Supporting information for:

Revealing the genetic structure of a trait by sequencing a population under selection

Leopold Parts[1], Francisco A. Cubillos[2], Jonas Warringer[3,4], Kanika Jain[2], Francisco Salinas[2], Mikael Molin[3], Suzannah J. Bumpstead[1], Amin Zia[5], Jared Simpson[1], Michael A. Quail[1], Alan Moses[5], Edward J. Louis[2], Richard Durbin[1] and Gianni Liti[2].

[1]The Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK. [2]Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham, NG7 2UH, UK. [3]Department of Cell and Molecular Biology, University of Gothenburg, 41390 Gothenburg, Sweden. [4]Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences (UMB), PO Box 5003, 1432 Ås, Norway. [5]Department of Cell & Systems Biology, University of Toronto, Canada, M5S 2J4.

Supporting Methods

Generating intercross lines

Overall strategy. We devised a simple strategy that forces yeast cells through multiple rounds of random mating and sporulation to create advanced intercross lines (AILs) (Fig. S1). This step can improve genetic mapping in two ways: increasing resolution by reducing linkage, and unlinking nearby QTLs. The experimental design has three main features. First, it requires little genetic strain manipulation, thus making the method straightforward to extend to natural populations and not limiting it to the laboratory strains. We have previously created over 50 strains in this way that can be used to create AILs (Cubillos et al. 2009). Second, our approach allows systematically undergo many rounds of sporulation and random mating. Finally, in addition to pools of haploid segregants, diploid heterozygous segregants can be obtained, making the yeast a better model for complex trait genetics in diploid organisms.

Strain selection. We selected a North American oak isolate (YPS128) and a West African palm wine isolate (DBVPG6044) as parent strains for their marked phenotypic growth differences at high temperature (Liti et al. 2009) (Fig. S2a). We started from strains that are stable haploid (*ho::HphMX4*), isogenic (except for opposite mating type) and unable to grow in the absence of uracil (Table S1). We replaced the *LYS2* with *URA3* gene in the DBVPG6044 strain (Cubillos et al. 2009). This gene replacement restores the ability of growing in the absence of uracil and makes it unable to grow without lysine.

F1 cross. The two parental strains YPS128 LYS+ and DBVPG6044 URA+ were crossed in complete media (YPDA) and grown overnight (Fig. S1). Patches were replica plated in synthetic minimal media (MIN) to select for diploid F1 hybrids. F1 hybrids were isolated and stored at -80^{o}C. Two F1 hybrid replicas were grown overnight (spread over a whole Petri dish) and replica plated on KAc at 23^{o}C to be sporulated for 10 days. We monitored sporulation efficiency (% of sporulating cells) until it reached >90%. Cells from the whole plate were carefully collected and resuspended in 0.5 ml of sterile water and treated with an equal amount of ether and vortexed for 10 minutes to selectively kill unsporulated cells (Dawes and Hardie 1974). Cells were washed 4 times in sterile water, resuspended in 900 µl of sterile water and treated with 100 µl of Zymolase (10 mg/ml) to remove the ascus. Cell mixtures were vortexed for 5 minutes to increase spore dispersion and inter-ascus mating.

Further rounds of intercrosses. Cells were plated at high density to begin a second round of mating and meiosis (F2). The presence of two distinct markers (*LYS2* and *URA3*) at the same genomic position prevents them from co-segregating in haploid cells, and thus allows selecting for diploid cells segregating at that locus. In order to confirm this system, we dissected 20 tetrads from the F6 pool and all of them had the correct 2:2 segregation of the LYS/URA markers. The sporulated pool was treated with zymolase and plated at high density to start the

next generation of intercross as described in the paragraph above. The result of this strategy is a large pool of segregants for sensitive and high resolution QTL mapping.

Ploidy. We can generate either haploid or heterozygous diploid pool of segregants. For the haploid pool, we used spores from different generations (F1, F6, F12 and F18) treated with ether and zymolase as described above. These pools consist of cells of both mating types (*Mat* a and *Mat* α), and thus have the potential to mate and form diploids. To check whether mating had taken place during the selection experiment, we isolated 960 segregants at the end of the heat selection (T2.5) and found that all of them were able to mate with the haploid tester strains, thus indicating they remained haploid during the selection. Furthermore, we genotyped 288 of these segregants at 11 loci distributed in 8 different chromosomes and confirmed that the segregants remained haploid. Our results indicate that the haploid segregants did not mate during the heat selection, suggesting that the propagation of the cells under heat stress prevents mating events. The pool of segregants with mixed mating types provides a further advantage compared to the *Mat* a pool previously described (Ehrenreich et al. 2010) which is unable to map QTLs in regions linked to the *MAT*, *LYP1* and *CAN1* loci. For the heterozygous diploid pool, we forced one extra round of mating and selection for *LYS+/URA3+* cells, resulting in a pure diploid pool that is stable during subsequent propagation step, as meiosis can only be induced by exposing the cells to a specific media.

Selection experiment

Pools of population size of 10-100 million cells (estimated by plating serial dilutions) were collected from the sporulation media and treated with ether and zymolase as described above. Spores were plated in complete media (YPDA) and incubated either at permissive temperature (23^{o}C) or at restrictive temperature (40^{o}C) until full growth was obtained. Each plate was incubated for 48 hours, after which, all cells were carefully collected and resuspended in distilled water. Ten percent of the suspended cells were used for the next re-plating in fresh media. The rest of the cells (90%) were used for DNA extraction, followed by high coverage genome sequencing described below. We tested a different dilution factor (1/100) for replating, and measured the changes in allele frequencies at the XIII subtelomeric QTL by Taqman assay. We found that the two dilution factors resulted in similar changes in allele frequencies (results not shown). We used the 1/10 dilution for all other experiments to maintain a larger population size through the selection experiment preventing the risk of clonality. At this dilution, the pool undergoes through at least 3.25 generations between re-platings in addition to an estimated 10-20 generations during the first two re-platings for a total of 25-35 generations by T3.

Genotyping

To calculate the percentage of recombinant haplotypes, we obtained genotypes at 24 loci by real time PCR coupled to high resolution melting (HRM) using the Corbet Rotorgene and Quantace PCR HRM mix. We used the data for 96 F1 segregants previously described (Cubillos et al. 2011), and genotyped 96 additional segregants from each of the F6, F12, and F18 at three regions (Dataset 1). We used the same genotyping approach to genotype 19

strongest QTLs in 96 segregants isolated from F12 replica 1 pool after 336 hours of heat selection (T3.5, Dataset 2).

Furthermore, we genotyped 960 F12 segregants isolated from the F12 replica 2 pool after 240 hours of heat selection (T2.5) using the Sequenom iPLEX™ Gold Assay (Sequenom® Inc.). Assays for Sequenom genotyping were designed using the eXTEND suite and MassARRAY Assay Design software version 3.1 (Sequenom® Inc.). Amplification was performed in a total volume of 5µL containing ~0.06-0.4ng genomic DNA, 100nM of each PCR primer, 500µM of each dNTP, 1.25 x PCR buffer (Qiagen), 1.625mM $MgCl_2$ and 1U HotStar Taq® (Qiagen). Reactions were heated to 94 °C for 15 min followed by 45 cycles at 94°C for 20 s, 56°C for 30 s and 72°C for 1 min, then a final extension at 72°C for 3 min. Unincorporated dNTPs were SAP digested prior to iPLEX™ Gold allele specific extension with mass-modified ddNTPs using an iPLEX Gold reagent kit (Sequenom® Inc.). SAP digestion and extension were performed according to the manufacturer's instructions with reaction extension primer concentrations adjusted to between 0.7-1.8µM, dependent upon primer mass. Extension products were desalted and dispensed onto a SpectroCHIP using a MassARRAY Nanodispenser prior to MALDI-TOF analysis with a MassARRAY Analyzer Compact mass spectrometer. Genotypes were automatically assigned and manually confirmed using MassARRAY TyperAnalyzer software version 4.0 (Sequenom® Inc.). Genotype data are presented in Datasets 1-2, and primer sequences are available upon request.

DNA isolation, library preparation, and sequencing

DNA was extracted using the phenol chloroform protocol as previously described (Borts et al. 1986). Multiplexed PCR-free Illumina sequencing libraries were prepared starting from 2 µg of genomic DNA from each strain or pool. Essentially this followed the protocol described in (Kozarewa et al. 2009), except that modified bottom adapter strands were used that introduced a unique tag sequence between SP2 primer site and P7 sequence (Table S10), allowing libraries prepared from 12 samples to be pooled and each samples' unique sequence subsequently deconvoluted according to the specific tag sequence present. The PCR free adapters were made by annealing T_no_PCR oligo AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T (* indicates phosphorothioate) with a specific bottom oligo that contains the indexing tag sequence given in Table S10. Fragments with 200-300bp inserts were gel purified and sequenced using standard Illumina SBS v4 chemistry for 2 x 76 cycles plus an extra 7 cycles of sequencing using the reverse complement of the SP2 primer to determine the tag sequence of each cluster. Parental strains were sequenced using 2 x 108 cycles.

Sequencing data handling

Sequencing reads were mapped to the S288c reference genome obtained from the SGRP project website (Carter 2006) using BWA (Li and Durbin 2009), with option "-n 8" to allow mapping of divergent reads from the founder strains. Pileup files comprising the genotypes of mapped reads were created for segregating sites inferred from both low-coverage capillary sequencing (Liti et al. 2009) and the parental strain shotgun sequence mapping to the S288c assembly using samtools (Li et al. 2009). For allele frequency inference, sites that were not

segregating in the initial population, corresponding to likely false positive variant calls, were filtered out, as well as sites that were noted as heterozygous in either parental strain, indicative of copy number variation or non-unique sequence in general. Furthermore, for allele frequency inference, we filtered the variants to have minimum distance of at least 200 bases to ensure that any single read does not contribute disproportionately due to spanning many variants. The mapping pipeline is available upon request.

Parental strain analysis

The parental strain sequence was mapped similarly to the selection experiment. We used the samtools variant caller with default settings to call differences from the reference sequence, and used these data to update the list of segregating sites used in the allele frequency analysis. The variants not observed were added to the list, and variants observed from capillary data, but not supported by the short reads were discarded.

We performed *de novo* assembly of the parental strains to assess the amount of sequence present in these strains that is not found in the S288c reference genome. We used the overlap-based assembler SGA (String Graph Assembler) to construct contigs from the sequence data. This assembler is based on the FM-index and allows efficient detection of overlaps between reads (Simpson and Durbin 2010). First, we filtered the sequence data to remove low quality reads by discarding any read that had more than 4 bases with a Phred-scaled quality value less than 4. We then corrected base calling errors in the reads by finding inexact overlaps between the reads. We required an overlap of at least 50bp with no more than 5% mismatches between the sequences to consider them overlapped. The consensus sequence for each read was called from the set of overlapping reads. After correction, the reads were assembled using the string graph assembly algorithm described in (Simpson and Durbin 2010) by finding the set of exact overlaps of length at least 50bp between the corrected reads.

Segregant analysis

To analyse the genetic background of two individual F6 segregants, we mapped the sequencing reads to the genome as described earlier, and classified every segregating site to stem from one of the two parental strains, or a ‘no-call’. A site was called to be from one parent, if it was covered by at least 15 sequencing reads with base and mapping qualities at least 30, and 80% of them had the parental allele. We conservatively refrained from making a call at low-coverage variants, subtelomeric regions up to 30kb, and variants with ambiguous mapping data. We called a recombination event if a region of at least 2kb from one parent was followed by a region of at least 2kb from the other, and at least 5 calls were made in both regions. This results in a conservative estimate of recombination events, as it discards non-crossovers shorter than 2kb, and recombination in subtelomeric regions.

We observed less recombination events than expected if at each meiotic generation a new set of crossover occurred independently of the previous ones. It is well known that the distribution of crossovers is not random but instead these events tend to cluster in specific regions (recombination hotspots (Szekvolgyi and Nicolas 2010)), thus many of the crossovers will take place at the same positions. Under the model of uniform distribution of recombination events, and 30 recombinations in F1, we would expect ~105 recombination events by F6. As we

observed 66 on average, we estimate that (66-30)/(105-30) = 48% of the events to reuse the recombination hotspots (perhaps favoured by intra-ascus mating), and we have 106 detectable events on average by F12. Thus, we further estimate that 1 centimorgan of genetic map in the F12 generation corresponds to 1.1 kilobases. It is also possible that there is a recombination preference in the heterozygous diploids within regions inherited from the same strain, and such events cannot be detected. We also expect to observe less recombination events taking place between sequences from different parental backgrounds in regions selected early during the intercross, as these regions will have long tracts of sequence from one parent in the pool. Furthermore, we have filtered out closely spaced events (within 2 kb), which could be originated by short non-crossovers or by re-usage of the same crossover. Some of these issues could be addressed by using a different cross with higher recombination frequency, or mutant strains that exhibit different recombination patterns (Szekvolgyi and Nicolas 2010).

Allele frequency inference

Under a simple model, there is an unobserved WA allele frequency f_l at each locus l; we want to infer the posterior distribution of f_l after observing the sequence data. We assume all reads to come from different segregants after filtering segregating sites to be distant, thus every segregant i has one allele a_i observed at some locus l' distance d_i away from l. We take d to be infinity if the loci are on different chromosomes. For that segregant, there is an unobserved allele b_l at locus l, and the probability that these loci are linked, with no recombination event occurring during the intercross between them, is $q_i = \exp(-d_i r)$, where r is the recombination rate. We took $r = 30\ (1 + (g-1)/2)$, where g is the number of intercross rounds, as there is on average 30 crossovers per tetrad, and every intercross after the first one has a 50% chance of introducing a switch between parental haplotypes. The likelihood of the allele frequency at locus l is thus

$P(D \mid f_l) = \prod_i P(a_i \mid f_l)$, where

$$P(a_i \mid f_l) = P(a_i, b_i = \text{'WA'} \mid f_l) + P(a_i, b_i = \text{'NA'} \mid f_l) =$$

$$= P(a_i \mid b_i = \text{'WA'})P(b_i = \text{'WA'} \mid f_l) + P(a_i \mid b_i = \text{'NA'})P(b_i = \text{'NA'} \mid f_l) =$$

$$= q_i^{a_i = \text{'WA'}} (1-q_i)^{a_i = \text{'NA'}} f_l + q_i^{a_i = \text{'NA'}} (1-q_i)^{a_i = \text{'WA'}} (1-f_l) \approx$$

$$\approx q_i f_l^{a_i = \text{'WA'}} (1-f_l)^{a_i = \text{'NA'}}$$

Here, we have discarded likelihood terms that require a recombination event, as we will filter q_i to be large. We calculate the posterior (beta) distribution of f_l by applying Bayes rule: $P(f_l \mid A) \propto P(A \mid f_l)\, P(f_l) = \prod_i P(a_i \mid f_l)\, P(f_l)$ where the beta prior $P(f_l)$ is uninformative, and we filter $q_i > 0.9$ (0.75 for Fig. 3a-b for smoothness). We further filtered out loci for which the difference between inferred posterior WA allele frequency, and frequency of the WA allele mapped to the locus was greater than 0.1 to downweight the effect of outliers, and repeated the smoothing. This inference procedure corresponds to a smoothing approach within a fixed window, discarding outliers, with the width determined by the recombination rate (~6kb for $q_i > 0.9$), and has the effect of discriminating against extreme allele frequencies. The posterior mean and confidence intervals were obtained from the approximated Beta distribution.

QTL mapping

Allele frequency change. We called a QTL if the inferred allele frequency changed in the same direction by at least 10% in both biological replicas, and the change was larger than four times the average standard deviation of the inferred allele frequencies. One QTL was called in any

50kb window, corresponding to the variant with largest combined allele frequency change over two replicas. We assessed the significance of the calls using the null distribution of allele frequency changes from the control experiment, where the initial pool was propagated in permissive temperature alongside the selected pool for 144h (T2). Due to repeptitive nature of the subtelomeric regions, and resulting low sequencing coverage, we did not consider loci within 30kb of the end of chromosomes. We fit a normal distribution to the allele frequency changes at the 26,871 loci assessed (Fig. S9), and calculated the probability of observing a change of at least 10% in either direction to be less than 10^{-7}. After Bonferroni-correcting for the 26,871 tests, the p-value remained less than 10^{-3}. We further required the allele frequency change at the locus to be at least four times the average standard deviation of the allele frequency posteriors assayed in the two compared experiments. This filter for large changes relative to the uncertainty in the inferred allele frequencies had the effect of discarding regions with low sequence coverage, usually around subtelomeres and centromeres.

Copy number and missing sequence. We mapped all reads to artificial chromosomes, each containing exactly one gene with 100 flanking bases, and recorded their average sequencing coverage every 100 bases. We used that to infer a copy number for each gene as the average gene coverage normalised by the average sequencing coverage (Table S5). We also mapped the reads to the assembled contigs from parental sequence data that did not map to the S288c reference; no large allele frequency changes were observed (data not shown).

Individual allele analysis

Tolerability of mutations. We ran SIFT (Ng and Henikoff 2003; Ng and Henikoff 2006) to predict the effect of all SNPs in coding regions of 7 genes associated with strong heat resistance QTLs along the two WA and NA lineages (Table S7). The main idea of SIFT is that a SNP at an evolutionary conserved locus can affect the function of corresponding protein and therefore is predicted to be intolerable. The conservation at any locus is measured by the amount of diversity of amino-acid (AA) bases observed at that locus. The diversity of AA bases at each locus is represented by a position weight matrix (PWM) that is constructed using an alignment of protein sequences homologues with the protein of interest. The alignments are built recursively by searching protein databases (e.g. Uni-Prot Tr-EMBL 39.6 in this case). For detailed methods of constructing PWMs as well as a number of heuristics used to perform the predictions by SIFT refer to original paper (Ng and Henikoff 2003).

Signatures of selection. We constructed an alignment of the *IRA1* gene sequences for the WA, NA, S288c and *S. paradoxus* strains and inferred the evolutionary tree (Fig. S4b) for *IRA1* using phyML (Guindon and Gascuel 2003). As can be seen from Table S11 no significant divergence between the WA and the NA lineages indicating significant differences in selection pressure along these lineages is observed (comparing to what is expected genome-wide, Fig. 1c of reference (Liti et al. 2009)). We further compared the ratio of amino acid replacements to synonymous polymorphisms in WA lineage to NA lineage, and found no significant difference (lineage WA=0.48 vs. lineage NA=0.75, P=0.4, Fisher's exact test, see also Table S11).

QTL validation

Reciprocal hemizygosity. We validated that the WA and NA versions of *IRA1* and *IRA2* affect growth in heat by reciprocal hemizygosity (Steinmetz et al. 2002). *IRA1* and *IRA2* were deleted individually in the hybrid strain YCC22F using the *URA3* gene as a marker (primers listed in Table S12). Genes were deleted using the standard single-step PCR gene deletion method (Wach et al. 1994). We genotyped independent transformants using HRM-PCR to determine which allele was deleted and measured the effect of the hemizygous deletion with different assays. We performed a temperature growth assay by plating serial dilutions of cells in YPDA and incubated the plates at 30^{o}C and 40^{o}C for 48 hours (Fig. 4a).

Growth curves. We also measured growth curves (Warringer and Blomberg 2003; Liti et al. 2009) in defined media with amino acids for the parental, hybrid and reciprocal hemizygous strains at a range of high temperatures (40^{o}C, 40.5^{o}C, 41^{o}C, 41.5^{o}C, 42^{o}C). Growth curves for each strain were measured in at least 8 replicas, and growth lag, growth rate and growth efficiency variables were extracted and log(2) transformed as previously described (Warringer and Blomberg 2003; Liti et al. 2009)

To test for an epistatic interaction between the *IRA1* and *IRA2* WA alleles, we evaluated the null hypothesis that $X_{observed} = X_{expected}$ where $X_{observed}$ is the observed generation time difference between hemizygotes with WA alleles for *IRA1* and *IRA2* and NA alleles for *IRA1* and *IRA2* and where $X_{expected}$ is the corresponding expected generation time difference, given the generation times of hemizygotes with *IRA1* and *IRA2* alleles of mixed ancestry and an additive model for allele interactions. Formally:

$X_{observed} = Y_{IRA1=WA,\ IRA2=WA} - Y_{IRA1=NA,\ IRA2=NA}$

$X_{expected} = (Y_{IRA1=WA,\ IRA2=NA} - Y_{IRA1=NA,\ IRA2=NA}) + (Y_{IRA1=NA,\ IRA2=WA} - Y_{IRA1=NA,\ IRA2=NA})$.

Y denotes the average of the log(2) transformed generation times of the indicated hemizygotes at 40^{o}C. $X_{observed}$ was found to be significantly higher than $X_{expected}$ (3.3 fold difference, $p < 10^{-29}$, two-sided t test), rejecting the null hypothesis and demonstrating an epistatic interaction between the WA alleles for *IRA1* and *IRA2* at 40^{o}C.

Competitive growth. Finally, we performed a competition experiment that resembled the selection regime imposed on the pool and competed the reciprocal hemizygous hybrids against each other (e.g. for *IRA1* naΔ/WA *vs.* waΔ/NA). For each competition experiment, equal numbers of cells were mixed, a pre-selection sample taken for allele frequency analysis and a dilution (1/10) plated for selection. Competition was performed at either 30^{o}C or 40^{o}C for 96 hours (with re-plating a 1/10 dilution after 48 hours) and a post-selection sample taken at final time point of 96 hours. Pyrosequencing was used to assess the allele frequency in the pools using specific polymorphisms within the IRA genes as a target.

Pyrosequencing

Pyrosequencing (*PyroMark24*, Qiagen) was used to assess changes in allele frequencies after selection in pool of segregants derived from different generations (F1, F6, F12 and F18) and used to infer the allele frequency of 2 strong QTLs (*IRA1* and *HKR1*) and loss of linkage in nearby regions (Fig. S3). We measured allele frequencies at QTL locus, and 20kb upstream and downstream the QTL. For visualisation, we calculated the ratio of beneficial allele frequency at a locus to its frequency at the QTL. We used a biotinylated universal primer (sequences in

Table S13) (Pacey-Miller and Henry 2003) and PCR conditions previously described in (Colella et al. 2007).

cAMP determination

cAMP determination was performed using TCA to extract metabolites (Nikawa et al. 1987) with slight modifications as outlined below. Briefly, $2x10^8$ cells (20 ml of culture at an OD_{600} 0.5) were pelleted in a cold centrifuge, washed and resuspended in 1 ml cold milliQ-water. Samples for normalization were withdrawn (OD_{600} determination) and subsequently metabolites were extracted (Gustafsson 1979) by adding 1.2 ml TCA (0.5 M) and occasional vigourous vortexing while samples were kept on ice for 15 min. TCA was removed by ether extraction. cAMP in the extracts was determined by the LANCE cAMP 384 kit (Perkin-Elmer, cat# AD0262) according to instructions from the manufacturer in 40 µl total reactions and by comparing to the standards supplied. The values for cAMP obtained were normalized to dry weight as obtained from a dry weight-OD600 standard curve.

Supporting Text

Simulation experiments

We simulated data from a simple generative model to explore the influence of QTL allele frequency and many rounds of crossing on mapping resolution, sensitivity to detect allele frequency changes, potential of adaptive mutations to dominate a haploid pool, as well as effects of more than one allele in haploid and diploid pools. Many aspects of using advanced intercross lines for trait mapping have been explored in earlier work (Darvasi and Soller 1995).

Influence of QTL allele frequency and number of intercross rounds on mapping resolution. The mapping resolution depends only on the physical distance corresponding to the genetic distance, which is determined by local recombination rate. For our purposes, we only consider recombination events that take place between tracts of sequence from different parental backgrounds. We observed 30 recombination events on average in 96 F1 segregants, giving an estimate of 3000 cM = 12 Mb, or 4 Kb per cM. However, as discussed above, the true correspondence varies considerably across the genome. Assuming independent recombinations in each generation, the genetic map expands with each intercross round. An x cM locus in the F1 generation corresponding to K base pairs will correspond to $2K/(n+1)$ base pairs in the Fn generation. The additional factor of 2 is due to each recombination in further ($n>1$) intercross rounds having a 50% chance of crossing between tracts of material from different parental background. Thus, by F6, we expect to increase the mapping resolution by a factor of 3.5 compared to F1, and a factor of 6.5 by F12. However, as we estimate 50% of

recombination events taking place in a small portion of the genome, we estimate the increase in resolution from F1 to F12 to be closer to a factor of 3.

We simulated data to visualise and quantify this effect for our approach. We considered a segregant pool selected long enough to reach equilibrium, with allele frequencies no longer changing. This means the QTL effect size per generation is inconsequential, and we thus operated with its final allele frequency alone. For simplicity, we assumed the pool before selection to have 50% NA allele frequency at all loci, and the QTL locus to have NA allele frequency $fq > 0.5$.

The NA allele frequency of locus l that is x cM away from the QTL is $fl = fq$ P(even # of recombinations between l and q) + (1 - fq) P(odd # of recombinations between l and q). Assuming a standard Poisson model for number of recombinations, and uniform distribution of the events, we can calculate fl exactly. We considered 500 loci spaced at 0.2cM intervals on both sides of the QTL (corresponding roughly to a locus every 200 bp in F12, see above), and calculated fl for each of them. Assuming 80x sequencing coverage (lower of the T2 samples we use most in the analyses), we then calculated the distribution of the sampled allele frequency at the locus as f'l ~ Beta(75 fl, 75(1 - fl)). Finally, we used the allele frequency estimation approach outlined in main text, and calculated the posterior allele frequency in the average case by combining the sampled allele frequency distributions, which corresponds to calculating a weighted sum of the parameters of the beta distributions within a 10cM window from each locus.

We considered two aspects of mapping resolution. First, we looked at the total size wi of the interval that we could detect to be changing more than 10% in allele frequency. Second, we calculated the length of the peak interval wp we are confident the causative locus lies in. For our purposes, we defined this to be the length of the region for which the QTL peak allele frequency change is within one standard deviation of the posterior mean allele frequency of sites in the region. We calculated wi and wp for QTLs of final allele frequencies of 0.99, 0.9, 0.8 and 0.65 (Fig. S10, Table S14). QTLs with higher fq had larger mapped intervals, but narrower peaks. In the case of strong QTLs that nearly fix (fq=0.9), the inferred peak interval was 6.6 cM, corresponding to about 7kb in the F12 cross, or about 35 segregating sites. This is in concordance with the actual observed mapped interval sizes (median 6.4 kb, Table S4). Higher sequencing coverage, better inference techniques, and more intercross rounds can narrow the interval even more.

Sensitivity to detect allele frequency changes. We simulated allele frequency changes of 1-10% for one locus, and calculated allele frequencies for 500 loci spaced at 0.2 cM intervals surrounding it as for QTLs. We then sampled 80 alleles at each locus from the calculated fl, followed by allele frequency inference from the sampled data. We repeated this procedure 1000 times, and estimated the normal distribution of inferred allele frequency changes. For a simulated change of 5%, we observed 2/1000 estimated changes smaller than 3%. Thus, extrapolating, we expect to see estimated changes of at least 3% at the 21 QTL loci if the true allele frequency is changing at least 5% at least 95% of the time. Similarly, we expect to see estimated changes of at least 5% if the true change is at least 7% at least 95% of the time. We observed 19/21 QTL loci changing less than 3% in allele frequency between T2 and T3, and two chancing up to 4%. We thus estimate that for 19 QTLs, the allele frequency changes are no

larger than 5%, and for the remaining two, no larger than 7%, while the alleles are still at least 15% from fixation.

Adaptive mutations. We provide three lines of computational evidence for lack of new adaptive mutations with large effect on intercross pool allele frequency during selection.

Firstly, the fitness requirement of adaptive mutations to dominate the pool is too high. A single adaptive mutation begins at very low initial frequency, $f_1 = 1/N$, where we take N, the total number of segregants in the pool to be 10^7. The doubling times for the segregants range from 1.5 hours in permissive condition (or for fit segregants in restrictive condition) to 2 hours for unfit segregants in restrictive condition. Let us assume the adaptive mutation rises to same frequency as the total frequency of haplotypes with beneficial alleles at the two loci that reach fixation - the *IRA1* and chrXIII subtelomeric loci (initial frequency $f_0 = 1/4$) all of which have doubling times $t_0 \sim 1.5$ hours. Over $T = 288$ hours of selection, the following identity must then hold for the doubling time t_1 of the adaptive mutation: $f_1 2^{T/t1} >= f_0 2^{T/t0}$, or $t_1 <= T/(\log_2 f_0 - \log_2 f_1 + T/t_0)$. Plugging in numbers for f_1, f_0, T, this gives $t_1 <= 1.34 = 0.9\, t_0$. Thus, in order to rise to appreciable frequencies in the very large pool, the haplotype with the adaptive mutation must grow 10% faster in restrictive condition than the segregants do in the permissive condition. If such mutations were possible, they would be more likely to rise during the many months of intercross rounds, not during the span of four days. However, in this case, the allele, not the haplotype, will be selected for, as further intercross rounds separate the adaptive mutation from the haplotype on which it arose.

Dominating adaptive mutations would drive the pool allele frequencies to extremes. In the very long run, the haplotype with the adaptive mutation will be the only one left in the pool, as no recombination happens during selection. As the frequency of the adaptive mutation rises in the pool, the pool loses heterozygosity and genetic complexity, and the frequency of the NA allele at all segregating loci will be driven to 0 or 1. If a haplotype with an adaptive mutation is present at high frequency in the pool, we would expect to see an allele frequency change from the initial pool at all loci towards the genotype of that haplotype, which we do not observe. Adaptive mutations would continue to rise in frequency after 192 hours. We do not observe global allele frequency changes after 192 hours. However, as outlined above, haplotypes with adaptive mutations should continue to rise in frequency in the pool. These three lines of evidence point to little contribution from adaptive mutations to the final segregant pool allele frequency makeup. Adaptive mutations for sporulation, mating, or growth can arise during intercross, and could be traced. However, for QTL mapping, we are conditioning our analysis on all the segregating sites present in the pool at the beginning of selection, regardless of whether they were present in the parental strains.

Finally, genotyping 960 segregants at the 19 QTL loci from the selected pool as described in main text yielded 787 unique haplotypes, with most abundant one represented six times. To assess the unexpectedness of this result, we created 1000 simulated datasets of the same size, sampling genotypes at each QTL locus according to its allele frequency. The average number of unique haplotypes per simulated dataset was 786, and 631 of 1000 (63%) of the datasets had at least one haplotype represented at least six times. Thus, our observations are consistent with the null model of no loss of complexity or overrepresentation of a single haplotype.

Effects of selection on allele frequency. We simulated allele frequency changes under simple assumptions for various scenarios. While standard (e.g. (Hartl and G. 2007)), the results give intuition for allele frequency changes observed.

Haploid individuals. We fixed the initial allele frequency of any locus to be 0.5 for simplicity, and calculated its change over generations in a deterministic way. For a one locus trait, the individuals with genotype '1' were assumed to have a fitness advantage *s*, which changed the rate at which they survived to the next generation, with the frequency $f_{l,t}$ of locus *l* at generation *t* was taken to be
$(1 + s)f_{l,t-1}/((1 + s)f_{l,t-1} + (1- f_{l,t-1})) = (1+s)/(1+sf_{l,t-1})f_{l,t-1}$. If $s > 0$, f_l increases, and if $s < 0$, it decreases in a near-geometric manner. For these one locus haploid pools, the beneficial allele asymptotically approaches fixation, with the speed depending on the magnitude of the selection coefficient (Fig. S11).

In case two loci are contributing, the calculation remains almost unchanged, but now the effect of selection is assumed to act only on the '11' genotype. In this case, if s > 0, the haplotypes with '11' genotype are fitter than the others, and again are driven to fixation. However, if s < 0, the '11' genotype is selected against, and will be purged from the pool in the long run. Both alleles will still be present at each locus (Fig. S11a). We hypothesize such interactions within a chromosome to be responsible for the lack of fixation upon ~40 generations of selection. The usual intuition behind this is that fitness depends on functioning of a specific pathway. While any single mutation does not alter the functionality of the pathway, there are many possible combinations of genotypes that render it defective. These combinations are selected against, producing a change in allele frequency, but not fixation of any allele.

Diploid individuals. As the diploid individuals propagate clonally just like haploids, we have to trace the frequency of the genotypes, not alleles, since there is no further mixing of the haplotypes between individuals. We can treat a one locus trait in diploids, identically to a two-locus trait in haploids, and find that for traits where the beneficial allele behaves in an additive or recessive way, selection drives the frequency of beneficial allele to fixation, and for dominant beneficial alleles, the homozygous non-beneficial allele combination is selected against (Fig. S11b). We observed QTLs with final allele frequencies as well as their speed of change consistent with both recessive (*IRA1*) and dominant (chrXIII subtelomere) beneficial alleles (main text). However, when the QTL acts in an additive manner, the allele frequency change is identical to that of the haploid pool.

If interaction effects are responsible for the allele frequency change, the effect can again be dominant, additive, or recessive. The differences to a one-locus model are slower effect of selection, as the fittest haplotype has lower initial frequency, and less extreme final allele frequency in case the interaction effect is dominant, as there are less genotype combinations selected against (Fig. S11b).

The RAS signalling pathway regulates quantitative growth at high temperature

Of the 19 QTLs we mapped here, we have validated three with large effect size. We previously mapped and validated the QTL present in the subtelomeric region of chromosome XIII (Cubillos et al. 2011). Due to the difficulty of assembling the subtelomeres, we are unable to characterise this high temperature growth QTL. The only other QTL that reached fixation was the GTPase activating protein *IRA1,* a negative regulator of the RAS signalling pathway, containing some of the most prominent human oncogenes (Tanaka et al. 1989). Interestingly, additional genes of the Ras/cAMP/PKA pathway pathway were contained in intervals with sharp increases in NA allele frequencies (Fig. 4C). *IRA1* and *IRA2* are of specific interest as they are highly conserved orthologs of the human disease gene NF1 (Ballester et al. 1990), which causes neurofibromatosis type 1, and as mutations in patients with neurofibromatosis also have similar detrimental effects on the yeast Ira1p activity (Gil and Seeling 1999). Thus, we validated by reciprocal hemizygosity (Steinmetz et al. 2002) that the naturally evolved *IRA1* and *IRA2* alleles indeed affect high temperature mitotic growth (Fig. 4 and Fig. S5a-b). These genes affect both growth rate (population doubling time) and efficiency (change in population density) with *IRA1* having a stronger effect compared to *IRA2*, consistent with their strength as inferred by the final allele frequency (Fig. S5a-b). The effect was restricted to a surprisingly narrow temperature range, peaking at 41°C, but with no effect at 42°C, reflecting an extremely fine-tuned gene-by-environment interaction and emphasizing the need for high resolution phenotyping in QTL analysis.

We tested the idea that the IRA polymorphisms in the WA lineage are pleiotropic at multiple environmental conditions. We measured growth curves of four replicas in five conditions (caffeine 2.25 mg/ml, paraquat 400µg/ml, ethanol 7%, DTT 1.8 mM, 1.5 µg/ml). The conditions were selected from previous high throughput screening in the *S. cerevisiae* deletion collection mutants showing an effect of the IRA deletions compared to the wild type. Interestingly, the West African *IRA* alleles do not have pleiotropic consequences, even at environmental conditions where a compromised RAS activity has been reported to have strong phenotypic consequences (Park et al. 2005) (Fig. S5c). Our results indicate that the growth defect due to *IRA1* and *IRA2* WA alleles is specific to high temperature.

To experimentally investigate the *IRA1* and *IRA2* interactions, we constructed the four possible reciprocal hemizygous combinations of their alleles (Table S1). Growth curve analyses indicated a strong negative epistatic interaction between the WA alleles (Fig S6) consistent with a partially redundant role of the Ira gene products. We then measured the internal level of cAMP in the parental strains and the double reciprocal hemizygous deletion strains, and detected a 1.6-fold higher level in the WA compared to the NA strain consistent with its RAS hyperactivity. This difference is similar to the difference in basal cAMP levels between the *ras2Δ* and the WT in mid-exponential phase and in the laboratory strain BY (Caballero et al, manuscript submitted). Importantly, we also detected a 9-fold difference in cAMP level between the hybrid carrying the WA alleles of the Ira genes compared to hybrid carrying the NA alleles when cells were grown at high temperature (Fig. 4C) but not at a permissive temperature. These data are consistent with the WA versions of Ira1 and Ira2 becoming defective at high temperature and leading to high cAMP levels and PKA activity and explain the lack of pleiotropy at other

conditions at permissive temperature. The clear identification of the *IRA1* and *IRA2* alleles as a cause of low performance at high temperatures show that our method can directly map causative genes without any *a priori* information and without requiring further fine-mapping.

Computational analyses of *IRA1* and *IRA2* alleles

The strongest predictions for genes identified in the large effect size heat resistance QTL regions are presented in Table S7. For the QTL of interest, *IRA1*, the predictions by SIFT sorted by scores of intolerance are presented in Table S8. It can be seen from this table and Fig. S4A that the strongest prediction corresponding to the SNP in the WA lineage at *IRA1*-ChrII:522887 (G to A) is located in a well-conserved locus IRA1p-1246. By aligning *IRA1* to *IRA2* and NF1 in human and *Drosophila* we found that our mutation of interest (IRA1p:1246) mapped in a highly conserved region of the alignment in all homologues (not shown).

To investigate the evolutionary history of the QTL *IRA1* we constructed an evolutionary tree of the gene consisting of WA, NA, S288c, and *S. paradoxus* as the outgroup (Fig. S4B). We did not observe any significant evidence of change in selection between WA and NA lineages compared to what is expected genome-wide (*11*). Furthermore, we compared the ratio of amino acid replacement to synonymous polymorphism in WA lineage to NA lineage, and found no significant difference (lineage WA=0.48 vs. lineage NA=0.75, P=0.4, Fisher's exact test), which supported this observation (Fisher's exact test with p-value=0.446, Table S11).

Selection during intercross rounds

We further investigated the regions in chromosome I (30-50 kb) and chromosome V (110-180kb)) selected during the F1 to F6 intercross rounds and tested if they harbour sporulation efficiency QTLs. We used 40 F1 segregants from the WA x NA cross described in (Cubillos et al. 2011) and backcrossed them to the WA parental strain. We sporulated the cells in KAc for five days and calculated the percentage of sporulating cells by counting at least 200 cells for each sample in duplicate. We found a strong effect on sporulation efficiency for the chromosome V maker YND1 (position 160kb). The sporulation average for segregants carrying the YND1 heterozygous marker (A/W) was 43% compared to the 21% detected in the homozygous (W/W). This difference is highly significant ($p < 0.005$, t test for unpaired data with unequal variance) and is consistent with the presence of a previously uncharacterised slow sporulation QTL present in the WA lineage that contributes to the low sporulation efficiency described in the WA strain (Cubillos et al. 2009) and results in strong selection for the NA version of this chromosome V locus during the intercross.

Supplementary Figures

Fig. S1. Intercross strategy.

Strategy used to generate advanced intercross lines. Description of the product and the action for each step are indicated on the left and the right side respectively. Each round of intercross requires approximately two weeks.

Fig. S2. Allele frequency after paraquat selection

WA allele frequencies of the whole genome for one pool subjected to heat and paraquat stress for 192 hours (T2) compared to the control experiment.

Fig. S3. Loss of linkage upon several rounds of intercross assayed by pyrosequencing.

Increased recombination frequency results in loss of linkage between QTLs and nearby loci, and resulted in narrower mapped QTL peaks. Allele frequency was measured at three positions (-20, 0 and 20Kb) for two strong QTLs. The relative abundance of beneficial allele is the ratio of beneficial allele frequency at genotyped locus (e.g. +20kb) and the beneficial allele frequency at QTL site.

Fig. S4. Bioinformatics analysis of IRA1

a. A logo representation of position weight matrix (PWM) constructed based on an alignment of proteins homologous to Ira1 used by SIFT to perform the prediction. The alanine 1246 in Ira1p is well-conserved and a SNP in the WA lineage (A to S) is predicted to be intolerable.
b. Evolutionary tree constructed for IRA1 gene in the WA, NA, the reference and *S. paradoxus* as outgroup. No significant divergence between WA and NA lineages is observed compared to what is expected genome-wide (see(Liti et al. 2009), Fig. 1c).

Fig. S5. Growth curves for *IRA1* and *IRA2* reciprocal hemizygotes.

a-b. Growth curves in **a** show a difference in growth between WA and NA strains (left plot) and a temperature dependent effect of hemizygous knockouts of *IRA1* (center plot) and *IRA2* (right plot) from the WA/NA hybrid on growth with maximum final OD difference at 41-41.5^{o}C. Interestingly, both panel **a** and **b** also indicate haploisufficiency of *IRA2* showing the WA allele still contributing to the heat phenotype in the hybrid. Growth rate and efficiency values show in **b** were extracted from growth curves show in **a**.
c. Growth defect of WA *IRA1* and *IRA2* alleles is specific to high temperature. The graphs show average doubling time (growth rate) of four replicas in five conditions for which deletions of *IRA*

genes had an effect compared to the wild type in screening the *S. cerevisiae* deletion collection mutants.

Fig. S6. Phenotypic analysis of double reciprocal hemizygotes of *IRA1* and *IRA2* alleles.

A. Average doubling time of 20 *IRA1*/2 double hemizygotes grown at 40^{o}C. The double hemizygote *IRA1* WA/*IRA2* WA grows significantly slower ($p < 10^{-21}$, two-sided t-test with unequal variance) than each of *IRA1* WA/*IRA2* NA (23% slower), *IRA1* NA/*IRA2* WA (21% slower) and *IRA1* NA/*IRA2* NA (26% slower).

B. Doubling time defect of *IRA1* WA/*IRA2* WA compared to multiplicative expectation indicate strong negative epistatic interaction between the WA alleles.

Fig. S7. Summaries of allele frequency changes between consecutive timepoints.

Histogram (left column) and scatterplot (middle column) of changes in inferred allele frequency between control experiment and initial pool (first row), control experiment and first timepoint (T1) of heat selection (second row), T2 and T1 of heat selection (third row) and T3 and T2 of heat selection (last row). Right column shows histograms of allele frequency changes relative to the average standard deviation of the compared posteriors between the same set of samples. All data are from replica 2.

Fig. S8. Comparison of alternative assays for allele frequency estimation.

Every marker corresponds to the WA allele frequency of one QTL locus after heat stress measured by genotyping (y-axis) or sequencing (x-axis) in first replica (blue) or second replica (green). Sequencing allele frequencies are taken from timepoint T2 for both replicas. Genotyped allele frequencies for replica 1 are assayed from 96 segregants genotyped at 11 loci, and for replica 2 from 960 segregants genotyped at 24 loci (Dataset 2). Genotypes from sequencing data were inferred using a smoothing approach, which underestimates the extremity of the allele frequencies, and explains the difference between genotyping and sequencing-derived estimates at low WA allele frequencies.

Fig. S9. Control experiment allele frequency change summaries.

QQ plot (left panel) and histogram with a normal distribution fit (matching sample mean and standard deviation, red line) of allele frequency changes between control experiment T2 and initial pool. Sites from subtelomeric regions (up to 30kb from chromosome end) are excluded

due to repetitive nature of the regions resulting in low sequencing coverage and high variance posterior allele frequencies.

Fig. S10. Mapping resolution for simulated QTL peaks.

Mapping resolution, quantified by the width *wp* of the inferred peak in cM (see text above), for QTLs with different final allele frequencies *fq.* The thin line corresponds to the mean inferred allele frequency, dashed line to mean + one standard deviation, thick line to the region with allele frequency change from 50% above 10%, and the thickest line to the mapped QTL region. Dashed black line shows the 10% change cutoff.

Fig. S11. Changes in allele frequency in simulated data.

Haploid (solid lines) and diploid (dashed lines) pool allele frequency changes for 1-locus (a) and 2-locus effects (b). Initial allele frequency of a locus is 0.6. Individual lines correspond to different fitness modifiers (see Supporting Text), from top to bottom: +1, +0.3, +0.1, +0.03, -0.03, -0.3, -1.

Dataset 1

Genotypes of three regions for 96 segregants from each of F6, F12, and F18 generations isolated before selection to assess increase in segregants recombinant in the regions using HRM PCR.

Dataset 2

Genotypes for 960 individual segregants isolated after heat selection (T2.5) genotyped for 19 QTLs using the Sequenom platform and 96 segregants isolated at T3.5 genotyped for 11 QTLs using HRM PCR.

Dataset 3

Allele frequencies for different timepoints, cross generations, and ploidies for the 21 QTLs detected. Each window corresponds to 80kb, centered on the variant with the largest change in frequency. See Fig. 3d for details on legend. Top panel – haploid timepoints; middle panel – F6 and F12 generations; bottom panel – haploid and diploid pools. Dashed lines indicate boundaries for the inferred QTL region.

Supporting Tables

Supplementary Table 1. Strains used.

Name	Derived from	Genotype and notes	Ref.
YPS128	Wild type isolate	Isolated in Pennsylvania (1999) by P. Sniegowski from soil beneath *Quercus alba.*	(Sniegowski et al. 2002)
DBVPG6044	Wild type isolate	Isolated in West Africa (pre-1914) by A. Guilliermond from bili wine from *Osbeckia grandiflora*	(Liti et al. 2005)
NCYC3607	YPS128	*Mat a, ura3::KanMX4, ho::HphMX4*	(Cubillos et al. 2009)
NCYC3625	DBVPG6044	*Mat* α*, ura3::KanMX4, ho::HphMX4*	(Cubillos et al. 2009)
YFCL1	DBVPG6044	*Mat* α*, ura3::KanMX4, ho::HphMX4, lys2::URA3*	(Cubillos et al. 2009)
CC35	F1 Hybrid: NCYC3607 x YFCL1	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, LYS2/lys2::URA3*	This study
YCC22F	F1 Hybrid: NCYC3607 x YFC3625	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4*	(Cubillos et al. 2011)
YCC23F	F1 Hybrid NCYC3607 x YFC3625	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4*	(Cubillos et al. 2011)
YFCRH101	YCC22F	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(na)::URA3/IRA1(WA)*	This study
YFCRH102	YCC22F	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(wa)::URA3/IRA1(NA)*	This study
YFCRH103	YCC22F	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira2(na)::URA3/IRA2(WA)*	This study
YFCRH104	YCC22F	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira2(wa)::URA3/IRA2(NA)*	This study
YFCRH117	YFCRH101	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(na)::URA3/IRA1(WA), ira2(na)::NatMX/IRA2(WA)*	This study
YFCRH118	YFCRH102	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(wa)::URA3/IRA1(NA), ira2(na)::NatMX/IRA2(WA)*	This study
YFCRH119	YFCRH102	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(wa)::URA3/IRA1(NA), ira2(wa)::NatMX/IRA2(NA)*	This study
YFCRH120	YFCRH104	*Mat a/*α*, ura3::KanMX4/ura3::KanMX4, ho::HphMX4/ho::HphMX4, ira1(na)::NatMX/IRA1(WA), ira2(wa)::URA3/IRA2(NA)*	This study

Supplementary Table 2. Regions selected for during rounds 6-12 of the intercross.

Chr, Location	Combined allele frequency change (R1 + R2)	Region start (bp)	Region end (bp)	Region length (bp)	Number of genes in region	Region genes
2, 132336	0.23	81134	187145	106011	68	
2, 475979	0.38	469724	477009	7285	3	*LYS2, YBR116C, TKL2*
4, 408754	0.34	400287	420414	20127	13	
4, 469197	0.39	433335	527824	94489	51	
4, 573669	0.56	570208	579573	9365	5	*MAK21, YDR061W, LCB2, YDR063W, RPS13*
4, 709767	-0.37	703851	817213	113362	52	
4, 892054	0.25	890414	895198	4784	4	*UPC2, AHA1, YDR215C, ADR1*
4, 1452475	-0.27	1449398	1459253	9855	7	*LCD1, RPL37B, PLM2, SAM2, LPP1, SPG3, PSP1*
5, 26206	-0.21	25590	28600	3010	4	*YEL068C, YEL067C, HPA3, SIT1*
7, 934217	0.22	928940	939704	10764	5	*CCH1, CRM1, YGR219W, MRPL9, TOS2*
8, 57903	0.27	55475	62378	6903	3	*SNF6, RIM4, RMD11*
8, 256539	0.32	249529	280866	31337	15	
9, 79354	0.3	76340	90209	13869	10	
10, 451935	0.26	444889	453307	8418	5	*APL1, YJR005C-A, POL31, SUI2, YJR008W*
10, 604363	0.46	600041	618895	18854	11	
12, 391582	-0.4	387520	583104	195584	122	
12, 912538	-0.28	895661	921060	25399	15	
12, 970723	-0.35	959600	971646	12046	4	*YLR419W, URA4, RPN13, YLR422W*
14, 41217	0.28	16112	50093	33981	21	
15, 178731	0.45	174849	183754	8905	4	*IRA2, REX4, YOL079W, AVO1*
15, 234578	-0.27	66457	244782	178325	98	
15, 571166	-0.33	567019	601747	34728	15	

Supplementary Table 3. Average sequencing coverage of analysed samples.

Sample	Generation	Replica	Type	Ploidy	Condition	Timepoint	Average coverage at segregating sites	SRA Accession number
Initial_R1_F6_T0	6	1	Pool	Haploid	Permissive	0	23.8	ERS002654
Initial_R2_F6_T0	6	2	Pool	Haploid	Permissive	0	13.1	ERS002657
Heat_R1_F6_T2	6	1	Pool	Haploid	Heat 40C	2	19.3	ERS002655
Heat_R2_F6_T2	6	2	Pool	Haploid	Heat 40C	2	25.7	ERS002658
Initial_R1_F6_S1	6	1	Segregant	Haploid	Permissive	0	20.3	ERS002656
Initial_R2_F6_S1	6	2	Segregant	Haploid	Permissive	0	27.4	ERS002659
Mock_R1_F12_T2	12	1	Pool	Haploid	Permissive	2	115.4	ERS018247
Heat_R1_F12_T2	12	1	Pool	Haploid	Heat 40C	2	129.3	ERS018248
Mock_R2_F12_T2	12	2	Pool	Haploid	Permissive	2	105.7	ERS018249
Initial_R2_F12_T0	12	2	Pool	Haploid	Permissive	0	107.3	ERS018250
Heat_R2_F12_T1	12	2	Pool	Haploid	Heat 40C	1	54.8	ERS018251
Heat_R2_F12_T2	12	2	Pool	Haploid	Heat 40C	2	83.7	ERS018252
Heat_R2_F12_T3	12	2	Pool	Haploid	Heat 40C	3	65.9	ERS018253
Diploid-heat_R2_F12_T3	12	2	Pool	Diploid	Heat 40C	3	32.6	ERS018255
Diploid-heat_R1_F12_T2	12	1	Pool	Diploid	Heat 40C	2	88.6	ERS018256
Paraquat-R1_F12_T2	12	1	Pool	Haploid	Paraquat	2	150	ERS018258

Supplementary Table 4. Regions selected for upon heat selection for F12 pool after 96h.

Chr, Location	Combined allele frequency change (R1 + R2)	Region start (bp)	Region end (bp)	Region length (bp)	Number of genes in region	Region genes
1, 119382	0.34	118724	125287	6563	5	*YAL018C, PSK1, YAL016C-B, YAL016C-A, TPD3*
2, 519399	-0.7	515930	522327	6397	2	*YBR139W, IRA1*
4, 461085	-0.33	457812	463886	6074	5	*MAF1, SOK1, TRP1, YDR008C, GAL3*
4, 802325	0.22	797088	805655	8567	1	*SEC7*
4,1050116	-0.23	1046548	1051757	5209	2	*SSD1, DPL1*
4, 1307579	0.44	1305616	1314837	9221	2	*HKR1, ARO80*
5, 526677	-0.27	522500	536992	14492	7	*CCA1, RPH1, ADK2, RAD3, BRR2, YER172C-A, RAD24*
6, 150543	-0.27	148802	154639	5837	5	*LOC1, NIC96, YPI1, RPN11, SAD1*
7, 127600	0.32	125071	135730	10659	3	*MDS3, DSD1, GCN1*
7, 858516	0.88	856677	859960	3283	4	*TIM13, YGR182C, QCR9, UBR1*
9, 292345	-0.35	287155	295587	8432	7	*CST6, CKA1, CAP2, BCY1, YIL032C, ULP2, YIL030W-A*
10, 238929	-0.45	235471	240115	4644	3	*GSH1, LSB6, CHS6*
10, 423960	-0.48	420008	425480	5472	5	*YJL009W, CCT8, YJL007C, CTK2, CYR1*
12, 137312	0.41	134931	140165	5234	4	*YLL007C, YLL006W-A, MMM1, SPO75*
12, 735715	-0.34	733797	737572	3775	2	*MET17, ACO1*
13, 749977	-0.32	743628	753953	10325	6	*BCH1, DFG5, RNT1, CUS1, YHM2, RPL20A*
13, 895955	-0.59	891318	898276	6958	4	*PSE1, NIP1, YMR310C, GLC8*
14, 480653	0.54	476889	485871	8982	6	*EOS1, TPM1, NIS1, APJ1, MKS1, IMP4*
14, 685970	-0.22	683792	688334	4542	4	*SSK2, PPG1, HUB1, ABZ1*
15, 182170	-1.31	176359	183754	7395	4	*IRA2, REX4, YOL079W, AVO1*
15, 1031113	-0.8	1028004	1032237	4233	4	*RAD17, RPS12, MRS6, GPB1*

Supplementary Table 5. Copy number variable genes upon heat selection.

Gene	F12 T2 copy number	F12 T0 copy number	Change T0-T2
Q0045	0.4	36.5	-36.1
Q0250	2	37	-35
Q0255	1.7	36.1	-34.4
Q0060	0.3	31.9	-31.6
Q0115	1.2	32	-30.8
Q0275	1.8	32.3	-30.5
Q0105	3.3	32.8	-29.5
Q0050	0.2	27.7	-27.5
Q0120	1.6	28.3	-26.7
Q0070	0.2	26	-25.8
Q0085	1.9	26.1	-24.2
Q0065	0.2	22	-21.8
Q0182	0.7	18.3	-17.6
Q0032	0.9	12.3	-11.4
Q0142	0.3	11.3	-11
YLR162W	44.5	55.4	-10.9
Q0140	3.3	13.2	-9.9
Q0130	2.7	11.4	-8.7
Q0144	2.2	10.8	-8.6
Q0143	0.7	7.9	-7.2
Q0080	0.1	6.2	-6.1
YDR366C	11.5	17.6	-6.1
Q0110	0.7	6	-5.3
Q0010	13.7	18	-4.3
Q0092	0	3.5	-3.5
Q0017	0.1	2.5	-2.4
YEL074W	4.1	5.9	-1.8
YIR044C	1.1	2.9	-1.8
YIL174W	0.7	1.9	-1.2
YJL225C	2.1	3.3	-1.2
YNL337W	1.6	2.8	-1.2
YOL166C	1.6	2.8	-1.2
YHR216W	3.4	4.4	-1
YLR465C	2.6	0.9	1.7
YDR340W	8.3	3.9	4.4

Supplementary Table 6. Allele frequencies at QTL loci during selection.

Chr	Location	Initial R2 F12 T0	Mock R2 F12 T2	Heat R2 F12 T1	Heat R2 F12 T2	Heat R2 F12 T3	DipHeat R2 F12 T2	DipHeat R2 F12 T3	Mock R1 F12 T2	Heat R1 F12 T2	Paraquat R1 F12 T2
1	119382	53.7	52.6	63.2	70.6	71.5	57.5	60.8	48	64	42.1
2	519399	36.2	37.1	15.7	0.3	1.2	18.3	2.1	33.7	0.6	39
4	461085	84.1	84	75.9	66.2	64.8	70.8	53.9	71.8	56.3	59.5
4	802325	34.1	32.5	41.1	44	46.1	39.9	54.2	40.9	51.4	51.6
4	1050116	38.4	37.7	31	25.3	25	34.5	27.3	37.5	26.5	33.9
4	1307579	43	42.3	52.5	63.7	65.8	50.7	65.3	40.8	63	33
5	526677	54.9	56.8	54.5	41.3	42.1	47.4	32.2	50	38.3	41.4
6	150543	60.5	64	59.7	49.4	48.5	52.7	40.6	56.8	44.2	56.9
7	127600	39	39.1	45.4	54.3	58.3	38.7	44	37.7	54.7	51.1
7	858516	43.1	43	67.1	86.2	86.3	63.9	86.5	40.6	85	38.5
9	292345	59.8	62.7	53	40.4	40.9	50.5	35.7	51.5	38.9	38.4
10	238929	58.7	60.5	48	35	34.6	43.8	28.2	47.2	27.8	45.5
10	423960	65.8	68.7	58.1	43.9	43.2	57.1	44.8	71.8	48.2	72.3
12	137312	35	35.4	43.6	58.2	56.2	43.3	54.9	37.9	56.4	29.3
12	735715	39.6	42.8	35.5	23.3	25.6	32.6	22.9	38.2	23.6	36.7
13	749977	43.1	45.8	38.3	31.2	33.9	36	25.2	49	32	44.3
13	895955	27.9	27.6	5.5	1.6	2.9	25	10.6	35.1	2.1	23.6
14	480653	46.9	47	60.6	71.5	74.8	61.3	81.7	40.4	69.4	50.8
14	685970	39.2	38.9	37.1	28.6	29.2	40.2	36.8	48	36.8	46.5
15	182170	85	87.8	60.3	17.2	16	58.2	17.7	77.5	16.7	67.1
15	1031113	62.2	61.9	40.3	19.3	21.4	41.3	18.5	58.3	20.4	51.8

Supplementary Table 7. Top SIFT scores for 7 strongest QTLs.

Gene	QTL region	Chr locus	Ref (S288c)	DBVPG6044 (WA)	YPS128 (NA)	Protein locus	Ref (S288c) AA	DBVPG6044 (WA) AA	YPS128 (NA) AA
GPB2	ChrI 36-49 kb	40524	A	T	.	422	K	I	.
IRA1	ChrII 518-523kb	522887	G	T	.	1246	A	S	.
HKR1	ChrIV 1306-1312 kb	No prediction							
GPA2	ChrV 93-207kb	196109	A	G	.	315	T	A	.
GSH1	ChrX 229-237kb	235664	C	T	.	230	S	F	.
GPB1	ChrXV 1032-1037kb	1031596	C	A	.	862	T	N	.
IRA2	ChrXV 168-183kb	172081	G	C	.	338	A	P	.

Strongest predictions by SIFT for 7 well-characterized genes associated with heat resistance QTL regions. All the predictions are associated with SNPs in WA lineage. "." means equal to S288c reference.

Supplementary Table 8. SIFT analysis of *IRA1* non-synonymous variants.

ChrII locus	Ref (S288c)	DBVPG6044 (WA)	YPS128 (NA)	Protein locus	Ref (S288c) AA	DBVPG6044 (WA) AA	YPS128 (NA) AA	SIFT score	Tolerable-Intolerable
522887	G	T	G	1246	A	S	.	0	Intolerable
517820	G	A	G	2935	V	M	.	0.01	Intolerable
518671	T	C	T	2651	V	A	.	0.02	Intolerable
520745	T	A	T	1960	F	I	.	0.18	Tolerable
520381	T	C	T	2081	V	A	.	0.18	Tolerable
523279	T	.	C	1115	L	.	P	0.19	Tolerable
520334	A	G	A	2097	I	V	.	0.19	Tolerable
520084	G	C	G	2180	G	A	.	0.19	Tolerable
520246	G	A	G	2126	S	N	.	0.25	Tolerable
518563	C	T	C	2687	A	V	.	0.27	Tolerable
524465	G	A	G	720	A	T	.	0.31	Tolerable
524102	G	.	A	841	D	.	N	0.35	Tolerable
522736	C	T	.	1296	S	L	.	0.39	Tolerable
523094	G	A	.	1177	A	T	.	0.4	Tolerable
521572	T	C	.	1684	I	T	.	0.49	Tolerable
521279	A	G	.	1782	I	V	.	0.64	Tolerable
517394	G	A	.	3077	E	K	.	0.68	Tolerable
524309	C	T	.	772	P	S	.	0.72	Tolerable
519853	A	G	.	2257	N	S	.	0.8	Tolerable
526588	T	.	C	12	F	.	S	0.84	Tolerable
520667	A	G	.	1986	N	D	.	0.86	Tolerable
521896	G	C	.	1576	S	T	.	0.95	Tolerable
519341	G	A	.	2428	V	I	.	1	Tolerable
526602	A	.	T	7	Q	.	H	?	?
525205	C	T	.	473	T	I	.	?	?
525008	G	A	.	539	A	T	.	?	?

Predictions by SIFT for the QTL IRA1. "." means equal to S288c reference.

Supplementary Table 9. Most significant 2-locus interactions.

nominal p	Bonferroni-corrected p	FDR at cutoff	WA/WA	NA/WA	WA/NA	NA/NA	Locus 1	Locus 2
0.0007	0.116	0.116	98	177	156	484	chr10 235663	chr12 730764
0.0024	0.398	0.199	5	153	83	697	chr15 1032447	chr15 172081
0.0026	0.433	0.144	155	250	250	267	chr7 131690	chr12 140165

Interaction p-value was calculated using two-sided Fisher's exact test in R (function fisher.test), on the two-locus genotype frequencies from 960 segregants (Dataset 2) as described in SI. Bonferroni-corrected p-value (166 tests), and false discovery rate (FDR) using the nominal p-value as a cutoff are given, along with genotype frequencies. No other interactions are significant at 0.3 FDR.

Supplementary Table 10. Sequencing primers used in the multiplexed library creation.

Primer Name	Sequence (5'-3')
inoPCR_B1	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCATCACGTTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B2	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCCGATGTTTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B3	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCTTAGGCATATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B4	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCTGACCACTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B5	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCACAGTGGTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B6	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGCCAATGTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B7	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCCAGATCTGATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B8	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCACTTGATGATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B9	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGATCAGCGATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B10	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCTAGCTTGTATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B11	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGGCTACAGATCTCGTATGCCGTCTTCTGCTT*G
inoPCR_B12	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCCTTGTACTATCTCGTATGCCGTCTTCTGCTT*G

Supplementary Table 11. Synonymous and non-synonymous polymorphisms in *IRA1*.

Strain	Ka	Ks	Ka/Ks
S. paradoxus	152	707	0.21
S288c	6	4	1.5
DBVPG6044 (WA)	14	29	0.48
YPS128 (NA)	3	4	0.75

Synonymous and non-synonymous polymorphism data between S288c, DBVPG6044 (WA) and YPS128 (NA) lineage and their outgroup *S. paradoxus*. No significant divergence between WA and NA lineages is observed from the Ka/Ks ratios.

Supplementary Table 12. Primers used for *IRA1* and *IRA2* deletion.

Primer name	Target coordinates	Sequence (TARGET, BARCODE, *ura3*)
IRA1-FW-YPS128	517266-517344	CAACAAATATAAAACAAAATATAATTATAAGGAAAAACGTATATAATCACTGCA ATACTCTAATTTAAAATGGACTCagcttttcaattcaattcatcat
IRA1 FW1-DBVPG6044	517266-517344	CAACAAATATAAAACAAAATATAATTATAAGGAAAAACGTATATAATCACTGCA ATACTCTAATTTAAAATCACTCGagcttttcaattcaattcatcat
IRA1-RV	526622-526702	TTTTGCCCTGCAAATAGAGCTTCAAACTTAACATTCTTCTTCAGCATATAACAT ACAACAAGATTAAGGCTCTTTCTAAAAagcttttttctttccaatt
IRA2-FW-YPS128	171000-171070	TTTCCCCCAACGTTACACCATTTTTTGATATCAACTAAACTGTATACATTATCT TTCTTCAGGGAGAAGCAGGACTCagcttttcaattcaattcatcat
IRA2-FW-DBVPG6044	171000-171070	TTTCCCCCAACGTTACACCATTTTTTGATATCAACTAAACTGTATACATTATCT TTCTTCAGGGAGAAGCACACTCGagcttttcaattcaattcatcat
IRA2-RV1	180309-180389	AGAAAACCCTAACATGAGATATGTACATTCATGCTTACAGATAGATATTGATA TTTCTTTCATTAGTTTATGTAACACCTCagcttttttctttccaatt

Sequence coordinates for chromosome II (*IRA1*) and XV (*IRA2*) are based on the S288c genome from the SGD database. The target sequence in the primers is shown in upper case, the unique sequence tag used for the real time PCR is underlined and the sequence that amplifies the selectable marker (*URA3*) is in lower case.

Supplementary Table 13. Primers used for pyrosequencing.

Chr	Kb	Forward primer	Reverse Primer	Sequencing primer
II	497	CTTTTCTGTTTCCAGAGATTTCAA	GCCCCCGCCCGTTAAATGAAAAAATCGGTTTGTTG	TCTATATCTGAATTTGAACTAGTA
II	517	GCCCCCGCCCGCTGAGTCTGAACTTAGCCAGTTGA	TTGTTACTTGAAGTCGGCTTCG	GGTACCTTTAAATAATACACTAAAA
II	538	GCCCCCGCCCGTTGACTTCAGGGGGGAAACAA	TGCGTGCATTCTACCCTTTGA	GCTCTTCCTCCCAAGATA
IV	1286	ACCTTGGGAACATCGGACTATT	GCCCCCGCCCGCATGACATCGTGTTATAAGAGTTCAGA	GTGTTGGTTTTTCACAAT
IV	1306	TTAGTCACAATCCGAAGTCCG	GCCCCCGCCCGGTACCGCTGAACTATCGGATG	CAGATAGCGATACAACCT
IV	1321	CCGGTGGCAAAAGTAAGAACAA	GCCCCCGCCCGGCCCACTGTTTGCTTTTTGGTA	AAAAGCTAGACCTAGATTGA
XV	171	CGATTTATCATTGCCTACTTTTGC	GCCCCCGCCCGGGCCTAGCCATTATCCAAAACAT	CATTACAGCGCCAAA

Supplementary Table 14. QTL mapping resolution for simulated data.

fq	**0.99**	**0.9**	**0.8**	**0.65**
***wi* (cM)**	159.6	139.2	110.4	41.2
***wp* (cM)**	4.8	6.6	8.4	12.8

The influence of the final allele frequency *fq* on the mapping resolution quantified by the width of the interval inferred to be changing in allele frequency (*wi*), and the width of the QTL peak (*wp*). See Supporting Text for details on the simulation procedure. To give intuition for correspondance with physical map, under the assumption of 100 uniformly distributed recombination events per segregant for 12^{th} generation, one centimorgan corresponds to roughly one kilobase.

Supporting References

Ballester, R., Marchuk, D., Boguski, M., Saulino, A., Letcher, R., Wigler, M., and Collins, F. 1990. The NF1 locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins. *Cell* **63**(4): 851-859.

Borts, R.H., Lichten, M., and Haber, J.E. 1986. Analysis of meiosis-defective mutations in yeast by physical monitoring of recombination. *Genetics* **113**(3): 551-567.

Carter, D.M. 2006. Saccharomyces Genome Resequencing Project. http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* **35**(6): 2013-2025.

Cubillos, F.A., Billi, E., Zorgo, E., Parts, L., Fargier, P., Omholt, S., Blomberg, A., Warringer, J., Louis, E.J., and Liti, G. 2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol Ecol*.

Cubillos, F.A., Louis, E.J., and Liti, G. 2009. Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. *FEMS Yeast Res* **9**(8): 1217-1225.

Darvasi, A. and Soller, M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**(3): 1199-1207.

Dawes, I.W. and Hardie, I.D. 1974. Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether. *Mol Gen Genet* **131**(4): 281-289.

Ehrenreich, I.M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J.A., Gresham, D., Caudy, A.A., and Kruglyak, L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**(7291): 1039-1042.

Gil, R. and Seeling, J.M. 1999. Characterization of Saccharomyces cerevisiae strains expressing ira1 mutant alleles modeled after disease-causing mutations in NF1. *Mol Cell Biochem* **202**(1-2): 109-118.

Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**(5): 696-704.

Gustafsson, L. 1979. The ATP pool in relation to the production of glycerol and heat during growth of the halotolerant yeast Debaryomyces hansenii. *Archives of Microbiology* **120**(1): 15-23.

Hartl, D.L. and G., C.A. 2007. *Principles of Population Genetics*. Sinauer Associates, Inc.

Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., and Turner, D.J. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* **6**(4): 291-295.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V. et al. 2009. Population genomics of domestic and wild yeasts. *Nature* **458**(7236): 337-341.

Liti, G., Peruffo, A., James, S.A., Roberts, I.N., and Louis, E.J. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the Saccharomyces sensu stricto complex. *Yeast* **22**(3): 177-192.

Ng, P.C. and Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**(13): 3812-3814.
-. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**: 61-80.
Nikawa, J., Cameron, S., Toda, T., Ferguson, K.M., and Wigler, M. 1987. Rigorous feedback control of cAMP levels in Saccharomyces cerevisiae. *Genes Dev* **1**(9): 931-937.
Pacey-Miller, T. and Henry, R. 2003. Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. *Anal Biochem* **317**(2): 166-170.
Park, J.I., Grant, C.M., and Dawes, I.W. 2005. The high-affinity cAMP phosphodiesterase of Saccharomyces cerevisiae is the major determinant of cAMP levels in stationary phase: involvement of different branches of the Ras-cyclic AMP pathway in stress responses. *Biochem Biophys Res Commun* **327**(1): 311-319.
Simpson, J.T. and Durbin, R. 2010. Efficient construction of an assembly string graph using the FM-index. *Bioinformatics* **26**(12): i367-373.
Sniegowski, P.D., Dombrowski, P.G., and Fingerman, E. 2002. Saccharomyces cerevisiae and Saccharomyces paradoxus coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res* **1**(4): 299-306.
Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., and Davis, R.W. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**(6878): 326-330.
Szekvolgyi, L. and Nicolas, A. 2010. From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J* **277**(3): 571-589.
Tanaka, K., Matsumoto, K., and Toh, E.A. 1989. IRA1, an inhibitory regulator of the RAS-cyclic AMP pathway in Saccharomyces cerevisiae. *Mol Cell Biol* **9**(2): 757-768.
Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. 1994. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10**(13): 1793-1808.
Warringer, J. and Blomberg, A. 2003. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in Saccharomyces cerevisiae. *Yeast* **20**(1): 53-67.