Method

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

Leopold Parts,[1,6] Francisco A. Cubillos,[2] Jonas Warringer,[3,4] Kanika Jain,[2] Francisco Salinas,[2] Suzannah J. Bumpstead,[1] Mikael Molin,[3] Amin Zia,[5] Jared T. Simpson,[1] Michael A. Quail,[1] Alan Moses,[5] Edward J. Louis,[2] Richard Durbin,[1] and Gianni Liti[2,6]

[1]The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; [2]Centre for Genetics and Genomics, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, United Kingdom; [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), 1432 Ås, Norway; [5]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario M5S 2J4, Canada

One approach to understanding the genetic basis of traits is to study their pattern of inheritance among offspring of phenotypically different parents. Previously, such analysis has been limited by low mapping resolution, high labor costs, and large sample size requirements for detecting modest effects. Here, we present a novel approach to map trait loci using artificial selection. First, we generated populations of 10–100 million haploid and diploid segregants by crossing two budding yeast strains of different heat tolerance for up to 12 generations. We then subjected these large segregant pools to heat stress for up to 12 d, enriching for beneficial alleles. Finally, we sequenced total DNA from the pools before and during selection to measure the changes in parental allele frequency. We mapped 21 intervals with significant changes in genetic background in response to selection, which is several times more than found with traditional linkage methods. Nine of these regions contained two or fewer genes, yielding much higher resolution than previous genomic linkage studies. Multiple members of the RAS/cAMP signaling pathway were implicated, along with genes previously not annotated with heat stress response function. Surprisingly, at most selected loci, allele frequencies stopped changing before the end of the selection experiment, but alleles did not become fixed. Furthermore, we were able to detect the same set of trait loci in a population of diploid individuals with similar power and resolution, and observed primarily additive effects, similar to what is seen for complex trait genetics in other diploid organisms such as humans.

[Supplemental material is available for this article. The sequence data from this study have been submitted to the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) under study accession no. ERP000500, with individual sample accession numbers provided in Supplemental Table S3.]

A central challenge of modern genetics is to identify genes and pathways responsible for variation in quantitative traits. In the last decade, efforts of global international collaborations have revealed numerous loci that influence disease risk in humans by genotyping and phenotyping very large cohorts of individuals. However, the effects of single alleles are almost all modest, and explain only a small portion of the heritable variability (Manolio et al. 2009). Furthermore, while trait loci are found, association peaks generally span a large region, and do not point to the underlying mechanism responsible for the association. Thus, studies in model organisms, where consequences of genetic variation can be analyzed using reverse genetic tools, have been important for understanding the genetics of complex traits (Yvert et al. 2003; Deutschbauer and Davis 2005; Perlstein et al. 2006; Nogami et al. 2007; Demogines et al. 2008; Sinha et al. 2008; Smith and Kruglyak 2008; Gerke et al. 2009; Liti et al. 2009b; Romano et al. 2010).

Mapping the effect of naturally occurring alleles on traits is not straightforward even in model organisms (Hunter and Crawford 2008). Designed crosses often use substantially manipulated laboratory strains (Brem et al. 2002; Steinmetz et al. 2002) and produce

segregants that have to be laboriously genotyped and phenotyped. Linkage analysis on the resulting individuals can suffer from low resolution due to a limited number of crossover events (Darvasi and Soller 1995), but more rounds of crossing alleviate the problem (Wang et al. 2003). Developing and maintaining sufficiently large outbred populations to resemble human cohorts used in association mapping is costly (Valdar et al. 2006).

Recently, analysis of a very large pool of recombinant yeast strains has been used to identify quantitative trait loci (QTLs) for multiple traits without characterizing individual segregants (Segrè et al. 2006; Ehrenreich et al. 2010; Wenger et al. 2010). While many QTLs were detected, the problem of finding all responsible loci and localizing the trait genes within the linkage regions, which typically span many genes, remains. Furthermore, such analyses in yeast have previously been limited to haploid samples, in which genetic architecture may differ from that in diploids. Here, we present a precise and sensitive approach to QTL mapping, extending the method recently proposed by Ehrenreich et al. (2010) to sensitively identify trait loci at high resolution, in some cases down to single genes, in both haploid and diploid populations.

Results and Discussion

Strategy for high resolution QTL mapping

We used a three-step process for QTL mapping (Fig. 1). First, we generated very large pools of progeny between two phenotypically
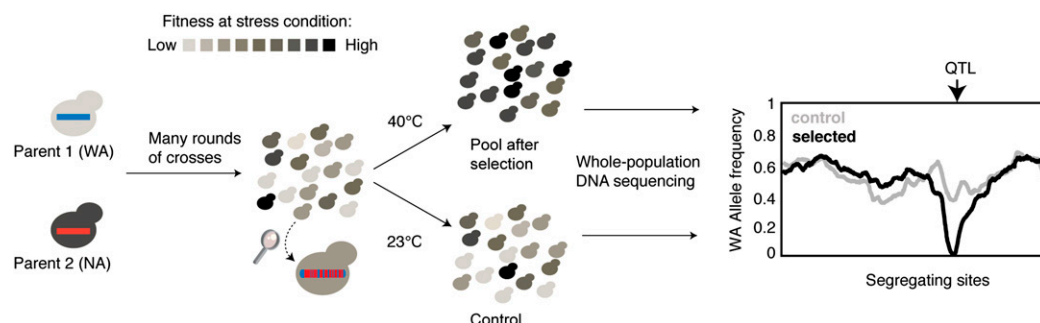
Figure 1. Overall strategy. A three-step QTL mapping strategy by crossing two phenotypically different strains for many generations to create a large segregating pool of individuals of various fitness, and growing the pool in a restrictive condition that enriches for beneficial alleles that can be detected via sequencing total DNA from the pool.

different yeast strains. We chose YPS128, a heat tolerant North American (NA) oak tree bark strain, and DVGBP6044, a heat sensitive West African (WA) palm wine strain as parents, and placed a different selectable marker at the same genomic position in each (Supplemental Table S1). We then systematically forced the yeast cells through multiple rounds of random mating and sporulation (Methods; Supplemental Fig. S1; Supplemental Material SI), creating advanced intercross lines (AILs) with reduced linkage between nearby loci. We produced haploid as well as heterozygous diploid pools of sixth and 12th generation progeny (F6 and F12 AILs), consisting of 10–100 million random segregants each.

We applied selective pressure to half of each pool by growing it asexually in a restrictive condition (40°C), to enrich for fit individuals with beneficial alleles. In parallel, we grew the other half of the pool in a permissive condition (23°C) as a control. Finally, we sequenced the pools before and at multiple timepoints during selection to directly assess the changes in parental allele frequencies throughout the genome. The entire procedure was performed in two biological replicates starting from the same F1 hybrid.

Genetic content of AIL pools before selection

We confirmed that the AIL pools were well-suited for QTL mapping. First, we established that nearly all loci were segregating before selection, with both parental alleles represented. Sequencing total DNA followed by allele frequency estimation (Methods) showed that >99% of the mappable genome was segregating in

the F6 pool, and 97% in the F12 pool with minor allele frequency >10% (Supplemental Material SI). Some other parts of the genome with minor allele frequencies >10% were also selected for during the intercross rounds without reaching fixation, likely due to alleles favoring sporulation, mating, or resistance to selection steps used in the cross (Supplemental Material SI). This allowed us to map several regions involved in these processes as a byproduct of our approach, including a previously uncharacterized sporulation QTL in chromosome V (Supplemental Table S2; Supplemental Material SI).

Using many rounds of crosses should expand the genetic map due to reduced linkage between nearby loci (Fig. 2A; Darvasi and Soller 1995). To confirm this, we genotyped 30 markers in 96 individual segregants from each of three generations, F1, F6, and F12, in three regions to assess the change in recombination fraction between adjacent markers (Data set 1). The genetic distance (measured as 100 times the average number of recombination events) between two chromosome XIII loci separated by 204 kb increased from 88 in F1 to 125 in F6 and 180 in F12 (Fig. 2A). This increase is less than expected under the assumption of independent recombination events and random mating, and is likely due to the considerable variation in recombination rates along the genome, with about half of the events concentrated in recombination hotspots (discussed in Supplemental Material SI). We further sequenced two segregants from the F6 pool at low coverage (Supplemental Material SI) and observed 64 and 68 recombination events (Fig. 2B), an 125% increase compared to 30 events observed
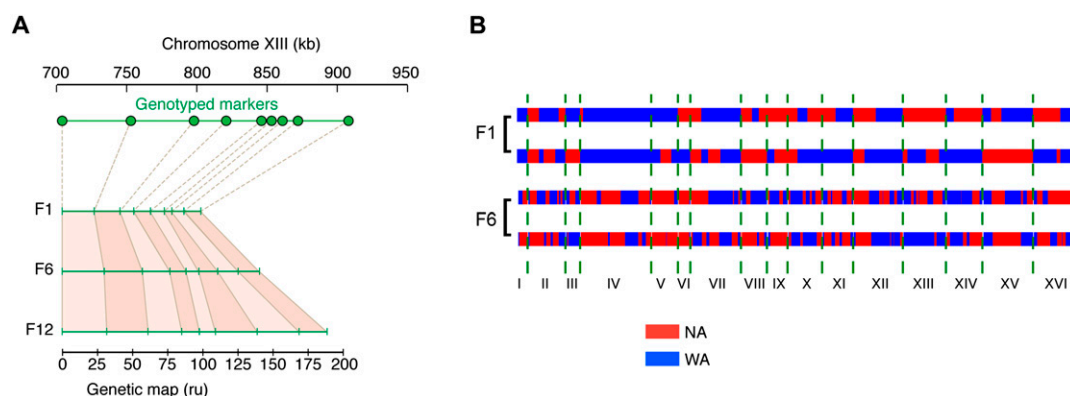
Figure 2. Recombination landscape after multiple rounds of intercrosses. (*A*) Expansion of the genetic map, measured in recombination units (ru) of 100 times the average number of recombination events from first to 12th generation (*bottom*) of a 200-kb chromosome XIII locus genotyped at nine markers (*top*). (*B*) Genetic background of two segregants from a first (F1) and sixth (F6) generation cross shows a sharp increase in recombination events.

on average in a set of 96 F1 segregants genotyped at ~200 evenly spaced loci (Cubillos et al. 2011).

Changes in population allele frequency in response to heat selection indicate protective alleles

To identify the alleles underlying variation in high temperature tolerance, we sequenced DNA from the F12 haploid pool to an average genome coverage of 25× to 150× (Supplemental Table S3) after 0 (T0), 96 (T1), 192 (T2), and 288 (T3, at least 25 generations, discussed in Supplemental Material SI) hours of growth at 40°C. There were 21 regions where the inferred allele frequency of the T2 pool was significantly different compared to the control experiment propagated at 23°C in both of the two biological replicates (Fig. 3A–C; Supplemental Table S4; Methods). We designated these regions as QTLs, with corresponding false positive rate
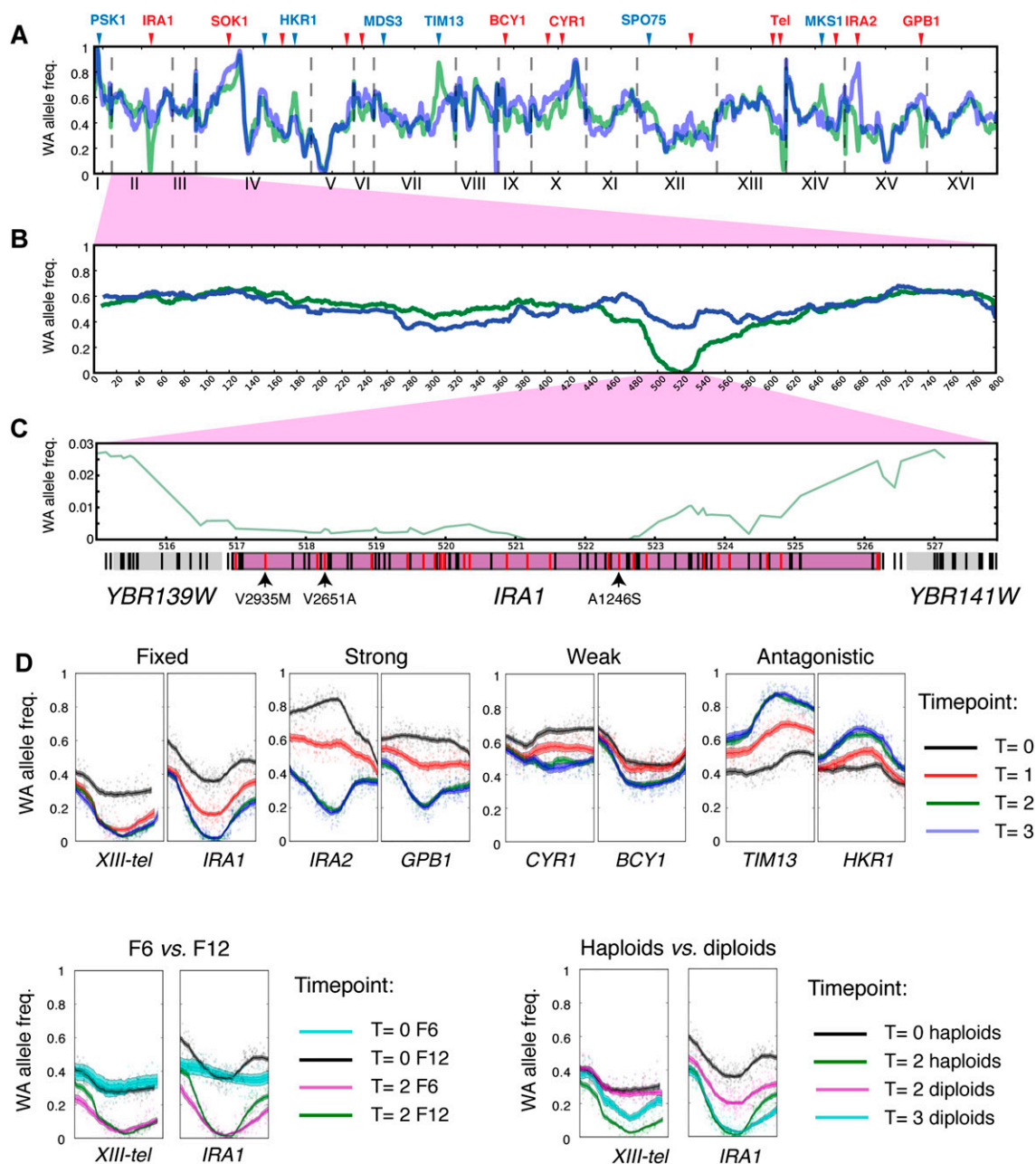
Figure 3. Changes in allele frequencies pinpoint QTLs. (*A–C*) WA allele frequency of whole genome (*A*), chromosome II (*B*), and *IRA1* region (*C*) of the F12 pool before (blue) and after (green) selection. Lines in gene regions in *C* denote segregating sites (black) and nonsynonymous segregating sites (red). The sites with intolerable mutations determined by SIFT analysis (Supplemental Material SI; Data set S3) are highlighted with arrows and designated with the amino acid change. (*D*) Individual examples of mapped QTLs that show differences in QTL strength, beneficial allele, effect of intercross rounds, and ploidy. Each window spans 80 kb and is centered on the locus with the largest allele frequency change in F12 T2 across two replicas. Shaded regions indicate 90% and 95% confidence intervals of the allele frequencies (Supplemental Material SI).

$<10^{-3}$ determined from the changes of allele frequencies in the control (Methods). As we chose this particular cutoff conservatively, there remains a set of lower confidence QTLs with smaller allele frequency changes that we do not consider here.

The NA allele was selected in two thirds (14/21) of the cases, consistent with it being the more heat resistant strain (Liti et al. 2009a). However, seven WA alleles were also selected, indicating antagonistic variants whose effect could only be observed when decoupled from the rest of the genetic context. The selected alleles are specific to the heat stress condition, as the same pool propagated at 23°C had no significant changes in allele frequency after 192 h, and selection under oxidative stress (paraquat, 1.5 mM) yielded a different set of loci with large allele frequency changes (Supplemental Fig. S2).

There is a risk that long term culturing under stress conditions will select for adaptive mutations rising to high frequencies and dominating the pool. However, theoretical models suggest this result is unlikely (Supplemental Material SI). We genotyped 960 segregants from the F12 haploid pool after 240 h of selection (T2.5) at 24 loci (Methods, Data set 2), and observed 787 unique haplotypes, with no haplotype (meaning here the set of QTL genotypes) represented more than six times. This is consistent with expectation under the model of independent selected segregants; if an adaptive clone had become dominant, its haplotype would have risen to high frequency (Supplemental Material SI). This suggests that the selection we observe is against haplotypes that cannot survive or grow at high temperature, rather than for a haplotype that performs better than all the others.

In addition to changes in chromosomal allele frequencies, all the mitochondrial genes were greatly reduced in copy number upon heat selection (Supplemental Table S5). Interestingly, 960/960 of the genotyped segregants exhibited a petite, nonrespiration phenotype when grown using a nonfermentable carbon source (glycerol and ethanol), indicating the loss of a mitochondrial genome, most probably as a response to the accumulation of reactive oxygen species (ROS) during heat stress (Davidson and Schiestl 2001; Rikhvanov et al. 2001). In contrast, none of the 96 F6 or F12 segregants isolated before heat selection exhibited the petite phenotype.

Prolonged artificial selection on AILs improves sensitivity and resolution

The additional value in using prolonged artificial selection is twofold. First, it allows alleles with smaller fitness effects to rise in frequency and become detectable (Fig. 3D). Only 10 of the 21 QTLs had significantly changed in allele frequency compared to the control experiment during the first 96 h under selection (T1), indicating that a longer experiment is required to find all QTLs (Supplemental Table S6). Second, extended selection can drive allele frequencies to an equilibrium, with no further significant changes taking place apart from random drift. Indeed, we observed that the largest inferred allele frequency change between T2 and T3 was 4%, and frequencies of 19/21 beneficial alleles changed <3% (Supplemental Table S6). We estimate by resampling that no changes >5% take place after T2 even though frequencies of 19 QTL alleles are >13% from fixation (Supplemental Material SI; Supplemental Table S6). Possible further interpretations of this are discussed below.

Our method is more sensitive than conventional linkage mapping. Analysis of 96 genotyped and phenotyped F1 segregants from the same cross has previously found one strong QTL for growth in high temperature at the right end of chromosome XIII (Cubillos
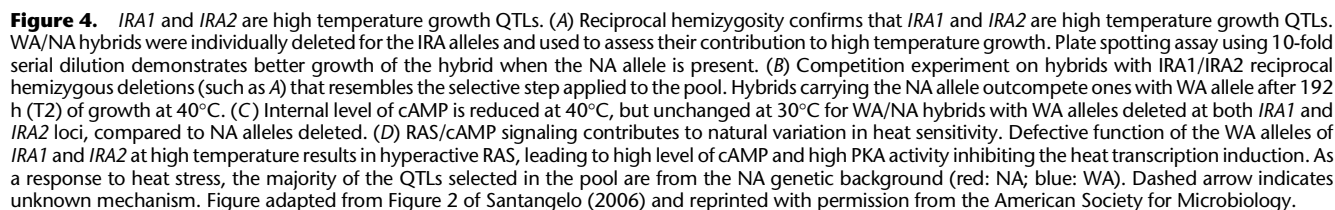
et al. 2011) (log-odds [LOD] score: 12.8, variance explained 66%). However, none of the other 20 QTL regions had a LOD score above the 5% FDR cutoff.

The advantage of using AILs with reduced linkage is evident from narrow mapped intervals, in some cases localizing to single genes (Fig. 3C,D; Supplemental Fig. S3). We designated mapped intervals as the segregating sites for which the allele frequency change is within one standard deviation of the change of the QTL peak (Supplemental Material SI; Supplemental Table S4). The resulting regions had a median size of 6.4 kb, and overlapped a median number of four genes (Supplemental Table S4), in contrast to regions selected during the intercross, which had a median size 16.3 kb, overlapping a median number of 10.5 genes (Supplemental Table S2). For example, for a chromosome II QTL, we could visually map the selected variant down to a small region of the *IRA1* gene (Fig. 3C), that also harbors the strongest candidate genetic variant from bioinformatic analysis (Supplemental Tables S7, S8; Supplemental Fig. S4). This resolution is in contrast to that from previous studies based on crosses between strains, including Ehrenreich et al. (2010), which typically map to large regions containing many genes, and can be further improved by using more advanced statistical methods for estimating the selected interval. The effects of many rounds of crossing and prolonged selection on mapping resolution have been quantitatively addressed in earlier work (Darvasi and Soller 1995) and are explored in simulation studies in the Supporting Information.

The RAS/cAMP signaling pathway regulates quantitative growth at high temperature

The two loci that were fixed in the selected population were *IRA1* on chromosome II and a subtelomeric locus of chromosome XIII (Fig. 3). *IRA1* is a GTPase-activating protein that negatively regulates the RAS signaling module (Tanaka et al. 1989). The RAS pathway is a general hotspot for natural variation (Smith and Kruglyak 2008), a known target of adaptive mutation (Kao and Sherlock 2008), and a regulator of protein kinase A (PKA) that in turn controls the synthesis and activity of stress response genes (Santangelo 2006). Interestingly, three additional regulators of the same RAS/cAMP module (*IRA2*, *GPB1*, and *CYR1*), as well as some of its targets (*BCY1*), were contained in QTL intervals with an increase in NA allele frequencies in the F12 pool, indicating the involvement of PKA signaling pathway components both upstream and downstream of cAMP in the heat resistance phenotype (Fig. 4D).

We validated the three strongest mapped QTLs. The QTL at the right end of chromosome XIII was previously confirmed by decreased growth rate of the F1 hybrid of parents upon truncation of the NA subtelomere (Cubillos et al. 2011). However, due to lack of assembled sequence in the region in the founder strains for this experiment, we could not identify the responsible gene. We validated by reciprocal hemizygosity (Steinmetz et al. 2002), deleting both alleles from a diploid hybrid in turn, that *IRA1* and *IRA2* alleles affect high temperature growth. The effect was evident from a plating assay, growth curves, and competition experiments (Fig. 4A,B; Supplemental Fig. S5A,B). These genes affect both growth rate (doubling time) and efficiency (final density) of segregants, with *IRA1* having a stronger effect than *IRA2*, consistent with the difference in their final allele frequencies (Supplemental Fig. S5A,B). Interestingly, the same *IRA1* and *IRA2* alleles do not have a strong differential effect on growth in other stress conditions (Supplemental Fig. S5C), even though RAS activity is involved in the response to these stresses (Park et al. 2005).

Figure 4. *IRA1* and *IRA2* are high temperature growth QTLs. (*A*) Reciprocal hemizygosity confirms that *IRA1* and *IRA2* are high temperature growth QTLs. WA/NA hybrids were individually deleted for the IRA alleles and used to assess their contribution to high temperature growth. Plate spotting assay using 10-fold serial dilution demonstrates better growth of the hybrid when the NA allele is present. (*B*) Competition experiment on hybrids with IRA1/IRA2 reciprocal hemizygous deletions (such as *A*) that resembles the selective step applied to the pool. Hybrids carrying the NA allele outcompete ones with WA allele after 192 h (T2) of growth at 40°C. (*C*) Internal level of cAMP is reduced at 40°C, but unchanged at 30°C for WA/NA hybrids with WA alleles deleted at both *IRA1* and *IRA2* loci, compared to NA alleles deleted. (*D*) RAS/cAMP signaling contributes to natural variation in heat sensitivity. Defective function of the WA alleles of *IRA1* and *IRA2* at high temperature results in hyperactive RAS, leading to high level of cAMP and high PKA activity inhibiting the heat transcription induction. As a response to heat stress, the majority of the QTLs selected in the pool are from the NA genetic background (red: NA; blue: WA). Dashed arrow indicates unknown mechanism. Figure adapted from Figure 2 of Santangelo (2006) and reprinted with permission from the American Society for Microbiology.

We further constructed all four possible double hemizygous combinations of *IRA1* and *IRA2* and grew them at 40°C. The doubling time for the strain with WA alleles was 3.3-fold longer compared to the multiplicative expectation (Supplemental Material SI), indicating a negative epistatic interaction between these alleles, which is consistent with partially redundant function (Supplemental Fig. S6). Furthermore, at 40°C we detected a ninefold higher level of internal cAMP in the hybrid carrying WA alleles of the IRA genes compared to the hybrid carrying NA alleles ($P < 0.005$, one-sided t-test) (Fig. 4C). This is consistent with RAS hyperactivity in the WA strain due to loss of negative regulation from IRA genes, resulting in a higher level of internal cAMP, contributing to its heat sensitivity (Fig. 4D). The RAS/cAMP signaling pathway has previously been implicated in the accumulation of ROS under stress (Hlavata et al. 2003). Our results are consistent with this functionality being under selection alongside the selection for removal of the mitochondria.

Different QTL haplotypes are maintained after selection

It is surprising that for 19 QTLs both alleles remained segregating in the pool after 192 h under selection (T2), and did not significantly change in frequency in a further 96 h in the restrictive condition (Supplemental Figs. S7, S8; Supplemental Table S6). This suggests that we have saturated for individual alleles with strong independent effects that are present in the founding strains. It also indicates that all the haplotypes remaining in the pool have nearly equal fitness in this stress condition, or are so rare even by T3 that change in their frequency does not have a major effect on the average pool genotype. These observations are consistent with an abundance of negative epistatic interactions, where particular combinations of alleles are selected against.

To test this explanation, we isolated and genotyped 960 segregants from the F12 pool after 240 h of selection (T2.5), and looked for scarcity and abundance of specific allele combinations at 19 QTLs (Supplemental Table S9, Data set 2, and Supplemental Material SI). None of the interchromosomal two-locus genotype combinations was significantly different from the expectation under independence after correcting for multiple testing (lowest multiple testing corrected $P > 0.1$, two-sided Fisher's exact test, Supplemental Table S9).

The apparent lack of allele fixation and strong interchromosomal interactions after 12 d under selection could be explained by various models involving one or more loci. First, it is possible that individual beneficial alleles are still changing in frequency, albeit slowly, and would fix if the selection was carried out for a much longer period. Second, there could be two-locus epistasis within a linked region that we are unable to detect from the genotyping data. Finally, the data are also consistent with complex control and interactions involving multiple genes (Phillips 2008).

QTL mapping in a heterozygous diploid population

Importantly for drawing comparisons with genetic studies of complex traits in humans and other diploid organisms, 18 of 21 heat resistance QTL alleles significantly changed in frequency in the pool of heterozygous diploid individuals (Supplemental Table S6; Data set 3). The process of selection was slower for the diploid pool, as allele frequencies continued to change between T2 and T3 (Supplemental Table S6; Fig. 3D). Examining the diploid pool allele frequency after selection at T3, the NA chrXIII QTL allele is dominant with only the homozygous deleterious genotype (WA/WA) being removed from the pool, and the NA *IRA1* allele is recessive with the beneficial allele (NA/NA genotype) being fixed. For 10 of the other 16 detected loci it is surprising that the allele frequencies after 12 d of selection on the diploid segregants were within 5% of the haploid pool after selection (Supplemental Table S6; Data set 3), consistent with the selected alleles having additive effects as observed for most human GWAS hits (Manolio et al. 2009). However, as we did not observe data from consecutive timepoints without significant allele frequency changes for the diploid pools, it is also possible that the allele frequencies had not yet reached equilibrium.

Extensions and applications of the method

It is straightforward to apply our method to any selectable trait, including ones that do not affect fitness. For example, cell sorting (Ehrenreich et al. 2010) to select for cell size or GFP expression on specific promoters, ageing the population for detecting chronological life span QTLs (Fabrizio and Longo 2003) or washing the plate to detect cell adhesion traits (Reynolds and Fink 2001), can all be used. A similar approach may also be adapted to other model genetic systems, including *Drosophila melanogaster* and *Caenorhabditis elegans*, which are amenable to crossing in bulk and large experimental population sizes. In these other models, the ratio of recombination rate to gene density determines the ability to identify responsible genes.

The heterozygous diploid intercross pool can be used for other genetic studies, such as dissecting mechanisms that contribute to heterosis (Lippman and Zamir 2007), and extended to include more of the genetic diversity in the species. We expect to be able to cross a larger number of parental strains to expand the range of standing variation segregating in the pool. Therefore, we have the potential to establish an artificial outbred yeast population that can be used as a model for natural diploid genome-wide association studies as carried out in humans.

Methods

Intercross

Parental strains YPS128 (*MAT*a, *ho::HphMX4*, *ura3::KanMX4*) and DBVPG6044 (*MAT*α, *ho::HphMX4*, *ura3::KanMX4*, *lys2::URA3*) were crossed in complete media (YPDA) and grown overnight. Patches were replica plated in synthetic minimal media (MIN) to select for diploid F1 hybrids. F1 hybrids were isolated and stored at −80°C. Two F1 hybrid replicas were grown overnight and replica plated on KAc at 23°C to be sporulated for 10 d until 90% sporulation efficiency. Cells were collected and resuspended in 0.5 mL of sterile water, treated with an equal amount of ether and vortexed for 10 min to kill unsporulated cells (Dawes and Hardie 1974). Cells were washed four times in sterile water, resuspended in 900 μL of sterile water, and treated with 100 μL of Zymolase (10 mg/mL) to remove ascus. Cell mixtures were vortexed for 5 min to increase spore dispersion and inter-ascus mating. For the heterozygous diploid pool, we forced one extra round of mating and selection for *LYS+/URA3+* cells. Full details are presented in Supplemental Material SI.

Selection experiment

Pools of 10–100 million cells were collected from sporulation media and treated with ether and zymolase. Spores were plated in YPDA and incubated until full growth was obtained. Each plate was incubated for 48 h, and then resuspended in distilled water. Ten percent of the cells were used for next replating, and the rest for DNA extraction.

Genotyping

SNP genotypes were obtained by real time PCR coupled to high resolution melting (HRM) using the Corbet Rotorgene and Quantace PCR HRM mix. Sequenom genotyping of 960 segregants was performed using the iPLEX Gold Assay (Sequenom Inc.) (Supplemental Material SI).

DNA isolation, library preparation, and sequencing

DNA was extracted using the phenol chloroform protocol. Multiplexed PCR-free Illumina sequencing libraries were prepared as in Kozarewa et al. (2009) with modifications (Supplemental Material SI). Fragments with 200–300-bp inserts were gel-purified and sequenced using standard Illumina SBS v4 chemistry for 2×76 cycles plus extra seven cycles to determine the tag sequence of each cluster.

Sequencing data handling

Sequencing reads were mapped to the S288c reference genome obtained from the SGRP project website (http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html) using BWA (Li and Durbin 2009), with option "-n 8." Pileup files comprising the genotypes of mapped reads were created for segregating sites inferred from both low-coverage capillary sequencing (Liti et al. 2009a) and the parental strain shotgun sequence mapping to the S288c assembly, and sites further filtered.

Parental strain analysis

The parental strain sequence was mapped similarly to the selection experiment. We used the SAMtools (Li et al. 2009) 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.

Segregant analysis

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 30 kb, and variants with ambiguous mapping data. We called a recombination event if a region of at least 2 kb from one parent was followed by a region of at least 2 kb from the other, and at least five calls were made in both regions.

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 ai 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 $qi = \exp(-d_i r)$, 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

$$P(D \mid fl) = \prod_i P(ai \mid fl), \text{ where}$$

$$P(ai \mid fl) = P(ai, bi = \text{'WA'} \mid fl) + P(ai, bi = \text{'NA'} \mid fl)$$

$$= P(ai \mid bi = \text{'WA'})P(bi = \text{'WA'} \mid fl) + P(ai \mid bi = \text{'NA'})P(bi = \text{'NA'} \mid fl)$$

$$= qi^{ai = \text{'WA'}}(1 - qi)^{ai = \text{'NA'}}fl + qi^{ai = \text{'NA'}}(1 - qi)^{ai = \text{'WA'}}(1 - fl)$$

$$\approx qi\,fl^{ai = \text{'WA'}}(1 - fl)^{ai = \text{'NA'}}.$$

Here, we have discarded likelihood terms that require a recombination event, as we will filter qi to be large. We calculated the posterior (beta) distribution of f_l by applying Bayes rule: $P(f_l \mid A) \propto P(A \mid f_l)\,P(f_l) = \prod_i P(ai \mid f_l)\,P(f_l)$, where the beta prior $P(f_l)$ is uninformative, and we filter $qi > 0.9$ (0.75 for Fig. 3A,B for smoothness). This inference procedure corresponds to a smoothing approach within a fixed window with the width determined by the recombination rate ($\sim$6 kb for $qi > 0.9$), and has the effect of discriminating against extreme allele frequencies. The posterior mean and confidence intervals were obtained from the approximated Beta distribution.

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 50-kb window, corresponding to the variant with the 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 144 h (T2). Due to the repetitive nature of subtelomeric regions resulting in a lack of assemblies and low sequencing coverage, we did not consider loci within 30 kb of the end of chromosomes. We fit a normal distribution to the allele frequency changes at the 26,871 loci assessed (Supplemental Fig. S9), and calculated the probability of observing a change of at least 10% in either direction to be $<10^{-7}$. After Bonferroni-correcting for the 26,871 tests, the P-value remained $<10^{-3}$.

FI segregants

We used standard marker regression for 200 genotyped markers and heat growth rate phenotype to map QTLs significant at 5% false discovery rate (FDR) using a standard linear model and 1000 permutations in rQTL (Broman et al. 2003).

Interaction tests

We tested for scarcity and abundance of two-locus genotype combinations by the two-sided Fisher's exact test using the fisher. test function in R on the two-locus genotype counts for each pair of genotyped loci.

Reciprocal hemizygosity

IRA1 and *IRA2* were deleted individually or in the four possible combinations (Supplemental Table S1) in hybrid strain YCC22F using the standard single-step PCR gene deletion method (Wach et al. 1994). We performed a temperature growth assay by plating serial dilutions of cells in YPDA and incubated the plates at 30°C and 40°C for 48 h.

Competitive growth

We competed the reciprocal hemizygous hybrids (naΔ/WA vs. waΔ/NA) by mixing equal numbers of cells, and growing at either 30°C or 40°C for 96 h. Pyrosequencing was used to assess the allele frequency in the pools.

Phenotyping

Individual yeast segregants and reciprocal hemizygous were phenotyped using high-resolution microcultivation instruments Bioscreen C (Growth curve Oy, Finland) for quantitative growth as previously described (Warringer and Blomberg 2003; Liti et al. 2009a).

cAMP determination

Intracellular cAMP was determined using a commercially available kit (LANCE cAMP 384 kit, Perkin-Elmer). Values presented are average values determined in three to six replicate cultures per strain and error bars indicate SEM.

Acknowledgments

References

Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296:** 752–755.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19:** 889–890.

Cubillos FA, Billi E, Zörgö E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G.2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol Ecol* **20:** 1401–1413.

Darvasi A, Soller M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141:** 1199–1207.

Davidson JF, Schiestl RH. 2001. Cytotoxic and genotoxic consequences of heat stress are dependent on the presence of oxygen in *Saccharomyces cerevisiae*. *J Bacteriol* **183:** 4580–4587.

Dawes IW, Hardie ID. 1974. Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether. *Mol Gen Genet* **131:** 281–289.

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. doi: 10.1371/journal.pgen.1000123.

Deutschbauer AM, Davis RW. 2005. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet* **37:** 1333–1340.

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.

Fabrizio P, Longo VD. 2003. The chronological life span of *Saccharomyces cerevisiae*. *Aging Cell* **2:** 73–81.

Gerke J, Lorenz K, Cohen B. 2009. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323:** 498–501.

Hlavata L, Aguilaniu H, Pichova A, Nystrom T. 2003. The oncogenic RAS2(val19) mutation locks respiration, independently of PKA, in a mode prone to generate ROS. *EMBO J* **22:** 3337–3345.

Hunter KW, Crawford NP. 2008. The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. *Annu Rev Genet* **42:** 131–141.

Kao KC, Sherlock G. 2008. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nat Genet* **40:** 1499–1504.

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

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

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

Lippman ZB, Zamir D. 2007. Heterosis: Revisiting the magic. *Trends Genet* **23:** 60–66.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. 2009a. Population genomics of domestic and wild yeasts. *Nature* **458:** 337–341.

Liti G, Haricharan S, Cubillos FA, Tierney AL, Sharp S, Bertuch AA, Parts L, Bailes E, Louis EJ. 2009b. Segregating *YKU80* and *TLC1* alleles underlying natural variation in telomere properties in wild yeast. *PLoS Genet* **5:** e1000659. doi: 10.1371/journal.pgen.1000659.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747–753.

Nogami S, Ohya Y, Yvert G. 2007. Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS Genet* **3:** e31. doi: 10.1371/journal.pgen.0030031.

Park JI, Grant CM, Dawes IW. 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:** 311–319.

Perlstein EO, Ruderfer DM, Ramachandran G, Haggarty SJ, Kruglyak L, Schreiber SL. 2006. Revealing complex traits with small molecules and naturally recombinant yeast strains. *Chem Biol* **13:** 319–327.

Phillips PC. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9:** 855–867.

Reynolds TB, Fink GR. 2001. Bakers' yeast, a model for fungal biofilm formation. *Science* **291:** 878–881.

Rikhvanov EG, Varakina NN, Rusaleva TM, Rachenko EI, Kiseleva VA, Voinikov VK. 2001. [Heat shock-induced changes in the respiration of the yeast *Saccharomyces cerevisiae*]. [Article in Russian]*Mikrobiologiia* **70:** 531–535.

Romano GH, Gurvich Y, Lavi O, Ulitsky I, Shamir R, Kupiec M. 2010. Different sets of QTLs influence fitness variation in yeast. *Mol Syst Biol* **6:** 346. doi: 10.1038/msb.2010.1.

Santangelo GM. 2006. Glucose signaling in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **70:** 253–282.

Segrè AV, Murray AW, Leu JY. 2006. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol* **4:** e256. doi: 10.1371/journal.pbio.0040256.

Sinha H, David L, Pascon RC, Clauder-Munster S, Krishnakumar S, Nguyen M, Shi G, Dean J, Davis RW, Oefner PJ, et al. 2008. Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* **180:** 1661–1670.

Smith EN, Kruglyak L. 2008. Gene-environment interaction in yeast gene expression. *PLoS Biol* **6:** e83. doi: 10.1371/journal.pbio.0060083.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416:** 326–330.

Tanaka K, Matsumoto K, Toh EA. 1989. IRA1, an inhibitory regulator of the RAS-cyclic AMP pathway in *Saccharomyces cerevisiae*. *Mol Cell Biol* **9:** 757–768.

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* **38:** 879–887.

Wach A, Brachat A, Pohlmann R, Philippsen P. 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10:** 1793–1808.

Wang X, Le Roy I, Nicodeme E, Li R, Wagner R, Petros C, Churchill GA, Harris S, Darvasi A, Kirilovsky J, et al. 2003. Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. *Genome Res* **13:** 1654–1664.

Warringer J, Blomberg A. 2003. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* **20:** 53–67.

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. doi: 10.1371/journal.pgen.1000942.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. 2003. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* **35:** 57–64.

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

Leopold Parts, Francisco A. Cubillos, Jonas Warringer, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2011/03/18/gr.116731.110.DC1
References	This article cites 40 articles, 8 of which can be accessed free at: http://genome.cshlp.org/content/21/7/1131.full.html#ref-list-1
Open Access	Freely available online through the *Genome Research* Open Access option.
License	Freely available online through the Genome Research Open Access option.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.