The molecular basis of phenotypic variation in yeast
Justin C Fay

The power of yeast genetics has now been extensively applied to phenotypic variation among strains of *Saccharomyces cerevisiae*. As a result, over 100 genes and numerous sequence variants have been identified, providing us with a general characterization of mutations underlying quantitative trait variation. Most quantitative trait alleles exert considerable phenotypic effects and alter conserved amino acid positions within protein coding sequences. When examined, quantitative trait alleles influence the expression of numerous genes, most of which are unrelated to an allele’s phenotypic effect. The profile of quantitative trait alleles has proven useful to reverse quantitative genetics approaches and supports the use of systems genetics approaches to synthesize the molecular basis of trait variation across multiple strains.

Addresses
Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University, St. Louis, MO, United States

Corresponding author: Fay, Justin C (jfay@genetics.wustl.edu)

Introduction
Most variable phenotypes have a complex genetic basis. While it has become relatively easy to coarsely map the genetic basis of such traits, identifying causative genes and genetic changes within those genes remains challenging. As such, the power of yeast genetics has made *Saccharomyces cerevisiae* an attractive model for dissecting complex traits and revealing the molecular bases of a number of traits in exquisite detail [1].

Perhaps the greatest challenge to quantitative trait mapping is being able to map any type of change, for example, single nucleotide polymorphism (SNP), insertion/deletion polymorphism (InDel), or change in chromosome structure, and to determine how it affects a trait, for example, changes in protein structure or gene expression. In many organisms, meeting this challenge is difficult due to technological limitations. For example, noncoding changes are more difficult to identify and their effects may not always be recapitulated outside of their native context. *S. cerevisiae* has proven adept at meeting these challenges and revealing the types variants in a population that contribute to quantitative trait variation.

Currently, over 100 quantitative trait genes (QTGs) and half as many quantitative trait nucleotides (QTNs) have been identified in yeast. Here, I review insights that have emerged from the consensus of these studies and highlight remaining challenges that need to be addressed. While the QTN program has been justly criticized as not being reflective of evolutionary change [2], the high resolution genetic analysis conducted in yeast is relevant to our understanding of how trait variation is generated and maintained in a population.

The case for QTN: linked QTGs and multiple QTNs
Genomics has made quantitative trait locus (QTL) mapping a tenable means of identifying the genetic basis of phenotypic variation. As a result, the number of QTL and their effect sizes have been documented for many traits. However, there has always been a concern that single QTL of large effect might be caused by multiple linked QTGs of smaller effect [3]. The first example of such in yeast involved a major effect high temperature growth QTL that was elegantly shown to be caused by three out of 15 genes in a 32 kb region [4]. The genes were identified using reciprocal hemizygosity analysis (Figure 1), which has since become the standard for efficient and robust identification of QTGs in yeast. More recent studies have documented other examples of linked QTGs [5–7], which intriguingly also occur in the same region on chromosome XIV but involve different genes. Currently, this 75 kb region harbors 10 known QTGs linked to a variety of traits and represents a hotspot of quantitative trait variation (Figure 2). In a similar vein, multiple QTNs have been shown to occur within a single QTG. Such genes include *FLO11* [8], *HO* [9], *IME1* [10], *MLH1* [11], *PCAI* [12], *PMS1* [11], *RADS* [13], *AQY1* and *AQY2* [14], highlighting the importance of carefully dissecting causal variants within a QTG. As summarized in Figure 3, out of 110 quantitative trait alleles that have been identified in yeast, half (54) have been delineated to specific nucleotide changes.

Types of changes
Mirroring the relative abundance of different types of DNA polymorphism [15], most mapped variants are single nucleotide changes (Figure 2 and Figure 3). However, chromosome rearrangements, copy number variants (CNVs) and InDels have also been identified.
potentially more often than what one might expect based on the genome abundance of these types of polymorphisms relative to SNPs [15]. In the case of the tandemly duplicated EnA locus, sodium/lithium tolerance is associated with a recent introgression of an S. paradoxus allele into some but not all strains of S. cerevisiae [15,16], a novel source of variation that is governed by reproductive barriers rather than mutation rate.

Coding versus noncoding changes

Similar to other organisms [17], the majority (69%) of QTNs lie within protein coding sequences (Figure 3). However, contrary to other organisms there is arguably little bias towards successfully mapping coding relative to noncoding QTN in yeast. Nevertheless, the relative abundance of protein coding changes is nearly identical to the 72% of the yeast genome that encodes proteins [18], which is substantially higher than most plant and animal genomes.

Small versus large effects

The vast majority of alleles that have been identified generate moderate to large phenotypic effects. In the context of QTL mapping, there is undoubtedly a bias towards identifying alleles of large effect; they are the first to be pursued and also the easiest to resolve to single genes or genetic changes. Even so, alleles of moderate to large effect can explain most variation in a cross. In two studies of sporulation efficiency, 88% of variation in a cross [10] and 92% of the parental difference [19] was explained by QTN in three genes, only one of which, RME1, was shared between the two studies. More generally, a modest number of QTL was found to explain an average of 88% of additive genetic variation across 46 traits [20]. Nevertheless, quite a few of the QTL identified in this latter study can be considered small effect loci as they would not have been found with even a moderate number of 100 segregants. Thus, large effect alleles, while commonly found, don’t preclude the existence of numerous alleles of small effect.

Recent studies have successfully targeted and identified alleles of small effect. In these studies, small effect alleles were either linked to large effect alleles or masked by interactions with them such that they were only identified by first fixing the alleles of large effect, either through backcrossing [21,22] or allele replacement [23,24]. The observation that QTL were found at or nearby three previously discovered large effect QTGs adds further evidence of multiple linked QTGs or QTNs [23]. However, targeting alleles of small effect is not easy; their effects are more difficult to distinguish from subtle differences in genetic background that can arise between nearly identical strains. For example, a non-complementation screen using the yeast deletion collection faithfully
recovered alleles of large effect but failed to identify alleles of small effect from a genome-wide screen [25].

**Common versus rare variation**

Resolving quantitative trait alleles to individual nucleotides makes it possible to assess their frequency in the population. The majority of QTN that have been identified are only found in one or a small number of strains (Figure 3). Of relevance, a fair number of mapping populations were derived from crosses using laboratory strains and identify alleles that are specific to laboratory strains. Even excluding such alleles, more rare or low frequency alleles are found than common alleles.

In addition to an abundance of rare QTN, common QTN can behave as rare alleles due to interactions with genetic background. In one case, the effect of an MKT1 allele was shown to be absent when measured in certain backgrounds [26]. In another case, a hypomorphic allele of CYS4 affecting colony color was entirely suppressed in some strain backgrounds [27]. The strong influence of genetic background is also evident from epistatic interactions between QTL within a single cross and from pairwise or higher-order interactions in combined cross analyses, where QTL are often found to be context-dependent [22,28,29].

**The molecular mechanisms and consequences of QTGs**

The identification of even a single QTG can provide insight into which biochemical pathways and cellular processes give rise to variation in a trait. While each trait is likely different, a number of examples in yeast provide insight into how DNA sequence changes create variation at the molecular level which is then translated into phenotypic variation.

In some cases, the mechanism by which a QTG influences a trait is already well understood, for example, transporters (PCAI, PHO84, AQY2, MPR1, CTR3, TRK1) and growth in the presence of compounds they transport (Table S1). In other cases the mechanism is not so obvious. Perhaps one of the most intriguing examples is that of MKT1 (Maintenance of Killer Toxin), required for propagation of the M2 dsRNA L-A virus [30]. A laboratory strain allele of MKT1 has been linked to high temperature growth [4], sporulation [19], DNA damage [13], petite formation [6], ethanol tolerance [7] and drug resistance [28]. The extensive pleiotropy of MKT1 is also present at the molecular level as shown by its influence on the expression of many genes [31]. Similar results have been found for other QTGs; alleles of AMV1 [32], IRA2 [31], CYS4 [27], and SRY1 [33] all affect the expression of dozens to hundreds of genes (Table 1).
The extensive changes in gene expression caused by QTG alleles raises the question of how many of these changes are relevant to the production of the phenotype. In the case of an amino acid change in CYSC that causes rust-colored colonies and differential expression of 75 genes, rust coloration was only affected by deletion of eight out of the 75 genes. Similar to SSX1, most of the expression differences caused by the QTG likely represent a downstream response to the phenotypic effects of the QTL: cysteine deficiency for CYSC [27] and a defect in amino acid sensing for SSX1 [33]. In another study involving copy number changes of CUP2, deletion of three out of 18 genes influenced by CUP2 dosage altered copper resistance [34]. One intriguing observation is that differentially expressed mediators of QTL phenotypes are much better predicted when their expression differences persist across multiple environments compared to those that are environment-specific [35]. One implication of these results is that gene expression can be used as a molecular signature of a QTL’s effect, in some cases without the presence of the phenotype.

Reverse quantitative genetics
There is growing potential for the use of reverse genetics approaches to understanding quantitative trait variation. Genome sequences are available for multiple strains of yeast and a number of approaches have been developed to predict functional polymorphism [15]. While difficult, predicting a phenotype based on genotype and other prior information is one of the ultimate goals of quantitative genetic analysis. Certainly this has been accomplished for segregants of a cross, but it is much more difficult when applied to previously unseen genotypes, even when the search is limited to a small genomic interval and a specific phenotype as is the case for refining a QTL to a QTG.

Cross-species conservation provides one means of predicting which SNPs within coding and noncoding sequences are likely to affect fitness. QTN within coding sequences often disrupt highly conserved amino acid positions and are predicted to be deleterious [15]. Predicting changes in gene expression based on promoter variants is more challenging. The combination of cross-species conservation, that is, phylogenetic footprinting, and transcription factor binding site models has been used to predict cis-regulatory SNPs and differentially expressed genes [15,36–39]. However, the sensitivity and specificity of predicted variants in coding and noncoding regions is informative but not perfect, similar to results from other organisms [40]. Indeed, multiple amino acid polymorphisms predicted to affect ZRT1 function showed no effect on growth in zinc limiting conditions [41]. The ZRT1 result illustrates a drawback to the reverse genetics approach: it is hard to know whether the lack of a phenotype is due to a small effect size, measuring the right phenotype, or a false positive prediction.

Predicting phenotype from genotype is greatly facilitated by S. cerevisiae’s well characterized deletion collection. While using the deletion collection has drawbacks, for example strain dependent effects of a deletion [42], the approach has now been successfully applied to genome based predictions [43]. In this study, a strain’s growth phenotype was predicted based on deleterious coding variants throughout its genome combined with corresponding fitness assays from the yeast deletion collection. Certain phenotypes, such as growth on galactose and glycerol, were generally well predicted by this approach.

Conclusions
Quantitative genetic analysis in yeast has yielded a general description of QTGs and QTN within those genes. Together, these studies have shown that:

i) there are multiple cases of linked QTGs and multiple QTN within a QTG,
ii) an appreciable fraction of quantitative trait alleles (34%) are not caused by SNPs,
iii) most alleles are caused by changes in conserved protein coding sequences,
iv) QTGs cause numerous downstream changes in gene expression, most of which are not required for the phenotype.

Despite these efforts, rare alleles and epistasis make it difficult to completely dissect a quantitative trait such that phenotypes can be reliably predicted for any genotype in the population. The combination of a quantitative trait allele’s expression signature along with advances in predicting functional variants has the potential to help resolve the molecular basis of trait variation present in a population rather than just segregants from a cross.

Acknowledgement
This work was supported in part by a National Institutes of Health grant (GM080669) to JCF.

Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Phenytype</th>
<th>Differentially expressed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMN1</td>
<td>Flocculation</td>
<td>12</td>
</tr>
<tr>
<td>CYS4</td>
<td>Colony color and drug resistance</td>
<td>75</td>
</tr>
<tr>
<td>SSY1</td>
<td>Resistance to a leucine analog</td>
<td>103</td>
</tr>
<tr>
<td>IRA2</td>
<td>High temperature growth</td>
<td>372</td>
</tr>
<tr>
<td>MKT1</td>
<td>Many</td>
<td>1159</td>
</tr>
</tbody>
</table>
Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gd.2013.10.005.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


28. While many changes in gene expression are found linked to quantitative trait alleles, this was the first study to show that most of these expression changes are not required for the production of the altered phenotype.


This study accomplishes this goal and also provides insight into the evolutionary forces acting on this rearrangement.


Knowing how each QTG affects a phenotype is key to developing a comprehensive understanding of the molecular basis of trait variation. This study shows that gene expression changes likely to mediate a QTG’s phenotype persist across multiple environments whereas those that are environment-specific are less likely to mediate a QTG’s phenotype.


41. Engle EK, Fay JC: ZRT1 harbors an excess of nonsynonymous polymorphism and shows evidence of balancing selection in *Saccharomyces cerevisiae*. G3 (Bethesda) 2013 http://dx.doi.org/10.1534/g3.112.005082.


The ability to predict phenotype from genotype is now one of the major barriers personalized medicine. This study demonstrates a remarkable ability to predict a variety of growth phenotypes using genome sequence variants predicted to be deleterious and the phenotypes of those genes carrying the sequence variants when deleted.