

# Chapter 4

## Two Flavors of Bulk Segregant Analysis in Yeast

Maitreya J. Dunham

### Abstract

Genetic mapping methods typically rely upon genotyping many individuals in a mapping population. In contrast, bulk segregant analysis looks for biases in genotype in phenotyped pools of segregants. For relatively strong and genetically simple traits, it can be a fast, inexpensive approach. Although it is technically possible to use many genotyping platforms, microarray-based methods are convenient for their genome-wide coverage, ease of use, and quantitative output. Also, precise knowledge of polymorphic sites is not required. I present two methods for bulk segregant analysis using microarrays, one based on hybridization differences between polymorphisms, and the other using an enzymatic method for enriching identical by descent segments of the genome. The first method requires specialized array platforms, while the second, genomic mismatch scanning (GMS), is compatible with any microarray. Although the methods presented are with yeast, most steps are equivalent for other organisms.

**Key words:** Bulk segregant analysis, Genomic mismatch scanning, Pooled mapping

---

### 1. Introduction

Bulk segregant analysis is a pool-based mapping approach that was first developed (1) and implemented using modern genomic technologies (2, 3) in plant systems and later in yeast (4, 5). Segregants from a mapping cross are binned into phenotype pools. These entire pools are then genotyped en masse. Regions of the genome that segregate independently will maintain a 50% allele frequency in all pools. DNA linked to the trait of interest will show a bias towards one parental genotype in one phenotype pool and the other parental genotype in the other pool. Pool genotype frequencies can be obtained genome-wide in several ways, including SNP detection arrays, genomic mismatch scanning (GMS), array-based RAD tags, and deep sequencing.

Because bulk segregant analysis examines the average allele frequency of a pool of individuals, the resolution typically does

not match individual genotyping approaches where every additional recombination breakpoint is informative. However, considerably less genotyping is required, making bulk segregant analysis attractive for quickly and cheaply locating strong QTLs or Mendelian traits. In many cases, the resolution is adequate for prioritization of obvious candidate loci. At a minimum, a single cross and a single array can generate quality mapping data.

Two genotyping strategies are presented here, in parts 3.3/3.4 and 3.5/3.6. The genetic manipulations (Subheading 3.1) and DNA preparations (Subheading 3.2) are the same for both protocols.

Subheadings 3.3 and 3.4 describe genotyping and analysis using Affymetrix tiling microarrays (see Fig. 1 for procedure outline), which contain overlapping 25-mers at a four or eight base tile across the yeast genome. Polymorphisms that generate a mismatch at the complementary 25 base sequences in the mapping strains lead to decreases in hybridization at so-called Single Feature Polymorphisms (SFPs). Affymetrix yeast expression arrays are also adequate for bulk segregant analysis and follow the same procedure. Recent reports (6, 7) describe custom Agilent microarrays targeted to specific segregating polymorphic sequence that can also be used similarly (though, obviously, with Agilent-specific hybridization procedures). Calibration hybridizations using the parental strain backgrounds are recommended for this approach.

Subheadings 3.5 and 3.6 use GMS (8–10). DNA from each pool is hybridized to DNA from each parental strain, and a series of enzymatic manipulations remove heterohybrid DNA molecules that are not perfectly matched. This identity by descent fraction is then hybridized to an array. This approach is compatible with any microarray, including home-printed PCR products, and does not require initial calibration hybridizations.

For both methods, prior knowledge of polymorphism location and identity are not required, though sufficient polymorphism distribution and density are important. For tiling microarrays, these can easily be discovered *de novo* or borrowed from prior results. GMS relies on polymorphisms in large PstI fragments as opposed to individually queried SNPs.

Next-generation sequencing approaches have also been developed for bulk segregant analysis (6, 11–13), although these are beyond the scope of this. Subheadings 3.1 and 3.2 would still be applicable to the new methods, as would the post-data-processing analytical methods in Subheadings 3.4 and 3.6. Sequencing holds the promise of allowing both linkage and detection of mutations or polymorphisms simultaneously (11, 13), making it an attractive route for further investigation.

Choices for data analysis methodology are varied and not well explored in the literature. They will also differ markedly depending on the computational resources available. For simplicity, I have included straightforward methods for analyzing and

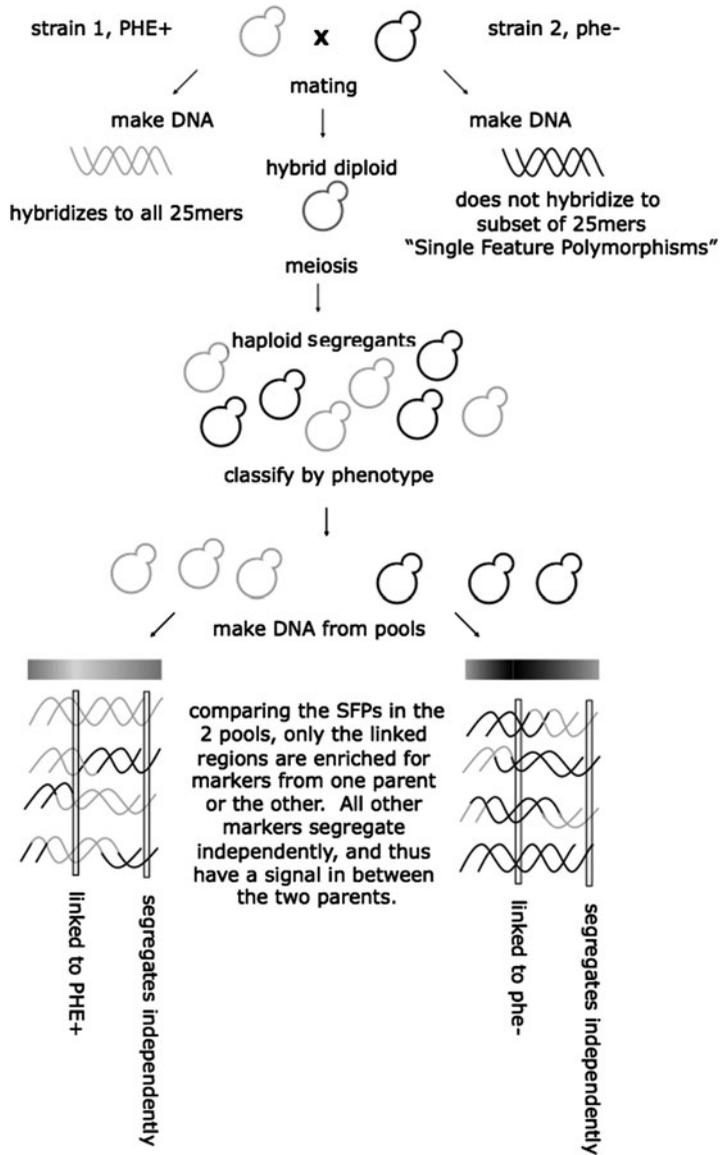


Fig. 1. Overview of bulk segregant using a Mendelian trait PHE+/phe- and Affymetrix array genotyping.

visualizing the data. Because fine-scale mapping will generally be required to confirm results anyway, this reflects a typical use of bulk segregant analysis as a gross mapping method to be coupled with targeted follow-up. For this reason, I recommend performing a positive control mapping cross when first learning the methods, both to ensure the many steps have worked properly and to see what a true positive result looks like amid a somewhat user-specific noise term. I have included several examples of mapping datasets throughout.

---

## 2. Materials

### 2.1. Genetic Cross

1. Standard yeast growth and sporulation media. These should be tailored to the exact strain backgrounds utilized in order to equalize growth and maximize sporulation efficiency.

### 2.2. DNA Preparation

1. Genomic DNA buffer set (Qiagen).
2. Lyticase (Sigma) or zymolyase (Seikagaku Corporation): 1,000 U/mL in water. Make day of use.
3. Proteinase K (Roche): 20 mg/mL in water. Make day of use.
4. RNase A (Sigma): 100 mg/mL in water.
5. Midi or maxi genomic columns depending on desired yield (Qiagen).

### 2.3. Genotyping Pools Using Affymetrix Tiling Microarrays

1. 10× one-phor-all buffer: 100 mM Tris-acetate pH 7.5, 100 mM magnesium acetate, 500 mM potassium acetate.
2. DNase I (Invitrogen): diluted to 0.15 U/μL in 1× one-phor-all buffer.
3. 2% agarose gel in 1× TAE.
4. DNA ladder spanning 50 bp range (NEB).
5. BioArray Terminal labeling kit (Enzo).
6. 12× MES: 1.22 M MES hydrate, 0.89 MES sodium salt in water. Filter sterilize. pH should be between 6.5 and 6.7. Store 4°C in dark. If solution discolors yellow, discard.
7. 2× hybe buffer: 200 mM MES, 2 M NaCl, 40 mM EDTA, 0.02% Tween-20 in water. Store at 4°C in dark.
8. 10 mg/mL salmon sperm DNA (Invitrogen), control oligo B2 (Affymetrix).
9. Affymetrix Yeast tiling array. Arrays intended for expression analysis are also adequate.
10. Wash buffer A: 0.9 M NaCl, 60 mM NaH<sub>2</sub>PO<sub>4</sub>, 6 mM EDTA, 0.01% Tween-20 in water. Filter sterilize. May be reused if filtered after each use.
11. Wash buffer B: 100 mM MES, 0.1 M NaCl, 0.01% Tween-20 in water. Filter sterilize. Store 4°C in dark. May be reused if filtered after each use.
12. 2× stain buffer: 200 mM MES, 2 M NaCl, 0.1% Tween-20 in water. Filter sterilize. Store 4°C in dark.
13. Streptavidin, R-phycoerythrin conjugate solution (SAPE): 10 μg/mL R-streptavidin (Invitrogen), 2 mg/mL BSA, 1× stain buffer in water. Make day of use. Store 4°C in dark until needed.

14. Antibody solution: 0.1 mg/mL goat IgG (Sigma), 2 mg/mL BSA, 3  $\mu$ g/mL biotinylated anti-streptavidin antibody from goat (Vector Labs), 1 $\times$  stain buffer in water. Make day of use. Store 4°C in dark until needed.

#### **2.4. Genotyping Pools Using Genomic Mismatch Scanning**

1. PstI, 10 $\times$  buffer 3, and 100 $\times$  BSA (NEB).
2. For DNA extraction and precipitation: phenol/chloroform 50/50 mix, using buffer-saturated pH 7–8 phenol, chloroform, 3 M NaOAc, and 100% ethanol. Take appropriate precautions using phenol/chloroform.
3. TE pH 8.
4. Dam methylase, 10 $\times$  dam methylase buffer, and SAM (NEB).
5. MboI, DpnI, and Sau3A (NEB).
6. 5 N NaOH, 3 M MOPS buffer, formamide.
7. 2 $\times$  FPERT buffer: 4 M sodium thiocyanate, 20 mM Tris-HCl pH 7.9, 0.2 mM EDTA.
8. Exonuclease III and 10 $\times$  Exo III buffer (Promega).
9. 0.5 M EDTA, 8 M LiCl.
10. Single Stranded Affinity Matrix (SSAM, see Note 1 for more information), 0.45  $\mu$ m filter cartridges (Millipore).
11. Linear acrylamide (Ambion).
12. 10 $\times$  ENH: 500 mM Hepes pH 8, 200 mM KCl, 40 mM MgCl<sub>2</sub>, 10 mM DTT.
13. PPD: 20 mM KPi pH 7.4, 50 mM KCl, 0.1 mM EDTA pH 8. Divide into 225 mL aliquots and freeze. Before use, add 2.5  $\mu$ L 0.1 M DTT and 25  $\mu$ L 10 mg/mL BSA (nonacetylated) to a 225  $\mu$ L aliquot of PPD.
14. MutH, MutL, MutS (USB).

---

### **3. Methods**

#### **3.1. Genetic Cross**

1. Identify two strains with the following characteristics: (a) can be crossed, (b) the cross results in reasonable spore viability, (c) the trait of interest segregates, and (d) the strains are divergent at the nucleotide level throughout the genome. See Note 2 for further information on good strain choices.
2. Cross the strains and isolate segregants by tetrad analysis or random spore analysis. I prefer tetrad analysis for the increased information about the underlying genetic architecture. Because bulk segregant analysis relies on the average genotype of pooled

segregants, increasing pool size offers diminishing returns. Modeling pools of different sizes (data not shown) indicates that simple traits can be reliably mapped with 30 segregants in each phenotype class, and we have obtained reasonable results with as few as 23 in a pool (4). I typically use 50–100. Independent segregants (i.e., not from the same tetrad) are preferred, but not required. It should be noted that other applications of this technique use much larger pool sizes, which may be required for very complex traits. These large populations of segregants may be generated by FACS or counterselection of appropriate traits (5, 6).

3. Phenotype the segregants for the trait of interest and pool them accordingly. In the simplest case of a Mendelian trait, this would generate two pools. Quantitative traits may be mapped using the high and low tails of the phenotype distribution.

### **3.2. DNA Preparation**

1. Equal representation of each segregant is important for accurate genotype frequencies. Grow each segregant in a separate overnight culture, typically individual 18 mM test tubes containing 2.5 mL YPD, though 96-well plates may also be used. Culture conditions should be chosen to ensure approximately equal saturation densities among the segregants.
2. Combine all the segregants with a shared phenotype into a single flask to create each pool.
3. Prepare DNA from each pool and each parental strain. At least 20–40  $\mu\text{g}$  high molecular weight DNA of high purity is desired. I have gotten the most consistent results using spooled DNA made with the Qiagen genomic DNA preparation kit (see Note 3 for further tips).

### **3.3. Genotyping Pools Using Affymetrix Tiling Microarrays**

See Fig. 1 for overview. Before beginning, preheat hybridization oven to 45°C. Prepare heat blocks at 37 and 99°C. I prefer to perform all steps in 1 day, though possible pause points are indicated.

1. Bring 10  $\mu\text{g}$  DNA to 15.8  $\mu\text{L}$  total volume in 10 mM Tris pH 8. Add 2  $\mu\text{L}$  10 $\times$  one-phor-all buffer and 1.2  $\mu\text{L}$  25 mM  $\text{CoCl}_2$ . Mix and spin down.
2. Quickly add 1  $\mu\text{L}$  (0.15 U) DNase I. Flick to mix, touch spin to collect volume, and incubate at 37°C for exactly 5 min. Move to 99°C for 15 min.
3. Run 1  $\mu\text{L}$  on a 2% gel for 50 min at 50 V. Also run a similar amount of undigested DNA to check quality. The remaining reaction may be stored at 4°C, while the gel is run or frozen indefinitely. The DNA should be all in a smear ~50 bp. This result is crucial for the success of the protocol. If the digestion

has not proceeded appropriately, see Note 4 for advice on optimizing this step.

4. Perform end-labeling reaction as directed by the Enzo labeling kit: to the entire volume of fragmented DNA, add 20  $\mu\text{L}$  5 $\times$  reaction buffer, 10  $\mu\text{L}$  10 $\times$   $\text{CoCl}_2$ , 1  $\mu\text{L}$  biotin-ddUTP, 2  $\mu\text{L}$  terminal deoxynucleotide transferase, and 48  $\mu\text{L}$  water to a total of 100  $\mu\text{L}$ . Incubate 37 $^\circ\text{C}$  15 min to 1 h. Ice. Add 5  $\mu\text{L}$  0.2 M EDTA to stop reaction.
5. Once labeling reaction is stopped, add 27  $\mu\text{L}$  water, 150  $\mu\text{L}$  2 $\times$  hybe buffer, 15  $\mu\text{L}$  10 mg/mL BSA, 3  $\mu\text{L}$  10 mg/mL salmon sperm DNA, and 5  $\mu\text{L}$  3 nM control oligo B2. Incubate 99 $^\circ\text{C}$  5 min.
6. During incubation, load array with 200  $\mu\text{L}$  1 $\times$  hybe buffer: first, place a 200- $\mu\text{L}$  filter pipette tip in one septum of the array. Holding the array vertically with the vent pipette tip towards the top, slowly pipette in the solution through the other septum. Remove tips. An air bubble should be present in the chamber, and no solution should be present in the vent tip. Place array in holder and load into hybridization oven. Rotate at 60 rpm for 10 min.
7. Returning to probe solution, incubate at 45 $^\circ\text{C}$  for 5 min. Spin at max speed in a microcentrifuge for 5 min to pellet debris.
8. Remove buffer from array using a similar strategy as described above. Load 200  $\mu\text{L}$  hybridization mixture, avoiding volume at bottom of tube. Seal septa using tape or stickers.
9. Place array in holder and load in hybridization oven. Incubate at 45 $^\circ\text{C}$  and 60 rpm for 20 h.
10. Prepare wash buffers A and B. For each array, prepare two fresh aliquots of 600  $\mu\text{L}$  SAPE and one 600  $\mu\text{L}$  aliquot antibody solution.
11. Ensure fluidics have been appropriately washed and primed. Perform washes as directed by your particular Affymetrix wash station. In my setup, this is the EukGE-WS2v4 protocol. Once complete, seal septa with correction fluid.
12. Scan array as directed by your Affymetrix scanner setup.

### **3.4. Analysis of Affymetrix Data**

1. Using Affymetrix software, process data from each array to generate CEL files. Extract the normalized intensity data at each feature (see Note 5 for a special caution about array geometry which may affect normalization).
2. If you have run calibration arrays with each parental strain, use the SNP Scanner package (14) or another method to identify polymorphisms between the two strains (SFPs) which show distinct differences in hybridization intensity. This may not be necessary if the linkage signal is strong enough.

3. Recover the intensity data from features covering the polymorphic locations between the strains.
4. Calculate the  $\log_2$  ratio comparing the data from each pool. This value may have some noise in it as the differential hybridization at each SFP varies. If more resolution is required, it may be useful to scale the data at each SFP such that one parental intensity value equals  $-1$  and the other parental intensity equals  $1$ . This scaling also gives an intuitive view of genotype frequency.
5. Plot genotype frequency data in chromosomal order. This is conveniently done using the Integrated Genome Browser available from Affymetrix. IGB can also generate ratios from intensity tracks if desired.
6. Scan chromosome plots for deviations in genotype frequency. Unlinked regions of the genome will segregate independently and be present at 50% representation from each parent in both pools. Therefore, these regions of the genome will have a  $\log_2$ -ratio or scaled genotype frequency of 0. Regions linked to the phenotype will show biased inheritance and will deviate from 0. See Fig. 2 for an example. In general, the highest portion of the peak is the best indicator of the QTL. See Note 6 for anecdotes otherwise.
7. Follow-up gross linkage data with fine mapping methodology of your choice. I have had success using strains from the yeast deletion collection as a source of mapping markers.

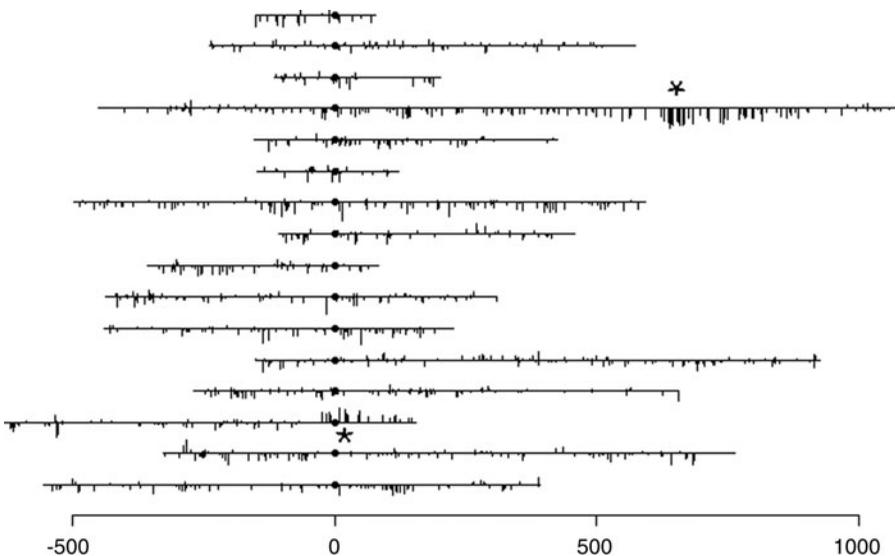


Fig. 2. Example mapping result using Affymetrix microarrays showing a primary locus on chromosome 14 and a modifier locus from the other parent on chromosome 4 (marked with stars). *Small circles* represent centromeres. Figure modified from Chiang (21).

### **3.5. Genotyping Pools Using Genomic Mismatch Scanning**

The GMS protocol consists of many steps which span several days. First, DNA from the parent strains and pools is digested with PstI. Parental DNA is methylated. Methylated parental DNA is hybridized to each pool DNA in a controlled reaction, then digested with methylation-sensitive restriction enzymes to generate hemimethylated heterohybrid DNA. Bacterial mismatch repair enzymes nick this DNA at sites of mismatches, and this nicked DNA is digested with exonuclease and removed from the reactions. The resulting pool DNAs are identical by descent with the matched parent strain and can be competitively hybridized to a microarray to determine the genotype ratio at each PstI fragment.

1. PstI digest a minimum of 40  $\mu\text{g}$  DNA from each parent strain and from each pool: bring DNA to 200  $\text{ng}/\mu\text{L}$  in a reaction with 5  $\text{U}/\mu\text{g}$  PstI, 1 $\times$  NEB Buffer 1, and 1 $\times$  BSA. Incubate 37°C for 2 h. Add 2.5  $\text{U}/\mu\text{g}$  PstI and allow reaction to proceed another 2 h. Digested DNA may be stored at  $-20^\circ\text{C}$  indefinitely.
2. Check 2  $\mu\text{L}$  digested DNA and a similar mass undigested DNA on a 0.8% gel to ensure complete digestion.
3. Extract DNA with phenol/chloroform, followed by chloroform alone. Precipitate DNA using sodium acetate and ethanol precipitation. Dissolve lightly dried pellet in 100  $\mu\text{L}$  TE. Digested DNA may be stored at  $-20^\circ\text{C}$  indefinitely.
4. Methylate parental DNAs: to 20  $\mu\text{g}$  DNA, add 10  $\mu\text{L}$  10 $\times$  dam methylase buffer, 0.25  $\mu\text{L}$  SAM (may be diluted 10 $\times$  if preferred), and 10  $\mu\text{L}$  dam methylase in a total volume of 100  $\mu\text{L}$ . Incubate at 37°C for 4 h, spiking reaction with an additional 0.25  $\mu\text{L}$  SAM every hour. This step may be scaled up if many samples from the same parents will be processed.
5. Extract and precipitate DNA as above.
6. Determine the quality of the methylation by digesting 1  $\mu\text{L}$  pre- and postmethylated DNA with MboI, DpnI, and Sau3A in separate reactions. Sau3A should cut all species; MboI will cut only unmethylated DNA; DpnI will cut only methylated DNA.
7. In this step, DNA from each pool and each parent is hybridized using a controlled method called formamide phenol emulsion reassociation technique (FPERT). Mix 20  $\mu\text{g}$  methylated parent DNA with 20  $\mu\text{g}$  pool DNA in a total volume of 200  $\mu\text{L}$ . This leads to four reactions total for a two-pool mapping population. Add 14  $\mu\text{L}$  5 N NaOH to denature the DNA. Mix every 5 min for 15 min at room temperature. Neutralize reaction with 29  $\mu\text{L}$  3 M MOPS buffer. Check that pH is close to 8.5 by pipetting a small drop onto pH paper. If not, adjust pH. The next steps are most easily performed in 1.5 mL screw-cap tubes. Split each reaction between two tubes. To each, add 32  $\mu\text{L}$  formamide, 45  $\mu\text{L}$  water, and 200  $\mu\text{L}$  2 $\times$  FPERT buffer.

Mix well. Add 150  $\mu\text{L}$  phenol to each tube. If the solution is clear, continue to add phenol in 25  $\mu\text{L}$  increments until solution becomes cloudy. Agitate solution for 24 h at room temperature. Agitation must be sufficient to maintain the emulsion. Extract each tube with 150  $\mu\text{L}$  chloroform followed by extraction with 400  $\mu\text{L}$  chloroform. Add 1 mL ethanol to each tube. Incubate  $-80^\circ\text{C}$  for 30 min. Spin at  $4^\circ\text{C}$  in a microfuge at max speed for 30 min. Carefully remove the supernatant, wash the pellet with 70% ethanol, and allow to air dry briefly. Pellet may be a clear, viscous liquid. If so, dissolve pellet in 0.5 mL 70% EtOH and spin at max speed 10 min. Repeat until viscous pellet is fully dissolved after spin, and all that remains is “typical” salt/DNA precipitate. Dissolve each pellet in 45  $\mu\text{L}$  TE and recombine split reactions. Quantify DNA by measuring A260 of a 1:25 dilution. The spectrum may have a peak at 230 nm and a shift upwards of the 260 nm peak.

8. To select for just the hemimethylated heterohybrid DNA, digest with DpnI and MboI, which will digest the methylated and unmethylated homohybrids: mix 20  $\mu\text{g}$  DNA, 20  $\mu\text{L}$  10 $\times$  NEB buffer 3, 40 U DpnI, and 40 U MboI in 1 $\times$  TE to 195  $\mu\text{L}$ . Incubate  $37^\circ\text{C}$  for 1 h. Ice reaction while preparing the ExoIII mix. For each reaction, add 60  $\mu\text{L}$  10 $\times$  ExoIII buffer and 60 U ExoIII to water for a total volume of 600  $\mu\text{L}$ . Add one 600  $\mu\text{L}$  aliquot ExoIII mix to each digestion reaction, working as quickly as possible. Incubate 15 min at 37 C. Stop reaction with 8  $\mu\text{L}$  0.5 M EDTA and 80  $\mu\text{L}$  8 M LiCl. Add 30  $\mu\text{L}$  SSAM that has been well resuspended before use. Incubate at room temperature for 10 min, mixing every 2 min. Spin down the SSAM using gentle centrifugation in a benchtop minicentrifuge for 1 min. Divide the supernatant between two 0.45  $\mu\text{m}$  filter cartridges. Spin at  $2,000 \times g$  for 3 min. Add 1  $\mu\text{L}$  linear acrylamide and mix well. Add 900  $\mu\text{L}$  ethanol to each tube. Incubate  $-80^\circ\text{C}$  for 30 min. Spin at maximum RPM at  $4^\circ\text{C}$  for 30 min. Wash pellet with 70% ethanol and briefly airdry pellet. Resuspend each pellet in 26  $\mu\text{L}$  TE. Measure DNA concentration.
9. Nick mismatch DNA: bring 1  $\mu\text{g}$  DNA to 25  $\mu\text{L}$  with TE. Add 5  $\mu\text{L}$  10 $\times$  ENH, 3  $\mu\text{L}$  PPD, 0.5  $\mu\text{L}$  100 mM ATP, 10 ng MutH, 100 ng MutL, and 2.5  $\mu\text{g}$  MutS in a total volume of 50  $\mu\text{L}$ . (See Note 7 for more information about enzyme amounts.). Incubate at room temperature for 20 min. Stop reaction by heating to  $80^\circ\text{C}$  for 5 min and then placing on ice.
10. Deplete nicked DNA: mix 5  $\mu\text{L}$  10 $\times$  ExoIII buffer, 45  $\mu\text{L}$  water, 20 U ExoIII. Add to nicked DNA. Incubate at  $37^\circ\text{C}$  for 10 min. Stop reaction with 1  $\mu\text{L}$  0.5 M EDTA and 5  $\mu\text{L}$  8 M LiCl. Add 5  $\mu\text{L}$  SSAM and incubate for 10 min at room temperature, mixing every 2 min. Spin down SSAM for 1 min

in a benchtop centrifuge at  $2,000 \times g$ . Transfer supernatant to a 0.45- $\mu\text{m}$  filter cartridge and spin through at  $2,000 \times g$  for 3 min. Add 1  $\mu\text{L}$  linear acrylamide. Add 240  $\mu\text{L}$  ethanol and incubate at  $-80^\circ\text{C}$  for 30 min. Spin at maximum g force at  $4^\circ\text{C}$  for 30 min. Wash pellet with 70% ethanol and air-dry pellet. Resuspend DNA in 15  $\mu\text{L}$  TE. Check concentration.

- Hybridize DNA to preferred microarray. See Note 5 for advice on array configuration. Although single color microarrays will likely work, the methods described in Subheadings 3.3 and 3.4 are more amenable to that platform. Two-color platforms such as spotted arrays or Agilent arrays allow direct comparison of the parental representation in each pool. For example, assuming a Mendelian trait, the class of segregants with one parental phenotype will have the linked region enriched in the reaction using that parent and depleted in the reaction vs. the other parent, leading to a clear representational difference. Amplification is not typically required if sufficient yield is obtained.

### 3.6. Analysis of GMS Data

- If a two-color platform was used, genotype ratio data are obtained directly. For single color arrays, preprocessing may be required to generate equivalent ratio representation of the data.
- Plot  $\log_2$  ratio data in genome order. I like the Karyoscope function of Java Treeview (15) for this application. Custom coordinates files can be created for any array platform. Deviations in each pool will be accentuated by subtracting the data from the two pools from one another.
- Scan the genome for imbalances in genotype frequency. See Fig. 3 for an example.

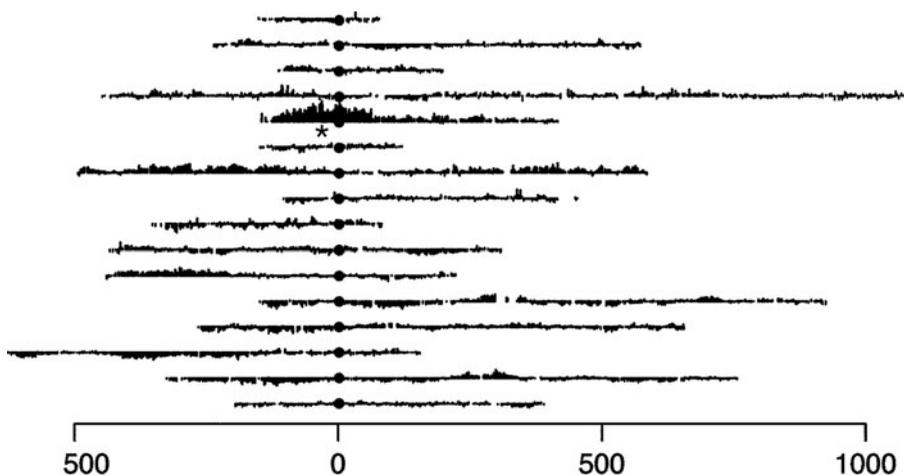


Fig. 3. Example mapping result using GMS and spotted PCR product arrays. Data are graphed across chromosomes using the karyoscope function of Java Treeview (15). *Small circles* represent centromeres. The primary locus is marked with a star.

4. Follow-up gross linkage data with fine mapping methodology of your choice. I have had success using strains from the yeast deletion collection as a source of mapping markers.

---

## 4. Notes

1. SSAM, which was once produced by Clontech and may still be found in some labs, may be substituted with any appropriate ssDNA-specific agent such as BNDC.
2. Validated pairs of strains include S288C derivatives (including true S288C strains as well as patchwork genotypes like W303 and CEN.PK) vs. SK1 (4, 5, 16), Y55 (4), D273-10B (17), RM11 (6, 18), and a variety of wine strains (12). Many other combinations are possible. Though all published examples so far seem to use one S288C-related parent (as illustrated in Fig. 1), pairs where neither strain is of the S288C background would likely perform better with tiling array approaches given the additional informative sites. Strain compatibility can be checked computationally for divergence by comparing genome sequences (e.g., (19)) or tiling microarray-based SNP profiles (e.g., (20)) Note that pairs of strains that share ancestry are not good choices unless there is reason to think the trait lies in a divergent segment of the genome. For example, W303,  $\Sigma$ 1278b, and CEN.PK contain large patches of the S288C genome. For mutation mapping, a control cross using the wt strain is first required to determine the background segregation of the trait of interest. For example, if a mutation causes altered growth on a particular substrate, it could not be mapped using a pair of strain backgrounds in which that trait already segregates strongly. In some cases, the phenotype caused by the mutation can be discerned even in the presence of background variation, but in my experience this can lead to inadvertent mapping of modifier loci (see Fig. 2 for an example).
3. The Qiagen Genomic DNA kit performs more consistently on divergent strain backgrounds with a couple of modifications. The enzyme used for digestion may perform differently on strains with varying cell wall content and flocculation features. Experimenting with lyticase vs. zymolyase may help, as does gentle agitation during the incubation. Also, we double the incubation time for all enzymatic digestion steps.
4. DNase I digestion is a critical step and tends to be a main source of variations between people and labs. Handle a small number of samples to allow for precise timing. Some researchers may find it convenient to digest several samples for 5–15 min and choose the best sample.

5. One caution about array layout: special considerations may be required for platforms in which features are spatially placed on the array in a nonrandom way with respect to chromosome location. For example, the Affymetrix yeast expression array is largely in genome order, as are early layouts of the Agilent yeast expression array. If a spatial smoothing normalization is used, this may incorrectly dilute linkage signals, and, conversely, hybridization artifacts may create false signals. If designing a custom array for bulk segregant analysis, it is highly recommended to place the features randomly on the surface.
6. I have published one example in which linkage fell towards the shoulder of a peak as opposed to the center (17). Additional unpublished work from my lab contains at least two other examples of this phenomenon. With the increased resolution of recombination maps, and better appreciation for related strain background dependencies, perhaps some reason will eventually present itself. Also, in at least one published case (12), the causative variant was carried in a large DNA insertion not present in the other strain, leading to a somewhat confusing pattern.
7. Mismatch repair enzyme concentration may benefit from further optimization as unit activities may be variable. See ref. (10) for optimization options.

---

## Acknowledgments

The author would like to thank Matthew Brauer for joint development of these methods. Appreciation also goes to Mark Rose and Elbert Chiang for collaboration on mapping data shown in Fig. 2. Caroline McCoach, Pat Brown, and David Botstein contributed to development of GMS protocols, and USB provided preproduction MutHLS enzymes. This work was supported in part by P50 GM071508 to the Lewis-Sigler Institute at Princeton University.

## References

1. Michelmore R, Paran I, Kesseli R (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci U S A* 88:9828–9832
2. Borevitz J, Liang D, Plouffe D, Chang H, Zhu T, Weigel D, Berry C, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res* 13:513–523
3. Hazen S, Borevitz J, Harmon F, Pruneda-Paz J, Schultz T, Yanovsky M, Liljegren S, Ecker J, Kay S (2005) Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol* 138:990–997
4. Brauer MJ, Christianson CM, Pai DA, Dunham MJ (2006) Mapping novel traits by array-

- assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* 173:1813–1816
5. Segre A, Murray A, Leu J (2006) High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol* 4:e256
  6. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039–1042
  7. Gresham D, Curry B, Ward A, Gordon DB, Brizuela L, Kruglyak L, Botstein D (2010) Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. *Proc Natl Acad Sci U S A* 107:1482–1487
  8. McAllister L, Penland L, Brown P (1998) Enrichment for loci identical-by-descent between pairs of mouse or human genomes by genomic mismatch scanning. *Genomics* 47:7–11
  9. Nelson S, McCusker J, Sander M, Kee Y, Modrich P, Brown P (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nat Genet* 4:11–18
  10. Smirnov D, Bruzel A, Morley M, Cheung V (2004) Direct IBD mapping: identical-by-descent mapping without genotyping. *Genomics* 83:335–345
  11. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jorgensen J-E, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods* 6:550–551
  12. Wenger JW, Schwartz K, Sherlock G (2010) Bulk segregant analysis by high throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genet* 6:e1000942
  13. Birkeland SR, Jin N, Ozdemir AC, Lyons RH, Weisman LS, Wilson TE (2010) Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole genome sequencing. *Genetics* 186(4):1127–1137
  14. Gresham D, Ruderfer DM, Pratt SC, Schacherer J, Dunham MJ, Botstein D, Kruglyak L (2006) Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* 311:1932–1936
  15. Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20:3246–3248
  16. Boer VM, Amini S, Botstein D (2008) Influence of genotype and nutrition on survival and metabolism of starving yeast. *Proc Natl Acad Sci U S A* 105:6930–6935
  17. Fiumera H, Dunham M, Saracco S, Butler C, Kelly J, Fox T (2009) Translocation and assembly of mitochondrially coded *Saccharomyces cerevisiae* cytochrome c oxidase subunit Cox2 by Oxal1 and Ymel1, in the absence of Cox18. *Genetics* 182:519–528
  18. Demogines A, Smith E, Kruglyak L, Alani E (2008) Identification and dissection of a complex DNA repair sensitivity phenotype in baker's yeast. *PLoS Genet* 4:e1000123
  19. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ (2009) Population genomics of domestic and wild yeasts. *Nature* 458:337–341
  20. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458:342–345
  21. Chiang EC (2005) Transcriptional regulation of *cdc31-30::kanMX* as mechanism for its suppression. Princeton University, Princeton