

Whole genome genetic-typing in yeast using high-density oligonucleotide arrays

E. A. WINZELER^{1*}, B. LEE, J. H. McCUSKER² and R. W. DAVIS

¹Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305-5307, USA.

²Department of Microbiology, 3020, Duke University Medical Center, Durham, NC 27710, USA.

SUMMARY

Genome sequence information in combination with new technologies has allowed researchers to approach genetic problems in new ways. High-density oligonucleotide arrays were used to probe the genome content of the yeast *Saccharomyces cerevisiae*. We show that these arrays, containing oligonucleotides complementary to the sequenced strain of *S. cerevisiae*, can be used to identify open reading frames that are missing or present in higher or lower copy number in related isolates of *S. cerevisiae*. We apply this method to the characterization of the genome of a strain derived from a clinical isolate of *S. cerevisiae*. Our results show that the telomeres are the regions with the most variability between the two strains.

Key words: oligonucleotide arrays, telomeres, yeast, hybridization.

INTRODUCTION

One of the issues facing those researchers who find themselves in the post-genome era is the question of sequence diversity. Barring any large decrease in costs, in few cases (notably that of *Escherichia coli* strain O157, and *Mycobacterium tuberculosis* clinical isolate CSU93,) will more than one strain of a particular species be sequenced. However, there may be thousands of different strains of a particular organism that are clinically or economically important. Methods are needed that allow researchers to determine on a wholesale level what the differences are between a sequenced strain or individual and other members of the species and where these differences lie.

An emerging and powerful method for monitoring genomes is array-based hybridization. In this massively parallel technique, miniature, ordered collections of nucleic acid probes are synthesized on (Pease *et al.* 1994) or attached to a solid surface (Shalon, Smith & Brown, 1996). The same sort of data that traditionally have been collected using other hybridization methods, such as southern and northern blots, can be acquired with the advantage that many thousands of different probes can be tested simultaneously. The arrays are thus ideal tools for an era in which thousands of gigabases of sequence information are available for experimental interrogation. The use of arrays for monitoring gene expression has been described (Schena *et al.* 1995; Lockhart *et al.* 1996; DeRisi, Iyer & Brown, 1997; Wodicka *et al.* 1997; Cho *et al.* 1998). In this paper we show that the arrays can also be used to monitor genome content; that they are able to detect the

absence of a particular DNA sequence in a strain, and that they can be used to detect changes in the copy number of this piece of DNA in the same strain. We use this method to catalogue some of the genomic differences between two strains of yeast: S96, a variant derived from the sequenced strain, and YJM789, a strain derived from a clinical isolate.

MATERIALS AND METHODS

High density oligonucleotide arrays were purchased from Affymetrix. Yeast cells were grown in YEPD to late log phase at 30 °C. Genomic DNA was purified using Qiagen genomic DNA 100 µg columns according to the manufacturer's protocol. Zymolyase and protease digestion times were extended from 30 to 45 min. DNA was re-suspended in 400 µl TE, reprecipitated, and re-suspended in 30 µl deionized H₂O. Yeast genomic DNA (10 µg) was digested in 0.15 Units DNase I (Gibco BRL PCR grade) in 1X One-Phor-All buffer (Pharmacia) containing 1.5 mM CoCl₂ for 5 min at 37 °C. The reaction was stopped by heating the samples to 100 °C for 15 min. Digestion was checked by examining 1 µl of the reaction product on a 2% agarose gel containing a 1:10000 dilution of SYBR-II green (Molecular Probes, Eugene, OR). The procedure was repeated if the majority of the product was not digested to a size of less than 100 bases (it was observed that the reproducibility of the reaction was highly sensitive to contaminants in the DNA preparation, such as EDTA). The DNA fragments were labelled by incubating the samples with 25 U terminal transferase (Boehringer Mannheim) and 1 nmole Biotin-N6-ddATP (NEN) for one hour at 37 °C. The entire sample was hybridized to the array in 200 µl volume containing 6X SSPE (Accugene), 0.005% Triton-X

* Author to whom correspondence should be addressed.

100 detergent, 20 μg fragmented denatured Salmon Sperm DNA (Gibco-BRL) and 1 nmole of a 3'-biotin control oligonucleotide that hybridizes to the border features on the array. Samples were heated to 100 °C for 10 min, and then cooled on ice before being applied to the array. Samples were hybridized for 2 h at 42 °C. The arrays were washed, stained with phycoerythrin-streptavidin (Molecular Probes) and scanned at an emission wavelength of 560 nm at 7.5 μM resolution using an Affymetrix GeneChip Scanner as previously described (Wodicka *et al.* 1997).

Grids were aligned to the scanned images using the known feature dimensions of the array. The hybridization intensities for each of the elements (both PM and MM) in the grid were determined by the 75th percentile method in the Affymetrix GeneChip[™] software package. The average of PM-MM intensity differences were calculated for all probes for every gene with probes on the array as previously described (Wodicka *et al.* 1997).

RESULTS

Oligonucleotide arrays were designed for studying gene expression in yeast (Wodicka *et al.* 1997). These manufactured arrays contain multiple 25 base oligonucleotide probes for each annotated open reading frame in the genome (Fig. 1). Using both empirical evidence and theoretical models (Lockhart *et al.* 1996) probe sequences were chosen to have good hybridization properties, and to be as unique as possible when compared to the completed yeast genome sequence. The probes are synthesized in a high-density format using photo-sensitive oligonucleotide chemistry and photolithography (Fodor *et al.* 1991; Pease *et al.* 1994). In addition to having probes complementary to each annotated gene (designated PM probes for perfect match), the arrays also contains probes that have a single base mismatch (MM) in the central region of the probe relative to the perfect match probe (Fig. 1 A). The hybridization intensity from the MM probe serves as a local background control for the PM probe (Lockhart *et al.* 1996). When fluorescently labelled RNA or cDNA is hybridized to these arrays, the abundance of a particular RNA molecule in the cell can be determined within a factor of two. The arrays are sensitive enough to detect less than one copy of an mRNA molecule per mammalian cell, and one mRNA molecule per 20 cells in yeast (Lockhart *et al.* 1996; Wodicka *et al.* 1997).

It has been shown that genomic DNA or cDNA can be hybridized to the arrays as well as RNA (Cho *et al.* 1998; Wodicka *et al.* 1997). To test whether these arrays could be used to detect deletions in a genome, DNA samples from two strains of yeast were prepared and hybridized to a partial genome array containing probes to yeast chromosomes V and

VIII. Both yeast isolates were derived from S288C, the sequenced strain of yeast and the one for which the array was designed. However, one strain carried a deletion in a single open reading frame, YEL060C, located on chromosome V. It was predicted that all regions on the array would show hybridization except the region containing probes to the deleted gene. After hybridization the data were analysed. The hybridization intensity (average difference PM-MM) for every gene on the array for the YEL060C deletion strain and its isogenic parent (YJM826) are plotted in Fig. 2A. The data showed roughly equivalent hybridization intensity values for the two strains for all genes probed by the array, except YEL060C. The presence or absence of a gene was determined by calculating whether the PM hybridization signal for the majority of probes for that gene were higher than that of the MM control. Using these criteria, sequences from all genes on chromosomes V and VIII were detected in both strains, except for those from YEL060C, which registered as absent in the deletion strain and present in its parent.

Next we tested if sequence copy number changes could be detected using the arrays by hybridizing different molar amounts of each of the two genomes, simultaneously. DNA from the deletion strain was mixed with DNA from the parent strain using the ratios shown in Fig. 2B. The DNA samples were prepared and hybridized to the arrays. The hybridization intensity (PM-MM) for the probes on the array corresponding to the deleted gene were computed and are plotted against the amount of deletion strain DNA present in the mix (Fig. 2B). The results clearly show a linear arrangement between the hybridization intensity (average difference) and the copy number of the gene, suggesting that increases or decreases in gene dosage could be easily detected with the arrays.

Finally, this method was used to examine the degree of genetic heterogeneity in a natural isolate of yeast. A strain of yeast was selected that was a haploid segregant of a clinical isolate taken from the lung of an AIDS patient. This yeast strain (YJM789) has the ability to grow at physiological temperatures (37 °C), and to undergo colony morphology switching; traits which may contribute to its ability to infect immuno-compromised individuals (McCusker *et al.* 1994a, b). As this strain was derived from a very different environment relative to the sequenced strain, we anticipated that differences in the genome might be evident.

YJM789 DNA was prepared and hybridized to a set of arrays containing probes to the entire genome. This array type contains 20 or more oligonucleotide probes for every annotated open reading frame. Excluding the rDNA and *CUP1* repeats, the largest gap is 41 325 bases wide at position 510 000 on Chromosome XII. The resulting hybridization pattern was examined and compared

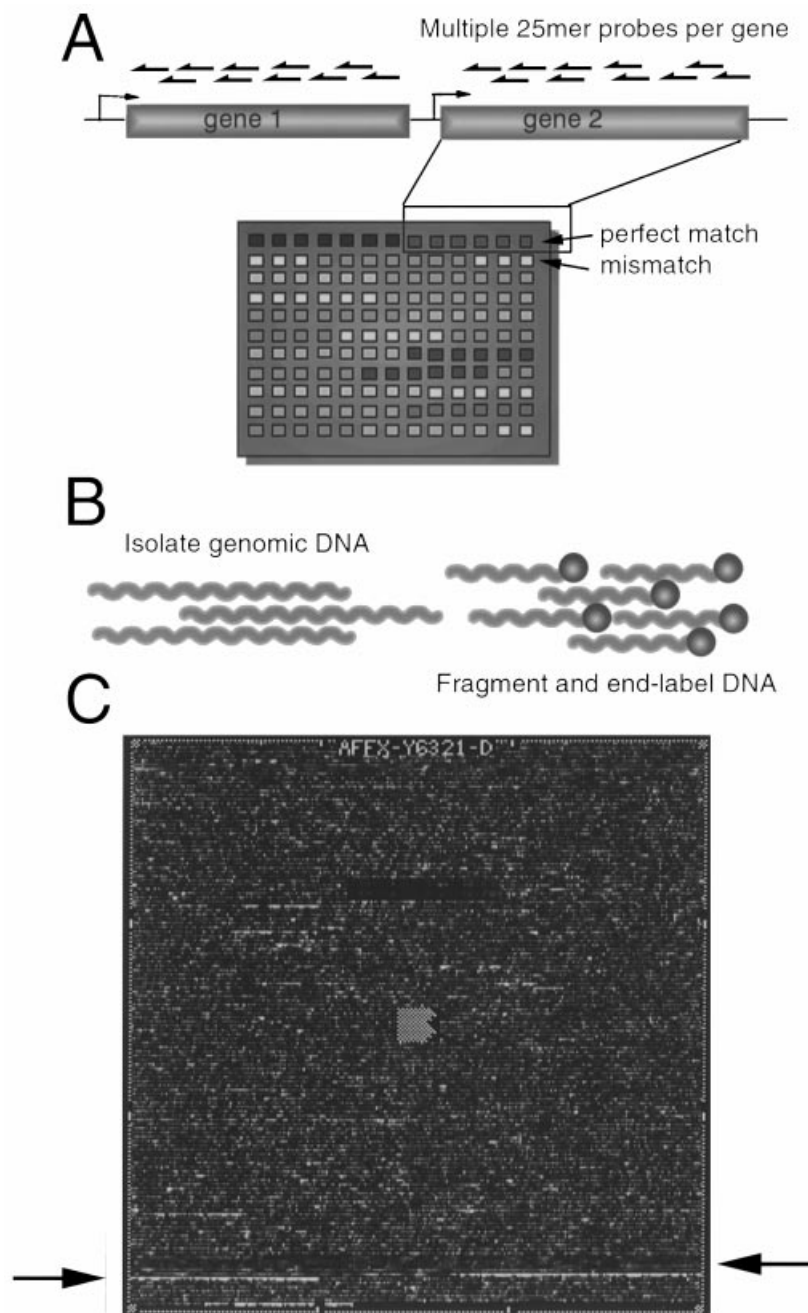


Fig. 1. A. Genome analysis using high-density arrays. 20 or more 25 mer oligonucleotide probes (small arrows) were selected from the completed yeast sequence for each annotated open reading frame in the yeast genome. These oligonucleotide probes were then synthesized on a glass surface. The location and identify of each different probe group on the array is known and each group occupies a $50 \mu\text{M}$ area containing about 10^7 molecules each. Up to 65 000 different probe groups can be synthesized on a single array. In addition to sequences designed to be perfectly complementary to yeast coding sequence (PM probes), a probe containing a single base mismatch (MM) is synthesized in a physically adjacent position. Probes complementary to intergenic regions, which are relatively small in yeast (less than one-third of the genome is noncoding) were not included on the array. B. Experimental protocol: DNA is isolated from different strains of yeast and fragmented using *Dnase I* as previously described (Winzeler *et al.* 1998). Fragmentation improves the hybridization properties of the target, and increases the number of ends available for labeling. The fragments are end-labeled using terminal transferase and biotin-ddATP and then hybridized to the array. After hybridization the arrays are washed, stained with a streptavidin phycoerythrin conjugate, and then scanned using a modified laser confocal scanning device. C. A scanned image of an array. Not every probe hybridizes with the same efficiency but the pattern is reproducible. Arrows show where probes to genes that were not included in the hybridization mix are located (usually controls derived from bacterial sequences) or where probes to genes having more than one copy (ribosomal proteins and so forth) in the genome are located.

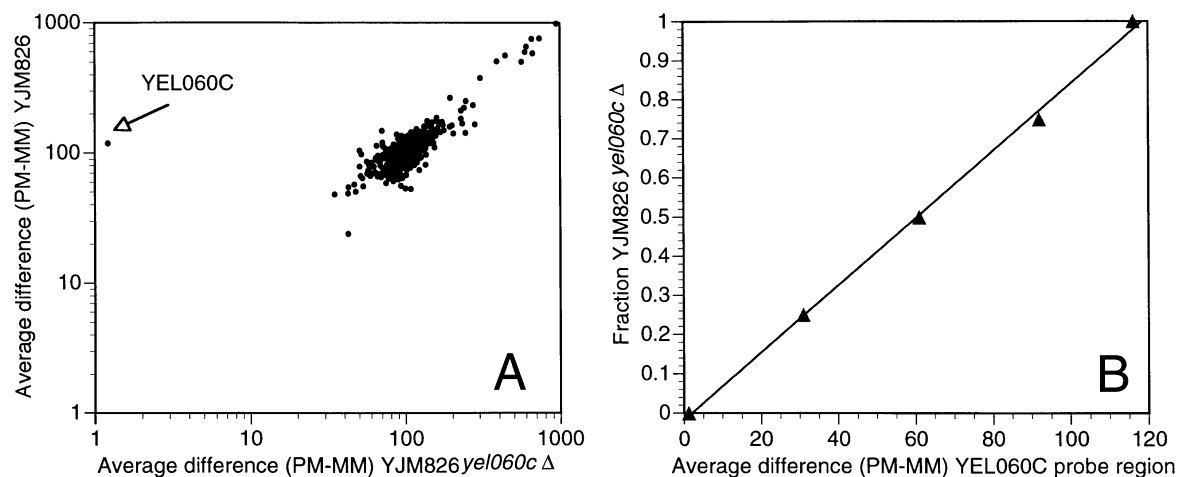


Fig. 2. Detecting DNA deletions in yeast. A. Analysis of hybridization patterns for two arrays. Two partial genome arrays containing probes to yeast chromosomes V and VIII were separately hybridized with DNA from YJM826 and a YJM826 derivative carrying a deletion in YEL060C as described in Fig. 1. After hybridization and scanning a grid was fit to the array images using the known dimensions of the array. The hybridization intensity was computed for each group of oligonucleotide probes on the array. For each gene, the average PM-MM signal intensity difference was calculated for all probes pairs giving a signal above background (PM > MM). Integrated PM-MM data for the 405 probed genes on the chromosome V and VIII array is shown for the two strains. B. Detecting copy number changes using high-density arrays. DNA from YJM826 *yel060CΔ* was mixed with YJM826 DNA at the ratios shown. The hybridization intensity was calculated by averaging the PM-MM signal difference for all probes pairs that were complementary to the YEL060C open reading frame (120) and whose PM/MM signal ratio was greater than background.

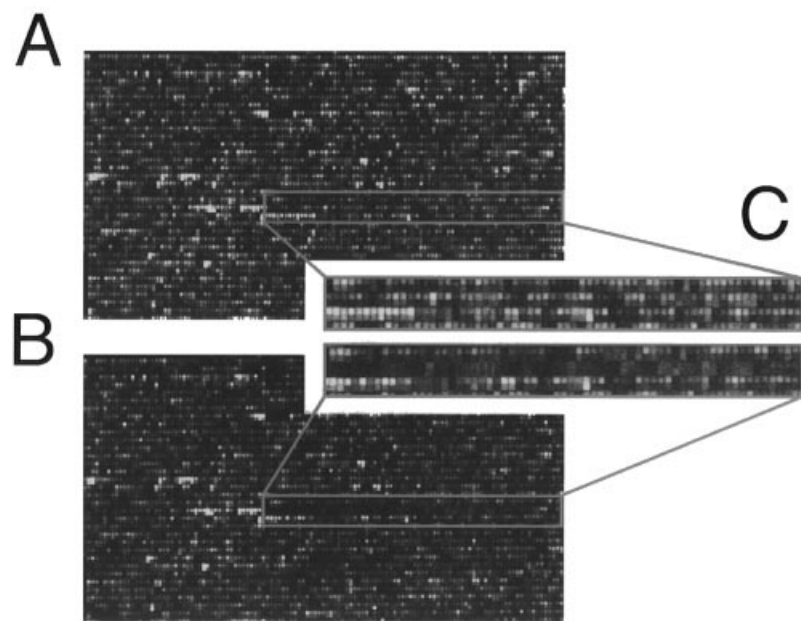


Fig. 3. A. Comparison of array images hybridized with DNA from a yeast strain (S96) isogenic with S288C, and a clinical isolate of *S. cerevisiae*, YJM789. The samples were hybridized to a set of high-density oligonucleotide arrays containing probes to the entire genome. After hybridization the arrays were stained with a streptavidin-phycoerythrin conjugate and then scanned with a modified confocal laser scanning device. C. A close-up of a portion of the array containing probes complementary to the YOL162W, YOL163W, YOL164W region.

to an array that had been hybridized with DNA from an S288C-derived strain, S96 (Fig. 3). Fractions of the two arrays are shown in Fig. 3A and 3B. Regions on the array exhibiting little or no specific fluorescent signal were identified when the two images were viewed. Fig. 3C shows a close-up of the region of the

array containing probes to the YOL162W, YOL163W and YOL164W open reading. A list of the 58 genes (almost 1% of the 6200 annotated open reading frames) that were visually determined to be completely absent is given in Table 1. Many of these open reading frames have no known function.

Table 1. Deleted open reading frames YJM789

ORF	Gene	Brief ID
YAL055W	—	Hypothetical protein
YAL020C	<i>ATS1</i>	Alpha-tubulin suppressor
YAR027W	<i>FUN55</i>	Strong similarity to YAR028w, YCR007c, YGL053w, YAR031w, FUN59p and YGL051w
YAR028W	—	Strong similarity to Fun55p, YGL053w, YCR007c, YAR031w, Fun59p and YGL051w
YAR029W	<i>FUN57</i>	Strong similarity to YAR031w, YGL053w, Fun55p, Fun59p and YGL051w
YAR030C	—	Hypothetical protein
YAR031W	—	Strong similarity to YGL053w, Fun59p, YGL051w, Fun55p, and YCR007c
YAR033W	<i>FUN59</i>	Strong similarity to YGL051w, YGL053w, YAR031w, Fun55p, YAR028w and YCR007c
YAR050W	<i>FLO1</i>	Cell wall protein involved in flocculation
YAR053W	—	Hypothetical protein
YAR060C	—	Strong similarity to hypothetical protein YHR212c
YAR061W	—	Putative pseudogene
YCR004C	<i>YCP4</i>	Strong similarity to <i>S. pombe</i> protein obr1
YDL248W	<i>COS7</i>	Strong similarity to subtelomeric encoded proteins
YER186C	—	Weak similarity to hypothetical protein YMR316w
YER188W	—	Hypothetical protein
YFL063W	—	Strong similarity to subtelomeric encoded proteins
YFL046W	—	Weak similarity to middle part of <i>C. elegans</i> myosin heavy chain A
YFR057W	—	Weak similarity to Cha4p
YGL259W	—	Pseudogene
YGL053W	—	Strong similarity to hypothetical proteins YAR031w, YGL051w, YAR028w, YAR033w and YCR007c
YGL052W	—	Questionable ORF
YGL051W	—	Strong similarity to YAR033w protein
YGR287C	—	Strong similarity to maltase
YGR288W	<i>MAL13</i>	Maltose pathway regulatory protein
YGR289C	<i>AGT1</i>	General alpha-glucoside permease
YGR290W	—	Hypothetical protein
YGR295C	<i>COS6</i>	Strong similarity to subtelomeric encoded proteins
YHL048W	<i>COS8</i>	Strong similarity to subtelomeric encoded proteins
YHL047C	—	Similarity to <i>C. carbonum</i> toxin pump
YHL045W	—	Strong similarity to subtelomeric encoded proteins
YHL008C	—	Similarity to <i>M. formicicum</i> formate dehydrogenase
YHR054C	—	Weak similarity to YOR262w
YHR139C-A	—	Hypothetical protein
YIL082W	—	—
YIL082W-A	—	—
YIL080W	—	Ty3-2 orf C fragment
YIR042C	—	Weak similarity to <i>B. licheniformis</i> hypothetical protein P20
YIR043C	—	Putative pseudogene
YJL219W	<i>HXT9</i>	Hexose transport protein
YJL218W	—	Strong similarity to <i>E. coli</i> galactoside O-acetyltransferase
YJL217W	—	Hypothetical protein
YJL216C	—	Strong similarity to Mal62p
YJL114W	—	—
YJL113W	—	—
YLL063C	—	Hypothetical protein
YLR155C	<i>ASP3A</i>	L-asparaginase II
YLR156W	—	Identical to hypothetical proteins YLR161w and YLR159w
YLR157C	<i>ASP3B</i>	L-asparaginase II
YLR158C	<i>ASP3C</i>	L-asparaginase II
YLR159W	—	Identical to hypothetical proteins YLR161w and YLR156w
YLR160C	<i>ASP3D</i>	L-asparaginase II
YLR161W	—	Identical to hypothetical proteins YLR156w and YLR159w
YLR162W	—	Hypothetical protein
YOL164W	—	Similarity to <i>Pseudomonas</i> alkyl sulfatase
YOL163W	—	Similarity to <i>P. putida</i> phthalate transporter
YOL162W	—	Strong similarity to hypothetical protein YIL166c
YOR153W	<i>PDR5</i>	Pleiotropic drug resistance protein

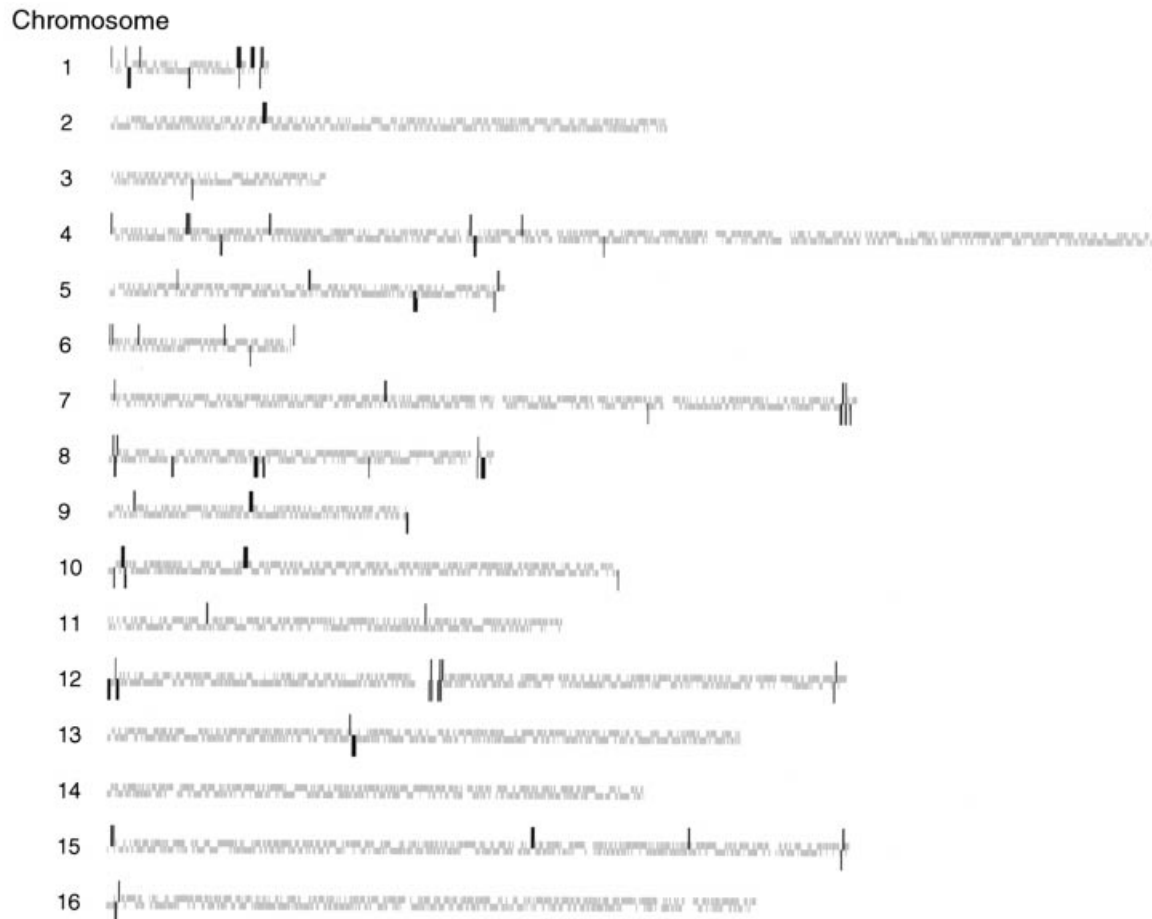


Fig. 4. A genome-wide map of (i) genes with increased copy number and (ii) genes with decreased copy number or which are deleted in YJM789. Open reading frames encoding by the Watson and Crick strands are shown above and below the centre line for each chromosome, respectively. In addition to those genes determined to be absent in YJM789 (listed in Table 1), genes whose hybridization intensity (average PM-MM) increased (through a putative duplication event) by more than a factor of 2.5 or whose hybridization intensity (average PM-MM) decreased by more than a factor of 2.5 (through a partial disruption of the gene, or in the cases of repetitive elements, through a drop in copy number) are indicated in black and by increased size. The position and length of all other genes are shown in light gray. The genome coordinates are derived from SGD (Cherry *et al.* 1997).

Surprisingly, the S288C strain is known to be unable to ferment maltose (Charron, Dubin & Michels, 1986), even though it has several genes predicted to encode maltase and its regulatory subunits. Hybridization to probes for some of these, apparently non-functional genes, was not detected with the YJM789 sample. As YJM789 can ferment maltose, these data suggest that the genes are not missing, but modified. All three copies of the cell wall asparaginase (*ASP3*) were missing. A multidrug resistance pump, *PDR5*, also registered as absent.

To determine whether the deleted genes mapped to specific regions of the genome, the locations of the deleted genes as well as those genes whose hybridization intensity showed a greater than 2.5 fold change between the two strain (up or down), were determined and plotted (Fig. 4). Of these changes, 55% were located within 50 kb of the telomeres of

the chromosomes (versus 12% of all genes). These results are not surprising since the subtelomeric regions of chromosomes from many eukaryotes are known to be polymorphic. However, in a number of cases, deletions were also observed in regions located well away from telomeres.

CONCLUSIONS

Our analysis shows that high-density oligonucleotide arrays offer a rapid way to identify some of the genomic differences between strains. These differences may prove an excellent way to both identify and characterize different strains or isolates. In addition many of these differences may be physiologically relevant. For example, the deletion of the gene encoding the multidrug resistance pump, *PDR5*, has been shown to be the cause of the

YJM789's strain increased sensitivity to the drug cycloheximide (Winzeler *et al.* 1998). The type of omission may allow us to say something about the environment from which a yeast strain was originally isolated. For example, since some of the sugar transporters in the S288c genome are missing in YJM789, one might infer that the S288c genome is rich in genes encoding these enzymes. S288c is mostly derived from strain EM93, which was isolated from a rotting fig near Merced California (Mortimer & Johnston, 1986), presumably an environment rich in sugars. Clearly more strains will need to be examined in order to see if such inferences are further supported.

Many of the genes identified as absent were located in the subtelomeric regions of the chromosomes. The importance of these regions is underscored by the fact that many commercially relevant yeast genes, such as those encoding sucrases, maltases and flocculation proteins, exist in multiple copies and are found at multiple subtelomeric locations. Finally, the significance of subtelomeric regions is not limited to yeast. The *var* genes, which are important for antigenic variation in *Plasmodium falciparum* (Su *et al.* 1995), often are located in the polymorphic subtelomeric regions of this organism's chromosomes (Fischer *et al.* 1997; Rubio, Thompson & Cowman, 1996). It has been proposed that rearrangements of these genes may represent a mechanism by which the parasite might evade the host's immune system (Corcoran *et al.* 1988; de Bruin, Lanzer & Ravetch, 1994; Lanzer, de Bruin & Ravetch, 1993). Thus, both organisms may use recombination between heterologous chromosome as a way to adapt to new environments.

For this analysis, only relatively large deletions were detected. However, because multiple oligonucleotide probes exist per gene, using a more sophisticated analysis, deletion or rearrangement breakpoints within a gene might be found. Regardless, the method describes a way to quickly determine some of the differences between two related strains. Its disadvantage is that it clearly cannot be used to identify those sequences that are absent in the sequenced strain and present in the alternate strain. The challenge will be to use other approaches such as subtractive cDNA hybridization, pulsed field gels, or other methods to fish out those novel sequences. Finally, the biggest hurdle will be to correlate the observed genomic differences with those phenotypic attributes that allow heterogenous strains to be distinguished.

ACKNOWLEDGEMENTS

We thank members Lisa Wodicka and David J. Lockhart for helpful advice and materials. Elizabeth Winzeler is supported by the John Wasmuth fellowship in Genomic Analysis (HG00185-01). This work was supported by NIH grant 1R01 HG01633.

REFERENCES

- CHARRON, M. J., DUBIN, R. A. & MICHELS, C. A. (1986). Structural and functional analysis of the MAL1 locus of *Saccharomyces cerevisiae*. *Molecular and Cell Biology* **6**, 3891–3899.
- CHERRY, J. M., BALL, C., WENG, S., JUVIK, G., SCHMIDT, R., ADLER, C., DUNN, B., DWIGHT, S., RILES, L., MORTIMER, R. K. & BOTSTEIN, D. (1997). Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**, 67–73.
- CHO, R. J., CAMPBELL, M. J., WINZELER, E. A., STEINMETZ, L., CONWAY, A., WODICKA, L., WOLFSBERG, T. G., GABRIELIAN, A. E., LANDSMAN, D., LOCKHART, D. J. & DAVIS, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell* **2**, 65–73.
- CORCORAN, L. M., THOMPSON, J. K., WALLIKER, D. & KEMP, D. J. (1988). Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell* **53**, 807–813.
- DE BRUIN, D., LANZER, M. & RAVETCH, J. V. (1994). The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. *Proceedings of the National Academy of Sciences, USA* **91**, 619–623.
- DERISI, J. L., IYER, V. R. & BROWN, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
- FISCHER, K., HORROCKS, P., PREUSS, M., WIESNER, J., WUNSCH, S., CAMARGO, A. A. & LANZER, M. (1997). Expression of *var* genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. *Molecular and Cell Biology* **17**, 3679–3686.
- FODOR, S. P. A., READ, J. L., PIRRUNG, M. C., STRYER, L., LU, A. T. & SOLAS, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773.
- LANZER, M., DE BRUIN, D. & RAVETCH, J. V. (1993). Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome. *Nature* **361**, 654–657.
- LOCKHART, D. J., DONG, H., BYRNE, M. C., FOLLETTIE, K. T., GALLO, M. V., CHEE, M. S., MITTMANN, M., WANG, C., KOBAYASHI, M., HORTON, H. & BROWN, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* **14**, 1675–1680.
- MCCUSKER, J. H., CLEMONS, K. V., STEVENS, D. A. & DAVIS, R. W. (1994a). Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. *Genetics* **136**, 1261–1269.
- MCCUSKER, J. H., CLEMONS, K. V., STEVENS, D. A. & DAVIS, R. W. (1994b). *Saccharomyces cerevisiae* virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42 degrees C and form pseudohyphae. *Infection and Immunology* **62**, 5447–5455.
- MORTIMER, R. K. & JOHNSTON, J. R. (1986). Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43.
- PEASE, A. C., SOLAS, D., SULLIVAN, E. J., CRONIN, M. T.,

- HOLMES, C. P. & FODOR, S. P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of the National Academy of Sciences, USA* **91**, 5022–5026.
- RUBIO, J. P., THOMPSON, J. K. & COWMAN, A. F. (1996). The *var* genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO Journal* **15**, 4069–4077.
- SCHENA, M., SHALON, D., DAVIS, R. W. & BROWN, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
- SHALON, D., SMITH, S. J. & BROWN, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research* **6**, 639–645.
- SU, X. Z., HEATWOLE, V. M., WERTHEIMER, S. P., GUINET, F., HERRFELDT, J. A., PETERSON, D. S., RAVETCH, J. A. & WELLEMS, T. E. (1995). The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100.
- WINZELER, E. A., RICHARDS, D. R., CONWAY, A. R., GOLDSTEIN, A. L., KALMAN, S., McCULLOUGH, M. J., McCUSKER, J. H., STEVENS, D. A., WODICKA, L., LOCKHART, D. J. & DAVIS, R. W. (1998). Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197.
- WODICKA, L., DONG, H., MITTMANN, M., HO, M.-H. & LOCKHART, D. J. (1997). Genome-wide expression monitoring in *Sacchaomyces cerevisiae*. *Nature Biotechnology* **15**, 1359–1367.