Published in final edited form as:

Cold Spring Harb Symp Quant Biol. 2009 ; 74: 145–153. doi:10.1101/sqb.2009.74.013.

Genetic Dissection of Complex Traits in Yeast: Insights from Studies of Gene Expression and Other Phenotypes in the BY×RM Cross

I.M. Ehrenreich^{1,2,3}, J.P. Gerke^{1,2,3}, and L. Kruglyak^{1,2,3}

¹Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544

²Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544

³Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544

Abstract

The genetic basis of many phenotypes of biological and medical interest, including susceptibility to common human diseases, is complex, involving multiple genes that interact with one another and the environment. Despite decades of effort, we possess neither a full grasp of the general rules that govern complex trait genetics nor a detailed understanding of the genetic basis of specific complex traits. We have used a cross between two yeast strains, BY and RM, to systematically investigate the genetic complexity underlying differences in global gene expression and other traits. The number and diversity of traits dissected to the locus, gene, and nucleotide levels in the BY×RM cross make it arguably the most extensively characterized system with regard to causal effects of genetic variation on phenotype. We summarize the insights obtained to date into the genetics of complex traits in yeast, with an emphasis on the BY×RM cross. We then highlight the central outstanding questions about the genetics of complex traits and discuss how to answer them using yeast as a model system.

Many phenotypes in nature display continuous variation and complex genetic inheritance (Falconer and Mackay 1996). Mapping the genetic basis of such traits is difficult due to the combined effects of environmental variation and multiple genetic loci (Lynch and Walsh 1998). This complexity poses a challenge for answering questions about the number of loci governing a trait, the allele frequencies and distribution of effects of these loci, and the prevalence of genetic and gene–environment interactions.

Because of the historical challenge of mapping complex traits to their multiple underlying loci, it is unknown whether general rules exist for how complex traits are specified at the genetic level. Theory predicts that a wide range of genetic architectures can exist, depending on the fitness effects of the underlying loci and the selective pressures experienced by a population (Orr 2005). The current challenge lies in empirically defining the distribution of genetic architectures that actually exist, which can only be achieved through the genetic dissection of a large number of complex traits.

The identification of the individual genes and polymorphisms underlying complex traits is essential for a full understanding of genetic architecture. For example, if a single genomic

region causes variation in five traits, then only the identification of the causative gene(s) underlying a quantitative trait locus (QTL) will determine whether this architecture is due to five genes with single effects or one gene with effects on five traits. Furthermore, the identification of the specific DNA sequence polymorphisms that cause trait differences makes it possible to determine how trait variation arises at the molecular level, as well as how such causal polymorphisms are distributed across a population. Obtaining molecular resolution of complex traits remains a great challenge. However, this challenge is not insurmountable, and upcoming research holds great promise for uncovering the basic principles of trait variation.

The single-celled brewer's yeast *Saccharomyces cerevisiae* has many qualities that make it the organism of choice for the study of a large number of quantitative traits at molecular resolution. Its rapid generation time and ease of growth and maintenance, combined with the ability to keep frozen stocks indefinitely, mean that entire populations of strains can be repeatedly tested for different phenotypes under a variety of conditions. Once QTLs are identified, the small genome, high recombination rate, and the ease of site-directed mutagenesis of yeast (Storici et al. 2001) enable fine mapping and precise functional tests to determine causative genes and nucleotides.

Our lab has used a cross of two yeast strains to systematically dissect the genetic architecture of complex traits and in some cases identify causative alleles at gene and nucleotide resolution. The parents of this cross are BY4716, a common laboratory strain, and RM-11, a strain isolated from a California vineyard. These strains differ on average at 1 nucleotide every 200 base pairs (Ruderfer et al. 2006). More than 100 haploid meiotic segregants from the BY×RM F₁ hybrid provide independent recombinations of the two parental genomes (Fig. 1). These segregants have been genotyped genome-wide at thousands of markers and assayed for global gene expression and many other phenotypes. These data have provided general insights into the genetic architectures underlying complex traits. Here, we review the main lessons learned so far and examine how these lessons can guide future research on the genotype–phenotype relationship in yeast.

INSIGHTS FROM THE BY×RM CROSS

Using the BY×RM cross, our lab was the first to combine classical linkage analysis with DNA microarrays to dissect the genetic basis of global gene expression (Brem et al. 2002). Measuring gene expression across the genome enables the analysis of thousands of quantitative traits in parallel, thereby providing a route for the systematic study of genetic architecture (Fig. 2). In addition, gene expression provides a direct connection between DNA sequence and phenotype, and specific tests can determine if a gene's own sequence causes variation in its expression.

Once the basic principles of the genetics of gene expression were understood in yeast, this new approach to studying quantitative genetic variation was quickly expanded to other organisms (for review, see Rockman and Kruglyak 2006). Mapping the loci that cause gene expression variation in the BY×RM segregants has revealed a number of general insights into the genetic basis of gene expression variation and complex traits in general. Studying traits other than gene expression levels in the BY×RM cross has made it possible to connect expression variation to other phenotypes and build a more comprehensive picture of the relationship between genotype and phenotype.

Gene Expression Traits Show Complex Inheritance

A central finding of the BY×RM experiments is that most gene expression traits are complex, influenced by both multiple interacting genetic factors and the environment. Gene expression variation in the BY×RM mapping population is highly heritable, with a median heritability of 84% (Brem et al. 2002). More than 75% of all transcripts map to at least one QTL in the

environments studied so far (Brem et al. 2002; Smith and Kruglyak 2008). However, the combined effects of the detected linkages explain less than 30% of the genetic variance for more than half of the transcripts in the genome, implying that most of the causal loci remain undetected (Brem and Kruglyak 2005). This assertion is strengthened by two observations: (1) transgressive segregation is rampant in BY×RM segregants, meaning that multiple alleles of opposite effect are present in BY and RM (Fig. 2B), and (2) many highly heritable transcripts show no linkages, suggesting that we do not have the power to detect the causal genetic basis of many expression differences (Brem and Kruglyak 2005). Together, these results indicate that gene expression variation is primarily under genetic control, but the genetic basis of this control is due to a potentially large number of loci (Brem et al. 2002; Brem and Kruglyak 2005; Smith and Kruglyak 2008). Modeling of the transcript variation found in this cross supports a complex genetic basis for gene expression variation, suggesting that 50% of all transcripts have at least five additive QTLs and 20% of all transcripts have at least 10 additive QTLs (Brem and Kruglyak 2005).

Additional studies of the BY×RM cross have shown that both genetic and gene–environment interactions are prominent in shaping gene expression variation. Purely synthetic interactions, in which polymorphisms in multiple genes cause a phenotypic change without any of these genes having additive effects on their own, appear to be rare in the BY×RM cross (Storey et al. 2005). Nonetheless, genetic interactions are common in these data, with more than 57% of all transcripts influenced by a genetic interaction (Brem et al. 2005; Storey et al. 2005). In the majority of these interactions, one of the loci showed an additive effect, whereas the other had an effect too modest to be detected without accounting for the interaction (Brem et al. 2005; Storey et al. 2005). Recent evidence suggests that genetic interactions affecting gene expression can involve more than two loci (Litvin et al. 2009).

In addition to genetic interactions, many transcripts exhibit gene–environment interactions. Global gene expression was recently measured across the BY×RM segregants in two different carbon sources: glucose and ethanol. Nearly half of all transcripts (47%) were found to have a gene–environment effect (Smith and Kruglyak 2008). Many different types of gene–environment interactions were detected, including loci with effects only under a single condition, loci with effects in the same direction but of different magnitudes under the two conditions, and loci with effects in the opposite direction across the two conditions (Smith and Kruglyak 2008). Taken together, these results show that gene expression variation, like most complex trait variation, is due to a mixture of additive, nonadditive, and genotype-by-environment effects.

The Molecular Mechanisms of Gene Expression Variation

Because transcripts are transcribed from genes with specific genomic locations, linkages for expression traits can be separated into two classes: (1) local linkages that arise from polymorphisms in or near the differentially expressed genes themselves and (2) distant linkages that give rise to expression variation at unlinked transcripts (Fig. 2C). Local linkages are common in the BY×RM cross, with more than 25% of all genes in the yeast genome showing local linkage (Ronald et al. 2005). Classic *cis-trans* tests using allele-specific expression assays in a diploid hybrid of BY and RM demonstrated that between 52% and 78% of local linkages were due to *cis*-regulatory polymorphisms at the differentially expressed genes themselves (Ronald et al. 2005). Not all *cis* effects could be explained by polymorphisms in transcription-factor-binding sites, suggesting that variants in a number of sequence classes, including within transcripts and 3'UTRs (untranslated regions), can give rise to *cis* effects. Ronald et al. (2005) also showed that local linkages can act in *trans*, because allele-specific measurements for multiple transcripts were inconsistent with *cis* effects on expression. In addition, an amino acid polymorphism in *AMN1* was shown to cause an increase in *AMN1*'s expression through

a feedback loop. Population genetic modeling based on the *cis* linkages in the BY×RM cross suggests that if gene expression were mapped across a wider range of *S. cerevisiae* isolates, every gene in the genome might be found to have a *cis*-regulatory polymorphism (Ronald and Akey 2007).

Consistent with the complex genetic architecture of gene expression variants, extensive distant linkages are observed in the BY×RM cross. Many of these *trans* effects stem from a small number of hot spots that influence the transcript abundance of tens to hundreds of genes (Fig. 3) (Brem et al. 2002; Yvert et al. 2003; Smith and Kruglyak 2008). Eight hot spots, each being linked to 7–94 genes of related function, were observed in the initial study of this cross (Brem et al. 2002), and subsequent studies identified additional hot spots (Smith and Kruglyak 2008). Positional cloning and bioinformatic analyses of distant linkages at both *trans* hot spots and other loci have shown that these effects are often not due to transcription factor polymorphisms (Yvert et al. 2003).

Analysis of global gene expression in BY×RM segregants in both glucose and ethanol showed that distant and local linkages respond very differently across environments. Local effects were more consistent across environments, whereas distant effects exhibited greater sensitivity to environment. Thirteen *trans* hot spots were found in each environment, but only seven were common between the two environments, illustrating the potential importance of distant linkages and, more specifically, *trans* hot spots in mediating transcriptional responses to the environment (Smith and Kruglyak 2008). Strikingly, 78% of the time, distant linkages had opposite effects in glucose and ethanol, suggesting that distant linkages frequently display gene–environment interactions (Smith and Kruglyak 2008).

Genetic Dissection of Other Types of Traits in the BY×RM Cross

The BY×RM segregants have been used to map a variety of higher-order phenotypes reflecting different aspects of yeast cell biology and physiology. Along and in combination with the genetics of gene expression, the study of these traits has shed light on basic principles regarding the genetics of complex traits.

Gene expression differences should affect cellular phenotypes through changes in protein abundance. Measuring genetic variation in protein abundance is therefore a logical next step in the analysis of complex traits in yeast. Studying protein abundance, like gene expression, also allows the comparison of the effects of local and distant linkages. Mass spectrometry enables the analysis of many proteins in parallel. Foss et al. (2007) used this technique to measure variation in the abundance of ~200 proteins among the BY×RM segregants. The detected protein abundance linkages mirror the gene expression linkages in several ways. As with variation in gene expression, variation in protein abundance shows high heritability. The average heritability in protein abundance was 0.62, and ~100 linkages were identified. Despite high heritability, 38% of the peptides analyzed showed no significant linkages. Because most traits under the control of a single locus should show linkage with the statistical power available, this result means that like gene expression variation, variation in protein abundance likely has a complex genetic basis. Also consistent with previous results in the BY×RM cross, there are clear examples of transgressive segregation in protein abundance, and four hot spots were observed in which the abundance of at least five proteins links to the same genomic location.

The results of Foss et al. (2007) provide interesting comparisons between variation in transcript levels and protein abundance. There is significant overlap between genes that varied in both the transcript and protein data sets. But this overlap is not complete, because only 43% of genes with protein abundance differences between BY and RM showed detectable differences in gene expression. In addition, only three of the gene expression hot spots resulted in protein

abundance hot spots, and one protein abundance hot spot did not have a corresponding hot spot for gene expression. These results show that there are many cases where a gene expression difference does not cause a change in protein levels, and there are also cases where differences in protein abundance arise only from posttranscriptional effects.

Another way to analyze a large number of traits at once is through the application of libraries of small molecules. Perlstein et al. (2007) measured differences in the growth of the BY×RM segregants in response to a panel of 100 molecules that included various classes of compounds and some FDA-approved medications. Linkage analysis identified 124 QTLs involved in small-molecule response. Resistance to small molecules shows a substantial genetic component, because more than half of the compounds link to at least one QTL. Many compounds link to multiple QTLs, indicating a complex genetic basis. There are clear examples of transgressive segregation, and there is striking overlap between hot spots for small-molecule resistance and gene expression variation. Eight regions of the genome link significantly to at least four small molecules, and seven of these regions overlap with hot spots for gene expression variation. These results suggest that gene expression variation, protein abundance, and small-molecule resistance have similarly complex genetic architectures, and in some cases they may share a common functional basis. Accordingly, a missense mutation in *PHO84*, a phosphate transporter, underlies one of the small-molecule hot spots and may explain gene expression variation linking to this region (Perlstein et al. 2007).

Other traits studied in the BY×RM cross include telomere length (Gatbonton et al. 2006), morphological variation (Nogami et al. 2007), noise in gene expression (Ansel et al. 2008), and sensitivity to DNA-damaging agents (Demogines et al. 2008). In all cases, the genetic basis of variation is complex. This observation is consistent with results from additional crosses in other strain backgrounds, which have identified complex genetic variation in ethanol production (Hu et al. 2007), high-temperature growth (Steinmetz et al. 2002; Sinha et al. 2008), and sporulation efficiency (Deutschbauer and Davis 2005; Gerke et al. 2009). Just as with multicellular organisms, the genotype–phenotype relationship in yeast is often due to allelic variation at multiple genes that exhibit pleiotropy, gene–gene interaction, and gene–environment interaction.

How Prevalent Are Hot Spots?

One of the most striking observations across the many data sets generated for the BY×RM cross is the importance of the QTL hot spots. Seven of eight hot spots governing small-molecule resistance overlap with hot spots for gene expression variation, as do three of the four protein abundance hot spots. Additionally, causative alleles in hot-spot regions are known to affect small-molecule resistance, sensitivity to DNA-damaging agents, 16 cell size and shape parameters, sporulation efficiency, and high-temperature growth (Fig. 3). These results strongly suggest that some complex traits in the BY×RM cross share a functional basis due to a set of polymorphisms with widespread effects.

To date, most of the hot spots in the BY×RM cross have been found to be due to derived alleles only present in the BY background. This suggests that the polymorphisms that cause hot spots may be at low frequency in general, potentially because they are deleterious under most conditions (Ronald and Akey 2007). This hypothesis is consistent with comparisons of gene expression variation with sequence divergence in multiple species, which indicate that gene expression variation is subject to stabilizing selection (Lemos et al. 2005; Bedford and Hartl 2009). This trend is also seen in mutation-accumulation experiments, which have shown that random mutations affect gene expression to a greater degree than polymorphisms that reach fixation between species (Rifkin et al. 2005). Together, these studies suggest that mutations with large effects on gene expression are typically culled from populations. Perhaps some hot spots have been maintained because of lineage-specific positive selection that negates or

outweighs any deleterious effects of having a major regulatory change (Ronald and Akey 2007; Lang et al. 2009). This scenario is especially plausible when considering the BY strain, which has been suggested to have undergone elevated rates of evolution due to manipulation in the lab (Gu et al. 2005). Because there have been no studies to date to examine genetic architecture across a large set of traits in any yeast cross besides BY×RM, the true prevalence of hot spots in *S. cerevisiae* remains to be empirically determined.

Summary

Important quantitative genetic principles and molecular mechanisms underlying trait variation have been identified through the genetic dissection of global gene expression and other trait variation in the BY×RM cross. Because complex traits in yeast resemble those of higher organisms, we expect yeast to continue to be a valuable model system for quantitative and evolutionary genetics. However, at present, we have only scratched the surface in terms of using yeast to understand complex traits. In the following sections, we discuss more specifically the enigmas that remain and what resources and methods will be necessary to elucidate them.

LOOKING FORWARD

The BY×RM cross has provided many insights into the genetic architecture of trait variation between two strains. One challenge now is to determine how genetic architectures vary across an entire population (Fig. 4). For example, in a set of parallel crosses between multiple strains, would we observe the same QTLs governing a trait or do different loci arise in each case? Do traits tend to be governed by rare or common polymorphisms? To what extent do the answers to these questions depend on population structure? Understanding how the basis of complex traits varies in a population is central to improving the methods used to dissect the genetic basis of common human diseases.

The yeast population provides a rich resource of variation among strains. *S. cerevisiae* strains have been isolated from a wide variety of habitats such as vineyards, palm wine, rotting fruit, oak trees, sake fermentations, Bertram palms, and immunocompromised humans. An important feature of the yeast population is that the strains from some of these habitats, including vineyard and sake strains, appear to be genetically isolated (Fay and Benavides 2005). Others, such as the clinical isolates, appear to derive from various sources but are likely adapted to a common niche (Schacherer et al. 2009). Recently, comprehensive surveys of sequence variation were performed for many yeast strains (Liti et al. 2009; Schacherer et al. 2009).

The environments in which yeast can be found impose unique stresses and selective pressures, and this variation is reflected in the phenotypic diversity among yeast isolates from different sources. For example, vineyard isolates have likely experienced selection for resistance to the antifungal agent copper sulfate (Fay et al. 2004). Genetic evidence suggests that oak tree strains exhibit high sporulation efficiency and freeze-tolerance relative to other strains due to selective pressures in woodland environments (Gerke et al. 2006; Kvitek et al. 2008). A systematic study across 52 yeast strains demonstrated a high degree of phenotypic variation in both gene expression and response to stresses (Kvitek et al. 2008). *S. cerevisiae* provides an opportunity to link molecular genetic variation, phenotypic variation, natural selection, and population structure across a variety of strains and environments.

Other yeast species offer alternative demographic scenarios not observed in *S. cerevisiae*. *S. paradoxus* is often found in the same soil and oak tree samples as *S. cerevisiae* (Sniegowski et al. 2002; JP Gerke, unpubl.). Unlike *S. cerevisiae*, however, *S. paradoxus* has not been the subject of domestication by humans, and it displays a fundamentally different population structure. Whereas *S. cerevisiae* displays a population structure that correlates primarily with

habitat (Fay and Benavides 2005; Schacherer et al. 2009), genetic differentiation in *S. paradoxus* correlates primarily with geographical distance (Liti et al. 2009). Some isolates of *S. paradoxus* display increased reproductive isolation (Sniegowski et al. 2002). The attributes that make *S. cerevisiae* an ideal model for molecular quantitative genetics—a small genome and efficient homologous recombination—are also present in *S. paradoxus*. Studies of genetic and phenotypic variation among the geographically diverged *S. paradoxus* strains will provide a useful complement to research in *S. cerevisiae*.

APPROACHES TO DISSECTING COMPLEX TRAITS IN YEAST

Yeast represents a powerful system for studying complex traits because it permits the use of all conventional mapping methods and is especially well suited to high-throughput genotyping and phenotyping. High recombination rate, small genome, and the ease of gene replacement enable rapid dissection of QTLs to individual genes. We present a brief summary of the approaches used to map QTLs in yeast, along with their respective advantages and disadvantages.

Methods for Mapping Complex Traits to Genomic Loci

Linkage mapping—As described earlier in the discussion of the BY×RM cross, linkage mapping is a powerful approach for identifying QTLs. Linkage mapping uses related individuals with known pedigrees to identify causal genomic loci (Lynch and Walsh 1998). Within the context of laboratory experiments, two inbred strains are typically crossed to produce segregating progeny. Most commonly, interval mapping (Lander and Botstein 1989), which defines regions of the genome that exhibit linkage to the trait of interest, is used to identify QTLs in a mapping population of segregants. The limitations of linkage mapping are that only a small number of parental genomes (typically two) are compared, and that identified QTLs often span a large number of genes. Indeed, even in yeast, QTLs typically span dozens of genes or more, making it difficult to localize the QTLs to the underlying quantitative trait genes (QTGs) through this approach alone.

Association mapping—Association mapping identifies correlations between genetic markers and traits using panels of individuals sampled from natural populations. Linkage disequilibrium (LD) determines the extent of the associated genomic region. LD decays rapidly in yeast (Schacherer et al. 2009), meaning that this approach should have a very high mapping resolution. Association mapping samples a much larger fraction of the genetic variation in a species than typical linkage mapping experiments. The major limitation of association mapping is its high false-positive rate, which arises from noncausal genotype–phenotype correlations due to population structure. Numerous methods exist to control for population structure while performing statistical tests for association (Price et al. 2006; Yu et al. 2006), but these structured association mapping techniques also cause a high rate of false negatives, often eliminating the signal of association from true positive loci that are strongly correlated with population structure (Zhao et al. 2007).

Hybrid linkage/association approaches—The respective advantages of linkage and association mapping can be combined by creating multiparent inbred line populations (Churchill et al. [The Complex Traits Consortium] 2004; Macdonald and Long 2007; Kover et al. 2009) or multiple two-parent mapping populations that share a common parent (Yu et al. 2008). In such populations, linkage mapping can be used to identify QTLs, and association mapping can be used to refine the QTLs to smaller genomic intervals. In yeast, which has an innately high recombination rate, it may be that multiparent mapping populations will allow causal loci to be mapped to a high resolution, perhaps to specific genes.

Bulk segregant analysis—A final, underused approach to mapping QTLs is bulk segregant analysis (BSA). In BSA, segregants with extreme phenotypes are genotyped collectively, and loci with significantly biased allele frequencies represent candidate QTLs (Michelmore et al. 1991). This approach has been previously used in yeast to map an auxotrophy, a growth defect on acetate, a locus involved in flocculation (Brauer et al. 2006), loci involved in adaptation to fluctuating carbon sources (Segre et al. 2006), resistance to DNA-damaging agents (Demogines et al. 2008), and a locus involved in resistance to leucine starvation (Boer et al. 2008). A major limitation in this approach is that it requires very precise, quantitative genotyping for a large number of markers. Previously, this had been a challenge; however, with easy custom array design and next-generation sequencing, there are now multiple options to measure allele frequencies in DNA pools. The advantage of this approach is that unlike other approaches, it permits a much larger number of individuals to be surveyed, which means that it is likely to be more powerful in detecting the multiple loci with modest effects that underlie complex traits. Despite its promise, BSA has yet to be used effectively to map multiple loci underlying a complex trait.

Methods for refining QTLs to Genes or Causal Polymorphisms

Reciprocal hemizygosity—In this approach, two nearly identical strains are compared. Both strains are diploids descended from the same two parental strains and are hemizygous for a putative QTG. The only difference between the strains is which parent's allele of the putative QTG they possess. Comparison of these two hemizygous strains makes it possible to test the effects of the two alleles in genetic backgrounds that are otherwise isogenic. Deletion collections exist in yeast in which there is a DNA bar-coded single-gene knockout strain for nearly every nonessential gene in the genome (Winzeler et al. 1999). These deletion collections allow reciprocal hemizygosity tests to be conducted on a genome-wide scale, either by screening with individual knockout strains en masse or by competing all knockout strains in heterozygous diploid pools (Steinmetz and Davis 2004). One drawback of this approach is that it does not provide nucleotide-level resolution.

Allele swaps—Perhaps the greatest advantage of yeast is the ease with which targeted gene replacement can be done. Homologous recombination can be used to create lines that differ at a specific site but are otherwise isogenic. If necessary, allele swap lines can be generated for all genes in a QTL and used to delineate the causal gene(s) underlying the QTL. Various allele replacement lines can then be intercrossed to directly measure the effects of allelic interactions. A drawback of this approach is that if genetic interactions exist, allele replacements may need to be done in a specific genetic background.

Introgression and targeted recombination—An additional approach to isolating QTLs is by recurrent crossing into a different background, using either phenotypic or marker-assisted selection. This approach can increase the power to identify small-effect QTLs by eliminating other loci with effects and by isolating a QTL allele from one parent in a genetic background that is largely composed of the other parent. This iterative crossing approach can be used to break down the amount of flanking parent-of-origin sequence, whittling down a QTL to a smaller genomic region. Either independently or in concert with introgression, selectable flanking markers can be used to identify segregants with recombination events specifically in a QTL. These targeted recombinants can be used to rapidly fine map a QTL.

NEEDED RESOURCES FOR THE FUTURE

Attaining a more comprehensive understanding of complex traits will require identifying a large fraction of the polymorphisms that encode trait differences within and between yeast

species. Accomplishing this will require new mapping populations and high-throughput genotyping and phenotyping approaches.

A prerequisite to creating new mapping populations is the identification of a large number of genome-wide markers. Obtaining such markers is now simple: It can be done by resequencing 10 or so strains in a single lane of an Illumina Genome Analyzer 2 run. In addition, an already published population genomics project has produced low-coverage genome sequences for a number of strains for both *S. cerevisiae* and *S. paradoxus* (Liti et al. 2009). Existing genome-wide polymorphism data provide a foundation for understanding the genetic diversity and population structure of these species (Liti et al. 2009; Schacherer et al. 2009). Additionally, next-generation sequencing technologies can be used for rapid targeted resequencing of strains with phenotypes of interest or genotyping of new mapping populations.

A central consideration going forward is whether yeast needs an infrastructure to promote community collaboration in the dissection of complex traits, as exists for *Arabidopsis thaliana* (Weigel and Mott 2009), *Drosophila* (Mackay et al. 2008), and mouse (Churchill et al. [The Complex Traits Consortium] 2004). In a sense, the simplicity of yeast diminishes the need for collaboration. At the same time, a concerted community effort to dissect the genetic basis of many traits to as many causal polymorphisms as possible in as many backgrounds as is feasible could result in a basic understanding of complex traits that may be impossible to achieve in any other organism.

The cloning of individual QTGs has been accomplished in a wide range of organisms, and the challenge now is to advance to a new level of understanding of complex traits. To achieve this, we must identify all QTGs for many genetically complex traits across a diverse set of strains and determine the underlying variants at the nucleotide level. Such a comprehensive undertaking is essential if we are to understand mechanistically the genetic architecture and evolutionary implications of complex traits, and if we are to harness natural variation for applied purposes.

Acknowledgments

The authors thank Rachel Brem, Gael Yvert, Erin Smith, Eric Foss, James Ronald, Josh Akey, Ethan Perlstein, John Storey, Jackie Whittle, and other Kruglyak lab members and collaborators for their contributions to the research based on the BY×RM cross. This work was supported by National Institutes of Health grant MH059520 and a James S. McDonnell Foundation Centennial Fellowship to L.K., National Institutes of Health grant GM071508 to the Lewis-Sigler Institute, and the Howard Hughes Medical Institute.

REFERENCES

- Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrman S, Francois J, Yvert G. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet* 2008;4:e1000049. [PubMed: 18404214]
- Bedford T, Hartl DL. Optimization of gene expression by natural selection. *Proc Natl Acad Sci* 2009;106:1133–1138. [PubMed: 19139403]
- Boer VM, Amini S, Botstein D. Influence of genotype and nutrition on survival and metabolism of starving yeast. *Proc Natl Acad Sci* 2008;105:6930–6935. [PubMed: 18456835]
- Brauer MJ, Christianson CM, Pai DA, Dunham MJ. Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* 2006;173:1813–1816. [PubMed: 16624899]
- Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci* 2005;102:1572–1577. [PubMed: 15659551]
- Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002;296:752–755. [PubMed: 11923494]

- Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 2005;436:701–703. [PubMed: 16079846]
- Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 2004;36:1133–1137. (The Complex Traits Consortium). [PubMed: 15514660]
- Demogines A, Smith E, Kruglyak L, Alani E. Identification and dissection of a complex DNA repair sensitivity phenotype in baker's yeast. *PLoS Genet* 2008;4:e1000123. [PubMed: 18617998]
- Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet* 2005;37:1333–1340. [PubMed: 16273108]
- Falconer, D.; Mackay, T. Introduction to quantitative genetics. Pearson Prentice Hall; Harlow, England: 1996.
- Fay JC, Benavides JA. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet* 2005;1:66–71. [PubMed: 16103919]
- Fay JC, McCullough HL, Sniegowski PD, Eisen MB. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol* 2004;5:R26. [PubMed: 15059259]
- Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, Kruglyak L. Genetic basis of proteome variation in yeast. *Nat Genet* 2007;39:1369–1375. [PubMed: 17952072]
- Gatbonton T, Imbesi M, Nelson M, Akey JM, Ruderfer DM, Kruglyak L, Simon JA, Bedalov A. Telomere length as a quantitative trait: Genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS Genet* 2006;2:e35. [PubMed: 16552446]
- Gerke JP, Chen CT, Cohen BA. Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* 2006;174:985–997. [PubMed: 16951083]
- Gerke J, Lorenz K, Cohen B. Genetic interactions between transcription factors cause natural variation in yeast. *Science* 2009;323:498–501. [PubMed: 19164747]
- Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* 2005;102:1092–1097. [PubMed: 15647350]
- Hu XH, Wang MH, Tan T, Li JR, Yang H, Leach L, Zhang RM, Luo ZW. Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. *Genetics* 2007;175:1479–1487. [PubMed: 17194785]
- Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 2009;5:e1000551. [PubMed: 19593375]
- Kvitek DJ, Will JL, Gasch AP. Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLoS Genet* 2008;4:e1000223. [PubMed: 18927628]
- Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989;121:185–199. [PubMed: 2563713]
- Lang GI, Murray AW, Botstein D. The cost of gene expression underlies a fitness trade-off in yeast. *Proc Natl Acad Sci* 2009;106:5755–5760. [PubMed: 19299502]
- Lemos B, Meiklejohn CD, Caceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution* 2005;59:126–137. [PubMed: 15792233]
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. Population genomics of domestic and wild yeasts. *Nature* 2009;458:337–341. [PubMed: 19212322]
- Litvin O, Causton HC, Chen BJ, Pe'er D. Modularity and interactions in the genetics of gene expression. *Proc Natl Acad Sci* 2009;106:6441–6446. [PubMed: 19223586]
- Lynch, M.; Walsh, B. Genetics and analysis of quantitative traits. Sinauer; Sunderland, MA: 1998.
- Macdonald SJ, Long AD. Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* 2007;176:1261–1281. [PubMed: 17435224]

- Mackay TF, Richards S, Gibbs R. Proposal to sequence a *Drosophila* genetic reference panel: A community resource for the study of genotypic and phenotypic variation. White Paper. 2008
- Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci* 1991;88:9828–9832. [PubMed: 1682921]
- Nogami S, Ohya Y, Yvert G. Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS Genet* 2007;3:e31. [PubMed: 17319748]
- Orr HA. The genetic theory of adaptation: A brief history. *Nat Rev Genet* 2005;6:119–127. [PubMed: 15716908]
- Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nat Genet* 2007;39:496–502. [PubMed: 17334364]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909. [PubMed: 16862161]
- Rifkin SA, Houle D, Kim J, White KP. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 2005;438:220–223. [PubMed: 16281035]
- Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet* 2006;7:862–872. [PubMed: 17047685]
- Ronald J, Akey JM. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. *PLoS ONE* 2007;2:e678. [PubMed: 17668057]
- Ronald J, Brem RB, Whittle J, Kruglyak L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet* 2005;1:e25. [PubMed: 16121257]
- Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L. Population genomic analysis of outcrossing and recombination in yeast. *Nat Genet* 2006;38:1077–1081. [PubMed: 16892060]
- Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 2009;458:342–345. [PubMed: 19212320]
- Segre AV, Murray AW, Leu JY. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol* 2006;4:e256. [PubMed: 16856782]
- Sinha H, David L, Pascon RC, Clauser-Munster S, Krishnakumar S, Nguyen M, Shi G, Dean J, Davis RW, Oefner PJ, et al. Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* 2008;180:1661–1670. [PubMed: 18780730]
- Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression. *PLoS Biol* 2008;6:e83. [PubMed: 18416601]
- Sniegowski PD, Dombrowski PG, Fingerman E. *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* 2002;1:299–306. [PubMed: 12702333]
- Steinmetz LM, Davis RW. Maximizing the potential of functional genomics. *Nat Rev Genet* 2004;5:190–201. [PubMed: 14970821]
- Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 2002;416:326–330. [PubMed: 11907579]
- Storey JD, Akey JM, Kruglyak L. Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol* 2005;3:e267. [PubMed: 16035920]
- Storici F, Lewis LK, Resnick MA. In vivo site-directed mutagenesis using oligonucleotides. *Nat Biotechnol* 2001;19:773–776. [PubMed: 11479573]
- Weigel D, Mott R. The 1001 Genomes project for *Arabidopsis thaliana*. *Genome Biol* 2009;10:107. [PubMed: 19519932]
- Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999;285:901–906. [PubMed: 10436161]

- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 2006;38:203–208. [PubMed: 16380716]
- Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. *Genetics* 2008;178:539–551. [PubMed: 18202393]
- Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 2003;35:57–64. [PubMed: 12897782]
- Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M. An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet* 2007;3:e4. [PubMed: 17238287]

Figure 1.

Crossing scheme used to create the BY×RM mapping population. Haploid parental strains are crossed to form a heterozygous diploid (“hybrid”) that is then sporulated to produce recombinant haploid offspring. More than 100 segregants were sampled from the BY×RM cross.

Figure 2.

Primer on the genetics of global gene expression. (A) Example of a transcript where BY and RM differ, and a single locus has a major effect on the expression level of the segregants. (B) Example of a transcript showing transgressive segregation, where BY and RM have comparable expression levels but the segregants exhibit much more variation than is seen in the parents. (C) Plot of local (blue) and distant (red) linkages across the entire genome. A number of locations in the genome show linkage to many distant transcripts (vertical bands). These loci are referred to as hot spots. (A–C are illustrations and do not represent actual data.)

B

Gene	Cellular function	Hot-spot types	Organismal phenotypes affected
<i>AMN1</i>	Daughter cell separation	Gene expression, Protein levels	Cell clumping
<i>LEU2</i>	Leucine biosynthesis	Gene expression	Ability to make leucine
<i>MAT</i>	Mating-type determination	Gene expression	Mating type
<i>GPA1</i>	Mating pathway signaling	Gene expression	Cell morphology, mating efficiency
<i>HAP1</i>	Oxygen response	Gene expression	Growth rate in laboratory environments
<i>PHO84</i>	Phosphate transport	Gene expression, protein levels, small-molecule resistance	Resistance to tetrachloroisophthalonitrile and pentachlorophenol
<i>MKT1</i>	Posttranscriptional regulation	Gene expression, small-molecule resistance	High-temperature growth, sensitivity to 4-nitroquinoline, sporulation efficiency
<i>IRA2</i>	Regulator of RAS pathway	Gene expression	Signaling levels through RAS/PKA pathway

Figure 3. Compendium of the hot spots observed in the BY×RM cross. The locations (A) and descriptions (B) of a subset of the hot spots in the genome. This plot is a modified version of a plot in Smith and Kruglyak (2008). Causal genes for each of these hot spots are named, although in some cases it is unclear whether coincident hot spots across traits are due to different closely linked genes or to the pleiotropic effects of a single gene. For instance, the *AMN1* gene expression hot spot is likely due to multiple genes with effects on different transcripts (Yvert et al. 2003). Red horizontal line in A indicates the threshold for calling a locus a hot spot. References for B are provided in the text.

Figure 4.

Genetic architecture across a population. (A) QTLs causing variation in a trait segregate with varying allele frequencies across 15 strains in a population. The upward red arrows and downward blue arrows are opposite effect alleles that cause relative increases or decreases in the trait value of an individual. (B) As a result of the large number of loci affecting the trait in the population, in an intercross, both the number and locations of QTLs will depend on the two parental strains chosen for the cross. In the three inter-crosses chosen here, QTLs “C” and “I” appear to be common, but identifying the causal polymorphisms in the other 12 strains would reveal that the minor alleles are actually rare.