeruginosa</i>	✓	✓	✓	
<i>*Haemophilus influenzae</i>	X	✓	X	X
<i>*Escherichia coli</i>	✓	X	✓	X
<i>Chlamydia pneumoniae</i>		✓		
<i>Vibrio cholerae</i>	✓		✓	
<i>Streptomyces coelicor</i>	✓		✓	
Conclusion	P, A, B	F, A, B	B	F,P, A,B

*-denotes sequence of complete genome was analysed.

Table 2.1

Several members of the archaea and bacteria possess a Thi5p homologue in addition to, or instead of, ThiCp (Figure 2.1). It is interesting to note that in prokaryotes, the gene encoding the Thi5 protein homologue always lies adjacent to a putative ABC transporter and in the same orientation, suggesting that they may be co-transcribed as part of an operon. The *Haemophilus influenzae* genome is notable in that *thiC* appears to be absent from the complete genome sequence (Table 2.1). The *H. influenzae* Thi5p homologue shares 37% similarity with the entire *S. cerevisiae* protein (Figure 2.1). In addition in *H. influenzae*, the *THI5* homologue is flanked by an ABC transporter and on the other side, an open reading frame with similarity to a hydroxymethylpyrimidine phosphate kinase (HMP-P kinase). For two species of *Chlamydia*, *C. trachomatis* and *C. pneumoniae*, the *THI5* homologue and the ABC transporter appear to be contained within a single open reading frame which would give a bifunctional product, therefore a gene fusion event may have taken place. It is not known why, in these organisms, the Thi5p homolog is so tightly associated with the permease. However, this might provide a clue about the kind of protein that Thi5p may interact with in *S. cerevisiae*.

Of the plants, *Arabidopsis thaliana* was the only organism where the complete genome sequence was publicly available. *A. thaliana* possesses just a ThiCp homologue and contains no Thi5p homologue. ThiCp-like proteins are found in *Poa secunda* (bluegrass) and *Oryza sativa* (rice). The absence of an apparent Thi5p homologue indicates that plants probably utilise the bacterial route to HMP formation.

Homologues of the *E. coli* thiazole biosynthetic enzyme, ThiGp, are only found in the bacterial phylogenetic domain (Table 2.1). In contrast, homologues of the *S. cerevisiae* enzyme, Thi4p, are readily identified in members of the other phylogenetic domains including the bacterium *Thermotoga maritima*. However, it is difficult to assign which thiazole pathway these organisms possess as it seems that plants use the starting substrates of the bacterial pathway but may carry out an enzymatic step that involves Thi4p (section 2.1).


```

C. pneumoniae 1 ~KALSVLALIPIMLIPWKGNSKSPD...KKNLTSITLLDWTNPNNHIPLYAGVAKGYF
C. trachomatis 1 ~~~~~FALIPITVLPCIFYLKDDPKLAAPVPTKSFTLLDWTNPNNHIPLYAGVEKGFF
H. influenzae 1 ~~~~~MKTIRYFSFVMGLMLILPS..FA...KEKISTMLDWYVNPDAHAATIVAQCKGFF
B. halodurans 1 MKSFKIISLLLATLFLASCQTNTASDDEA...LETVEVMLDWYPNAVHTFLYVAIENGYF
T. maritima 1 ~~~~~MKRLIFIAMFLISVLA.LFA....EEVTVLDWYPNTNHTGLYVAKDLGFY
S. cerevisiae 1 ~~~~~~MS...TDKIITFLLNWQPTPYHIPHIAQTKGYF
A. fulgidus 1 ~~~~~~MRRLLLTLLAATLLGCSQAPQEKLKVGIL....PIEDSIPVVADEQGGF

C. pneumoniae 57 KQHGLDLQKNTDSSAVPHVLFEQVDMALYHALGIMKTSIKGMPQIVGRLI.DSSLO
C. trachomatis55 VDEGISLTLOKNTDTCSSIPHLLEKVDYTLYHSLGVLEKJAVRGAPVQVAGRLI.DSSLO
H. influenzae 51 EKNNEVEITEPADPALPPKLAAGKVDLAVSYQPQYQQVAEGLPIVRVGSLLI.SNPLN
B. halodurans 58 AEEGLDVIVFPTNPTDPIQLTASGATPLALSYQPDVILARSKDLPVSVASVV.RSPLN
T. maritima 47 REEGLDVKIVQPSRLTAE.QLVASGKAFFGVSYQEAFTLSRGEPMPIVSTAAII.QHNTS
S. cerevisiae 31 KEQGLDMAILEPTNPSTVTELGSGKVDMLKAMHTLAARKGFVPTSVASLI.DEPFT
A. fulgidus 48 KKHGLDVEVIQFSSAVERDAALTAGEVDAVITDPLAVILLRDKGYDVRIASLCIGKIPKE

C. pneumoniae116 GF..LYRSQDPIYK.FEDLNGKVLGFCINNSRDLNRLLETILNRNGVVPSEVK..NVSSDL
C. trachomat 114 GL..LYRKNEGIEK.LEDLNGRVLGFCINDSKNLPNLLEALRKHHVVPSEIK..NVSADM
H. influenzae110 SV..VVLKKSNI.KSLADLKGKKGYSVSGFED..CLLDTMLHSIGLSNKDVELNVNWSL
B. halodurans117 HV..MFLAEQDF.DSPADLVGLTVGYPGIPVNE.PILKTMVEAAGGDYEQVHLMVGFEL
T. maritima 105 GF..AWLKDEGI.KSVKDWEGKRYGWSGSPTEK.ATIEYIMRKYGADPSKVIFVNVG...
S. cerevisiae 90 GV..LYLKGSGITEDFQSLKGGKIGYVGE.FGK.IQIDELTKHYGMKPEYTAVRCGMNV
A. fulgidus 108 GVFAILLAPNSDIDSVEDLDGKEIAVSINTIEYA.LDVMLSEYGLNYIKISIPKPIRM

C. pneumoniae171 ISPMLLNKIDFLYG.AFYNIEGVKLQT....LGMPVKCFLSDTCDLPTGP...QLIVFT
C. trachomat 169 ISPMITYQIDFLYG.GFYNVEGVITIAL....KGTPTGCFLSDTYGSPTGP...QLITCG
H. influenzae166 SPSLLTGQVDAVIG.AFRNF.ELNQ.LALEKQEG..IAFF.PEQYG.VP..AYDELILVA
B. halodurans173 GASIVSGRADAVVG.TYINH.EYFV.L...KHEGHDISYFNPVDYG.VP..EYDELVLIS
T. maritima 158 QMDFFACTLNDVDFAWIFY.GWDG.VA.SKVKGIEIEELPLKDIDEVF..DYTPVLIT
S. cerevisiae146 AKYIECKIDAGTIGIECMQQVELEEYLAQGRPASDAKMLRIDKLACLGCCCTVLYIC
A. fulgidus 167 QA.LLNCNVETAV.....LPEPLASYAV.SKGAKLILSD...AMLNESITQTVIVF

C. pneumoniae222 KKGTKASEPEIVEAFQKALQESIIESKDHPEDAFKLYAKETKSIPKNLYQEYLQWEETFP
C. trachomat 220 KKGSPAMTPQTLQSLQKALSRLDFCREYPOEAFIYVEATKDSPKVLSDEHAQWEVTLF
H. influenzae217 NKNSVTD..KKTSAFILTALEQATSYLQAHNPNEAQAQFVSYPNELNTPNLQLA.WKDTLP
B. halodurans224 NEAYVEESGEVLAFFWRAALKCYEWMVENPDEALNVLLT.NQDEANFPLIQEV.EEESLS
T. maritima 213 SESLMKQKPLVRRFLRATAKGYEYALQHPVEAAKILLKYAPE.....LDEKI.VVESQK
S. cerevisiae206 NDEFLKKNPEKVRKFLKAIKKATDYVLADPVKAWKEYIDEKP.QLNNDLSYKQ.YQRCYA
A. fulgidus 213 RGDV...DDEKVQKFLEAYREAVREINANKEKYREKFIEIARVPEN.....

C. pneumoniae282 LLAQSQPLSKDLVDKLLLETIIKRYPELASEVAKFSLNDLYNPSPLEEQSV*~~~~~
C. trachomat 280 LLAKTQEPLSRELLESLLVTLSTTCPDLRTSIDTFSIETLIS.DVSETIASS*~~~~~
H. influenzae274 FLANK....PR....OLDAKRYQQMAEFMQQKGLIPKAL.ALKEYAVEIE~~~~~
B. halodurans282 ILLEKMEIN.PNGPFGGQDAESWEEVISWLD AHDWLEQPV.VAEDAFSSITD~~~~~
T. maritima 267 YLAGQYKA.DAEKWGYQKEEVWRYAEWLHSMGFLKEM.DVTKAFTNEFLP*~~~~~
S. cerevisiae264 YFSSSLYN.VH.....RDWKKVTGYGKRLATLPP..DYVSNYTNEYLSWPEPEEVSD
A. fulgidus 256 .LAGSYPIPDYPEPQQLPKEYYDRYLEWAMQRGLIEKVPVYEEAIR*~~~~~

C. pneumoniae333 ~~~~~~
C. trachomat 331 ~~~~~~
H. influenzae315 ~~~~~~
B. halodurans331 ~~~~~~
T. maritima 317 ~~~~~~
S. cerevisiae313 PLEAQRMAIHQEKCRQEGTFKRLALPA*
A. fulgidus 301 ~~~~~~

```

Figure 2.1 Alignment of *S. cerevisiae* Thi5p with several homologues of the bacteria and archaea identified as described in Table 2.1. Alignments were created using default settings on PILEUP (see Materials and Methods). Similar residues are indicated by grey shading, identical residues that are the most conserved between all

2.2.2 Fungal Thi5p homologues.

Numerous fungal homologues of *S. cerevisiae* Thi5p have been recovered from BLASTP searches of the Genbank database. Additionally, a French-based consortium, called Génolevures, have recently partially sequenced *THI5* homologues in a number of hemiascomycetes (Souciet *et al.*, 2000), <http://cbi.labri.u-bordeaux.fr/Génolevures/Genolevures.php3>). These include *Kluyveromyces marxianus*, *Pichia farinosa* and *Zygosaccharomyces rouxii*. Apart from *S. cerevisiae*, none of these yeasts or the other fungi listed in Figure 2.2 have been shown to possess more than one copy of the *THI5* homologue. A more detailed analysis of the copy number among the yeasts will be described in Chapter 3.

Figure 2.2 shows alignments of a selection of the fungal homologues and includes the *K. lactis* and *S. kluyveri* sequences from Chapter 4. A striking feature of these proteins is that they are very highly conserved. The basidiomycete *Uromyces fabae* possesses a Thi5p homologue that is 60% identical (68% similar) to that of the ascomycete *S. cerevisiae*. The closest relative to *S. cerevisiae* is *S. kluyveri* which exhibits 85% identity over their respective Thi5 proteins. All four *S. cerevisiae* gene products Thi5p, Thi11p, Thi12p and Thi13p have no recognisable motifs that are featured in the PROSITE dictionary (release 16.0, GCG Wisconsin™ package). There is a predicted transmembrane domain (amino acids 190-206) and two repeats KRLAI/_LP (284-290, 334-340) identified by the PSORT server (<http://psort.nibb.ac.jp/>) and MIPS server (<http://mips.gsf.de/proj/yeast/CYGD/db/index.html>), respectively. The membrane domain is highly conserved and is identical between most of the organisms featured in Figure 2.2. In comparison, the repeats show a much smaller degree of conservation and are therefore of less significance. Figure 2.3 shows a similarity plot of the fungal alignment of Figure 2.2. The dashed line represents the average similarity of the full length peptides whilst the plot itself compares the degree of similarity at each amino acid position. Highly conserved residues are

<i>A. parasiticus</i>	1	MSTDKITFLTNWHATPYHAPVYLAQSKGYFKEEGKVAILEPNPSDVTETIGSGKVDMGFKAMIHTLAA
<i>B. fuckeliana</i>	1	MSTDKITFLTNWHATPYHAPVYLAQKLGYFADEGIKVAILEPNPSDVTETIGSGKVDMGFKAMIHTLAA
<i>N. crassa</i>	1	MSTDKITFLANWHATPYHAPVYLAQAKGFFKEOGIOVAILEPNPSDVTETIGSGKVDLGFKAMIHTLAA
<i>C. albicans</i>	1	MSTNKITFLLNWEAAPYHIPPVYLANIKGYFKDENLDTAILEPSNPSDVTETIGSGKVDMLKAMVHTLAA
<i>C. tropicalis</i>	1	MSTDKITFLLNWEVAPYHIPPVYLASQKGYFKEEGIDVAILEPSNPSDVTETIGSGKVDMLKAMVHTLAA
<i>S. cerevisiae</i>	1	MSTDKITFLLNWQPTPYHIPPVYLAQIKGYFKEOGIDMAILEPINPSDVTETIGSGKVDMLKAMIHTLAA
<i>S. kluyveri</i>	1	MSTDKITFLLNWQPTAYHIPPVYLAQIKGYFKEONLDTAILEPINPSDVTETIGSGKVDMLKAMIHTLAA
<i>K. lactis</i>	1	MSSDKITFLLNWQAAPYHIPPVYLASMKGYFKEOGIDVAILEPSNPSDVTETIGSGKVDMLKAMIHTLAA
<i>S. pombe</i>	1	MSTNKITFLTNWEATPYHIPPVYLAQTRGYVEREGIEVAILEPINPSDVTALIGSGKVDMLKAMIHTLAA
<i>U. fabae</i>	1	MSTDKISVLLNWHATPYHIPPVYLAQSKGFFAKEGIKVAILEPNPSDVTETIGSGKADLCKKAMIHTLAA

<i>A. parasiticus</i>	71	KARNFPVTSIGSLLDEPFTGVVYLKDSGITEDFRSLKGKIGYVGEFGKIQIDELTKYVGMTADDYTAVR
<i>B. fuckeliana</i>	71	KARDFPVTSIGSLLDEPFTGVVYLKDSGITIDFRSLKGKIGYVGEFGKIQIDELTKYYGMEPTDYEAVR
<i>N. crassa</i>	71	KARNYPVLSIGSLLDEPFTGVVYLKESGITUDFRSLKGKRIGYVGEXGKIQIDELTSHYGLTPADYTAVR
<i>C. albicans</i>	71	KARGLPVTSIGSLLDEPFTGICYLEGSGGITSDFQSLKGKRIGNVGEFGKIQIDELTKHYGMTPDYVAVR
<i>C. tropicalis</i>	71	KARGYPVTSIGSLLDEPFTGICYLEGSGGITSDFQSLKGKRIGYVGEFGKIQIDELTKHYGMTPDYVAVR
<i>S. cerevisiae</i>	71	KARGFPVTSVASSLLDEPFTGVVYLKSGGITEDFQSLKGKIGYVGEFGKIQIDELTKHYGMKPEDYTAVR
<i>S. kluyveri</i>	71	KARGFPVTSVASSLLDEPFTGIVYLKNSGITEDFQSLKGKRIGYVGEFXXKIQIDELTKHYGMTPDYVAVR
<i>K. lactis</i>	71	KARGYDVTSIASSLLDEPFTGVVYLEGSGGITQDFNSLKGKIGYVGEFGKIQIDELTKHYGMKPTDYEAVR
<i>S. pombe</i>	71	KARGYPVTSEGSLLDEPFTGIVTLKGNGI.NDFKDKGKRIGYVGEFGKIQIDDLCSKEGLSPSDYTAIR
<i>U. fabae</i>	71	KARGFPVTSIGSLLDEPFTGVVYLEGSGGITSDFRSLKGKRIGYVGEFGKIQIDELTKYYGMTSKDYTAVR

<i>A. parasiticus</i>	141	CGMNVTKAIIIRGIDIDAGIGLENVQMVLAEWLASQNRPRDDVQIVRIDQLAELGCCCFCSSILYIANDAFI
<i>B. fuckeliana</i>	141	CGMNVSKAIIIEGSIHAGIGLENVQMVLEEWLANQGRPRSDVQMLRIDELAELGCCCFCSSILYIANDAFI
<i>N. crassa</i>	141	CGMNVSKAIIKGEIDAGIGLENFQMVLEEWLASQGRDKADVQMLRIDELAELGCCCFCSSILYICNETFI
<i>C. albicans</i>	141	CGMNVAKYIIIEGTIDCGIGIECTQQVELEEALKEQKDSNDKMLRIDKLAELGCCCFCSTILYIANDKFI
<i>C. tropicalis</i>	141	CGMNVAKYIIIEGQIDCGIGIECTQQVELEEALKKQKDPNGAKMLRIDKLAELGCCCFCSTILYIANDKFI
<i>S. cerevisiae</i>	141	CGMNVAKYIIIEGKIDAGIGIECMQQVELEEYLAKQGRPASDAKMLRIDKLACLGCCCFCSTILYICNDEFL
<i>S. kluyveri</i>	141	CGMNVCKYIIIEGKIHAGIGIECTQQVELEEYCKAQGRPVTDVQMLRIDKLACLGCCCFCSTILYICNDKFL
<i>K. lactis</i>	141	CGMNVAKYIIIEGKIDAGIGIECMQQVLEELKEKGRDPKDAKMLRIDELACLGCCCFCSTILYIANDKFL
<i>S. pombe</i>	140	CGMNVAPAIINGEIDGIGIECMQQVELERWCVSQGRPRSDVQMLRIDRLANLGCCCFCSTILYIAHDEFT
<i>U. fabae</i>	141	CGMNVSKAIIIEGTIDAGIGLENVQMVLEEWCKANNRPASDVKMLRIDELAELGCCCFCSSILYIANEDWL

<i>A. parasiticus</i>	211	AANPEKVQKFMRAVKKATDYVLAEPAAAFEEYVDMKPIMGTPVNRKIFERSAYFSDRLKNVSRDWAKVT
<i>B. fuckeliana</i>	211	AQNPEKVRKFMRAVKKATDHVLGSPSDAYKTYIDIKPEMASPVNRKIFERSAYFSKDLKNVSRDWEKVT
<i>N. crassa</i>	211	SQNPDEVRAFMRAVKKATDFVLANPDAAWKEYVDFKPVMGTELNRKIFERSAYFSEDLKNVERDWSKVT
<i>C. albicans</i>	211	AENSQAVKKFLKATKATDYMLAHPREAWAEYGNFKPTMTDLNKKFQRCYAYFSESLLNVHRDWRKVN
<i>C. tropicalis</i>	211	AENPKKKFLKAVKKATDYMLANPKQAWAEYGNFKPNMOSELNKKFSRCYAYFSDSLNVHRDWRKVN
<i>S. cerevisiae</i>	211	KKNPEKVRKFLKATKATDYVLADPVKAWKEYIDFKPQLNNLSYKQYQRCYAYFSSSLNVHRDWRKVT
<i>S. kluyveri</i>	211	AENGEKVRKFLTAIKKATDYVLNEPEKAWAEYIDFKPELNNLKYKQYQRCYAYFSSSLNVHRDWRKVT
<i>K. lactis</i>	211	AENPVKVKKFLNAIKKATDLVLQDPEQAWKYVDFKPOLNDPLSYKQYQRCYAYFSSSLNVHRDWRKVT
<i>S. pombe</i>	210	AKHPDKKKAFLRAHHSATLDMLKDPVQTYKEYTHFKREMGSELHREQFERCEYFSDHISNVPRDWNKVT
<i>U. fabae</i>	211	KCHPKETAAFMRAVKACADAMFADPRGSWAECVKVPAMNTPLNRIMFDRSENYMSQDLINVS RDWNKVT

<i>A. parasiticus</i>	281	NYGKRLGILDADFQPNYTNQYLSWTLDA...SDPLGDKRMALQKEVACEGG.FKRLQVASSA*~~~~
<i>B. fuckeliana</i>	281	KYGKRLGVINEDFSPNYTNEFLWKLEGD...SADPTGDKRMVVLQEEVADNGG.FKRLDVKLAA*~~~~
<i>N. crassa</i>	281	KYGQRLGVLSPDFTPNYTNXFLDWALQPE...SSDPTGXQKRMVELQKNVACCGG.FHRLPLWDQA*~~~~
<i>C. albicans</i>	281	NYGKRLDILPENYPNYTNEYLSWPEPKE...VDDPEKAQDLMLKHQEECKTCGG.YKRLVLA*~~~~~
<i>C. tropicalis</i>	281	NYGKRLDILPADYPNYTNEYLSWPEPQE...VSDPLKAQELMAKHQEECKTCGG.YKRLVLSGI*~~~~~
<i>S. cerevisiae</i>	281	GYGKRLAILPPDYVSNYTNEYLSWPEPEE...VSDPLEAQRMAIHQEKCRQEGT.FKRLALPA*~~~~~
<i>S. kluyveri</i>	281	GYGKRLDILPPDYQANYTNEYLSWPEPSE...VSDPLEAQRMLGVHQEKCRDCG.GFNRLDVSAI*~~~~~
<i>K. lactis</i>	281	AYGKRLDILPTDFKSNYTNEYLSWEEPT...VDDPLAAQRLLKAHQEACRDQGGFFKRVELK*~~~~~
<i>S. pombe</i>	280	NYSKRLGILPDFEPNCTNGYLFWELDPDEK...DPMCKQEAIAEIQDEIKQKGGVFSGNSLRYVEPANL*
<i>U. fabae</i>	281	NYSKRLGIVPDFVSNYTNFVQWEVAPEGGEAEGLAQOQEMKTLOAHVAEHGG.VLQAVTA*~~~~~

Figure 2.2 Alignment of several fungal Thi5p homologues. The proteins applied to the alignment are those retrieved from a BLASTP search using *S. cerevisiae* Thi5p as query with the exception of the *S. kluyveri* and *K. lactis* sequences which were obtained as part of this study (Chapter 4). The location of a predicted trans-

represented as peaks. The peak height is representative of the degree of conservation between all the fungal sequences. Peaks above the dashed line identify the regions of the proteins that are the most similar. In Figure 2.3 the highest peak is seen at the centre of the plot, position 190-206, which corresponds to the predicted transmembrane domain described above. The amino acid residues N-terminal to this domain are generally more conserved between the various homologues than the amino acid residues of the C-terminal region. The most key residues for enzyme function may therefore reside in the N-terminal portion of the Thi5 protein.

2.3 Analysis of the genome environment of the *S. cerevisiae* *THI5* gene family.

2.3.1 Comparison of the *THI5*, *THI11*, *THI12* and *THI13* promoters in *S. cerevisiae*.

The four reading frames represented by the *THI5*, *THI11*, *THI12* and *THI13* genes are very highly conserved. Over their entire length (1023 bp) there are just twenty seven nucleotide differences between the four genes (Chapter 1, section 1.5.2). Twenty four of these differences represent synonymous codon changes. This indicates that there is a selection pressure working against variability in the encoded proteins.

A comparison of the promoter regions of the *THI5* gene family has been carried out. Figure 2.4 displays an alignment of the DNA sequences that are found upstream of the four genes. Highlighted in black are those bases that are identical between the four sequences whereas bases that are shared by two or three sequences are highlighted grey. From the alignment it can be seen that the high sequence conservation between the four genes appears to extend upstream to approximately 400 bases into the promoter regions. The degree of total identity then declines with increasing distance from the -400 region whereby identity between two sequences for any given nucleotide is mostly observed. Upon close examination of the sequence data, it can be seen that there is a conservation bias. Outside the blocks of total identity (black highlight), sequence conservation is seen between chromosomes

PLOTSIMILARITY of: pileup.msf{*} 1 to 350
Window: 10 October 25, 2001 20:54

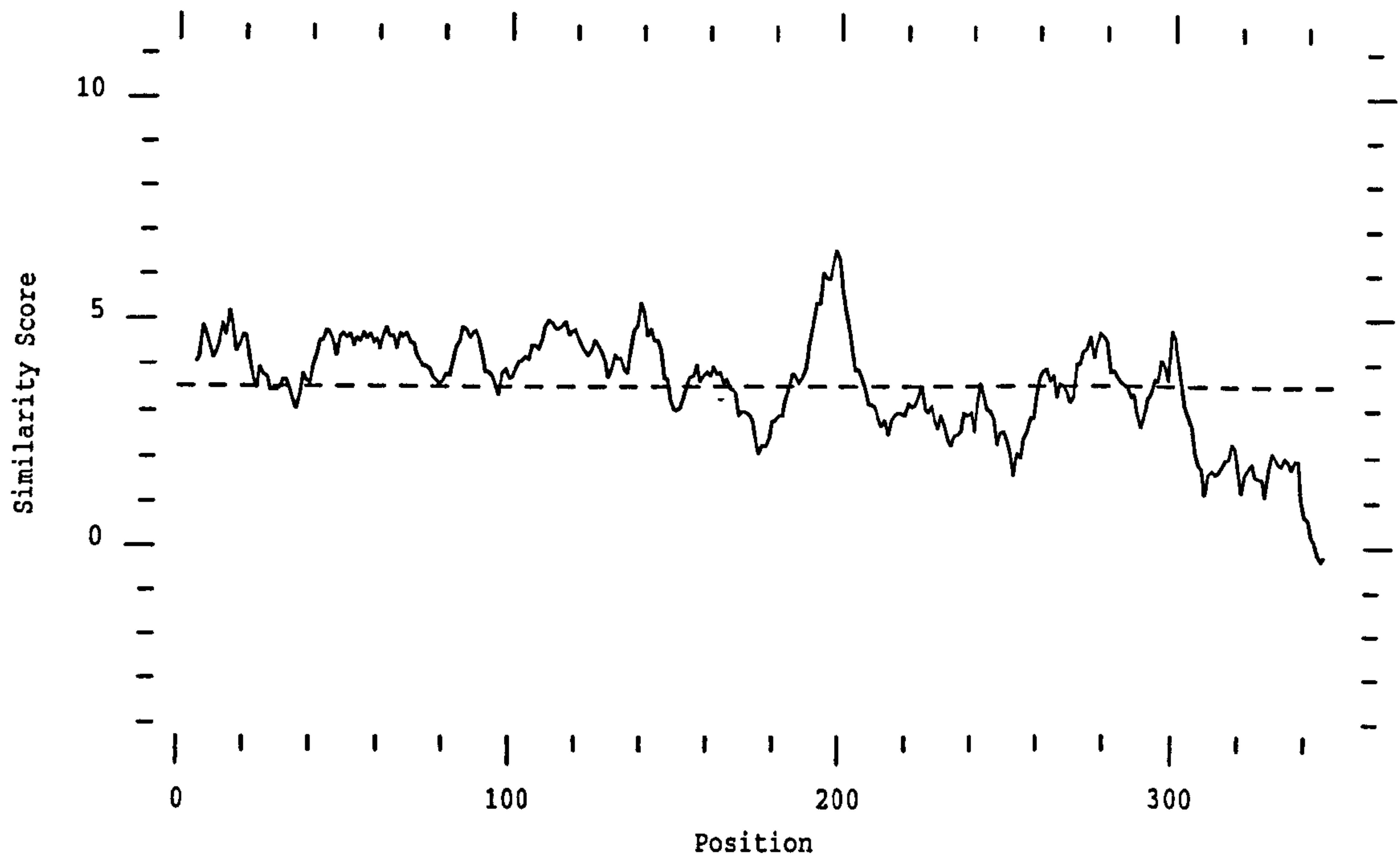


Figure 2.3 Similarity plot of figure 2.2 created using default settings of PLOTSIMILARITY (GCG Wisconsin Package™). 'Position' is the amino acid residue of the aligned protein sequences from 1 (N-terminal residue) to 350 (C-terminal residue). For explanation of plot see text.

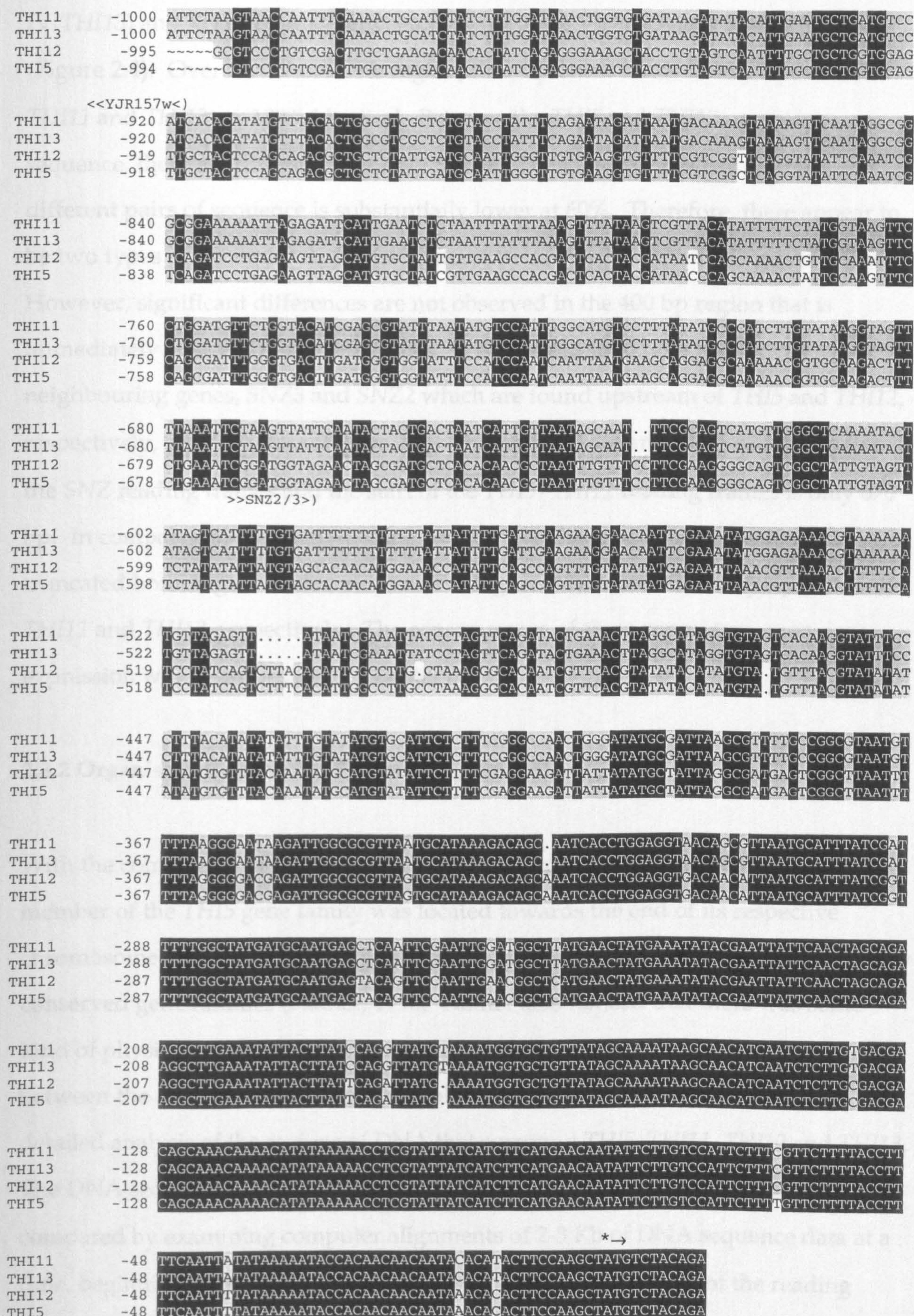


Figure 2.4 Alignment of the promoters of the *THI5* gene family in *S. cerevisiae* created using PILEUP. 1 Kb of DNA sequence is covered by the alignment. An asterisk (*) indicates the position of the translational ATG start site of the four genes.

IV(*THI13*) and X(*THI11*) and between chromosomes VI(*THI5*) and XIV(*THI12*) (Figure 2.4). Over the whole 1 Kb region, the sequences immediately upstream of *THI11* and *THI13* are 100% identical. Between the *THI5* and *THI12* upstream sequence, there are differences at only 8 sites. The identity between the two different pairs of sequence is substantially lower at 60%. Therefore, there appear to be two types of promoter that exist adjacent to members of the *THI5* gene family. However, significant differences are not observed in the 400 bp region that is immediately adjacent to the four genes. It is interesting to note that the neighbouring genes, *SNZ3* and *SNZ2* which are found upstream of *THI5* and *THI12*, respectively, are very close (Figure 2.4). The physical distance between the end of the *SNZ* reading frames and the start of the *THI5/THI12* reading frames is only 670 bp. In comparison, the small reading frame *YJR157w* on chromosome X and its truncated homologue on chromosome IV (Figure 2.5), is found 920 bp upstream of *THI11* and *THI13*, respectively. The consequences of these promoters upon gene expression will be addressed in Chapter 6.

2.3.2 Organisation of the *THI5* genes in *S. cerevisiae*.

With the completion of the *S. cerevisiae* genome project it was noticed that each member of the *THI5* gene family was located towards the end of its respective chromosome and neighbouring open reading frames were themselves part of highly conserved gene families (Hather, 1996). Hather also noticed that there was some kind of physical organisation of these gene families which was partly shared between the four chromosomes. To explore this further, I have undertaken a detailed analysis of the regions of DNA that surround *THI5*, *THI11*, *THI12* and *THI13*. The DNA sequences both upstream and downstream of each of the four genes were compared by examining computer alignments of 2-3 Kb of DNA sequence data at a time, beginning with the sequences immediately adjacent to each of the reading frames. Any significant open reading frames were identified by their gene names or systematic codes from the *Saccharomyces* Genome Database (<http://genome-www.stanford.edu/>

Figure 2.5 Physical map illustrating the duplicated regions that neighbour each of the four members of the *THI5* gene family. The gene block highlighted in blue is the sugar metabolism gene cluster and highlighted in yellow is the pyridoxine gene cluster (see text). Each cluster is duplicated once and the percentage identities at the DNA level (which includes the intergenic regions) is shown. The insertion of either of the two clusters has resulted in discrete boundaries of DNA sequence conservation. These ‘translocation points’ have been included in the diagram and their SGD co-ordinates are listed below;

Chromosome IV, 2939 and 15804.

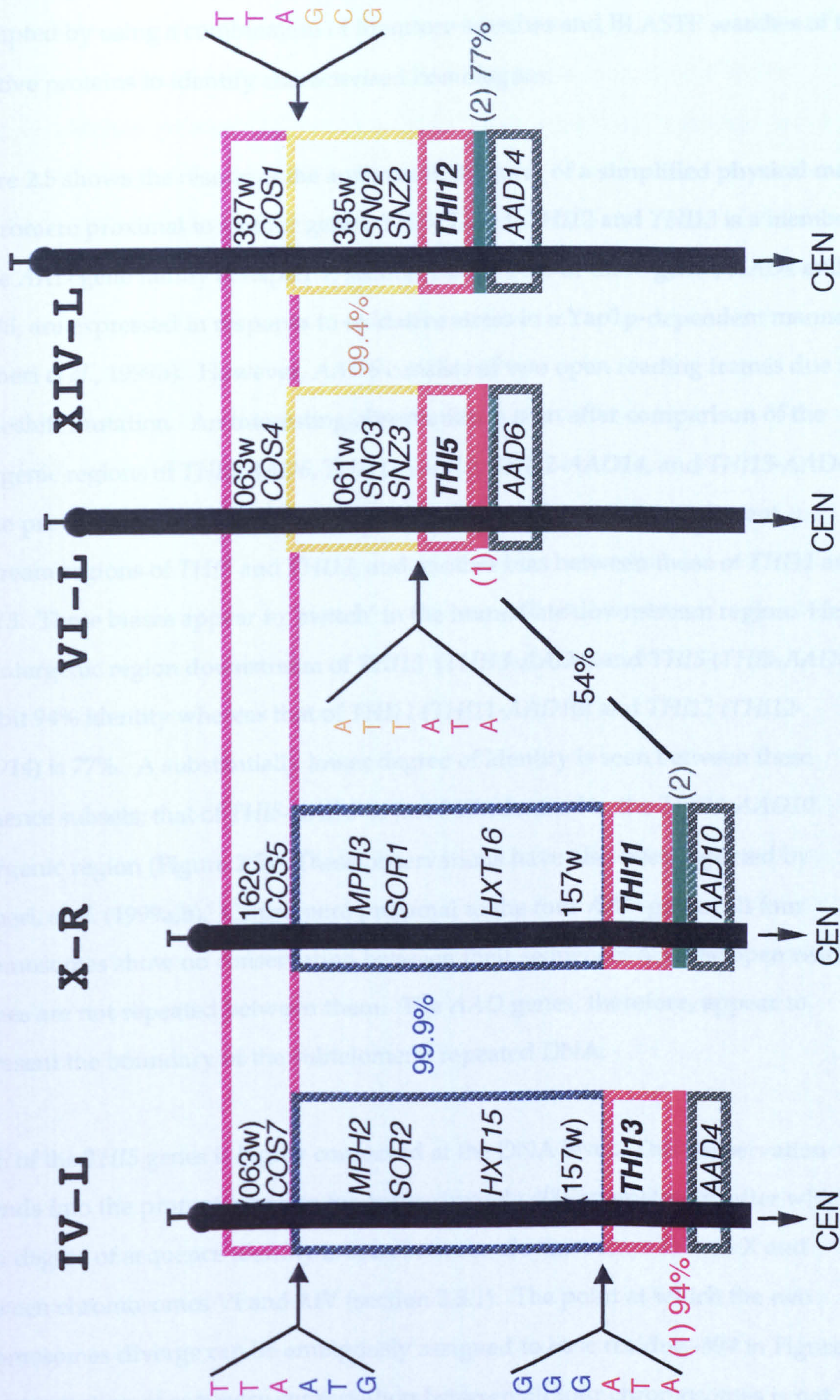
Chromosome VI, 7557 and 12529.

Chromosome X, 742551 and 729685.

Chromosome XIV, 9467 and 14432.

Also shown is the percentage identities between the *AAD-THI* intergenic regions which are labelled (1) *AAD4-THI13*, *AAD6-THI5* and (2) *AAD10-THI11*, *AAD14-THI12*.

note: The homologue of YJR157w on chromosome IV has no allocated systematic code on the *Saccharomyces* Genome Database. It is referred to as (157w) in the figure.



Saccharomyces/). A functional assignment of each open reading frame was attempted by using a combination of literature searches and BLASTP searches of the putative proteins to identify characterised homologues.

Figure 2.5 shows the results of the analyses in the form of a simplified physical map. Centromere proximal to all four genes, *THI5*, *THI11*, *THI12* and *THI13* is a member of the *AAD* gene family (Chapter 1, section 1.2.4). Two of these genes, *AAD4* and *AAD6*, are expressed in response to oxidative stress in a Yap1p-dependent manner (Delneri *et al.*, 1999b). However, *AAD6* consists of two open reading frames due to a frameshift mutation. An interesting observation is seen after comparison of the intergenic regions of *THI5-AAD6*, *THI11-AAD10*, *THI12-AAD14*, and *THI13-AAD4*. In the previous section, I reported a conservation bias between the adjacent upstream regions of *THI5* and *THI12*, and another bias between those of *THI11* and *THI13*. These biases appear to ‘switch’ in the immediate downstream region. Here, the intergenic region downstream of *THI13* (*THI13-AAD4*) and *THI5* (*THI5-AAD6*) exhibit 94% identity whereas that of *THI11* (*THI11-AAD10*) and *THI12* (*THI12-AAD14*) is 77%. A substantially lower degree of identity is seen between these sequence subsets; that of *THI5-AAD6* is just 54% identical to the *THI11-AAD10* intergenic region (Figure 2.5). These observations have also been reported by Delneri, *et al.* (1999a,b). Centromere proximal to the four *AAD* genes, all four chromosomes show no conservation between their sequences and the open reading frames are not repeated between them. The *AAD* genes, therefore, appear to represent the boundary of the subtelomeric repeated DNA.

Each of the *THI5* genes is highly conserved at the DNA level. This conservation extends into the promoter region for approximately 400 bp upstream, after which a high degree of sequence identity is seen between chromosomes IV and X and between chromosomes VI and XIV (section 2.3.1). The point at which the two chromosomes diverge can be ambiguously assigned to base residue -399 in Figure 2.4. The resumption of sequence conservation between all four chromosomes is not observed until the codon corresponding to the final leucine residue (TTA) of each of

the COS genes. These points of divergence have been called 'translocation points' and are shown in Figure 2.5. Between these points there exists one of two duplicated blocks of genes. Each block, which includes intergenic regions, is very highly conserved with its partner. These blocks have been named due to the functional relationship between the genes that exist within them. The block of genes adjacent to *THI11* and *THI13* (Figure 2.5, blue) is the 'sugar metabolism cluster'. *HXT15* and *HXT16* encode proteins that resemble hexose transporters. Both protein products appear to have a role in the uptake of mannose, fructose and glucose, but not galactose (Wieczorke *et al.*, 1999). *SOR1* and *SOR2* appear to encode a zinc-containing alcohol dehydrogenase that looks like sorbitol dehydrogenase (Gonzalez *et al.*, 2000). Both *MPH2* and *MPH3* encode maltose transporters that are responsible for residual uptake of glucose in an *S. cerevisiae* mutant that is deleted for all the *HXT* genes (*HXT1-17*) and *GAL2*. Deletion of *MPH2*, *MPH3* and another homologue (called *AGT1*) within this mutant background results in the complete abolishment of glucose consumption (Wieczorke, *et al.*, 1999). The function of *YJR157w* and its homologue on chromosome IV is unclear.

The block of genes adjacent to *THI5* and *THI12* (Figure 2.5, yellow) is the 'pyridoxine cluster' based on the roles of *SNO2/3* and *SNZ2/3* in the biosynthesis of this vitamin (Chapter 1, section 1.6.1). The determination of a role for *YDL061w* and *YNL335w* in pyridoxine formation awaits the construction of a mutant strain containing deletions for both genes. The COS members and adjoining genes are members of much larger gene families of unknown function (Chapter 1, section 1.2.4).

2.5 Discussion.

Numerous homologues of the *S. cerevisiae* Thi5p enzyme have been identified in organisms outside of the fungal kingdom. Some of these, for example *H. influenzae*, appear to possess the protein instead of the more common ThiC enzyme. It is not known whether these organisms use their Thi5p homologue in HMP biosynthesis. However, the *H. influenzae* homologue is located adjacent to a gene that appears to

encode the HMP-P kinase. An interesting experiment would be to see if a clone of the *H. influenzae* gene can rescue the thiamin auxotrophy of an *S. pombe nmt1* mutant. It would also be useful to know what the starting substrates are in these organisms, especially in *H. influenzae*.

To date, there is no direct evidence that the fungal homologues are involved in HMP biosynthesis, including the four putative isozymes of *S. cerevisiae*. However, expression of one or more of the *S. cerevisiae* genes and *Aspergillus parasiticus nmt1* have been shown to be regulated by thiamin (Cary & Bhatnagar, 1995; Hather, 1996). To date, *S. pombe nmt1* is the only homologue which is known to be involved in HMP formation and the *U. fabae* homologue is known to rescue an *nmt1* mutant (Sohn *et al.*, 2000). Based on the present published data and the high degree of conservation, it is expected that the Thi5/*nmt1* homologues of the fungi are all HMP biosynthetic enzymes. A computer alignment of these proteins has identified a conserved putative transmembrane domain in the centre of the primary sequence.

It seems that there are two identified pathways to HMP biosynthesis which use very different enzymes and starting substrates. One is represented by the presence of Thi5 and appears to exist throughout the fungi. The other is represented by *E. coli* ThiC and occurs outside of the fungal domain. However, in prokaryotes both pathways seem to co-exist. Some organisms use one or the other pathway, others appear to possess both. A similar situation has been found for pyridoxine biosynthesis (Mittenhuber, 2001). The *SNO* and *SNZ* genes of *S. cerevisiae* have been shown in our laboratory to be involved in pyridoxine biosynthesis, whereas the *E.coli* pathway is different and involves the *Pdx* genes (Marsh, 2000). *SNO* and *SNZ*, but not the *Pdx* genes, are distributed throughout the fungi. Like *THI5*, they are also found among some bacteria either with or instead of the *Pdx* genes. However, a comparison with my own data of Table 2.1 show that *SNO/SNZ* and *THI5* are not co-distributed among the prokaryotes.

From analysis of the currently available sequence data, only *S. cerevisiae* has been shown to possess *THI5* as a gene family. Each of the four genes, *THI5*, *THI11*, *THI12* and *THI13*, are distributed in the *S. cerevisiae* genome in a non-random fashion. Each member is located within the subtelomeric portion of four different chromosomes and share these regions with other gene families. A detailed comparison between these portions has found these gene families to exhibit a high degree of physical organisation. Each chromosome end is almost identical apart from the insertion of one of two blocks of functionally related genes. The structured environment of the *THI5* gene family may not be unique to *S. cerevisiae* and is investigated further in Chapter 4. A model that illustrates the evolution of these particular subtelomeric regions will be presented in Chapter 7.

Chapter 3.

A phylogenetic analysis of *THI5* copy number among the hemiascomycetes.

3.1 Introduction

It has been discussed in the previous chapter that apart from the sequenced genome of *S. cerevisiae* S288C, no other fungus has yet been shown to possess more than one copy of a *THI5* homologue. S288C is just one strain of many that are classified as *S. cerevisiae* and is derived from several strains, EM93, EM126, NRRL-210 and the commercial baking yeasts FLD, LK and American Yeast Foam. Of these one in particular, EM93, is thought to have donated 88% of the S288C genome (Mortimer and Johnston, 1986).

The various *S. cerevisiae* strains, along with several other species including *S. bayanus*, *S. pastorianus* and *S. paradoxus*, are classified within a small subgroup of the *Saccharomyces* genus termed the *sensu stricto* complex. Recently, three geographically isolated species, two from Japan (*S. kudriavzevii*, *S. mikatae*) and one from Brazil (*S. cariocanus*) have been characterised and included within the *Saccharomyces sensu stricto* (Naumov *et al.*, 2000). In contrast to the close genome relationships of this subgroup, the *Saccharomyces sensu lato* species are not as closely related and, based on modern taxonomy methods are found to be intermixed with non-*Saccharomyces* species (James *et al.*, 1997).

This chapter describes the results of a survey that has examined the extent of the *THI5* multicopy state among a selection of such hemiascomycetes. A number of interesting and unexpected observations have led to the hypothesis explaining the genetic basis of thiamine auxotrophy among some typical yeast strains isolated from natural environments.

3.2 Survey of *THI5* copy number.

Southern blot hybridisation analysis was used to determine the copy number of *THI5* in various yeasts. The restriction endonucleases, *XhoI*, *PstI* and *EcoRI*, were chosen for the digestion of genomic DNA isolated from species of the *Saccharomyces* and *Kluyveromyces* genera. These enzymes were selected because, after Southern blot analysis of S288C, they give rise to four *THI5* hybridising fragments of different sizes with each corresponding to a different locus (*THI5*, *THI11*, *THI12* and *THI13*). Although, for several yeasts, the absolute copy number could not be determined, this approach proved to be adequate for establishing whether *THI5* is present as a single copy gene or a gene family. The actual size of each fragment for S288C and the restriction maps are shown in Figure 3.1.

3.2.1 The parental strains of *S. cerevisiae*. S288C.

Southern blot analysis was carried out on two of the S288C parental strains, EM93 and American Yeast Foam. For each restriction enzyme, multiple bands were seen upon hybridisation with a *THI5* probe (Figure 3.2a). Genomic DNA digested with *XhoI* endonuclease yielded hybridising fragments of identical sizes between S288C and EM93. However, the 14.0 Kb band of the *THI11* locus was substantially fainter for EM93 (Figure 3.2b). *PstI* digests of these two strains also results in identical restriction fragments. The first band in the EM93 *PstI* lane appeared to be a doublet corresponding to the 10.7 Kb *THI13* and 9.9 Kb *THI12* restriction fragments in S288C. Also present was the 7.4 Kb *PstI* fragment of *THI5* and a faint 4.3 Kb *THI11* band (Figure 3.2b). The 2.8 Kb *THI12*, 3.2 Kb *THI5* and 3.9 Kb *THI11* *EcoRI* restriction fragments were found to be of identical sizes in both strains. In addition was a unique 9.2 Kb *EcoRI* restriction fragment which was only visualised on an overexposed autoradiograph (Figure 3.2b). Interestingly, for each of the three EM93 restriction digests, there appeared a fifth band, which is not seen in S288C. These are the 10.4 Kb and 4.8 Kb restriction fragments of *XhoI* and *PstI* respectively, and a 6.6 Kb *EcoRI* fragment which is also present in American Yeast Foam. Taken together

Figure 3.1 Restriction maps of the S288C *THI5*, *THI11*, *THI12* and *THI13* genes including 10 Kb flanking DNA. The predicted sizes of restriction fragments that cover the region of the *THI5* probe are shown below.

Gene	<i>Xho</i> I	<i>Pst</i> I	<i>Eco</i> RI
<i>THI5</i>	3.8	7.4	3.2
<i>THI11</i>	14.0	4.3	3.9
<i>THI12</i>	10.2	9.9	2.8
<i>THI13</i>	7.0	10.7	8.5

0 10 20 Kb

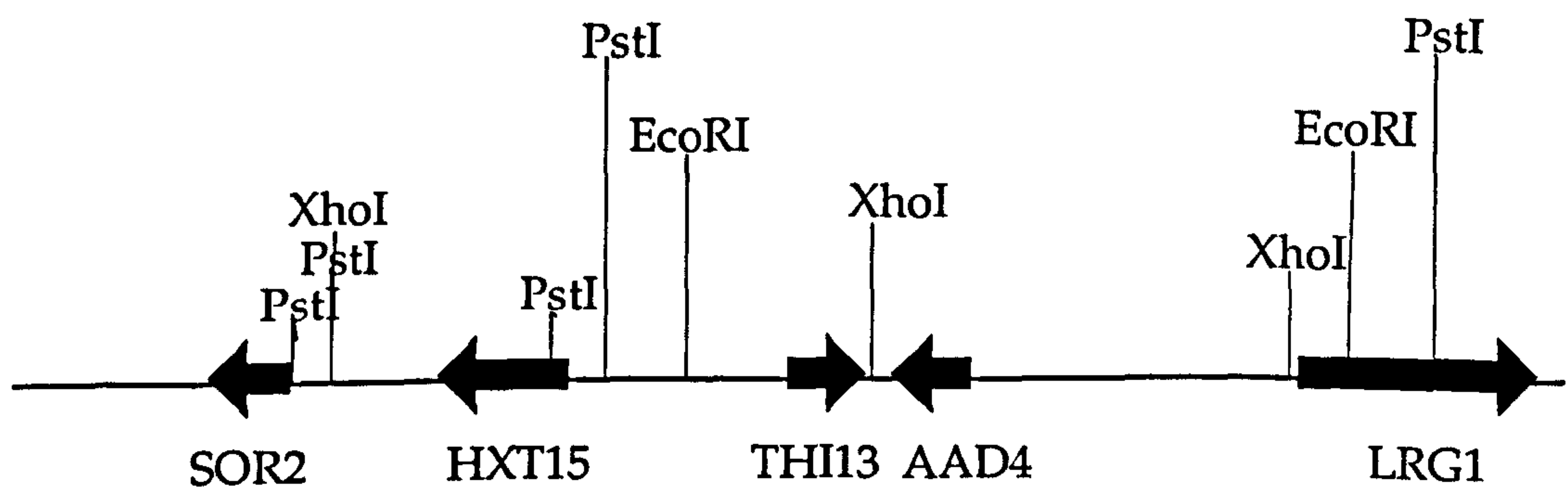
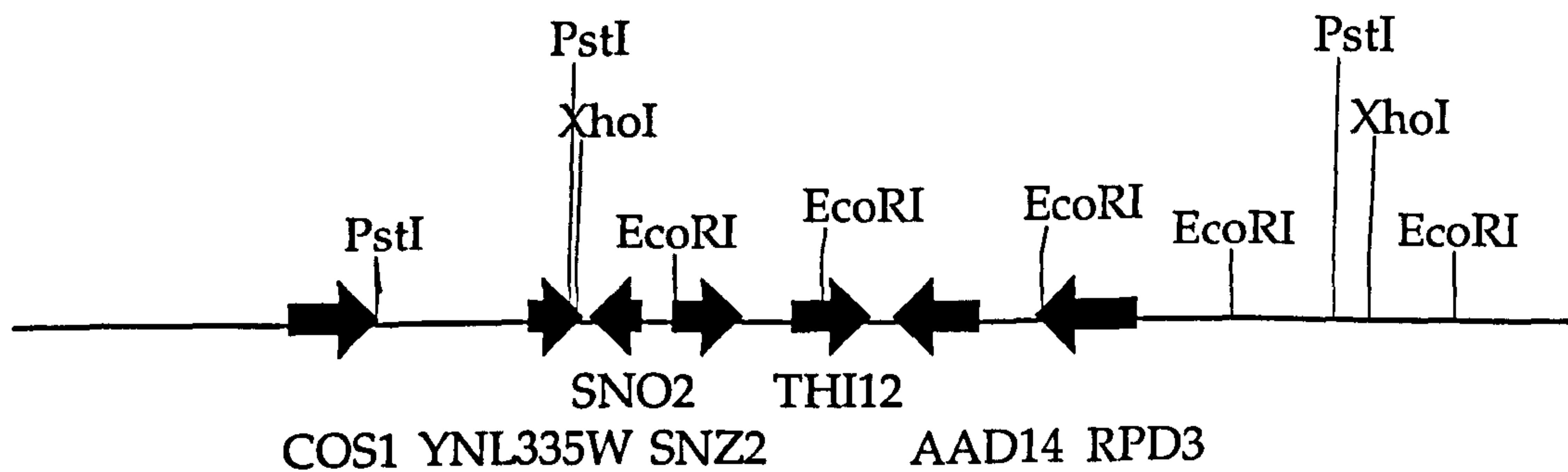
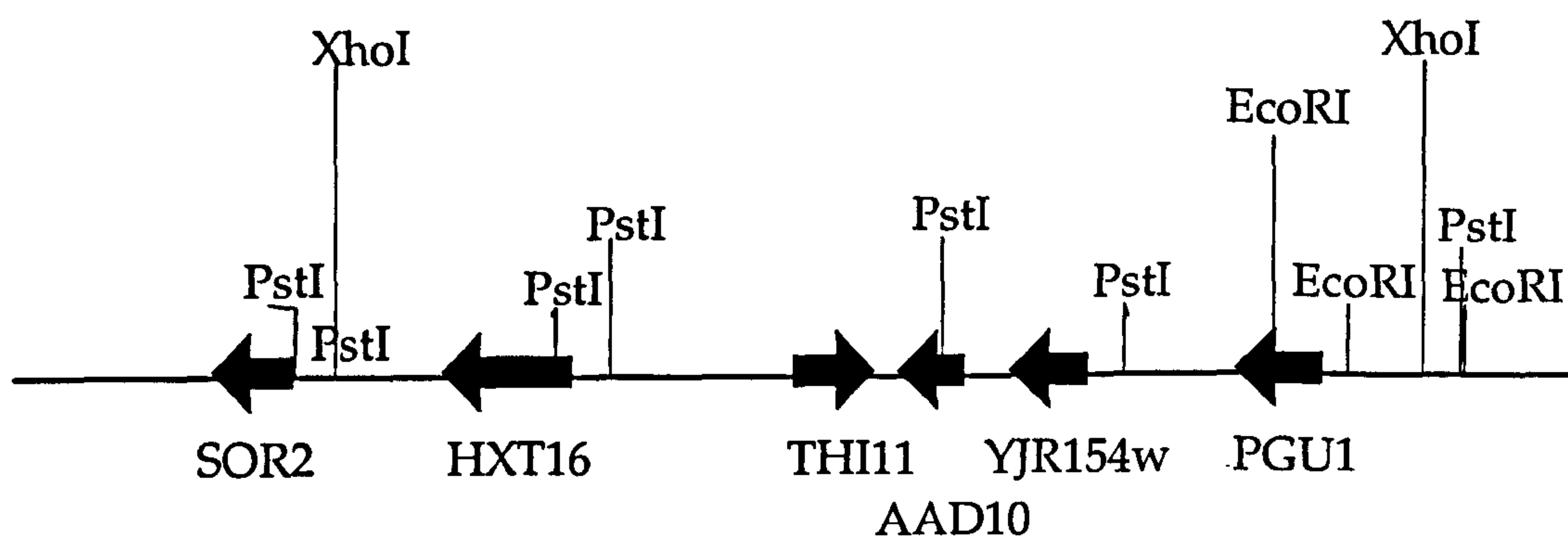
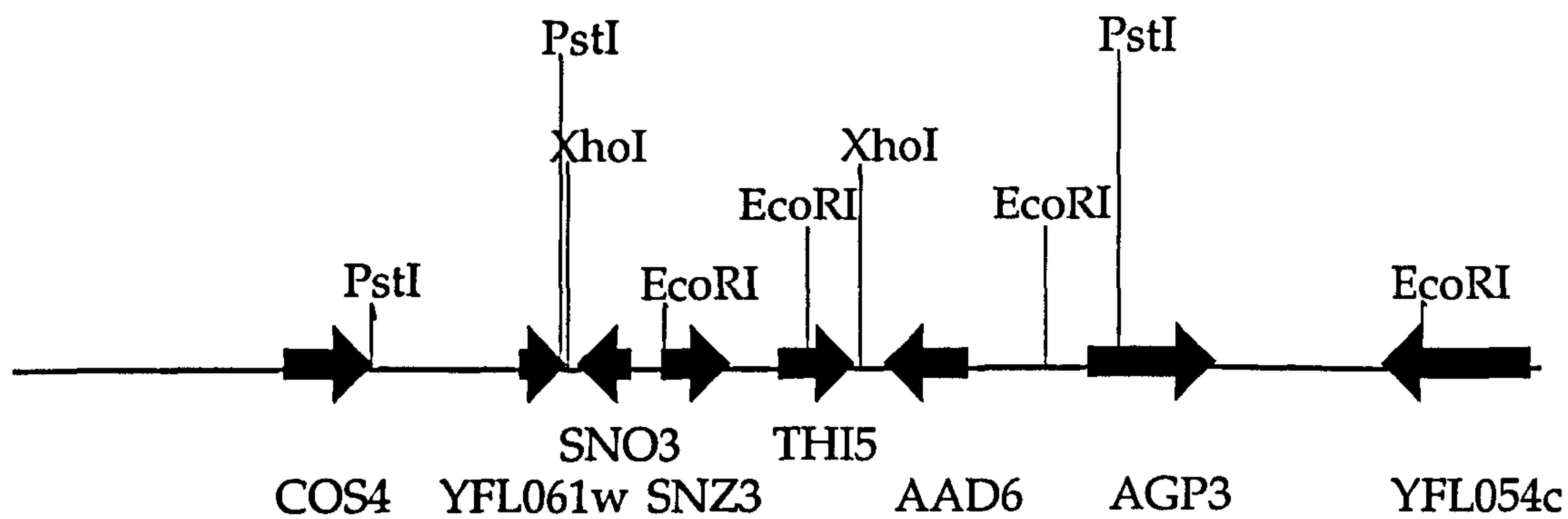
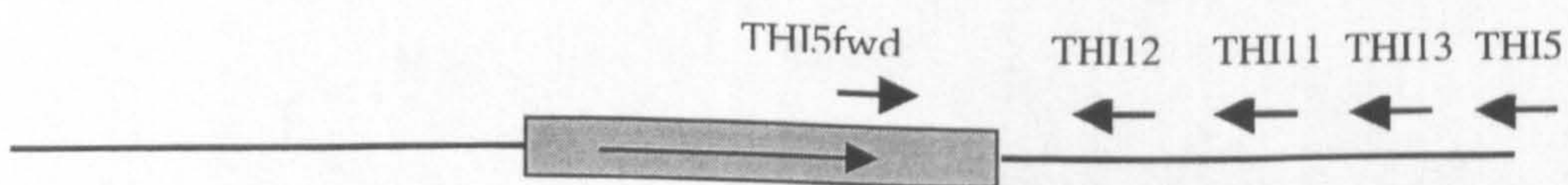
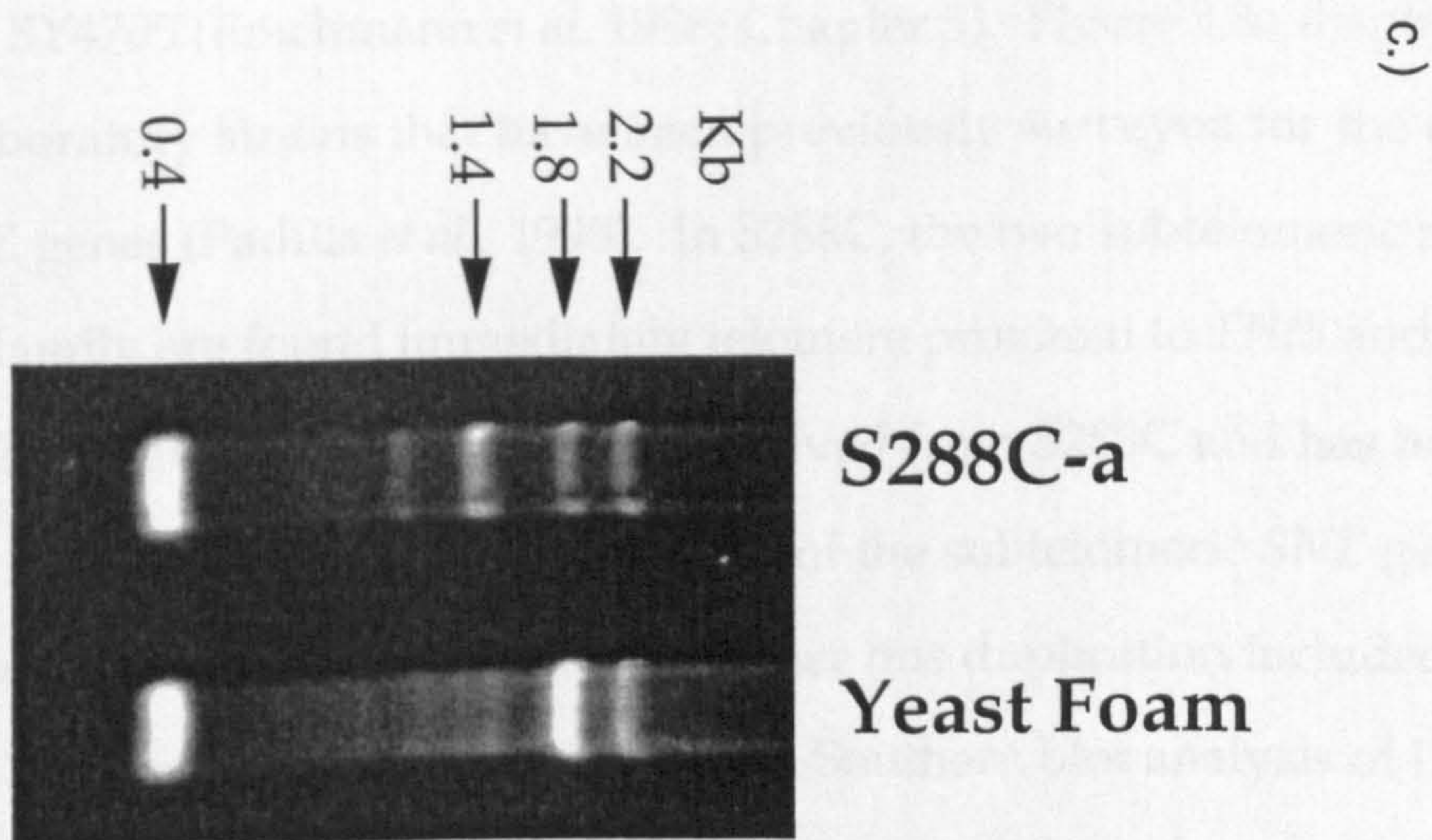
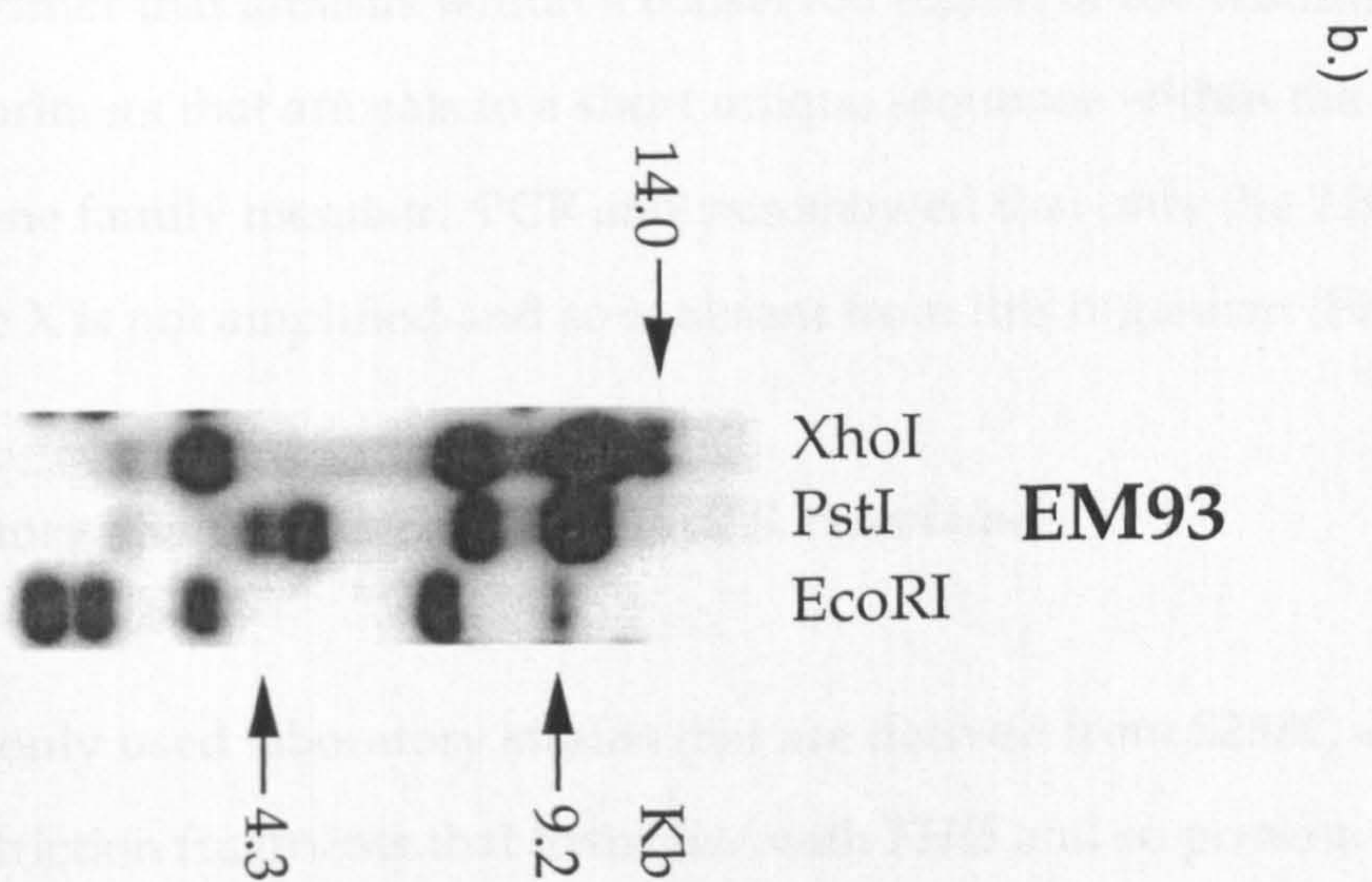
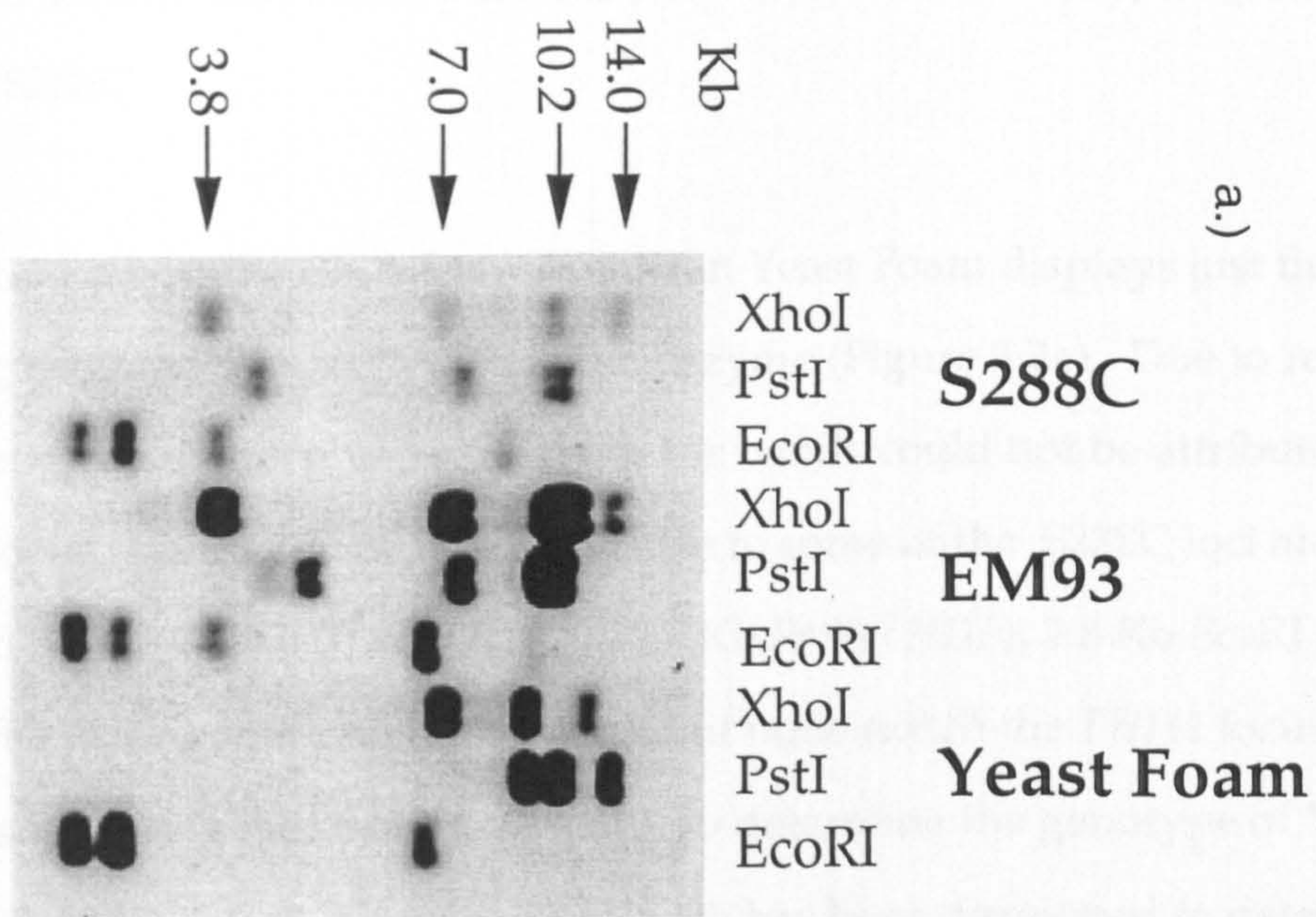


Figure 3.2 *THI5* Southern blot hybridisation and PCR analysis of ancestral strains of S288C.

- a.) EM93 (NCYC 292) and American Yeast Foam (NCYC 232). Probe = pRH4. Hyb. temp. = 65°C. Sizes of the S288C *Xho*I bands are shown.
- b.) Over-exposed autoradiograph of the EM93 Southern blot. Arrows indicate a fifth band with fainter intensity to the other four.
- c.) PCR genotyping of American Yeast foam showing the presence of *THI5* loci.
- Each locus is resolved through PCR from a common primer that anneals within the gene. The other primer is one of four oligonucleotides that anneals to a unique DNA sequence at different positions downstream of each gene family member. The expected band sizes after electrophoresis of the reaction products are 0.4 Kb (*THI12*), 1.4 Kb (*THI11*), 1.8 Kb (*THI13*) and 2.2 Kb (*THI5*).





these results suggest that EM93 possesses a five-membered family, with an extra gene to those of S288C.

The restriction digests of DNA from American Yeast Foam displays just three *THI5* hybridising fragments for each restriction enzyme (Figure 3.2a). Due to restriction fragment length polymorphisms, many of the bands could not be attributed to a particular locus. Bands corresponding in size to some of the S288C loci are represented by the 7.0 Kb *XhoI* (*THI13*), 10.7 Kb *PstI* (*THI13*), 2.8 Kb *EcoRI* (*THI12*) and 3.2 *EcoRI* (*THI5*) restriction fragments, but none match the *THI11* locus fragments suggesting this may be absent. To determine the genotype of Yeast Foam, a PCR-based approach was used which has been developed to detect the individual members of the *THI5* gene family. This involves DNA amplification between a primer that anneals within a conserved region of the reading frames and one of four primers that anneals to a short unique sequence within the 3' region of a particular gene family member. PCR analysis showed that only the *THI11* gene of chromosome X is not amplified and so is absent from this organism (Figure 3.2c).

3.2.2 Laboratory and Commercial strains of *S. cerevisiae*.

Some commonly used laboratory strains that are derived from S288C all have identical restriction fragments that hybridise with *THI5* and so possess the four genes of the sequenced genome. These are strains W303 (Corti, 1997), YPH499 (Hather, 1996) and BY4705 (Brachmann *et al.* 1998; Chapter 5). Figure 3.3a displays the results for two laboratory strains that have been previously surveyed for the copy number of the *SNZ* genes (Padilla *et al.*, 1998). In S288C, the two subtelomeric members of this gene family are found immediately telomere proximal to *THI5* and *THI12* (Chapter 2). DS10 is a laboratory strain derived from S288C and has been shown by CHEF gel analysis to possess an extra copy of the subtelomeric *SNZ* gene on chromosome II. I decided to examine whether this duplication included the adjoining *THI5* and *THI12* genes. However, Southern blot analysis of DS10 has been inconclusive as each restriction digest only displayed those fragments seen for S288C

Figure 3.3 Southern blot hybridisation analysis to examine *THI5* copy number of various *S. cerevisiae* strains.

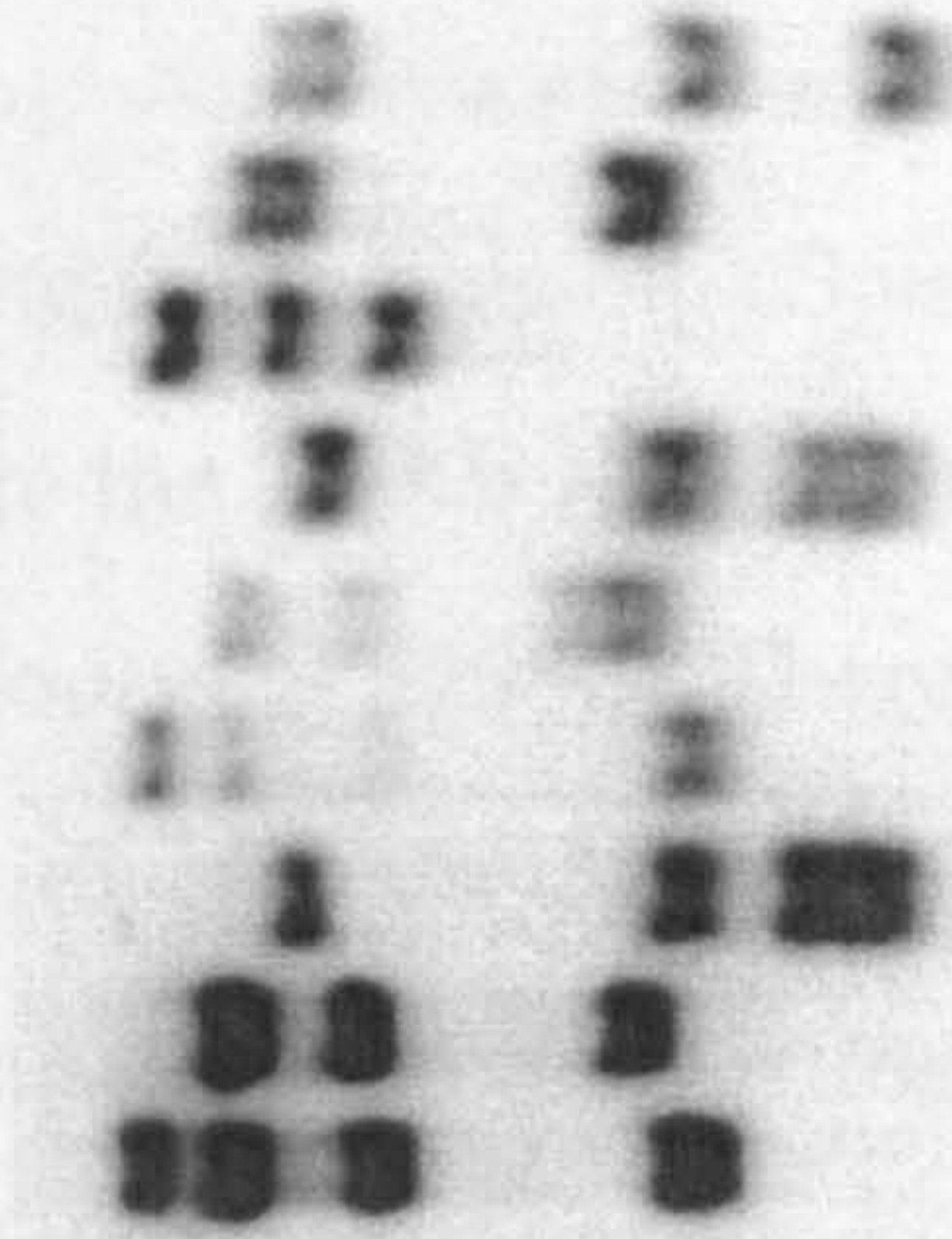
- a.) Two laboratory strains, DS10 and MW1076 pseudohyphal haploid (Padilla *et al.* 1998). Probe = pRH4. Hyb. temp. = 65°C.
- b.) *S. cerevisiae* 'type' strain (CBS1171), Ale (NCYC 1681) and Lager (NCYC 1324). Probe = 620bp ESP30 *Eco*RI fragment of pUP27.1 (covers C-terminal portion of *THI5* and 204 bp of 3' untranslated sequence). Hybridisation temp. = 65°C.
- c.) Yeast isolated from a bottle of 'Fullers 1845™' beer. Probe = 677bp *Cla*I/*Xho*I fragment of pRH4 (covers *THI5* coding sequence only). Hyb. temp. = 65°C.

a.)

S288C
XhoI
PstI
EcoRI

DS10
XhoI
PstI
EcoRI

Pseudohyphal
XhoI
PstI
EcoRI



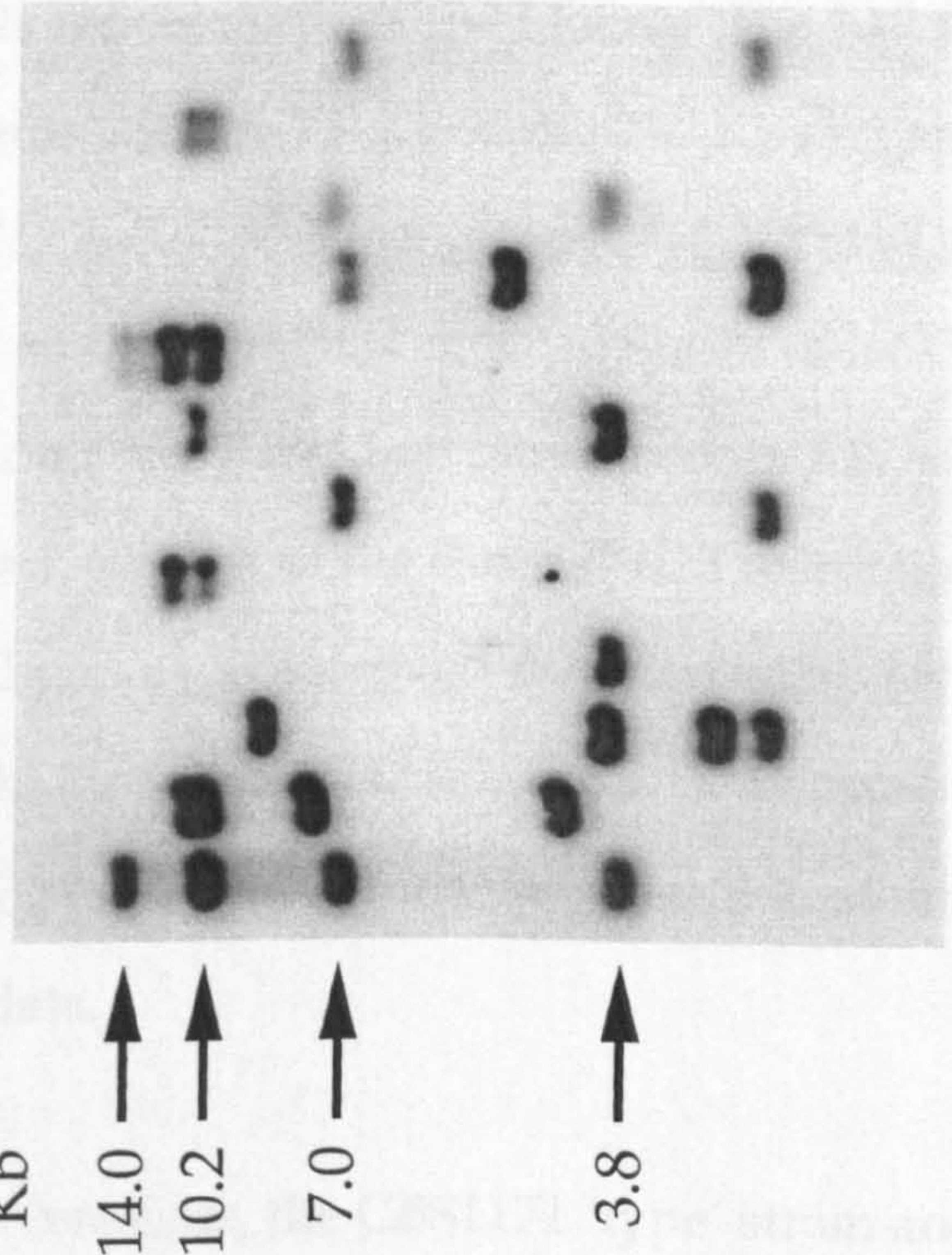
b.)

S288C
XhoI
PstI
EcoRI

CBS1171
XhoI
PstI
EcoRI

Alc NCYC 1681
XhoI
PstI
EcoRI

Lager NCYC 1324
XhoI
PstI
EcoRI



c.)

S288C
XhoI
PstI
EcoRI

Fullers 1845
XhoI
PstI
EcoRI



after hybridisation with the *THI5* probe. This approach relied on restriction fragment length polymorphisms to distinguish a possible fifth member. Such polymorphisms will not likely be present for very recent duplications, therefore, only through a blot of whole chromosomal separations would a fifth gene be resolved.

The second strain, MW1076, is a laboratory strain derived from Σ 1278b that shows pseudohyphal growth (Liu, *et al.*, 1996). It has been reported by Padilla *et al.* that this strain lacks both subtelomeric copies of *SNZ*. Figure 3.3a shows that each restriction digest yielded three hybridising fragments. Based on S288C, the 14.0 Kb *Xho*I, 4.3 Kb *Pst*I and 3.9 Kb *Eco*RI restriction fragments represent the *THI11* locus. The 10.2 Kb *Xho*I, 9.9 Kb *Pst*I and 2.8 Kb *Eco*RI fragments and the 7.0 Kb *Xho*I, 10.7 Kb *Pst*I and 8.5 Kb *Eco*RI fragments represent the *THI12* and *THI13* loci respectively. There were no fragments corresponding to *THI5* and so this gene has probably been deleted along with the *SNZ* member on chromosome VI. *THI12* on chromosome XIV was found to be present in this strain. However, as each of the three *THI12* restriction fragments encompasses the adjacent *SNZ* gene, the observed fragment sizes for *THI12* would be expected to be different to those of S288C (Figure 3.1). Because these fragments migrated identically, *SNZ2* (and *SNO2*) must be present in strain MW1076, so contradicting the published data.

A survey of three commercial strains of *S. cerevisiae*, the CBS1171 'type' strain and two NCYC strains (1681, 1324) used in the production of ales and lagers (formerly *S. uvarum*; Goodey & Tubb (1982)), illustrates that *THI5* is present as a gene family but with fewer copies than S288C (Figure 3.3b). All three strains possess the 10.7 Kb (*THI13*) and 9.9 Kb (*THI12*) *Pst*I hybridising fragments on a Southern blot. In addition, compared with S288C, they share the 3.8 Kb *Xho*I (*THI5*), 2.8 Kb *Eco*RI (*THI12*) and a 6.9 Kb *Eco*RI fragment not found in S288C. There was no evidence of any *THI11* restriction fragments that hybridised with the probe. As most of the DNA restriction enzyme digests yield just two positive hybridising fragments then it is likely that these yeasts possess two copies of *THI5*. The three bands of the *Eco*RI

digest of DNA from strain NCYC1681 indicates that this yeast may possess a third copy or alternatively may be due to a *EcoRI* restriction site within the probed region of one of the two genes. To determine whether the NCYC1681 ale strain accurately represents a commercial yeast used in the current production of real ales, samples from two different bottled beers were obtained. After recovery of the yeast culture from the samples, the *THI5* copy number was examined in the same way as before. Both the Fullers 1845™ (Figure 3.3c) and the Caledonian yeast (data not shown) produced identical restriction fragments to NCYC1681 on a Southern blot.

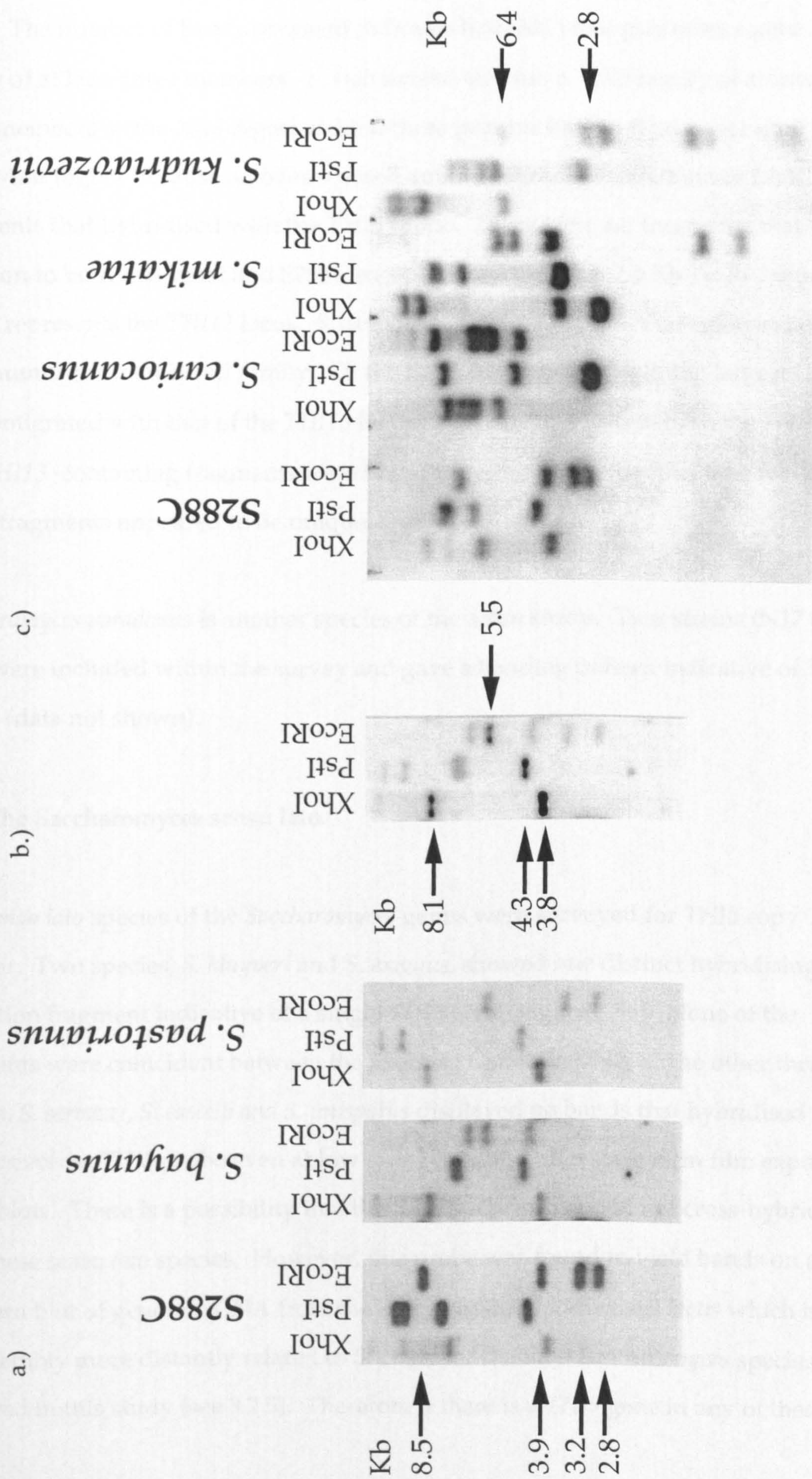
3.2.3 Other yeasts of the *Saccharomyces sensu stricto* complex.

Further Southern blot hybridisation analysis of *THI5* copy number was carried out for *Saccharomyces* species of the *sensu stricto* subgroup. The survey included *S. pastorianus*, *S. bayanus* and the recently classified yeasts *S. cariocanus*, *S. mikatae* and *S. kudriavzevii*. Both *S. bayanus* and *S. pastorianus* displayed either 2 or 3 restriction fragments for each genomic digest (Figure 3.4a). Only a minority of fragments appeared to be similar to S288C, the 3.8 Kb *XhoI* fragment of *THI5* and the 4.3 Kb *PstI* fragment of *THI11*. In addition, there was a faint *S. bayanus XhoI* band of 13.9 Kb which is almost identical to the S288C 14.0 Kb (*THI11*) fragment. For *S. pastorianus* there was the 9.9 Kb *PstI* and the 2.8 Kb *EcoRI THI12* fragments. The Southern blots of *S. bayanus* and *S. pastorianus* were electrophoresed on the same agarose gel. Each digest produced at least one fragment that is of exactly the same size and can be visualised when one autoradiograph is superimposed onto the other (Figure 3.4b). Two of the fragments, 8.1 Kb *XhoI* and 5.5 Kb *EcoRI*, were not found for S288C. These coincidental bands provide further support to the hypothesis that *S. pastorianus* is a hybrid of *S. bayanus* and a *S. cerevisiae* strain (Tamai, *et al.*, 1998).

For the yeasts that have been recently included within this subgroup, numerous restriction fragments were found to hybridise to the *THI5* probe (Figure 3.4b). Very few fragments migrated identically to those of S288C. *S. mikatae* displayed four hybridising bands for the *XhoI* digests, three for *PstI* digests and five for the *EcoRI*

Figure 3.4 *THI5* Southern blot hybridisation analysis of the *Saccharomyces sensu stricto*. Probe = pRH4.

- a.) *S. pastorianus* (CBS1538) and *S. bayanus* (CBS380). Sizes represent the S288C *EcoRI* bands. Hyb. temp. = 65°C.
- b.) Superimposed blots of *S. pastorianus* and *S. bayanus* using the overlay effect of Photoshop 5.5 with 90% opacity. Bands that migrate identically during electrophoresis have an additive dark black appearance with their sizes shown adjacent.
- c.) *S. cariocanus* (ufrJ50816), *S. mikatae* (IFO 1815) and *S. kudriavzevii* (IFO 1802). hyb. temp. = 55°C.



digests. Only the 3.8 Kb *Xho*I (*THI5*) fragment of S288C was present within this yeast. The number of bands obtained indicates that this yeast possesses a gene family of at least three members. *S. kudriavzevii* also has a *THI5* family of at least three members as the *Xho*I digest yielded three positive bands. The upper limit is somewhat higher because an overexposed autoradiograph showed seven *Eco*RI fragments that hybridised with the *THI5* probe. There were no fragments that were common to both this yeast and S288C except for possibly the 2.8 Kb *Eco*RI fragment which represents the *THI12* locus. *S. cariocanus* displayed bands that again indicates a minimum three membered family. Of the three fragments of *Pst*I, the largest (10 Kb) comigrated with that of the *THI13* locus in S288C. In addition both the *THI12*- and *THI13* -containing fragments were found to comigrate in the *Xho*I lane whilst all *Eco*RI fragments appeared to be unique.

Saccharomyces paradoxus is another species of the *sensu stricto*. Two strains (N17 and N12) were included within the survey and gave a banding pattern indicative of 5 copies (data not shown).

3.2.4 The *Saccharomyces sensu lato*.

Five *sensu lato* species of the *Saccharomyces* genus were surveyed for *THI5* copy number. Two species, *S. kluyveri* and *S. exiguus*, showed one distinct hybridising restriction fragment indicative of a single *THI5* gene (Figure 3.5a). None of the fragments were coincident between the species. Genomic DNA of the other three species, *S. servazzi*, *S. castelli* and *S. unisporus* displayed no bands that hybridised with the *S. cerevisiae* *THI5* probe even at low stringency and after long term film exposure of the blots. There is a possibility that the *S. cerevisiae* probe will not cross-hybridise with these *sensu lato* species. However, this probe was found to yield bands on a Southern blot of genomic DNA from the dairy yeast *Kluyveromyces lactis* which is considerably more distantly related to *S. cerevisiae* than the *Saccharomyces* species surveyed in this study (see 3.2.5). Therefore if there is a *THI5* gene in any of these

Figure 3.5 *THI5* Southern blot hybridisation analysis of several species of the *Saccharomyces sensu lato* and *Kluyveromyces*. Hyb. temp. = 55°C.

- a.) *S. kluyveri* (CBS3082), *S. exiguus* (CBS379) and *S. servazzi* (CBS4311). Probe = pRH4. Sizes of S288C *Xho*I bands are shown.
- b.) *K. lactis* (MSK110), *K. polysporus* (CBS2163), *K. drosophilarum* (CBS2896) and *K. thermotolerans* (CBS 6924). Probe = 800 bp *Hind*III fragment of pDW5a which covers the C-terminal portion of the *K. lactis* *KMOL2* gene (Chapter 4, figure 4.3a).
Sizes of the *K. lactis* bands are shown for each restriction digest.
- c.) The *Kluyveromyces* species probed with *S. cerevisiae THI5* (pRH4).

a.)

S288C
XhoI PstI EcoRI
S. kluyveri
XhoI PstI EcoRI
S. exiguus
XhoI PstI EcoRI
S. servazzii
XhoI PstI EcoRI

Kb
14.0
10.2
7.0
3.8

b.)

XhoI PstI EcoRI
K. lactis
XhoI PstI EcoRI
K. polysporus
XhoI PstI EcoRI
K. dros.
XhoI PstI EcoRI
K. therm.
XhoI PstI EcoRI

Kb
20.0
16.0

1.7

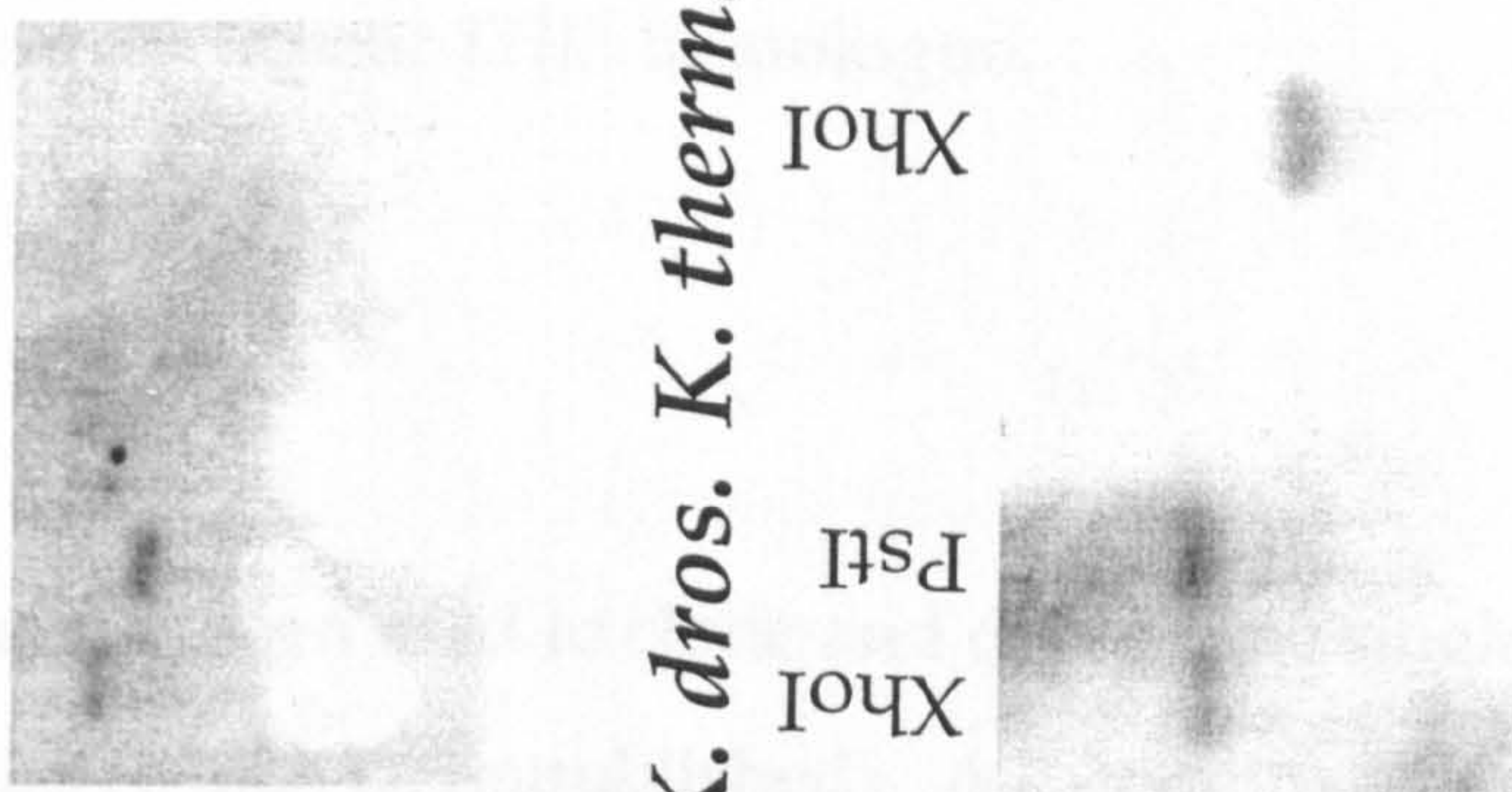
c.)

K. lactis K. poly.

XhoI PstI
XhoI PstI

K. dros. K. therm.

XhoI PstI
XhoI PstI



three species, it would be expected that the probe would have displayed positive fragments. No such fragments signifies an absent *THI5* homologue.

3.2.5 The *Kluyveromyces* genus.

MSK110 is a *K. lactis* genetic strain that has been used to clone and disrupt the single copy *THI5* homologue, termed *Kl. THI5* (Walsh D, unpublished). A C-terminal portion of *Kl. THI5* was used as a probe for a Southern blot of the *Kluyveromyces* species *K. drosophilarum*, *K. polysporus* and *K. thermotolerans*. Like *K. lactis*, *K. drosophilarum* produced a single band that hybridises with the *Kl. THI5* probe and therefore is likely to possess just a single copy of the gene (Figure 3.5b). Although the sizes of the *XhoI* fragments were found to be different for the two yeasts, both the single *PstI*(16.0 Kb) and *EcoRI*(1.7 Kb) fragments were identical so indicating that the environment of the *THI5* homologues within the genome is probably very similar. *K. polysporus* and *K. thermotolerans* displayed no bands that hybridise with the *Kl. THI5* probe. The absence of apparent *Kl. THI5* homologues in these yeasts may be due to a lack of cross-hybridisation with the *K. lactis* probe. With reference to an 18S ribosomal RNA phylogeny of ascomycetes, it can be seen that *K. polysporus* lies very much within the *Saccharomyces sensu lato* and *K. thermotolerans* is more closely related to *S. kluyveri* than *K. lactis*. With this in mind, the Southern blot was repeated and hybridised with the *S. cerevisiae THI5* probe. Bands were seen for *K. lactis*, *K. drosophilarum* and also *K. thermotolerans* (Figure 3.5c). However, again, no *THI5* homologue was detected for *K. polysporus* despite use of a large excess of digested genomic DNA on the agarose gel.

3.3 Further analysis of hemiascomycetes that lack a *THI5* gene.

There are several yeasts, which are classified outside of the *Saccharomyces sensu stricto*, that failed to give a positive signal when hybridised with the *THI5* gene probe suggesting that a *THI5* homologue is absent. All these yeasts have one common attribute when their nutritional and growth requirements are compared on the CBS

database (http://www.cbs.knaw.nl/search_ydb.html). They are documented as thiamin auxotrophs. In order to investigate which stage of thiamin biosynthesis is defective, a *S. cerevisiae* *THI4* probe was used for Southern blot analysis of *Xho*I digests of DNA from the thiamin auxotrophic yeasts described above (Figure 3.6a). In each case there was a positive signal indicating the presence of a *THI4* homologue. However, it is not known whether the gene is functional in thiazole biosynthesis. This early observation led to the hypothesis that these yeasts are only defective in HMP synthesis and that the rest of the thiamin biosynthetic pathway is likely to be functional.

To test this hypothesis, the three *Saccharomyces* thiamin auxotrophs and *K. polysporus* were observed for growth on minimal agar medium supplemented with or without thiamin and with either HMP or HET replacing the vitamin (Table 3.1). The results clearly confirmed the thiamin auxotrophy and showed that this could also be alleviated by an exogenous supply of the HMP precursor. Therefore, although unable to produce HMP, the remainder of the thiamin biosynthetic pathway is functional in these organisms.

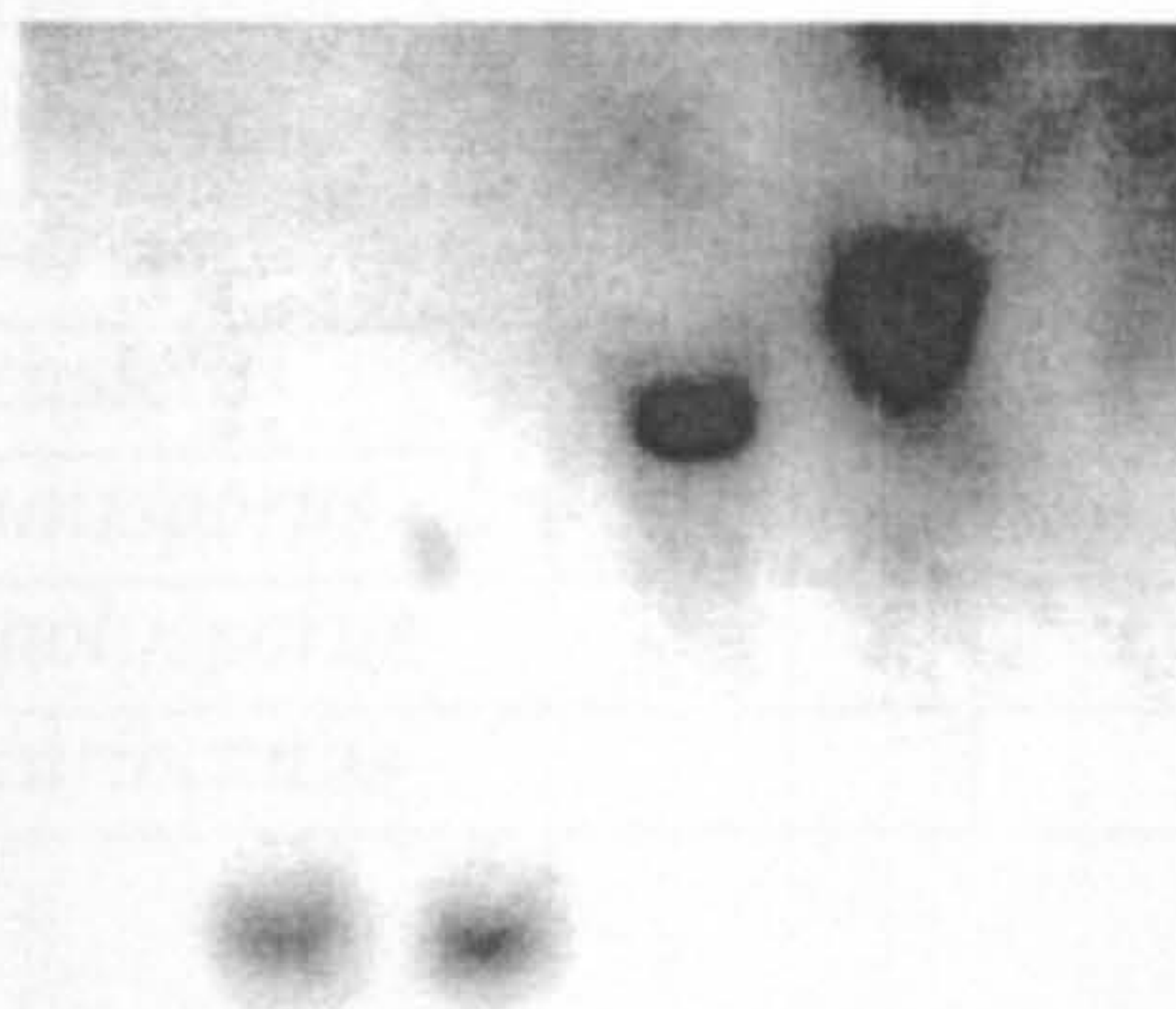
3.4 *S. cariocanus* is a thiamin auxotroph.

The thiamin and precursor requirements have since been determined for all the other yeasts used in this study for which there are no recorded data. Only *S. cariocanus*, a member of the *sensu stricto* complex, was shown to be auxotrophic for thiamin. This yeast contained multiple copies of *THI5* and was able to synthesise HMP. However, *S. cariocanus* was unable to synthesise thiamin without the supplementation of the thiazole precursor, HET (Table 3.1).

HMP auxotrophy appears to be the result of the loss of *THI5* from the genome (see 3.3). Therefore, HET auxotrophy in *S. cariocanus* could be a result of the loss of *THI4*. However, Southern blot analysis using the *S. cerevisiae* *THI4* probe indicated that the homologue is present (Figure 3.6b). The HET auxotrophy could be due to a

a.)

XhoI S288C
XhoI *S. servazzi*
XhoI *S. castellii*
XhoI *S. unisporus*



b.)

XhoI S288C-a
PstI
EcoRI
XhoI *S. cariocanus*
PstI
EcoRI
XhoI *S. mikatae*
PstI
EcoRI

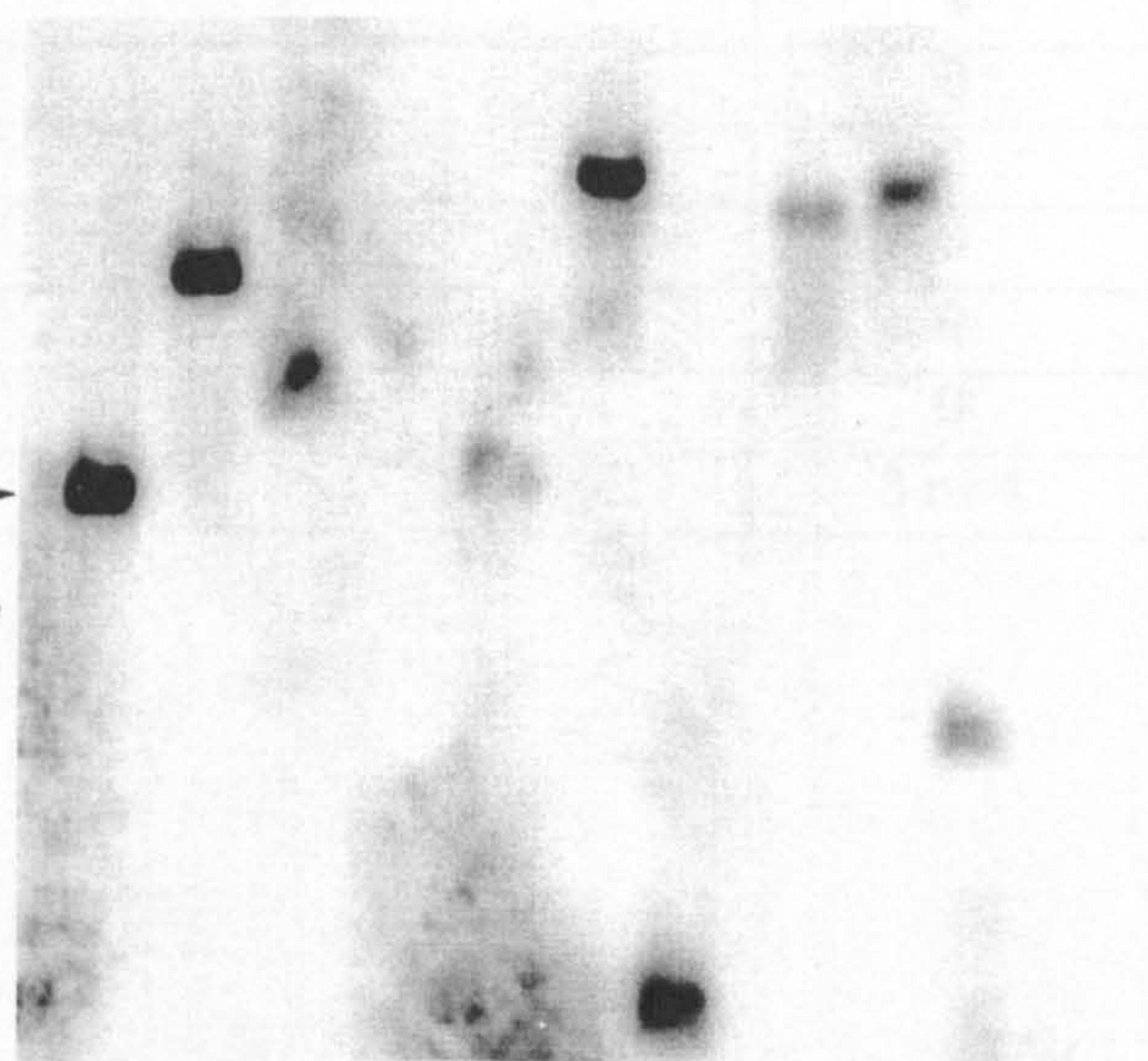


Figure 3.6 Southern blot detection of *THI4*. Probe = 1.1 Kb *EcoRI* fragment of pUP9. Hyb. temp. = 55°C.

a.) The thiamine auxotrophs of the *Saccharomyces sensu lato* and *Kluyveromyces*.

b.) The *sensu stricto* auxotroph *S. cariocanus* and the prototroph *S. mikatae*.

Species	w/o thi	+thi	+HET	+HMP	<i>THI5</i> copies
<i>S. cerevisiae</i> S288C	✓	✓	✓	✓	4
<i>S. cerevisiae</i> RWY15*	S	✓	S	✓	0
<i>S. servazzi</i>	X	✓	X	✓	0
<i>S. castelli</i>	X	✓	X	✓	0
<i>S. unisporus</i>	X	✓	X	✓	0
<i>K. polysporus</i>	X	✓	X	✓	0
<i>S. cariocanus</i>	X	✓	✓	X	3 or 4

Table 3.1 Precursor requirements of yeasts that exhibit thiamine auxotrophy. Growth was scored on Wickerham's medium supplemented with or without thiamine and the precursors hydroxymethylpyrimidine and hydroxyethylthiazole. After 5 days incubation ✓ denotes growth, X denotes no growth, S denotes severely restricted growth compared to S288C. Also shown is the number of *THI5* genes present within the yeast's genome.

**S. cerevisiae* RWY15 is a laboratory strain derived from S288C that has all four copies of *THI5* deleted(Chapter 5).

mutation of *S. cariocanus* *THI4* or, alternatively, it could be the result of a mutation in another gene which could lead to the discovery of another enzyme involved in precursor biosynthesis.

3.5 Discussion.

It appears that the multicopy state of *THI5* is exclusive to those yeasts of the *Saccharomyces sensu stricto*. However, the copy number varies, even between different strains and closely related species. The processes that gives rise to such a varied copy number are unknown. One plausible explanation is that ectopic recombination events between the subtelomeres may cause gene copy expansion or contraction. Differences in copy number have been observed for other subtelomeric gene families; the variable number of *MAL* loci in brewing strains and the variable *RTM* copy number among both brewing and wine strains of the *sensu stricto* (Denayrolles *et al.*, 1997; Jespersen *et al.*, 1999). Such copy number variation and RFLPs within the subtelomeres could be a useful tool in order to identify and distinguish different strains of the same species. This has already been exploited in one study which examined thirty six *S. cerevisiae* strains isolated from vineyards in Italy. It was found that genetic analysis combined with RFLPs using a subtelomeric probe gave maximum discrimination of strains and so thirty four were discovered to be unique (Cavalieri *et al.*, 1998).

The principal progenitor of S288C, EM93, has been shown to display an extra *THI5* hybridising fragment for each genomic DNA digest. One band is substantially fainter than the other four bands on a Southern blot. There are two possible explanations. Firstly, it may be due to two alleles of the same *THI5* gene but with restriction site polymorphisms, although EM93 is not thought to be polyploid (Mortimer, 2000). Alternatively, the extra band may represent an additional, more diverged member of this gene family. This might be a more centromere proximal copy, which is not uncommon for subtelomeric gene families in S288C. It seems that EM93 could have easily contributed the four genes seen in S288C. American Yeast

Foam lacks *THI11* and a good extension from this study would be to survey the other four parents to see if they contain the full complement of four *THI5* genes.

There is a large number of thiamin auxotrophs outside of the *sensu stricto* complex, especially among members of the *Saccharomyces sensu lato*. This phenotype is common among yeasts generally; out of the total yeasts on the CBS database for which there are data (2,558 yeasts) 58% are thiamin auxotrophs. From the phenotypic data of yeasts examined in this study, it was observed that the thiamin auxotrophy is due to the inability to synthesise HMP. The fact that exogenous HMP is successfully incorporated into the thiamin molecule signified that the HMP-kinase, encoded by the *THI20/THI21* homologue, was functional. Coupled with genetic data, it seems that this auxotrophy is solely due to the absence of a *THI5* homologue and it would be interesting to see if the HMP requirement can be cured by the introduction of the *S. cerevisiae* or *S. kluyveri* *THI5* clone (Chapter 4). If this is the case, the *THI5* gene could be used as a selective marker in the transformation of *S. servazzi*, *S. castelli* and *S. unisporus*.

THI5 is present as either a gene family (*Saccharomyces sensu stricto*), a single copy gene, or is totally absent from the genome. The reasons for this gene amplification within the *sensu stricto* are unknown. Gene expansion within the subtelomeres is often directly related to its metabolic role in an organism's environment. For example, amplification of the RTM genes is only observed in those *sensu stricto* yeasts which are regularly propagated on molasses (distilling, brewing and baking) and is either absent or only present as a single copy in laboratory strains. Likewise, wine yeasts possess no subtelomeric copies of *SUC* due to their role in fermenting grape juice, which contains glucose and fructose instead of the sucrose of molasses (Ness & Aigle, 1995). *THI5* may therefore be amplified as a function of the thiamin requirement under these various environments. Alternatively, selection for multicopy *THI5* within the *sensu stricto* could be due to different roles of the members in HMP biosynthesis either as a result of subtly variable enzyme activities or as a result of differences in regulation. Such differences are seen within the *PRS*

gene family. Only a single gene has yet been identified in *S. pombe* or *C. albicans*, whereas *S. cerevisiae* possesses five genes. Through experimental evidence, it has been suggested that each PRS member makes a different contribution to the production of intracellular 5-phosphoribosyl-1(α)-pyrophosphate in *S. cerevisiae* (Carter *et al.*, 1997).

The reasons behind multicopy selection of *THI5* can only be answered once the redundancy and function of each member is known. This will be addressed in Chapter 5.

Chapter 4

Determination of *THI5* gene order conservation between the *Saccharomyces sensu stricto*, *Saccharomyces kluyveri* and *Kluyveromyces lactis*.

4.1 Introduction.

This chapter describes the results of a comparative analysis of gene order among the *Saccharomyces sensu stricto* and two yeasts outside of this subgroup; *S. kluyveri* and *K. lactis*. In Chapter 3 it has been shown that whilst the *sensu stricto* members possess *THI5* in multiple copies, *S. kluyveri* and *K. lactis* contain just a single homologue. The genomes of these two yeasts are thought to share a high degree of gene linkage conservation (75%) with those members of the *sensu stricto* and so are likely to give clues as to the location of the ancestral *S. cerevisiae* *THI5* gene that gave rise to the present four membered gene family (Keogh *et al.*, 1998).

The *S. kluyveri* genomic library used in this study is from the diploid CBS 3082 type strain isolated from *Drosophila pinicola* (deposited to CBS by H.J. Phaff 1951). This strain is estimated to possess between five and seven chromosomes and a genome size of 9.55 Mb, 1.5 Mb less than the total size of the *S. cerevisiae* genome. In addition, it has been suggested that the proposed whole genome duplication event, which may have given rise to the sixteen chromosomes observed in *sensu stricto* yeasts, occurred after this lineage diverged from *S. kluyveri* (Keogh *et al.*, 1998). Recently, The French-based 'Genolevures' consortium have partially sequenced the genomes of thirteen hemiascomycetes which includes *S. kluyveri* CBS 3082. DNA sequencing of random sequence tags has yielded 2.5 Mb of data but does not include the *THI5* homologue observed on a Southern blot (Neuveglise *et al.*, 2000).

The *K. lactis* strain used was CBS 2359 and is regarded as the reference strain of the *K. lactis* research community. It possesses six chromosomes with a total genome size of between 10-12 Mb. This same strain has been included in the Génolevures sequencing project described above. So far, almost 3000 open reading frames have

been identified and, like *S. kluyveri*, no *THI5* homologue has been sequenced (Bolotin-Fukuhara *et al.*, 2000).

4.2 The *Saccharomyces sensu stricto*.

It has been shown that *THI5* is present as a gene family within the *Saccharomyces sensu stricto*. However, the absolute copy number varies, even between different strains of the same species. It is not known whether the *THI5* homologues in *sensu stricto* yeasts are arranged within the highly organised structure of gene families as seen for *S. cerevisiae* S288C, neither is it known whether these homologues are situated at the subtelomeres.

By using degenerate primers, gene order conservation between *THI5* and its adjacent genes was investigated by PCR. The primers were designed to amplify the intergenic regions between *SNZ-THI5* and *THI5-AAD* by annealing to conserved portions within the genes. Electrophoresis of the PCR products shows that the reactions yield amplified DNA fragments of similar size to those of S288C (Figure 4.1). The one exception is *S. bayanus* which showed no amplification between the *THI5* and the adjoining *AAD* member. This may be a consequence of DNA polymorphisms within the primer binding region or could be due to this particular yeast having no *AAD* gene downstream of *THI5*. The detection of amplified DNA in the other strains shows that the gene order is highly conserved. This indicates that the subtelomeric arrangement of gene families, as seen in the sequenced genome, is likely to exist throughout the *sensu stricto*.

4.3 *Saccharomyces kluyveri*.

4.3.1 Isolation of rescuing clones.

A *S. cerevisiae* strain, RWY16 (*MAT α ade2 Δ ::hisG thi13::HIS3 thi12::LEU2 lys2 Δ 0 met15 Δ 0 thi11::TRP1 ura3 Δ 0 thi5::KANMX4*), has been constructed and contains

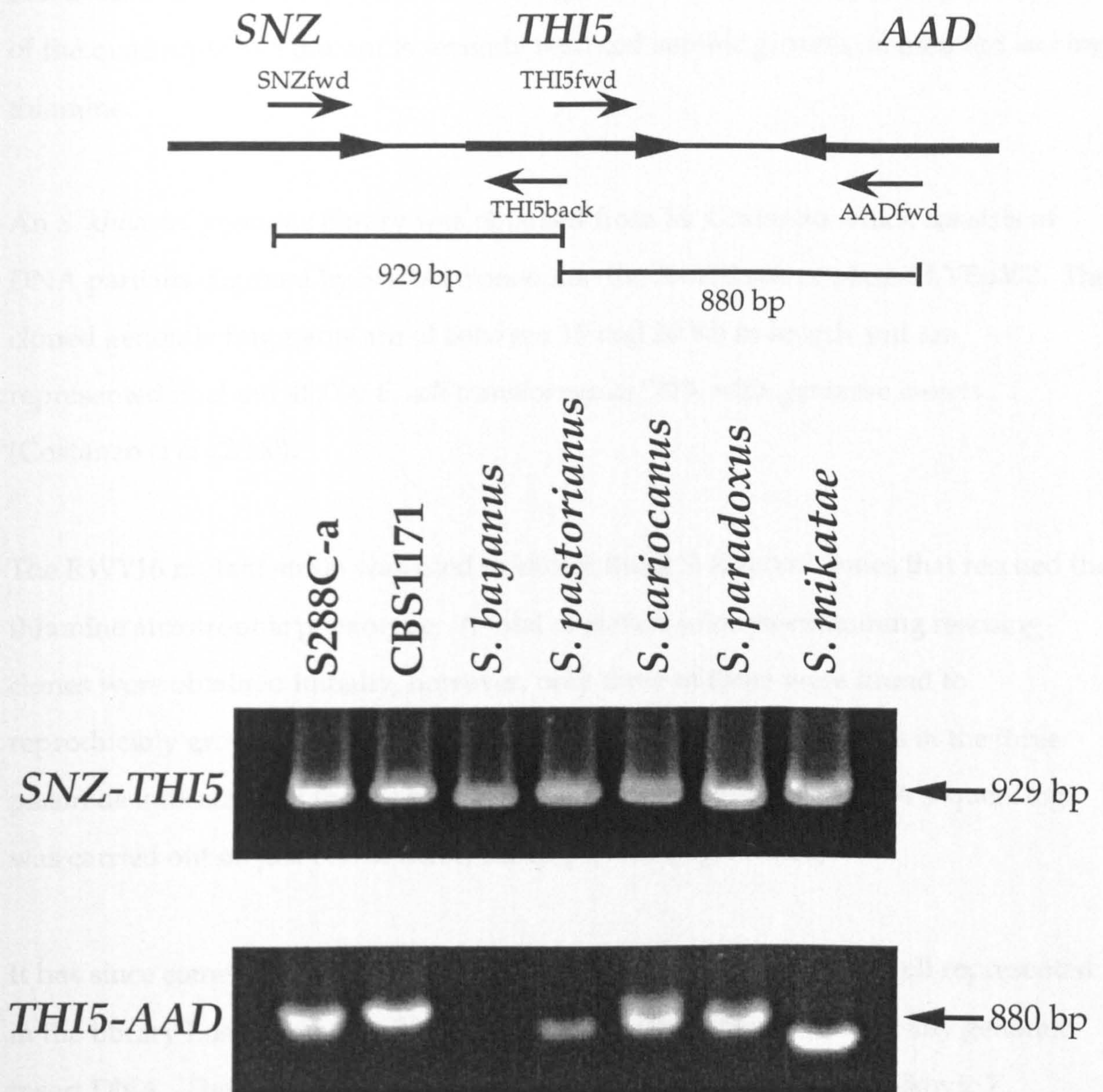


Figure 4.1 Degenerate PCR analysis to determine *SNZ-THI5-AAD* gene order conservation between members of the *Saccharomyces sensu stricto*. The PCR products were run on an agarose gel stained with ethidium bromide as shown above. For each of the two PCR reactions a *Saccharomyces sensu lato* member, *S. exiguus*, was used as a negative control. No bands were visualised for this yeast (data not shown). The strategy for primer selection is described in Materials and Methods.

deletions of all four members of the *THI5* gene family (Chapter 5). The phenotype of the quadruple *thi5* mutant is severely retarded aerobic growth on medium lacking thiamine.

An *S. kluyveri* genomic library was obtained from M. Costanzo which consists of DNA partially digested by *Sau3AI* cloned into the *Bam*HI site of plasmid YEp352. The cloned genomic fragments are of between 15 and 20 Kb in length and are represented in about 40,000 *E. coli* transformants, 70% with genomic inserts (Costanzo *et al.*, 2000).

The RWY16 mutant strain was used to isolate those *S. kluyveri* clones that rescued the thiamine auxotrophic phenotype. A total of eleven colonies containing rescuing clones were obtained initially, however, only three of these were found to reproducibly grow on medium lacking thiamine. Restriction analysis of the three plasmids indicates that these clones are identical and subsequent DNA sequencing was carried out on just one of these, called pRW3a (Figure 4.2a).

It has since come to my attention that the *S. kluyveri* genome is not well represented in the library I used and only 30% of the *E. coli* transformants contain any genomic insert DNA. The average insert size appears to be less than 6 Kb (Gojkovic Z. personal communication).

4.3.2 DNA sequencing and contig construction of *S. kluyveri* rescuing clone, pRW3a.

pRW3a contains an insert of 5.0 Kb spanned almost entirely by three *Eco*RI fragments of 2.8 Kb, 1.1 Kb and 1 Kb (Figure 4.2). These *Eco*RI fragments were subcloned, yielding pRW2.8, pRW1.1 and pRW1 respectively. DNA sequencing of these subclones and the *Eco*RI sites of pRW3a was performed according to the strategy shown in Figure 4.2b. Following contig assembly, all open reading frames of longer than 200 bp were identified and the translated sequence of each was

Figure 4.2 Cloning and DNA sequencing of the *THI5* homologue plus flanking DNA in *S. kluyveri*.

- a.) Restriction map of rescuing clone pRW3a. This is a partial *Sau3AI* fragment of *S. kluyveri* genomic DNA cloned into the *Bam*HI site of YEp352. *S. kluyveri* genomic DNA insert is represented by a thick line.
- b.) Strategy used to sequence the entire 5.0 Kb genomic insert of pRW3a. The three *Eco*RI fragments that span the insert were subcloned into pUC19. Initial DNA sequencing was carried out using primers that annealed to the plasmid DNA. Forward (fwd) primer is NEBiolabs #1212, reverse (rev) primer is NEBiolabs #1201. Further sequencing used primers that annealed to the insert DNA of the subclones where the sequence had already been determined. Contigs were assembled to give the full DNA sequence. The translated products of all open reading frames that are larger than 70 amino acids were queried using BLASTP against the Genbank database. Two open reading frames returned known peptide sequences with a BLAST score of above 200. For each open reading frame, the two highest results are given below. DNA sequencing of the putative *THI5* homologue, *Sk.THI5*, was repeated using primers skdn and skup.

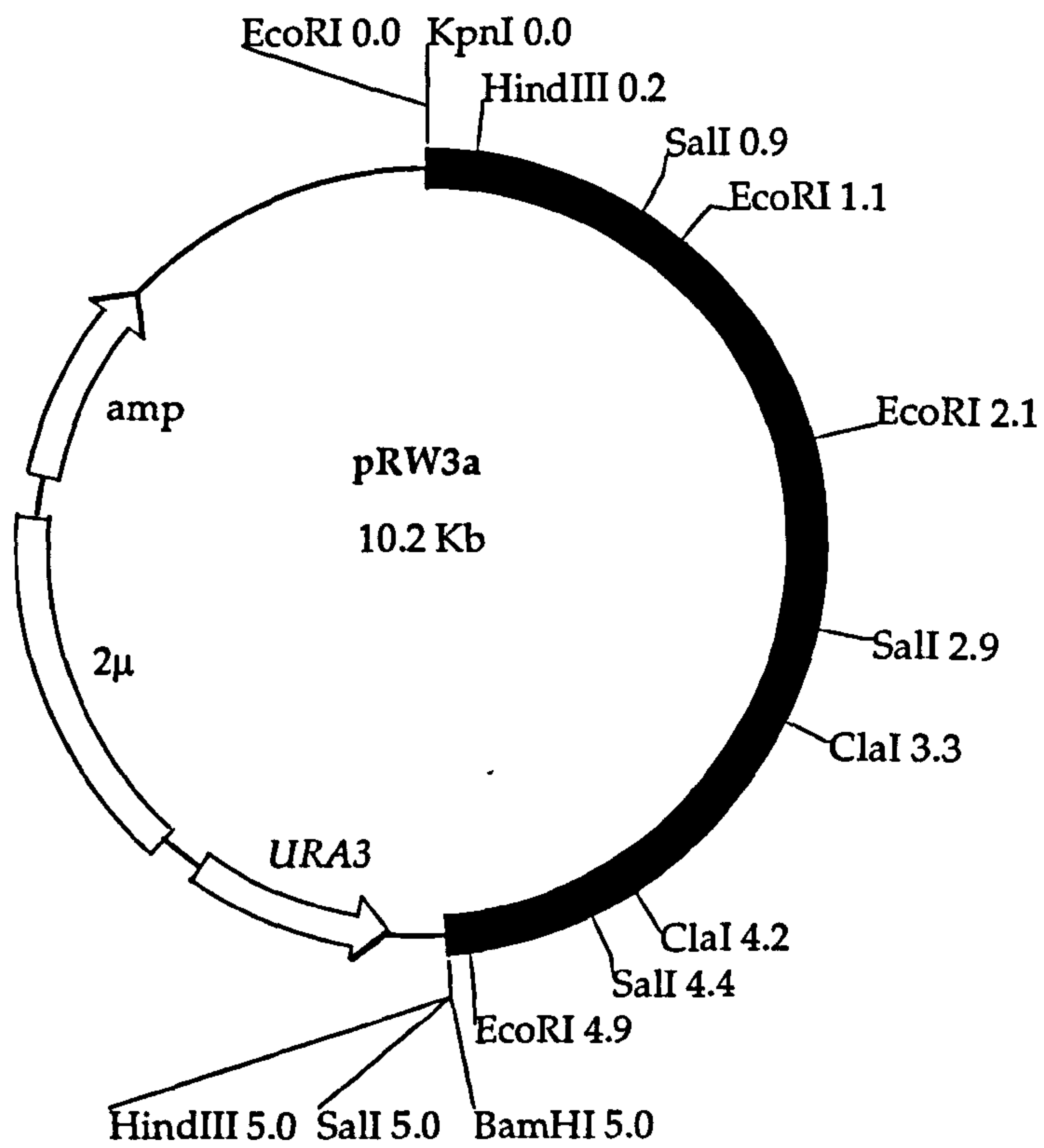
Query=SkORF1

	Score	E
Sequences producing significant alignments:	(bits)	Value
ref NP_013114.1 Ppr1p [Saccharomyces cerevisiae]	507	e-142
emb CAA62160.1 (X90564) PPR1 protein [Saccharomyces cerevisiae]	270	4e-71

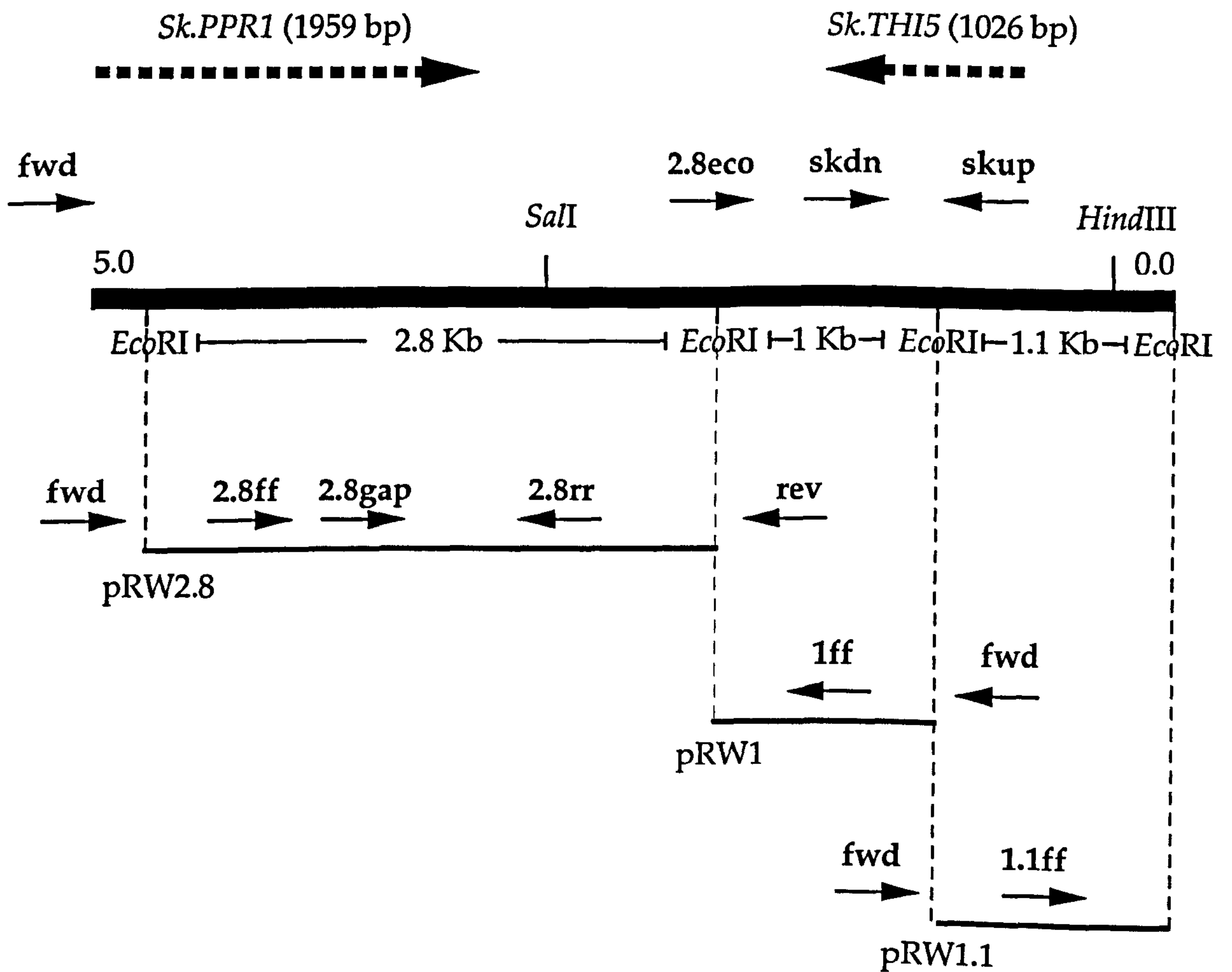
Query=SkORF2

	Score	E
Sequences producing significant alignments:	(bits)	Value
ref NP_010037.1 Thi13p [Saccharomyces cerevisiae] >gi 2132467 p...	611	e-174
ref NP_116597.1 Thi5p [Saccharomyces cerevisiae]	610	e-174

a.)



b.)



compared to the Genbank database. The full sequence is displayed in Appendix A. Two open reading frames of 1026 bp and 1959 bp were recognised as homologues of the *S. cerevisiae* genes *THI5* and *PPR1* respectively (Figure 4.2b, appendix A). These ORFs have been named *Sk.THI5* and *Sk.PPR1* with both genes having putative roles in pyrimidine synthesis.

4.3.3 The ORF adjacent to *Sk.THI5* resembles a *S. cerevisiae* DNA binding transcription factor involved in pyrimidine nucleotide biosynthesis.

Sk.THI5 shows 82% nucleotide identity to all four genes of the *S. cerevisiae* *THI5* family. The *S. kluyveri* gene product exhibits 85% identity and 88% similarity with each *S. cerevisiae* isozyme. This high protein sequence conservation is consistent with other fungal homologues and has been discussed in Chapter 2.

The product of *Sk.PPR1* displays 55% similarity and 46% identity with *S. cerevisiae* Ppr1p, however, based on the *S. cerevisiae* sequence, a region encoding 265 amino acids is missing from the N-terminal portion of the *Sk.PPR1* sequence. This N-terminal portion has been sequenced elsewhere as part of the Génolevures project and is shown in the alignment along with the data from this study and the full *S. cerevisiae* peptide (Figure 4.3). The Génolevures protein sequence is identical to the overlapping sequence obtained in this study apart for 30 amino acids of the former. This is most likely due to DNA sequencing errors as my own translated sequence data appears to better match the *S. cerevisiae* peptide.

Further evidence to support the functional assignment of the *S. kluyveri* ORF comes from studies of mutant alleles of *S. cerevisiae* *PPR1*. *PPR1* encodes a transcription factor (see Discussion) and two constitutive and three non-inducible mutations of the *S. cerevisiae* gene have been identified (Flynn & Reece, 1999; Reece, 2000). The non-inducible mutations all map to the DNA binding region (amino acids 43, 57 and 64). Residues at positions equivalent to K43 and L64 are conserved between the *S. cerevisiae* and *S. kluyveri* peptides whilst a conservative substitution (L-A) exists for

Figure 4.3 Alignment of Ppr1p from *S. cerevisiae* with the *S. kluyveri* homologue.

Alignments were created using default settings on PILEUP (see Materials and Methods). Sc.Ppr1p is the *S. cerevisiae* full length peptide. Sk.Ppr1p is the *S. kluyveri* predicted product of the open reading frame sequenced as part of this study.

Sk.Ppr1p(2) covers the N-terminal portion of the *S. kluyveri* protein and was obtained from two overlapping genomic clones sequenced as part of the Genolevures project (RSTs AU0AA002G10T1/ AU0AA015G05D1, <http://cbi.labri.u-bordeaux.fr/Genolevures/Genolevures.php3>).

*denotes important conserved residues identified by mapping mutations that affect Sc.Ppr1p function (see text).

ScPpr1p	1	~~~~~MKQKKFNSKSNRDLSSKRGDSPNIGISKSRACKRCRLKKIKCDQEFPSCKKCAKLEVPC
SkPpr1p	1	~~~~~
SkPpr1(2)	1	MSANKKRRNSNENISSNIDISYSKQKKRPSSSIIGISRSIAACKRCRLKKVKCDQKFPSCSKCASANEPC
ScPpr1p	62	VSLDPATGKDVPRSYVFFLEDRLAVMMFVLKEYGMDPTKI.....RGNIPATSDDEPFDLKKYS
SkPpr1p	1	~~~~~
SkPpr1(2)	71	VSLDPATGRDVPRSYVIFLEDRLAEMMKLKECGINPLEVPGKYSSNE*KSQGNIPATSEDXPCDFNLIE
ScPpr1p	121	SVSSLGEEGILPHNGLLADYLVQKGNMASSAITSKSMASPTINVQRKEFLVNSKKQDGSALLPETGSP
SkPpr1p	1	~~~~~
SkPpr1(2)	140	..QKLRSSEHELSDDNAMAGYIINNGTSMQKGVSIDES.....STEKPRKSTNSTISNQTATSLELGES
ScPpr1p	191	MTSDARAEELRRCNKEISALGTMRESSFNSEFLGDSSGISFAKLIFTATNFRQDSGDDVIDEDIKQREQKY
SkPpr1p	1	~~~~~
SkPpr1(2)	201	SRDLESLGAMKQNKQKNSMG....NASNSYLGDSGIXFAKLIFTAVKF....SPDAVEPDQEQESSVX
ScPpr1p	261	NGYAEAENNPHEFDPLELPFPHAAEVMTSRFFVDINSQLPLHRELFLKKYFEPYGPWNPNIALASDOTG
SkPpr1p	1	~~~~~DPLYLPTKDEAEGLLFKYFAHSNSQLPVLHREFFLKKYFEPVYGSLTEGVSLASDHTV
SkPpr1(2)	263	TEFDEVSES..FDPLYLPTKDEAEGLLFKYFAHSNSQLPVLHREFFLKKYFEPVYGSLTEGVSLASDHTV
ScPpr1p	331	INSAFEIPITSAFSAHTEPKRENVTEKIDVCSSVDVPWYDTWEISQKVNMRPIVELP.....TKFHIPYF
SkPpr1p	67	INREFKLP.....SIH.....QIEGNDETWFETWSSNRNSNLK.....IP.....LRYHVPLF
SkPpr1(2)	331	INREFKLP.....SIH.....QIEGNDETWFETWSSNRNSNLKIPLRYPRTVVFSKHH..FF
ScPpr1p	396	FLNIIFAIGHATQVLKSDITTVATYKRRATKYIASLFSSSDRLEALAGTLLMVIYSIMRPNVPGVWYTMG
SkPpr1p	111	FLNIIFSIGESTRVLDNAEKSRSNEKKRAVHYIDTLVCSNDRLEALAGVLLTAIYSIMRPNVPGVWYIMG
SkPpr1(2)	382	HWRIYTCLGFKC*~~~~~
ScPpr1p	466	SVLRLTVDLGLHSEKINKNYDAFTREIRRLFWCVYSLDRQICSYFGRPFPGIPEESITTRYPSLLDDSF
SkPpr1p	181	SALRLAVDLGLHAEKINKNYDPFTREIRRLFWCTYALDRQICVHFGRPFPGIPDENITAKEPSTLDDALI
SkPpr1(2)	394	~~~~~
ScPpr1p	536	TLTNREIDDYSDLPSPNPSSKVIALAMYKIRRIQASIVRIIYAPGAELPRRFMDLESWRIETYNELERWF
SkPpr1p	251	TTAADSIDDYSQVKSSMASYKVISLAIKIRKIQTSAQVLYAPNGRLPSHFTDLEKWRAVMDYELDNWY
SkPpr1(2)	394	~~~~~
ScPpr1p	606	QVDVPKNFEMMNCKFNISWFDLNYHYSKSIYGLSPKFPTLNDTAFKIVLDSIKGTIDVEYNLCVNKKIG
SkPpr1p	321	NKVTPTKTHRKMNCEFKVELFQLNYCHSKIMLYGLSPKSLSLTSYGYEVVHDSSTNAINAYKLCKEGNIN
SkPpr1(2)	394	~~~~~
ScPpr1p	676	YTWVAVHNMFMGTMTYLYVNEYSKNNINCCQEKVSEYTEKVLIVLKNLIGFCESAKTCYTSYKILSSVVI
SkPpr1p	391	YTWVAVHNLFMAGMSYLYTIYNSKSGSMETSDFKEHSSKLLHRLMSDWSCTAAKNKYKIFQVLSAAVT
SkPpr1(2)	394	~~~~~
ScPpr1p	746	RLKFMQIN.DAKGIFSDSNPLTSQANRMSSYDKKTNVLGFDDGTFDNKVFNRNINFEKAPFDIPLDEFFT
SkPpr1p	461	RLKFEDVEMKESGRIETAVPPEKKCG..DDNRQKTETLTSLSLNRNNNEMSTVTQENALGD.TLDQFFV
SkPpr1(2)	394	~~~~~
ScPpr1p	815	ELEKHSNVSQFNTLDVSEGNQVINESASTNTSSALNCQSYTNNQDIMDILFQVTSGSVWDEFFVRSNGN
SkPpr1p	528	ELDKLSPFSETTRSQVSHSNFDSPELENRNTIS..GTXQYIN...ALSSTAQNTLGKI.DETTPYSQPSN
SkPpr1(2)	394	~~~~~
ScPpr1p	885	EGESSYDISKGNSESGGIF*~~~~~
SkPpr1p	592	MVKENY.....KRVDELGLSRKYPSTKDGQRVYDMIYQVSTEAVWDQFFSNLGGGNGDVGGQPDLF
SkPpr1(2)	394	~~~~~

residue 57 (Figure 4.3). Both sites of the constitutive mutations, which map to L233 and N884 within the activation domain, are identical in *S. kluyveri*.

4.4 *Kluyveromyces lactis*.

4.4.1 DNA sequencing and contig construction of *K. lactis* clones pDW5a and pDW5b.

The *K. lactis* *THI5* homologue, named *Kl.THI5*, has previously been cloned and sequenced (Walsh D, unpublished data). *Kl.THI5* is 1022 bp encoding a protein with 81% identity to the *S. cerevisiae* Thi5p. The DNA sequence data also include 560 bp of the *Kl.THI5* promoter. Two overlapping plasmid clones span the entire gene; pDW5a contains the C-terminal portion of *Kl.THI5* and approximately 3.3 Kb DNA downstream, pDW5b contains the promoter and N-terminal portion of *Kl.THI5* along with 2.6 Kb DNA upstream. The maps of these plasmids are illustrated in Figure 4.4. The insert DNA on both clones was sequenced using a primer walking strategy that reads into the DNA flanking the *Kl.THI5* gene (Figure 4.4b). Assembly of contigs along with the existing data yielded continuous DNA sequence totalling 5493 bp (Appendix A).

4.4.2 Identification of adjacent ORFs to the *THI5* homologue, *Kl.THI5*.

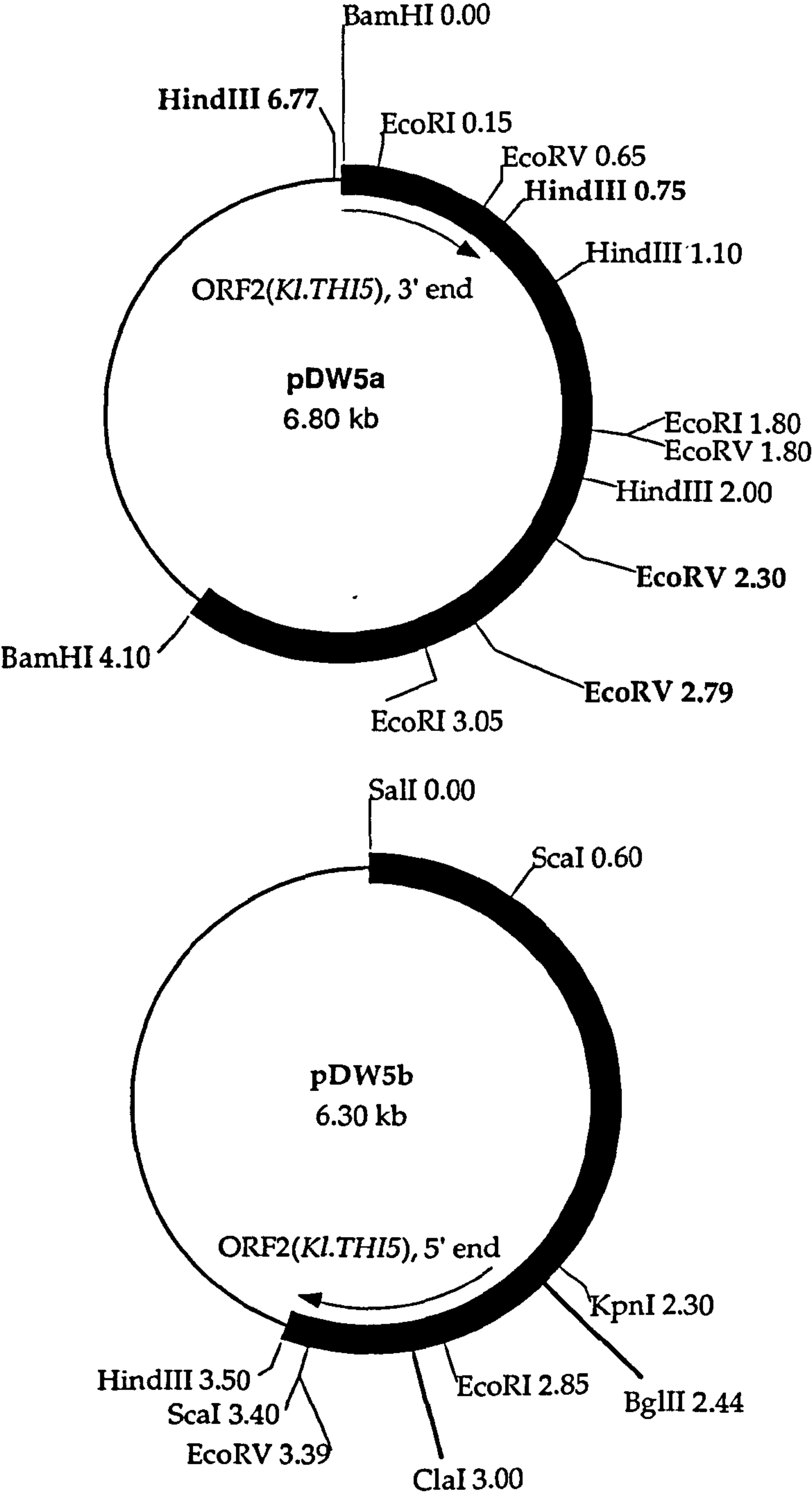
A search for genes within the *K. lactis* sequence data reveals the presence of three open reading frames with the same orientation (Figure 4.4b). ORF2 represents the previously sequenced *Kl.THI5* gene.

The translated peptide sequence of the adjacent upstream ORF (ORF1) was used to query the Genbank database. Numerous results were obtained covering dihydropyrimidinases, ureases, allantoinases and dihydroorotases from various organisms. The reason why such a broad range of enzymes are returned from the database is that all possess a urease motif (Gojkovic *et al.*, 2000; Jabri *et al.*, 1995). The

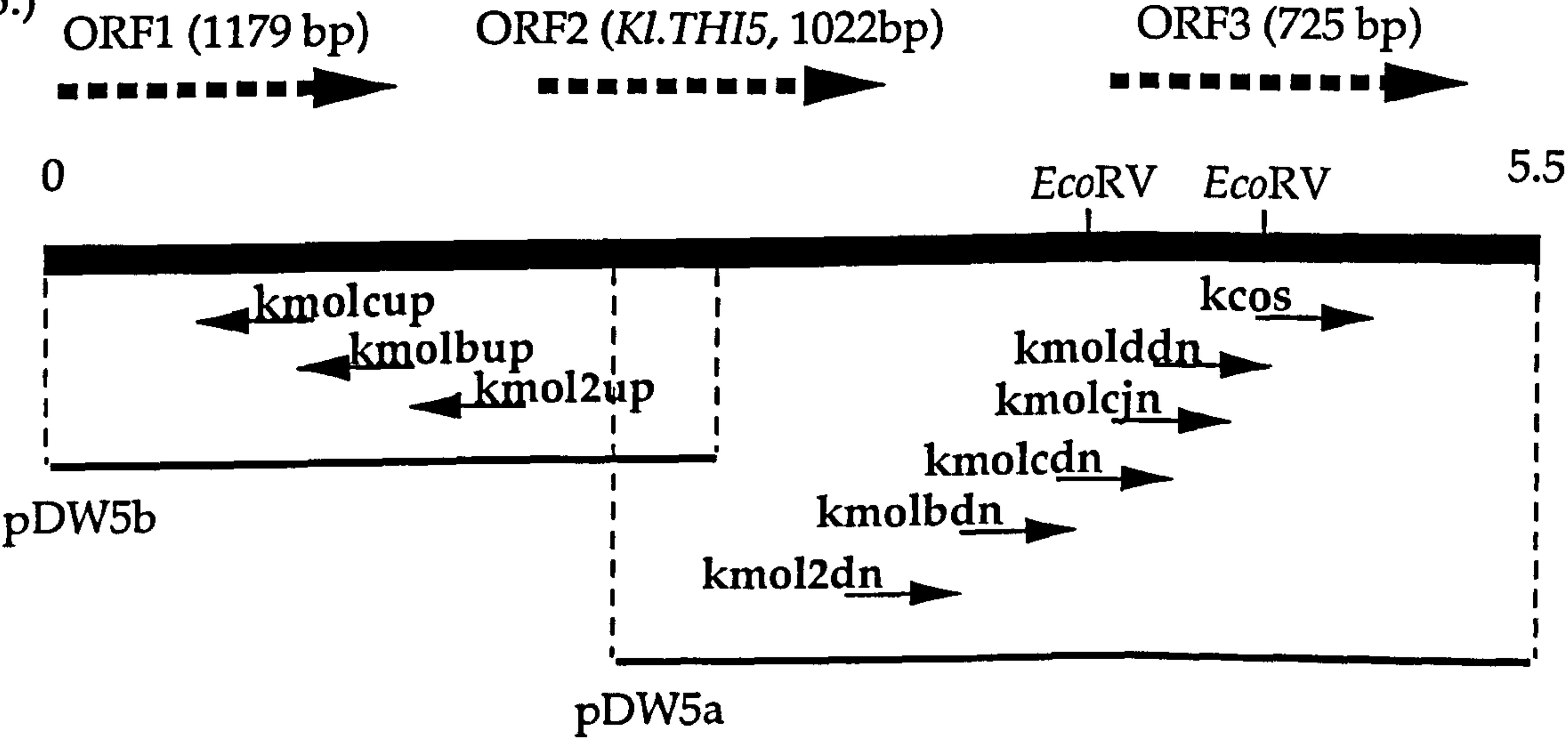
Figure 4.4 DNA sequencing of the regions flanking the *THI5* homologue in *K. lactis*.

- a.) Restriction maps of the *K. lactis* *Kl.THI5* clones pDW5a and pDW5b. pDW5a is a 4.0 Kb *Bam*HI fragment of genomic DNA cloned into pUC18. The probe used for Southern blot analysis (Chapter 3, Figure 3.4b) is the 0.8 Kb *Hind*III fragment. The probe used for *Kl.COS* copy number (this chapter) is the 0.49 Kb *Eco*RV fragment. pDW5b is a 3.5 Kb *Sal*I/*Hind*III genomic fragment cloned into pUC18. Both genomic inserts were previously isolated as positive clones upon colony hybridisation of a *K. lactis* λ EMBL3 library probed with the *S. cerevisiae THI5* gene.
- b.) Strategy used to sequence the entire 5.5 Kb genomic DNA that is spanned by the inserts of pDW5a and pDW5b. Both clones were sequenced using a primer walking strategy originating from the *Kl.THI5* gene. This gene has been previously sequenced by D. Walsh (unpublished data). Following contig assembly, three open reading frames of larger than 100 amino acids were identified. These were named ORF 1,2 and 3.

a.)



b.)



closest four matches with highest BLASTP scores are dihydropyrimidinases from *Saccharomyces kluyveri* (score 538), *Agrobacterium radiobacter* (score 204), and *Bacillus subtilis* (score 197). A hypothetical protein of *Streptomyces coelicolor*, probably another dihydropyrimidinase, is the third highest match with a score of 201. A BLASTP query of just the translated *S. cerevisiae* genome database produces only one match (score 78); an allantoinase, which is encoded by the *DAL1* gene.

The ORF1 translation along with the closest match from the entire Genbank database (*S. kluyveri* Pyd2p) and *S. cerevisiae* (Dal1p) are compared in Figure 4.5. ORF1 shows 66% nucleotide identity to *PYD2* and is 71% identical at the protein level. Pyd2p has been shown to be involved in the catabolism of pyrimidines, which allows their use as the sole nitrogen source and is a property shared by only a minority of fungi which does not include *S. cerevisiae* (Gojkovic, *et al.*, 2000). Based on this homology, ORF1 is considered to be a homologue of *PYD2* of *S. kluyveri* and has been renamed *Kl.PYD2*. It is not known if, like *S. kluyveri*, *K. lactis* can catabolise pyrimidines (Gojkovic Z. personal communication).

The assignment of ORF3 has not been as straightforward as for the other genes from *K. lactis* and *S. kluyveri*. A standard BLASTP search using the translated product of ORF3 returns a list of homologues with the closest match having a very low score of 45. A more thorough database search was conducted of the entire DNA sequence downstream of *Kl.THI5* which eliminates the effects of sequencing errors or mutations which may alter the reading frame of any putative translated products. This involves a protein-protein search of all six reading frame translations (BLASTX; Gish & States (1993)). The results returned identical matches for the same reading frame as ORF3.

4.4.3 Determining the role of ORF3.

The results of the database searches are shown in Table 4.1. On close inspection of the data it can be seen that ORF3 could be a homologue of a *S. cerevisiae* gene family

SkPyd2p	1	~~~~MPIYDLIIKNGIICTASDIYAAE....IAVNNGKVQIIAASIDPSLGSEVIDAEGA
KlOrf1p	1	~~~~~
Dal1p	1	MPINAITSDHVIINGANKPATIVYSTESGTIIDVLEGSVVMKETEITKYEIHTLENVSPC
* *		
SkPyd2p	53	FITPGGITDAHVHVDEPLKLLGIVVDTMEHATRSVAVAGGTTTVVAFSTQDVSKKGPSALAE
KlOrf1p	1	~~~~~
Dal1p	61	TILPGLVDSHVHLNEPGRTSWEGFET...GTQAAISGGVTTVVDMPLNALP...PTTNVE
* *		
SkPyd2p	113	SVKLDVDEYSEQTLYCDYGLHLILFQIEKPSVEARELLDVQLQAAYNDYGVS.SVKMFMT
KlOrf1p	1	~~~~~VSPRVKMFMT
Dal1p	115	NFRIKL.EAAEGQMWCDVGFWGGLVPHNLPDIPLVKAGVR...CFKGFLLDSGVVEEF..
*		
SkPyd2p	172	YPGLQISDYDIMSAMYATRKNNGFTTMIHAENGDMVKWMIEALEEQGLTDAYYHGV.SRPS
KlOrf1p	11	YPGLQISDYDILTMTQATRKNNGFTTMIHAENGDMVKWMIEQLEEQNMTSPYFHGV.SRPI
Dal1p	169	.PPIG.KEY.IEEALKVLAEDTMMMFHAE...LPKAHEQQQPEQSHREYSSFLSSRPD
*		
SkPyd2p	231	IVEGEATNRAITLATTMD...TPILFVHVSSPQAAEVIKQAQTKGLKVYAETCPQYALLS
KlOrf1p	70	IVEGEATNRAIVLAKMMD...TPILFVHVSSPEAIDTIRSAQTKGLKVYAETCPQYALLS
Dal1p	223	SFEIDAINLILECLRANGPVPVHIVHLASMKAIPLIRKARASGLPVTETCFHYL...
*		
SkPyd2p	288	DAITRCHHHGGEVESYGVGIDLSSISESPFT..NPDDRFIGSKYI.CSPPIRPEGTQKSIW
KlOrf1p	127	DIDTKCHHELEQQQSTCSDANKSEPRVKLTAKEDDKFIGAKAI.CSPPIRPEHCQESVW
Dal1p	280C.....IAAEQIPD...GATYFKCCPPIRSESNRQGLW
*		
SkPyd2p	345	KCMNNGTFTIVGSDHCSYNYYEKTSTASKHRAFDPENNKNGEFRIYPNGLPGVCTRMPLL
KlOrf1p	186	KAMNNGTFTIVGSDHCAYNYYDKTS..GKLDAF...KDNKNGEFKYIPNGMPGVCTRMPLL
Dal1p	310	DALREGVIGSVVSDH.....SPCTP....ELKNLQKGLFFDSWGGIASVGLGLPLM
*		
SkPyd2p	405	YDYGYLRCNLTSMMLVEIQCTNPAKVYGMYPQKGSILPGVSDADLVLLWYPDDSKKEYNS
KlOrf1p	242	FDYGVLLQKLANMTKFVEIQCTNPAKLYGLYPKKGSILPGVSDADLVVWYPPNYKG..CK
Dal1p	357	FTQG.....CSLVDIVTWCKNTSHQVGLSHQKGTIAPGY.DADLVVF.....DTAS
*		
SkPyd2p	465	KPKLITNKLMEHNCDYTPFEGIEIKNWPRYTIVKGKIVYKEGEIILKENADGKYLKRGKSF
KlOrf1p	300	KPKTYTNDLLVHGCYTPYEGFEINNWPREFITIVKGSIVYKEGEIILKENAKGTYYKRGKSG
Dal1p	403	KHK.ISSSVYFKNKLTAYNGMTVKGTVLKTIIRGQVYTNANGVSKTPLGQTL LDSRR~
*		
SkPyd2p	525	MCTPKNEWVTEWRPKYES~~~~~
KlOrf1p	360	LCTANDEWVTEWRPQYLNQKSG
Dal1p	461	~~~~~

Figure 4.5 Alignment of the *K. lactis* ORF1 peptide with dihydropyrimidinase from *S. kluyveri* (Sk.Pyd2p) and allantoinase from *S. cerevisiae* (Dal1p).

*denotes residues that are possibly important in binding of a metallo-cofactor. This is based on a visual comparison between the above proteins and the alpha subunit of urease from *Klebsiella aerogenes* (Jabri *et al.* (1995)

Protein	Organism	BLASTP score	Region of ORF3 peptide.(total=241aa)
1. Cos10p	<i>S. cerevisiae</i>	44.7	167-209
2. Ycr007cp(COS)	<i>S. cerevisiae</i>	43.9	83-224
3. Yhl044wp(COS)	<i>S. cerevisiae</i>	42.7	63-209
4. Yir043cp(COS)	<i>S. cerevisiae</i>	42.4	125-202
5. Cos2p	<i>S. cerevisiae</i>	42.4	152-202
6. Cos7p	<i>S. cerevisiae</i>	42.0	152-233
7. Cos8p	<i>S. cerevisiae</i>	40.4	66-207
15. Yar028wp(COS)	<i>S. cerevisiae</i>	38.9	167-223
16. Yar033wp(COS)	<i>S. cerevisiae</i>	37.7	160-219
17. Ygl051wp(COS)	<i>S. cerevisiae</i>	37.7	160-240
18. Prm9p(COS)	<i>S. cerevisiae</i>	37.4	146-210
19. Yar027wp(COS)	<i>S. cerevisiae</i>	37.0	167-208
22. Cytochrome b	<i>R. chensinensis</i>	32.3	57-127
24. Yhl042wp(COS)	<i>S. cerevisiae</i>	32.0	121-202
25. Putative Ebs1p-like protein	<i>K. lactis</i>	31.2	137-190
26. Membrane protein pirS49598	<i>L. lactis</i>	30.8	88-146
27. Ecm34p	<i>S. cerevisiae</i>	30.8	152-192

Table 4.1 BLASTP results using translated *K. lactis* ORF3 as query against the Genbank non-redundant protein database. All results apart from several COS gene family members are shown. Shaded members are considered to be the closest match due to a relatively high score covering the majority of the ORF3 peptide sequence. *R. chensinensis* is from the *Rana* (frog) genus, *L. lactis* is a *Lactococcus* species.

*This table and figure 4.6 show that ORF3 is only distantly related to the members of the COS family and may not share any functional relationships.

of unknown function, named COS (COnserved Sequences). This gene family consists of twenty two members of which several are very highly conserved and are located within the subtelomeres (Chapter 1, Figure 1.1). Figure 4.6 displays an alignment of four COS gene products and includes Ycr007cp which is the most closely related member to the ORF3 peptide (34% similarity, 25% identity). The degree of similarity and identity varies within the *S. cerevisiae* COS family with members sharing as little as 38% similarity, 22% identity (Cos7pxYhl044wp) through 55% similarity, 47% identity (Cos5pxCos10p) and upto 86% similarity, 83% identity (Cos4pxCos2p). It is not surprising therefore that the *K. lactis* homologue possesses such low values and may itself be part of an averagely conserved gene family. With this in mind, ORF3 has been named *Kl.COS*. However, it has not been ruled out that this may be a novel gene.

4.4.4 Analysis of *Kl.COS* copy number in *K. lactis*.

Four highly conserved members of the *S. cerevisiae* COS gene family are found upstream of each *THI5* member. A model will be described in which these genes may have been adjacent, like *K. lactis*, before an insertion of either the *SNO/SNZ* gene block or the block of genes encoding proteins involved in sugar metabolism (Chapter 7). It has been shown that *Kl.THI5* in *K. lactis* is present as a single copy gene, therefore, it would be interesting to know whether the same is true for *Kl.COS*.

The Southern blot survey of *THI5* copy number in various organisms (Chapter 3) used a probe from *S. cerevisiae* which easily cross-hybridised with the *S. kluyveri* homologue (82% identity) and marginally with *K. lactis Kl.THI5* (75%). For this reason Southern blot analysis was used to try to determine the copy number of *Kl.COS* in *K. lactis* which would easily detect other family members of greater than 80% identity at the DNA level. *K. lactis* genomic DNA was digested with a range of enzymes and was hybridised with a probe made from a portion of the *Kl.COS* gene of pDW5a. The blot is displayed in Figure 4.7. The results suggest that like *Kl.THI5*,

COS10	1	~~~~~MGEETRLSS.KGLET
COS2	1	~~~~~MKENELKNEKSV DVLSF.KQLES
YCR007C	1	~~~~~MQPHLDNNSNDDVKLDTLGEQNVLS
YHL044W	1	~~~~~MSSELLISNSKPRPEGLRK.LCE
ORF3	1	MSLIKNIADVITLTSTFHKAKINDKTNTYTKAMSKPEKDANDVHELLPVKEIPAPPYSEG

COS10	16	HSSFCLPEDIFRNRLTWLCSE..IRNS.GF.WIWSLSWLPLAVWGMASWTFYPLLTS.V
COS2	23	.QKIVLPQDLFRSSFTWFCYE..IYKSLAF.RIWMLLWLPLSVWVKLSNNCIVPLIVSLL
YCR007C	28	AENITLPEDTFKSYMTYLLYEMAHYKPMIFSFLALSVSILIVVIFHNVKACDVVFGFSIF
YHL044W	23	GETVILPRDITPSKCAVFLKQNIWFISYIFIHIIITLILNRLAISAHGNTLIIILA.ALL
ORF3	61	GE...LGRDFSSNIILHSLRANKTCFNMVKITTVACLLYGLALYEL..DIGFALFSIFM

COS10	71	V.LILALIAILPVVQIQYSKYTLSNQ..L.....TQLSKEI IKSAPG.AYSADWDAVAIH
COS2	79	V.LFLGPIFVLVICGLS.RKRSLSKQ..L.....IQFCKEITENTPS.SDPHDWEVVAAN
YCR007C	88	VTSLFLSTLIP.....FNVIISDEGFR.....IKLLLEVITHRPA.VKGKEWRAITDN
YHL044W	82	ITISLFLLLLLPYLSCSRYKLRCLDDCK.....FKLLAEVITHKPN.MDLSTWDRIAYD
ORF3	116	TTAVAFLFCLFPLHAYDRNTEPSGSVFKDIGNALFFLKTTHDFKPSYIDMKQWDIVARV

COS10	122	FNSYLYENKAWKTAHFFFN GTDCQEA FRKTILE.PAVLRRQNE DPRFSSIEILVP.....
COS2	129	LNSYLYENNWNTRYFFFNAMVCQEA FRTTLE.PFSLKK.DKAAKVKSFKDSVP.....
YCR007C	136	MNQYLLDNGLWSTRYYFYSSERCYKFFRFLVKEKPPGVNVNSSVKDATSTQIDAPANEAS
YHL044W	136	MNQEVYDRRICADRSFFYDGSYCYQVFKLV.ATPYLVNSN...MNSIYADLEMRSNGAT
ORF3	176	VNKHLYEAGSWPNNRFFYDGKRCQAMFIALVKKPTDPLDDYISKANETLKI SMDELWNS

COS10	176YTEDAVQVYFTKVNAQWRLLIHGKKECKLTMLENVKLPKETYR
COS2	182YIEEALGVYFTEVEKQWKLFNTEKSWSPVGLED AKLPKEAYR
YCR007C	196	NEVIKCFSSFSSDPI.FEAYFVKAVEVEKQAQQEYWRKQYPDADIP~~~~~
YHL044W	192	N..INDSGNSSLHIELGTIYIFKALAVERN SVDKYWEDKYPEMGTVTV~~~~~
ORF3	236	TAASSA*~~~~~

COS10	218	CKLAWSCQRIRISFSPLDFLPDMSDFFICANLSPA.CLV..SYWLIDIFFRMIDDFQNI R
COS2	224	FKLTWFLKRISNIFMLIPFL...NFLCCIVVSRGMCLLLRTL YLGWILFMLVQGFQNI R
YCR007C	240	~~~~~
YHL044W	236	~~~~~
ORF3	242	~~~~~

COS10	275	PKSMKVDDKMQYLSDIINEQGASPEKWDTIARKTNTRLFEKRVWKNEEFFFDGTDCAFF
COS2	280	VLIMSMEHKMQFLSTIINEQESGANGWDEIARKMNRYLFEKKAWKNEEFFFDGIDCEWFF
YCR007C	240	~~~~~
YHL044W	236	~~~~~
ORF3	242	~~~~~

COS10	335	ERNFSSLLFSKKSASPRSLNVELWKYIQEAQLSCN YEPLP
COS2	340	NHFFYRVLSAKKSMWPLPLNVELWPYIKEAQLSRSEVLLV
YCR007C	240	~~~~~
YHL044W	236	~~~~~
ORF3	242	~~~~~

Figure 4.6 Alignment of the *K. lactis* ORF3 peptide with the four proteins that return the highest BLASTP scores.

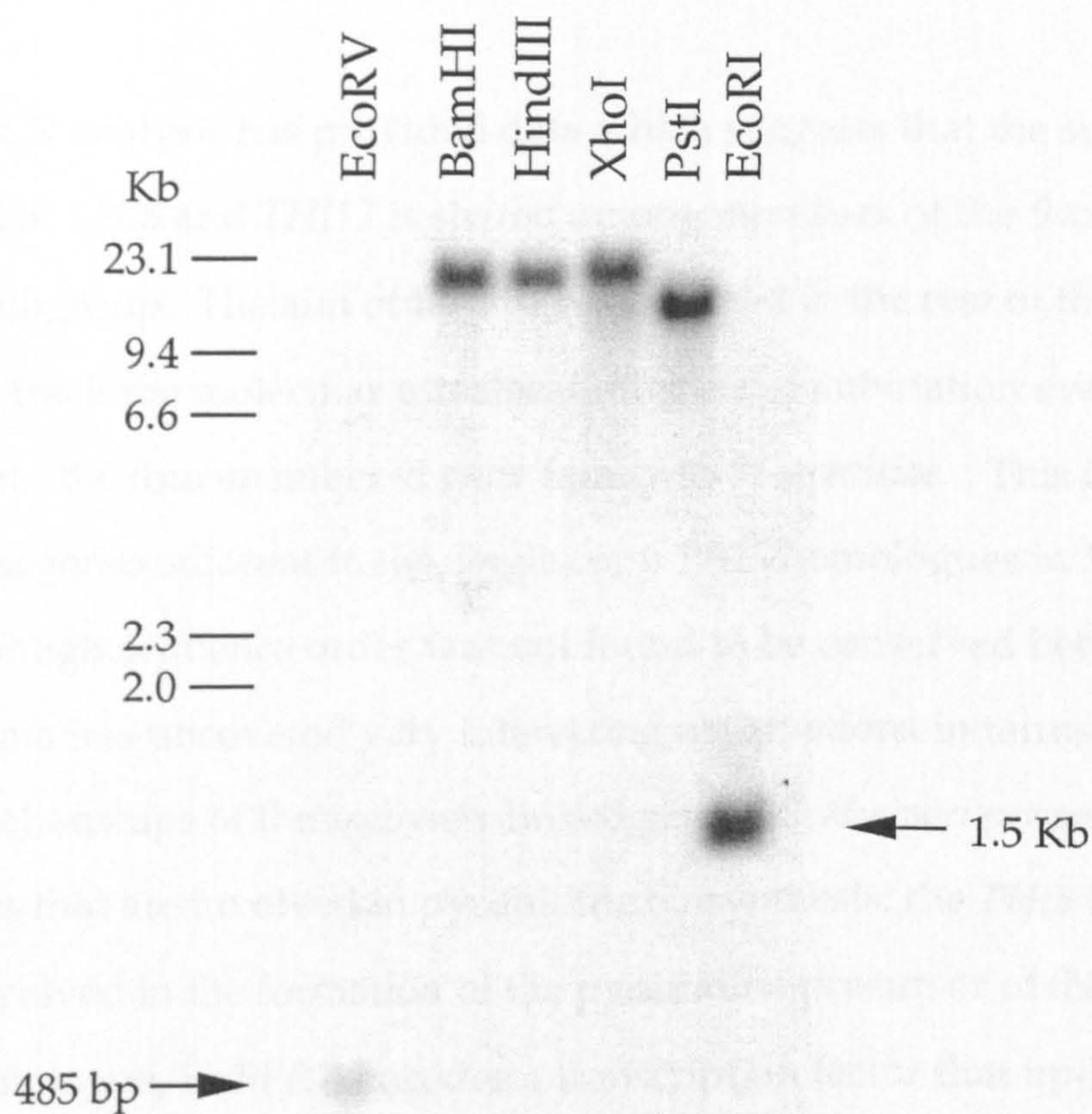


Figure 4.7 Southern blot analysis of *Kl.COS* copy number in *K. lactis*. The probe is a 485bp *EcoRV* fragment of pDW5a (figure 4.3) and the hybridisation temp. is 65°C. The blot was repeated with a lower hybridisation temp. of 55°C and over-exposure of the autoradiograph. The result shows identical bands on both blots.

Kl.COS is not present as a gene family. If this finding is confirmed, the disruption of this single gene in *K. lactis* would allow for extensive phenotypic studies in order to determine the true function of *COS*, one of the largest gene families in *S. cerevisiae* and which still remains unstudied.

4.5 Discussion.

Degenerate PCR analysis has provided data which suggests that the subtelomeric environment of *THI5* and *THI12* is shared among members of the *Saccharomyces sensu stricto* subgroup. The aim of the study described in the rest of this chapter was an attempt to track the molecular translocation and recombination events that culminated into the four-membered gene family in *S. cerevisiae*. This involved identifying the genes adjacent to the single copy *THI5* homologues in *S. kluyveri* and *K. lactis*. Although sequence order was not found to be conserved between the three species, the data has uncovered very interesting observations in terms of the functional relationships of these closely linked genes. *S. kluyveri* possesses two adjacent genes that are involved in pyrimidine biosynthesis; the *THI5* homologue, *Sk.THI5*, is involved in the formation of the pyrimidine precursor of thiamine, whilst the *PPR1* homologue, *Sk.PPR1*, encodes a transcription factor that up-regulates several *URA* genes in response to pyrimidine nucleotide starvation (Losson & Lacroute, 1981). These *URA* genes encode enzymes that synthesise uridine 5'-monophosphate (UMP) from the amino acid glutamine. The biosynthesis of pyrimidine nucleotides and HMP for thiamine is known to occur by entirely different pathways (Chapter 1), although, in each case, glutamine is one of the starting substrates. Despite the evidence, it is reasonable to suggest that these two genes may encode enzymes with overlapping roles and there is published data that could support this. Firstly, we know that there may exist an alternative minor biosynthetic route to HMP synthesis which does not proceed through pyridoxine (Chapter 1). It is feasible that this route may originate from an intermediate of the nucleotide pathway described above. Secondly, more convincing evidence has come from studies that have looked at sensitivity to 6-azauracil. 6-azauracil is an inhibitor of

both pyrimidine and purine nucleotide biosynthesis and causes lowering of intracellular nucleotide pools (Hampsey, 1997). A *ppr1* mutant results in 6-AU sensitivity, suppressed by the addition of uracil to the medium. Two other mutants, *snz1* and *sno1*, are also sensitive to 6-AU and again suppressed by the addition of uracil to the medium (Padilla *et al.*, 1998). The wild-type genes have been found to encode enzymes involved in the biosynthesis of pyridoxine, the substrate for HMP formation (Marsh and Meacock, unpublished).

A similar relationship exists in *K. lactis*. Upstream of *Kl.THI5* is a homologue of the *S. kluyveri* *PYD2* (dihydropyrimidinase) gene which confers the ability of this yeast to use pyrimidines as the sole nitrogen source (Gojkovic *et al.*, 2000). This property is rare among the fungi and does not include *S. cerevisiae*. Hence, in *K. lactis*, there are two adjacent genes, one involved in the breakdown of pyrimidines and one involved in their formation (*Kl.THI5*). The adjacent gene that lies upstream of *S. kluyveri* *THI5* has not been identified. A strong candidate, based upon the findings in *K. lactis*, is the *PYD2* gene. The *S. kluyveri* DNA sequence data has been published for this region and includes part of an open reading frame immediately downstream of *PYD2* (Gojkovic *et al.*, 2000). Instead of *Sk.THI5*, this open reading frame appears to encode a portion of a Cus1p homologue of *S. cerevisiae* which has a role in assembly of the spliceosome. Therefore, the environments around the *S. kluyveri* and *K. lactis* *THI5* genes appear to be very different.

As suggested for *Sk.PPR1* and *Sk.THI5* in *S. kluyveri*, *Kl.PYD2* and *Kl.THI5* may have related functional roles in *K. lactis*. This yeast does not have the four copies observed in *S. cerevisiae* and the precise purpose of multiple copies is not known but they may satisfy an increased requirement for HMP (discussed in later chapters). Therefore, to help boost intracellular HMP levels, *K. lactis* may convert degradation products of pyrimidines into HMP and therefore the dihydropyrimidinase activity may be very important. If this were the case, a *kl.thi5* mutant would not be a thiamine auxotroph if there was an exogenous pyrimidine supply. This has already been carried out in a *K. lactis* laboratory strain that contains a disrupted *Kl.THI5* gene (Walsh and

Meacock, unpublished). This strain is not a thiamine auxotroph but exhibits slightly retarded growth on medium lacking thiamine compared to wild-type (Chapter 6).

The relationship between *Kl.TH15* and *Kl.COS* is unclear as the function of the multiple *S. cerevisiae* homologues of the latter gene is not known. In *S. cerevisiae* many of the twenty two *COS* genes are subtelomeric of which four (*COS1*, *COS4*, *COS5* and *COS7*, Figure 1.1) were likely to be originally found adjacent to each *TH15* member before the insertion of the two different duplicated blocks of genes (Chapter 7). Therefore conservation of gene order between *K. lactis* and *S. cerevisiae* seems to be partially conserved and like *Kl.TH15*, *Kl.COS* may be present as a single copy gene. Clues as to the function of *S. cerevisiae* *COS* comes from a microarray study with the purpose of identifying genes that respond to zinc depletion in a Zap1p dependent manner (Lyons *et al.*, 2000). Transcription of several *COS* members appears to be induced as a result of zinc deficiency but not to the levels seen for genes known to be highly expressed during these conditions (e.g. the high affinity Zn uptake transporter encoded by *ZRT1*). However, similar to *ZRT1*, these *COS* members do possess a ZRE-like sequence (zinc responsive element recognised by Zap1p transcription factor) within their promoters. It is not known, as suggested above, whether *COS* plays a part in zinc homeostasis.

Taken together, the sequence data have uncovered interesting relationships between the *TH15* homologues and their neighbours in *S. kluyveri* and *K. lactis* which represent three pyrimidine metabolic pathways.

Chapter 5

A functional analysis of the *THI5* gene family in *S. cerevisiae*.

5.1 Introduction.

As reported in Chapter 3, the existence of *THI5* as a gene family is exclusive to those yeasts of the *Saccharomyces sensu stricto*. Yeasts outside of this subgroup possess *THI5* as a single copy gene. Of all the Thi5p homologues, only *S. pombe nmt1p* has been shown to be required for HMP synthesis (Chapter 1). Based on the studies of the single *S. pombe nmt1* gene it is presumed that the protein products of *THI5*, *THI11*, *THI12* and *THI13* of *S. cerevisiae* S288C are also involved in HMP synthesis. The redundancy of the *S. cerevisiae* family is currently unknown, although the promoter regions of *THI5* and *THI12* both give thiamin-regulated expression of a *lacZ*-reporter gene (Burrows, 1997).

To determine the functional redundancy and contribution of each of the four genes to HMP formation, this chapter describes the construction and phenotypic analysis of all combinations of the single, double and triple deletion mutants. This eventually leads to the single strain which possesses deletions of all four genes. To create the mutants, a PCR-mediated gene replacement strategy has been adopted. This technology has been successfully applied to the concurrent knockout of larger families in *S. cerevisiae* including the six-membered *AAD* gene family and all twenty hexose transporter genes (Delneri *et al.*, 1999; Wieczorke *et al.*, 1999).

5.2 Creation of mutant strains containing total deletions of the *THI5* loci.

5.2.1 Deletion strategy.

The deletion of the four members of the *THI5* gene family in all fifteen possible single, double, triple and quadruple combinations was performed in the *S.*

cerevisiae laboratory strain BY4705. This strain is derived from S288C and has retained the four *THI5* genes (Figure 5.2). The advantages of using BY4705 is that it contains completely deleted alleles of seven of the most commonly used auxotrophic marker genes (Brachmann *et al.*, 1998). This makes BY4705 a good choice for the deletion of members of gene families using different nutritional markers and, because the genomic alleles are total deletions, there is no undesirable marker locus gene conversion.

The complete replacement of each *THI5* locus with a different auxotrophic nutritional marker gene used a PCR-generated targeting cassette as illustrated in Figure 5.1 (Lorenz *et al.*, 1995). The cassette contains the marker flanked by forty base pairs corresponding to the extreme ends of the target locus. Integration occurs by homologous recombination between both ends of the cassette and the corresponding gene termini resulting in a replacement of a *THI5* member with the nutritional marker. The outcome of this replacement is a total gene deletion with the exception of forty base pairs at each terminus. In the case of the *THI5* family, which is highly conserved at the DNA level, it is not possible to target a particular member specifically and so the cassette can integrate at any of the four genes. This random targeting assumes that there is little or no bias to a particular locus due to potential differences in chromatin accessibility.

5.2.2 Creation and confirmation of all four single gene deletions, strains RWY1-RWY4.

The first round of deletions was carried out by employing the *TRP1* targeting cassette, which was transformed into strain BY4705. After selection to tryptophan prototrophy genomic DNA of twenty four randomly-chosen positive transformants was each digested with *XhoI* restriction endonuclease and electrophoresed. This was followed by Southern blot hybridisation using the *THI5* gene probe which, for BY4705, gives rise to four positive *XhoI* fragments of different sizes (14 Kb, 10.2 Kb, 7.0 Kb and 3.8 Kb), each corresponding to one of the four *THI5* genes (Chapter 3,

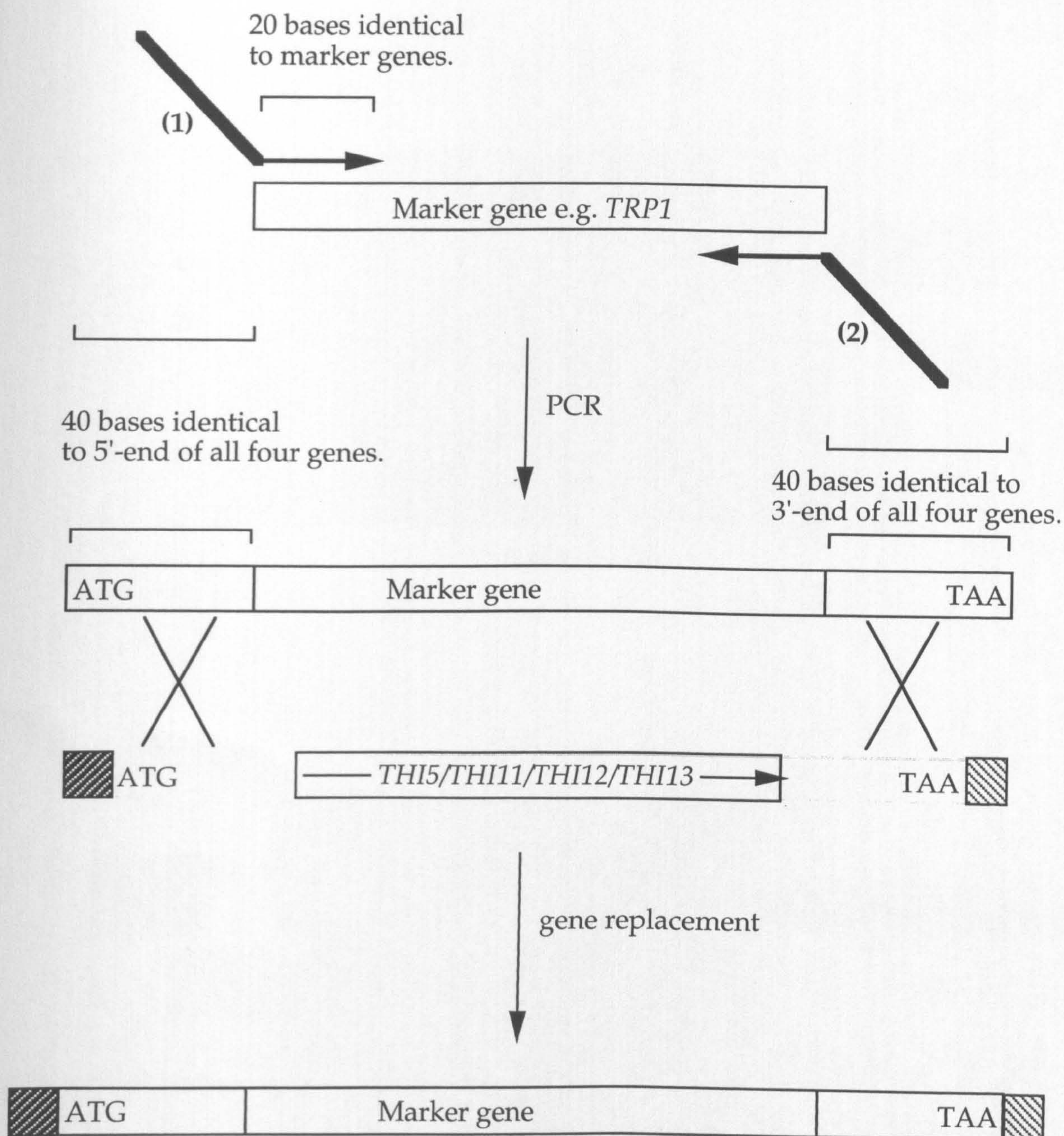
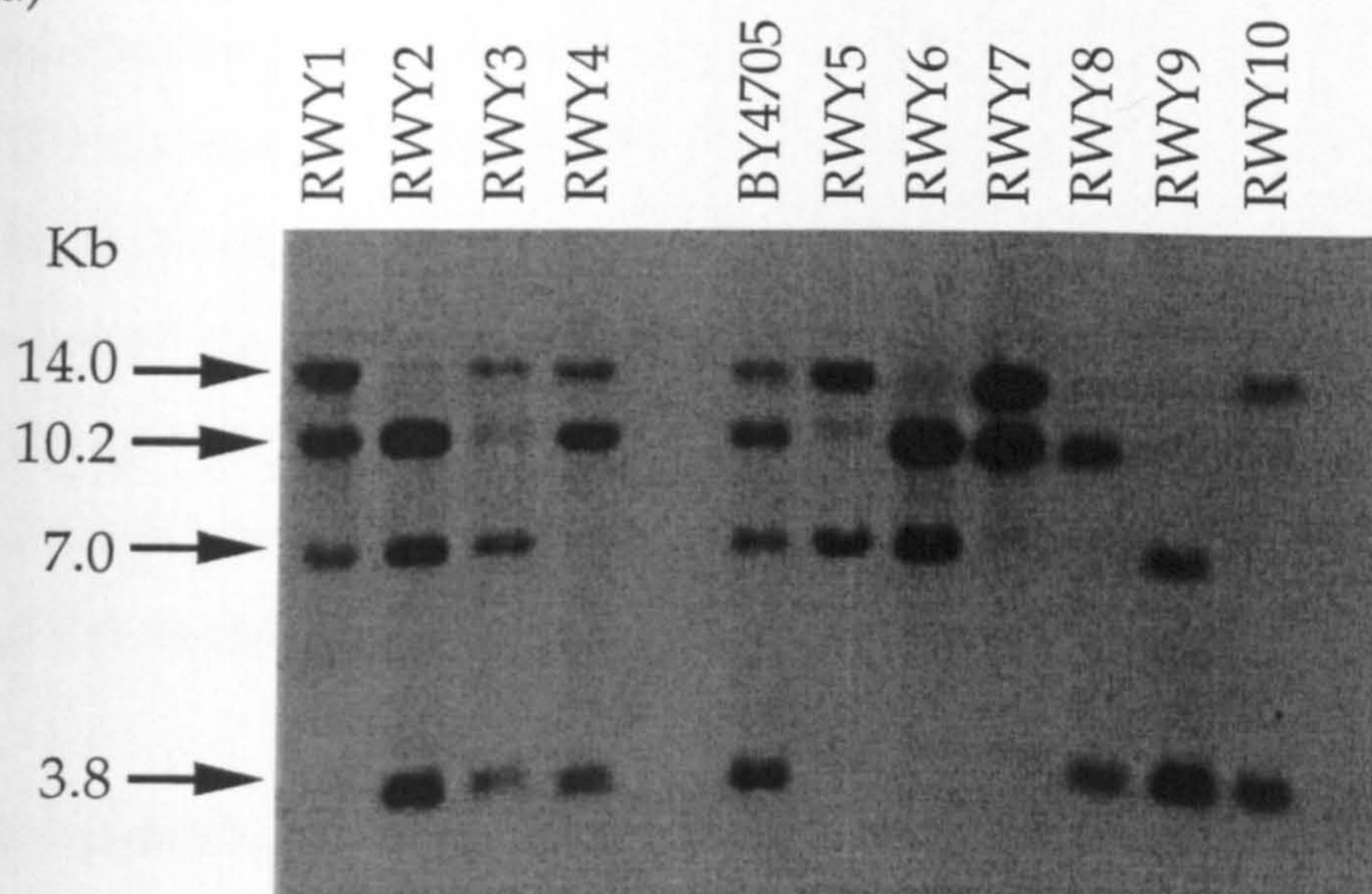


Figure 5.1 Construction of *THI5* disruption cassettes (*THI5* targeting cassettes).

The cassettes are synthesised by PCR of the marker gene using the pRS41X series of plasmids as template DNA (Brachmann *et al.* (1998)). A set of universal primers can amplify any nutritional marker from the plasmids. The oligonucleotides used are (1) *THI5*up and (2) *THI5*down. After transformation of the linear DNA into yeast these cassettes can integrate at any of the four *THI5* loci. After integration only 40 bp at the extreme ends of the gene remains.

a.)



b.)

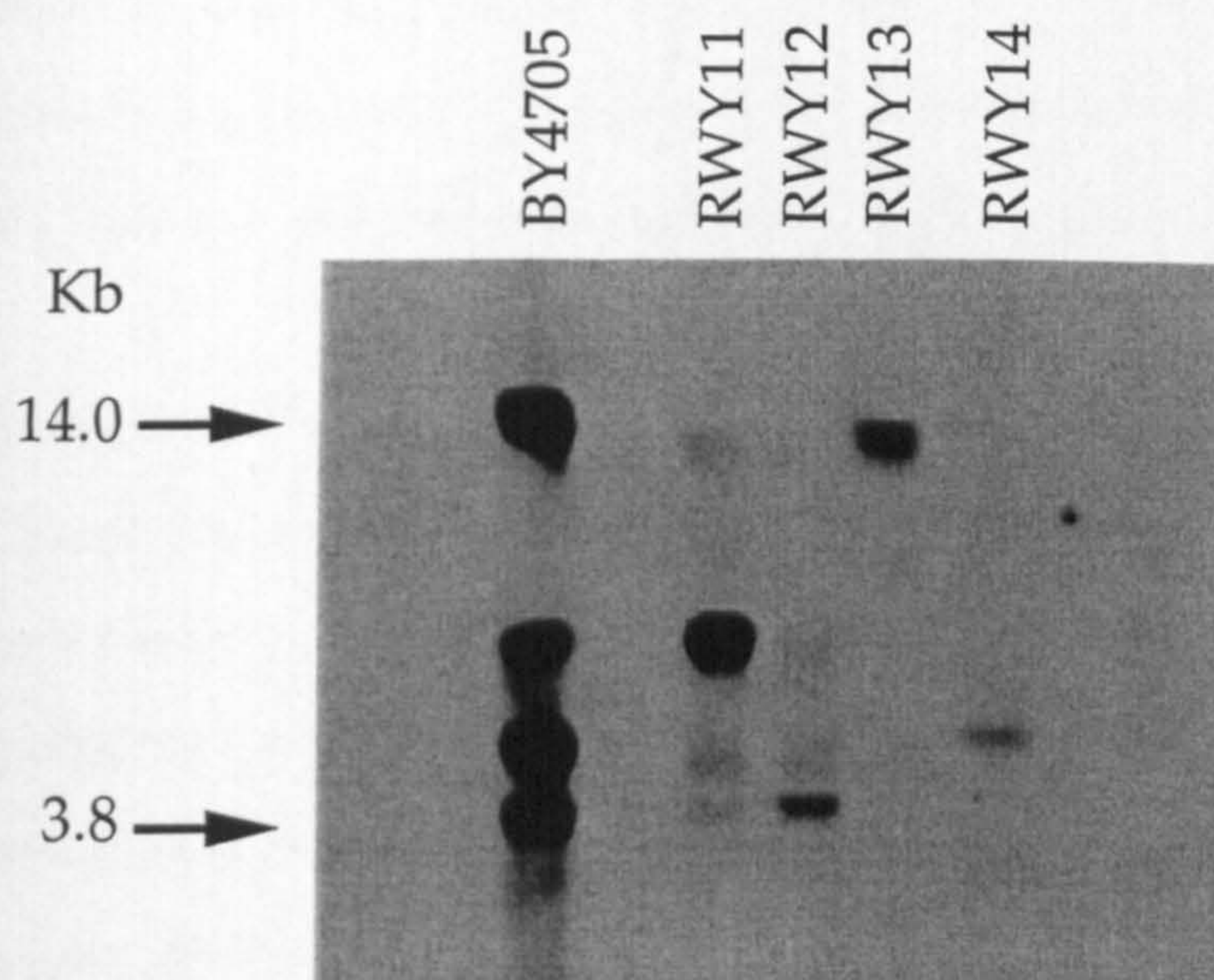


Figure 5.2 Confirmation of genotypes of the RWY mutant strains. Genomic DNA from each strain was digested with *Xho*I restriction endonuclease followed by electrophoresis. Southern blot hybridisation was carried out using the pRH4 *THI5* probe. Each of the genes *THI5*, *THI11*, *THI12* and *THI13* can be distinguished by different sized restriction fragments.

THI5 = 3.8 Kb, *THI11* = 14.0 Kb, *THI12* = 10.2 Kb, *THI13* = 7.0 Kb.

Shown here are the results after a.) the first and second round of targeted gene replacements b.) the third round of replacements.

Figure 3.1). Each transformant was scored for the presence/absence of each gene. Four transformants were found to be deleted for *THI5*, three for *THI11*, five for *THI12* and three for *THI13*. The remaining nine transformants possessed all four *THI5* family members and so these transformants must have arisen from either contaminating plasmid template DNA or integration of the *TRP1* cassette elsewhere in the genome. Confirmation of the mutant genotypes are shown in Figure 5.2a. The single gene disruption strains are RWY1 ($\Delta thi5$), RWY2 ($\Delta thi11$), RWY3 ($\Delta thi12$) and RWY4 ($\Delta thi13$).

5.2.3 Creation and confirmation of the six double mutants and the four triple mutants, strains RWY5-RWY14.

Construction of the single mutants has shown that there is no significant targeting bias of the disruption cassette and so this approach of random targeting to the *THI5* loci was used to create the other mutant strains.

The second round of deletions used the *HIS3* targeting cassette, which was transformed into each of the single mutants. Integration at a previously disrupted locus was selected against by using minimal defined medium with no added histidine and tryptophan. The genotypes of a sample of HIS^+ , TRP^+ transformants was determined by Southern blot hybridisation and it was found that all combinations of double mutants had been generated (Figure 5.2a). Although these strains have different genotypes, their nutritional requirements (HIS^+ , TRP^+) are identical. A genealogical tree describes the routes to the construction of these and all the other mutant strains and is illustrated in Figure 5.3.

The third round of deletions used the *LEU2* targeting cassette to generate all four triple mutants (figures 5.2b and 5.3). These strains contain just one member of the gene family and are designated RWY11, which possesses the wild-type *THI13* gene, RWY12 (*THI11*), RWY13 (*THI5*) and RWY14 (*THI12*). The triple mutants, like the six

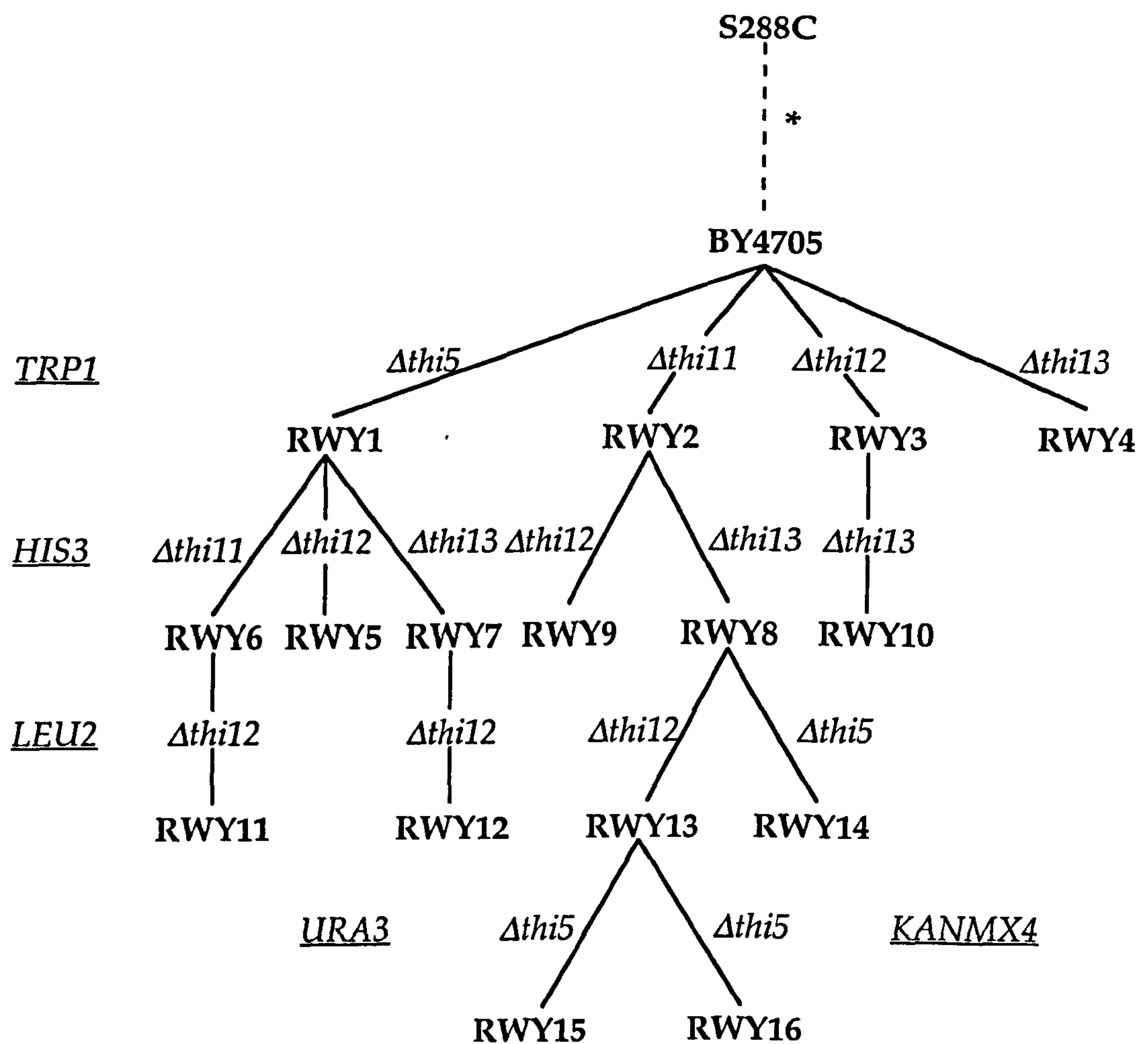


Figure 5.3 Genealogy of the RWY mutant strains. Each solid line indicates a replacement of the specified gene.

*The full ancestry of BY4705 is described by Brachmann *et al.* (1998).

double mutants, have the same overall phenotype (TRP1⁺, HIS⁺ and LEU⁺) despite having different genotypes.

5.2.4 The construction of the quadruple mutant strains, RWY15 and RWY16.

The final deletion used both the *URA3* and *KanMX4* cassette, which were transformed into mutant strain RWY13 to create strains RWY15 and RWY16 respectively. A PCR-based approach was designed and used to screen and confirm the genotypes. This involved designing primers that amplified the DNA between an integrated disruption cassette and a conserved region of the four promoters. Each cassette primer annealed at a different distance to this promoter region. In this way, all four mutant loci could be detected upon electrophoresis of the PCR reactions. Figure 5.4 shows use of the PCR analysis to confirm the genealogy of the mutant strains, RWY15 and RWY16. A list of all mutants along with their genotypes is displayed in Table 5.1. Strain RWY16 was constructed to allow the stable maintenance of THI⁺ clones from a *S. kluyveri* genomic library based on the *URA3*⁺ vector YEp352 as described in Chapter 4. RWY15 is the strain used in all subsequent phenotypic analyses.

5.3 Phenotypic analysis of the RWY mutant *S. cerevisiae* strains.

5.3.1 Phenotypic analysis with regards to thiamine biosynthesis.

The product of the single *S. pombe* *THI5* homologue, *nmt1*, is involved in the formation of the HMP moiety of thiamin (Chapter 1). It is assumed that at least one *S. cerevisiae* homologue has an identical role and deletion of all four *THI5* genes should prevent HMP biosynthesis. Phenotypic studies of the four triple deletion mutant strains (RWY11-14) and the single quadruple mutant (RWY15) was carried out in order to confirm the predicted HMP auxotrophy of RWY15 and to determine the functional redundancy of this gene family. Also, the positions of the four *S. cerevisiae* Thi5 isozymes in the HMP pathway was determined relative to the

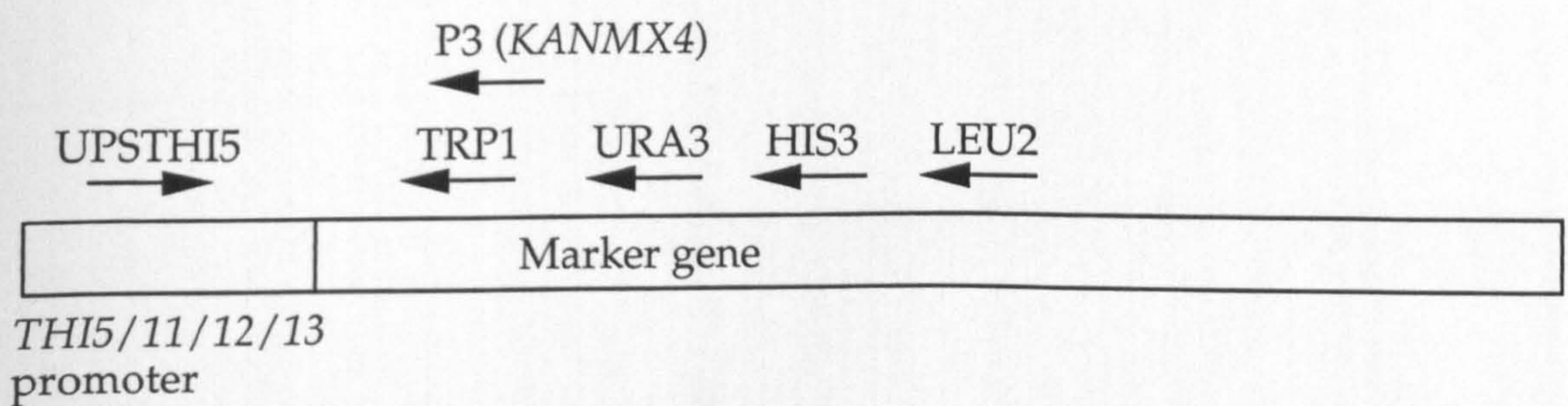
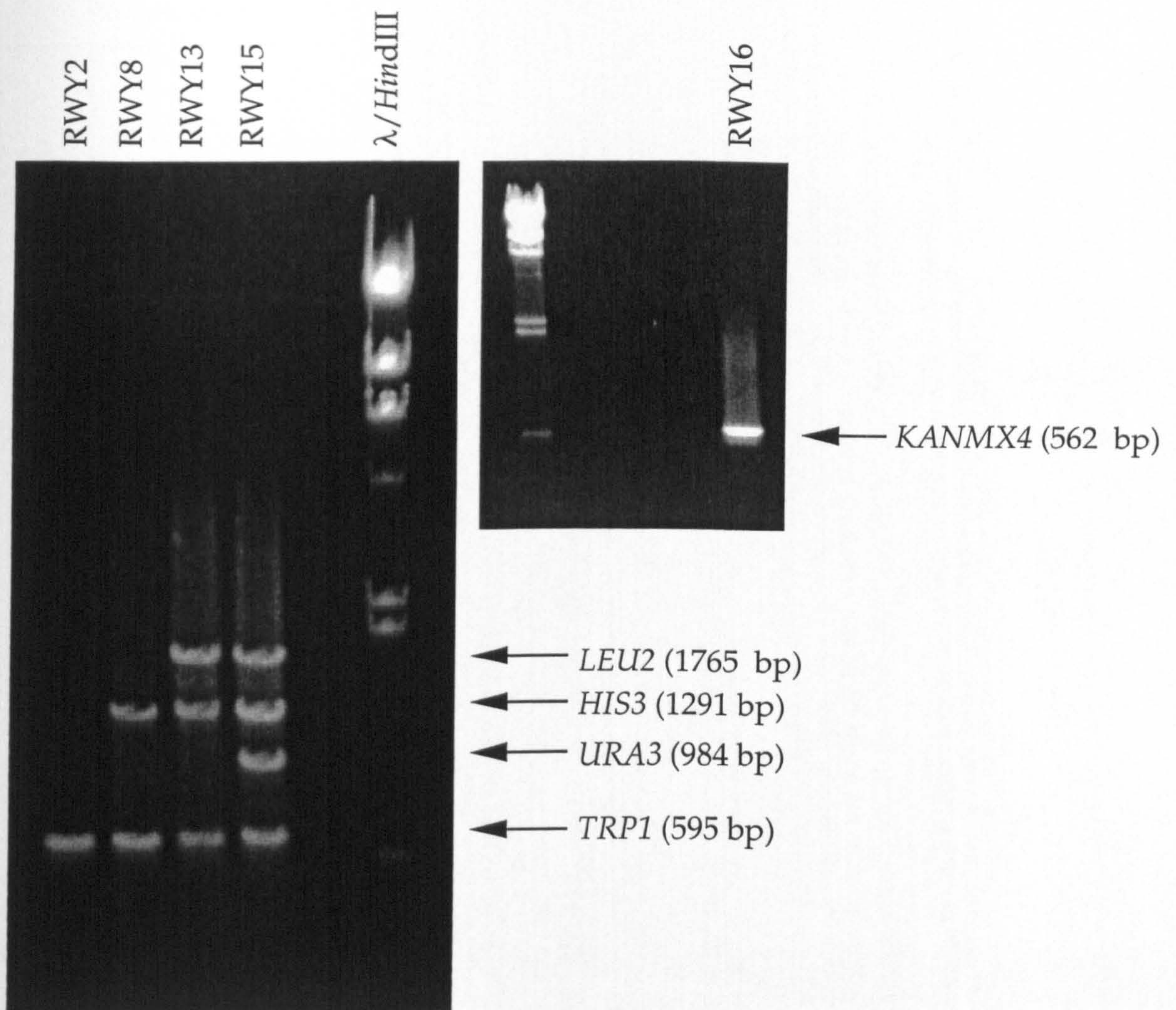


Figure 5.4 A PCR genealogy of RWY15. Four separate PCR reactions were performed for each mutant strain and the products were pooled and electrophoresed together (above). Each mutant allele can be detected by the different size fragments. Also shown is the presence of a *KANMX4* marker gene disruption in RWY16.

UPSTHI5 is the upstream primer common to all four promoters. The second primer is unique to each marker gene.

Strain	Genotype
BY4705	MAT α ade2 Δ ::hisG his3 Δ 200 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 ura3 Δ 0
RWY1	MAT α ade2 Δ ::hisG his3 Δ 200 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY2	MAT α ade2 Δ ::hisG his3 Δ 200 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0
RWY3	MAT α ade2 Δ ::hisG his3 Δ 200 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi12::TRP1 ura3 Δ 0
RWY4	MAT α ade2 Δ ::hisG his3 Δ 200 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi13::TRP1 ura3 Δ 0
RWY5	MAT α ade2 Δ ::hisG his3 Δ 200 thi12::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY6	MAT α ade2 Δ ::hisG his3 Δ 200 thi11::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY7	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY8	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0
RWY9	MAT α ade2 Δ ::hisG his3 Δ 200 thi12::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0
RWY10	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi12::TRP1 ura3 Δ 0
RWY11	MAT α ade2 Δ ::hisG his3 Δ 200 thi11::HIS3 leu2 Δ 0 thi12::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY12	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 thi12::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi5::TRP1 ura3 Δ 0
RWY13	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 thi12::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0
RWY14	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 thi5::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0
RWY15	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 thi12::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0 thi5::URA3
RWY16	MAT α ade2 Δ ::hisG his3 Δ 200 thi13::HIS3 leu2 Δ 0 thi12::LEU2 lys2 Δ 0 met15 Δ 0 trp1 Δ 63 thi11::TRP1 ura3 Δ 0 thi5::KANMX4

Table 5.1 The complete list of *S. cerevisiae* thi5 gene family mutants and their genotypes.

precursor and cofactor pyridoxine, an issue which has not yet been addressed in *S. pombe*.

Figure 5.5 compares the growth of these mutants on Wickerham's glucose medium supplemented, as indicated, with the vitamins thiamin and pyridoxine and the HMP precursor. Plate A contained both thiamin and pyridoxine. All the triple mutants and RWY15 showed identical growth to each other and to the BY4705 parent. Plate B was supplemented with pyridoxine and had no added thiamin. The triple mutants RWY11, which contains just *THI13*, RWY12 (*THI11*), RWY13 (*THI5*) and RWY14 (*THI12*) all exhibited identical growth to BY4705. However, RWY15 showed minimal growth indicative of thiamin auxotrophy. Therefore, each of the four genes can function in thiamin biosynthesis. The addition of HMP to this medium (plate C) resulted in growth of RWY15 comparable to the triple mutants and BY4705, indicating that the four gene products all function in the biosynthesis of the HMP moiety.

For Wickerham's medium without added pyridoxine but supplemented with thiamin, normal growth was seen for all strains (Plate D). Minimal growth, as displayed in plate B, is exhibited by RWY15 on medium lacking both pyridoxine and thiamin, with normal growth on the same medium supplemented with HMP (data not shown). Taken together, these data prove that Thi5p, Thi11p, Thi12p and Thi13p are functionally redundant in terms of HMP biosynthesis and that the enzymes function in the formation of this precursor from pyridoxine.

5.3.2 Effect of alternative carbon sources on thiamin auxotrophy of RWY15.

Almost all the published data examines thiamine biosynthesis with glucose as the sole carbon source. Under aerobic conditions, glucose is preferentially catabolised through the fermentative route which is due to the Crabtree effect (De Deken, 1966). The fate of pyruvate either towards fermentation products or through the TCA cycle involves the thiamin-dependent enzymes, Pdc and Pdh (Chapter 1). Other carbon

Figure 5.5 Phenotypic analysis of the triple mutants RWY11-RWY14 and the quadruple *THI5* mutant RWY15 with regards to pyridoxine and thiamin biosynthesis. Yeast was previously grown overnight in liquid SD medium containing the full range of amino acids and bases present in the SCM powder mix (see materials and methods). Cells were harvested at approximately 1×10^8 cells/ml. 5 μ l aliquots of 1/10 serial dilutions beginning with 1×10^8 cells/ml were spotted onto Wickerham's SCM agar medium supplemented as indicated with the vitamins pyridoxine(Pyr) and thiamin(Thi) and the HMP precursor. All agar plates were incubated at 28°C for 2 days. The arrow on plate B indicates the positions of the RWY15 serial dilutions.

Plate A: +Pyr +Thi

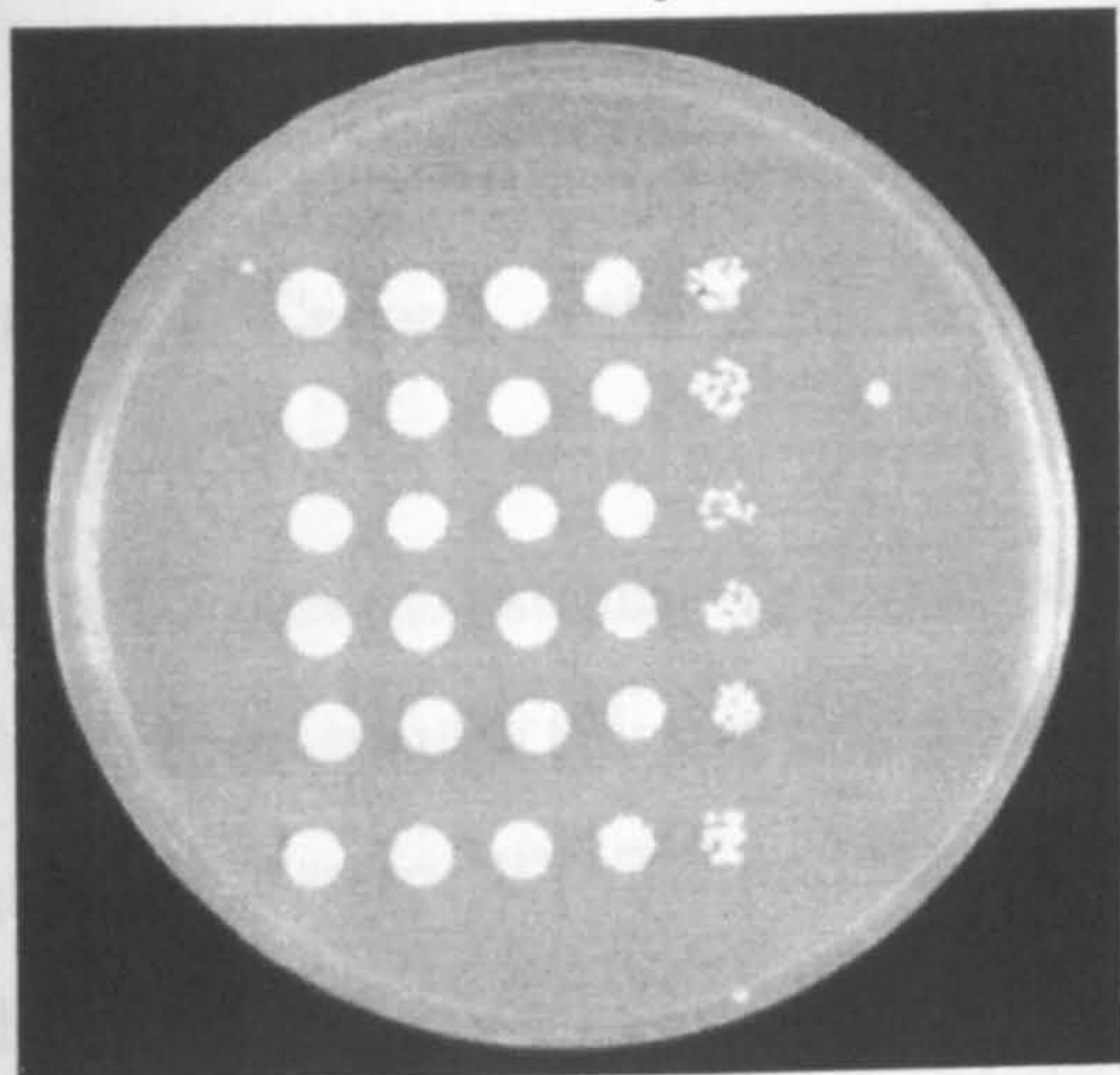


Plate B: + Pyr -Thi

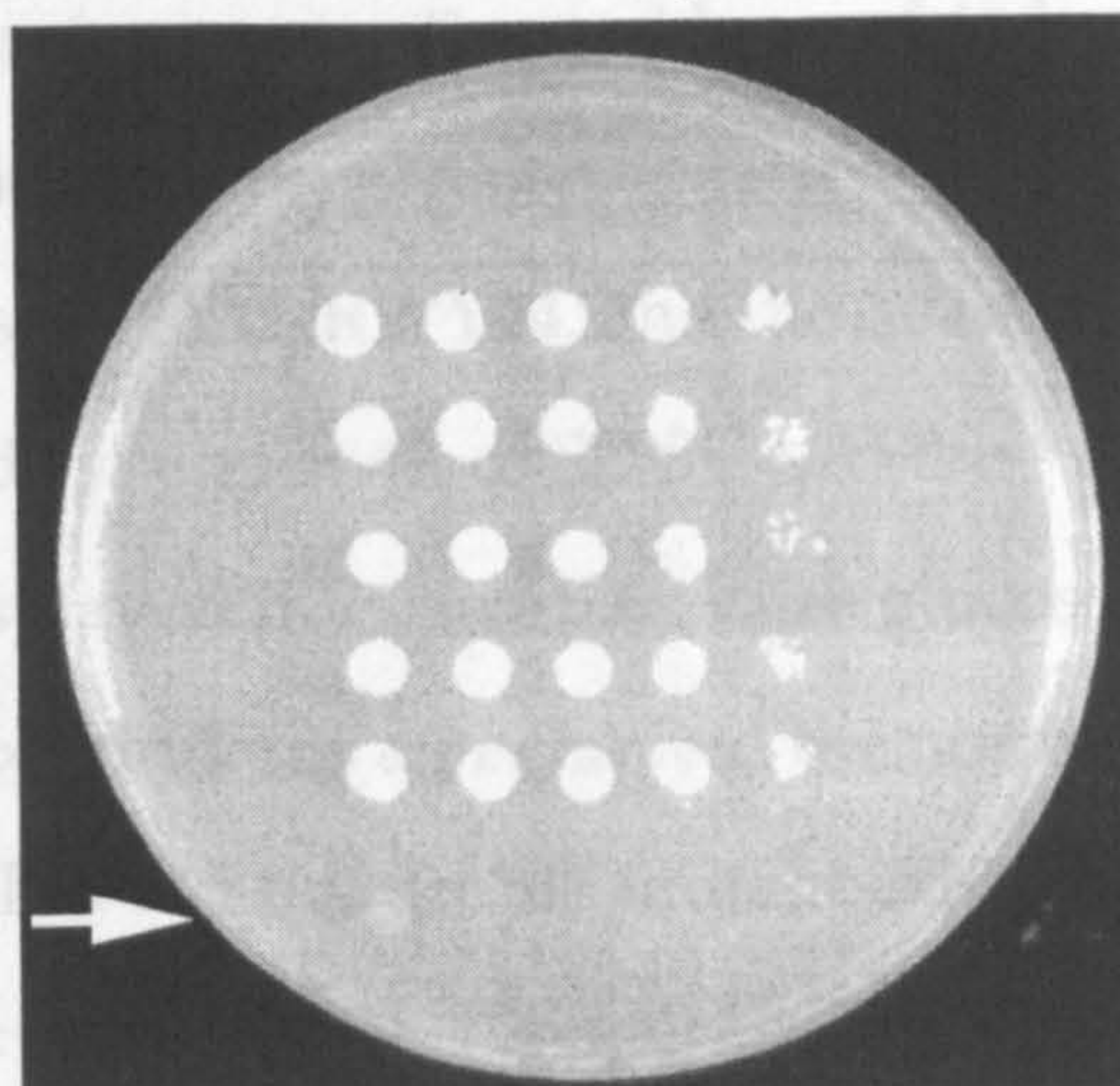


Plate C: +Pyr -Thi +HMP

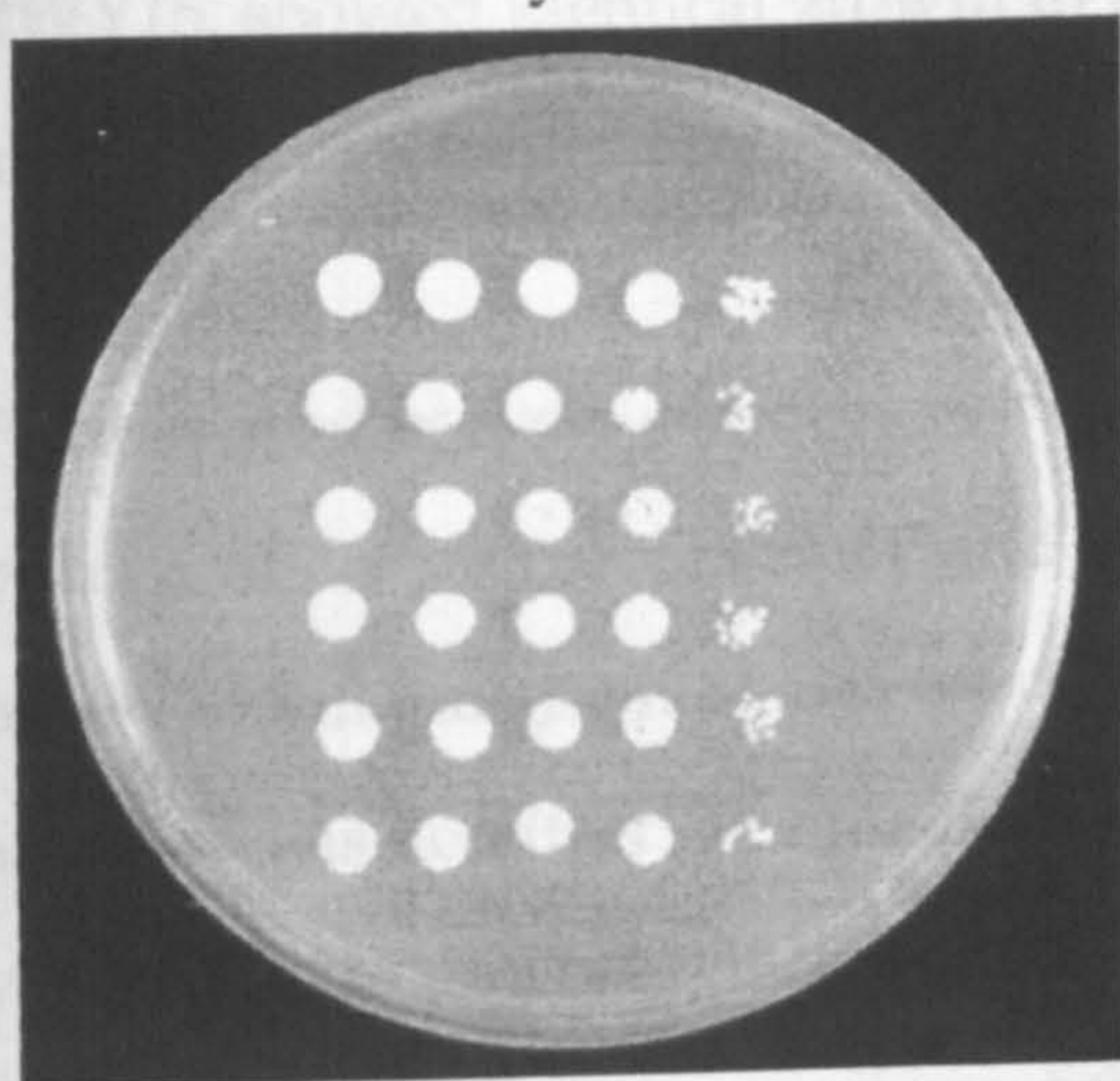
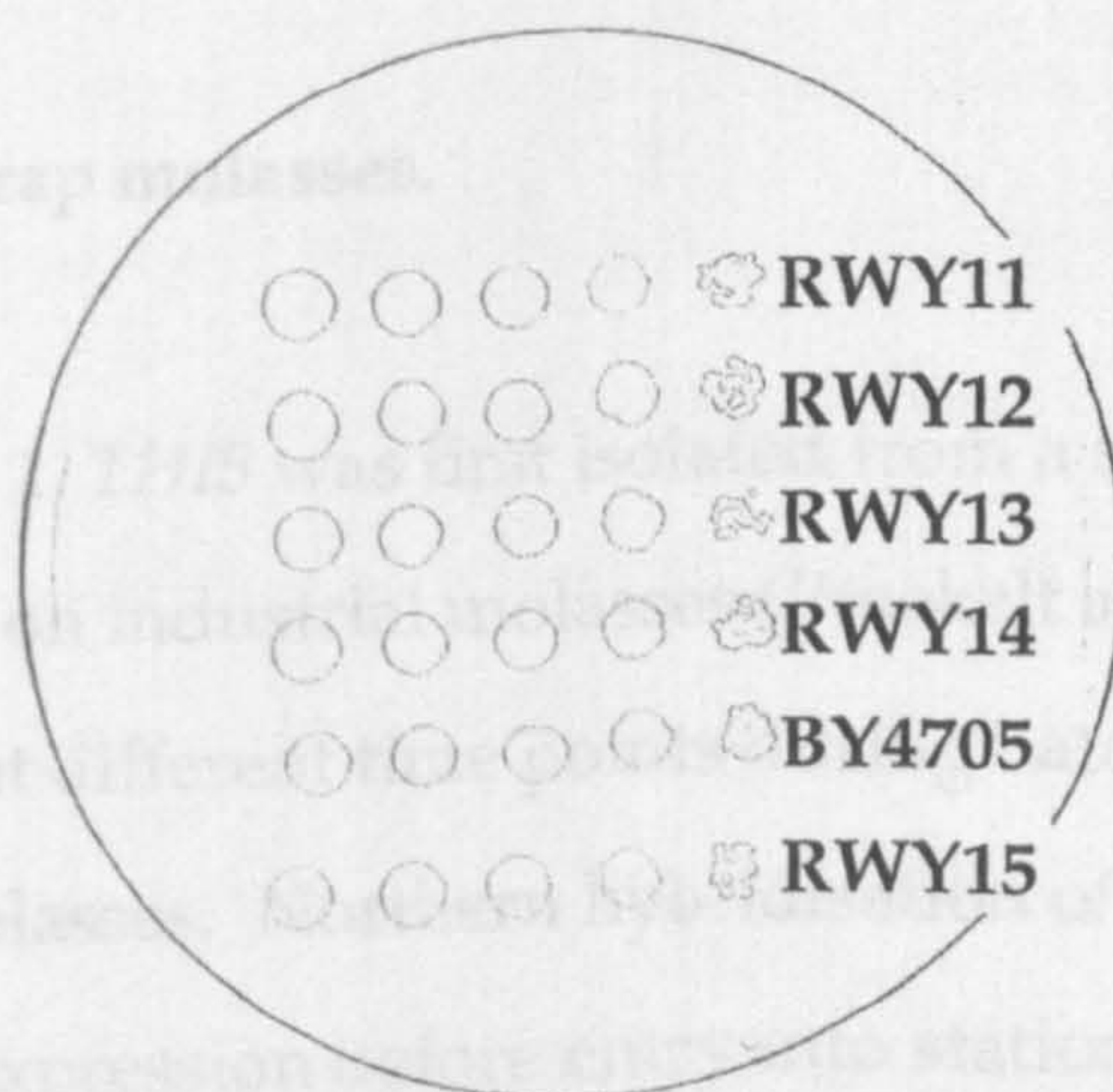
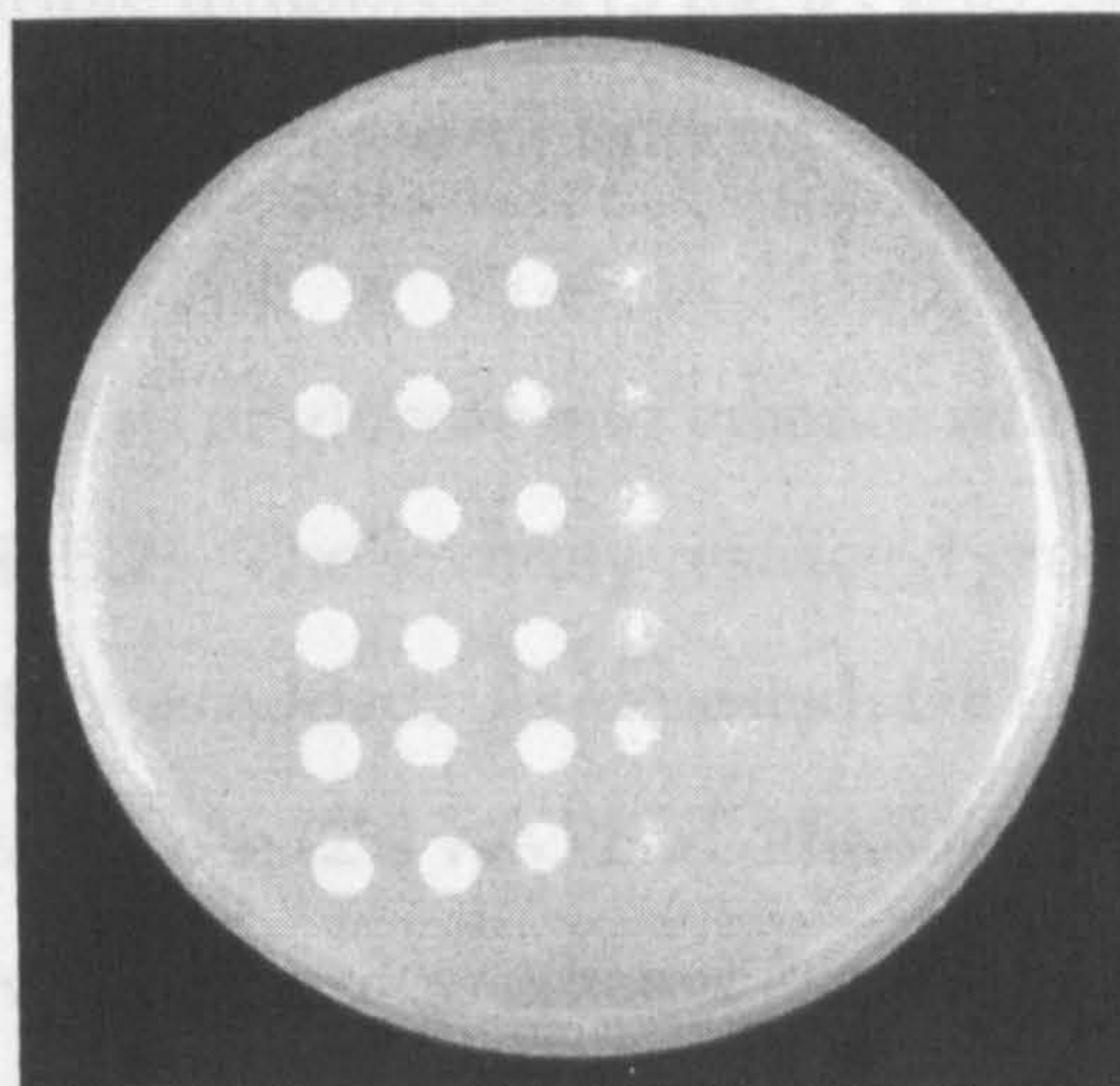


Plate D: -Pyr +Thi



sources can greatly alter the metabolic flux through these and other pathways which may affect the thiamin requirement. The role of *THI5*, *THI11*, *THI12* and *THI13* in HMP biosynthesis was examined on Wickerham's agar medium with no added thiamin and containing one of three sole carbon sources, glucose, galactose and ethanol. Growth of the triple deletion mutants, RWY11-RWY14, and the quadruple mutant, RWY15, were compared with the wild-type growth of S288C and the BY4705 parental strain. The agar plates are displayed in Figure 5.6. With glucose as the carbon source (Figure 5.6, plate A), the triple mutants all exhibited the normal growth seen for BY4705 and S288C whilst the RWY15 quadruple mutant exhibited the apparent thiamin auxotrophy described in section 5.3.1. On galactose medium (plate B), the pattern of growth was the same as for glucose. However, surprising results were obtained on ethanol medium (plate C). The quadruple knockout strain, RWY15, showed identical growth to the four triple mutants and to the BY4705 parent indicative of thiamin prototrophy. All these observations were fully reproducible.

Further studies were performed with combinations of glucose and ethanol with each other and the precursors HMP and HET (Table 5.2). The severely restricted growth of RWY15 in glucose medium was lifted if HMP was added. As expected, HET addition has no effect. For medium with both glucose (2% v/v) and ethanol (2% v/v) or a low glucose concentration (0.5% v/v) compared to ethanol (2% v/v), RWY15 exhibits the same apparent thiamin auxotrophy. Normal growth occurs in the same media supplemented with HMP but not HET.

5.3.3 Growth on blackstrap molasses.

As described in Chapter 1, *THI5* was first isolated from a screen of genes expressed only at stationary phase on industrial molasses (Praekelt and Meacock, 1992). Total RNA has been isolated at different time points during batch culture of *S. cerevisiae* grown on blackstrap molasses. Northern hybridisation of the RNA with a *THI5* probe shows transient expression before entry into stationary phase (Hather, 1996).

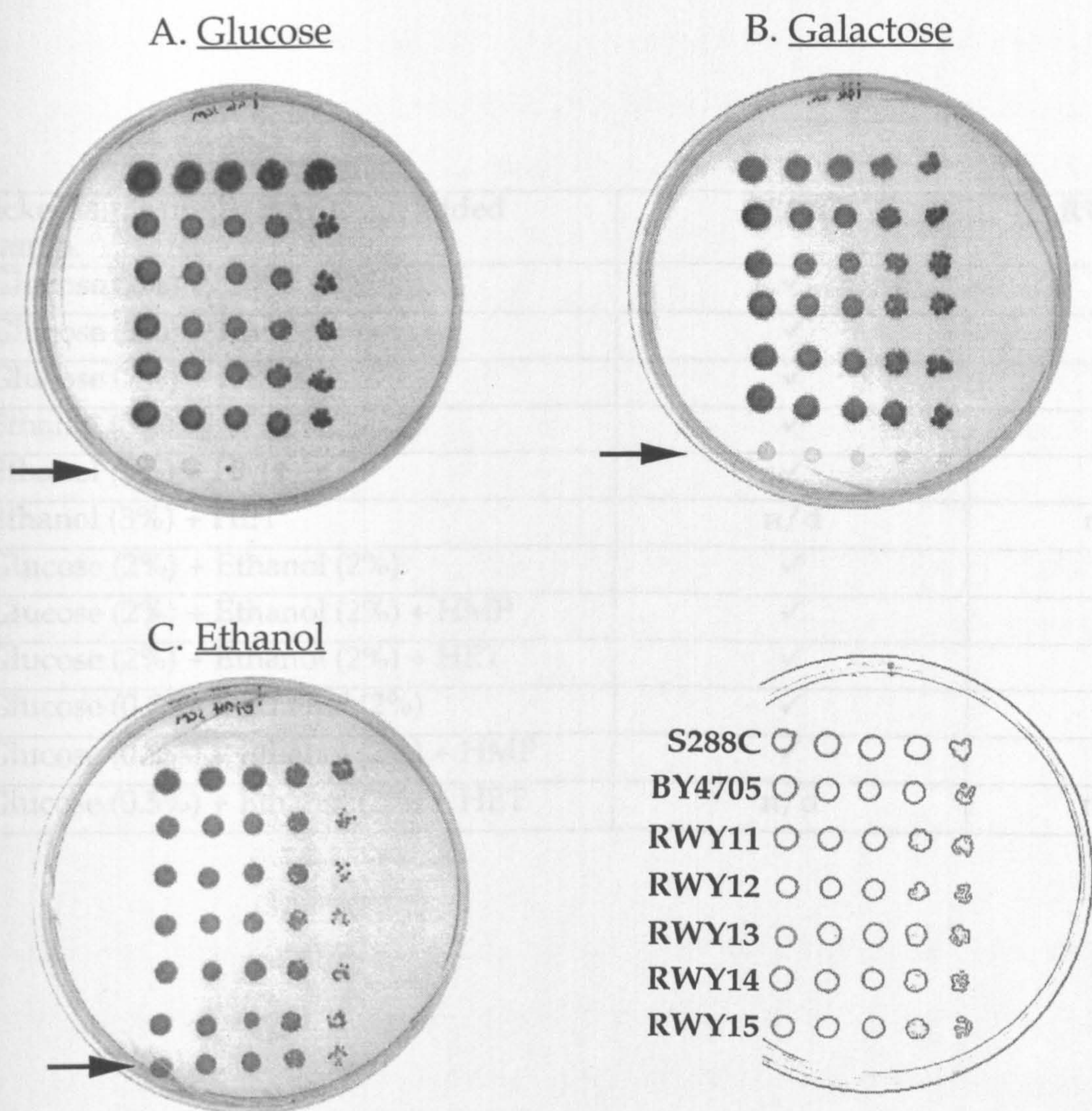


Figure 5.6 Growth of mutants RWY11-RWY15 compared with BY4705 parent and S288C on different carbon sources. Yeast was previously grown overnight in liquid YPD medium to late exponential phase. Cells were washed and diluted to an optical density of 0.25 at 600 nm. 5 μ l aliquots of 1/5 serial dilutions were spotted onto Wickerham's agar medium containing all vitamins (except thiamin) and the full range of amino acids and bases as present in the SCM powder mix (see materials and methods). Glucose (plate A), galactose (plate B) or ethanol (plate C) was added to the medium which constituted the sole carbon source for yeast propagation. All agar plates were incubated at 28°C for 4 days. Arrows indicate the position of RWY15 inoculation.

Wickerham's medium without added thiamin.	S288C	RWY15
+ Glucose (2%)	✓	X
+ Glucose (2%) + HMP	✓	✓
+ Glucose (2%) + HET	✓	X
+ Ethanol (3%)	✓	✓
+ Ethanol (3%) + HMP	✓	✓
+ Ethanol (3%) + HET	n/d	n/d
+ Glucose (2%) + Ethanol (2%).	✓	X
+ Glucose (2%) + Ethanol (2%) + HMP	✓	✓
+ Glucose (2%) + Ethanol (2%) + HET	✓	X
+ Glucose (0.5%) + Ethanol (2%)	✓	X
+ Glucose (0.5%) + Ethanol (2%) + HMP	✓	✓
+ Glucose (0.5%) + Ethanol (2%) + HET	n/d	n/d

Table 5.2 Phenotypic comparisons between *S. cerevisiae* strains S288C and RWY15. Strains were previously grown overnight in liquid YPD medium to late exponential phase. Cells were washed and diluted to an optical density of 0.25 at 600 nm. 1/5 serial dilutions were spotted onto Wickerham's SCM agar medium containing all vitamins (except thiamin) and containing the carbon sources indicated. Values in brackets represent the final percentage w/v for glucose or v/v for ethanol. The medium was supplemented with HMP or HET where indicated. All agar plates were incubated at 28°C for 3 days. Growth was scored for each strain. ✓ denotes wild type growth, X denotes minute growth. n/d – not done.

With the full range of *THI5* mutants currently available, I decided to see whether disruption of all four genes results in an obvious growth defect on this medium. Growth of the triple *THI5* mutants, RWY11-RWY14, and the single quadruple mutant, RWY15, were compared to the BY4705 parental strain and S288C when grown on agar plates containing blackstrap molasses. The plates were incubated for 4 days at 28°C and the results are displayed in Figure 5.7. S288C exhibited slightly more growth than the other strains tested, including BY4705. This may be due to the presence of several auxotrophic mutations in the latter strain. However, all the nutritional requirements were included in the medium. More importantly, each of the triple mutants and the quadruple mutant RWY15 showed the same degree of growth as the BY4705 parent. Therefore, growth is not obviously perturbed by the concurrent deletion of *THI5*, *THI11*, *THI12* and *THI13*.

5.4 Discussion.

The construction of *S. cerevisiae* strains containing all combinations of *THI5* deletions has been performed and their genotypes confirmed. The four genes *THI5*, *THI11*, *THI12* and *THI13* are functionally redundant in terms of HMP formation for thiamin biosynthesis. The mutant strain RWY15, which possesses deletions of all four members, exhibits an apparent thiamin auxotrophy on glucose. This phenotype is labelled as 'apparent' because there is published data that describes evidence of a minor alternative HMP biosynthetic pathway. The consequence of this second pathway in terms of the *THI5* gene family and the RWY15 mutant will be addressed in the next chapter.

The HMP precursor is synthesised from pyridoxine (reference). Phenotypic analysis of the *S. pombe nmt1*⁻ mutant has always been carried out on minimal medium equivalent to Wickerham's medium but without added pyridoxine. Although never suggested, the *nmt1* protein could always have been a pyridoxine biosynthetic enzyme encoded by a gene that is regulated by thiamin. Such enzymes exist in *S. cerevisiae*, encoded by the subtelomeric members of the *SNO* and *SNZ* gene families

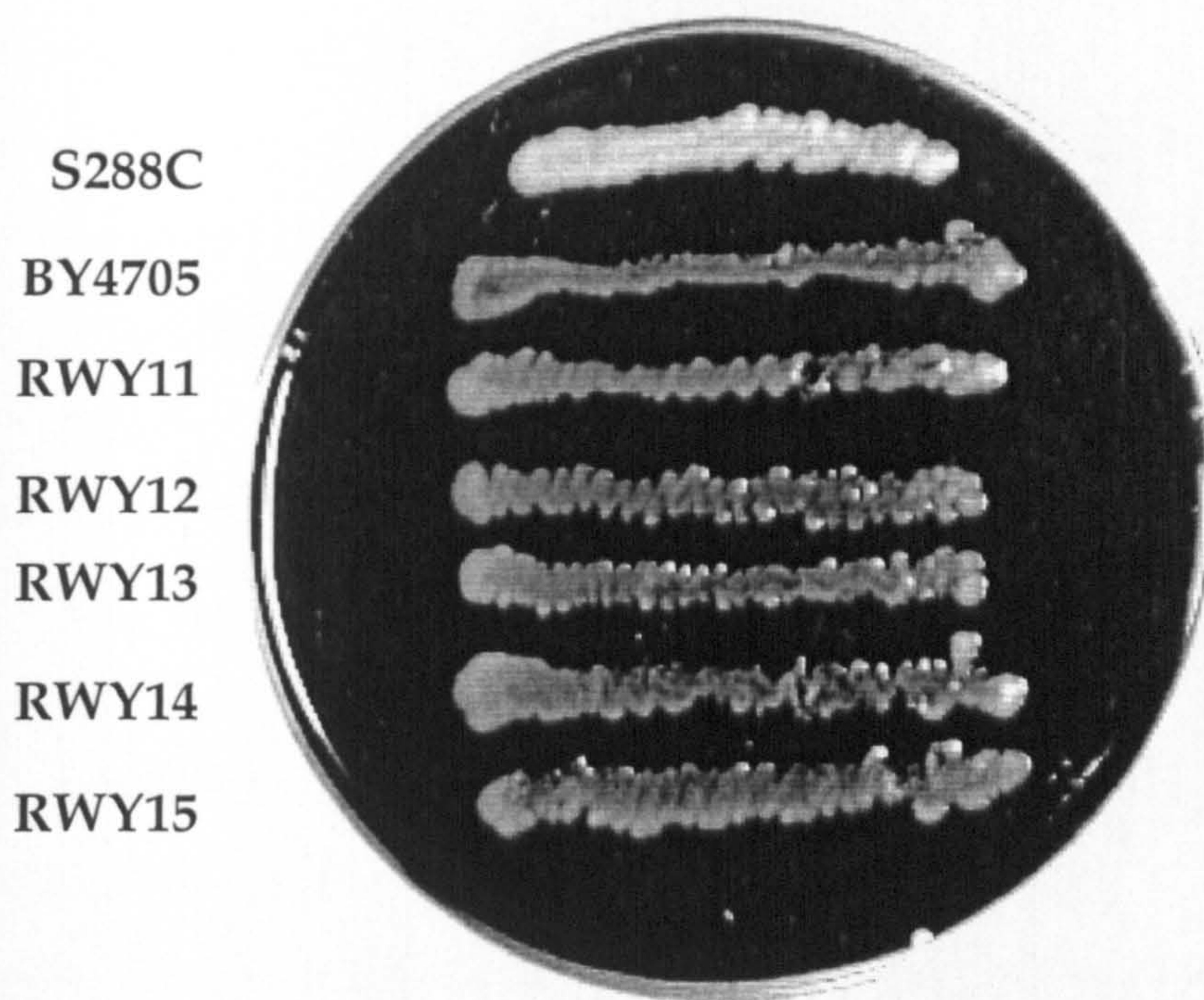


Figure 5.7 Growth of the triple mutant strains RWY11-RWY14 and the quadruple mutant RWY15 compared to the BY4705 parental strain and S288C on blackstrap molasses. The molasses agar medium was prepared as described in materials and methods and contains the SCM powder mix. Each yeast strain was streaked onto the molasses agar from a single colony pregrown on YPD agar medium and then incubated at 28°C for 4 days. A control was prepared to check that growth is not due to residual intracellular thiamin obtained from the YPD medium. A YPD colony of RWY15 and S288C was used to inoculate a Wickerham's agar plate with no added thiamin and was incubated alongside the molasses agar (data not shown). After four days on the Wickerham's agar, large colonies were seen for S288C and no growth for RWY15 as previously described in section 5.3.1.

(Marsh and Meacock, unpublished). For the *S. cerevisiae* strain RWY15, the presence or absence of added pyridoxine in the medium has no effect upon the apparent thiamin/HMP auxotrophy. Therefore, each of the four Thi5p isozymes catalyses a step somewhere between pyridoxine and HMP.

Studies of fungal thiamin biosynthesis routinely use glucose as the carbon source. It is expected that the thiamin requirement will not change appreciably for other carbon sources due to the number of key central metabolic enzymes that use the active cofactor (Chapter 1). To check this assumption, growth of each of the triple mutants and the quadruple mutant, RWY15, was compared with the BY4705 parental strain on minimal medium without added thiamin. This medium contained either glucose, galactose or ethanol as the carbon source. Growth on galactose was severely perturbed for the RWY15 strain which contains deletions of all four genes, *THI5*, *THI11*, *THI12* and *THI13*. The triple mutants exhibited wild-type growth signifying that this gene family is functionally redundant which is in agreement with the results obtained for glucose. Both galactose and glucose catabolism mostly proceeds through glycolysis to pyruvate. The two main fates of this central metabolite, either towards ethanol production or into the TCA cycle, each involve a step that is catalysed by a thiamin-dependent enzyme.

Other metabolic routes which require the active cofactor will also add to the thiamin drain but to a lesser extent, especially in the media used in these experiments as it contained the full range of amino acids and several bases that make up the SCM powder mix. Therefore enzyme activities, like that of Ilv2p, will be dispensable. This may explain in part the observation of the wild-type growth of RWY15 on minimal medium with ethanol as carbon source. Ethanol utilisation can completely bypass the pyruvate branch as ethanol is metabolised via acetate through the citric acid cycle whilst macromolecular biosynthesis through gluconeogenesis proceeds via oxaloacetate and phosphoenol pyruvate (Walker, 1998) pp223-224. Even in the absence of the SCM mix, RWY15 exhibited wild-type growth indicating a greatly reduced thiamin requirement (data not shown). It is possible that in this strain, the

alternative HMP biosynthetic pathway mentioned earlier in this section may be sufficient for thiamin formation on ethanol medium.

Growth of the RWY15 strain was subsequently examined for growth on the same Wickerham's medium but containing two carbon sources, glucose and ethanol. This strain showed the same retarded growth phenotype as for medium containing just glucose. As this sugar is always preferentially metabolised prior to other available carbon sources, there will be an initial drain on the thiamin pool possibly due to flux through pyruvate towards fermentation products. This pool will not be sufficiently replenished in RWY15.

Taken together, these data suggest that the enzyme activity encoded by the *THI5*, *THI11*, *THI12* and *THI13* genes is dispensable for certain carbon sources such as ethanol but indispensable for others like glucose and galactose. Therefore, the pattern of *THI5* gene expression will likely reflect the substrate being metabolised. Based on these observations, we can predict the expression profile of *THI5* in a batch culture of wild-type *S. cerevisiae* which contains glucose as the sole carbon source in medium lacking thiamin. Glucose assimilation to ethanol until the diauxic shift would be accompanied by an abundance of *THI5* mRNA. These transcripts may not be detected at the post-diauxic shift when the ethanol is subsequently metabolised. This experiment has been carried out and will be presented as part of the further physiological analyses in Chapter 6. The results are in total agreement with this hypothesis.

The *S. cerevisiae* *THI5* gene was first isolated due to its differential expression between exponential and stationary phase in molasses batch culture (Chapter 1). The transient appearance of *THI5* mRNA occurs towards the end of exponential growth followed by a rapid decline (Hather, 1996). Growth of the *THI5* triple mutants and the RWY15 quadruple mutant was examined on blackstrap molasses, the medium used in the original experiments. None of the mutants exhibited

severely defective growth indicating that *THI5* is not essential for yeast propagation on this medium.

Blackstrap molasses is an industrial by-product of the refining of sugar cane. It is commonly used as animal feed and in the large-scale production of yeast (U.S. Sugar Corporation, <http://www.suga-lik.com/blackstrap.html>). Its composition has been studied and is found to contain thiamin at approximately 0.9 mg/kg (Curtin, L.V., National Feed Ingredients Association report, <http://www.ifas.ufl.edu/~ona/mol.pdf>). In the experiments described previously and in this chapter, the concentration of molasses was 100g/l and equates to a thiamin concentration of approximately 0.26 μ M. This is ten times lower than the amount in supplemented Wickerham's medium (2 μ M). It is likely that expression of *THI5* at the end of exponential growth of *S. cerevisiae*, on blackstrap molasses, is due to the exogenous depletion of this vitamin. This has been examined experimentally for the *THI4* gene, encoding the enzyme involved in HET formation. Like RWY15, a *S. cerevisiae thi4* mutant strain exhibits identical growth to wild-type on molasses medium (Praekelt and Meacock, 1992). The pattern of *THI4* gene expression during growth on molasses can be mimicked using Wickerham's medium supplemented with 0.15 μ M thiamin, which is comparable to the amount in the blackstrap molasses (Praekelt *et al.*, 1994).

Chapter 6

Physiological characterisation of *THI5*, *THI11*, *THI12* and *THI13*.

6.1 Introduction.

The physiological analyses presented in this chapter were designed to address three questions:

- (1) What is the nature of the severely restricted growth phenotype that was observed for the *S. cerevisiae* RWY15 mutant strain when inoculated onto thiamin-deficient medium (Chapter 5)?
- (2) Is it possible to confirm the presence/absence of a previously proposed alternative HMP biosynthetic pathway (Chapter 1, section 1.5.5)?
- (3) What can we learn about the regulation of each of the four *THI5* genes by making use of the mutant strains constructed in Chapter 5 and by analysing the activity of promoter:reporter constructs?

6.2 Phenotypic comparison of *S. cerevisiae* strain RWY15 and the *K. lactis klthi5* mutant.

6.2.1 Phenotypic analysis of the *K. lactis klthi5* mutant.

The single-copy *THI5* homologue of *K. lactis*, called *KLTHI5*, has been previously disrupted (Walsh D, unpublished). As no part of this work has been described elsewhere, a brief account of the disruption of this gene will be presented here.

The strain used for the construction of a *klthi5* mutant was the *K. lactis* tryptophan auxotroph MSK-110 (*uraA trp1:URA3*), a derivative of the CBS 2359 reference strain described in Chapter 4 (Stark & Milner, 1989). A disruption vector was made where the 0.55 Kb *ClaI*/*BglII* fragment of pDW5b (Chapter 4) was removed and replaced by a 1.7 Kb *ClaI*/*KpnI* fragment of pJ2B3 (Stark and Milner, 1989) carrying the *K. lactis TRP1* locus. Linearised DNA was transformed into MSK-110 and selected to

tryptophan prototrophy. Integration at the *KLTHI5* locus was confirmed by Southern blot hybridisation. This resulted in a strain where the whole N-terminal half of the *KLTHI5* gene has been deleted. Initial phenotypic analysis by D Walsh was performed on Wickerham's medium without thiamin and apparently showed that this mutant strain was a thiamin auxotroph. This auxotrophy was rescuable by addition of either thiamin or HMP and so shared an identical phenotype with the *S. pombe nmt1* mutant.

From here, I decided to check this observation and found that after two days incubation on the same medium without thiamin, there was no growth of the *klthi5::TRP1* mutant whilst the parental MSK-110 strain exhibited distinct colonies. However, after three day's incubation residual growth could be seen for the mutant and by four days distinct colonies were observed. After seven day's incubation the mutant showed colonies of similar size to the MSK-110 parent (data not shown). It was suggested (E. Schweingruber, personal communication to P Meacock) that any residual growth is probably due to thiamin released by lysis of dead yeast cells. To investigate this, the MSK-110 parental and the *klthi5* mutant were incubated for five days on Wickerham's medium followed by inoculation of a single colony of each strain onto fresh medium with incubation for a further five days. This was repeated twice consecutively and the results are displayed in Figure 6.1. This cultivation and re-inoculation should deplete any residual thiamin in the yeast which may have come from lysed cells; it can be seen that there is noticeable growth of the *klthi5* mutant but with smaller colonies than the parental strain. Taken together, these observations appear to indicate that the *klthi5* mutant is not a strict thiamin auxotroph but instead exhibits a slow growth phenotype without the vitamin.

6.2.2 The slow growth phenotype of RWY15.

The slow growth of the *klthi5* mutant was compared with the *S. cerevisiae* quadruple *THI5/11/12/13* mutant RWY15. RWY15 and a wild-type strain, S288C were inoculated onto Wickerham's agar medium lacking thiamin. After a normal

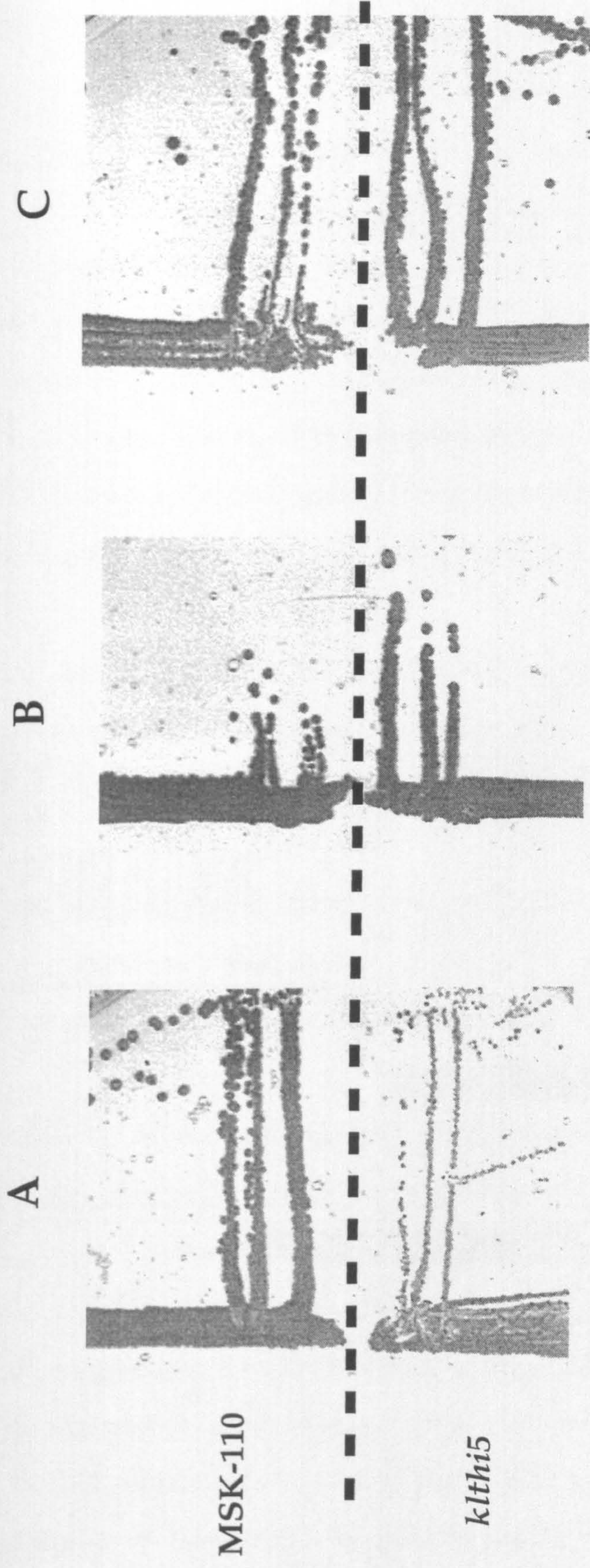


Figure 6.1 Phenotypes of the *K. lactis thi5* mutant strain (*klthi5*) with respect to thiamin biosynthesis and comparison with the parental *K. lactis* strain (MSK-110). Yeast were inoculated onto Wickerham's SCM agar glucose medium supplemented with either HMP, thiamin or without thiamin. After 5 days incubation a single colony from each yeast was inoculated onto identical fresh medium and incubated for a further 5 days. This process of re-inoculation was repeated twice. Shown above are inverted photographs showing growth of the two *K. lactis* strains on minimal defined medium without added thiamin (A), with HMP (B) and with thiamin (C).

incubation period of two to three days, RWY15 exhibited the thiamin auxotrophy as reported for the *S. pombe nmt1* mutant (Chapter 5). After 7 days, minute colonies of RWY15 were visible whilst very large colonies were observed for S288C (data not shown). At ten days, one of these minute colonies was streaked onto fresh thiamin-deficient Wickerham's agar medium and both plates were incubated for a further six days. Figure 6.2 shows the growth of S288C and RWY15 from the original agar plate after a total of sixteen days incubation and from the re-inoculated plate which had been incubated for six days. RWY15 shows a very severely restricted growth phenotype compared to the wild-type strain. This phenotype is very different to that observed for the *K. lactis klthi5* mutant (compare figures 6.1 and 6.2). Whilst it can be concluded that the *K. lactis* mutant is not a strict thiamin auxotroph, it is difficult to classify accurately the phenotype of RWY15. No differences in cell morphology were observed between RWY15 and S288C.

In order to investigate the RWY15 phenotype under more controlled conditions, I decided to follow growth in liquid batch cultures. The data are presented in the following section.

6.3 Aerobic vs. anaerobic thiamin biosynthesis in *S. cerevisiae*.

6.3.1 The growth of RWY15 compared to parent strains.

In Chapter 1, I reported the recent work of Tanaka *et al.* (Tanaka *et al.*, 2000) who present data indicating an alternative route to HMP formation, during anaerobiosis, which does not proceed via pyridoxine. The effects of aerobic and anaerobic conditions on the observed growth of *S. cerevisiae* strains S288C, BY4705 and RWY15 were investigated with regards to thiamin biosynthesis. The strains were inoculated, in duplicate, on Wickerham's glucose agar medium with or without thiamin, or with HMP or HET replacing the vitamin. Both sets of plates were incubated at the same temperature with one set placed inside an apparatus that generates carbon dioxide and hydrogen, thus displacing oxygen so that cells were under anaerobic conditions.

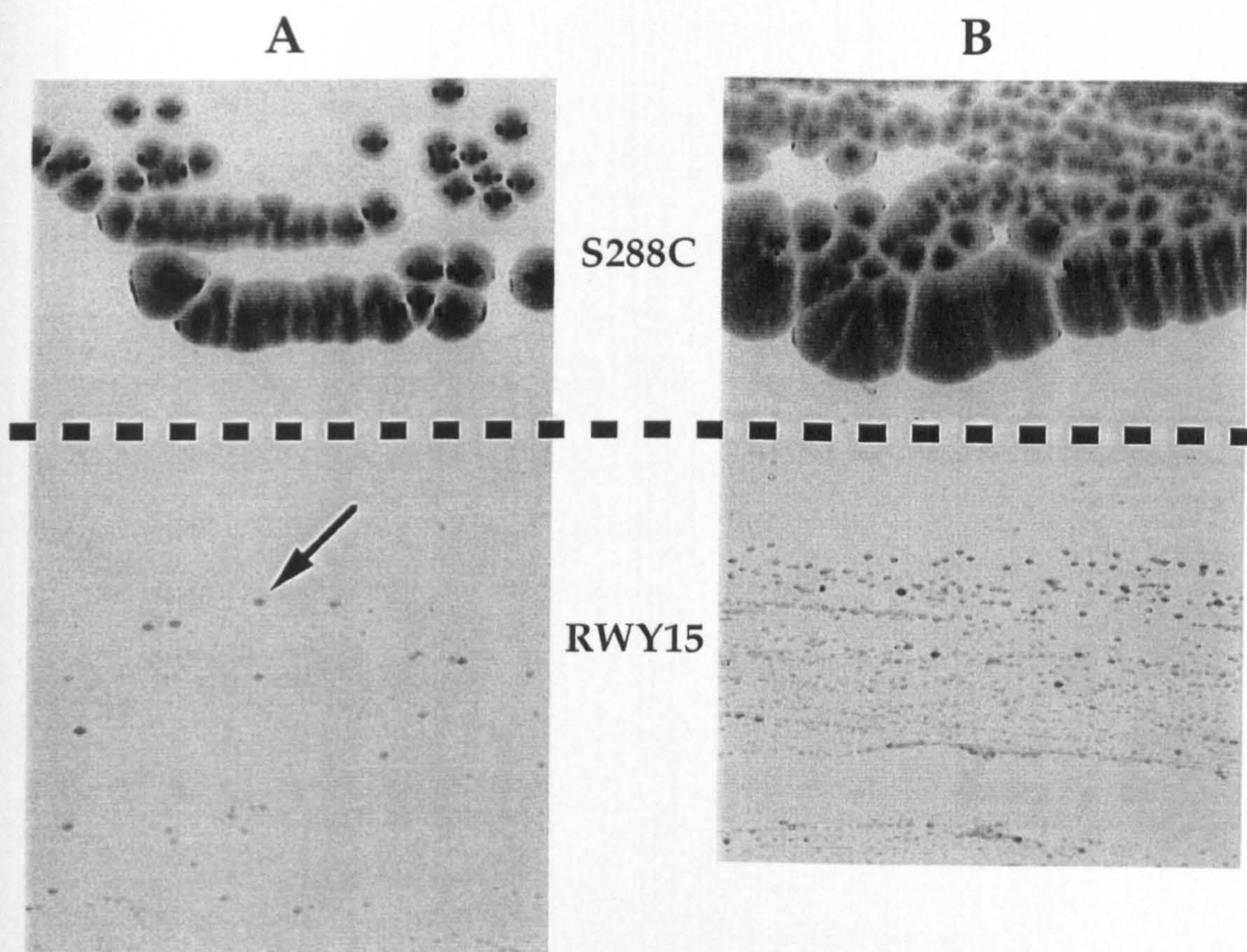


Figure 6.2 Investigation of the slow growth phenotype of RWY15 on Wickerham's agar glucose medium without thiamin. Shown on plate B is the growth of S288C and RWY15 after 16 days incubation at 28°C. After ten days incubation of plate B, a colony from each yeast was innoculated onto fresh medium and incubated for 6 days (plate A). The colours of the plate photographs have been inverted to aid the visualisation of the RWY15 colonies. One such colony is highlighted by an arrow on plate A.

The growth resulting after seven days incubation is shown in Figure 6.3. On medium with added thiamin all three strains displayed similar growth in both aerobic and anaerobic conditions (row A). Without added thiamin and in an aerobic environment, RWY15 exhibited the expected minimal growth due to the deficiencies in HMP formation described in Chapter 5 (row B, column 1). Normal growth was seen for the other strains. However, when incubated on the same medium in an anaerobic environment, all three strains showed the same severely restricted growth as seen for aerobic RWY15. This suggests that under such conditions, all three strains are making thiamin independently of Thi5p. This slow growth which, surprisingly, is shared by S288C and BY4705, was not rescued by supplementation of the medium with HET (Figure 6.3, row C). Instead, normal anaerobic growth of all strains was restored by HMP (row D). All the triple mutants (RWY11-RWY14) exhibited identical growth to each other and to BY4705 under the conditions tested (data not shown).

6.3.2 Monitoring growth in aerobic and anaerobic batch cultures.

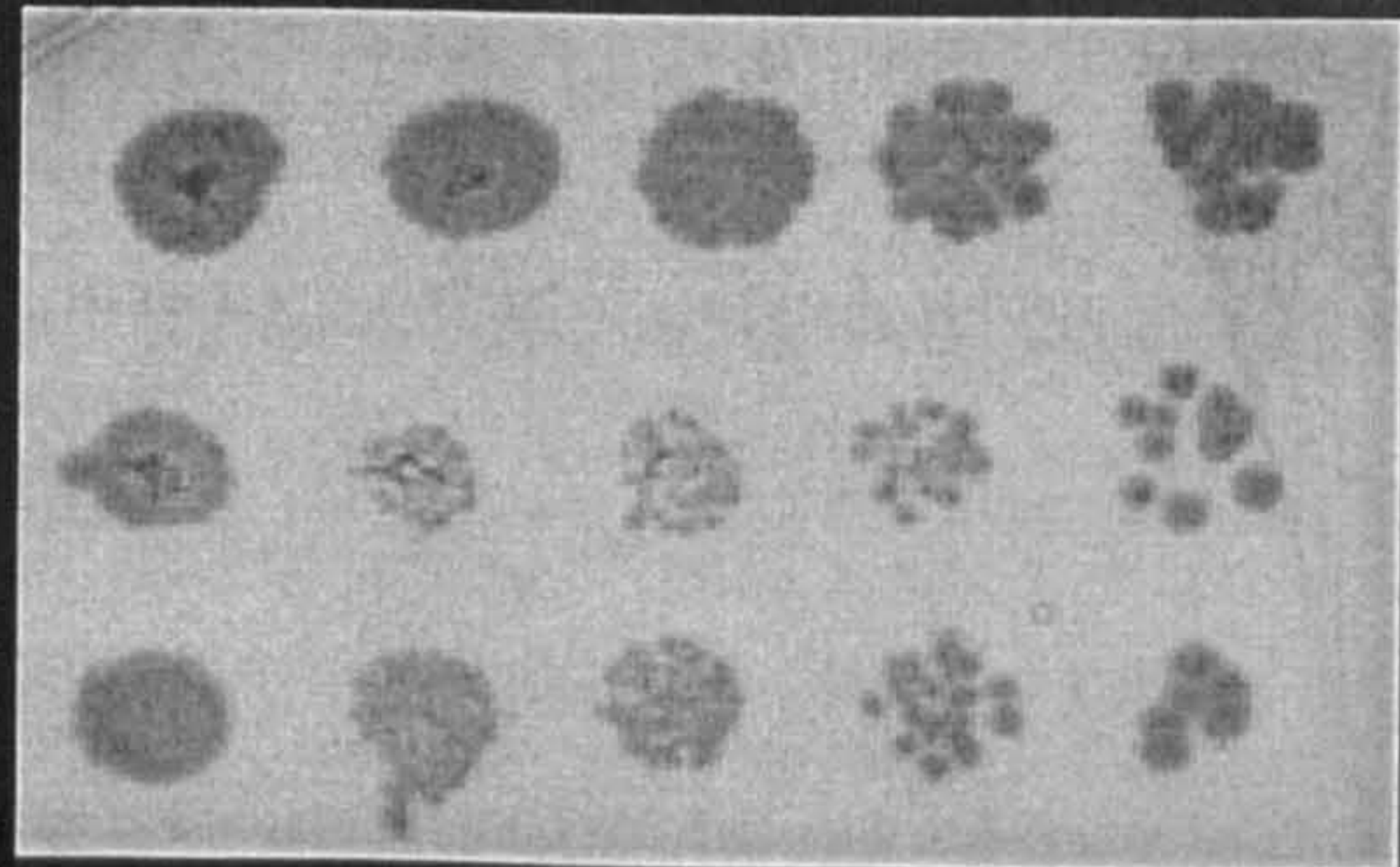
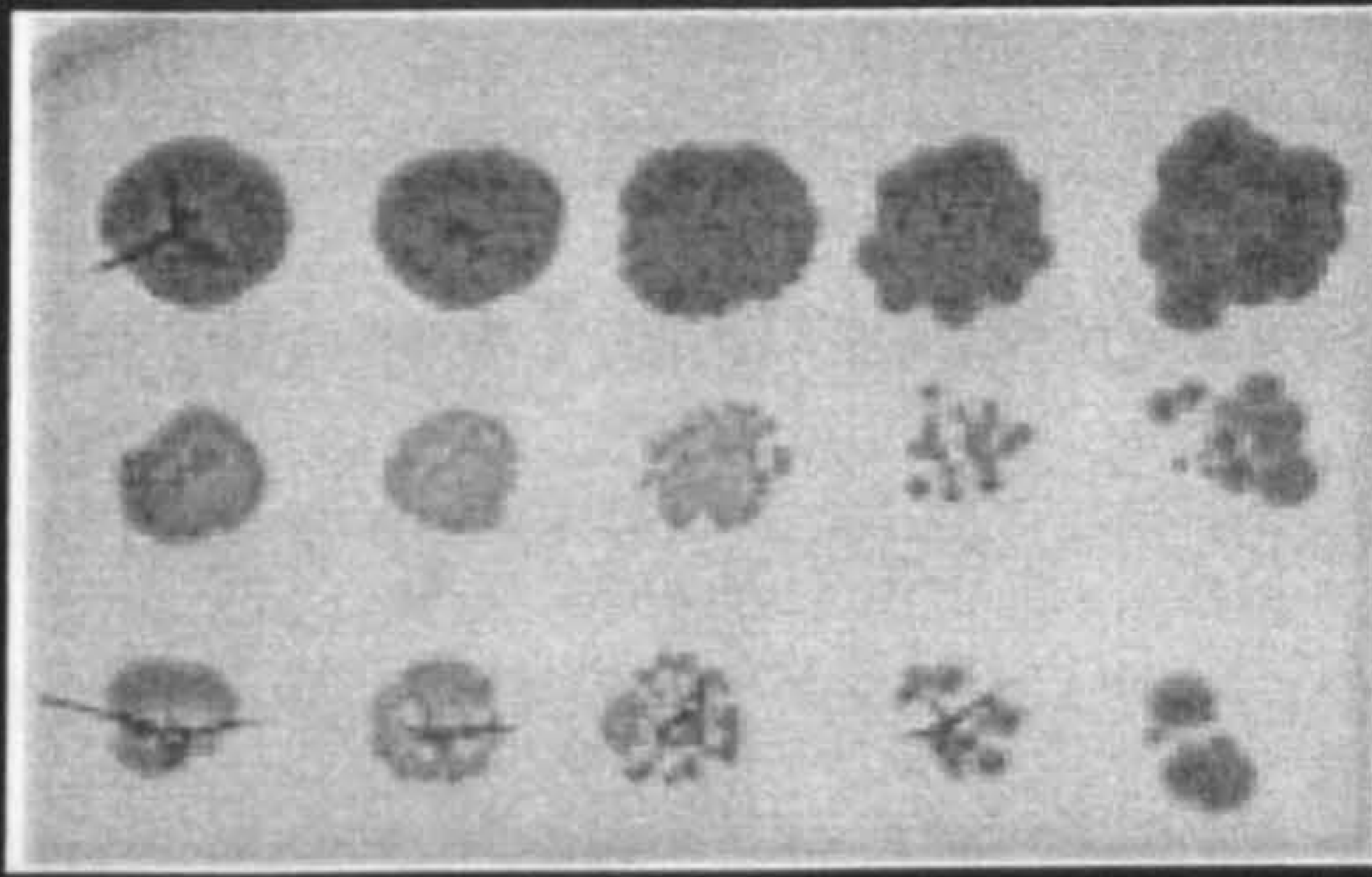
Growth of the wild-type *S. cerevisiae* strain, S288C, along with the RWY15 mutant strain, was further examined in liquid batch cultures under aerobic and anaerobic conditions. These batch cultures were performed in a 1.5 litre LH 502D fermenter which had significant advantages over standard cultivation in shake flasks. Firstly, the fermenter provided much more reproducibly controllable growth conditions (temperature, mixing rate and aeration) and allowed sampling from the closed vessel without risking contamination of the contents. Also, anaerobic conditions could be achieved by passing a constant supply of oxygen-free nitrogen gas through the growth medium. This allowed the maintenance of anaerobiosis even during sampling. Where necessary, complete aerobiosis of the vessel contents was achieved by sparging the culture with compressed air and using a high impeller speed to ensure good mixing (see Materials and Methods).

Figure 6.3 Phenotypic investigation of *S. cerevisiae* strains S288C, BY4705 and RWY15 with regards to the aerobic/anaerobic biosynthesis of thiamin and its precursors. Strains were previously grown overnight in liquid YPD medium. Cells were washed and then diluted to an OD600 of 0.25 and 5 µl aliquots of 1/5 serial dilutions were spotted onto Wickerham's SCM agar glucose medium supplemented as indicated with thiamin, HET or HMP. All agar plates contained oleic acid (in the form of tween 80) and ergosterol which are additional nutritional requirements of *S. cerevisiae* during anaerobiosis. The inoculation of each plate was performed in duplicate with one plate being placed in a closed gas generating chamber which maintains an anaerobic environment (see materials and methods). In addition, two YP agar plates containing the non-fermentable carbon source glycerol (YPGly) were inoculated with S288C. One YPGly plate was incubated in the anaerobic apparatus. The purpose of the YPGly plate was to confirm the maintenance of an anaerobic environment within the apparatus. In the absence of oxygen, S288C will not grow on glycerol. All agar plates were incubated at 28°C for 7 days. S288C showed no growth on the YPGly plate that was placed in the gas generating chamber but showed substantial growth on the YPGly that was placed in an aerobic environment. This indicated that oxygen had been successfully displaced within the chamber.

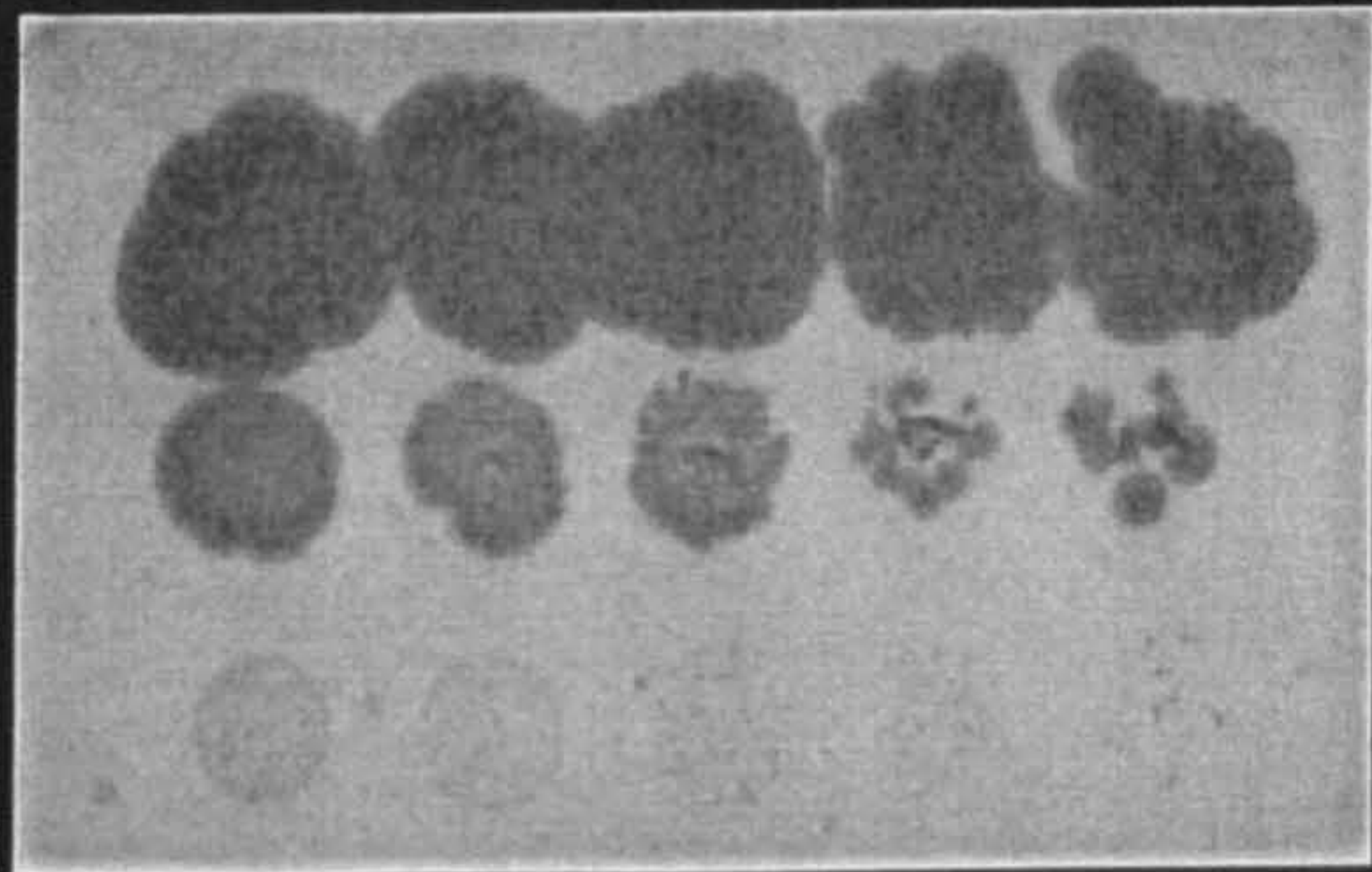
Aerobic

Anaerobic

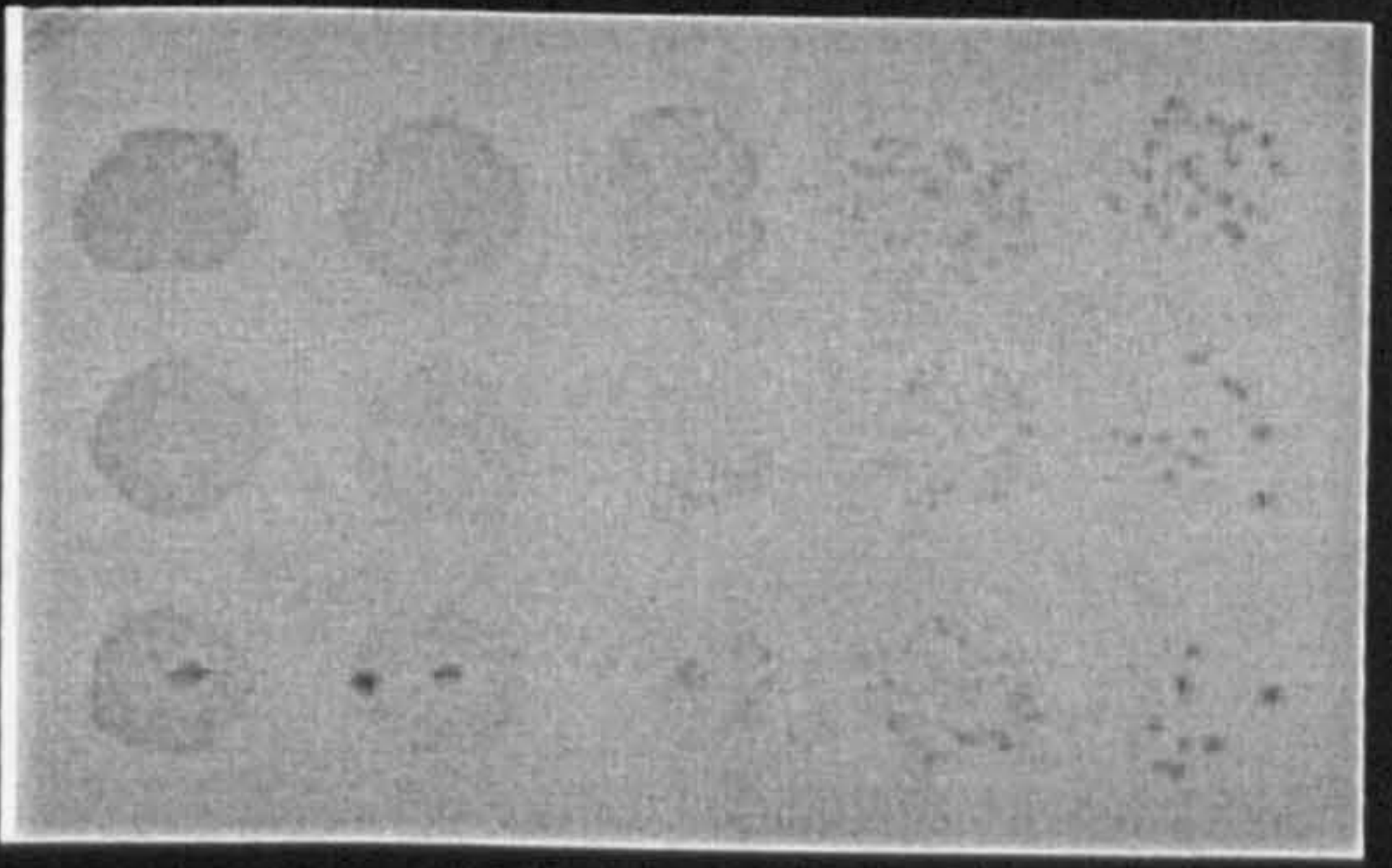
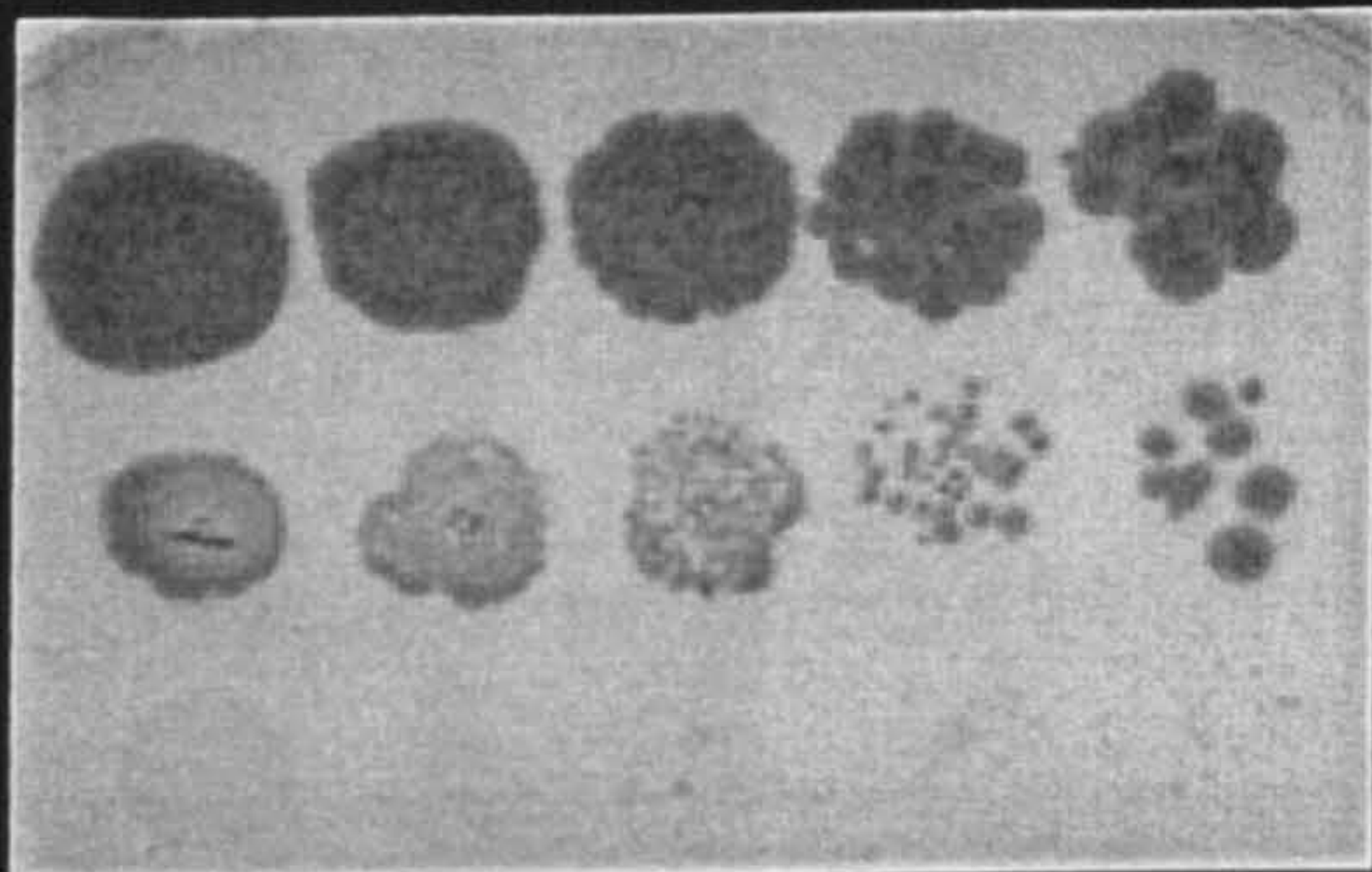
A



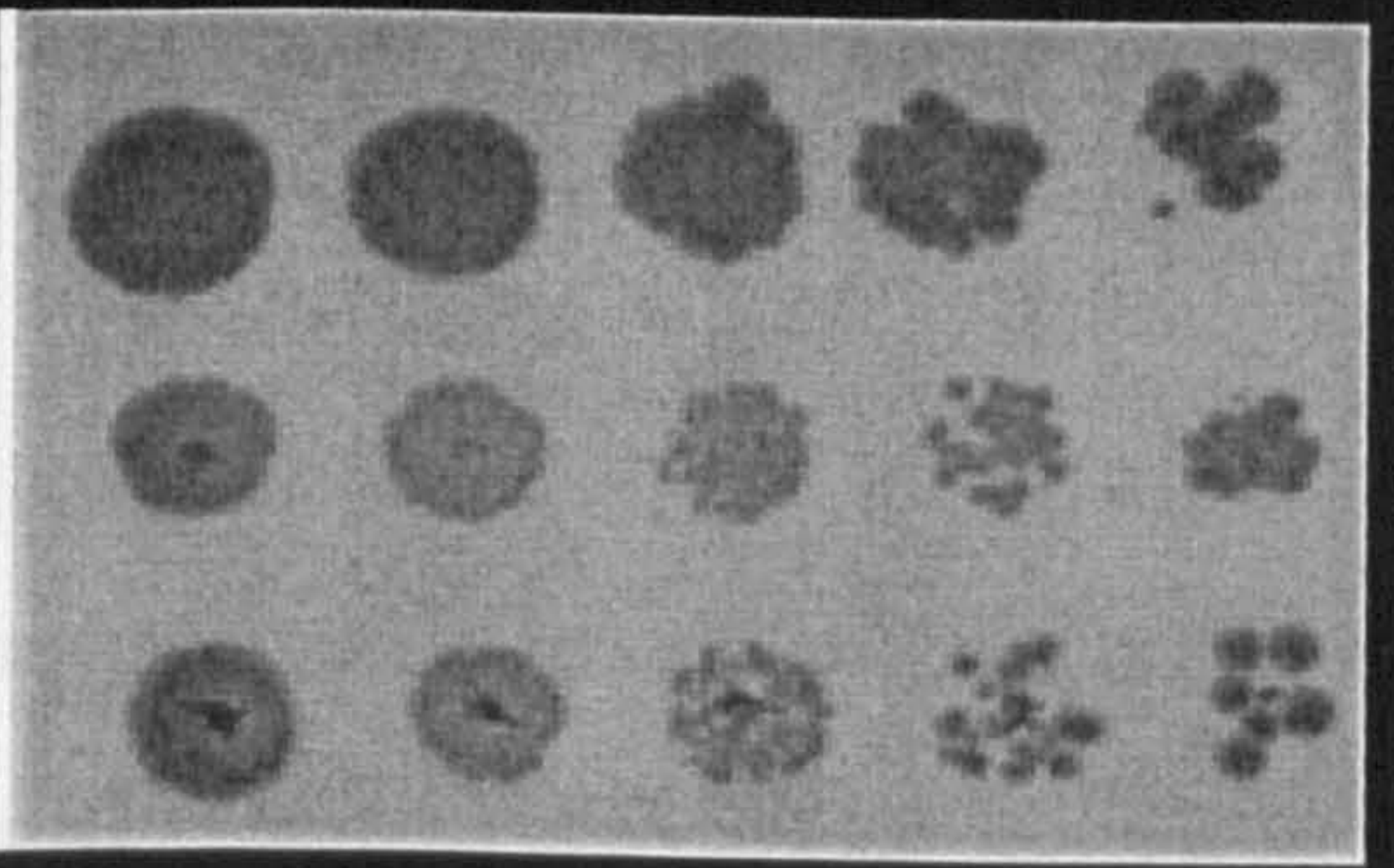
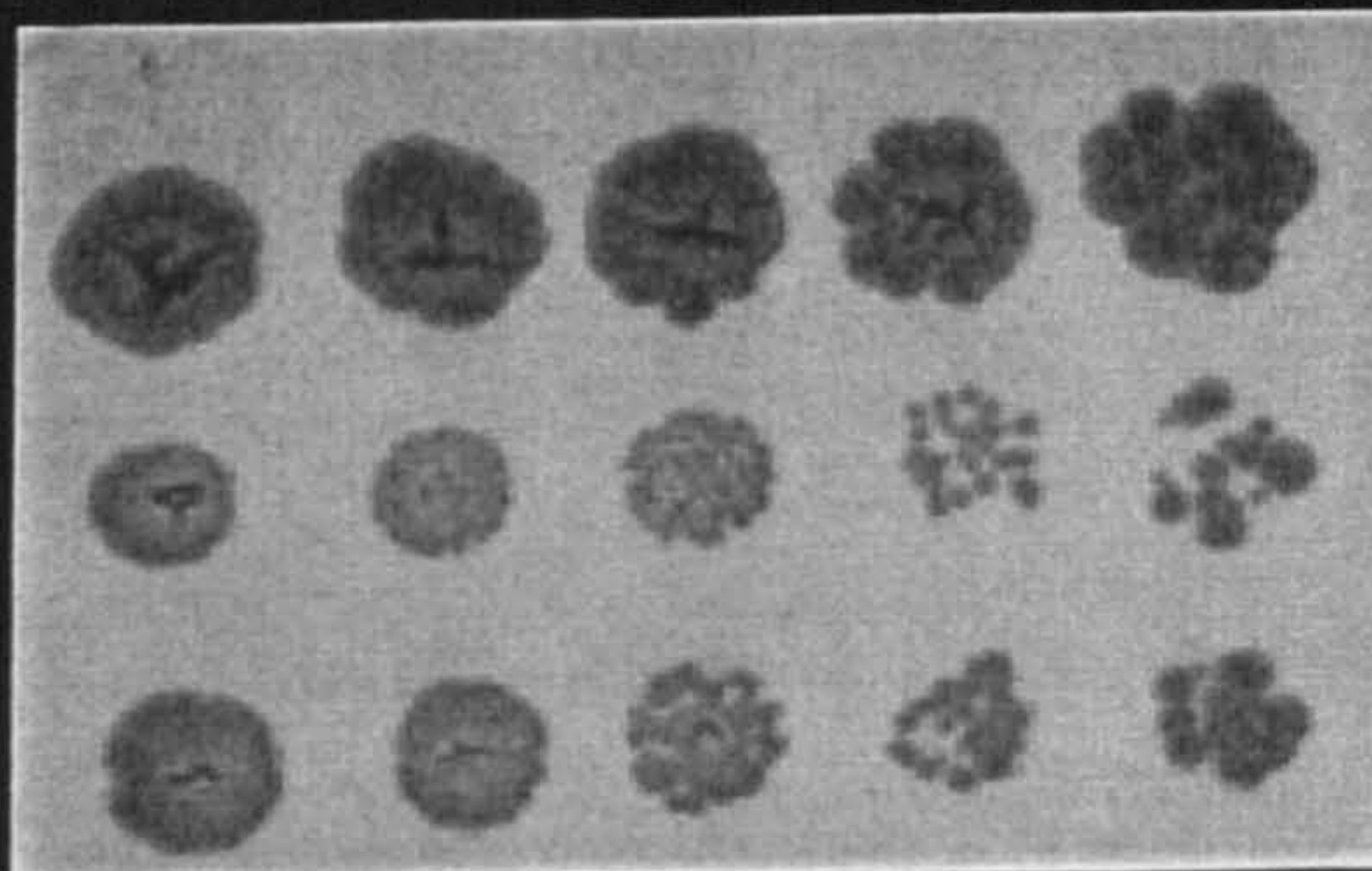
B



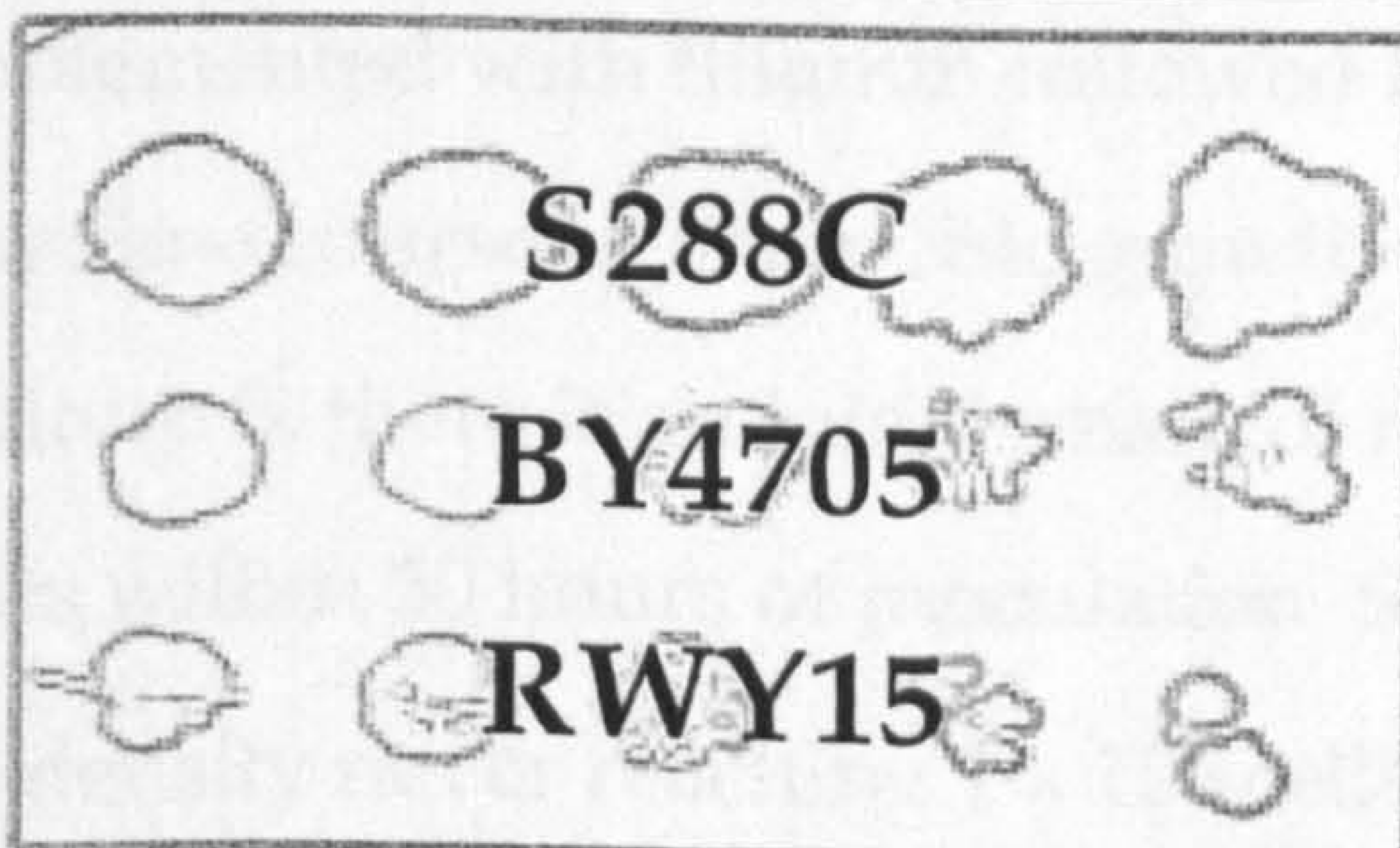
C



D



A= + thi
B= w/o thi
C= + HET
D= + HMP

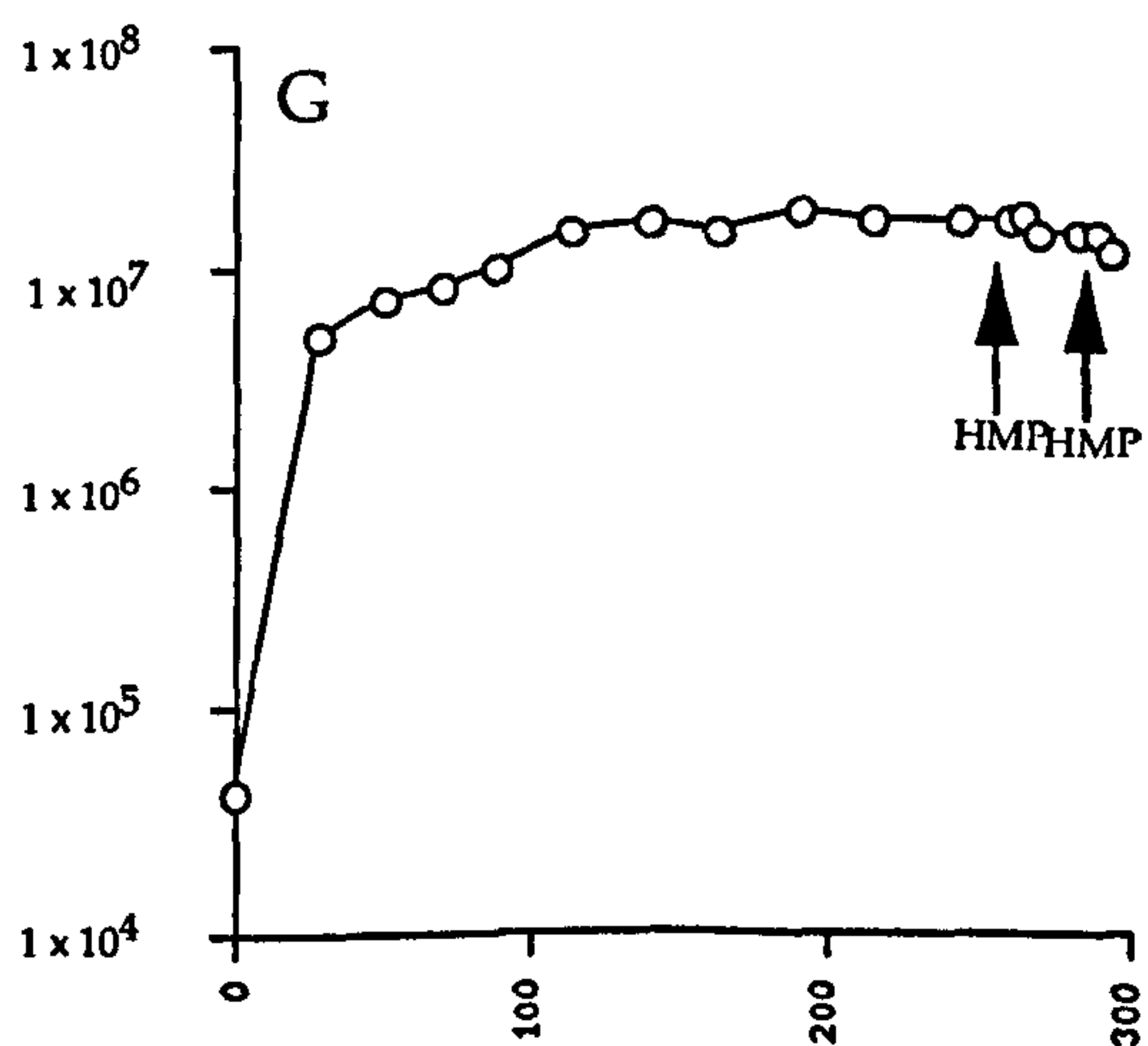
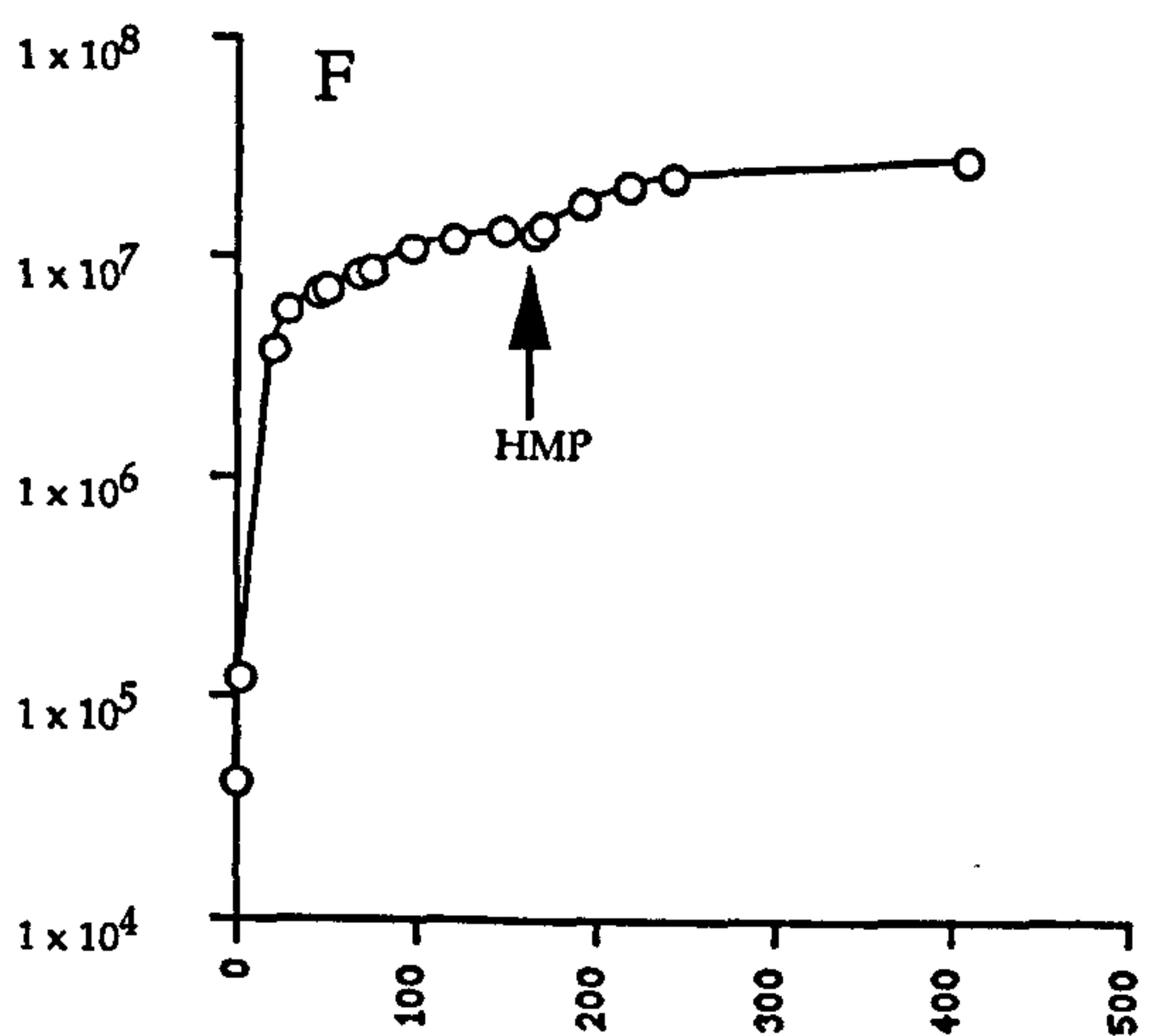
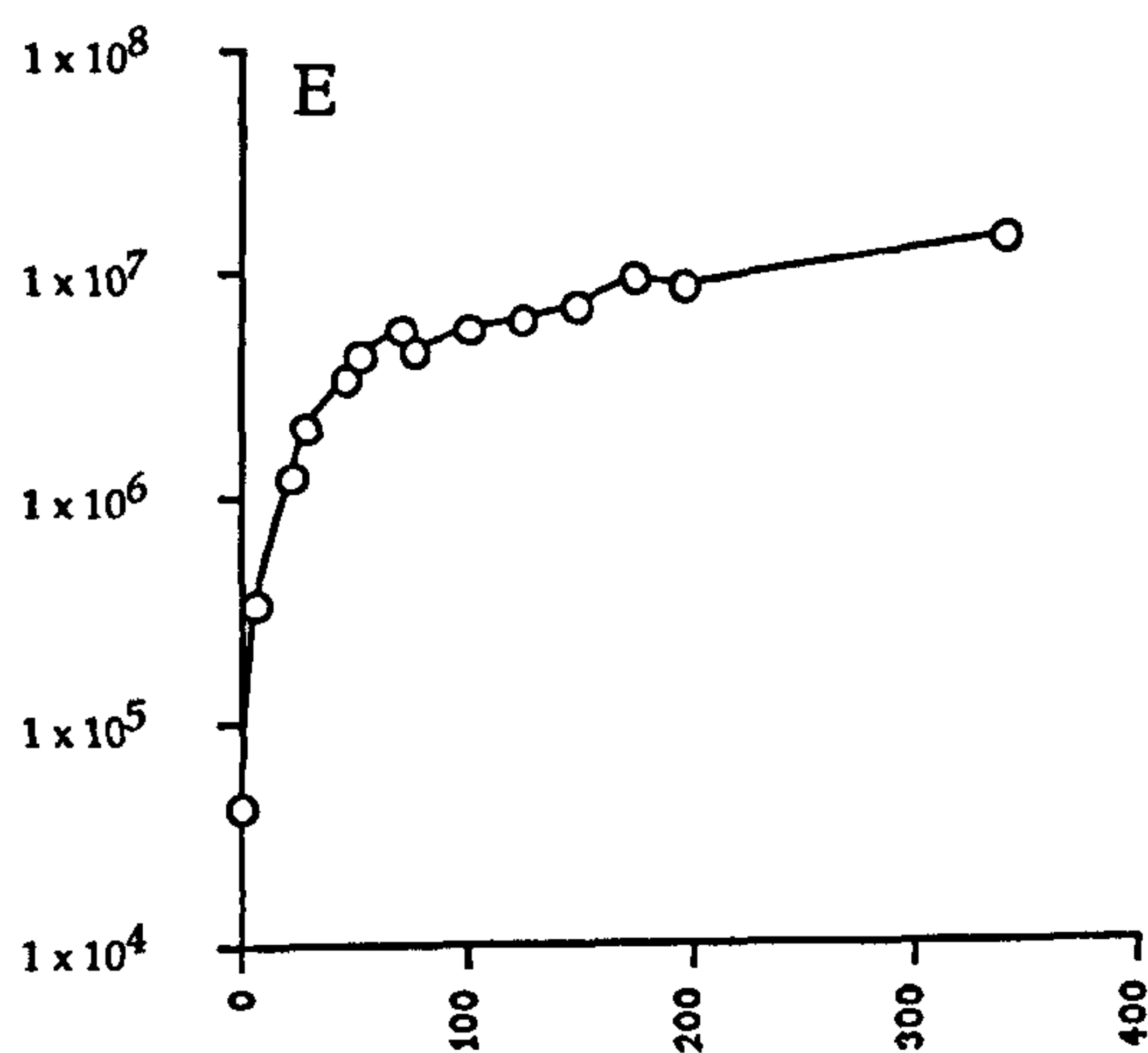
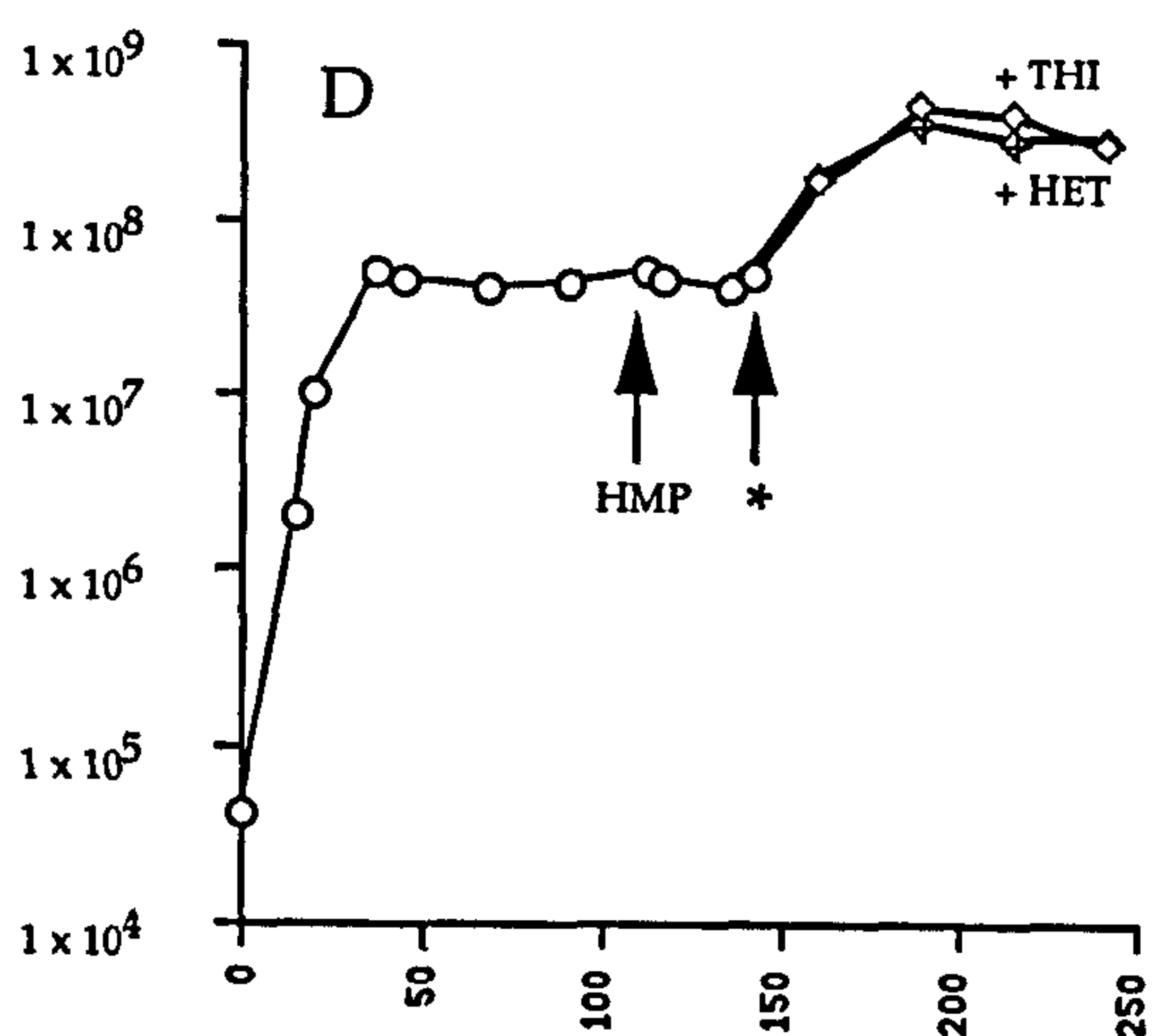
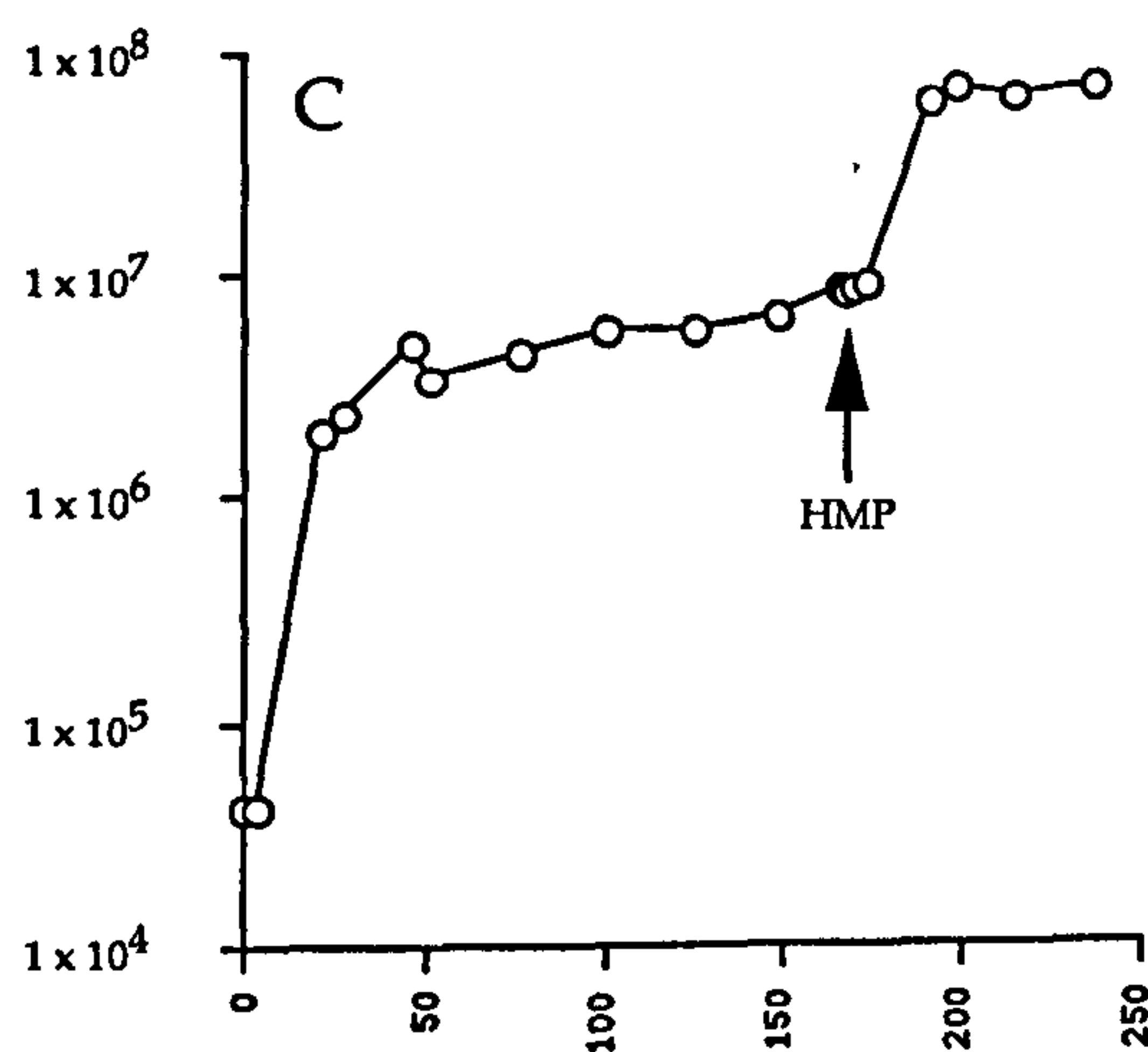
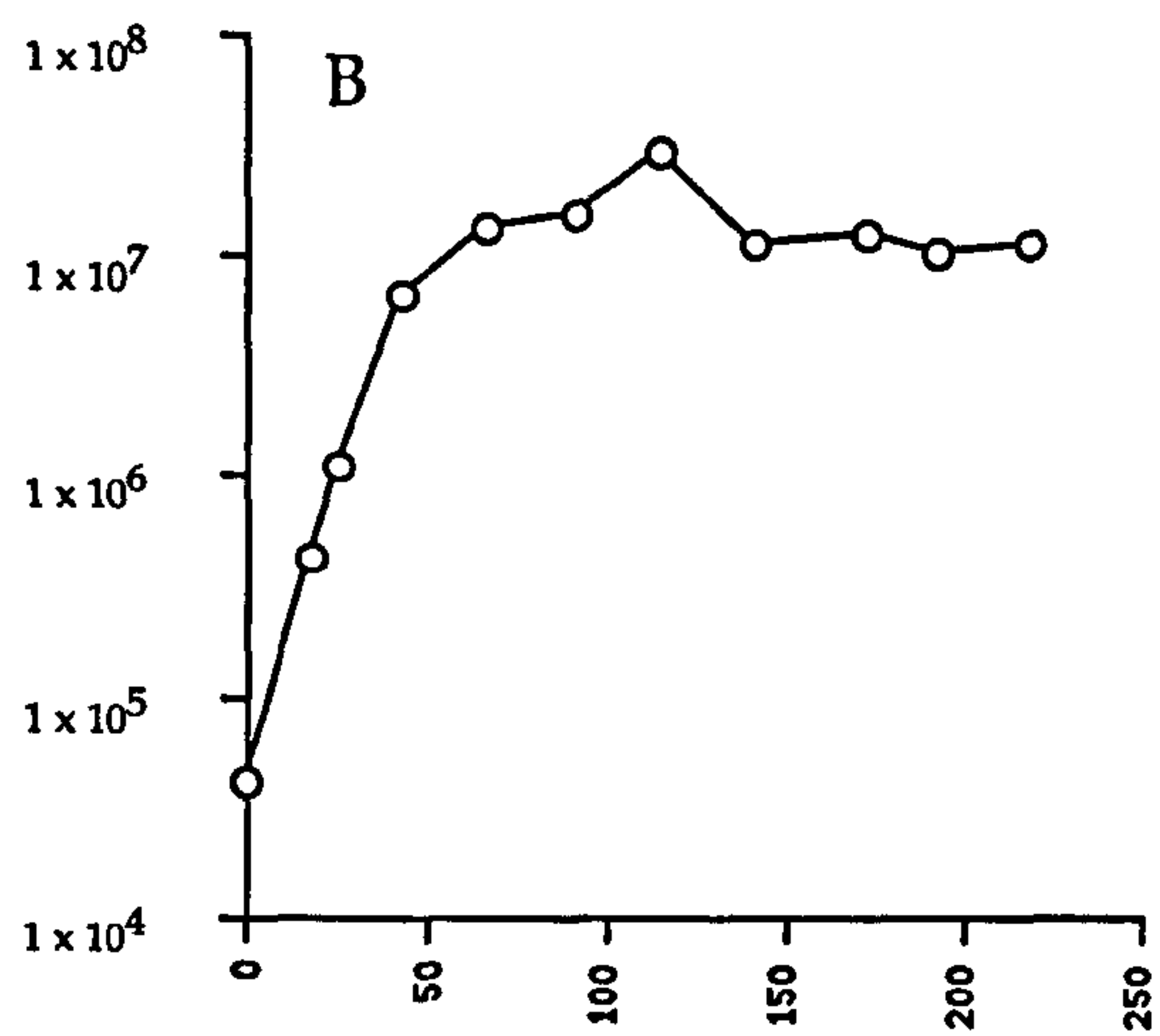
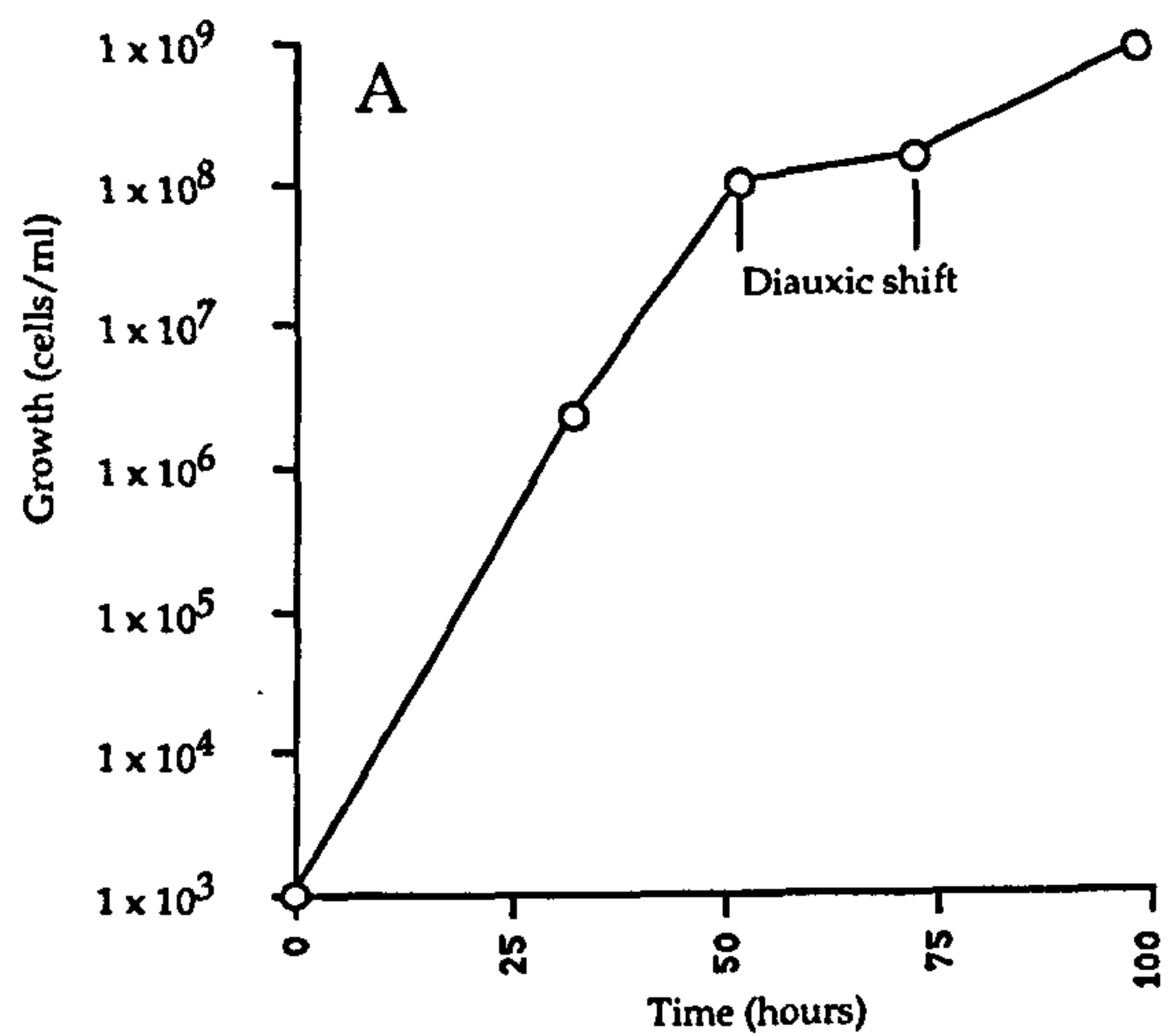


The growth of S288C and RWY15 was monitored during a series of aerobic and anaerobic batch fermentations. The results of these experiments are displayed in graphical format in Figure 6.4. Graphs A, B and C show data on the growth of S288C in thiamin-deficient Wickerham's glucose medium. This medium contained all the other vitamins including pyridoxine and those amino acids and bases that were required by the RWY15 mutant strain. Graph A displays growth of S288C under fully aerobic conditions. Under these conditions, the culture shows typical growth kinetics whereby rapid growth due to glucose metabolism occurs from an inoculating concentration of 1×10^3 cells/ml up to 1×10^8 cells/ml in 50 hours. During this time period, the yeast are producing ethanol as a consequence of the Crabtree effect (Chapter 1). This is followed by a period of negligible growth where the depletion of glucose in the medium causes a change in global gene expression within *S. cerevisiae*. The duration of this response is about 20 hours after which growth resumes due to the metabolism of ethanol. As ethanol is a non-fermentable carbon source, it is expected that under anaerobic conditions, cell density will not increase above 1×10^8 cells/ml.

Graph B displays the growth kinetics of S288C during anaerobiosis. The culture density begins to plateau within 50 hours and fails to reach the concentration of 1×10^8 cells/ml observed in aerobiosis. In fact, growth slows at below 1×10^7 cells/ml and the cell concentration never rises above 3×10^7 cells/ml.

Both batch cultures A and B were inoculated from starter cultures which were also grown in thiamin-deficient Wickerham's medium. To see if the prior accumulation of thiamin has any effect on the anaerobic growth of S288C in the fermenter, a starter culture was grown in medium supplemented with thiamin followed by washing of the cells. Graph C displays the time course of anaerobic growth in thiamin-deficient medium. As seen for culture B, there is an initial phase of rapid growth followed by a plateau which occurs within 50 hours of inoculation. Severely restricted growth continued, with the cell density never reaching 1×10^7 cells/ml. At 166 hours, HMP was added to the fermentation. This was followed by a sharp

Figure 6.4 Liquid batch fermentations of strains S288C (cultures A, B and C) and RWY15 (D, E, F, G) in aerobic (A, D) or anaerobic (B, C, F, G) conditions. Growth of the yeast strains was monitored in a LH 502D fermenter containing 1.5 litres of Wickerham's glucose medium without thiamin and containing adenine, lysine, methionine, antifoam A, tween 80 and ergosterol. The time course of growth (represented as cells/ml) is displayed in graphical format. Prior to inoculation, yeast was grown aerobically for 24 hours in the Wickerham's medium. For cultures C, D, E, F and G, thiamin was present in the starter culture at a final concentration of 2 μ M. Cells were harvested, washed and inoculated into the fermenter to a density of 4×10^4 cells/ml (1×10^3 cells/ml for culture A). Aerobic conditions (cultures A and D) or anaerobic conditions (B, C, E, F, G) were maintained as described in Materials and Methods. Where indicated, HMP was added at 1.4 mg/l (final concentration, 10 μ M) except for culture G where there were two additions of HMP at 2 mg/l. The asterisk in culture D signifies the point where the fermenter culture was split into two independent cultures in shake flasks with the addition of thiamin (final concentration, 2 μ M) into one and the addition of HET (2 μ M) into the other.



stimulation of the growth rate with the maximal cell concentration reaching 6.5×10^7 cells/ml within 30 hours of HMP addition. Taken together, the S288C growth data suggests that the formation of HMP is limiting during anaerobiosis.

Graphs D, E, F and G (Figure 6.4) show the growth data of the RWY15 mutant strain which lacks all four members of the *THI5* gene family. Under anaerobic conditions (E, F, G), growth of the RWY15 culture began to plateau within 50 hours of inoculation after which a very small amount of growth was seen to occur over a long length of time. A final cell density of not more than 3×10^7 cells/ml was reached after 400 hours. These growth curves followed an identical trend to those seen for the anaerobic growth of a S288C culture, as described above. Based on the observations for S288C, the addition of HMP to the fermentation would be expected to be accompanied by a rapid increase in cell density. However, the addition of this precursor had no effect on the growth rate (culture F), even when added in copious quantities (G).

A fully aerobic RWY15 culture (D) was seen to exhibit similar growth kinetics to those of an anaerobic culture except for an appreciably higher maximal cell concentration (4×10^7 cells/ml) at 45 hours. As with the anaerobic cultures, the addition of HMP had no effect upon the growth plateau but the subsequent addition of either HET or thiamin was seen to cause a rapid increase in cell density. It would be interesting to see if the addition of HET alone or thiamin has an identical effect under anaerobic conditions.

6.4 Regulation of the *THI5* gene family.

6.4.1 Northern blot hybridisation of the triple mutants.

It is already known that for cultures of S288C grown on Wickerham's glucose medium with added thiamin, northern blot analysis failed to detect a positive signal for *THI5* mRNA (Hather, 1996). However, mRNA transcripts of 1.4 Kb are readily

detected from cultures grown without the addition of thiamin. This regulation by thiamin has also been seen for the *S. pombe* homologue *nmt1* (Chapter 1). As all four mRNAs (*THI5*, *THI11*, *THI12* and *THI13*) cross-hybridise with the *THI5* probe on a northern blot (Burrows, 1997), a positive signal represents the total pool of mRNA transcribed from one or more genes. It is almost certain that each member is expressed in response to thiamin depletion as all of them have been found to be functionally redundant (Chapter 5).

Northern blot hybridisation analysis has been carried out for each of the triple mutants (RWY11-RWY14) to confirm that each member is expressed. The relative quantification of the transcripts was also attempted. Total RNA was isolated from each mutant, and the BY4705 parent, grown to mid-exponential phase on Wickerham's medium with and without thiamin. Equal quantities of RNA were electrophoresed in duplicate followed by hybridisation of the blot with either the *THI5* or actin ³²P-labelled probe. Figure 6.5 shows detection of transcripts for every triple mutant grown on minimal medium without thiamin. As each of these mutants contained a different intact *THI5* family member, all four genes must be expressed in response to thiamin depletion. As expected, RNAs from the same strains grown in thiamin supplemented medium gave no visible signals on the autoradiograph (Figure 6.5a). Quantification of the mRNA from cells grown in thiamin-depleted medium was achieved using a Phosphorimager™ (see Materials and Methods) which scans a screen that has been previously exposed to the northern blot and provides information on band intensities according to the levels of emitted radiation. The intensity units of the bands in Figure 6.5b are displayed in the legend. The data have been corrected through use of the actin signal to take account of any unequal loading of the RNA gel. From the data, it can be clearly seen that none of the mutants produced the amount of *THI5/11/12/13* mRNA that was present in BY4705. Also, by comparing the data from the triple mutants, it was seen that each of the four genes contribute very similar levels of mRNA to the total pool.

Figure 6.5 Northern blot hybridisation analysis of the triple mutants. Total RNA was isolated from aerobic batch cultures of strains BY4705 and RWY11-14. Strains were grown in Wickerham’s glucose medium supplemented with or without thiamin to a density of $1\text{-}2\times10^7$ cells/ml.

- a.) Northern blot of BY4705 and the triple mutants RWY11 (possessing just *THI13*), RWY12 (*THI11*), RWY13 (*THI5*) and RWY14 (*THI12*). Also shown is the quadruple mutant RWY15 grown in medium + thiamin. Probe=pRH4 (see materials and methods)
- b.) Image output from Imagequant™ Phosphorimager software. Probe=pRH4. Each lane was loaded with total RNA from cultures grown without added thiamin.
- The intensity units of each lane are shown in the table below. Corrected units are based on actin mRNA standardisation from a duplicate blot. Expression ratio is the ratio of intensities to that of RWY12 (lane C). This experiment was repeated and is represented along with this data under ‘average expression ratio’.

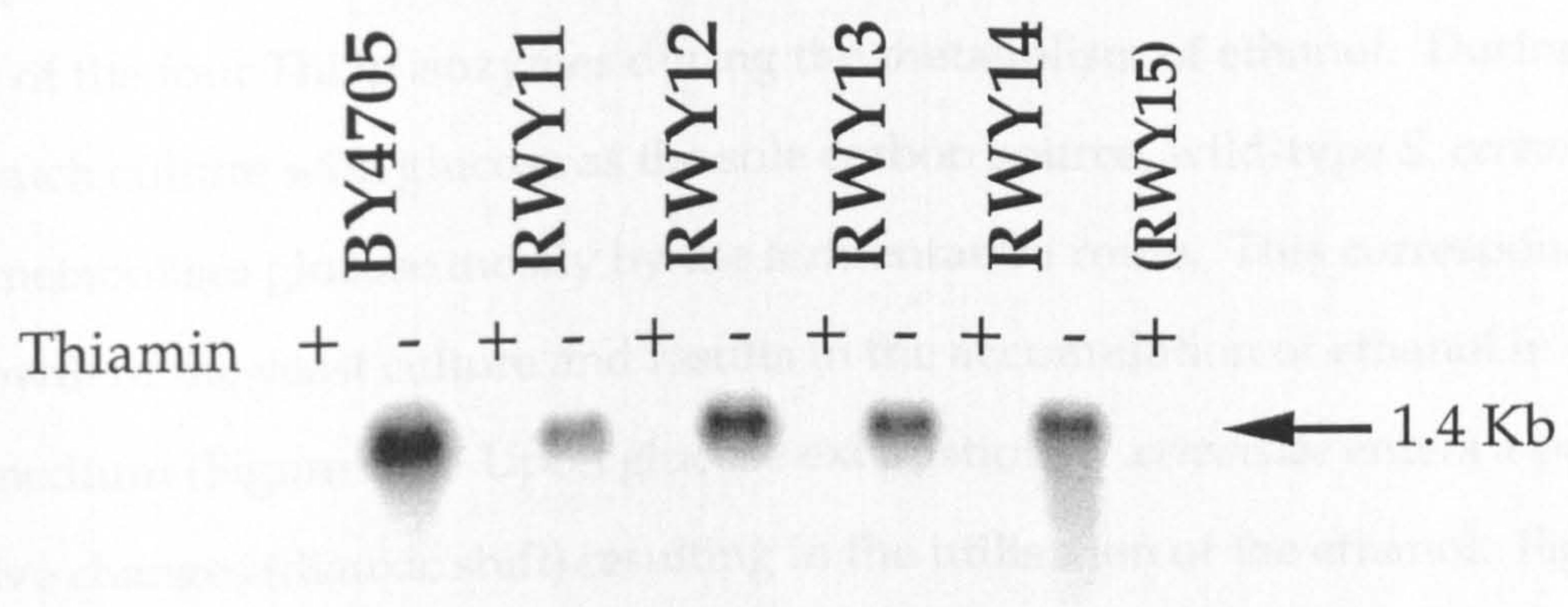
Lane	Intensity units (THI5 probe)	Intensity units (Actin)	Corrected intensity (THI5 probe)	Expression ratio	Average exp. ratio
A BY4705	209008	5089	339011	3.9	3.1
B RWY11	100934	7472	111532	1.3	1.3*
C RWY12	86991	8253	86991	1.0	1.0
D RWY13	112256	6463	143351	1.6	1.2
E RWY14	85170	7494	93772	1.1	0.9

* No further data were obtained for lane B and so average ratio is from a single experiment.

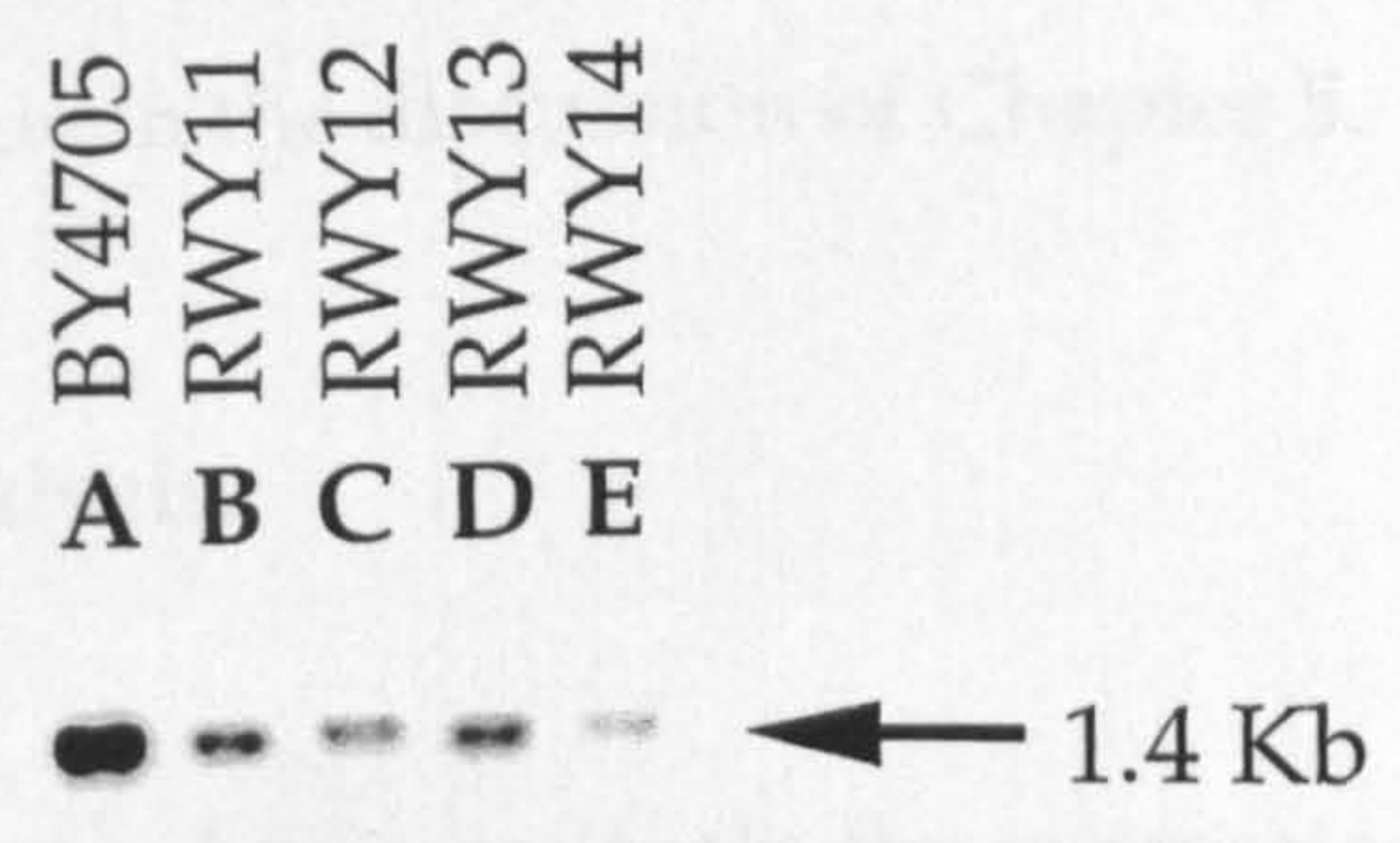
5.4.2 Expression of THF5 in S. cerevisiae during glucose batch culture

The apparent growth morphology of the RWY15 mutant strain has been shown to be carbon source dependent (Chapter 5). With ethanol as the carbon source, normal growth on this carbon source is observed for RWY15 (Chapter 5, Figure 5.6).

This suggests that H^+ formation for ethanol biosynthesis is not dependent on the



b.)



Slot blot RNA hybridisation was used to investigate the expression of each of the THF5 genes in response to exogenous thiamin and its precursors. Total RNA was isolated from S288C, BY4705 and each of the RWY mutants (RWY11-RWY14) after growth to late exponential phase (2.5×10^8 cells/ml). The culture medium was replaced with 1% ethanol medium, 1% thiamine, 1% 5-Hydroxyethyl methylthioethylamine or 1% 5-Hydroxyethyl methylthioethylamine. Samples were taken at 0, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432, 67108864, 134217728, 268435456, 536870912, 1073741824, 2147483648, 4294967296, 8589934592, 17179869184, 34359738368, 68719476736, 137438953472, 274877906944, 549755813888, 1099511627776, 2199023255552, 4398046511104, 8796093022208, 17592186044416, 35184372088832, 70368744177664, 140737488355328, 281474976710656, 562949953421312, 1125899906842624, 2251799813685248, 4503599627370496, 9007199254740992, 18014398509481984, 36028797018963968, 72057594037927936, 144115188075855872, 288230376151711744, 576460752303423488, 1152921504606846976, 2305843009213693952, 4611686018427387904, 9223372036854775808, 18446744073709551616, 36893488147419103232, 73786976294838206464, 147573952589676412928, 295147905179352825856, 590295810358705651712, 1180591620717411303424, 2361183241434822606848, 4722366482869645213696, 9444732965739290427392, 18889465931478580854784, 37778931862957161709568, 75557863725914323419136, 151115727451828646838272, 302231454903657293676544, 604462909807314587353088, 1208925819614629174706176, 2417851639229258349412352, 4835703278458516698824704, 9671406556917033397649408, 19342813113834066795298816, 38685626227668133590597632, 77371252455336267181195264, 154742504910672534362390528, 309485009821345068724781056, 618970019642690137449562112, 1237940039285380274899124224, 2475880078570760549798248448, 4951760157141521099596496896, 9903520314283042199192993792, 19807040628566084398385987584, 39614081257132168796771975168, 79228162514264337593543950336, 158456325028528675187087900672, 316912650057057350374175801344, 633825300114114700748351602688, 1267650600228229401496703205376, 2535301200456458802993406410752, 5070602400912917605986812821504, 10141204801825835211973625643008, 20282409603651670423947251286016, 40564819207303340847894502572032, 81129638414606681695789005144064, 162259276829213363391578010288128, 324518553658426726783156020576256, 649037107316853453566312041152512, 1298074214633706907132624082305024, 2596148429267413814265248164610048, 5192296858534827628530496329220096, 10384593717069655257060992658440192, 20769187434139310514121985316880384, 41538374868278621028243970633760768, 83076749736557242056487941267521536, 166153499473114484112975882535043072, 332306998946228968225951765070086144, 664613997892457936451903530140172288, 1329227995784915872903807060280344576, 2658455991569831745807614120560689152, 5316911983139663491615228241121378304, 10633823966279326983230456482242756608, 21267647932558653966460912964485513216, 42535295865117307932921825928971026432, 85070591730234615865843651857942052864, 170141183460469231731687303715884105728, 340282366920938463463374607431768211456, 680564733841876926926749214863536422912, 1361129467683753853853498429727072845824, 2722258935367507707706996859454145691648, 5444517870735015415413993718908291383296, 10889035741470030830827987437816582766592, 21778071482940061661655974875633165533184, 43556142965880123323311949751266331066368, 87112285931760246646623899502532662132736, 174224571863520493293247799005065324265472, 348449143727040986586495598010130648530944, 696898287454081973172991196020261297061888, 1393796574908163946345982392040522594123776, 2787593149816327892691964784081045188247552, 5575186299632655785383929568162090376495104, 11150372599265311570767859136324180752990208, 22300745198530623141535718272648361505980416, 44601490397061246283071436545296723011960832, 89202980794122492566142873090593446023921664, 178405961588244985132285746181186892047843328, 356811923176489970264571492362373784095686656, 713623846352979940529142984724747568191373312, 1427247692705959881058285969449495136382746624, 2854495385411919762116571938898990272765493248, 5708990770823839524233143877797980545530986496, 11417981541647679048466287755595961091061972992, 22835963083295358096932575511191922182123945984, 45671926166590716193865151022383844364247891968, 91343852333181432387730302044767688728495783936, 182687704666362864775460604089535377456991567872, 365375409332725729550921208179070754913983135744, 730750818665451459101842416358141509827966271488, 1461501637330902918203684832716283019655932542976, 2923003274661805836407369665432566039311865085952, 5846006549323611672814739330865132078623730171904, 11692013098647223345629478661730264157247460343808, 23384026197294446691258957323460528314494920687616, 46768052394588893382517914646921056628989841375232, 93536104789177786765035829293842113257979682750464, 187072209578355573530071658587684226515959365500928, 374144419156711147060143317175368453031918731001856, 748288838313422294120286634350736906063837462003712, 1496577676626844588240573268701473812127674924007424, 2993155353253689176481146537402947624255349848014848, 5986310706507378352962293074805895248510699696029696, 11972621413014756705924586149611790497021399392059392, 23945242826029513411849172299223580994042798784118784, 47890485652059026823698344598447161988085597568237568, 95780971304118053647396689196894323976171195136475136, 191561942608236107294793378393788647952342390272950272, 383123885216472214589586756787577295904684780545900544, 766247770432944429179173513575154591809369561091801088, 1532495540865888858358347027150309183618739122183602176, 3064991081731777716716694054300618367237478244367204352, 6129982163463555433433388108601236734474956488734408704, 12259964326927110866866776217202473468949912977468817408, 24519928653854221733733552434404946937899825954937634816, 49039857307708443467467104868809893875799651909875269632, 98079714615416886934934209737619787751599303819750539264, 196159429230833773869868419475239575503198607639501078528, 392318858461667547739736838950479151006397215279002157056, 784637716923335095479473677900958302012794430558004314112, 1569275433846670190958947355801916604025588861116008628224, 3138550867693340381917894711603833208051177722232017256448, 6277101735386680763835789423207666416102355444464034512896, 12554203470773361527671578846415332832204710888928069025792, 25108406941546723055343157692830665664409421777856138051584, 50216813883093446110686315385661331328818843555712276103168, 100433627766186892221372630771322662657637687111424552206336, 200867255532373784442745261542645325315275374222849104412672, 401734511064747568885490523085290650630550748445698208825344, 803469022129495137770981046170581301261101496891396417650688, 1606938044258990275541962092341162602522202993782792835301376, 3213876088517980551083924184682325205044405987565585670602752, 6427752177035961102167848369364650410088811975131171341205504, 12855504354071922204335696738729300820177623950262342682411008, 25711008708143844408671393477458601640355247900524685364822016, 51422017416287688817342786954917203280710495801049370729644032, 102844034832575377634685573909834406561420991602098741459288064, 205688069665150755269371147819668813122841983204197482918576128, 411376139330301510538742295639337626245683966408394965837152256, 822752278660603021077484591278675252491367932816789931674304512, 1645504557321206042154969182557350504982735865633579863348609024, 3291009114642412084309938365114701009965471731267159726697218048, 6582018229284824168619876730229402019930943462534319453394436096, 13164036458569648337239753460458804039861886925068638906788872192, 26328072917139296674479506920917608079723773850137277813577744384, 52656145834278593348959013841835216159447547700274555627155488768, 105312291668557186697918027683670432318895095400549111254310977536, 210624583337114373395836055367340864637790190801098222508621955072, 421249166674228746791672110734681729275580381602196445017243910144, 842498333348457493583344221469363458551160763204392890034487820288, 1684996666696914987166688442938726917102321526408785780068975640576, 3369993333393829974333376885877453834204643052817571560137951281152, 6739986666787659948666753771754907668409286105635143120275902562304, 13479973333575319897333507543509815336818572211270286240551805124608, 26959946667150639794667015087019630673637144422540572481103610249216, 53919893334301279589334030174039261347274288845081144962207220498432, 107839786668602559178668060348078522694548577690162289924414440996864, 215679573337205118357336120696157045389097155380324579848828881993728, 431359146674410236714672241392314090778194310760649159697657763987456, 862718293348820473429344482784628181556388621521298319395315527974912, 1725436586697640946858688965569256363112777243042596638790631055949824, 3450873173395281893717377931138512726225554486085193277581262111899648, 6901746346790563787434755862277025452451108972170386555162524223799296, 13803492693581127574869511724554050904902217944340773110325048447598592, 27606985387162255149739023449108101809804435888681546220650096895197184, 55213970774324510299478046898216203619608871777363092441300193790394368, 110427941548649020598956093796432407239217743554726184882600387580788736, 220855883097298041197912187592864814478435487109452369765200775161577472, 441711766194596082395824375185729628956870974218904739530401550323154944, 883423532389192164791648750371459257913741948437809479060803100646309888, 1766847064778384329583297500742918515827483896875618958121606201292619776, 3533694129556768659166595001485837031654967793751237916243212402585239552, 7067388259113537318333190002971674063309935587502475832486424805170479104, 14134776518227074636666380005943348126619871175004951664972849610340958208, 28269553036454149273332760011886696253239742350009903329945699220681916416, 56539106072908298546665520023773392506479484700019806659891398441363832832, 113078212145816597093331040047546785012958969400039613319782796882727665664, 226156424291633194186662080095093570025917938800079226639565593765455331328, 452312848583266388373324160190187140051835877600158453279131187530910662656, 904625697166532776746648320380374280103671755200316906558262375061821325312, 1809251394333065553493296640760748560207343510400633813116524750123642650624, 3618502788666131106986593281521497120414687020801267626233049500247285301248, 7237005577332262213973186563042994240829374041602535252466099000494570602496, 14474011154664524427946373126085988481658748083205070504932198000989141204992, 28948022309329048855892746252171976963317496166410141009864396001978282409984, 57896044618658097711785492504343953926634992332820282019728792003956564819968, 115792089237316195423570985008687907853269984665640564039457584007913129639936, 231584178474632390847141970017375815706539969331281128078915168015826259279872, 463168356949264781694283940034751631413079938662562256157830336031652518559744, 926336713898529563388567880069503262826159877325124512315660672063305037119488, 1852673427797059126777135760139006525652319754650249024631321344126610074238976, 3705346855594118253554271520278013051304639509300498049262642688253220148477952, 7410693711188236507108543040556026102609279018600996098525285376506440296955904, 14821387422376473014217086081112052205218558037201992197050570753012880593911808, 29642774844752946028434172162224104410437116074403984394101141506025761187823616, 59285549689505892056868344324448208820874232148807968788202283012051522375647232, 118571099379011784113736688648896417641748464297615937576404566024103044751294464, 237142198758023568227473377297792835283496928595231875152809132048206089502588928, 4742843975160471364549467545955

6.4.2 Expression of *THI5* in *S. cerevisiae* during glucose batch culture.

The apparent thiamin auxotrophy of the RWY15 mutant strain has been shown to be carbon source dependent (Chapter 5). With ethanol as sole carbon source, normal growth on thiamin-deficient medium is observed for RWY15 (Chapter 5, Figure 5.6). This suggests that HMP formation for thiamin biosynthesis is not dependent on the existence of the four Thi5p isozymes during the metabolism of ethanol. During an aerobic batch culture with glucose as the sole carbon source, wild-type *S. cerevisiae* initially metabolises glucose mostly by the fermentative route. This corresponds to rapid growth of the yeast culture and results in the accumulation of ethanol in the growth medium (Figure 6.6). Upon glucose exhaustion, *S. cerevisiae* enters a period of adaptive changes (diauxic shift) resulting in the utilisation of the ethanol. Figure 6.6 shows the expression profile of the *THI5* gene family during an aerobic batch culture of S288C in thiamin-deficient glucose medium. Total RNA was isolated from samples of the culture during the initial period of exponential growth and at the entry and exit of the diauxic shift. Northern blot analysis of the electrophoresed RNA using the *THI5* probe clearly shows that the mRNA is abundant until the diauxic shift, with the complete absence of any signal at the point of diauxic exit. These data fit the hypothesis made in the discussion of Chapter 5.

6.4.3 Slot blot hybridisation analysis.

Slot blot RNA hybridisation was used to investigate the expression of each of the *THI5* genes in response to exogenous thiamin and its precursors. Total RNA was isolated from S288C, BY4705 and each of the triple mutants (RWY11-RWY14) after growth to late exponential phase ($2-5 \times 10^7$ cells/ml). The culture medium was Wickerham's glucose supplemented with or without thiamin, HET, HMP or both precursors together. Equal amounts of each of the thirty RNA samples was applied to two membranes, one was subsequently incubated with the *THI5* probe and the other with a probe to detect actin mRNA (*ACT1*). The intensities of each of the slots hybridised with *THI5* was measured using the Phosphorimager™ and data were

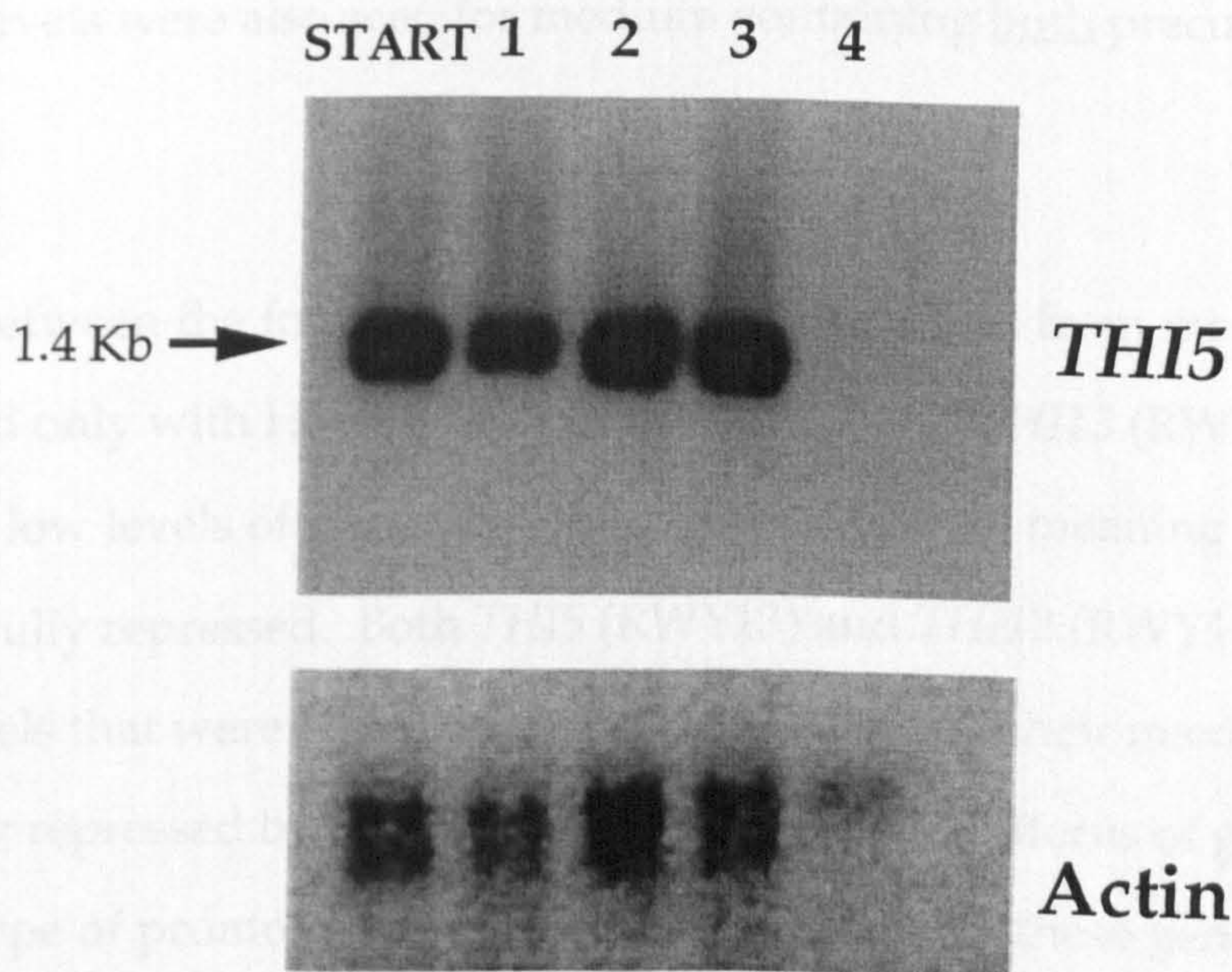
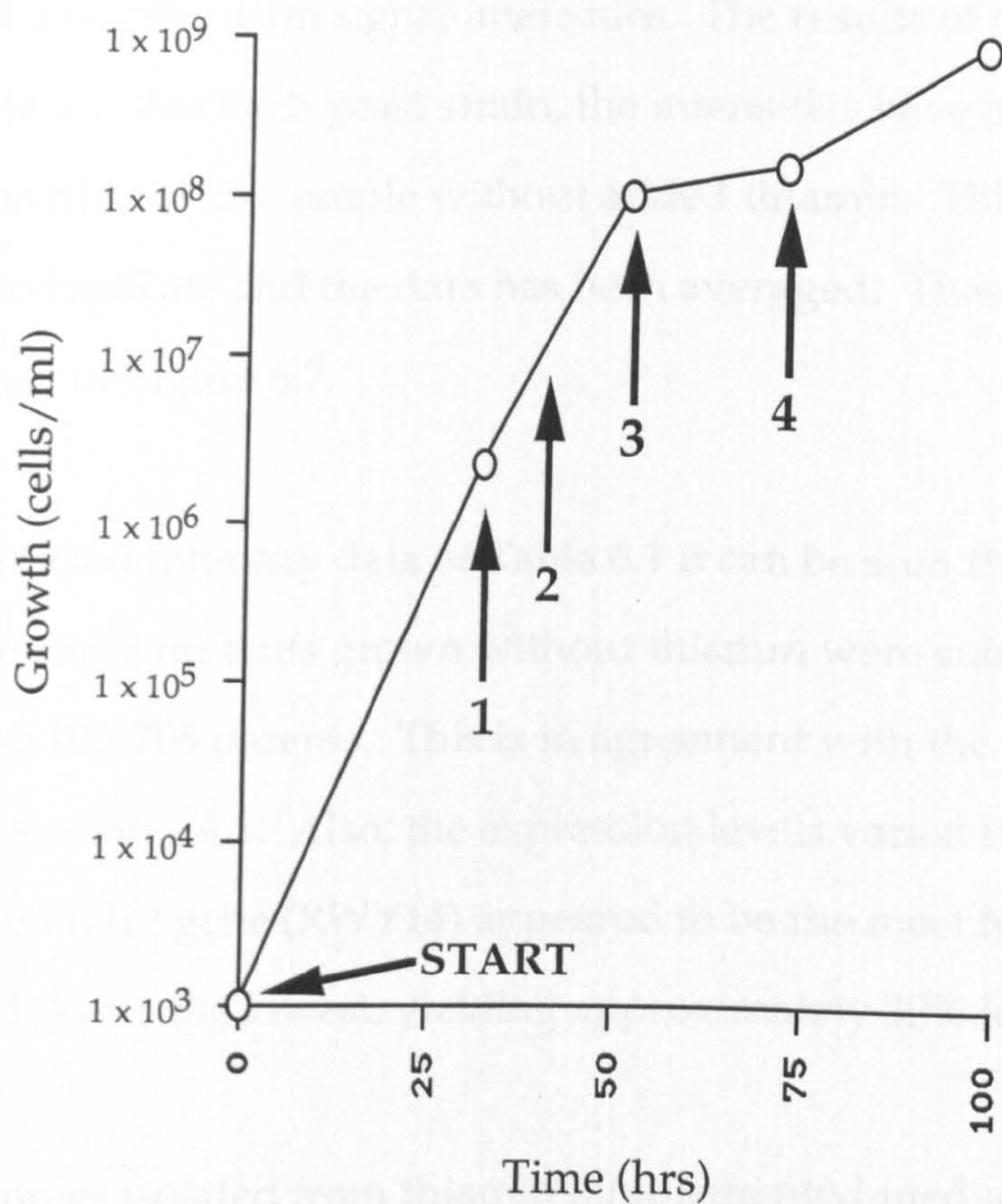


Figure 6.6 *THI5* expression during the course of an anerobic S288C batch culture in Wickerham's glucose medium without thiamin. Total RNA was isolated from a culture sample which was taken at the time points shown above. The RNA was electrophoresed followed by northern hybridisation using the pRH4 *THI5* probe. The same blot was then hybridised with an actin probe.

standardised using the actin signal intensities. The results of one such experiment are listed in Table 6.1. For each yeast strain, the intensities have been expressed as a percentage relative to the sample without added thiamin. This experiment was performed in duplicate and the data has been averaged. These figures are displayed in the bar chart of Figure 6.7.

From the corrected intensity data of Table 6.1 it can be seen that the transcript levels of each of the triple mutants grown without thiamin were substantially lower than the S288C and BY4705 parents. This is in agreement with the northern blot data presented in section 6.4.1. Also, the expression levels varied between the triple mutants. The *THI12* gene (RWY14) appeared to be the most highly expressed whilst *THI13* (RWY11) was the lowest, yielding approximately 30% less mRNA.

For those samples isolated from thiamin-supplemented media, basal expression was seen for each strain totalling not more than 8% of maximal expression (Figure 6.7). Similar low levels were also seen for medium containing both precursors, HET and HMP.

Differences between the four genes were seen for samples from media supplemented only with HMP. The *THI11* (RWY12) and *THI13* (RWY11) genes both exhibited the low levels of transcription (described above) meaning that these genes were almost fully repressed. Both *THI5* (RWY13) and *THI12* (RWY14) yielded transcript levels that were approximately 80% lower than their maximum and so were not fully repressed by HMP. These differences in patterns of gene expression mirrors the type of promoters that are found upstream of these genes. Both *THI11* and *THI13* possess 100% DNA sequence identity for over 1 Kb upstream of their ATG translational start site. This common promoter region shows differences to that of *THI5* and *THI12* which are similar to each other. For 1 Kb upstream of *THI5* and *THI12*, there are differences at just 8 nucleotides (Chapter 2, section 2.3.1).

Table 6.1 Slot blot hybridisation of RNA from *S. cerevisiae* strains S288C, BY4705, RWY11, RWY12, RWY13 and RWY14. All strains were initially inoculated into Wickerham's glucose medium without thiamin. After overnight incubation each yeast culture was then inoculated (to a final cell density of 3×10^5 cells/ml) into the same medium with or without thiamin or with the precursors HMP or HET replacing the vitamin (24 cultures in total). Cells were harvested once a cell density of $2-5 \times 10^7$ cells/ml was reached and total RNA was isolated. 2 μ g of each RNA sample was applied to the membrane using the slot blot apparatus (see Materials and Methods). This was performed in duplicate with one membrane subsequently hybridised with the pRH4 *THI5* probe and the other hybridised with the actin probe. The intensity units of each 'slot' hybridised with *THI5* was determined using the PhosphorimagerTM apparatus. These units were standardised with respect to the intensity of the corresponding 'slot' that was hybridised with the actin probe and are shown in the Table (corrected *THI5*). Figures in brackets compare the units of maximal expression in thiamin-depleted media (BY4705 has been set at 100%). These values were then expressed as a percentage of the corrected intensity value after growth in medium without thiamin (% compared to w/o thi). This experiment was repeated and included analysis of a culture grown in the presence of both HET and HMP. The percentages from both experiments (normalised to w/o thi data) are represented as average values (average %).

RNA sample	Corrected THI5	% compared to w/o thi	Average %
S288C +thi	2706	2	4
w/o thi	(123%) 157800	100	100
+HMP	22538	14	18
+HET	115131	73	72
+HMP+HET			4
BY4705 +thi	7302	6	7
w/o thi	(100%) 128538	100	100
+HMP	34489	27	27
+HET	75819	59	60
+HMP+HET			4
RWY11 +thi	3838	6	7
w/o thi	(47%) 59977	100	100
+HMP	4671	8	8
+HET	25656	43	40
+HMP+HET			4
RWY12 +thi	2134	3	4
w/o thi	(61%) 78082	100	100
+HMP	4539	6	8
+HET	64637	83	82
+HMP+HET			6
RWY13 +thi	2418	4	5
w/o thi	(49%) 62938	100	100
+HMP	14457	23	21
+HET	59489	95	98
+HMP+HET			4
RWY14 +thi	1467	2	4
w/o thi	(67%) 86064	100	100
+HMP	16269	19	21
+HET	79669	93	82
+HMP+HET			2

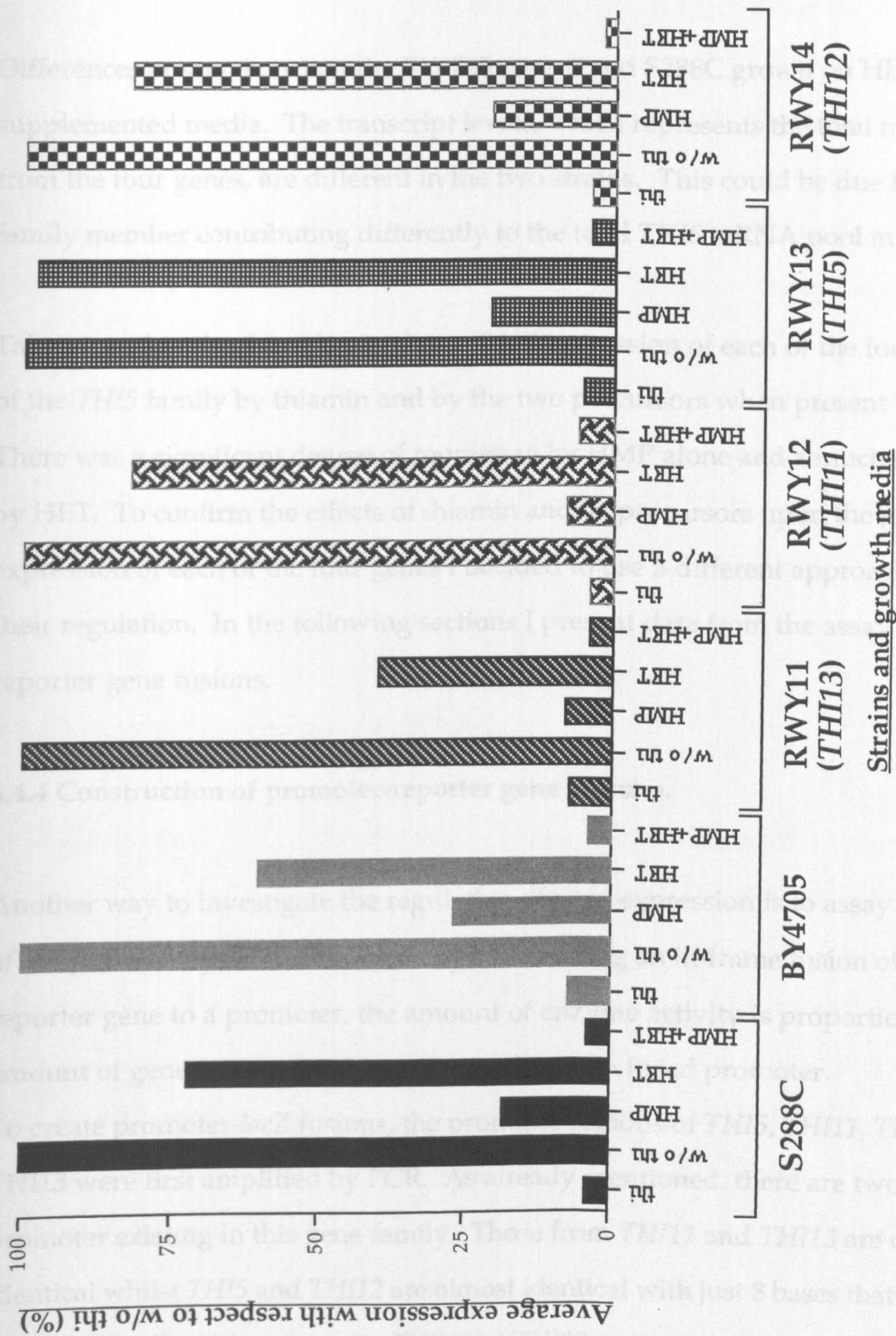


Figure 6.7 Graphical representation of the average percentage values listed in table 6.1. The gene present in each of the triple mutant strains is written in brackets.

In response to HET only, maximal expression of *THI5* was not significantly altered whilst *THI11* and *THI12* expression is slightly lower (20% reduction). It is *THI13* which appears to be most repressed by HET, showing a 60% reduction in mRNA levels.

Differences are seen between the BY4705 parent and S288C grown on HET or HMP-supplemented media. The transcript levels, which represents the total mRNA pool from the four genes, are different in the two strains. This could be due to each gene family member contributing differently to the total *THI5* mRNA pool in S288C.

Taken together the data clearly showed full repression of each of the four members of the *THI5* family by thiamin and by the two precursors when present together. There was a significant degree of repression by HMP alone and a much smaller effect by HET. To confirm the effects of thiamin and its precursors upon the gene expression of each of the four genes I decided to use a different approach to examine their regulation. In the following sections I present data from the assay of promoter-reporter gene fusions.

6.4.4 Construction of promoter-reporter gene fusions.

Another way to investigate the regulation of gene expression is to assay the activity of a reporter enzyme such as lacZ. By constructing an in-frame fusion of the reporter gene to a promoter, the amount of enzyme activity is proportional to the amount of gene expression that is driven from the fused promoter.

To create promoter-*lacZ* fusions, the promoter regions of *THI5*, *THI11*, *THI12* and *THI13* were first amplified by PCR. As already mentioned, there are two types of promoter existing in this gene family. Those from *THI11* and *THI13* are completely identical whilst *THI5* and *THI12* are almost identical with just 8 bases that differ (Chapter 2). For this reason the *THI5* and *THI12* promoters (680 bp) were amplified from their respective genomic clones, pRH11 (Hather, 1996) and pSK1.1 (Vandyck *et*

al., 1995). As no clone was available for *THI11* or *THI13*, the common promoter DNA was amplified from S288C genomic DNA.

The primers were designed so that they could be subsequently cloned into the *EcoRI*/*Bam*HI sites of plasmid pUP34 in such a way as to maintain the correct reading frame from the ATG translational start site between *THI5/11/12/13* and the *lacZ* gene. This was achieved by engineering a *Bam*HI restriction site into the primer which anneals to all four genes as displayed in Figure 6.8. Both the upstream primers of *THI11/13* and *THI5/12* had an *EcoRI* site linked to the 5' end. After cloning the PCR fragments into the *EcoRI*/*Bam*HI sites of pUC19, DNA sequencing was carried out from the *EcoRI* site into the insert DNA to check the identity of each promoter (data not shown). These promoter fragments were then subcloned into the much larger pUP34 *URA3 lacZ*-reporter plasmid followed by DNA sequencing across the site of the promoter-*lacZ* fusion (Figure 6.8 legend). The pUP34 clones have been named pHL4, pHL5 and pHL6 containing the promoters of *THI5*, *THI12* and *THI11/13*, respectively.

6.4.5 Assays of the reporter plasmids.

Expression of the three promoter-*lacZ* constructs was examined in *S. cerevisiae* strain W303a (*MATa*, *ade2-1*, *can1-100*, *leu2-3-112*, *trp1-1*, *ura3-1*, *his3-11-15*) which has been the preferred strain for other *lacZ* expression studies into thiamin regulation. This strain has been shown to possess the four *THI5* genes found in S288C (Chapter 3). W303a transformed to uracil prototrophy with either pHL4, pHL5 or pHL6 was grown in Wickerham's glucose medium supplemented without thiamin or with thiamin, HMP or HET. β -galactosidase assays were then performed. The results are displayed as averaged values in Table 6.2 and are compared graphically in Figure 6.9.

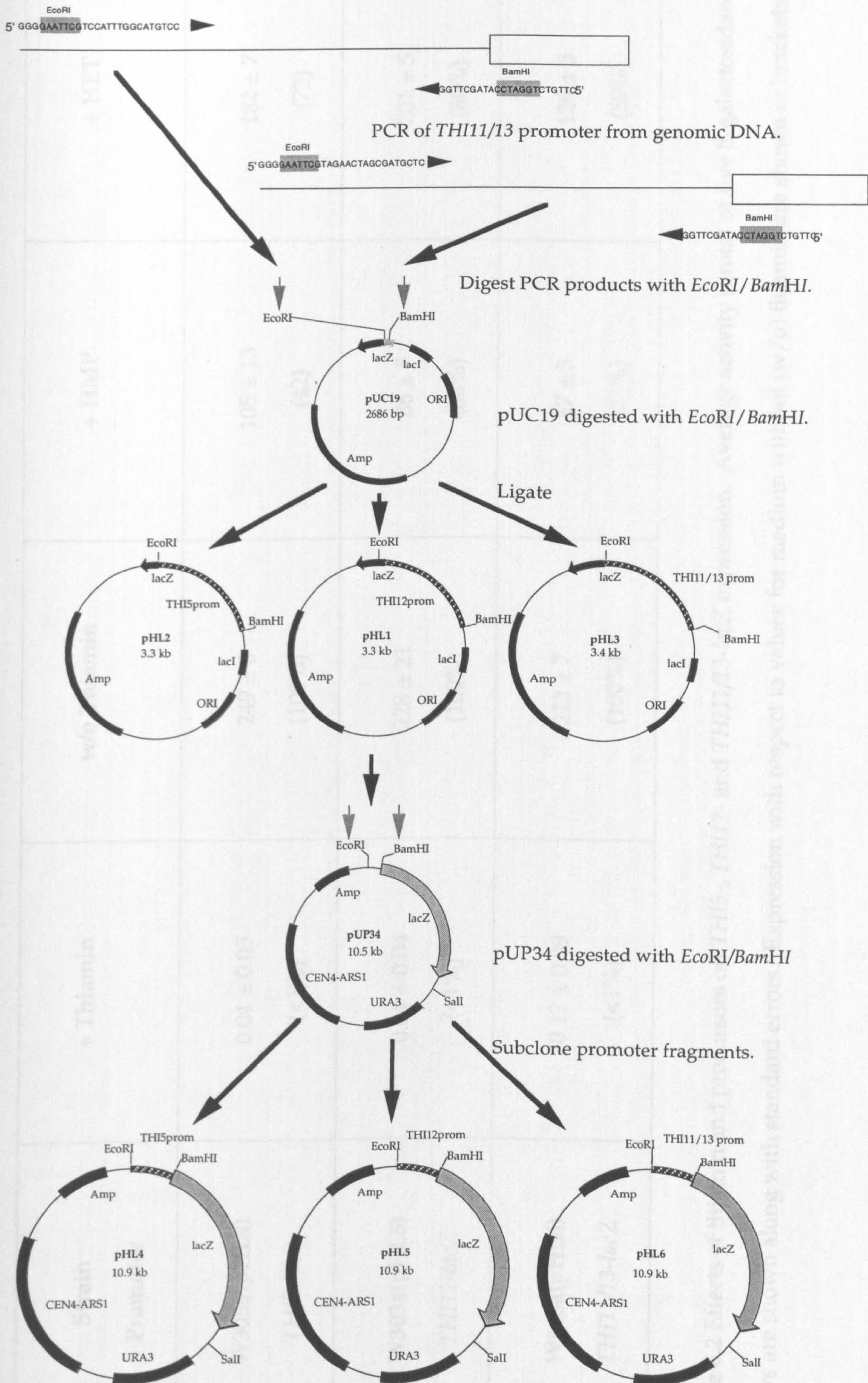
For all three constructs grown in thiamin-supplemented medium, basal expression of the reporter gene was seen. In each case, this was less than 1% of the maximal activity observed from cultures grown in unsupplemented medium (w/o thiamin)

Figure 6.8 Construction of promoter:*lacZ* gene fusions. Construction of the reporter plasmids pHL4 (containing the *THI5* promoter), pHL5 (*THI12* promoter) and pHL6 (*THI11/THI13* promoter) is described in the text (Section 6.4.4).

DNA sequence analysis was carried out across the fusion junction of these three plasmids and is shown below. The result is an in-frame fusion between the start codon of the *THI5*, *THI12*, *THI11/13* gene and codon 10 of *lacZ*.

pHL4	ACACACTTCCAAGCTATGGATCCCGTCGTT	
pHL5	ACACACTTCCAAGCTATGGATCCCGTCGTT	
pHL6	ACATACTTCCAAGCTATGGATCCCGTCGTT	
		M N P V V
<i>THI5</i>	ACACACTTCCAAGCTATG	
<i>lacZ</i>	ATGACCATGATTACGGATTCAC	
		M T M I T D S L A V V

PCR of *THI5* and *THI12* promoters from pRH11 and pSK1.1 respectively



Strain Promoter	+ Thiamin	w/o Thiamin	+ HMP	+ HET
W303a(pHL4) <i>THI5-lacZ</i>	0.04 ± 0.05 (<1%)	249 ± 6 (100%)	105 ± 13 (42)	182 ± 7 (73)
W303a(pHL5) <i>THI12-lacZ</i>	0.04 ± 0.04 (<1%)	228 ± 21 (100%)	86 ± 5 (38%)	201 ± 5 (88%)
W303a(pHL6) <i>THI11/13-lacZ</i>	0.12 ± 0.09 (<1%)	213 ± 7 (100%)	57 ± 3 (27%)	126 ± 3 (59%)

Table 6.2 Effects of thiamin and precursors on *THI5*-, *THI12*- and *THI11/13-lacZ* expression. Average activity (units) of five β-galactosidase assays are shown along with standard errors. Expression with respect to values for medium without (w/o) thiamin are shown in brackets.

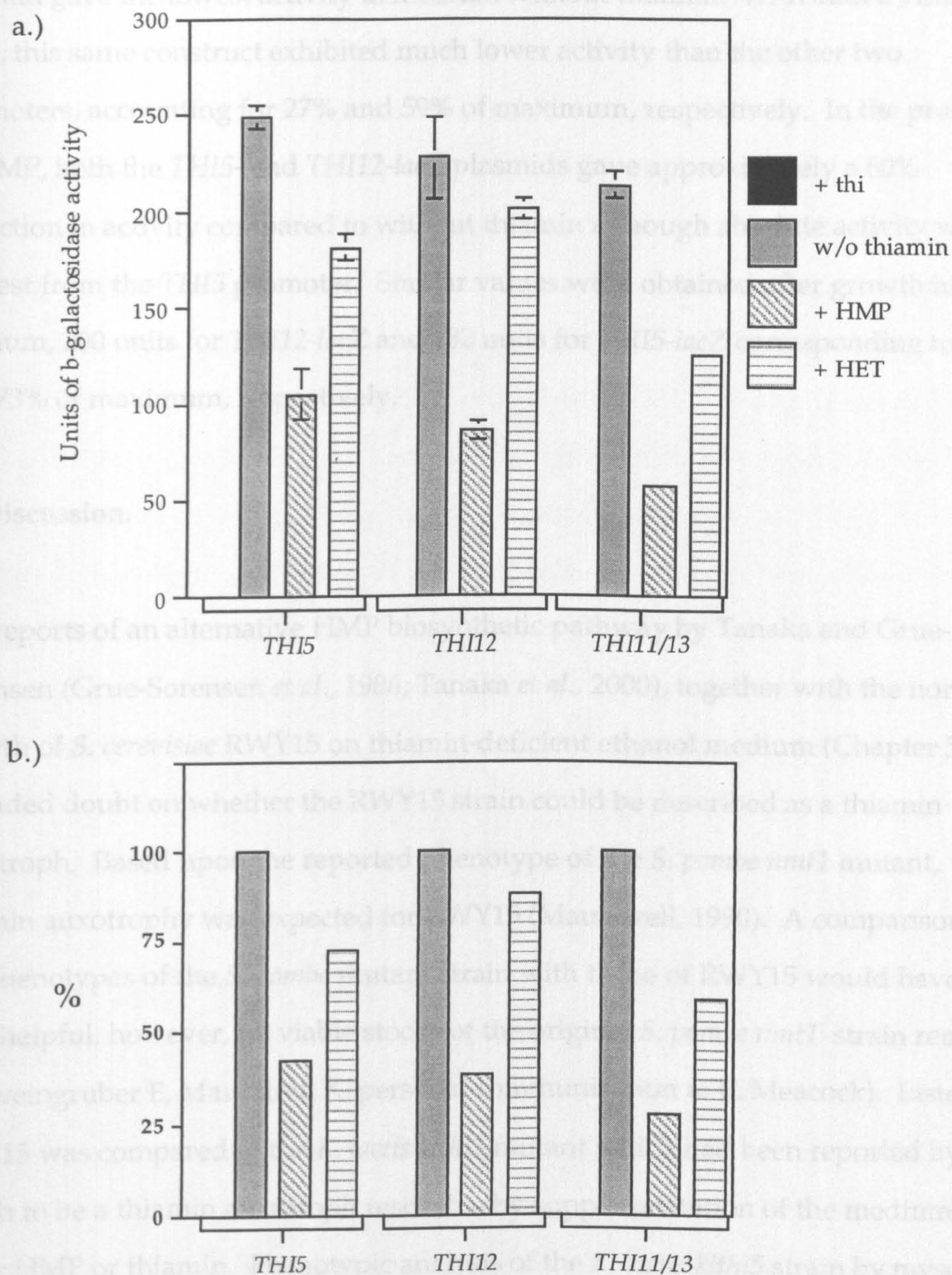


Figure 6.9 Effects of thiamin and precursors on *THI5*-, *THI12*- and *THI11/13-lacZ* expression. These are graphical representations of the data from table 6.2.

a.) B-galactosidase activity

b.) B-galactosidase activity expressed as a percentage compared to activity in medium without added thiamin.

which gave similar values of over 200 units. Although marginal, the *THI11/13-lacZ* plasmid gave the lowest activity in medium without thiamin. With added HMP or HET, this same construct exhibited much lower activity than the other two promoters, accounting for 27% and 59% of maximum, respectively. In the presence of HMP, both the *THI5-* and *THI12-lacZ* plasmids gave approximately a 60% reduction in activity compared to without thiamin although absolute activity was highest from the *THI5* promoter. Similar values were obtained after growth in HET medium, 200 units for *THI12-lacZ* and 182 units for *THI5-lacZ* corresponding to 88% and 73% of maximum, respectively.

6.5 Discussion.

The reports of an alternative HMP biosynthetic pathway by Tanaka and Grue-Sorensen (Grue-Sorensen *et al.*, 1986; Tanaka *et al.*, 2000), together with the normal growth of *S. cerevisiae* RWY15 on thiamin-deficient ethanol medium (Chapter 5), provided doubt on whether the RWY15 strain could be described as a thiamin auxotroph. Based upon the reported phenotype of the *S. pombe nmt1* mutant, thiamin auxotrophy was expected for RWY15 (Maundrell, 1990). A comparison of the phenotypes of the *S. pombe* mutant strain with those of RWY15 would have been very helpful, however, no viable stocks of the original *S. pombe nmt1*⁻ strain remain (Schweingruber E, Maundrell K, personal communication to P. Meacock). Instead, RWY15 was compared to the *K. lactis klthi5* mutant which had been reported by D. Walsh to be a thiamin auxotroph rescuable by supplementation of the medium with either HMP or thiamin. Phenotypic analysis of the *K. lactis klthi5* strain by myself found that, after a sufficiently long incubation period, this strain is not a strict thiamin auxotroph and was seen to grow almost like wild-type *K. lactis*. Long-term incubation of the *S. cerevisiae* RWY15 strain on the same medium resulted in very limited growth which may be a result of residual amounts of thiamin in the cells with no actual *de novo* biosynthesis taking place. However, upon transfer to fresh thiamin-deficient medium, RWY15 was still seen to exhibit a little growth. This limited growth might be due to HMP formation by the proposed alternative

pathway. This alternative pathway, as described by Tanaka, has been suggested to be responsible for the anaerobic biosynthesis of HMP. With this in mind, the growth of both wild-type *S. cerevisiae* and the RWY15 mutant was investigated on agar plates and in liquid cultures of thiamin-deficient medium. The agar plate studies showed that successful anaerobic propagation of *S. cerevisiae* is dependent upon the presence of HMP, and not HET. This suggests that the principal HMP biosynthetic pathway is oxygen-dependent. During anaerobic batch cultures of both S288C and RWY15, these strains showed initial rapid growth for up to 50 hours after inoculation followed by very slow growth. Further growth of S288C was seen upon addition of HMP. Curiously, HMP addition had no effect on the RWY15 mutant. For both strains, the initial growth could be due to the mobilisation of intracellular stores of thiamin or the active cofactor, ThDP. A decline in growth was therefore possibly due to depletion of intracellular ThDP. This suggests that under anaerobic conditions no *de novo* biosynthesis of thiamin occurs. However, the initial period of rapid growth seen for S288C and RWY15 in anaerobic conditions corresponds to approximately 7-8 doublings (generations) within about 50 hours from inoculation. In comparison, a *S. cerevisiae thi4* mutant that was inoculated into thiamin-free medium grew aerobically for only three generations. Also, it has been shown that the intracellular thiamin concentration falls rapidly to a basal level within 8 hours of inoculation of *S. cerevisiae* into thiamin-deficient medium (Praekelt, *et al.*, 1994). In Praekelt's experiment, the yeast had been pre-grown in 2 μ M thiamin medium which is identical to that used for the starter cultures in my experiments. This indicates that an alternative thiamin biosynthetic pathway may allow limited HMP formation for thiamin biosynthesis which provides enough thiamin for a further 4-5 generations before thiamin becomes limiting, resulting in the growth plateau at 50 hours. The small amount of residual growth after this time may be due to the alternative HMP pathway which produces, inefficiently, HMP allowing a small amount of thiamin biosynthesis. An inefficient alternative pathway would agree with the observations of Grue-Sorenson (Chapter 1).

During aerobic growth, RWY15 showed a similar growth curve to that seen for anaerobic conditions. This suggests that the inefficient alternative HMP biosynthetic pathway may be constitutively functional during both aerobic and anaerobic growth. Studies of S288C suggested that the main pathway, which involves *THI5*, is only functional during aerobiosis. These interpretations of the data are summarised in the model of Figure 6.10. The only paradox is that Tanaka *et al.* found no incorporation of labelled formate into the pyrimidine ring during anaerobiosis, whereas Grue-Sorenson's alternative pathway (which is aerobic) sees formate incorporation at a different site of the HMP molecule to the principal histidine and pyridoxine route. A number of experiments would help to clarify the existence of the other HMP pathway as illustrated in the model. To begin with, the intracellular ThDP levels could be determined and compared between aerobic and anaerobic batch fermentations of both S288C and RWY15. Secondly, if a *thi6* mutant strain shows the same course of growth as RWY15 and S288C during anaerobiosis, then growth of the latter two strains cannot be due to an alternative HMP pathway as any production of the precursor cannot proceed through the condensation reaction to form thiamin phosphate (Chapter 1).

Whereas RWY15 exhibited severely restricted growth on glucose media which may be indicative of thiamin auxotrophy, the *K.lactis klthi5* mutant was a thiamin prototroph. One possible explanation for this is based upon the *K. lactis* sequencing data reported in Chapter 4. Upstream of *klthi5* is a gene that encodes an enzyme that is not found in *S. cerevisiae*. This enzyme is homologous to dihydropyrimidinase which is involved in the assimilation of pyrimidines, a trait which is not common to the fungi generally. Degradation products of pyrimidine nucleotides such as uracil may be converted to HMP by a route independent of Klthi5p.

The investigation into the regulation of the four genes has produced interesting results. Previous studies have only focused on *THI5* and *THI12* expression which possess almost identical promoters that are quite different from the promoter shared by *THI11* and *THI13*. The data presented in this study show that all four *THI5*

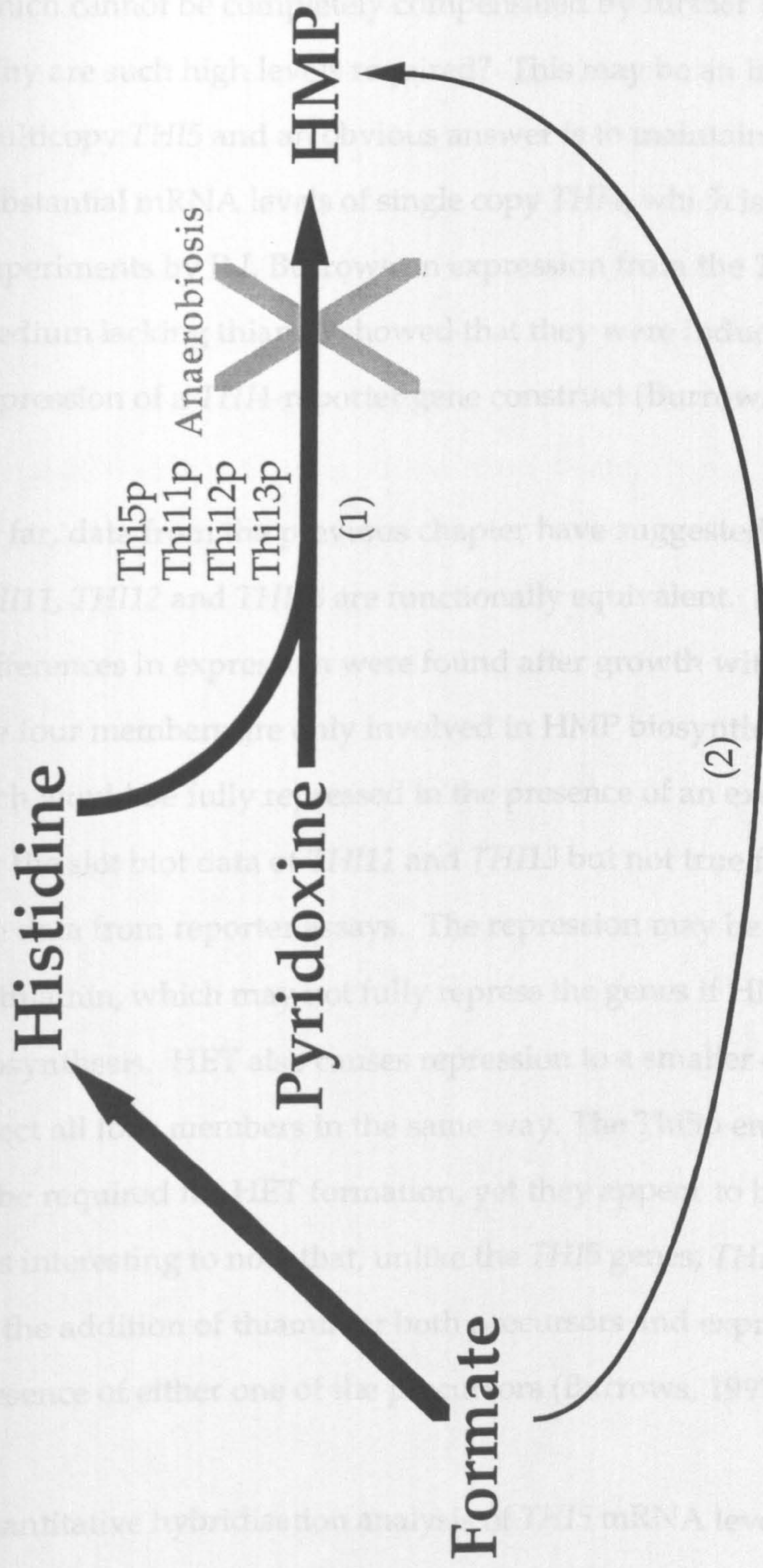


Figure 6.10 Model showing two biosynthetic routes to HMP formation. Grue-Sorenson *et al.* (1986) showed that formate incorporation in route (1) is approximately three times greater than for route (2). Tanaka *et al.* (2000) showed that this principal route (1) does not exist during anaerobic conditions. Data from this study have suggested that the Thi5 isozymes are dispensible for anaerobic growth where the inefficient route (2) is solely responsible for HMP synthesis (see text).

members, like *S. pombe nmt1*, are negatively regulated by thiamin. Also all members are fully repressed in the presence of both the precursors HET and HMP. The RNA analysis finds that each of the genes, when present on their own, does not produce the high levels of mRNA seen when all are present. This may (in part) explain why more than one gene is required in *S. cerevisiae*; to produce these high transcript levels which cannot be completely compensated by further upregulating just one gene.

Why are such high levels required? This may be an important reason for selection of multicopy *THI5* and an obvious answer is to maintain a comparable frequency to the substantial mRNA levels of single copy *THI4*, which is involved in HET biosynthesis. Experiments by R.J. Burrows on expression from the *THI5* and *THI12* promoters in medium lacking thiamin showed that they were induced to just 42% of the maximal expression of a *THI4*-reporter gene construct (Burrows, *et al.*, 2000).

So far, data from the previous chapter have suggested that the four genes *THI5*, *THI11*, *THI12* and *THI13* are functionally equivalent. However, noticeable differences in expression were found after growth with just one of the precursors. If the four members are only involved in HMP biosynthesis, it might be expected that each would be fully repressed in the presence of an exogenous supply. This was true for the slot blot data of *THI11* and *THI13* but not true for *THI5* and *THI12* or any of the data from reporter assays. The repression may be due to subsequent production of thiamin, which may not fully repress the genes if HMP uptake is slower than HMP biosynthesis. HET also causes repression to a smaller degree and again seems not to affect all four members in the same way. The Thi5p enzymes have not been shown to be required for HET formation, yet they appear to be regulated by this precursor. It is interesting to note that, unlike the *THI5* genes, *THI4* is only negatively regulated by the addition of thiamin or both precursors and expression is not altered in the presence of either one of the precursors (Burrows, 1997).

Quantitative hybridisation analysis of *THI5* mRNA levels gave different results to the β -galactosidase studies of the promoter-*lacZ* constructs. The RNA approach measures the expression as a result of chromosomal influences other than the

immediate promoter region but to detect the contribution of individual genes, the other three had to be deleted which may unnaturally alter the regulation of the remaining member. The background strain used to measure *lacZ* expression possesses all four members and promoter-reporter gene fusions yield more accurate expression data than mRNA quantitation by slot blot hybridisation. However, the chromosomal influences could be significant. For example, telomere-proximal to *THI5* and *THI12* are the thiamin-regulated *SNO* and *SNZ* promoters at a distance of only 1.6 Kb (Chapter 2, Figure 2.5).

The best approach to investigate the transcriptional regulation of the four genes would be to measure transcript levels quantitatively in a wild-type background such as the S288C strain. A modified reverse transcriptase-PCR method (RT-PCR) could be developed which would allow the detection of mRNA from individual genes by taking advantage of the small number of differences between their immediate 3'-untranslated regions.

General Discussion.

Thi5p homologues are distributed among the fungi, bacteria and archaea (Chapter 2, Table 2.1). The presence of a Thi5p homologue suggests that the pyrimidine moiety of thiamin (HMP) is synthesised from pyridoxine in all of these organisms. This hypothesis is probably true for the highly conserved fungal homologues now that both the *S. pombe* and *S. cerevisiae* proteins have been shown to be involved in HMP formation. It is not known whether any of the more diverged bacterial and archaeal Thi5p homologues have any involvement in thiamin biosynthesis. However, the *H. influenzae* homologue appears to be transcribed within an operon which includes a putative HMP-P kinase.

The presence of an *E. coli* ThiC protein homologue is associated with an alternative route to HMP biosynthesis which uses an intermediate of purine biosynthesis, called 5-aminoimidazole ribotide (AIR), as the substrate. The bacteria and archaea appear to possess either the pyridoxine or AIR pathways, or both pathways together. Only the pyridoxine (Thi5p) pathway has been retained in the fungal kingdom, whilst only the AIR (ThiC) pathway has been retained in the plantae kingdom.

An informatic analysis of the fungal Thi5p proteins has found a conserved transmembrane domain within the central portion of their primary sequences. This observation suggests a membrane location for Thi5p although the identity of the membrane is unknown and there are no obvious targeting signals within any of the fungal sequences. A useful approach to identify the subcellular location of Thi5p would be to visualise the destination of a Thi5p-GFP protein fusion in *S. cerevisiae*.

In *S. cerevisiae*, Thi5p is present as four highly conserved isozymes encoded by the *THI5*, *THI11*, *THI12* and *THI13* genes. All members of this gene family are subtelomerically located and they share these regions with members of other gene

families and are distributed in a highly organised manner (Chapter 2, section 2.3). The adjacent centromere proximal genes are all members of the *AAD* family. Telomere proximal is one of two duplicated blocks of genes which appear to encode functionally related proteins. These blocks are termed the 'sugar metabolism' cluster and the 'pyridoxine cluster' (Chapter 2). Each duplicate cluster is very highly conserved with its partner, even within the intergenic regions.

A model for the evolution of these four subtelomeric regions is shown in Figure 7.1. The model proposes that the ancestral gene of the *THI5* family was found adjacent to a *COS* gene and an *AAD* gene and that all three genes were subtelomerically located. It is interesting to see that in *K. lactis*, the single copy *THI5* homologue is also adjacent to the single copy *COS* homologue (Chapter 4, Figure 4.4). The *S. cerevisiae* ancestral conformation (*COS-THI5-AAD*) probably underwent a duplication by some kind of ectopic recombination event between itself and another subtelomeric region. As shown in Figure 7.1, a subsequent translocation event probably took place whereby the sugar metabolism cluster or the pyridoxine cluster was inserted between *THI5* and *COS*. These translocations might have been either as a whole 'block' unit or as the insertion of each gene separately. The whole-block translocation is the favoured mechanism due to the fact that *SNO* and *SNZ* homologues of the pyridoxine cluster are always found adjacent to each other (Padilla *et al.*, 1998). Following these translocations, another inter-subtelomeric recombination event probably took place between two other chromosomes to give the conformation seen in the present *S. cerevisiae* S288C genome (Chapter 2, Figure 2.5).

The existence of *THI5* as a gene family was found to be exclusive to those yeasts of the closely related *Saccharomyces sensu stricto* complex (Chapter 3). Yeasts outside of this subgroup possessed the homologue in either one or zero copies. Those that possess no copies are documented as thiamin auxotrophs and include a significant number of the *Saccharomyces sensu lato* species such as *S. servazzi*, *S. castelli* and *S. unisporus*. Data from this study suggested that the genetic basis of thiamin

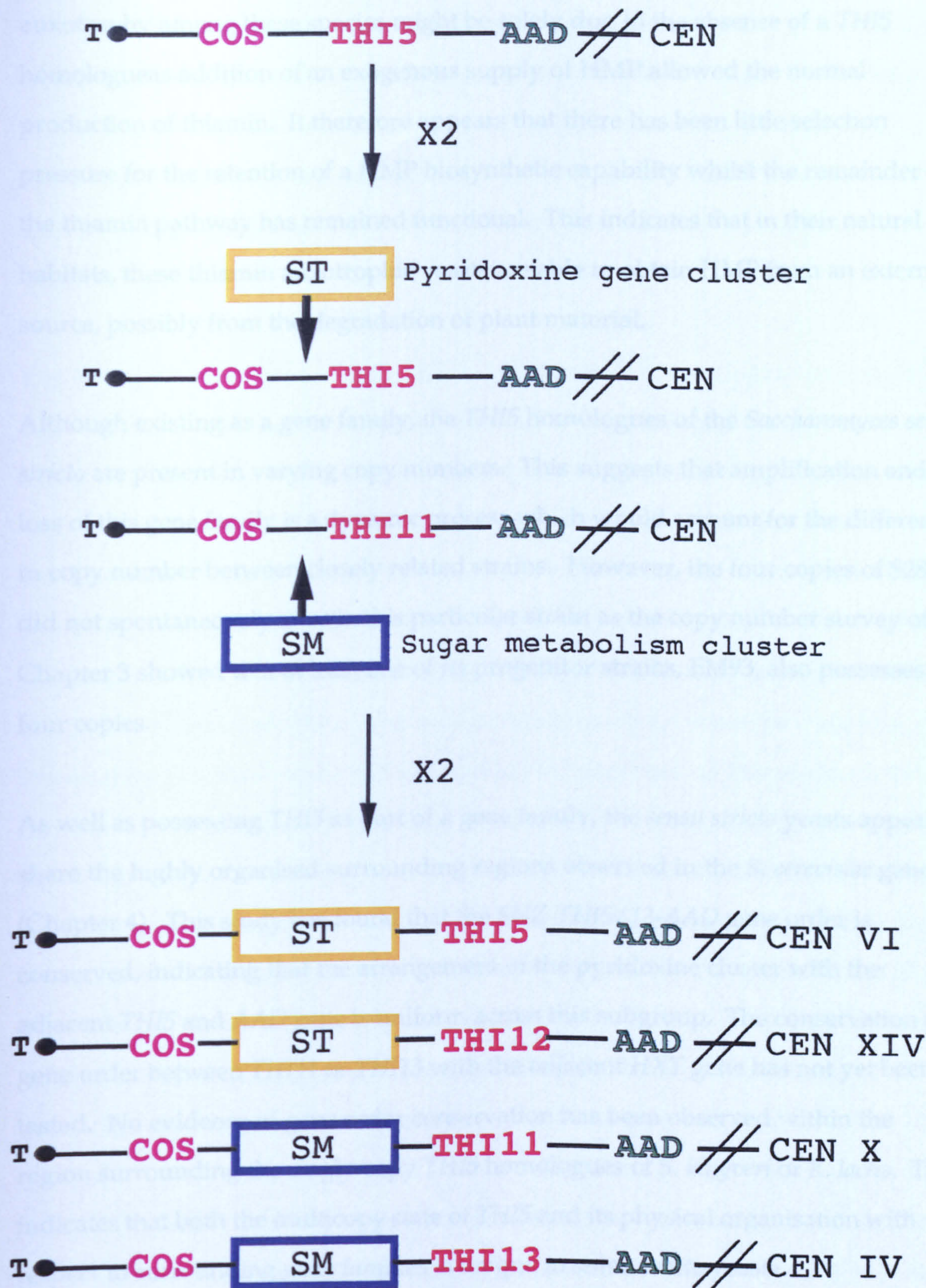


Figure 7.1 Model showing the evolution of the subtelomeric regions of *S. cerevisiae* chromosomes IV-L, VI-L, X-R and XIV-L.

auxotrophy among these species might be solely due to the absence of a *THI5* homologue. As addition of an exogenous supply of HMP allowed the normal production of thiamin. It therefore appears that there has been little selection pressure for the retention of a HMP biosynthetic capability whilst the remainder of the thiamin pathway has remained functional. This indicates that in their natural habitats, these thiamin auxotrophic yeasts are able to obtain HMP from an external source, possibly from the degradation of plant material.

Although existing as a gene family, the *THI5* homologues of the *Saccharomyces sensu stricto* are present in varying copy numbers. This suggests that amplification and loss of this gene family is a dynamic process which would account for the differences in copy number between closely related strains. However, the four copies of S288C did not spontaneously arise in this particular strain as the copy number survey of Chapter 3 showed that at least one of its progenitor strains, EM93, also possesses four copies.

As well as possessing *THI5* as part of a gene family, the *sensu stricto* yeasts appear to share the highly organised surrounding regions observed in the *S. cerevisiae* genome (Chapter 4). This study has found that the *SNZ-THI5/12-AAD* gene order is conserved, indicating that the arrangement of the pyridoxine cluster with the adjacent *THI5* and *AAD* gene is uniform across this subgroup. The conservation of gene order between *THI11* or *THI13* with the adjacent *HXT* gene has not yet been tested. No evidence of gene order conservation has been observed within the region surrounding the singly-copy *THI5* homologues of *S. kluyveri* or *K. lactis*. This indicates that both the multicopy state of *THI5* and its physical organisation with respect to surrounding gene families is unique to *sensu stricto* yeasts.

Although I found no homologous relationship of neighbouring genes between *S. kluyveri*, *K. lactis* and the *sensu stricto*, there is an interesting functional relationship. As will be discussed later in this chapter, in *S. cerevisiae* (and probably the whole *sensu stricto* complex) there is an emerging functional relationship between *THI5* and

THI12 and the neighbouring pyridoxine gene cluster. In *S. kluyveri*, *Sk.THI5* is found immediately downstream of a homologue of *S. cerevisiae* *PPR1* which encodes a transcription factor that regulates expression of genes encoding pyrimidine nucleotide biosynthetic enzymes (Chapter 4). In *K. lactis*, *Kl.THI5* is adjacent to a gene encoding a putative dihydropyrimidinase which is involved in pyrimidine degradation.

The four *S. cerevisiae* genes; *THI5*, *THI11*, *THI12* and *THI13* are highly conserved at the DNA level with the vast majority of differences resulting in synonymous codon changes. This means that there is a selection pressure to retain their integrity and copy number and so each gene must be important. The construction of all combinations of single, double, triple and quadruple mutants has been carried out successfully using a one-step gene replacement method (Chapter 5). The *S. cerevisiae* quadruple *thi5*, *thi11*, *thi12*, *thi13* mutant, called RWY15, shows thiamin auxotrophy on glucose minimal medium. This phenotype does not appear to be complete auxotrophy as a small amount of residual growth was seen on thiamin-depleted agar medium after a long incubation period (Chapter 6). In liquid medium, this slow growth was significant. Supplementation of the growth media with HMP, but not HET, restored normal growth of the RWY15 strain whereas pyridoxine alone did not restore growth indicating that the Thi5 enzyme is involved in synthesis of HMP from this vitamin. None of the four triple mutant strains exhibited any growth defects on thiamin-depleted glucose medium meaning this gene family is functionally redundant with regards to thiamin biosynthesis.

Each of the four *S. cerevisiae* members was found to be regulated by thiamin, HMP and, to a certain degree, HET. However, subtle differences in the regulation of the four members were observed with respect to the thiamin precursors, HMP and HET. The limitations of the experimental procedures have been discussed in Chapter 6 with recommendations for further investigation of their regulation.

One of the aims of this study was to investigate claims for the existence of an alternative HMP biosynthetic pathway which is postulated to be responsible for the formation of this precursor during anaerobiosis (Grue-Sorensen *et al.*, 1986; Tanaka *et al.*, 2000). It has been shown that the successful anaerobic propagation of *S. cerevisiae* is dependent upon the presence of exogenous HMP or thiamin but not HET (Chapter 6). The data also show that anaerobic HMP formation is slow and inefficient and is independent of the pyridoxine route via Thi5p or any of its isozymes. A number of key experiments have been suggested to give further confirmation of the existence of two pathways (Chapter 6 Discussion). If the alternative pathway is identified then the next question is to ask why the principal route to HMP formation is defective when oxygen is limiting. One possibility is an enzymatic step that involves a redox reaction in the principal pathway. Under anaerobiosis, the reducing- or oxidising- agent may not be sufficiently recycled in the cell and so *S. cerevisiae* could have acquired some kind of bypass. It is by using alternative redox agents that *S. cerevisiae* can cope with oxygen depletion (Chapter 1, section 1.2.1).

The scheme of Figure 1.6 (Chapter 1) showed *S. cerevisiae* HMP biosynthesis to be predominantly through enzymes that are encoded by gene families. From the characterisation of the *THI5* gene family as part of this study, a revised scheme is displayed in Figure 7.2 where these genes function at the step from pyridoxine to HMP-P. It is not known what form of pyridoxine is used as the substrate for HMP formation, but the best candidate is pyridoxal phosphate (Tazuya *et al.*, 1995). The glutamine amidotransferase reaction, which was previously recognised as being a consequence of Ade4p activity, now includes members of the *SNO* gene family in the revised scheme. This is because sequence analysis by (Padilla *et al.*, 1998) discovered that *SNO* showed a considerable degree of similarity to glutamine amidotransferases of bacteria. It would be interesting to see if purine biosynthesis, which uses PRA as substrate, is in anyway perturbed in the strain constructed by L. Marsh which contains deletions of all three *SNO* genes (Marsh, 2000).

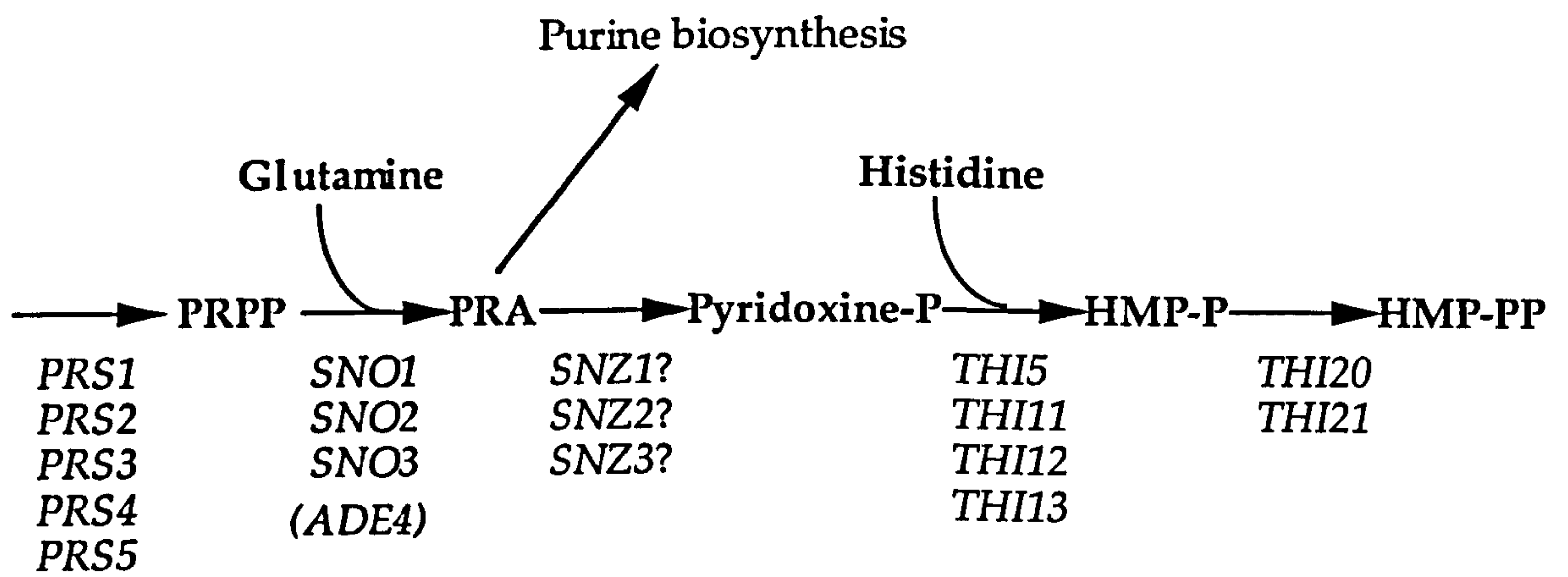


Figure 7.2 Proposed scheme of HMP-PP biosynthesis.

Concluding remarks from our understanding of HMP biosynthesis and the evolution and organisation of the *THI5* gene family of *S. cerevisiae* will be made in the following sections. The data discussed above and in the Introduction (Chapter 1) will be drawn upon to describe formally the HMP clusters of *S. cerevisiae* chromosomes VI and XIV and to discuss the likely mechanism that gave rise to subtelomerically repeated genes.

7.1 The HMP gene clusters of *Saccharomyces cerevisiae*.

The subtelomeric copies of the *SNO* and *SNZ* gene families appear to have a role in the formation of pyridoxine for entry into HMP biosynthesis (Chapter 1, Section 1.6). Consistent with this role, the subtelomeric *SNO* members, *SNO2* and *SNO3* have been shown to be regulated by thiamin (D. McKissock, personal communication). It therefore appears that the amplification of both the *SNO* and *SNZ* gene families allowed for the divergence of the subtelomeric members in terms of their regulation. These genes are found telomere proximal to *THI5* and *THI12* which, as a result of this study, are known to encode HMP biosynthetic isozymes. Therefore, three adjacent genes confer related functional roles in the biosynthesis of the pyrimidine moiety of thiamin diphosphate. The 'pyridoxine' gene clusters defined in Chapter 2 can now be extended to cover *THI5* and *THI12* and renamed the HMP gene clusters of chromosomes VI and XIV of *S. cerevisiae* (Figure 7.3). It will be interesting to see if other neighbouring genes are included in these HMP clusters, especially the physically linked conserved ORFs *YFL061w* and *YNL335w*.

It seems likely that the HMP cluster is responsible for the observed growth inhibition of *Saccharomyces sensu stricto* yeasts by thiamin in pyridoxine-free medium (Chapter 1, section 1.6.3; Kamihara and Nakamura, 1982). The addition of thiamin probably represses the gene expression of *SNO2/3* and *SNZ2/3* (along with *THI5/12*) causing a marked decrease in Sno and Snz enzyme activity which may not be restored by simply increasing *SNO1* and *SNZ1* gene expression. Alternatively, *SNO1* and *SNZ1* may not exist in many *sensu stricto* strains. The effects observed by Kamihara and

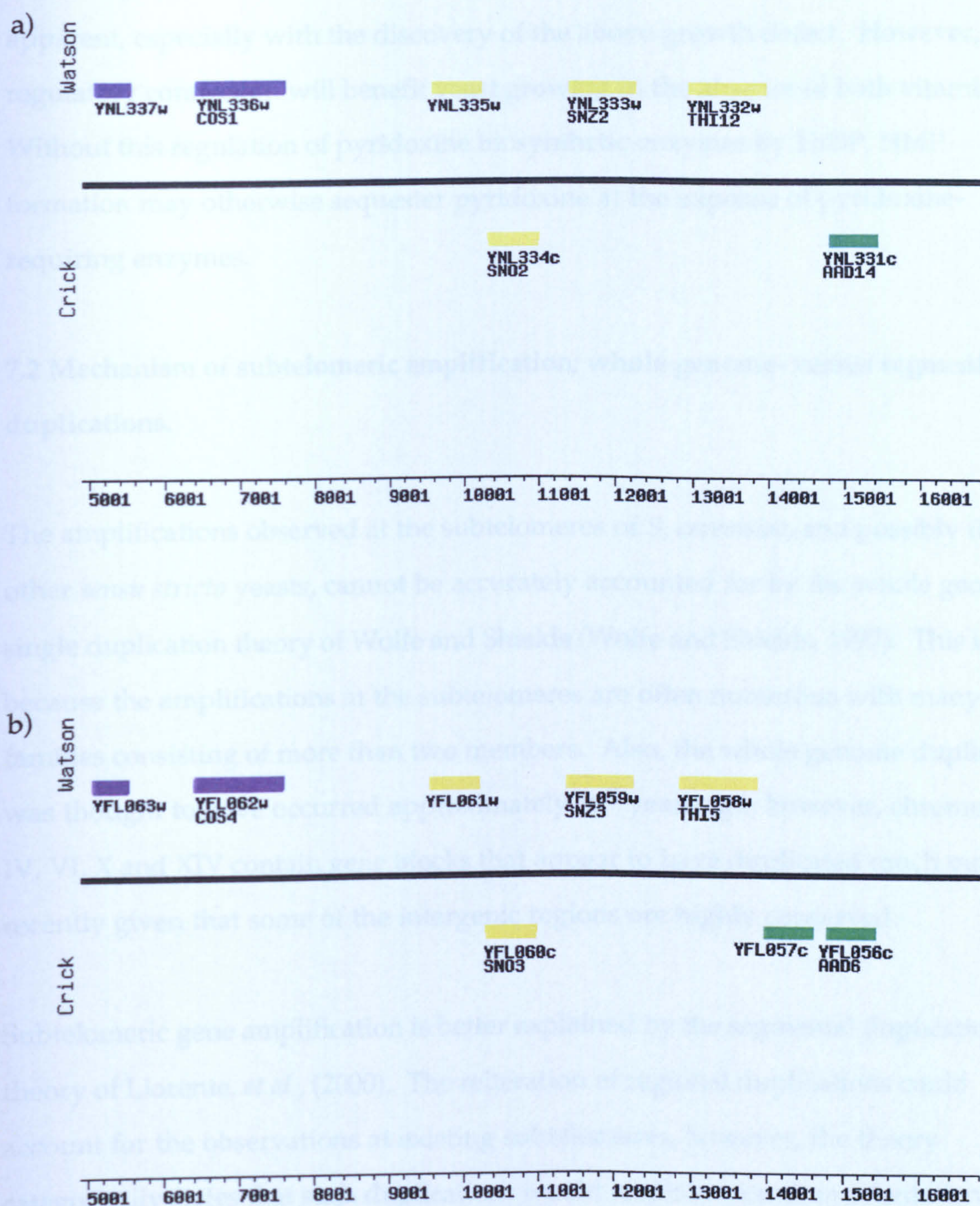


Figure 7.3 The HMP gene clusters (yellow) of chromosomes XIV (a) and VI (b). The scale shows chromosomal coordinates.

Nakamura are likely due to an 'induced' pyridoxine auxotrophy. The benefits of a regulatory link between thiamin and pyridoxine biosynthesis are not immediately apparent, especially with the discovery of the above growth defect. However, this regulatory connection will benefit yeast growing in the absence of both vitamins. Without this regulation of pyridoxine biosynthetic enzymes by ThDP, HMP formation may otherwise sequester pyridoxine at the expense of pyridoxine-requiring enzymes.

7.2 Mechanism of subtelomeric amplification; whole genome- versus segmental-duplications.

The amplifications observed at the subtelomeres of *S. cerevisiae*, and possibly the other *sensu stricto* yeasts, cannot be accurately accounted for by the whole genome single duplication theory of Wolfe and Shields (Wolfe and Shields, 1997). This is because the amplifications at the subtelomeres are often numerous with many gene families consisting of more than two members. Also, the whole genome duplication was thought to have occurred approximately 10^8 years ago, however, chromosomes IV, VI, X and XIV contain gene blocks that appear to have duplicated much more recently given that some of the intergenic regions are highly conserved.

Subtelomeric gene amplification is better explained by the segmental duplication theory of Llorente, *et al.*, (2000). The reiteration of regional duplications could account for the observations at existing subtelomeres, however, the theory categorically states that such duplications would result in a random orientation of the genes with respect to the centromere. The data provided in Chapter 2 (Figure 2.5) show that the orientation of duplicated genes with respect to the centromeres appear to be conserved. Another amplification mechanism, by gene conversion, must therefore exist which is probably restricted to the subtelomeric regions.

In conclusion, the *S. cerevisiae* *THI5* gene family has been shown to be functionally redundant in terms of HMP formation from pyridoxine. Apart from the six-

membered *AAD* family, no other subtelomeric gene family has been extensively studied in terms of its functional role and the functional redundancy of its members. With the increasing availability of strains and technologies that facilitate the deletion of multiple genes in *S. cerevisiae*, other gene families such as *PAU* (23 members) and *COS* (22 members) can be examined in this way.

Chapter 8

Materials and Methods.

8.1 Growth media and conditions.

Solid media contained 2% (w/v) Difco purified agar. All growth media were sterilised by autoclaving at 21°C, 15 psi. for 20 minutes. Where necessary, the addition of defined amounts of vitamins (2 µm filter sterilised) were added to the media after sterilisation. For studies requiring anaerobic conditions, all media were further supplemented with 0.9 ml/l tween 80 and 30 mg/l ergosterol (C. Grant, personal communication). *E. coli* strains were incubated at 37°C and yeast strains at 28°C. Anaerobic incubation of yeast on solid media was achieved by placing the agar plates into a BBL GasPak[®] chamber containing an Oxoid gas generating sachet (Unipath Ltd). Confirmation of anaerobic conditions was determined by the absence of growth of *S. cerevisiae* strain S288C on YP plates with glycerol (3% v/v) as sole carbon source.

8.1.1 Luria broth (LB).

The propagation of *E. coli* strains used complex LB medium consisting of 1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract and 0.5% (w/v) NaCl, pH adjusted to 7.2 with NaOH. Selection for ampicillin-resistant *E. coli* strains required the addition of ampicillin to the LB medium to a final concentration of 50 µg/ml.

8.1.2 Yeast peptone (YP) medium.

When no nutritional selection was required, all yeast strains were grown on complex YP medium. This consisted of 1% (w/v) yeast extract and 2% (w/v) bacto-peptone. For all yeast strains except *S. pastorianus* and *S. bayanus*, glucose was added as carbon source to a final concentration of 2% (w/v) (YPD). For *S. pastorianus* and *S. bayanus*, glucose was added to a final concentration of 10% (w/v).

8.1.3 Synthetic dextrose (SD) medium.

Nutritional selection of *S. cerevisiae* strains used SD minimal medium (0.67% (w/v) Difco Yeast Nitrogen Base without amino acids, 2% (w/v) glucose supplemented with the appropriate SCM drop-out powder mix. For nutritional selection along with positive selection for G418 resistance (conferred by the *KanMX4* module introduced into RWY16), SD medium cannot be used due to the high salt

concentration which affects the G418 selection (A. Wach, personal communication). YPD medium containing 300 mg/l G418 was used instead, followed by nutritional selection on SD.

8.1.4 Wickerham's minimal medium.

For phenotypic studies of yeast with regards to their thiamin or pyridoxine requirements, the following Wickerham's minimal medium was used (Wickerham, 1951):

Salts mixture [KH_2PO_4 (1g/l), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (0.5g/l), NaCl (0.5g/l), $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$ (0.5g/l)], $(\text{NH}_4)_2\text{SO}_4$ (2.5g/l) as a nitrogen source, trace elements [H_3BO_3 (8 μM), $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$ (2 μM), $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ (1 μM), $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ (1 μM), $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ (1 μM), KI (1 μM), $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0.1 μM)], and vitamins [nicotinic acid (65 μM), pantothenic acid (25 μM), inositol (110 μM), biotin (1 μM), *p*-aminobenzoic acid (4 μM), riboflavin (2 μM)]. The full SCM powder mix was also added except for the large-scale batch fermentations where the individual amino acids adenine (20 mg/l), lysine (30 mg/l) and methionine (20 mg/l) were present.

Where indicated, thiamin, HMP or HET was added to a final concentration of 2 μM and pyridoxine to 9 μM . The carbon source was glucose (2% w/v), galactose (2% w/v) or ethanol (3% v/v). Unless otherwise indicated, the carbon source used was always glucose.

8.1.5 Blackstrap molasses agar medium.

Molasses medium contained 100 g/l Blackstrap molasses, 2 g/l $(\text{NH}_4)_2\text{SO}_4$, 2 g/l KH_2PO_4 and 0.6 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 2 g/l SCM powder (Praekelt and Meacock, 1992).

8.1.6 The synthetic complete powder mix (SCM).

The full SCM powder mix contained the following amino acids and bases in equal ratio; alanine, arginine asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, inositol, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, uracil and valine. Adenine was added at 1/4 ratio, leucine at 2x ratio and *p*-aminobenzoic acid at 1/10 ratio. The mix was used at 2 g/l in solid or liquid medium. For liquid medium, the mix was dissolved in 10 ml water and filter sterilised. For prototrophic selection of yeast, the powder mix was made up as before omitting the relevant nutrient(s). This mix was called SCM drop-out powder (Rose, 1990).

8.2 Bacterial and Yeast strains.

The *E. coli* and yeast strains used in this study are listed along with their genotypes (where appropriate) in Table 8.1. The RWY series of *S. cerevisiae* strains are listed in Chapter 5, Table 5.1.

8.2.1 *E.coli* strain storage

10ml LB containing 50µg/ml ampicillin (if appropriate) was inoculated with a single bacterial colony and incubated overnight. 1ml of this culture was aliquoted into a cryogenic tube containing 1ml 50% (v/v) glycerol, the contents mixed, and frozen in a dry ice/IMS bath for 5 min. Tubes were then transferred to a -80°C freezer for permanent storage. Strain revival was carried out by inoculating LB agar with a loopful of the cell suspension and incubation at 37°C overnight.

8.2.2 *S. cerevisiae* strain storage

10ml YPD was inoculated with a single yeast colony and incubated overnight. 3 ml of this culture was aliquoted into a cryogenic tube containing 1 ml glycerol. Tubes were frozen in a dry ice/IMS bath for 5 min and transferred to a -80°C freezer for permanent storage. Strain revival was carried out by rapidly thawing the tube at 37°C for 2 min, inoculating YPD agar with a loopful of the cell suspension, and incubation at 28°C overnight.

8.3 Plasmids, vectors and genomic libraries.

All plasmids used in this study are listed in Table 8.2.

8.3.1 The *S. kluyveri* genomic library.

The *S. kluyveri* library was obtained from M. Costanzo, Cornell University and contains 15-20 Kb partial *Sau3AI* fragments of genomic DNA from *S. kluyveri* CBS 3082. These fragments have been cloned into the *Bam*HI site of plasmid YEp352 (*URA3, Amp^r*). The library is composed of about 40,000 *E. coli* transformants and approximately 70% contain inserts (Costanzo *et al.*, 2000). It is reported that the M13 reverse primer does not anneal to the library vector (T. Fox, personal communication). Other reports suggest that,

Strain	Genotype	Reference/Source
<i>E. coli</i> DH5α	φ80 <i>lacZ</i> Δ <i>M15</i> , <i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi-1</i> , <i>hsd R17</i> (<i>r_K⁻</i> , <i>m_K⁺</i>), <i>supE44</i> , <i>relA1</i> , <i>deoR</i> ,Δ(<i>lacZYA-argF</i>)U169	(Hanahan, 1983)
<i>S. cerevisiae</i>		
S288Ca	<i>MATa</i> , <i>mal</i> , <i>gal2</i> , [<i>cir+</i>]	(Johnston & Mortimer, 1984)
BY4705	<i>MATα</i> , <i>ade2Δ::hisG</i> , <i>his3Δ200</i> , <i>leu2Δ0</i> , <i>lys2Δ0</i> , <i>met15Δ0</i> , <i>trp1Δ63</i> , <i>ura3Δ0</i>	Brachmann <i>et al.</i> (1998)
W303a	<i>MATa</i> , <i>ade2-1</i> , <i>can1-100</i> , <i>leu2-3-112</i> , <i>trp1-1</i> , <i>ura3-1</i> , <i>his3-11-15</i>	R. Rothstein, Columbia University
MW554 (DS10)		M. Werner-Washburne, University of New Mexico
MW1076 (σ1278b derivative)	<i>MATa</i> , <i>ura3-52</i> , <i>leu2::hisG</i> , <i>his3::hisG</i> , <i>trp1::hisG</i>	M. Werner-Washburne, University of New Mexico
<i>K. lactis</i> MSK-110	<i>UraA</i> , <i>trp1::URA3</i>	(Stark and Milner, 1989)
<i>K. lactis</i> MSK-110 <i>klthi5</i>	<i>UraA</i> , <i>trp1::URA3</i> , <i>thi5::TRP1</i>	D. Walsh, University of Leicester

Strain	Reference/Source
<i>S. cerevisiae</i> Type	CBS 1171
<i>S. cerevisiae</i> Lager (formerly <i>S. uvarum</i>)	NCYC 1324
<i>S. cerevisiae</i> Ale (formerly <i>S. uvarum</i>)	NCYC 1681
<i>S. cerevisiae</i> Fullers	Bottled ale from retail outlet – ‘Fullers 1845’
<i>S. cerevisiae</i> Caledonian	Bottled ale from retail outlet -‘Caledonian Deuchars’
<i>S. cerevisiae</i> EM93	NCYC 1324
<i>S. cerevisiae</i> American Yeast Foam	NCYC 232
<i>S. bayanus</i>	CBS 380
<i>S. pastorianus</i> (formerly <i>S. carlsbergensis</i>)	CBS 1538
<i>S. paradoxus</i> N12	E. Louis, University of Leicester
<i>S. paradoxus</i> N17	E. Louis, University of Leicester
<i>S. cariocanus</i>	Naumov <i>et al.</i> (2000)
<i>S. kudriavzevii</i>	Naumov <i>et al.</i> (2000)
<i>S. mikatae</i>	Naumov <i>et al.</i> (2000)
<i>S. servazzi</i> ,	CBS 4311
<i>S. castelli</i>	CBS 4309
<i>S. unisporus</i>	CBS 398
<i>S. kluyveri</i>	CBS 3082
<i>S. exiguus</i>	CBS 379
<i>K. drosophilarum</i>	CBS 2896
<i>K. polysporus</i>	CBS 2163
<i>K. thermotolerans</i>	CBS 6924

Table 8.1 Bacterial and yeast strains used in this study. NCYC (National Collection of Yeast Cultures), CBS (Centraalbureau voor Schimmelcultures).

Plasmid	Genotype	Reference/Source
pRH4	<i>ori, Amp^r, THI5</i>	R.J. Hather (1996)
pUP27.1	<i>ori, Amp^r, THI5 cDNA</i>	U. Praekelt, University of Leicester
pUP9	<i>ori, Amp^r, THI4 cDNA</i>	U. Praekelt, University of Leicester
pBS-Actin	<i>ori, Amp^r, ACT1</i>	T Pillar, University of Leicester
pDW5a	<i>ori, Amp^r, Kl.THI5 (C-terminal portion)</i>	D. Walsh, University of Leicester
pDW5b	<i>ori, Amp^r, Kl.THI5 (N-terminal portion)</i>	D. Walsh, University of Leicester
pUC19	<i>ori, Amp^r, lacZ</i>	{Yanisch-Perron, 1985 #609}
pRW3a	<i>ori, Amp^r, 2μ ori, LEU2, (carrying a genomic clone of <i>S. kluyveri</i> THI5)</i>	This study
pRW1	1 Kb <i>Eco</i> R1 fragment of pRW3a carried on pUC19	This study
pRW1.1	1.1 Kb <i>Eco</i> R1 fragment of pRW3a carried on pUC19	This study
pRW2.8	2.8 Kb <i>Eco</i> R1 fragment of pRW3a carried on pUC19	This study
pRS400	<i>ori, Amp^r, lacZ, KanMX4</i>	Brachmann <i>et al.</i> (1998)
pRS413	<i>ori, Amp^r, lacZ, CEN6, ARS4, HIS3</i>	Brachmann <i>et al.</i> (1998)
pRS414	<i>ori, Amp^r, lacZ, CEN6, ARS4, TRP1</i>	Brachmann <i>et al.</i> (1998)
pRS415	<i>ori, Amp^r, lacZ, CEN6, ARS4, LEU2</i>	Brachmann <i>et al.</i> (1998)
pRS416	<i>ori, Amp^r, lacZ, CEN6, ARS4, URA3</i>	Brachmann <i>et al.</i> (1998)
pUP34	<i>ori, Amp^r, CEN4, ARS1, URA3, lacZ</i> (lacking its promoter)	U. Praekelt, University of Leicester
pHL1	<i>THI12</i> promoter carried on pUC19	This study
pHL2	<i>THI5</i> promoter carried on pUC19	This study
pHL3	<i>THI11/13</i> promoter carried on pUC19	This study
pHL4	<i>ori, Amp^r, CEN4, ARS1, URA3, lacZ</i> (<i>THI5-lacZ</i> fusion)	This study
pHL5	<i>ori, Amp^r, CEN4, ARS1, URA3, lacZ</i> (<i>THI12-lacZ</i> fusion)	This study
pHL6	<i>ori, Amp^r, CEN4, ARS1, URA3, lacZ</i> (<i>THI11/13-lacZ</i> fusion)	This study
pRH11	<i>ori, Amp^r, THI5</i>	R.J. Hather (1996)
pSK1.1	<i>ori, Amp^r, fl(+)</i> <i>ori, THI12</i>	(Vandyck, 1995)

Table 8.2 Bacterial and Yeast plasmids used in this study.

in reality, the library fragments average 6 Kb and approximately 30% of *E. coli* transformants contain inserts (Z. Gojkovic, personal communication).

8.4 Isolation of DNA and RNA.

8.4.1 Small scale *E. coli* plasmid preparations.

Plasmid DNA was isolated from 2ml culture of *E. coli* according to the protocol described by Serghini *et al.* (Serghini *et al.*, 1989). Cells were disrupted by vortexing in a phenol solution. This was followed by spinning to get rid of cell debris, precipitating the nucleic acids and then treating with RNase enzyme.

8.4.2 Medium scale, highly purified *E. coli* plasmid preparations.

Plasmid DNA was isolated using Qiagen™ plasmid mini-prep kits, according to the manufacturers instructions. DNA purified in this way was used as template for DNA sequence reactions.

8.4.3 Large scale, highly purified *E. coli* plasmid preparations.

Plasmid DNA was isolated using Qiagen™-tip 100 columns. The method was carried out according to the manufacturers instructions.

8.4.4 Small scale, rapid *S. cerevisiae* plasmid preparations.

Solutions:

BME Buffer (0.9M Sorbitol, 0.05M Na₂PO₄ (pH 7.5), 0.1% v/v β-mercaptoethanol (added just before use).

Procedure:

Yeast cultures were grown up overnight in 5 ml of the appropriate selective media from a single colony. This culture was pelleted at 3000 rpm, washed once with distilled water, and resuspended in 800 µl of BME buffer. 25 µl of 10 mg/ml yeast lytic enzyme was added, mixed by gentle inversion and incubated at 37°C for 30-45 minutes until spheroplasts formed. Samples were then switched to 70°C for 20 minutes, followed by the addition of 200µl 5M potassium acetate and incubation on ice for 45 minutes. The resulting precipitate was pelleted at 13000 rpm for 10 seconds and the supernatant transferred to a fresh tube. 0.55 ml of isopropanol was added, the tube mixed and left at room temperature for 5 minutes. Plasmid DNA was then

pelleted at 13000 rpm for 10 minutes, washed in 70% ethanol, allowed to dry and resuspended in 50 µl TE buffer.

8.4.5 Rapid preparation of chromosomal DNA from *S. cerevisiae*.

The following method was used to isolate chromosomal DNA from *S. cerevisiae* laboratory strains only.

Solutions:

Breaking buffer (2% (v/v) Triton X-100, 1% (v/v) SDS, 100 mM NaCl, 10 mM Tris-Cl (pH 8), 1 mM EDTA (pH 8)).

Procedure:

S. cerevisiae cultures were grown overnight in 10 ml YPD to stationary phase. The culture was pelleted at 3000 rpm and the pellets washed once with water. The pellet was resuspended in 200 µl breaking buffer. Approximately 200 µl glass beads was added to the suspension followed by 200 µl phenol/chloroform. The glass bead suspension was vigorously mixed for exactly 8 minutes in a multi-vortex apparatus. 200 µl TE buffer was added followed by vortexing for 30 seconds. After centrifugation at 13000 rpm, 5 minutes the upper aqueous layer was isolated (the lysate) and extracted with chloroform/iso-amyl alcohol. The lysate was then precipitated with 1 ml ethanol and pelleted 13000 rpm for 3 minutes. The pellet was resuspended in 0.4 ml TE buffer followed by the addition of 3 µl 10 mg/ml RNase and incubated for 1 hour, 37°C. 10 µl of 4M ammonium acetate and 1 ml of cold ethanol was then added, followed by gentle mixing and 4 minute incubation at 4°C. The suspension was pelleted 13000 rpm for 3 minutes and the pellet washed with 70% ethanol. The pellet was dried and resuspended in 100-200 µl TE buffer.

8.4.6 Large scale preparation of highly purified chromosomal DNA from the hemiascomycetes.

The following method (based on that of (Cryer *et al.*, 1975)) was used to isolate chromosomal DNA from the various yeast species used in this study.

Solutions:

A: 1.2M sorbitol, 25mM EDTA, pH8.0

B: 1.2M sorbitol, 0.1M Na citrate, 10mM EDTA, pH8.0

C: 3% sarkosyl, 0.5M Tris-HCl, 0.2M EDTA, pH7.6

Procedure:

Yeast cells were grown for 24 hours in 50 ml YPD and pelleted at 3000 rpm for 5 minutes. Cells were resuspended in 5ml A, 175 μ l 1M DTT added, incubated at 30°C for 30 minutes with shaking, repelleted (6000 rpm, 5 minutes) and resuspended in 5ml B. 100 μ l 10mg/ml YLE was added, samples mixed and incubated at 37°C for 1 hour with shaking until cells were spheroplasting. After spheroplasting, cells were washed three times in 5ml 1.2M sorbitol (pelleting each time at 5,000 rpm, 5 minutes) then resuspended in 3 ml C. 100 μ l 2mg/ml proteinase K (made up in C) was added and cells incubated at 37°C for 45 minutes followed by 55°C for 45 minutes. Volume was then made up to 5 ml with TE and the lysate extracted three times with phenol/chloroform and once with chloroform/iso-amyl alcohol until the interface was clear. Some yeast species required up to six phenol/chloroform extractions. Nucleic acids were precipitated in 2 volumes cold ethanol for 15 min on ice, and pelleted at 5,000 rpm, 5 minutes, 4°C. DNA/RNA was very gently resuspended in 500 μ l TE, 20 μ l 10mg/ml RNase added, followed by incubation at 37°C for 1-2 hours. The DNA was then reprecipitated in cold ethanol, washed twice in 70% cold ethanol, air dried and resuspended in 250-500 μ l TE.

8.4.7 Preparation of total RNA from *S. cerevisiae*.

Total yeast RNA was isolated using the rapid phenol/SDS extraction protocol described by Schmitt *et al.* (Schmitt *et al.*, 1990).

Solutions:

All water was treated with di-ethyl-pyrocabonate prior to use.

AE: 50mM Na acetate, 10mM EDTA, pH5.3

Procedure:

A 10-20 ml sample of yeast culture was harvested by centrifugation (5000 rpm, 5 minutes, 4°C) and resuspended in 1ml AE. To 400 μ l aliquots were added 80 μ l 10% SDS, and the solution vortexed hard for 30 seconds. An equal volume of AE equilibrated phenol was added and re-vortexed for 30 seconds. Samples were incubated at 65°C for 4 minutes, immediately chilled in dry ice/IMS for 3 minutes until phenol crystals appeared, and the heating-freezing cycle repeated twice more. Samples were then centrifuged at 13000 rpm for 5 minutes, and the aqueous layer was phenol extracted twice. RNA was ethanol precipitated, washed in 70% ethanol and the pellet resuspended in 50 μ l water.

8.5 DNA transformation of bacteria and yeast.

8.5.1 *E.coli* transformation.

Competent *E. coli* cells were prepared using the calcium chloride based method of (Mandel & Higa, 1970). 100 µl of competent cells were added to plasmid DNA (10 ng in 25 µl) or DNA ligation mixes, followed by incubation on ice for 30 minutes. Cells were then heat-shocked at 42°C for 2 minutes, the volume made up to 1ml with LB and re-incubated at 37°C for 1 hour. Cells were pelleted in a microfuge (13000 rpm, 1 minute) resuspended in 100ml LB, plated onto selective medium and incubated at 37°C overnight.

8.5.2 High efficiency transformation of *S. cerevisiae*.

For transformation of either plasmids or the PCR-synthesised disruption cassettes into *S. cerevisiae*, a lithium acetate ultra-high efficiency protocol based on Gietz (Geitz *et al.*, 1995) was used.

Solutions:

TE: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0)

TEL: 100 mM lithium acetate in TE (pH 7.5)

PEG: 40% w/v PEG3350 in TEL.

Procedure:

Yeast cells were grown overnight in 10ml YPD, diluted to a concentration of 2×10^6 cells/ml in 50 ml fresh YPD and grown to 1×10^7 cells/ml. Cells were harvested at 3000 rpm for 7 minutes, resuspended in 1 ml water and pelleted at 13000 rpm for 1 minute in a microfuge. 1ml of TEL was added, the cells pelleted (13000 rpm, 1 minute) and resuspended gently in TEL to achieve a density of 2×10^9 cells/ml. To 50 µl cell aliquots was added transforming DNA (1-5 µg of disruption cassette or <500 ng of plasmid DNA), 50 µl 1 mg/ml single-stranded salmon sperm DNA and 300 µl freshly made PEG. Samples were incubated at 30°C for 30 minutes, then heat-shocked at 42°C for 15 minutes. The cells were pelleted, washed in water to remove residual PEG and cells were plated onto selective medium and incubated at 30°C for 3-4 days.

8.6 Nucleic acid electrophoresis.

8.6.1 DNA electrophoresis.

DNA fragments were separated by agarose gel electrophoresis, either in 0.8% (w/v) agarose (for separation of genomic restriction digestions) or 1% (w/v) agarose (for separation of all other fragments e.g plasmid DNA digestions) in 1 x TAE buffer, containing 0.5 mg/ml ethidium bromide as described by Maniatis *et al.* (Maniatis *et al.*, 1982). The separated DNA products were visualised under UV light.

8.6.2 RNA electrophoresis.

RNA was analysed by horizontal gel electrophoresis on formaldehyde denaturing agarose gels (Lehrach *et al.*, 1977).

Solutions:

MOPS (x10): 200 mM MOPS, 83 mM Na acetate, 1 mM EDTA (pH7.0).

Loading buffer: 0.4% (w/v) bromophenol blue, 0.4% (w/v) xylene cyanol, 50% (v/v) glycerol, 1 mM EDTA

Procedure:

A 1.5% (w/v) agarose gel made in MOPS, containing 3% (v/v) formaldehyde, was immersed in MOPS running buffer immediately prior to electrophoresis. RNA samples were prepared in sterile microfuge tubes by mixing the following:

RNA (up to 20 µg)	5 µl
MOPS (x5)	2.5 µl
Deionised formamide	10 µl
Formaldehyde	3.5 ul

To denature any double stranded RNA, samples were heated at 65°C for 10 minutes and immediately cooled on ice. 2.5 µl loading buffer and 1 µl 1 mg/ml ethidium bromide was added to each sample followed by electrophoresis at 90-100V for 2-3 hours. RNA was then visualised on a UV transilluminator.

8.7 Southern-, northern-, and slot-blot hybridisations.

8.7.1 Southern blot hybridisation.

DNA was transferred from agarose gels to nylon membranes (Hybond N⁺) by capillary action using the method of Southern (Southern, 1975). After blotting, DNA

was immobilised on the membrane in an Amersham UV crosslinker (RPN 2500), using the preset energy setting of $70,000 \mu\text{j cm}^{-2}$.

Pre-hybridisation and hybridisation of the membrane with a labelled probe was carried out according to the Amersham Gene Images™ protocol at the temperatures specified in the Results chapters. To make the probe, fluorescein dUTP was incorporated into DNA in a random prime labelling reaction using the Gene Images% random prime labelling kit (Amersham Pharmacia Biotech). Probes were used immediately without purification or stored at -20°C for up to three months in the dark. After hybridisation, membranes were washed (at the same temperature used during hybridisation) once in $1 \times \text{SSC}$, $0.1\% \text{ w/v SDS}$ and then four times in $0.5 \times \text{SSC}$, $0.1\% \text{ w/v SDS}$. Where indicated, lower temperatures were used for the washes.

Detection of filter-bound probe used a chemilluminescence method as featured and described in the CDP-Star™ detection module (Amersham Pharmacia Biotech). Filters were then sealed in clear plastic bags and exposed to X-ray film for 1 to 10 minutes at room temp.

8.7.2 Northern blot hybridisation.

Solutions:

SSC ($\times 10$): 1.5M NaCl , 0.15M Na citrate ($\text{pH} 7.0$)

Procedure:

Following electrophoresis, RNA gels required no treatment prior to blotting. The RNA was transferred onto the membrane overnight by capillary action from a reservoir of $10 \times \text{SSC}$ (Thomas, 1980). RNA was fixed on the membrane in an Amersham UV crosslinker (RPN 2500) using the preset energy setting of $70,000 \mu\text{j cm}^{-2}$.

8.7.3 Slot blot hybridisation.

Slot-blot hybridisation was used to provide rapid detection of RNA through the direct transfer of the sample to narrow 'slots' on Hybond-N⁺ membrane. Slot blot membranes were prepared using a 24-well microfiltration unit. A piece of Hybond-N⁺ and an equal size of 3MM paper (cut to the size of the apparatus) were soaked in $10 \times \text{SSC}$. The Hybond-N⁺ was placed on top of the 3MM paper and clamped into the centre of the microfiltration unit with the Hybond-N⁺ on top and touching the wells. Each well was rinsed with $100 \mu\text{l } 20 \times \text{SSC}$ and drawn through under vacuum, followed by $8 \mu\text{g RNA}$ in $400 \mu\text{l } 20 \times \text{SSC}$ drawn through under minimum vacuum.

Each well was then rinsed twice with 100 μ l 20 x SSC. RNA was fixed on the membrane in an Amersham UV crosslinker (RPN 2500) using the preset energy setting of 70,000 μ j cm⁻².

8.7.4 Preparation of high specific activity probes for northern- and slot-blot hybridisation.

DNA probes were radiolabelled using random hexanucleotide primers, based on the method of (Feinberg & Vogelstein, 1983).

Solutions:

O: 1.25M Tris-HCl, 0.125M MgCl₂ (pH8.0)

A: 1ml O, 18 μ l β -Mercaptoethanol, 5 μ l each of 100mM dATP, dTTP, dGTP

B: 2M HEPES-NaOH (pH6.6)

C: Hexanucleotides at 90 A₂₆₀/ml in TE (Pharmacia)

OLB: Mixture of A:B:C in the ratio 10:25:15

Procedure:

20–40ng of the DNA fragment being radiolabelled (in 15 μ l water), was denatured at 100°C for 5 minutes then immediately cooled on ice. To this was added 5 μ l OLB, 1 μ l (1unit) Klenow enzyme (Pharmacia), 1 μ l 10mg/ml BSA, and 2.5 μ l ³²P-dCTP, the tube contents mixed and incubated at room temperature for 4-7 hours. Unincorporated nucleotides were removed by passing the reaction mixture through a Sephadex G50 (pre-swollen with TE) column constructed in a Pasteur pipette plugged with polyallomer wool that was equilibrated with 100 μ l aliquots of TE. The crude radiolabelled probe mix was applied to the column and eluted with further 100 μ l aliquots of TE. Fractions of 100 μ l were collected from the bottom of the column in Eppendorf tubes, and the passage of probe through column monitored using a Geiger counter. A sharp rise in the radioactivity levels of fractions was usually observed after elution of 500 μ l, with the following five fractions (containing the highest cpm), being pooled for use in hybridisations. Prior to hybridisation, the radiolabelled probe was denatured at 96°C for 5 min.

8.7.5 Pre-hybridisation and hybridisation of RNA-bound membranes.

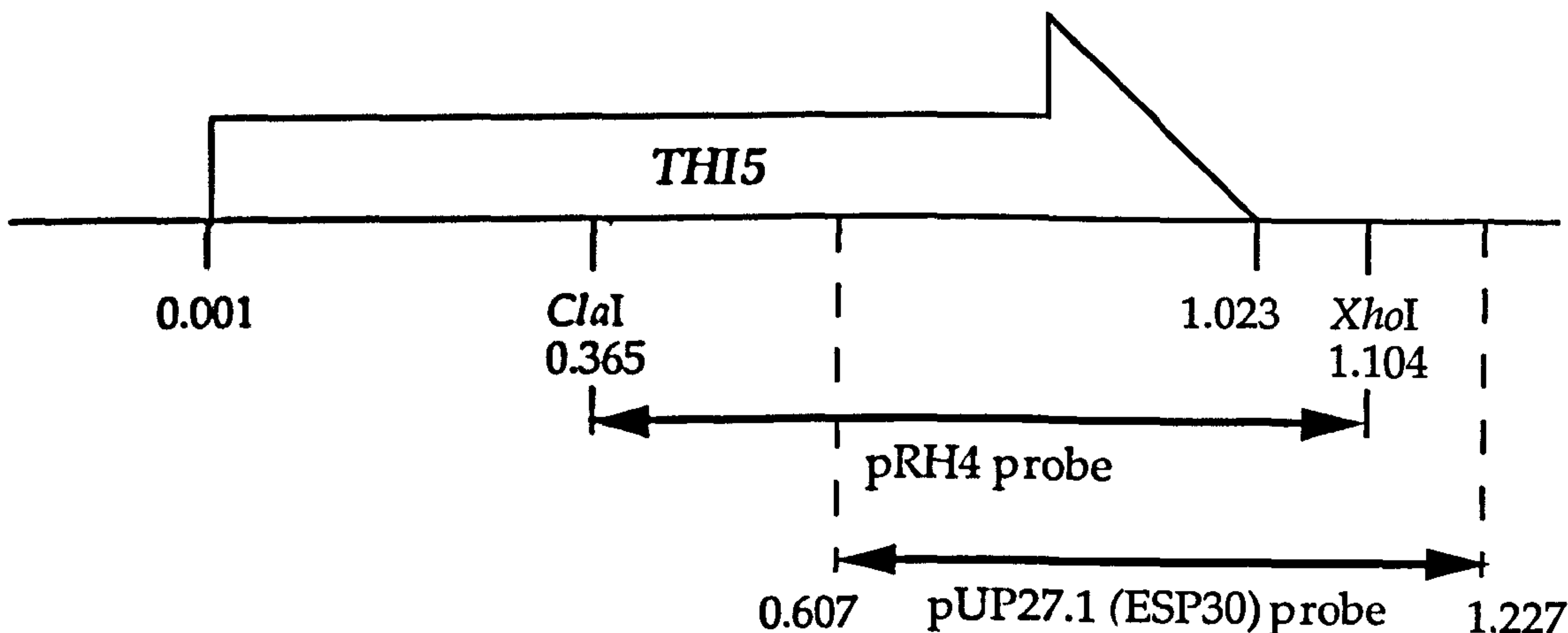
Filters were placed in glass tubes containing 20ml Church-Gilbert buffer (0.5M NaHPO₄, 7% w/v SDS, 1mM EDTA, pH7.4) (Church & Gilbert, 1984), and incubated at 65°C with rotation for 4-7 hours. Fractionated radiolabelled probe was then added to the filter (still in Church-Gilbert buffer) and re-incubated at 65°C overnight. Filters were washed four times in 20 ml 3xSSC, 0.1% w/v SDS at 65°C, partially dried by blotting on 3MM paper and wrapped in Saran wrap. Autoradiography was carried out at -80°C using X-ray film in an X-ray cassette. The relative radioactivity

of any bands or slots was measured by analysis on a Phosphorimager™ apparatus (see Section 8.13.3).

8.7.6 DNA probes used in this study.

S. cerevisiae *THI5* - 620bp *THI5* *Eco*RI fragment of pUP27.1(ESP30) or 677bp *Cla*I/*Xho*I fragment of pRH4.

S. cerevisiae *THI4* - 1.1 Kb *Eco*RI fragment of pUP9.



S. cerevisiae Actin - 342 bp *ACT1* *Eco*RI/*Hind*III fragment of pBS-Actin.

K. lactis *THI5* - 800 bp *Hind*III fragment of pDW5a.

K. lactis COS - 485bp *Eco*RV fragment of pDW5a.

8.8 DNA manipulations.

8.8.1 DNA restriction endonuclease digestion.

Restriction enzymes were supplied by Gibco-BRL or New England Biolabs with their appropriate reaction buffers and were stored at -20°C. Enzyme digestions were carried out according to the manufacturers' instructions.

8.8.2 DNA elution for agarose gels.

DNA fragments were recovered from agarose gels using the Qiagen™ gel extraction kit, according to the manufacturer's instructions.

8.8.3 DNA ligation.

Ligation reactions used the T4 DNA ligase enzyme supplied by Gibco-BRL (1U/ml). Generally, reactions were carried out in a total volume of 10 µl, containing 1 µl ligase, 2 µl T4 ligation buffer, plus DNA samples being ligated. These were incubated overnight at room temperature. The amount of DNA used to obtain the optimal vector:insert ratio was calculated based on the relative lengths of the two DNA fragments, according to Dugaiczyk *et al.* (1975).

8.9 Methods using the Polymerase Chain Reaction (PCR).

PCR was used for the amplification of DNA fragments and was performed using a Hybaid™ Omn-E thermal cycler. All primers were stored as stock solutions of 10 µM.

Solutions:

11.1X PCR buffer (500 mM Tris-Cl pH 8.8, 122 mM ammonium sulphate, 50 mM magnesium sulphate, 75 mM β-mercaptoethanol, 50 µM EDTA pH 8.0, 11.1 mM dATP, 11.1 mM dCTP, 11.1 mM dGTP, 11.1 mM dTTP, 1.26 mg/ml BSA).
Distilled water (dH₂O)

8.9.1 Reaction conditions for American Yeast Foam (Chapter 3).

A primer mix was made consisting of primers THI5fwd:THI12:THI11:THI5:THI13 mixed together in the ratio 6:2:2:3:2. The contents of the reactions and the PCR program are described below.

dH ₂ O=31.5 µl	94°C – 2 minutes
11.1X PCR=4.5 µl	<u>94°C – 1 minute }</u>
primer mix=3 µl	<u>55°C – 1 minute }</u> 35 cycles
genomic DNA=200 ng in 10 µl	<u>72°C – 1 minute 45 seconds }</u>
Taq DNA polymerase=1 µl	72°C – 10 minutes

8.9.2 Reaction conditions for degenerate PCR analysis (Chapter 4).

The primers were designed to anneal to conserved portions of the SNZ, THI5, and AAD genes in *S. cerevisiae*. As the SNZ gene family has two highly conserved subtelomeric members, the primer was designed to anneal to conserved portions between *S. cerevisiae* SNZ and *S. mikatae* SNZ (sequence kindly donated by P. Clifton, University of Washington). Further details of these primers are described in section 8.9.5.

dH ₂ O=32.5 µl	94°C – 2 minutes
SNZfwd=1 µl (10 µM stock)	<u>52°C – 1 minute</u> } <u>35 cycles</u>
THI5back=1 µl (10 µM stock)	<u>72°C – 55 seconds</u> }
genomic DNA=200 ng in 10 µl	72°C – 10 minutes
Taq DNA polymerase=1 µl	

The same conditions were used for amplification of the *THI5*-*AAD* intergenic region which used primers THI5fwd and AADfwd.

8.9.3 Reaction conditions for synthesis of disruption cassettes.

Linear DNA cassettes, for use in the targetted disruption of genes belonging to the *THI5* family, were prepared using PCR as described below. Cassettes containing different nutritional markers were synthesised using the same primers with the appropriate plasmid of the pRS4XX series of plasmids (Brachmann *et al.*, 1998)

dH₂O=up to 100 µl
 11.1X PCR buffer=9 µl
 THI5up primer=2 µl (10 µM stock)
 THI5down primer=2 µl (10 µM stock)
 pRS4XX plasmid template= 100 ng
 Taq DNA polymerase= 1 µl

94°C – 2 minutes
94°C – 1 minute }
55°C – 1 minute } 10 cycles
72°C – 1 minute per Kb of amplification, minimum 1 minute
94°C – 1 minute
65°C – 1 minute } 20 cycles
72°C – 1 minute per Kb of amplification, minimum 1 min.
 72°C – 1 minute

8.9.4 Reaction conditions for amplification of promoters.

Promoter regions of *THI5*, *THI11*, *THI12* and *THI13* were amplified prior to restriction digestion and subsequent cloning as follows;

dH₂O=up to 100 µl
 11.1X PCR buffer=9 µl
 THI5start=2 µl (10 µM stock)

RB5 or THI11prom=2 µl (10 µM stock)
 plasmid for *THI5/THI12*(pRH11 or pSK1.1)
 or genomic (for *THI11/13*) DNA template= 100 ng
 Taq DNA polymerase= 1 µl

For amplification of <i>THI12</i> promoter:	of <i>THI11/13/5</i> promoters:
95°C - 5 minutes	95°C - 5 minutes
<u>95°C - 30 seconds</u>	<u>95°C - 1 minute</u>
<u>57°C - 30 seconds</u>	<u>57°C – 1 minute</u>
<u>72°C – 30 seconds</u>	<u>72°C – 30 seconds</u>
72°C – 10 minutes	72°C – 10 minutes
<u>35 cycles</u>	

8.9.5 Oligonucleotide primers.

The oligonucleotides used for PCR and DNA sequencing are listed in Table 8.3.

8.10 DNA sequencing.

DNA sequencing was performed on an ABI model 377XL DNA sequencer (PNACL, University of Leicester) following sample preparation using the BigDye™ Ready Reaction Terminator Cycle Sequencing Kit (PE Biosystems). The reaction was prepared according to the manufacturer’s recommendations, except that 1 µg template DNA and 30 pmol primer were used in all cases. The cycling reaction was carried out in a Hybaid Omn-E thermal cycler, and consisted of 25 cycles of a denaturation step (96°C, 30 seconds), an annealing incubation (50°C, 15 seconds) and an extension incubation (60°C, 4 minutes). Extension products were purified using the ethanol/sodium acetate precipitation procedure as described in the manufacturer’s protocol.

8.11 Assay of β-galactosidase activity in yeast.

β-galactosidase assays were carried out using the method of Reynolds (Reynolds, 1989).

Solutions:

Z Buffer: 60mM Na₂HPO₄, 40mM NaH₂PO₄, 10mM KCl, 1mM MgSO₄, 50mM β-mercaptoethanol (added just before use), pH7.0
 ONPG: 4mg/ml σ-nitrophenol-β-D-galactoside in 100mM KPO₄, pH7.0

Procedure:

Oligonucleotides	Sequence (5'.....3')	Target sequence
THI5fwd	ACCAAAGATGTTACGCTTAC (+773) (+792)	THI5/11/12/13 reading frames
THI5	GACTTICTGGTTTGCCACCGCTCC (+2965) (+2941)	Unique THI5 3'-flanking sequence
THI12	CGCAGTTGCCTTATCCACCGAGTGTGTGC (+1234) (+1205)	Unique THI12 3'-flanking sequence
THI11	CGGATTATAAGGGGTATGATGTAGGC (+2156) (+2130)	Unique THI11 3'-flanking sequence
THI13	GGCCAGCTGACGAATTGACTGAGCGGC (+2602) (+2575)	Unique THI13 3'-flanking sequence
AADfwd	GGAACATGGCACTGARTCTG	Conserved AAD gene sequence
SNZfwd	AAACCCATGTCGACCTTACC	Conserved SNZ gene sequence
THI5back	GATGGGTGGTATYTCCATCC	THI5/11/12/13 reading frames
THI5up	ATGTCTACAGACAAGATCACATTTTGTGA ACTGGCAACCAACCCAGATTGTACTGAGA GTGCAC	Marker gene flanking sequence of pRS41X series of plasmids
THI5down	AGCTGGAAGAGCCAATCTCTTGAAAGTACC TTCCTGTCTGCATTTTCTGTGCGGTATTCA CACCG	Marker gene flanking sequence of pRS41X series of plasmids
UPSTHI5	GGCGCGTTA(AG)TGCATAAAGACAGC (-351) (-328)	THI5/11/12/13 promoter
TRP1	CAATGCCCTCCCTCTTGGCC	TRP1 disruption cassette
HIS3	GCTTCTCTTATGGCAACCGC	HIS3 disruption cassette
LEU2	CAGATGAGGCGCTGGAAGCC	LEU2 disruption cassette
URA3	TACCTGGGCCCACCACACCG	URA3 disruption cassette
P3	GCTTCGCGCCGTGCGGCCATC	KanMX4 disruption cassette
THI5start	CTTGTCTGGATCCATAGCTTGG	(Chapter 6, Figure 6.8)
THI11prom	GGGGAATTCGTCCATTGCGCATGTCC	(Chapter 6, Figure 6.8)
RB5	GGGGAATTCGTAGAACTAGCGATGCTC	(Chapter 6, Figure 6.8)
Sequencing		
<i>S. kluyveri</i>		
fwd	M13 forward primer NEBiolabs #1212	
rev	M13 reverse primer NEBiolabs #1201	
2.8rr	GATGCACAGGGTCTACTAGG	
2.8ff	CTTTGATCACTACTGCTGCG	
1ff	GTCTCCGCCATCTAAGCCAG	
1.1ff	ATCTGACGGCGTAGAACGTG	
2.8eco	ATCAGACTGTTGATATAACG	
2.8gap	ACACCGTCTTGAAATGTCTG	
skup	CACGTTCTACGCCGTCAGAT	
skdown	GTTGAAACCACCACAGTCTC	
<i>K. lactis</i>		
kcos	CTTTATGAAGCAGGCAGCTG	
kmol2up	CCGAACTTTCAGCAGACTTTGC	
kmol2dn	CATACCGCACAGTCATCACC	
kmolcup	TGCGTCCAATTTACCGCTCG	
kmolcdn	TGTTCAACCTACTCGTCAAC	
kmolddn	GATGTGCATGAGTTGCTTCC	
kmolbup	GGCTTCATGTGAAAGTTGACTG	
kmolbdn	CACGGTTGGGTTAGAAGTTC	

Table 8.3 Oligonucleotides used in this study.

Yeast strains containing promoter-*lacZ* reporter gene constructs were grown for 24 hours in 5 ml Wickerham's minimal medium supplemented as indicated with thiamin, HMP or HET. Cultures were inoculated into fresh, identical medium at a final optical density of OD₆₀₀=0.1 and allowed to grow until a cell density of OD₆₀₀=0.3 was reached. The cultures were pelleted in a centrifuge (5000 rpm for 5 minutes), the cells washed in water and resuspended in 1ml Z Buffer. At this point OD₆₀₀ values of the suspensions were recorded. 50-100 µl of this cell suspension was transferred to a fresh tube, the volume made up to 0.9 ml with Z buffer, followed by addition of 50 µl chloroform and 10 µl 0.1% SDS. Samples were vortexed for 15 seconds using a multi-tube vortex adapter and incubated at 30°C for 15 minutes to equilibrate. 180 µl ONPG was added to each tube, mixed by inversion, re-incubated at 30°C and timing begun. Reactions were stopped by the addition of 450 µl 1M Na₂CO₃ when the solutions had turned yellow and the time noted. Samples were centrifuged (13000 rpm for 5 minutes) and the OD₄₂₀ of the supernatant measured compared to a Z Buffer/ONPG blank. β-galactosidase activity values were calculated using the following equation:

$$\text{UNITS} = 1000 \times \text{OD}_{420} / T \times V \times \text{OD}_{600}$$

where T = reaction time at 30°C (minutes)

V = volume of cells used (ml)

8.12 Large-scale aerobic and anaerobic batch fermentations.

The large-scale batch fermentations were carried out in an LH fermenter model 502D equipped with a temperature control set at 28°C. The medium (except vitamins) was autoclaved inside the sealed fermenter vessel. 100 µl Antifoam A was added to prevent a foam developing during the batch cultivation. Vitamins were then added at the beginning of the fermentation along with the yeast inoculant by injection using a syringe placed in a sterile injector port on the apparatus. Samples were taken using a specialised port that partially filled a 50 ml sterile glass tube when vacuum was applied via a syringe attached to an air filter. Agitation of the vessel contents was maintained by an impeller drive set at 1000 rpm for aerobic cultures and 400 rpm for anaerobic cultures. For maintenance of fully saturated aerobic conditions, compressed air (0.5 l/min) was passed through two filters into the vessel just beneath the impeller drive unit. For maintenance of anaerobic conditions, the air-line was attached to a source of compressed oxygen-free nitrogen. For all fermentations the vessel contents were sparged with either air or nitrogen for 2 hours prior to inoculation.

8.13 Methods using computer-based analyses.

8.13.1 Computer-generated alignments.

Alignments of both DNA and peptide sequences were created using the PILEUP application of the Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. Default settings were used unless otherwise stated. Block shading of the alignments was carried out using the BOXSHADE WWW server (http://www.ch.embnet.org/software/BOX_form.html). The entire subtelomeric chromosomal sequences of chromosomes IV-L, VI-L, X-R and XIV-L were obtained from MIPS (<http://www.mips.biochem.mpg.de/proj/yeast/>).

Chromosomal alignments of 2-3 Kb segments were prepared using the PILEUP and BESTFIT applications of the Wisconsin Package using default settings. Determination of the boundaries of sequence similarities was performed visually.

8.13.2 Contig construction of DNA sequence data.

Overlapping DNA sequence data were joined together to create a single continuous sequence using the FRAGMENT ASSEMBLY tools of the Wisconsin Package. Open reading frames were identified using GeneJockey II™ (Biosoft).

8.13.3 Measuring intensities of bands on RNA northern- and slot- blots.

The relative emitted radioactivity of positively hybridising bands were measured using the Molecular Dynamics Phosphorimager™ apparatus, according to the manufacturer's instructions. Analysis of the intensities used the Imagequant™ software. Northern blot intensities were measured as the area of each peak emission whilst slot blot intensities were measured as the volume of the peak.

References

- Azumi, M. & Goto-Yamamoto, N. (2001). AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces sensu stricto* and its application to phenetic clustering. *Yeast* 18, 1145-54.
- Bakker, B. M., Bro, C., Kotter, P., Luttik, M. A., van Dijken, J. P. & Pronk, J. T. (2000). The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J Bacteriol* 182, 4730-7.
- Baleiras Couto, M. M., Eijssma, B., Hofstra, H., Huis in't Veld, J. H. & van der Vossen, J. M. (1996). Evaluation of molecular typing techniques to assign genetic diversity among *Saccharomyces cerevisiae* strains. *Appl Environ Microbiol* 62, 41-6.
- Begley, T. P. (1996). The biosynthesis and degradation of thiamin (Vitamin-B1). *Nat Prod Rep* 13, 177-185.
- Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J. & Xi, J. (1999). Thiamin biosynthesis in prokaryotes. *Arch Microbiol* 171, 293-300.
- Blandin, G., Durrens, P., Tekaia, F., Aigle, M., Bolotin-Fukuhara, M., Bon, E., Casaregola, S., de Montigny, J., Gaillardin, C., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Perrin, A., Potier, S., Souciet, J., Talla, E., Toffano-Nioche, C., Wesolowski-Louvel, M., Marck, C. & Dujon, B. (2000). Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Letters* 487, 31-6.
- Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F., Duchateau-Nguyen, G., Lemaire, M., Marmeisse, R., Montrocher, R., Robert, C., Termier, M., Wincker,

- P. & Wesolowski-Louvel, M. (2000). Genomic exploration of the hemiascomycetous yeasts: 11. *Kluyveromyces lactis*. *FEBS Letters* 487, 66-70.
- Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J. C., Hieter, P. & Boeke, J. D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115-132.
- Burrows, R. (1997). Regulation of thiamine biosynthesis in *Saccharomyces cerevisiae*. *PhD. Thesis*. University of Leicester.
- Burrows, R. J., Byrne, K. L. & Meacock, P. A. (2000). Isolation and characterization of *Saccharomyces cerevisiae* mutants with derepressed thiamine gene expression. *Yeast* 16, 1497-508.
- Byrne, K. (1998). Genetic analysis of thiamine metabolism in *Saccharomyces cerevisiae*. *PhD. Thesis*. University of Leicester.
- Carter, A. T., Beiche, F., Hove-Jensen, B., Narbad, A., Barker, P. J., Schweizer, L. M. & Schweizer, M. (1997). *PRS1* is a key member of the gene family encoding phosphoribosylpyrophosphate synthetase in *Saccharomyces cerevisiae*. *Mol Gen Genet* 254, 148-56.
- Cary, J. W. & Bhatnagar, D. (1995). Nucleotide sequence of a *Aspergillus parasiticus* gene strongly repressed by thiamine. *Biochim et Biophysic Acta* 1261, 319-320.
- Cavalieri, D., Barberio, C., Casalone, E., Pinzauti, F., Sebastiani, F., Mortimer, R. & Polsinelli, M. (1998). Genetic and molecular diversity in *Saccharomyces cerevisiae* natural populations. *Food Technol and Biotechnol* 36, 45-50.

Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proc Natl Acad Sci U S A* 81, 1991-5.

Corti, P. (1997). Functional analysis of a thiamine-regulated gene family in baker's yeast *Saccharomyces cerevisiae*. *MSc. Thesis*. University of Leicester.

Costanzo, M. C., Bonnefoy, N., Williams, E. H., Clark-Walker, G. D. & Fox, T. D. (2000). Highly diverged homologs of *Saccharomyces cerevisiae* mitochondrial mRNA-specific translational activators have orthologous functions in other budding yeasts. *Genetics* 154, 999-1012.

Cryer, D. R., Eccleshall, R. & Marmur, J. (1975). *Methods in cell biology*. New York: Academic Press.

de Barros Lopes, M., Soden, A., Martens, A.L., Henschke, P.A. & Langridge, P (1998). Differentiation and species identification of yeasts using PCR. *Int J Syst Bacteriol* 48, 279-286.

De Deken, R. H. (1966). The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* 44, 149-56.

del Portillo, H. A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C. P., Schneider, N. K., Villalobos, J. M., Rajandream, M. A., Harris, D., Pereira da Silva, L. H., Barrell, B. & Lanzer, M. (2001). A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410, 839-42.

Delneri, D., Gardner, D. C., Bruschi, C. V. & Oliver, S. G. (1999a). Disruption of seven hypothetical aryl alcohol dehydrogenase genes from *Saccharomyces cerevisiae* and construction of a multiple knock-out strain. *Yeast* 15, 1681-9.

Delneri, D., Gardner, D. C. & Oliver, S. G. (1999b). Analysis of the seven-member AAD gene set demonstrates that genetic redundancy in yeast may be more apparent than real. *Genetics* 153, 1591-600.

Denayrolles, M., de Villechenon, E. P., Lonvaud-Funel, A. & Aigle, M. (1997). Incidence of SUC-RTM telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* 31, 457-61.

Dickinson, J. R., Lanterman, M. M., Danner, D. J., Pearson, B. M., Sanz, P., Harrison, S. J. & Hewlins, M. J. E. (1997). A ^{13}C nuclear magnetic resonance investigation of the metabolism of leucine to isoamyl alcohol in *Saccharomyces cerevisiae*. *J Biol Chem* 272, 26871-26878.

Dickinson, J. R., Harrison, S. J., Dickinson, J. A. & Hewlins, M. J. E. (2000). An investigation of the metabolism of isoleucine to active amyl alcohol in *Saccharomyces cerevisiae*. *J Biol Chem* 275, 10937-10942.

Downs, D. M. & Roth, J. R. (1991). Synthesis of thiamine in *Salmonella typhimurium* independent of the *purF* function. *J Bacteriol* 173, 6597-604.

Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. (1975). Ligation of *EcoRI* endonuclease-generated DNA fragments into linear and circular structures. *J Mol Biol* 96, 171-184.

Ehrenschaft, M., Bilski, P., Li, M. Y., Chignell, C. F. & Daub, M. E. (1999). A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proc Natl Acad Sci U S A* 96, 9374-8.

Enjo, F., Nosaka, K., Ogata, M., Iwashima, A. & Nishimura, H. (1997). Isolation and characterization of a thiamin transport gene, *THI10*, from *Saccharomyces cerevisiae*. *J Biol Chem* 272, 19165-19170.

- Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analyt Biochem* 132, 6-13.
- Feldmann, H. (2000). Genolevures-a novel approach to 'evolutionary genomics'. *FEBS Letters* 487, 1-2.
- Flynn, P. J. & Reece, R. J. (1999). Activation of transcription by metabolic intermediates of the pyrimidine biosynthetic pathway. *Mol Cell Biol* 19, 882-8.
- Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E. & Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*.
- Geitz, R. D., Schiestl, R. H., Willems, A. R. & Woods, R. A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* 11, 355-360.
- Gish, W. & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet* 3, 266-72.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. (1996). Life with 6000 genes. *Science* 274, 546-52.
- Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. (2000). PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J Mol Biol* 295, 1073-87.

- Gonzalez, E., Fernandez, M. R., Larroy, C., Sola, L., Pericas, M. A., Pares, X. & Biosca, J. A. (2000). Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the *Saccharomyces cerevisiae* YAL060W gene product. Disruption and induction of the gene. *J Biol Chem* 275, 35876-85.
- Goodey, A. R. & Tubb, R. S. (1982). Genetic and biochemical-analysis of the ability of *Saccharomyces cerevisiae* to decarboxylate cinnamic acids. *J Gen Microbiol* 128, 2615-2620.
- Grue-Sorensen, G., White, R. L. & Spenser, I. D. (1986). Thiamin biosynthesis In *Saccharomyce cerevisiae* - Origin of the pyrimidine unit. *J Amer Chem Soc* 108, 146-158.
- Hampsey, M. (1997). A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast* 13, 1099-133.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557-80.
- Hather, R. J. (1996). Analysis of a set of growth genes in *Saccharomyces cerevisiae*. *PhD. Thesis*. University of Leicester.
- Himmeldirk, K., Sayer, B. G. & Spenser, I. D. (1998). Comparative biogenetic anatomy of vitamin B-1: A 13-C NMR investigation of the biosynthesis of thiamin in *Escherichia coli* and in *Saccharomyces cerevisiae*. *J Amer Chem Soc* 120, 3581-3589.
- Hohmann, S. (1993). Characterization of *PDC2*, a gene necessary for high-level expression of pyruvate decarboxylase structural genes in *Saccharomyces cerevisiae*. *Mol Gen Genet* 241, 657-666.

- Hohmann, S. & Meacock, P. A. (1998). Thiamin metabolism and thiamine diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: genetics and regulation. *Biochim et Biophys Acta* 1385, 201-19.
- Hua, Q., Araki, M., Koide, Y. & Shimizu, K. (2001). Effects of glucose, vitamins, and DO concentrations on pyruvate fermentation using *Torulopsis glabrata* IFO 0005 with metabolic flux analysis. *Biotechnol Prog* 17, 62-8.
- Hunter N, Chambers S.R., Louis E.J., Borts R.H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J* 15, 1726-1733.
- Iwashima, A., Kawasaki, Y. & Kimura, Y. (1990a). Transport of 2-methyl-4-amino-5-hydroxymethylpyrimidine in *Saccharomyces cerevisiae*. *Biochim et Biophys Acta* 1022, 211-214.
- Iwashima, A., Nishino, H. & Nose, Y. (1973). Carrier mediated transport of thiamine in baker's yeast. *Biochim et Biophys Acta* 330, 222-234.
- Iwashima, A., Nosaka, K., Nishimura, H. & Kimura, Y. (1986). Some properties of a *Saccharomyces cerevisiae* mutant resistant to 2- amino-4-methyl-5-b-hydroxyethylthiazole. *J Gen Microbiol* 132, 1541-1546.
- Jabri, E., Carr, M. B., Hausinger, R. P. & Karplus, P. A. (1995). The crystal structure of urease from *Klebsiella aerogenes*. *Science* 268, 998-1004.
- James, S. A., Cai, J., Roberts, I. N. & Collins, M. D. (1997). A phylogenetic analysis of the genus *Saccharomyces* based on 18S rRNA gene sequences: description of *Saccharomyces kunashirensis* sp. nov. and *Saccharomyces martiniae* sp. nov. *Int J Syst Bacteriol* 47, 453-60.

- Jespersen, L., Cesar, L. B., Meaden, P. G. & Jakobsen, M. (1999). Multiple alpha-glucoside transporter genes in brewer's yeast. *Applied and Env Microbiol* 65, 450-456.
- Johnston, J. R. & Mortimer, R. K. (1984). Pedigree of the yeast genetics stock center wild-type strains S288c and X2180 of *S. cerevisiae*. *Heredity* 52, 459-459.
- Jordan, F. (2000). Thiamin diphosphate-dependent decarboxylase active centers: Similarities and differences. *Abstracts of Papers of the Am Chem Soc* 219, 77-BIOL.
- Juilliard, J.-H. & Douce, R. (1991). Biosynthesis of the thiazole moiety of thiamin (vitamin B1) in higher plant chloroplasts. *Proc Natl Acad Sci U S A* 88, 2042-2045.
- Kamihara, T. & Nakamura, I. (1982). Regulation of respiration and its related metabolism by vitamin B1 and vitamin B6 in *Saccharomyces cerevisiae*. *Advances in Biochem Eng and Biotechnol* 29, 35-82.
- Kawasaki, Y. (1993). Copurification of hydroxyethylthiazole kinase and thiamine-phosphate pyrophosphorylase of *Saccharomyces cerevisiae*. Characterization of hydroxyethylthiazole kinase as a bifunctional enzyme in the thiamine biosynthetic pathway. *J Bacteriol* 175, 5153-5158.
- Kawasaki, Y., Nosaka, K., Kaneko, Y., Nishimura, H. & Iwashima, A. (1990). Regulation of thiamine biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol* 172, 6145-6147.
- Kaziro, Y. (1959). Studies on thiaminokinase from baker's yeast 1. Purification and properties. *J Biochem* 46, 1523-1539.
- Keogh, R. S., Seoighe, C. & Wolfe, K. H. (1998). Evolution of gene order and chromosome number in *Saccharomyces*, *Kluyveromyces* and related fungi. *Yeast* 14, 443-457.

Komiya, H., Miyazaki, M. & Takemura, S. (1981). The nucleotide sequence of 5S ribosomal RNA from *Schizosaccharomyces pombe*. *J Biochem (Tokyo)* 89, 1663-6.

Kurtzman, C. P. (1994). Molecular taxonomy of the yeasts. *Yeast* 10, 1727-40.

Leder, I. G. (1975). Thiamine, biosynthesis and function. In *Metabolic pathways*, pp. 57-85.

Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16, 4743-51.

Liu, H., Styles, C. A. & Fink, G. R. (1996). *Saccharomyces cerevisiae* S288C has a mutation in *FLO8*, a gene required for filamentous growth. *Genetics* 144, 967-78.

Llorente, B., Fairhead, C. & Dujon, B. (1999). Genetic redundancy and gene fusion in the genome of the baker's yeast *Saccharomyces cerevisiae*: functional characterization of a three-member gene family involved in the thiamine biosynthetic pathway. *Mol Microbiol* 32, 1140-1152.

Llorente, B., Malpertuy, A., Neuveglise, C., de Montigny, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., Durrens, P., Gaillardin, C., Lepingle, A., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekai, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., Weissenbach, J., Souciet, J. & Dujon, B. (2000). Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Letters* 487, 101-12.

Lorenz, M. C., Muir, R. S., Lim, E., McElver, J., Weber, S. C. & Heitman, J. (1995). Gene disruption with PCR products in *Saccharomyces cerevisiae*. *Gene* 158, 113-7.

Losson, R. & Lacroute, F. (1981). Cloning of a eukaryotic regulatory gene. *Mol Gen Genet* 184, 394-9.

Louis, E. J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 11, 1553-73.

Louloudi, M. & Hadjiliadis, N. (1994). Structural aspects of thiamine, its derivatives and their metal-complexes in relation to the enzymatic action of thiamine enzymes. *Coord Chem Rev* 135, 429-468.

Lyons, T. J., Gasch, A. P., Gaither, L. A., Botstein, D., Brown, P. O. & Eide, D. J. (2000). Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc Natl Acad Sci U S A* 97, 7957-62.

Machado, C. M., Praekelt, U. M., de Oliveira, R. C., Barbosa, A. C. C., Byrne, K. L., Meacock, P. A. & Menck, C. F. M. (1997). Dual role for the yeast *THI4* gene in thiamine biosynthesis and DNA damage tolerance. *J Mol Biol* 273, 114-121.

Mandel, M. & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J Mol Biol* 53, 159-162.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning. A Laboratory Course Manual*. New York: Cold Spring Harbour.

Mantsala, P. & Zalkin, H. (1984). Glutamine nucleotide sequence of *Saccharomyces cerevisiae* *ADE4* encoding phosphoribosylpyrophosphate amidotransferase. *J Biol Chem* 259, 8478-84.

Marsh, L. (2000). Analysis of the *SNO* multi-gene family in *Saccharomyces cerevisiae*. *MSc. Thesis*. University of Leicester.

- Martini, A. V. & Kurtzman, C. P. (1985). Deoxyribonucleic-acid relatedness among species of the genus *Saccharomyces sensu stricto*. *Int J Sys Bacteriol* 35, 508-511.
- Maundrell, K. (1990). *Nmt1* of fission yeast - a highly transcribed gene completely repressed by thiamine. *J Biol Chem* 265, 10857-10864.
- Maundrell, K. (1993). Thiamine-repressible expression vectors pREP and pRIP For Fission Yeast. *Gene* 123, 127-130.
- McAlister, L. & Holland, M. J. (1982). Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. *J Biol Chem* 257, 7181-8.
- Mittenhuber, G. (2001). Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J Mol Microbiol Biotechnol* 3, 1-20.
- Moller, K., Olsson, L. & Piskur, J. (2001). Ability for anaerobic growth is not sufficient for development of the petite phenotype in *Saccharomyces kluyveri*. *J Bacteriol* 183, 2485-9.
- Mortimer, R. K. (2000). Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research* 10, 403-9.
- Mortimer, R. K. & Johnston, J. R. (1986). Genealogy Of principal strains of the Yeast Genetic Stock Center. *Genetics* 113, 35-43.
- Muller, E. H., Richards, E. J., Norbeck, J., Byrne, K. L., Karlsson, K.-A., Pretorius, G. H. J., Meacock, P. A., Blomberg, A. & Hohmann, S. (1999). Thiamine repression and pyruvate decarboxylase autoregulation independently control the expression of the *Saccharomyces cerevisiae* *PDC5* gene. *FEBS Letters* 449, 245-250.

Nagy, M., Lacroute, F. & Thomas, D. (1992). Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *Proc Natl Acad Sci U S A* 89, 8966-70.

Naumov, G.I., Naumova, E. S., Hagler, A.N., Mendonça-Hagler, L.C. & Louis, E.J. (1995a). A new genetically isolated population of the *Saccharomyces sensu stricto* complex from Brazil. *Antonie Leeuwenhoek* 67, 351-355.

Naumov, G.I., Naumova, E. S., Louis, E.J. (1995b). Two new genetically isolated populations of the *Saccharomyces sensu stricto* complex from Japan, *J Gen Appl Microbiol* 41, 499-505.

Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J. & Roberts, I. N. (2000). Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int J Syst Evol Microbiol* 50, 1931-42.

Ness, F. & Aigle, M. (1995). RTM1: a member of a new family of telomeric repeated genes in yeast. *Genetics* 140, 945-56.

Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F., Gaillardin, C. & Casaregola, S. (2000). Genomic exploration of the hemiascomycetous yeasts: 9. *Saccharomyces kluyveri*. *FEBS Letters* 487, 56-60.

Nishida, H. & Sugiyama, J. (1993). Phylogenetic relationships among Taphrina, Saitoella, and other higher fungi. *Mol Biol Evol* 10, 431-6.

Nishimura, H., Kawasaki, Y., Kaneko, Y., Nosaka, K. & Iwashima, A. (1992a). Cloning and characteristics of a positive regulatory gene, *THI2* (*PHO6*), of thiamin biosynthesis in *Saccharomyces cerevisiae*. *FEBS Letters* 297, 155-158.

- Nishimura, H., Kawasaki, Y., Kaneko, Y., Nosaka, K. & Iwashima, A. (1992b). A positive regulatory gene, *THI3*, is required for thiamine metabolism in *Saccharomyces cerevisiae*. *J Bacteriol* 174, 4701-4706.
- Nishimura, H., Kawasaki, Y., Nosaka, K. & Kaneko, Y. (1997). Mutation *thi81* causing a deficiency in the signal transduction of thiamine pyrophosphate in *Saccharomyces cerevisiae*. *FEMS Microbiology Letters* 156, 245-9.
- Nishimura, H., Kawasaki, Y., Nosaka, K., Kaneko, Y. & Iwashima, A. (1991). A constitutive thiamine metabolism mutation, *thi80*, causing reduced thiamine pyrophosphokinase activity in *Saccharomyces cerevisiae*. *J Bacteriol* 173, 2716-2719.
- Nishimura, H., Sempuku, K. & Iwashima, A. (1982). Thiamine transport In *Saccharomyces cerevisiae* protoplasts. *J Bacteriol* 150, 960-962.
- Nosaka, K. (1990). High affinity of acid phosphatase encoded by *PHO3* gene in *Saccharomyces cerevisiae* for thiamin phosphates. *Biochim et Biophys Acta* 1037, 147-154.
- Nosaka, K., Kaneko, Y., Nishimura, H. & Iwashima, A. (1989). A possible role for acid-phosphatase with thiamin-binding activity encoded by *PHO3* in yeast. *FEMS Microbiol Letters* 60, 55-59.
- Nosaka, K., Kaneko, Y., Nishimura, H. & Iwashima, A. (1993). Isolation and characterization of a thiamin pyrophosphokinase gene, *THI80*, from *Saccharomyces cerevisiae*. *J Biol Chem* 268, 17440-17447.
- Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T. & Iwashima, A. (1994). Isolation and characterization of the *THI6* gene encoding a bifunctional thiamin-phosphate pyrophosphorylase hydroxyethylthiazole kinase from *Saccharomyces cerevisiae*. *J Biol Chem* 269, 30510-30516.

Osmani, A. H., May, G. S. & Osmani, S. A. (1999). The extremely conserved *pyroA* gene of *Aspergillus nidulans* is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers. *J Biol Chem* 274, 23565-9.

Padilla, P. A., Fuge, E. K., Crawford, M. E., Errett, A. & WernerWashburne, M. (1998). The highly conserved, coregulated *SNO* and *SNZ* gene families in *Saccharomyces cerevisiae* respond to nutrient limitation. *J Bacteriol* 180, 5718-5726.

Petersen, R. F., Nilsson-Tillgren, T. & Piskur, J. (1999). Karyotypes of *Saccharomyces sensu lato* species. *Int J Syst Bacteriol* 49 Pt 4, 1925-31.

Piskur, J. (2001). Origin of the duplicated regions in the yeast genomes. *TIGs* 17, 302-3.

Praekelt, U. M., Byrne, K. L. & Meacock, P. A. (1994). Regulation of *THI4* (*MOL1*), a thiamine biosynthetic gene of *Saccharomyces cerevisiae*. *Yeast* 10, 481-490.

Praekelt, U. M. & Meacock, P. A. (1992). *MOL1*, a *Saccharomyces cerevisiae* gene that is highly expressed in early stationary phase during growth on molasses. *Yeast* 8, 699-710.

Pryde, F. E., Gorham, H. C. & Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Curr Opin Genet Dev* 7, 822-8.

Pryde, F. E., Huckle, T. C. & Louis, E. J. (1995). Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of sub-telomeric repeat sequences. *Yeast* 11, 371-82.

Rachidi, N., Martinez, M. J., Barre, P. & Blondin, B. (2000). *Saccharomyces cerevisiae* PAU genes are induced by anaerobiosis. *Mol Microbiol* 35, 1421-30.

Raghuram, V., Lobo, Z. & Maitra, P. K. (1994). Pdc2, a yeast gene essential for synthesis of pyruvate decarboxylase, encodes a novel transcription factor. *J Genet* 73, 17-32.

Rainieri, S., Zambonelli, C., Hallsworth, J. E., Pulvirenti, A. & Giudici, P. (1999). *Saccharomyces uvarum*, a distinct group within *Saccharomyces sensu stricto*. *FEMS Microbiol Lett* 177, 177-85.

Reece, R. J. (2000). Molecular basis of nutrient-controlled gene expression in *Saccharomyces cerevisiae*. *Cell Mol Life Sci* 57, 1161-71.

Reynolds, A. (1989). Yeast vectors and assays for expression of cloned genes. In *Current Protocols in Molecular Biology*, pp. 13.6.1-13.6.4. Edited by F. M. Asubel. New York: Wiley Interscience.

Richards, E. (1996). Genetic analysis of two UV mutants of thiamine biosynthesis in the baker's yeast *Saccharomyces cerevisiae*.. *MSc. Thesis*. University of Leicester.

Rose, M. D., Winston, F. and Heiter, P. (1990). *Methods in Yeast Genetics: A Laboratory Manual*. New York: Cold Spring Harbour.

Rudenko, G., McCulloch, R., Dirks-Mulder, A. & Borst, P. (1996). Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*.

Schmitt, M. E., Brown, T. A. & Trumpower, B. L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* 18, 3091-2.

Schweingruber, A. M., Dlugonski, J., Edenharter, E. & Schweingruber, M. E. (1991). Thiamine in *Schizosaccharomyces pombe*: dephosphorylation, intracellular pool, biosynthesis and transport. *Curr Genet* 19.

Seoighe, C. & Wolfe, K. H. (1999). Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* 2, 548-54.

Serghini, M. A., Ritzenthaler, C. & Pinek, L. (1989). A rapid and efficient 'mini prep' for isolation of plasmid DNA. *Nucleic Acids Res* 17, 3604.

Singleton, C. K. (1997). Identification and characterization of the thiamine transporter gene of *Saccharomyces cerevisiae*. *Gene* 199, 111-21.

Sohn, J., Voegelé, R. T., Mendgen, K. & Hahn, M. (2000). High level activation of vitamin B1 biosynthesis genes in haustoria of the rust fungus *Uromyces fabae*. *Mol Plant Microbe Interact* 13, 629-36.

Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B., Durrens, P., Gaillardin, C., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekaiia, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. & Weissenbach, J. (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* 487, 3-12.

Southern, E. M. (1975). Detection of specific sequences amongst DNA fragments separated by gel electrophoresis. *J Mol Biol* 98, 503-517.

Stark, M. J. & Milner, J. S. (1989). Cloning and analysis of the *Kluyveromyces lactis* *TRP1* gene: a chromosomal locus flanked by genes encoding inorganic pyrophosphatase and histone H3. *Yeast* 5, 35-50.

Suomalainen, H. & Oura, E. (1971). Yeast nutrition and solute uptake. In *The Yeasts*, pp. 3-74. Edited by A. Rose & J. Harrison. New York: Academic Press.

Tamai, Y., Momma, T., Yoshimoto, H. & Kaneko, Y. (1998). Co-existence of two types of chromosome in the bottom fermenting yeast, *Saccharomyces pastorianus*. *Yeast* **14**, 923-933.

Tanaka, K., Tazuya, K., Yamada, K. & Kumaoka, H. (2000). Biosynthesis of thiamin under anaerobic conditions in *Saccharomyces cerevisiae*. *Biol & Pharm Bull* **23**, 108-111.

Tazuya, K., Adachi, Y., Masuda, K., Yamada, K. & Kumaoka, H. (1995). Origin Of the Nitrogen Atom Of Pyridoxine In *Saccharomyces cerevisiae*. *Biochim et Biophys Acta-General Subjects* **1244**, 113-116.

Tazuya, K., Azumi, C., Yamada, K. & Kumaoka, H. (1994). Origin of the N-1, C-2 and C-2' atoms of the pyrimidine moiety of thiamin in *Saccharomyces cerevisiae*. *Biochem and Mol Biol Int* **33**, 769-774.

Tazuya, K., Azumi, C., Yamada, K. & Kumaoka, H. (1995). Pyrimidine moiety of thiamin is biosynthesized from pyridoxine and histidine in *Saccharomyces cerevisiae*. *Biochem and Mol Biol Int* **36**, 883-888.

Tazuya, K., Yamada, K. & Kumaoka, H. (1989). Incorporation of histidine into the pyrimidine moiety of thiamin in *Saccharomyces cerevisiae*. *Biochim et Biophys Acta* **990**, 73-79.

Tazuya, K., Yamada, K. & Kumaoka, H. (1993). Pyridoxine is a precursor of the pyrimidine moiety of thiamin in *Saccharomyces cerevisiae*. *Biochem and Mol Biol Int* **30**, 893-899.

- Tazuya, K., Yamada, K., Nakamura, K. & Kumaoka, H. (1987). The origin of the sulfur atom of thiamin. *Biochim et Biophys Acta* 924, 210-215.
- Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc Natl Acad Sci U S A* 77, 5201-5.
- Toh-E, A., Kakimoto, S.-i. & Oshima, Y. (1975a). Two new genes controlling the constitutive acid phosphatase synthesis. *Mol Gen Genet* 141, 81-83.
- Vandyck, L., Pascualahuir, A., Purnelle, B. & Goffeau, A. (1995). An 8.2 Kb DNA segment from chromosome-XIV carries the *RPD3* and *PAS8* genes as well as the *Saccharomyces cerevisiae* homolog of the thiamine-repressed *nmt1* gene and a chromosome-III-duplicated gene for a putative aryl-alcohol dehydrogenase. *Yeast* 11, 987-991.
- Vaughan-Martini, A. & Martini, A. (1987). Three newly delimited species of *Saccharomyces sensu stricto*. *Antonie Leeuwenhoek* 53, 77-84.
- Vaughan-Martini, A. & Martini, A. (1993). A taxonomic key for the genus *Saccharomyces*. *Sys & App Microbiol* 16, 113-119.
- Viswanathan, M., Muthukumar, G., Cong, Y. S. & Lenard, J. (1994). Seripauperins of *Saccharomyces cerevisiae*: a new multigene family encoding serine-poor relatives of serine-rich proteins. *Gene* 148, 149-53.
- Walker, G. (1998). *Yeast Physiology and Biotechnology* : John Wiley and Sons Ltd.
- White, R. L. & Spenser, I. D. (1979). Thiamin biosynthesis in *Saccharomyces cerevisiae*. *Biochem J* 179, 315-325.

White, R. L. & Spenser, I. D. (1982). Thiamin biosynthesis in yeast - origin of the 5-carbon unit of the thiazole moiety. *J Am Chem Soc* 104, 4934-4943.

Wickerham, L. J. (1951). Taxonomy of yeast. *US Dep Agric Tech Bull* 1029, 11-56.

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C. P. & Boles, E. (1999). Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Letters* 464, 123-8.

Wightman, P., Quain, D.E. & Meaden, P.G. (1996). Analysis of production brewing strains of yeast by DNA fingerprinting. *Lett Appl Microbiol* 22, 90-94.

Wolfe, K. H. & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708-713.

Young, E. T., Sloan, J., Miller, B., Li, N., van Riper, K. & Dombek, K. M. (2000). Evolution of a glucose-regulated *ADH* gene in the genus *Saccharomyces*. *Gene* 245, 299-309.

Zurlinden, A. & Schweingruber, M. E. (1997). Identification of a DNA element in the fission yeast *Schizosaccharomyces pombe nmt1 (thi3)* promoter involved in thiamine-regulated gene expression. *J Bact* 179, 5956-5958.

Appendix A

DNA sequence of genomic clone pRW3a.

Saccharomyces kluyveri CBS 3082. Date: 15/11/01. 1-5253 bp.

```
1  TGCATGCCTG CAGGTCGACT CTAGAGGATC CTCTTTATCT GCCTACAAAG GATGAAGCGG
61  AAGGATTACT GTTCAAATAC TTTGCGCACT CCAACTCACA GTTACCTGTA TTACACAGAG
121 AGTTTTTCCT AAAGAAATAC TTCGAGCCTG TATACGGTTC TTTAACCGAA GGTGTATCCT
181 TAGCGTCAGA CCATACTGTT ATAAACAGGG AATTCAAGCT CCCTTCAATC CATCAAATTG
241 AGGGAAACGA TGAAACGCCA TGGTTTGAGA CGTGGAGTTC AAACAGAAAC AGTAACCTCA
301 AAATACCGCT GAGATACCAC GTACCGTTGT TTTTCTAAA CATCATTTTT TCCATTGGAG
361 AATCTACACG TGTCTTGGAT TCAAATGCTG AAAAGTCTAG AAAGTTCAAG AAGCGTGCAG
421 TACATTACAT TGATACATTA TATTGTTCTA ATGATAGGTT AGAGGCTCTT GCGGGTGTGT
481 TGTTGATCGC AATATACTCA ATCATGAGAC CCAACGTTCC TGGTGTGTGG TATATAATGG
541 GTTCTGCACT GAGATTAGCT GTTGATCTGG GACTACATGC AGAGAAACTA AACAAGAATT
601 ACGATCCCTT TACGAGAGAG CTTGTCGAC GTCTGTTCTG GTGTACGTAT GCCCTGGATA
661 GGCAAATCTG TGTCCATTTT GGCCGCCCCT TCGGTATTCC GGACGAAAAC ATTACTGCAA
721 AATTCCCCAG TACGTTGGAC GACGCTTTGA TCACTACTGC TCGGATAGC ATCGATGATT
781 ATTCACAGGT GAAAAGTTCC ATGGCTTCGT ACAAAGTGAT ATCCTTGGCC ATATTTAAGA
841 TCAGAAAAAT TCAGACGTCA ATAGCCCAGG TTTTATATGC GCCCAACGGT CGTCTACCGT
901 CACATTTTAC CGACCTAGAG AAATGGAGAG CAGTGATGGA TTATGAATTG GATACTGGT
961 ATAACAAAGT TACCCCAAAA ACTCATCGTA AAATGAATTG TGAGTTCAAA GTAGAACTAT
1021 TCCAACTGAA CTATTGCCAT TCAAAGATAA TGTTGTACGG GCTCTCTCCC AAATCCTTAT
1081 CTTTAACGTC TTATGGGTAC GAAGTTGTGC ATGATAGTAG TACAAACGCG ATAAATGCAT
1141 ACTATAAACT ATGCAAAGAG GGTAATATTA ATTACACTTG GGTTGCGGTC CACAATCTTT
1201 TCATGGCAGG TATGTCGTAT CTTTACACGA TCTACAATTC CAGTAAAGGC AGCATGGAAA
1261 CCAGTGACAA ATTTAAAGAG CACTCTTCTA AACTATTACA CCGTCTTGAA ATGTCTGATT
1321 GGTCTTGTGA TGCTGCGAAA AATTGTTACA AGATTTTCCA GGTTCTTTCA GCAGCTGTCA
1381 CCAGATTGAA GTTTGAGGAC GTTGAGATGA AGGAGTCAGG TAGAATAGAA ACCGCCGTTT
1441 CTCCCGAAAA GAAAGGTGGC GATGACAACA GGCAAAAGAC AGAAACTCTC ACTTCAGATT
1501 TATCCCTTAA CAGGAATAAT AATGAAATGT CAACGGTTAC CCAAGAAAAC GCACTAGGTG
1561 ACACTTTGGA CCAGTTCTTT GTTGAGTTGG ACAAACTGTC GCCCTTTTCT GAGACCACTA
1621 GAAGTCAGGT CTCCCATTC AATTTCGACT CACCAGAAAC TGAAAATAGA AACACTACAA
1681 GCGGTACTCA MCAGTATATC AATGCCTTGT CCAGTACGGC CCAGAACACC CTTGGAAAGA
1741 TCGATGAGAC AACACCGTAT TCCCAGCCTT CGAATATGGT TAAAGAAAAT TACAAGAGAG
1801 TAGATGAGCT TGGCTTAAGT AGGAAATATC CATCGACCAA AGATGGCCAG CGGGTATATG
1861 ATATGATTTA CCAAGTGAGC ACGGAAGCCG TCTGGGACCA ATTCTTTAGC AATCTTGGAG
1921 GTGGGAATGG TGATGTTGGA GGGCAGCCGG ATTTATTTTA AATATTTTTC TTTCCCTTTC
1981 TTTATATATT CACACCATCG TTCATCATAA CATTTTGTGG AGTCACAACCT GGAAACATAC
2041 TTTGTAAACT CTATCGTGTT GGAACGAAAT CCAACTATCA TCTGTCGACT AGTATTAGTA
2101 TTATTAGTGT ATTGTCATGT ACGGTGTTAG AAGATGACAC GAAGAATGAG GAACAGTCAT
2161 CAAATTTAGT GGAAGCTGAA ACGCAAGGAT GATGAACGAA AAAAAAAAC ATGAAAGGAA
2221 AGAGGAAGTA AGAATATTTG TACGGTAGAA TCATCGACTT TTCTTTCTAG ATTCTTAGAA
2281 AAATGAAGAG AACTCCTAGT AGACCCTGTG CATCAAATAT TATAGCTTTG ATCATGAGAA
2341 ATATGCGAAT ATTGGGACAA TTGTTGAGAT TCCATTTTTT GGTAAAGGCT ATAATAATTT
2401 GGTATATGAT TATACTAGAA GTTCTCCTCC TCTTTCTAGG AATCTCTAAA AGGAGTCGAT
2461 GATTCTATAT ACAATTATTC ATATATTCTC TTTCCTTTTA TATATTTTGG TTCATTATCC
2521 TATTACATTA TCAATCCTTG CGTTTCAGCT TCCACTGACT TAGATGACTG TTTCTCATTC
2581 TTTATGTCAC CTTATATCAC CGTGATGAT AATATACTAG CAACACGAAT ACTAGTCGAT
2641 GGCCGATAGT TGTATTTCTG TCCAACAAAT CAAAAACGG AATCCCGCAA AATCATCTCA
2701 AAATCCGCCC TATTTTCACT ATCAGACTGT TGATATAACG ATTTCCCTAG TTATCTGATG
2761 TTATTTATGT TACTCGACTT ATTACAAGCT CCAGGTAAAC TATCATGTTG TATATAGTTT
2821 TATTTCCCAA AGCCCATCGG AATTCCGGCA ATTTGGCCGA GTGGTTAAGG CGTGAGATTA
```


2881 GAATCTCGTT AGAGTTTCGA CTTATAATTA TCTCTTGGGC TCTGCCCCGCG CAGGTTTCGAA
2941 TCCTGCAGTT GTCGTTATTT CCCATATTTT TTCACAAACT GAACAACCTG TGCGGTACTG
3001 CCAAGCTGGC TTATTGTTAC TGGCAACTAT ATTTTACGGA TTTACAAGGT TATCTAAAGA
3061 CACTGGTTCC CCAGTAAATT TAGCCCAACG CACAGGCTTG TTAACCGATT CACTTGTTTCG
3121 GATAAAAATG CCAAAGCCGA AGGTTGATGC TTAGGATAGA ATTTAACTTA GATGATAAAC
3181 GTATAAAGTG ATCAATATCT TATACAGTCA AGTCCTATAT ACTAGCAATC AAAAATGGAT
3241 CATAGATCGT GTTCATAACA AAGATAGTCA TTGCTGGCTT AGATGGCGGA GACATCCAAT
3301 CTGTTGAAAC CACCACAGTC TCTACACTTT TCTTGGTGAA CACCCATCAA TCTTTGAGCT
3361 TCCAATGGGT CGGAAACTTC AGAAGGTTCT GGCCAAGACA AGTATTCGTT GGTGTAGTTG
3421 GCTTGGTAGT CTGGTGGCAA AATTTCCAAT CTCTTACCGT AGCCGGTAAC CTTCTTCCAG
3481 TCACGGTGGA CGTTGAACAA GGAGGAAGAG AAGTAAGCGT AACATCTTTG GTATTGCTTG
3541 TACTTCAGGT CGTTGTTCAA TTCTGGCTTG AAGTCAATGT ATTTCGGCCA GGCCTTTTCT
3601 GGTTCGTTCA AAACGTAGTC AGTAGCCTTC TTGATAGCAG TCAAGAATCT CTTAACCTTT
3661 TCACCGTTTT CGGCCAAGAA CTTGTCGTTG CAAATGTACA AGATGGTACA GAAACAACAA
3721 CAACCCAAAC AAGCCAGCTT GTCAATTCTC AACATTTGGA CGTCGGTGAC TGGTCTGCCT
3781 TGAGCCTTGC AGTATTCTTC CAATTCAACT TGTTGGATAC ATTCGATACC AATACCAGCA
3841 TGGATCTTAC CTTCCATGAT GTACTTACCG ACGTTCATAC CACATCTGAC AGCAGTGTAG
3901 TCATCTGGGG TCATACCGTA GTGCTTGGTC AATTCATCCA ATTGGATCTT ACMGAATTCA
3961 CCAACGTAAAC CAATACGCTT ACCCTTCAAA GATTGGAAGT CTTTCGGTGAT ACCGCTGTTC
4021 TTCAAGTACA ACAAACCGGT GAAAGGTTTCG TCCAACAAAG AAGCAACAGA GGTAAGTGGG
4081 AAACCTCTAG CTTTGGCAGC TAGGGTGTGG ATCATAGCCT TCAAACCCAT GTCGACCTTA
4141 CCAGAACCGA TCAATTCAGT GACATCAGAT GGGTTGGTTG GTTCCAATAG AGCAATGTCC
4201 AAGTTTTGTT CTTGAAGTA ACCCTTGGTT TGAGCCAAGT AGATTGGGAT GTGGTAGGCA
4261 GTTGGTTGCC AGTTCAACAA AAAGGTGATC TTATCAGTAG ACATATTAGC TTGTTTAGTT
4321 TGTTTGATTG TTTGGATGGT TTTATAATAA GAATAAATAT GAGAAAGATA GAAAAGTTCT
4381 AAGATACTGA AATATTGTGT CCCTTATATA GGTTCAAAT CCAGGGGTAT ACGACTCCAA
4441 GAAGAGTTCT TACTGTTACT TTTGGAGTGT GTACCATGCC ACCCAAACCA TTCTATCTAA
4501 CAGGCCTAAG CACCGCCTCT AAGGTCAAAA TTGTACGTAA TACGCGCTTT TCAGACATTG
4561 TCTACAACCA TTAGAATTTT ATCTGACGGC GTAGAACGTG GAAGAAGGCA TTTTCGACAG
4621 GTTTCATTTT TTAACAGCTT TTAAGTTGTC CATGTATTTT CTCGTAAGT AACTTTGGTA
4681 ACTTTAAGAA CTTACCCTTT TTTTTTCTCA GACCTGCTAC TTTTTTCACG ATTATGGTGC
4741 GCGCTATTTT AATGCGGCTA CAAGTCAAGC TTTTCAGATGA GCAAAGCAGG CGTCTTTTTT
4801 TTCGTTTCCT GCCATAACCC CGCTGGAAAG TCCGAGAGAG ACGGAGAAAT CGGTTTGAGA
4861 GTTTTGAGGC TCAAGGTTTA TAGTGTAAGA ATGTTTAGTC TATCAAAAAG AACGTGGGTG
4921 GATTAGCACC AGCTATTGTC GTATATTGTT GCGGGCCGTC ATTGTAACGC ATACGATCCC
4981 CGGGTACCGA GCTCGAATTC GTAATCATGT CATAGCTGTT TCCTGTGTGA AATTGTTATC
5041 CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT
5101 AATGAGTGAG CTAATCACA TTAATTGCGT TCGGCTCACT GCGGCTTTC AGTCGGGAAA
5161 CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCCCC GGGGAGAGGC GGTTTGCGTA
5221 TTGGGCGCTC TTCCGCTTTC TCGCTCACTG ACT

//

ORF translation from 3 to 1961.

1 HACRSTLEDP LYLPTKDEAE GLLFKYFAHS NSQLPVLHRE FFLKKYFEPV YGSLTEGVSL
61 ASDHTVINRE FKLPSIHQIE GNDETPWFET WSSNRNSNLK IPLRYHVPLF FLNIIFSIGE
121 STRVLDSNAE KSRNFKKRAV HYIDTLYCSN DRLEALAGVL LIAIYSIMRP NVPGVWYIMG
181 SALRLAVDLG LHAEKLNKNY DPFTRELRRR LFWCTYALDR QICVHFGRPF GIPDENITAK
241 FPSTLDDALI TTAADSIDDY SQVKSSMASY KVISLAIFKI RKIQTSIAQV LYAPNGRLPS
301 HFTDLEKWRA VMDYELDNWY NKVTPKTHRK MNCEFKVELF QLNYCHSKIM LYGLSPKSLS
361 LTSYGYEVVH DSSTNAINAY YKLCKEKNIN YTWVAVHNLF MAGMSYLYTI YNSSKGSMET
421 SDKFKEHSSK LLHRLEMSDW SCDAAKNCYK IFQVLSAAVT RLKFEDVEMK ESGRIETAVP
481 PEKKGDDNR QKTETLTSDL SLNRNNNEMS TVTQENALGD TLDQFFVELD KLSPFSETTR
541 SQVSHSNFDS PETENRNTTS GTXQYINALS STAQNTLGKI DETTPYSQPS NMVKENYKRV
601 DELGLSRKYP STKDGQRVYD MIYQVSTEAV WDQFFSNLGG GNGDVGGQPD LF

//

ORF translation from 3279 to 4304 on the reverse strand.

1 MSTDKITFLL NWQPTAYHIP IYLAQTKGYF KEQNLDIALL EPTNPSDVTE LIGSGKVDMG
61 LKAMIHTLAA KARGFPVTSV ASLLDEPFTG LLYLKNSGIT EDFQSLKGKR IGYVGEFXKI
121 QLDELTKHYG MTPDDYTAVR CGMNVGKYIM EGKIHAGIGI ECIQQVELEE YCKAQGRPVT
181 DVQMLRIDKL ACLGCCCCFCT ILYICNDKFL AENGEKVKRF LTAIKKATDY VLNEPEKAWA
241 EYIDFKPELN NDLKYKQYQR CYAYFSSSLF NVHRDWKKVT GYGKRLEILP PDYQANYTNE
301 YLSWPEPSEV SDPLEAQRLM GVHQEKCRDC GGFNRLDVSA I

//

DNA sequence of genomic clones pDW5a and pDW5b.

Kluyveromyces lactis CBS 2359. Date: 15/11/01. 1-5943 bp.

```
1  GTGTCACCTC GGGTCAAAGT CTTTATGACA TACCCTGGGC TACAAATTAG TGATTATGAT
61  ATTTTAACTA CGATGCAAGC CACCAGAAAG AATGGTTTCA CAACAATGGT TCATGCAGAG
121 AACGGAGATA TGGTTAAATG GATGATTGAG CAACTTGAAG AGCAAAATAT GACCTCCCCT
181 TACTTTTCATG GTGTCTCTAG ACCAACCATA GTTGAAGGAG AAGCTACAAA TAGAGCCATT
241 GTGCTAGCCA AAATGATGGA CCCCCAATC CTGTTTCGTT ATGTATCATC GCCGGAGGCC
301 ATTGATACAA TTAGGTCAGC GCAAACAAAA GGTCTCAAAG TATACGCCGA GACATGTCCA
361 CAGTATGCCC TTCTTTCTGA TATTGATACC AAATGTCACC ACGAATTAGA GCAGCAGCAG
421 AGTACTTGTT CGGATGCAAA TAAATCAGAA CCAAGGGTAA AACTAACTGC AAAGGAAATT
481 GATGATAAAT TTATAGGAGC AAAAGCAATA TGTTCCTCCT CAATCCGTCC GGAGCACTGT
541 CAAGAGTCAG TCTGGAAAGC CATGAACAAT GGAACATTTA CAATTGTTGG ATCTGATCAT
601 TGTGCCTATA ATTACTACGA CAAAACGAGC GGTAAATTGG ACGCATTTAA AGACAATAAG
661 AACGGAGAGT TCAAGTATAT TCCCAATGGT ATGCCGGGAG TTTGTACTAG AATGCCTTTG
721 CTATTTGATT ATGGCGTTCT TCAAGGGAAA CTTGCTAACA TGACCAAGTT CGTGGAGCTT
781 CAGTGTAATA ATCCCGCTAA ACTATATGGA CTGTATCCGA AAAAAGGTTC CATTCTTCCT
841 GCGTCAGTG ATGCTGATTT AGTGGTTTGG TATCCACCAA ATTACAAAGG TGGAAAGAAA
901 CCTAAACTG TAACAAATGA CTTGTTAGTT CATGGATGTG ATTACACACC ATATGAAGGG
961 TTTGAAATAA ATAAGTGGCC AAGGTTTACC ATTGTTAAAG GCAGCATAGT CTACAAAGAA
1021 GGCGAAATAA TCAAAGAAAA TGCAAAGGGC ACTTATGTAA AAAGAGGAAA GAGTCAACTC
1081 TGTACAGCTA ATGATGAATG GGTAACGGAG TGGAGACCAC AATACCTAAA TGGTCAAAAG
1141 AGTGGCTAAA ACTTAACCTT GTTTTATCTA TATATTGAAA CATGTAAAAA ATACAGTCAA
1201 CTTTCACATG AAGCCTTTAA GCATCGAAAA AATCACATGA TTTAGCAATC TGATCCCATT
1261 ATTAAATTGA TAGCTTCATT AATAGAGTTT AATAATAGGA TTTGTGTTTC TGTATATGT
1321 AATCCTGACA TTTAGTGAGA TACTTAATTT TTTCGTTTTA CTGCATATAA GGAAAAAGAT
1381 ACTACTGTTA GAGCATGAAA GGAATAGCTG CGACGCAACC GGTTCGGCT AGCTTATTAA
1441 TATAGCAACT TTTTACCGTT CCTTATACCG TTCCTGGGTG TGTAGTATGG TCAAGCTGCT
1501 TGTAATGAGC CAAAAACGAT AAGCGACGCG GTGTGAAATA AATAGTTACC CGGATATACA
1561 ATACATACTC TTGATAAAAT CCAGCCATTG CGCATTGTAT GGCAACTAAG TGTCTACAGT
1621 AATTAAAGAT ACCTAGGCAG TCTCTAAATA TTACATACTC TGTGCAAAGT CTGCTGAAAG
1681 TTCGGTTTAT TGAATGAAAA AAACAACGAA ATAATACCCA AATTCAATTG AGAGAAATGTC
1741 TTGACTATTG TTTCAAACCT CGGAATTTAA GCGTACCCTT TGCATTTCTG TTCTAGGGAA
1801 AACTGACACT ATAAGTTACA TTCACAACGT ACCCTAATAT AAAGTTTAGA TTTTATTGGC
1861 ATATACCAA TATACCAATT CTTGCATAAT AATAGTAGGG AGTTCAAGTC TTGATGTTTT
1921 GTTTCCTTTT TTTCCAAGAA ATAACTAGC GAATTTAGGA AATGCTCTTC CAGTTGCATA
1981 ATAAATGAAC AGCTAATGAG ATTGGACTTT TTTTGTATCT TTTAACTACG CAAATGGTTG
2041 AAAATCCTGC TAACGGTGCT AGGTTTATAG TATCCTTATG GTGTCAATAG GGGGTGGAAA
2101 ATCATCATCA GATGGTACCA CAATGAGTTC TTTTATAGGG AAGTGCCTCA AGTATCACAC
2161 ACAGCCATAA ACTCTGGTTC AAAACATATA AAAGACAGGT AACTGTCATT GCTGAAAGAG
2221 ATCTCAGTTG AAGATCATTG TAATATATCT GTTTATATTG TATAGTGTCC AACAATCACC
2281 AACCTTTAAG TTAAACAAAC ACAGAACACA ATGTCTTCAG ACAAATCAC ATTTTTATTA
2341 AACTGGCAAG CTGCCCCATA CCATGTCCCA GTCTATTTGG CTTCCATGAA AGGCTATTTT
2401 AAAGAACAAG GAATTGATGT CGCTATTCTA GAACCAAGTA ACCCATCTGA TGTGACTGAA
2461 TTGATTGGTT CGGGTAAAGT TGACATGGGT CTAAAGGCTA TGATCCATAC CCTTGCTGCT
2521 AAGGCCAGAG GTTACGATGT CACTTCTATC GCTTCTTTAT TGGACGAGCC ATTCACCGGT
2581 GTCTTGTA CTGGAAGGTAG CGGTATTACC CAAGACTTCA ACTCTTTGAA GGGTAAGAAA
2641 ATCGGATACG TTGGTGAATT CGGTAAGATT CAAATCGATG AATTGACCAA ACATTACGGT
2701 ATGAAACCAA CGGATTACGA AGCTGTCAGA TGTGGTATGA ACGTTGCCAA ATACATCATT
2761 GAAGGTAAGA TCGATGCTGG TGTGTTGTTT GAATGTATGC AACAAGTTCA ACTTGAAGAA
2821 TACTTGAAGG AAAAAGGCAG AGATCCAAAG GACGCTAAGA TGTGAGAAT TGACGAACCT
2881 GCATGTTTGG GATGTTGTTG TTTCTGTACC ATCTTGATCA TCGCCAATGA CAAATCTTTG
2941 GCTGAAAACC CTGTCAAGGT TAAGAAGTTC TTGAATGCTA TCAAGAAAGC AACTGACTTG
3001 GTTTTGCAAG ATCCTGAACA AGCATGGAAG GATTACGTTG ACTTCAAGCC TCAATTGAAC
```


3061 GACCCATTGA GTTACAAGCA ATTCCAAAGA TGTTTCGCTT ACTTCTCCTC TTCCTTGTAC
3121 AACGTCCATC GTGACTGGAG AAAGGTTACT GCTTACGGTA AGCGTTTGGA TATCCTTCCA
3181 ACTGATTTCA AATCAAAC TA CACTAACGAG TACTTGTCCT GGAAGAACC TACCGAAGTT
3241 GATGATCCAT TAGCTGCTCA AAGATTAATC AAAGCCCATC AAGAAGCTTG CAGAGACCAA
3301 GGTGGTTTCT TCAAGAGAGT TGAGCTAAAA TAAAACATCT CCACAACAAT ATGAGGTCAC
3361 TATCAATGCT CAATTCTTTT TATCTTGGTG TCTGGTTACC AGTCACCCAT ATGTAAGACC
3421 TTGCTCTCTT TCCATATAGT CAATTTTATG TTTAGTGAAA TAACGAATTA TAAACTGATT
3481 GAATCATACC GCACAGTCAT CACCTTCTAT TCTTCGTTGC ATTGTTYSKT TCCACTTCAG
3541 AGTATTTTAG TAATGAACCC AGACAGATAA CCCCTGATCA TAACAATTAT TAAGCTTGTG
3601 TTATCTTAAG AGCCGAATCG CAGCTGATGA TAGGTTCTTG CGTGTGTTGT AGATTATCCT
3661 GGAATGTATG CTTGGCCTAT GTAACCTTC CACATAAATA TGATCTAAAA AATAACACTA
3721 AAAGTATATC ATGCGAGATA TTCAACCTTA TACCTATCTA TGACTATAAT AAGTAGATGC
3781 AACATCTCCC AAATGCAAAT TTCAAAGCCA GAACCATGAG TGTCTTCCTA TTGAAGAATT
3841 AACTGATCTA ACTCTGATGT TAGAGCGTTG CAATGTGCTT TGTCCGCTAG TATTATTTTG
3901 CATCTTTTTA AAGCGAATAA GGTTCGGTTT ACAATTGTGA AGGTGAAGAT GGCCGTATTT
3961 TTCTTACTTC ACGGTGGGT TAGAAGTTCA TAGTGTGGCC TGAAC TAAGC AACGTACTCG
4021 AAATTAAATT ACTGCACAAA CTAAAGGGA AACAAGTATT CTTTAGGAAG AGAAGTCAGC
4081 AGAATCCACA GTTTTAAATA GACGATAGCA ACTGGCAGTC TTTTGACTGG AAATTCTCGC
4141 ATTTCTATTC ATAAATAATT AGAAAGAACA CCCAGAGTA TCATAGTTTC ACGAGATCTT
4201 TAAAACAGTT GATCAACTTG GATCTTCCTT TAGCCGACT CGTGATAACG AAGTATTATA
4261 TCAAGTTCCG AAAGTTTAGC CCATGGATTT AATAGACAAT TCGAATTCGT TTCCTAACGG
4321 GATGTCTGAA AGGTGGTAGG GTGCATTTGG AACCGACTTA TAAACGGGTA TTTGCTGGTC
4381 GATAAAATGT TCAACCTACT CGTCAACTAA TTGTCTGTTT TAAC TTTT TTAGATCA
4441 CTTTTTAAAC TTAAATATGA GTTAATTAGT TTTAATTCAC ACTAAATGAC TGTTCATTAT
4501 TTGAATATGA GGGTCAAGTT TCTTTATGGA CTCATGGATA ACCGATNAGC ACCAAGGTTT
4561 GTTGTTTCGNG TGT TTTCTTG TCATTCGTT ATTATTCATT TATTTTAGAG TTTGTACATT
4621 TAGAATAAGA AAGACAGTTT GAGTGCTTAA TTATTTAAAA GAAAAAATGT GCTGCGGGAA
4681 AGCTTAAATA GGTCAAAGGT CACGTAAAGA GGTACTTCAG AGNCTGCACT ATCGAAATAT
4741 TACTATGTAG GTATTACTGG TGGTATTTAC CGTGTGTTTT CTACAAATTT TCTATTATGT
4801 GATATCTGAC TTTCTTGTCT GATAATAGTC TTTTCCCTTA AGTCCATCAN GAAATGAGCT
4861 TAATTAATAA TATTGCAGAT GTCATTACTT TAACCAGTAC TTTTCACAAA GCAAAGATAA
4921 ACGACAAGAC NAACACATAC ACCAAGGCCA TGTCAAAGCC AGAAAAAGAT GCTAATGATG
4981 TGCATGAGTT GCTTCCTGTT AAGGAAATAC CAGCTCCCCC TTATTCNGAG GGTGGGGAGC
5041 TAGGACGTGA TTTCTCTTCT AATATTATTT TGCACTCCTT AAGAGCAAAT AAAACATGTT
5101 TTAACATGGT GAAGATAACG ACGGTAGTAG CGTGCTTACT TTATGGTTTG GCTCTATATG
5161 AGTTAGATAT AGGTTTTCGCA TTATTCTCCA TATTATGAC TACAGCAGTG GCCTTTTTAT
5221 TTTGCTTATT TCCAGTATTA CATGCGTATG ATCGTAATAC CGAACCTTCT GGTTCAGTAT
5281 TCAAAGATAT CGGTAACGCT CTATTTTTC TAAAGACCAT TCATGACTTC AAGCCTTCTT
5341 ATATTGACAT GAAGCAATGG GACATTGTTG CAAGGGTAGT AAACAAACAT CTTTATGAAG
5401 CAGGCAGCTG GCCAAATAAT CGATTCTTTT ATGATGGGAA AAGATGCCAG GCCATGTTCA
5461 TAGCCTTAGT AAAGAAACCA ACCGATCCTT TACTCGACGA CTACATTTTCG AAGGCAAATG
5521 AAAC TTTTGAA GATCTCGATG GACGAATTGT GGAATTCCAC TGCGGCAAGC TCAGCGTAAG
5581 AGGACGGTAT GCATATCGTA GTCATTGTC ATTTTCGTCAG CTGTCACACA AAGATTCATA
5641 TTTCTATCTA TTCAGAGTCG GCCAATTTTT GAGGGCTCTA TATTTAACGT GAATCATGTC
5701 TCGAACCGAA CATAAATACA TTGACAACCT CCACTATCGA ATTATACTTC CAGAGATCAA
5761 TTTGTTTCAG TGAATTAGTA AGGTGTCGGA GATTTAGCAT TATAAATATA CCGAAAATAA
5821 ACTCGATAAG ATATACTTCT GTAGTATTTA AATACCATAA ACATGAAGAT AAAGAAAGTA
5881 AGGGTACCAC TACTTGCTCC TGATAATAGC TGGGACGGGT ATCGAATTAC AAATTCAAAT
5941 TCA

//

ORF1 translation from 1 to 1149.

1 VSPRVKVFMT YPGLQISDYG ILTTMQATRK NGFTTMVHAE NGDMVKWMIE QLEEQNMTSP
61 YFHGVS RPTI VEGEATNRAI VLAKMMDTPI L FVHVSSPEA IDTIRSAQTK GLKVYAETCP
121 QYALLSDIDT KCHHELEQQQ STCSDANKSE PRVKLTAKEI DDKFIGAKAI CSPPIRPEHC
181 QESVWKAMNN GTFTIVGSDH CAYNYYDKTS GKLD AFKDNK NGEFKYIPNG MPGVCTR MPL
241 LFDYGV LQ GK LANMTKFVEL QCTNPAKLYG LYPKKG SILP GVS DADLVVW YPPNYKGGKK
301 PKTVTNDLLV HGCDYTPYEG FEINNWP RFT IVKGSIVYKE GEI IKENAKG TYVKRGKSQL

361 CTANDEWVTE WRPQYLNGQK SG

//

ORF2 translation from 2311 to 3333.

1 MSSDKITFLL NWQAAPYHVP VYLASMKG YF KEQGIDVAIL EPSNPSDVTE LIGSGKVDMG
61 LKAMIHTLAA KARGYDVTSI ASLLDEPFTG VLYLEGGSGIT QDFNSLKGKK IGYVGEFGKI
121 QIDELTKHYG MKPTDYEAVR CGMNVAKYII EGKIDAGVGI ECMQQVQLEE YLKEKGRDPK
181 DAKMLRIDEL ACLGCCCCFCT ILYIANDKFL AENPVKVKKF LNAIKKATDL VLQDPEQAWK
241 DYVDFKPQLN DPLSYKQFQR CFAYFSSSLY NVHRDWRKVT AYGKRLDILP TDFKSNYTNE
301 YLSWEEPTEV DDPLAAQRLI KAHQEACRDQ GGFFKRVELK

//

ORF3 translation from 4854 5579.

1 MSLIKNIADV ITLTSTFHKA KINDKXNTYT KAMSKPEKDA NDVHELLPVK EIPAPPYXEG
61 GELGRDFSSN IILHSLRANK TCFNMVKITT VVACLLYGLA LYELDIGFAL FSIFMTTAVA
121 FLFCLFPVLH AYDRNTEPSG SVFKDIGNAL FFLKTIHDFK PSYIDMKQWD IVARVVNKH
181 YEAGSWPNNR FFYDGKRCQA MFIALVKKPT DPLLDYISK ANETLKISMD ELWNSTAASS
241 A

//