Journal of Industrial Microbiology & Biotechnology (2002) 28, 186–191

www.nature.com/jim

The yeast lifecycle and DNA array technology

RM Williams

Department of Biochemistry, Stanford University Medical School, Stanford, CA 94305-5307, USA

The genome variability and meiotic gene expression patterns in two unrelated laboratory yeast strains, SK1 and W303, have been characterized using high-density oligonucleotide arrays. The statistical analysis and comparison of the data has allowed identification of: (1) genes with functional importance to meiosis and sporulation in yeast and (2) genes expressed in a strain-specific manner. The genome-wide data also reveal potential reasons why these strains display significant differences in the ability to make fertile spores. Molecularly tagged yeast deletion strains have been used to determine the contribution of each gene to the execution of the sporulation/germination pathway. The application of genetics and the new genomic technologies have allowed a quantum jump in our understanding of yeast molecular biology.
Journal of Industrial Microbiology & Biotechnology (2002) **28,** 186–191 DOI: 10.1038/sj/jim/7000200

Keywords: microarray; genotyping; expression analysis; yeast; genomics; fertility

Introduction

This report describes the application of functional genomics to a specific biological question: our understanding of meiosis and sporulation in yeast [7]. The techniques applied to this question include: strain genotyping, expression analysis and the massively parallel study of yeast deletion strains. A primary aim of this work is to identify all the genes in yeast making functional contributions to meiosis, sporulation and germination. A secondary aim is to determine why one yeast strain (SK1) should have a superior fertility to many of the alternative laboratory yeast strains (e.g., W303). In brief, the genomic analysis of these two unrelated yeast strains has produced: (1) a set of ~2000 genetic markers and ~40 gene deletions potentially linked to sporulation; (2) ~1600 genes expressed during meiosis and spore formation in either strain, with ~900 being expressed irrespective of strain background; and (3) ~300 deletion strains unable to undergo all the steps in the yeast lifecycle.

Background

Why study meiosis? From the point of view of functional genomics, this is an opportunity to test new technologies within the framework of an organism and a developmental process, which have been relatively well characterized during the past 40 years of yeast genetics [5]. Biologically, meiosis is a fundamentally important process, which is central to the reproduction of eukaryotes and underlies most of the genetic diversity on the planet. In addition, the key events in meiosis are conserved from yeast to man; thus, its study in yeast has some relevance to our understanding of reproductive health in man.

For yeast to undergo the meiosis/sporulation pathway, it must be diploid, indicated by the presence of the a/alpha heterodimer, and it must also be in the absence of a fermentable carbon source or complex nitrogen source. In the laboratory, this means that yeasts sporulate when they are transferred from a relatively rich pre-sporulation media (YPA yeast extract–peptone–acetate) into a medium of 2% acetate (e.g., SPII). The progression of the morphological events associated with sporulation is characterized by three waves of gene expression: early, middle and late. These events are summarized in Figure 1. Early meiosis includes the events of DNA synthesis and recombination. Middle meiosis is characterized by the chromosome segregation events, and late meiosis involves spore maturation.

The observed kinetics of this pathway varies considerably from strain to strain. Strain SK1 undergoes the pathway in a highly synchronous, rapid and efficient manner, yielding nearly 100% spores after 10–12 h in SPII media. In contrast, strain W303 is far less efficient and only produces 60% spores after 24 h in SPII medium. Both strains were used in the genotyping and expression analysis in an attempt to produce a corroborated list of genes with importance to meiosis and sporulation irrespective of strain background. These were expected to be the genes with significant functional importance to meiosis and sporulation.

The design and principle of the Affymetrix GeneChip have been described on many occasions and the reader is referred to previous

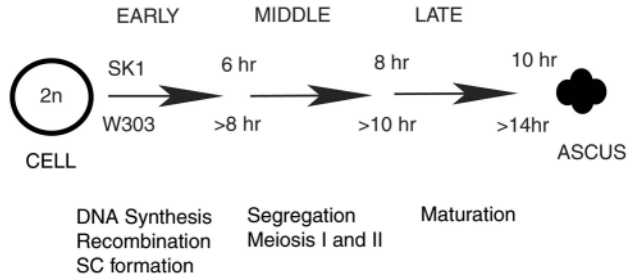


Figure 1 The kinetics and major morphological landmarks as yeast undergo meiosis and sporulation. Early meiotic events are controlled by the Ime1p and Ume6p factors and middle meiotic events are determined by the transcription factor Ndt80p. Specific late meiotic transcription factors are yet to be identified. MATα/MATα or MATa/MATa control diploid strains, isogenic to their MATa/MATα counterparts, fail to enter meiosis; as an alternative, these strains undergo a starvation response.

Correspondence: RM Williams, Diversa, 4995 Directors Place, San Diego, CA 92121, USA
Received 5 November 2000; accepted 26 June 2001

articles for a comprehensive description [6,13]. In brief, the high-density oligonucleotide array is composed of many thousands of individual oligonucleotide features, each of which is spatially addressable on the chip and contains a 25-mer oligonucleotide with a unique sequence. These 25-mer oligonucleotides are synthesized on the chip using a system of masks and chemical photolithography. The individual oligonucleotide sequences were chosen to be as unique as possible, and to be as representative as possible of the whole gene. Notably, the sequences used for the yeast genome arrays were based on the data from the strain S288c sequencing project. Features for a particular gene are laid out next to one another on the array with as many as 20 independent features being synthesized to represent each gene. The final determination of whether a particular gene sequence is present or absent in a complex mixture of nucleic acid (be it DNA or RNA) is decided by integration of hybridization data over all available features representing a particular gene.

Genotyping W303 and SK1

Initially, an overview of just how different the two yeast strains were at the level of genome content was generated by allelic variation scanning [12]. Genomic DNA was prepared from each strain, fragmented with DNaseI, labelled with a dideoxy adenine derivative with a biotin incorporated, hybridized to an array and stained with a streptavidin–phycoerythrin conjugate. Figure 2 shows scanned images of high-density oligonucleotide arrays with hybridized and labelled genomic DNA. Notably, almost every feature on the arrays is illuminated, confirming as expected that most of the sequences present in the samples were also present on the array. The comparison of the two scanned images reveals long blank patches on both of the arrays (marked with arrows labelled A and B). These correspond to strain-specific gene deletions. The deletion shown for W303 (A) covers the ORFs: ynr070w, ynr071w and ynr072w. The deletion shown for SK1 (B) covers the ORFs: yol163w, yol162w and yol161w. In summary, a total of 8 large deletions are found in W303 relative to S288c, and 39 deletions are found in SK1 relative to S288c. There are also a number of gene duplications unique to each strain, as well as many single feature differences. Interestingly, some of the genes deleted in SK1 are expressed in W303 during meiosis. This suggests that they could be guilty in part for the poor sporulation performance of the W303 strain.

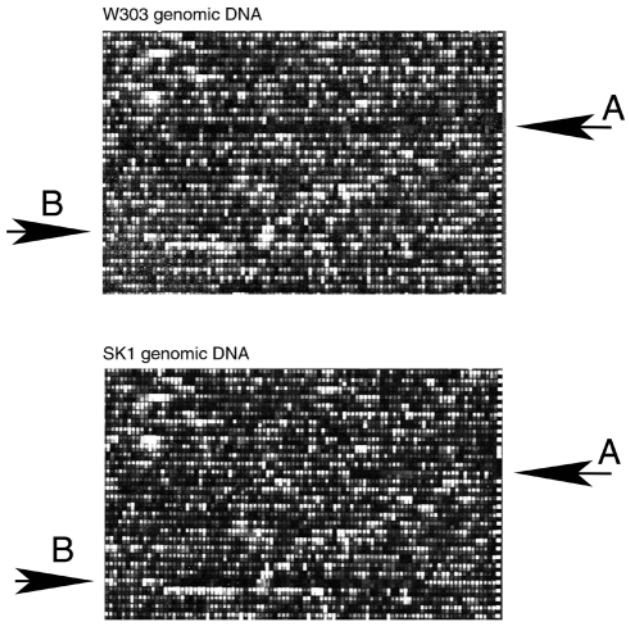


Figure 2 Scanned images showing the same regions of two independent high-density oligonucleotide arrays hybridized with fragmented total genomic DNA from W303 (top panel) and SK1 (bottom panel). The arrows point to (A) a deletion in only W303 and (B) a deletion in only SK1.

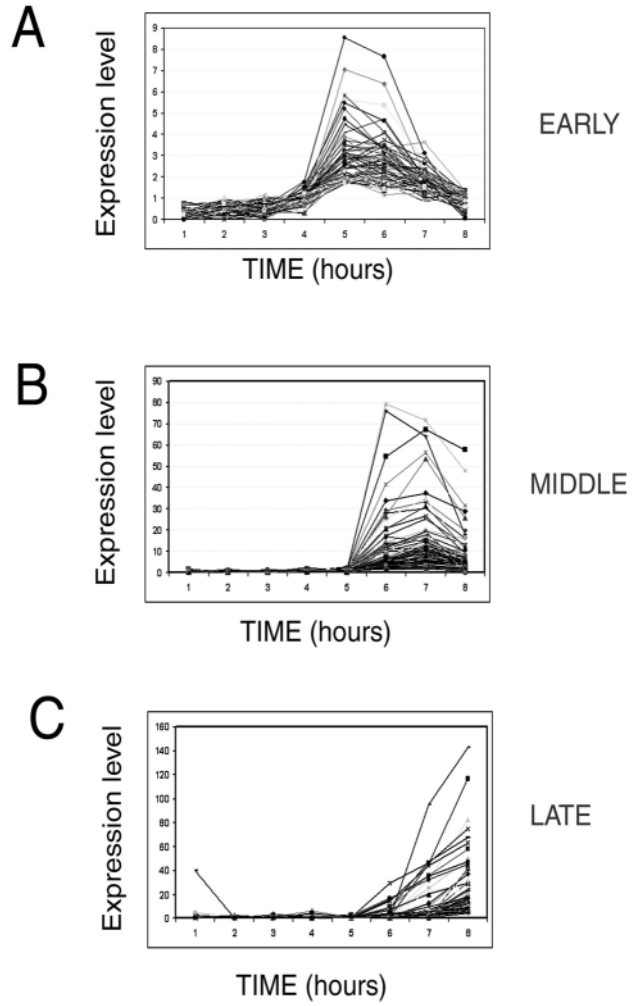


Figure 3 Waves of transcription during meiosis in SK1. Panel A: Selected genes showing early meiotic transcription profiles. Panels B and C: Genes with middle and late meiotic transcriptional profiles, respectively. The x-axis represents time and the y-axis is expression level.

Differences between the two strains at the level of individual feature hybridization can be deduced by comparing fluorescence intensities on a feature-by-feature and gene-by-gene basis. An automated scan was used to look for significant loss of hybridization for each array feature for all the genes in the genome. This revealed that between W303 and S288c, there are only 318 features throughout the genome with significantly different hybridization intensities. In contrast, between SK1 and S288c, there are 2025 features with significantly different hybridization intensities,

indicating that SK1 and S288c/W303 are essentially unrelated. Historically, these results make perfect sense; S288c (the sequenced strain) was used as the starting material in a series of crosses aimed at producing a strain with superior sporulation properties — the result of this procedure was W303. The genetic procedures unequivocally worked, presumably in part by introducing the 318 detected differences into S288c. Now identified, these differences await further characterization and an assessment of their relevance to meiosis.

In conclusion, there are a number of regions in the yeast genome linked to meiosis/sporulation efficiency; 318 are of particular interest. Genotyping using the oligonucleotide arrays is very informative about the genetic variation present in yeast, allowing the relationships between different strains to be rapidly evaluated.

Expression analysis of W303 and SK1 during meiosis

A yeast genomic expression study is carried out following a basic protocol: mRNA is isolated using hot phenol and glass beads. The mRNA is reverse-transcribed into cDNA, fragmented with DNaseI then treated as described above for the hybridization and detection of genomic DNA [1]. Once the array is scanned, the fluorescence intensities record a measure of the abundance of every mRNA present in the cell at the time of sample preparation. For the example in question, samples were prepared over a time course as a culture of diploid yeast underwent the meiosis/sporulation pathway.

Figure 3 shows a graphic representation of the differential transcription occurring in SK1 during meiosis/sporulation. The three main waves of gene expression are presented. Panel A shows genes with an early meiotic expression pattern, induced at 2–3 h and then later re-repressed. Panel B shows the middle genes, induced around 4–5 h and then re-repressed, and the late genes shown in panel C are induced 6–7 h into the time course.

The presentation of the expression data for the entire genome is more practically achieved using the clustergram format [2], as shown in Figure 4. Using four different time courses (26 different samples in total, including controls), the 26 point vectors describing the expression characteristics of each gene can be analyzed using pairwise correlation statistics, and genes with similar vectors placed next to one another in the clustergram [3]. Importantly, this gives insights into the potential function of previously uncharacterized genes using the "guilt by association" hypothesis. Employing all the available expression data to do this is a good idea as it defines with the greatest precision when a genes expression is either ON or OFF, or somewhere in between. Clusters characteristic of just meiosis-specific gene expression or just starvation-specific expression can be seen (see arrows).

Around 1600 genes were found to be significantly meiotically regulated in sporulating SK1 or W303 [7]. This number is compatible with an independent estimate based of genomic analysis of LacZ translational fusions during sporulation [8]. Gene-by-gene comparison in SK1 and W303 identified ~900 "core" loci meiotically regulated in both strains (including ~650 meiotically expressed genes not previously reported [2]) that comprise the vast majority of known sporulation genes and other open reading frames essential for the process. This strain-specific phenomenon could be accounted for by promoter sequence variation or differences in the transcription apparatus. It is possible that differences in the synchrony of the two cultures as they underwent the sporulation pathway also resulted in the decreased detection of some transcripts.

The statistical analysis of expression data

Commercial software for statistical analysis and clustering of expression data is now widely available. A package called GeneSpring from Silicon Genetics (Redwood City, CA) integrates

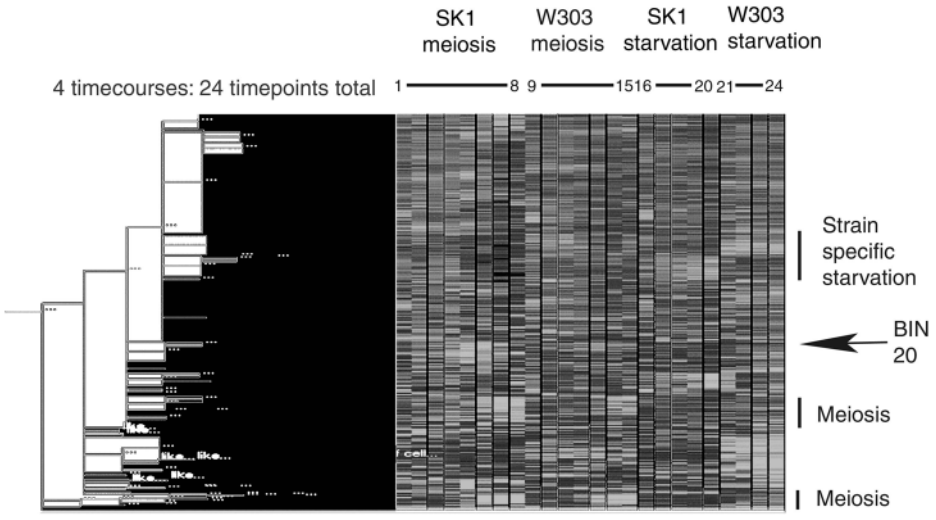
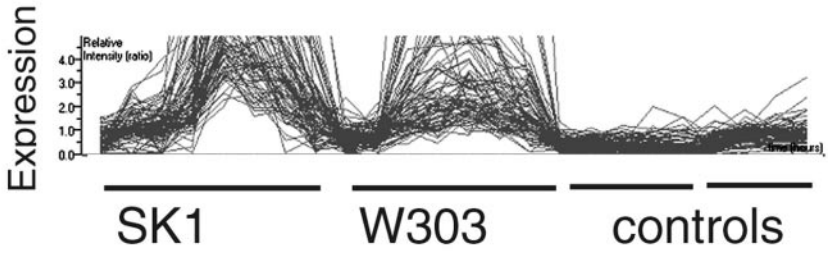


Figure 4 Dendrogram and accompanying clustergram generated after data analysis using pairwise correlation statistics; 24 data points for each of the ~6200 genes in the yeast genome are presented. These were gathered using two different strains undergoing both meiosis and starvation. SK1 meiosis time course: vertical columns 1–8; W303 meiosis time course: columns 9–15; SK1 starvation time course: columns 16–20; W303 starvation time course: columns 21–24. In this presentation, the light shading depicts high expression levels and the dark shading represents low expression levels. On the right side of the figure are vertical bars indicating the regions of the clustergram containing genes with expression patterns typical of meiosis or a strain-specific starvation response. The arrow indicates the approximate location in the clustergram of genes binned together in cluster bin 20 based on their similar expression characteristics (~90 early meiotic genes).

A



B

ORF	Distance		Sequence	
YDR285W	18	ATACCCTCGG	CGGCTAA	ATTTTTAGAG
YDR374C	227	TGTACCTCGG	CGGCTAA	TCTTTTTGGT
YER044C-A	94	TATTGCTGGG	CGGCTAA	ATGTTTCCGA
YFL003C	-58	TTAAGGTCGG	CGGCTAA	TTCTTCAAAC
YGL205W	266	CGGTCATTAG	CGGCTAA	TAGCCGTTGG
YHL022C	-166	CTGTCATTGG	CGGCTAA	TTCATTAGAG
YHR153C	86	GAAAAGTGGG	CGGCTAA	AACCGAGAAA
YHR172W	376	TCGTTGCTGA	CGGCTAA	GAAATTGGTA
YIL031W	415	AAATCGTCGG	CGGCTAA	TCTGAAAAGT
YIL072W	169	TTAACCTGGG	CGGCTAA	ATTGTACTTT
YIL073C	179	CAAGTTTATG	CGGCTAA	GAATTTAGAG
YIL073C	479	CAAACTTCGG	CGGCTAA	TTGCAAATAA
YKR005C	396	TCCTCTTTGG	CGGCTAA	CTCAAGAAGC
YLL030C	434	TAAACTCTGG	CGGCTAA	AAGATAATCA
YLR263W	164	AAGGGTTCAG	CGGCTAA	ATAAACTACG
YML089C	-123	CGAATTCCTC	CGGCTAA	CCAATGAAAT
YMR101C	142	TGGACCTCGG	CGGCTAA	AGCTATACTT
YMR133W	90	ACAAGGTGGG	CGGCTAA	CTTATAAAAA
YMR148W	418	GACCTTAGCG	CGGCTAA	TCAATTCTAC
YNL196C	419	ACTAATCCTT	CGGCTAA	CGTTCTAAAC
YOR351C	146	AGAATGATGG	CGGCTAA	ATTTCGGCGG

Figure 5 Panel A shows in a graphic format the expression profiles of the 90 genes allotted to cluster bin 20 by using K-means clustering. The profiles share the common properties of differential expression during meiosis in both strains, and no differential expression during starvation. Panel B lists 20 genes in this cluster with UME6p cognate DNA target site, its sequence and its location. Its potential importance was deduced due to its overrepresentation in the promoters of the genes in this expression cluster.

the breakdown of the genome into a predefined number of groups or bins, with the search for statistically significant promoter sequence elements. The breakdown of the whole genome into 40 bins by K-means clustering followed by a systematic search for common promoter elements reveals potentially significant elements in about 20% of the bins. For example, the 90 genes assigned to cluster bin 20 have their approximate location in the whole genome clustergram indicated in Figure 4 (arrow). These genes generally have expression patterns in both strains that: (1) are specific to early meiosis, (2) are not strain-dependent and (3) are not expressed in starvation. Figure 5, panel A shows a graphic representation of the expression data for these 90 genes. These observations indicate that this group of genes has a high likelihood of being functionally important to meiosis/sporulation in yeast.

The cluster bin 20 promoters were scanned for all possible 9-mer DNA sequences, and those sequences occurring with a high frequency were noted. Figure 5, panel B lists the identity of 21 of 90 genes in this cluster with the common promoter sequence: 5′-GGCGGCGGCTAA-3′, the location of this sequence element with respect to the translation start, and its sequence context. This particular sequence element occurs in these 90 promoters with a P value of 1.8×10^{-10} above that expected by random chance. A search of the yeast literature quickly reveals that this is the binding site for a well-known yeast transcription factor called Ume6p, which has a key role in the coordination of early meiotic events [10]. The clearly defined expression pattern of the genes in cluster bin 20, together with determination of the cognate transcription factor binding site, provides good evidence in support of any additional conclusions that might be drawn from the data.

The expression data can also be searched with the aim of identifying genes potentially involved in making one strain sporulate quickly and efficiently, and the other to sporulate slowly and relatively inefficiently. One way to do this search is by looking for genes with very low expression profile correlation coefficients between the two meiosis time course experiments, which also have known roles in important cellular events, such as the initiation of meiosis or synaptonemal complex formation. Such a search reveals that certain genes involved in DNA replication (CDC46), recombination (RAD51) and chromosome condensation (SMC1) have radically different expression patterns in the two strains. Notably, it is also known that the slower W303 strain suffers a defect in meiotic DNA replication. Identifying strain differences in this manner provides interesting leads for further research, but unfortunately cannot provide firm answers to specific problems.

In summary, the analysis and clustering of expression data can suggest functional roles for previously uncharacterized genes. These suggestions can be further supported by the determination, or detection, of regulatory sequences in the promoters of genes

clustered together. For example, the Ume6p DNA target site was unambiguously detected as a controlling element in the promoters of early meiotic genes.

Parallel genetics and the yeast lifecycle

Probably the best way to test the functional contribution of a gene to a particular process is by studying the behavior of a strain deleted for that gene. Yeast currently offers a unique opportunity to contrast the data from gene expression studies with the data from the systematic deletion of genes. The degree of correlation between differential expression and the phenotype of the deletion mutant can be assessed. It is sometimes expected that a gene with clear differential expression under a particular set of conditions is particularly important in the cells' adaptation to those conditions: this is an oversimplification [11].

Yeast deletion strains can be created easily with a widely used two-step gene replacement protocol. This process has been systematically applied to every gene in the yeast genome and 24,000 unique haploid and diploid (including the hetero- and homozygous deletion) strains generated. Molecular tags offer a way to follow deletion strains in a population [4,9]. A molecular tag is a 20-bp DNA sequence with the desired melting properties and little homology to other sequences in the genome. Two tags are introduced into each deletion strain at the time the deletion strain is constructed simply by synthesizing the 20-mer into the long oligonucleotide primer used to make the deletion cassette. The

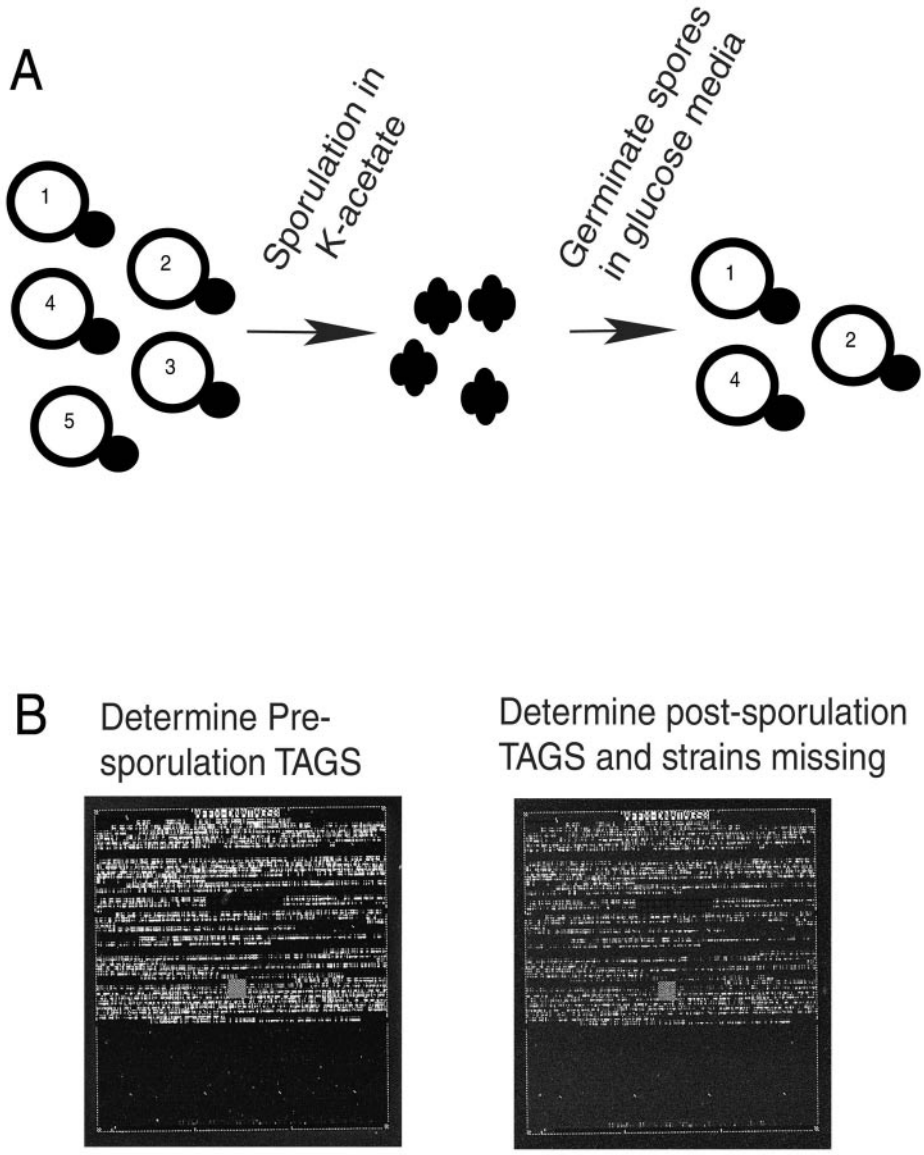


Figure 6 Panel A: The sketch shows a pool of different yeast deletion strains, each carrying a unique molecular identifying tag (e.g., tags 1,2 , 3, 4, 5). Initially, the started pool can be sampled, DNA prepared and the presence of the tags verified. Following sporulation and germination, the pool can be sampled again, DNA prepared, tags amplified and the absent strains determined by loss of the corresponding tags from the pool (tags 5 and 3 missing). Therefore, the deleted genes with the 5 and 3 tags are potential sporulation/germination genes. Panel B: Scans of high-density oligonucleotide tag arrays hybridized with tags from pools of deletion mutants before and after sporulation and germination. The deletion pool under investigation contained ~4000 differently tagged homozygous diploid deletion mutants.

molecular tags, and thus the presence of a deletion strain in a population, can be tracked by first isolating genomic DNA from a population of strains. Each molecular tag is surrounded by common polymerase chain reaction (PCR) priming sites. The tags are then amplified from genomic DNA using primers complementary to the tag priming sites. The PCR-amplified tags are detected by hybridization to a high-density array containing the tag complements. Unlike the expression array described earlier, this array does not contain genomic yeast sequences, only tag sequences, and thus could be used to track tags in any organism containing the molecular tags.

These genetic tools can be used to look in parallel at the phenotypes of the deletion strains with respect to sporulation and germination (Figure 6, panel A). This is particularly interesting because it allows isolation of both new sporulation and germination mutants in a relatively straightforward manner. Once a mutant has been observed disappearing from the pool, its phenotype can be verified by examining a fresh aliquot of the strain obtained from the stock in the freezer. Very few mutants specifically deficient in germination are currently known, a symptom of the fact that yeast genetics is rather difficult with unviable spores. The pooled deletion strains can be cultured as outlined in Figure 6, panel A and samples taken for genomic DNA preparation and PCR analysis at predetermined stages in the lifecycle. The results of the hybridization of the recovered tags to an array bearing the tag complements are shown in the scans in Figure 6, panel B. Each one of the fluorescent features corresponds to a molecular tag sequence and hence reveals information concerning the presence or absence of a particular strain in the pool. The tag populations in the pool pre- and post-sporulation/germination have been compared using specially written computer scripts and ~300 genes identified as functionally important to sporulation/germination (R.M.W. and A. Deutschbauer, manuscript in preparation). Importantly, many genes when deleted were found to have little or no effect on the ability of yeast to undergo sporulation and germination. These included genes scored as differentially transcribed in the expression study described above. Together, these observations support the idea that expression data alone are limited in their ability to functionally classify genes.

Conclusions

The combination of classical genetics, allelic variation scanning, expression analysis and parallel genetics is providing many new insights into complex cellular processes such as meiosis and sporulation in yeast. The challenge now is to try and use these data to build a real understanding of the actual biochemical mechanisms underlying these processes.

Acknowledgements

This research was carried out in the laboratories of R.W. Davis (Stanford University) and R.E. Esposito (University of Chicago). M. Primig, E. Winzeler and A. Conway are among co-contributors to this work. R.M.W. is supported by grant 1RO1HG01633 (awarded to R.W.D.). The yeast deletion strains were produced by the multiple laboratories involved in the international yeast deletion consortium.

References

1 Cho RJ, MJ Campbell, EA Winzeler, L Steinmetz, A Conway, L Wodicka, TG Wolfsberg, AE Gabrielian, D Landsman, DJ Lockhart and RW Davis. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 2: 65–73.
2 Chu S, J DeRisi, M Eisen, J Mulholland, D Botstein, P Brown and I Herskowitz. 1998. The transcriptional program of sporulation in budding yeast. *Science* 282: 699–705.
3 Eisen M, P Spellman, P Brown and D Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863–14868.
4 Giaever G, DD Shoemaker, TW Jones, H Liang, EA Winzeler, A Astromoff and RW Davis. 1999. Genomic profiling of drug sensitivities *via* induced haploinsufficiency. *Nat Genet* 21: 278–283.
5 Kupiec M, B Byers, R Esposito and A Mitchell. 1997. Meiosis and Sporulation in *Saccharomyces cerevisiae*. Cold Spring Harbour Laboratory Press, pp. 889–1036.
6 Lockhardt D, H Dong, M Byrne, M Follettie, M Gallo, M Chee, M Mittmann, C Wang, M Kobayashi, H Horton and E Brown. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14: 1675–1680.
7 Primig M, RM Williams, EA Winzeler, GG Tevzadze, AR Conway, SY Hwang, RW Davis and RE Esposito. 2000. The core meiotic transcriptome in budding yeasts. *Nat Genet* 26: 415–423.
8 Ross-Macdonald P, P Coelho, T Roemer, S Agarwal, A Kumar, R Jansen, K Cheung, A Sheehan, D Symoniatis, L Umansky, M Heidtman, F Nelson, H Iwasaki, K Hagers, M Gerstein, P Miller, G Roeder and M Snyder. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402: 413–418.
9 Shoemaker DD, DA Lashkari, D Morris, M Mittmann and RW Davis. 1996. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat Genet* 14: 450–456.
10 Strich R, RT Surosky, C Steber, E Dubois, F Messenguy and RE Esposito. 1994. UME6 is a key regulator of nitrogen repression and meiotic development. *Genes Dev* 8: 796–810.
11 Winzeler E and R Davis. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901–906.
12 Winzeler E, D Richards, A Conway, A Goldstein, S Kalman, M McCullough, J McCusker, D Stevens, L Wodicka, D Lockhart and R Davis. 1998. Direct allelic variation scanning of the yeast genome. *Science* 281: 1194–1197.
13 Wodicka L, H Dong, M Mittmann, M-H Ho and DJ Lockhart. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 15: 1359–1367.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

1. Ref. [5]: Please supply publisher location.
2. Proposed running head: Yeast lifecycle and DNA.
3. In the Figure 4 caption, Dentrogram was changed to Dendrogram. Is this correct?