Yeast: An Experimental Organism for 21st Century Biology

David Botstein*,[1] and Gerald R. Fink[†]
*Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, and [†]Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ABSTRACT In this essay, we revisit the status of yeast as a model system for biology. We first summarize important contributions of yeast to eukaryotic biology that we anticipated in 1988 in our first article on the subject. We then describe transformative developments that we did not anticipate, most of which followed the publication of the complete genomic sequence of *Saccharomyces cerevisiae* in 1996. In the intervening 23 years it appears to us that yeast has graduated from a position as the premier model for eukaryotic cell biology to become the pioneer organism that has facilitated the establishment of the entirely new fields of study called "functional genomics" and "systems biology." These new fields look beyond the functions of individual genes and proteins, focusing on how these interact and work together to determine the properties of living cells and organisms.

TABLE OF CONTENTS

doi: 10.1534/genetics.111.130765
[1]Corresponding author: Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544. E-mail: botstein@genomics.princeton.edu

TWENTY-THREE years ago, in an article in *Science* magazine, we speculated that yeast might be the ideal experimental organism for modern biology. We argued that the amalgam of recombinant DNA technology and classical biochemistry and genetics had created a revolution that gave biologists access to an array of new methods for connecting proteins and genes with their roles in the biology of an organism. We wrote that the reason that yeast could serve "as a model for all eukaryotic biology derives from the facility with which the relation between gene structure and protein function can be established" (Botstein and Fink 1988, p. 1440).

In this essay, we revisit the status of yeast as a model experimental system. We begin by providing a summary of the important contributions of yeast to the knowledge of eukaryotic biology that we anticipated in 1988. We then describe transformative developments that we did not anticipate, most of which followed the publication of the complete genome sequence of *Saccharomyces cerevisiae* in April 1996. We have made no effort to provide a formal review of the literature. The articles that we cite are intended as illustrative examples only.

In the 23 years since our last essay it appears to us that yeast has graduated from a position as the premier model for eukaryotic cell biology to become the pioneer organism that facilitated the establishment of entirely new fields of study called "functional genomics" and "systems biology." These new fields look beyond the functions of individual genes and proteins, focusing on how they interact and work together to determine the properties of living cells and organisms.

Functional Genomics: Gene–Protein–Function Association via Mutants

Probably the most important and enduring contribution of the model yeasts (*S. cerevisiae* and *Schizosaccharomyces pombe*) and the scientific communities that study the biology of these organisms has been the connection of genes and proteins with the functions that they provide to cells. As we indicated in 1988, the methods for introducing mutations, at will, into and out of the yeast genome, have made it particularly easy to study not only the biochemical function of gene products, but also the biological consequences of failure of the genes to function. Mutations were produced and introduced into yeast strains by each researcher as needed; this soon came to be seen as rate limiting.

Not long after the publication of the yeast genome sequence, the *Saccharomyces* community organized a cooperative effort that produced a nearly complete set of deletions of every open reading frame (*cf*. Winzeler *et al.* 1999 and Giaever *et al.* 2002). Each gene was replaced by a drug-resistance gene and marked with synthetic "barcode" sequences. These features made each deletion selectable, facilitating transfer by DNA transformation, and made each deletion distinguishable from all the others so that individual mutants can be followed in screens of the entire library of deletion mutants. Mutations in each yeast gene and many ensembles of mutations have been subjected to diverse biological assays, often leading to increased understanding of the biological roles of many of the genes. The library of deletion mutants and its derivatives over the years has been exploited with great effect in genome-scale experiments; many of the successful methods that we summarize are based on deletion libraries (see Scherens and Goffeau 2004 for a review). A similar deletion library for fission yeast (*S. pombe*) has recently become available (Kim *et al.* 2010).

Other comprehensive mutant libraries have been constructed and used to characterize yeast gene functions. The introduction of green fluorescent protein (GFP) fusion technology to visualize protein localization and interactions was used to provide localization information for most proteins of yeast (Ghaemmaghami *et al.* 2003; Huh *et al.* 2003). Libraries of fusions to other sequence tags have been constructed to facilitate immunoprecipitation and detection of protein interactions. Characteristically, the yeast community made these strains generally available, facilitating the work of all.

Today the Saccharomyces Genome Database (SGD; http://www.yeastgenome.org/) provides information about every yeast gene based not only on the literature, but also on the systematic study of every *Saccharomyces* gene about which anything has been learned.

Since 1996 the fraction of the nearly 5800 protein-coding *Saccharomyces* genes for which a rudimentary understanding of their biological role is known has risen from ~30% to ~85%. This fraction is much higher for this yeast than for any other eukaryote. Nearly 1000 yeast genes (*i.e.*, ~17%) are members of orthologous gene families associated with human disease (Heinicke *et al.* 2007). For the majority of these genes their mammalian homolog is functional in yeast and complements the yeast deletion mutant. In 1988, although we could point to only a handful of cases of interspecies functional complementation, they have long since become routine (see Dolinski and Botstein 2007 for a review).

Databases and Gene Ontology

Comparison of the yeast genomic sequences with those of other model systems, including the human, led quickly to the realization that both protein amino acid sequences and protein functions have been conserved well enough that annotations of function should frequently, if not always, be transferable from one eukaryotic species to another. Since functional information must ultimately be obtained by experiment, this realization emphasized the advantages of yeast as a model experimental system. Virtually all cellular-level technologies for assessing protein function have remained easier with yeast than with most other organisms. This is especially so for the high-throughput technologies that rapidly developed (see below).

Transfer between species of functional information about homologous proteins via sequence homology required better ways to organize and systematize not only the genome sequences, but also the functional information about genes linked to their amino acid sequences. The Gene Ontology Consortium (GO) (Ashburner *et al.* 2000; http://www.geneontology.org/) was formed to meet this need. The GO collaborators developed, and continue to maintain, three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner. The goal was to be able to describe what proteins do in ways that are biologically realistic yet compatible with computational methods. This goal has largely been met. GO annotations are a major feature of all the genomic databases and publications in genomics and are also used in many varieties of computational methods that group genes together according to their biological roles, as we describe below.

One of the early insights provided by GO was the realization of the need to improve on the traditional, but very loose rubric "function," when thinking about what a gene does for an organism. The GO collaborators decided on three roughly orthogonal categories: "biological process" (*e.g.*, DNA replication), "molecular function" (*e.g.*, DNA-dependent DNA polymerase), and "cellular component" (*e.g.*, nucleus). To illustrate, knowing that a gene is a protein kinase is relatively uninformative for biology until one sees that a deletion of this gene results in a DNA replication defect in yeast and that the protein itself is found in the nucleus. A researcher who comes upon the worm or human ortholog of this kinase will, after having consulted GO, have obtained a very promising hypothesis about what this protein might do for the worm or the human.

It should be no surprise that most of the GO annotations, especially those involving biochemistry and cell biology, derive from experiments with yeast. Increasingly, annotations derive from experiments done directly in other model organisms, which is necessary for biological phenomena not found in yeast. It is no longer uncommon to find SGD annotations for genes and proteins that derive from experiments done on another species that are transferred to yeast. Nevertheless, yeast as a model organism remains a disproportionate contributor to the shared knowledge of eukaryotic biology embodied and organized in the gene ontology (for more on this, see Dolinski and Botstein 2005, 2007).

Gene Expression and Regulatory Networks

The advent of the genomic sequences spurred considerable technology development aimed at comprehensive study of all the genes in the genome simultaneously. The earliest and simplest of these were DNA microarrays containing sequences from every open reading frame, which were used to detect labeled copies of mRNAs (see Brown and Botstein 1999 for an early review). Although this technology could be applied to any organism for which gene sequences were known, it was with yeast that this and other comprehensive functional genomic technologies were pioneered and validated. Even more important than the experimental tractability of yeast was the extensive and reliable knowledge of yeast gene functions already available by 1996 from the literature. When DeRisi *et al.* (1997) published their landmark study of gene expression during exponential growth and the diauxic shift, they could easily validate their new methods because the behavior of many genes under this growth regime had already been well studied.

Shortly thereafter, computational methods capable of analyzing and visualizing the ~100,000 individual measurements that compose a data set of this kind began to be introduced (Eisen *et al.* 1998; Alter *et al.* 2000). Here again, the well-documented prior knowledge of the behavior of a relatively few genes validated the more general results that could be extracted by the new analytic methods from, for example, studies of the cell division cycle (Spellman *et al.* 1998), sporulation (Chu *et al.* 1998), and responses to various stresses (Gasch *et al.* 2000). Some of these data sets have continued to serve the burgeoning computational biology, systems biology, and bioinformatics communities, providing test beds for an array of increasingly sophisticated analytical methods (*cf.* Botstein 2010).

For yeast biologists, the main results of the early gene expression studies was the discovery of a vast array of interlocking transcriptional regulatory networks and the transcription factors that control them. Genome-wide expression experiments on yeast validated the wide application of the technology. Gene expression technology became the basis for characterizing the functional genomics of everything from fly development to the definition of tumor subtypes. These data form the basis of a general worldview about system-level integration of transcriptional controls in eukaryotic cells that has awakened the interest of many physical and computational scientists.

The success of gene expression technologies naturally led to the development of a great variety of other genome-scale technologies. Virtually every one of these was developed and validated first with yeast—once again because of the well-documented prior knowledge for the functions of at least a few key genes—and then diffused into more general use. Notable among these methods were those for mapping the binding sites of transcription factors *in vivo* by chromatin precipitation followed by DNA microarray (ChIP-chip) (Iyer *et al.* 2001; Lieb *et al.* 2001) or by direct sequencing of the bound DNA (Chip-seq) (Robertson *et al.* 2007). Like the simpler gene expression methods, these methods could be used to follow natural processes dynamically and observe, for example, the periodic binding of factors involved in cell cycle progression. Also notable are methods that allow the genome-wide assessment of translation rates (Ingolia *et al.* 2009) or mRNA stability (Wang *et al.* 2002), each of which has the property of allowing dynamic assessment of regulatory changes.

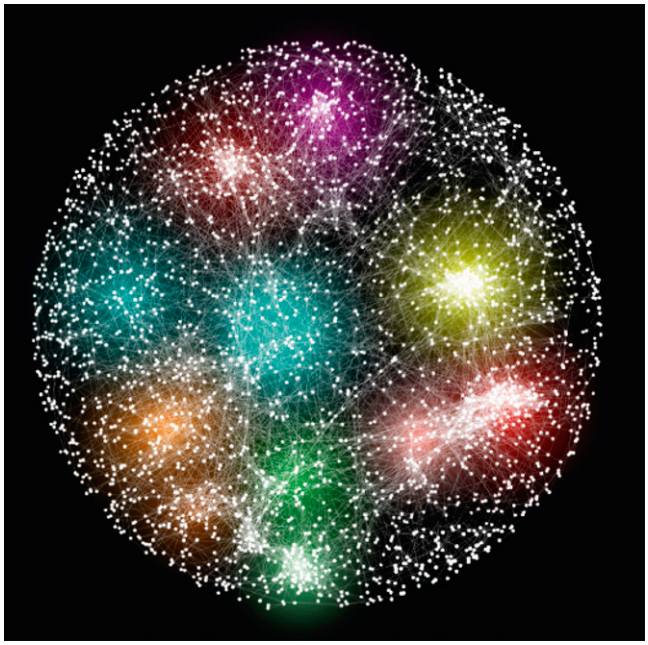

Figure 1 The synthetic gene array network as described by Costanzo *et al.* (2010) and analyzed and visualized by Baryshnikova *et al.* (2010). Each point represents a functional gene and each edge a functional connection, and the colored regions indicate subnetworks with similar GO process annotations.

Protein Interaction Networks

One of the great promises of functional analysis on the genomic scale was always the possibility of advancing beyond analysis of gene and protein functions one by one. Soon after the publication of the yeast genome sequence, a number of such technologies emerged. Each of these technologies identifies interactions among proteins or genes, which typically are visualized as a network. It is in the arena of understanding these networks of functional relationship that the main opportunities and challenges for understanding cellular biology at the system level lie. Yeast biology has led the way into this arena, and this appears to us to be the path forward for the future.

The first of these studies to appear was based on the two-hybrid method for detecting protein interactions (Fields and Song 1989). A number of large-scale efforts using variants of this approach have produced a large body of data. Although lower throughput versions of the two-hybrid method have been successfully employed to identify a great variety of protein–protein interactions, on a large scale the method has been plagued by large numbers of false-positive and false-negative signals, resulting in disappointingly little overlap between the most prominent of the published genome-scale networks (Uetz *et al.* 2000; Ito *et al.* 2001; reviewed by Fields 2009).

A much more reliable approach turned out to be a conceptually straightforward biochemical method that can nevertheless be implemented at high throughput. An epitope tag attached to every open reading frame enables precipitation of protein complexes. The identities of the coprecipitating proteins are determined by mass spectrometry. The method is reliable, and the proportion of false-positive signals is low, although it is clear that interactions below a threshold of affinity are unlikely to be detected (Krogan *et al.* 2006). Of course, care must be taken to ensure that the tagged protein is fully functional biologically, which can usually be ascertained by testing it for the ability to complement the cognate deletion mutation. Several large protein interaction data sets obtained by this method are in general agreement with each other and with other information. Like the other methods, affinity precipitation can, with some effort, be used to follow protein interactions dynamically.

Gene Interaction Networks

The study of double mutants became a mainstay of genetic analysis of biological function >50 years ago. Mutations that suppress or enhance the phenotypes of other mutations were used to discover much of what is known about molecular biology. Analysis of the phenotypes of double mutants was the basis for the analysis of metabolic, morphogenetic, and signal transduction pathways in all the organisms for which genetic manipulation is convenient. Indeed, the extraordinary facility with which mutations can be combined in yeast was one of the reasons that we were so optimistic about yeast as a model organism in 1988.

"Synthetic lethality" occurs when mutations in two different genes, each not lethal by itself, display a lethal phenotype when combined. Early studies in *Drosophila* and yeast interpreted synthetic lethality as an indication that the two genes have an essential function in common (Dobzhansky 1946; Sturtevant 1956; Novick *et al.* 1989). This notion led to the idea that one might screen for genes of similar function on this basis (Bender and Pringle 1991; see Guarente 1993 for an early review). Probably the most successful of the post-genome-sequence technologies in yeast has been the extension of this idea to a genomic scale (Tong *et al.* 2001; Costanzo *et al.* 2010).

The method, synthetic gene array (SGA) analysis, begins with the library of marked deletion mutants. Using robots to manipulate the thousands of strains of the library simultaneously, Tong *et al.* (2001) produced double mutants and determined their ability to grow on petri plates. They devised a clever scheme for crossing a strain carrying a mutation in a query gene to all the viable deletion mutants in the library and recovering haploid double-mutant progeny. This scheme has proved to be extremely robust.

Costanzo *et al.* (2010) studied 5.4 million double mutants using this SGA approach and produced quantitative genetic interaction profiles for ~75% of genes of *S. cerevisiae*. The resulting genetic interaction network (Figure 1) is remarkably coherent when examined either *in toto* or in sharper focus. Like the clustering and visualization schemes developed for DNA microarrays, analysis and visualization of SGA data provide an informative and intuitive picture of gene interactions.

Integrating Co-expression and Protein and Gene Interaction Networks

Much of the information about eukaryotic gene functions and interactions is derived from experiments with yeast, not only because of the advantages that we recognized in 1988, but also because experiments at the genome scale that are feasible with yeast are much more difficult to carry out with other eukaryotes. Nevertheless, each genome-scale method has issues of false-positive and false-negative signals. Although the results (seen as networks of co-expression, protein interaction, and gene interaction) are each reasonably self-consistent and (with the possible exception of two-hybrid results) reproducible, it is the integration of these data that provides the best view of a eukaryotic cell at the system level.

Bioinformatic methods for integrating genome-scale functional data have been in development for about a decade. As was the case with genome-scale experimental methods, yeast has been the obvious test bed for such methods because of the unparalleled fund of knowledge of the organism and because a large number of genome-scale data sets became available for yeast. To illustrate with data readily available on the internet, SGD currently provides a program (SPELL; Hibbs *et al.* 2007) to answer queries about gene expression on the basis of data from literally hundreds of data sets representing thousands of experimental conditions; Biopixie (http://avis.princeton.edu/pixie/index.php) uses a Bayesian network algorithm (Myers *et al.* 2005) to integrate the gene expression data with other kinds of genome-scale functional data, notably the results of SGA data, two-hybrid protein interaction data, and direct protein interaction data (see Wang *et al.* 2002 and Huttenhower *et al.* 2009 for recent reviews).

In the context of yeast as a model organism, it is worth mentioning that it appears to be straightforward to transfer to other organisms the bioinformatic methods devised for yeast data. Indeed, the major types of analysis—from the most basic analysis and display of gene expression and interaction data to the more complex algorithms used to infer gene and protein interactions—have been successfully applied to data on humans (*e.g.*, Huttenhower *et al.* 2009) as well as other model organisms, including plants (*e.g.*, *Arabidopsis*: Lee *et al.* 2011; Pop *et al.* 2010).

Leveraging Diversity to Understand Complex Inheritance

We have emphasized the role of yeast as a model for experiments that allow inferences of individual gene functions, gene and protein interactions, and network structures, many of which can be transferred to other eukaryotes, including humans, because of the high degree of evolutionary conservation of genome sequences. However, many inherited traits (in humans as well as other organisms) cannot be attributed to one or a few genes. These cases of complex inheritance have become a major roadblock to progress in understanding many common human diseases. The path forward in human genetics clearly involves the exploitation of the diversity in human genome sequences. Current efforts to use such tools have resulted in successful identification of common polymorphisms that correlated with disease, but unfortunately these polymorphisms collectively seem inadequate to account for the observed heritability (see Lander 2011 for a fuller discussion). It is clear that important elements of the rules for complex inheritance remain to be discovered.

Here again yeast has emerged as a model system. Brem *et al.* (2002) introduced the idea of using natural variation in yeast genomic sequences to model the inheritance of a simple class of quantitative trait: namely gene expression. They found (see also Yvert *et al.* 2003 and Ehrenreich *et al.* 2010) that many of these traits (called eQTL) exhibit complex inheritance. The logic of this approach is simple: if complex inheritance involves complex gene interactions, and if many of the genes of two organisms are highly conserved, then we could expect to learn much about the rules of complex inheritance by studying yeast, where any mechanistic ideas are easily followed up because of the experimental tractability and comprehensive knowledge of gene function that made yeast a leading model organism in the first place. Human geneticists (reviewed in Cheung and Spielman 2009) have already obtained substantial evidence that the methods pioneered in yeast can lead the way to understanding complex patterns of inheritance in humans and other eukaryotes.

Strengths and Weaknesses of Genome-Scale Experimentation and Inference: Experimental Validation Is Essential

As with any growing field, especially one that involves many new ideas and techniques, functional genomics has had its share of uncertainties and controversies. Many discussions have revolved around the reliability of genome-scale and high-throughput data, on the one hand, and the methods of inference, on the other. It is worth noting that many of these issues can be settled by direct experimental validation. An excellent recent example of this kind of experimental validation is the study of ~100 previously uncharacterized genes originally inferred, by bioinformatic analysis of mainly high-throughput data, to have mitochondrial functions. The great majority of the inferences proved accurate (Hibbs *et al.* 2009); about half of the genes are conserved in mammals, some of which are implicated in human diseases.

With the ability to validate predictions of protein function in yeast, the obvious advantage of the genomic approach is its efficiency and, potentially, its completeness. However, it should be kept in mind that many functions are required only in limited circumstances. To illustrate, in the cases of mitochondrial function, experiments on yeast grown in rich glucose media (where respiratory functions are known to be

dispensable) are unlikely to allow functional inference: indeed, the previous failure to characterize these genes may only reflect the preference of yeast researchers for such media. For there to be much more progress in finding the functions of the remaining uncharacterized yeast genes and gene interactions, a much wider exploration of the environmental universe of growth conditions and of genetic and environmental perturbations will be required. This applies not only to single gene characterizations, but also to the study of gene interactions and complex inheritance.

Evolution

There are two senses in which yeast is a useful model system for the study of evolution. One of these is our ability to infer, mainly from sequence, the course of evolution of the species and then perform experiments that illuminate these inferences. The other is in the use of yeast as an organism amenable to experimental studies of evolutionary adaptation to selective pressures because of its short doubling time.

Evidence for the theory of duplication and divergence

The large number of fungal genome sequences has provided firm support for the idea that the *S. cerevisiae* genome resulted from a whole-genome duplication (WGD) (Wolfe and Shields 1997; reviewed in Scannell *et al.* 2007; database: http://wolfe.gen.tcd.ie/ygob/; Byrne and Wolfe 2005). The following scenario for the evolution of the bakers' yeast genome is well supported by these data. About 100 million years ago a tetraploid yeast was formed by endo-reduplication of a diploid or by fusion of two yeast cells, each containing ~5000 genes. This ancestor lost ~85% of the duplicated copies, leaving the current species with ~6000 genes, ~10–20% of which are duplicated.

Nascent tetraploids created by the mating of two diploids are unstable: they lose whole chromosomes at a high rate, have increased DNA damage, and survive poorly in stationary phase (Andalis *et al.* 2004). These phenotypes may be a consequence of the increase in cell size associated with tetraploidy, as it has been shown that cell size alone alters the regulation of many genes and creates an unequal scaling between the spindle pole body, metaphase spindle length, and chromosome number (Galitski *et al.* 1999; Storchova *et al.* 2006; Wu *et al.* 2010).

These deleterious consequences would have affected the ancient tetraploid's relative fitness, so whole-genome duplication must have provided it with some advantage. A striking observation is that duplicate copies of 5 of the 10 genes for the glycolytic pathway have been retained. It has been proposed that the consequent increased dosage of glycolytic genes relative to those that have been lost provided an important selective advantage in a glucose-rich environment (Conant and Wolfe 2007). This increased gene dosage likely contributes to the largely unique ability of *S. cerevisiae* to maintain rapid growth rates on glucose even when oxygen levels are low.

This hypothesis that increased glycolytic flux was a result of the whole-genome duplication is attractive because the duplication event appears to have been coincident with the appearance of angiosperms, whose fruits have a high sugar content (Friis *et al.* 1987). It should be noted, however, that the ability to ferment glucose under aerobic conditions predates the WGD. Moreover, many additional alterations must have taken place to permit *S. cerevisiae* to grow in the high level of ethanol that results from this kind of metabolism, which is toxic to other microorganisms and could provide an additional selective advantage.

The fate of the duplicated genes has provided key insights into their evolutionary trajectory subsequent to the duplication. One possibility is "neo-functionalization," in which the preduplication ancestor performed one function, and, following duplication, one of the paralogs lost the old function and gained a new one (Conant and Wolfe 2008). Although the original notion by Ohno (1970) was that WGD provided the grist for evolution by permitting one of the paralogs to gain a new function, comparative genome studies suggest that most neo-functionalization does not occur by changes that result in novel catalytic functions, but rather by alterations in the regulatory responses of the paralogs (Wohlbach *et al.* 2009). Many duplicated genes, such as *CYC1* and *CYC7* and *COX5A* and *COX5B*, have retained their original catalytic function but respond to different environmental signals.

Another possible route after duplication is "subfunctionalization," where the ancestral gene had two functions, and after WGD one of the paralogs loses one of the functions and the second loses the other. A study that employed a comparative genomics approach coupled with a cross-species functional assay provided convincing evidence in support of subfunctionalization (Wapinski *et al.* 2010). It showed that the transcriptional activator Ifh1 and the repressor Crf1 that control ribosomal protein gene regulation in normal and stress conditions in *S. cerevisiae* are derived from the duplication and subsequent specialization of a single ancestral protein capable of carrying out both functions.

Experimental evolution studies with yeast

It has long been known that growing microorganisms under constant selection results in adaptive evolution, which has generally been observed by recovering heritable phenotypes that answer the selection applied. The availability of genome sequences and functional genomic technologies has made such studies more attractive for two reasons. First, one can follow the process of evolution itself, as it occurs in the laboratory over several hundred generations, using these molecular technologies. Studies of this kind have shown that repeated adaptive evolution studies under diverse conditions have a quite limited palate of outcomes, as assessed by the gene expression pattern of fitter variants that are recovered (Ferea *et al.* 1999; Gresham *et al.* 2008). Second, one can follow mutations that become enriched during the selective regime, by comparative genome

hybridization methods (Dunham *et al.* 2002) and ultimately by direct DNA sequencing (*cf*. Kao and Sherlock 2008).

One striking result of these studies is the observation of repeated chromosomal rearrangements that answer the selective pressure. These rearrangements are very similar to those observed in human tumor cells. It is already clear from these studies that the genetic mechanisms that underlie the adaptive evolution of yeast in the laboratory closely resemble those that must underlie the evolution of tumor cells during cancer progression.

Human Disease

In our 1988 *Science* essay, we anticipated that expression of heterologous proteins in yeast cells would facilitate the connection between structure and function in other organisms. "We think that conservation most strongly validates the use of yeasts as models for the primary deduction of functional and mechanistic aspects of proteins and protein systems shared by eukaryotes" (Botstein and Fink 1988, p. 442). What we did not anticipate was the enormous utility of such heterologous expression for assessment of function of human proteins associated with disease states.

Hereditary nonpolyposis colorectal cancer (HNPCC) is associated with defects in DNA mismatch repair. The history of this discovery about a major common human disease is instructive: an observation about instability of simple DNA repeats in tumors (Aaltonen *et al.* 1993) led to the deliberate isolation of mutations with this phenotype in yeast by Strand *et al.* (1993), who explicitly predicted that mutations causing HNPCC in humans might lie in the human orthologs of the three yeast genes (*PML1*, *MLH1*, and *MSH2*) that they identified. By the end of 1993, Fishel *et al.* (1993) had implicated the human ortholog of *MSH2* (now called hMSH2) as a causative gene for HNPCC. Mutations in either hMSH2 or hMLH1 underlie the great majority of HNPCC cases.

The functional characterization of defective human proteins in yeast can reveal aberrant enzyme functions that may not be apparent from assays in humans or from inspection of the protein sequence. An excellent example for this is hMSH2. Analysis in yeast of human *MSH2* mutant proteins from colon tumors revealed that some of them had defects expected to be caused by mutations known to affect catalytic function (Gammie *et al.* 2007). In addition, the analysis in yeast revealed that many of the human variants have lost crucial protein–protein interactions, others have reduced steady-state levels, and some influence the ATPase activity of the mismatch recognition complex. These defects, which appear to contribute to tumor formation, were unanticipated by direct study of humans.

Another benefit of studying human disease gene function in yeast is the potential for remediation of the deficiency responsible for the disease. Many missense mutations lead to decreased affinity for substrate or cofactor [altered Michaelis constant (Km)] that result in a metabolic defect because of insufficient product. It has been suggested that such Km defects could be compensated for by dietary intake of high concentrations of the substrate that could ameliorate the defect (Ames *et al.* 2002). For several human disorders, such as gyrate atrophy and CBS deficiency, which are caused by recessive mutations in B6-dependent enzymes (ornithine aminotransferase and cystathione synthase respectively), vitamin B-6-responsive patients have been identified (Clayton 2006). Recent work with yeast suggests that analysis of potential human CBS B6-responsive alleles in yeast strains carrying the cognate yeast mutations is extremely informative and can, like the Msh2 analysis, reveal defects unanticipated by other methods.

Prospects for the Future: Much Remains To Be Learned

It is clear that yeast has functioned as a model system that allows inference of individual gene functions, of gene and protein interactions, and of network structures through various kinds of experiments ranging from individual assays to high-throughput genome-scale experiments. We have emphasized how yeast has contributed to our understanding of basic biology in other eukaryotes and of human disease. We have every reason to expect that these contributions will continue and even grow in importance because the technology, especially for DNA sequencing, is continuing to improve rapidly.

It is difficult to predict the future, but we can sketch areas in which progress is likely to occur in the short to medium term. There is every expectation that the information in the yeast genome will continue to be mined. Although most of the open reading frames in the genome have now been annotated, there remains a dearth of information on the functions of noncoding RNAs. Moreover, the function of the many proteins shorter than 100 amino acids remains a challenge.

A fundamental unanswered question is the subcellular distribution of proteins and metabolites in the cell cycle. Currently, we have only a rudimentary cartography—a two-dimensional image of the pathways of metabolism and a catalog of corresponding protein functions. We are slowly obtaining a catalog of the protein–protein interactions, but the locations within a cell of most proteins and small molecules, so critical to their function, remain unknown. But imaging technology and analytical methods for molecular detection are improving rapidly, so we look forward to a three-dimensional understanding of the yeast cell that shows where each of the biological processes occurs.

Saccharomyces has already begun to play a central role in both the pharmaceutical and the industrial arenas. Yeast has a number of advantages for processes that require production on a large scale: the low cost of culture media and a history of efficient fermentation technology. Moreover, the yeast itself, a by-product of the production process, is a valuable commodity for animal feed and thus does not incur additional costs required for disposal.

Despite these advantages, yeast has been eschewed for the production of therapeutic proteins because the complex polysaccharides that decorate proteins secreted from yeast are immunogenic. A vaccine prepared from the hepatitis B antigen could be prepared from yeast because the protein is not secreted and forms particles easily separated from the cytosol. However, strains of *Picchia* with the entire yeast glycosylation pathway replaced with the one from humans (including the genes encoding the completely foreign proteins of the sialic acid pathway) have been constructed. This advance (Hamilton *et al.* 2003, 2006) has opened the door for the production of secreted proteins, such as monoclonal antibodies and cytokines.

The advantages of yeast alluded to earlier are even more critical in the production of biofuels, where the product is produced on an immense scale (Lam *et al.* 2010). Yeast is already used in Brazil and China to produce ethanol for fuel. A goal for the future is the conversion of plant polysaccharides (cellulose and xylans) to ethanol or other alcohols. Strains of yeast have been selected that produce ethanol from glucose at close to the theoretical maximum yield. However, the facts that yeast does not use pentoses as a carbon source, is inhibited by many of the breakdown products of lignin (*e.g.*, furfural), and does not have the pathways to produce branched chain alcohols (and is inhibited by them) present challenges for the future. Given the ease of the genetic manipulation of yeasts, it is likely that these problems will be solved.

Yeast is still the most facile organism for studying the relationship of genotype to phenotype in eukaryotic cells. Much is known about the transmission of the traditional carriers of information, DNA and RNA, during mitosis and meiosis, but little is known about the inheritance of organelles, macromolecules such as polysaccharides and lipids, and the myriad small molecules that populate the cells of living organisms. The recognition that the yeast [PSI]-factor is analogous to mammalian prions (Wickner 1994), is transmitted during mitosis and meiosis, and, like DNA, can transform cells has already added to our understanding of non-Mendelian inheritance.

Conclusion

In the 23 years since our last essay on this subject, *Saccharomyces* biology has progressed to the point where only ~15% of the *Saccharomyces* genes are completely without annotation (although it must be admitted that many of the annotations are really quite sketchy). Like many processes of discovery (indeed, like genome sequencing itself), annotating the last 15% may take even longer than it took to annotate the first 85%: completion of our understanding of the function of each and every gene will be an asymptotic process.

However, in the intervening time, yeast, more than any other organism, has led the way to another, potentially more important frontier beyond the functions of single genes and proteins: the "systems level." The goal is understanding the functions of ensembles of genes and proteins as they act to maintain metabolism and cellular homeostasis under a great diversity of environmental conditions and to provide for the regulation and organization of reproduction, cellular growth, and development. For the foreseeable future, the experimental advantages offered by yeast will serve to keep this model organism at the forefront of this new frontier.

Acknowledgments

We are grateful for support from National Institute of General Medical Sciences grants GM 046406 and GM-071508 to D.B. and GM40266 and GM035010 to G.R.F.

Literature Cited

Aaltonen, L. A., P. Peltomaki, F. S. Leach, P. Sistonen, L. Pylkkanen *et al.*, 1993 Clues to the pathogenesis of familial colorectal cancer. Science 260: 812–816.

Alter, O., P. O. Brown, and D. Botstein, 2000 Singular value decomposition for genome-wide expression data processing and modeling. Proc. Natl. Acad. Sci. USA 97: 10101–10106.

Ames, B. N., I. Elson-Schwab, and E. A. Silver, 2002 High-dose vitamin therapy stimulates variant enzymes with decreased coenzyme binding affinity (increased K(m)): relevance to genetic disease and polymorphisms. Am. J. Clin. Nutr. 75: 616–658.

Andalis, A. A., Z. Storchova, C. Styles, T. Galitski, D. Pellman *et al.*, 2004 Defects arising from whole-genome duplications in *Saccharomyces cerevisiae*. Genetics 167: 1109–1121.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botsteini, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25–29.

Baryshnikova, A., M. Costanzo, Y. Kim, H. Ding, J. Koh *et al.*, 2010 Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat. Methods 7: 1017–1024.

Bender, A., and J. R. Pringle, 1991 Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in Saccharomyces cerevisiae. Mol. Cell. Biol. 11: 1295–1305.

Botstein, D., 2010 It's the data! Mol. Biol. Cell 21: 4–6.

Botstein, D., and G. R. Fink, 1988 Yeast: an experimental organism for modern biology. Science 240: 1439–1443.

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

Brown, P. O., and D. Botstein, 1999 Exploring the new world of the genome with DNA microarrays. Nat. Genet. 21: 33–37.

Byrne, K. P., and K. H. Wolfe, 2005 The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15: 1456–1461.

Cheung, V. G., and R. S. Spielman, 2009 Genetics of human gene expression. mapping DNA variants that influence gene expression. Nat. Rev. Genet. 10: 595–604.

Chu, S., J. DeRisi, M. Eisen, J. Mulholland, D. Botstein *et al.*, 1998 The transcriptional program of sporulation in budding yeast. Science 282: 699–705.

Clayton, P. T., 2006 B6-responsive disorders: a model of vitamin dependency. J. Inherit. Metab. Dis. 29: 317–326.

Conant, G. C., and K. H. Wolfe, 2007 Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol. Syst. Biol. 3: 129.

Conant, G. C., and K. H. Wolfe, 2008 Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9: 938–950.

Costanzo, M., A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear *et al.*, 2010 The genetic landscape of a cell. Science 327: 425–431.

DeRisi, J. L., V. R. Iyer, and P. O. Brown, 1997 Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278: 680–686.

Dobzhansky, T., 1946 Genetics of natural populations. Xiii. Recombination and variability in populations of *Drosophila pseudoobscura*. Genetics 31: 269–290.

Dolinski, K., and D. Botstein, 2005 Changing perspectives in yeast research nearly a decade after the genome sequence. Genome Res. 15: 1611–1619.

Dolinski, K., and D. Botstein, 2007 Orthology and functional conservation in eukaryotes. Annu. Rev. Genet. 41: 465–507.

Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown *et al.*, 2002 Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 99: 16144–16149.

Ehrenreich, I. M., N. Torabi, Y. Jia, J. Kent, S. Martis *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464: 1039–1042.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein, 1998 Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95: 14863–14868.

Ferea, T. L., D. Botstein, P. O. Brown, and R. F. Rosenzweig, 1999 Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc. Natl. Acad. Sci. USA 96: 9721–9726.

Fields, S., 2009 Interactive learning: lessons from two hybrids over two decades. Proteomics 9: 5209–5213.

Fields, S., and O. Song, 1989 A novel genetic system to detect protein-protein interactions. Nature 340: 245–246.

Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins *et al.*, 1993 The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75: 1027–1038.

Friis, E. M., W. G. Chaloner, and P. R. Crane, 1987 *The Origins of Angiosperms and Their Biological Consequences*. Cambridge University Press, Cambridge, UK.

Galitski, T., A. J. Saldanha, C. A. Styles, E. S. Lander, and G. R. Fink, 1999 Ploidy regulation of gene expression. Science 285: 251–254.

Gammie, A. E., N. Erdeniz, J. Beaver, B. Devlin, A. Nanji *et al.*, 2007 Functional characterization of pathogenic human MSH2 missense mutations in *Saccharomyces cerevisiae*. Genetics 177: 707–721.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen *et al.*, 2000 Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11: 4241–4257.

Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle *et al.*, 2003 Global analysis of protein expression in yeast. Nature 425: 737–741.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002 Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387–391.

Gresham, D., M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai *et al.*, 2008 The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4: e1000303.

Guarente, L., 1993 Synthetic enhancement in gene interaction: a genetic tool come of age. Trends Genet. 9: 362–366.

Hamilton, S. R., P. Bobrowicz, B. Bobrowicz, R. C. Davidson, H. Li *et al.*, 2003 Production of complex human glycoproteins in yeast. Science 301: 1244–1246.

Hamilton, S. R., R. C. Davidson, N. Sethuraman, J. H. Nett, Y. Jiang *et al.*, 2006 Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313: 1441–1443.

Heinicke, S., M. S. Livstone, C. Lu, R. Oughtred, F. Kang *et al.*, 2007 The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. PLoS ONE 2: e766.

Hibbs, M. A., D. C. Hess, C. L. Myers, C. Huttenhower, K. Li *et al.*, 2007 Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics 23: 2692–2699.

Hibbs, M. A., C. L. Myers, C. Huttenhower, D. C. Hess, K. Li *et al.*, 2009 Directing experimental biology: a case study in mitochondrial biogenesis. PLoS Comput. Biol. 5: e1000322.

Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson *et al.*, 2003 Global analysis of protein localization in budding yeast. Nature 425: 686–691.

Huttenhower, C., M. A. Hibbs, C. L. Myers, A. A. Caudy, D. C. Hess *et al.*, 2009 The impact of incomplete knowledge on evaluation: an experimental benchmark for protein function prediction. Bioinformatics 25: 2404–2410.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman, and J. S. Weissman, 2009 Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218–223.

Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori *et al.*, 2001 A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98: 4569–4574.

Iyer, V. R., C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder *et al.*, 2001 Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409: 533–538.

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

Kim, D. U., J. Hayles, D. Kim, V. Wood, H. O. Park *et al.*, 2010 Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo *et al.*, 2006 Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440: 637–643.

Lam, F. H., F. S. Hartner, G. R. Fink, and G. Stephanopoulos, 2010 Enhancing stress resistance and production phenotypes through transcriptome engineering. Methods Enzymol. 470: 509–532.

Lander, E. S., 2011 Initial impact of the sequencing of the human genome. Nature 470: 187–197.

Lee, J. H., W. Terzaghi, and X. W. Deng, 2011 DWA3, an Arabidopsis DWD protein, acts as a negative regulator in ABA signal transduction. Plant Sci. 180: 352–357.

Lieb, J. D., X. Liu, D. Botstein, and P. O. Brown, 2001 Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat. Genet. 28: 327–334.

Myers, C. L., D. Robson, A. Wible, M. A. Hibbs, C. Chiriac *et al.*, 2005 Discovery of biological networks from diverse functional genomic data. Genome Biol. 6: R114.

Novick, P., B. C. Osmond, and D. Botstein, 1989 Suppressors of yeast actin mutations. Genetics 121: 659–674.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, London/New York.

Pop, A., C. Huttenhower, A. Iyer-Pascuzzi, P. N. Benfey, and O. G. Troyanskaya, 2010 Integrated functional networks of process, tissue, and developmental stage specific interactions in Arabidopsis thaliana. BMC Syst. Biol. 4: 180.

Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao *et al.*, 2007 Genome-wide profiles of STAT1 DNA association using

chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods 4: 651–657.
Scannell, D. R., G. Butler, and K. H. Wolfe, 2007 Yeast genome evolution: the origin of the species. Yeast 24: 929–942.
Scherens, B., and A. Goffeau, 2004 The uses of genome-wide yeast mutant collections. Genome Biol. 5: 229.
Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders *et al.*, 1998 Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9: 3273–3297.
Storchova, Z., A. Breneman, J. Cande, J. Dunn, K. Burbank *et al.*, 2006 Genome-wide genetic analysis of polyploidy in yeast. Nature 443: 541–547.
Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274–276.
Sturtevant, A. H., 1956 A highly specific complementary lethal system in *Drosophila melanogaster*. Genetics 41: 118–123.
Tong, A. H., M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader *et al.*, 2001 Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368.
Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson *et al.*, 2000 A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623–627.
Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag *et al.*, 2002 Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. USA 99: 5860–5865.
Wapinski, I., J. Pfiffner, C. French, A. Socha, D. A. Thompson *et al.*, 2010 Gene duplication and the evolution of ribosomal protein gene regulation in yeast. Proc. Natl. Acad. Sci. USA 107: 5505–5510.
Wickner, R. B., 1994 [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science 264: 566–569.
Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson *et al.*, 1999 Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285: 901–906.
Wohlbach, D. J., D. A. Thompson, A. P. Gasch, and A. Regev, 2009 From elements to modules: regulatory evolution in Ascomycota fungi. Curr. Opin. Genet. Dev. 19: 571–578.
Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.
Wu, C. Y., P. A. Rolfe, D. K. Gifford, and G. R. Fink, 2010 Control of transcription by cell size. PLoS Biol. 8: e1000523.
Yvert, G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss *et al.*, 2003 Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat. Genet. 35: 57–64.

Communicating editor: A. Hinnebusch