

Review

Why Are There Still Over 1000 Uncharacterized Yeast Genes?

Lourdes Peña-Castillo and Timothy R. Hughes¹

Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada

ABSTRACT

The yeast genetics community has embraced genomic biology, and there is a general understanding that obtaining a full encyclopedia of functions of the ~6000 genes is a worthwhile goal. The yeast literature comprises over 40,000 research papers, and the number of yeast researchers exceeds the number of genes. There are mutated and tagged alleles for virtually every gene, and hundreds of high-throughput data sets and computational analyses have been described. Why, then, are there >1000 genes still listed as uncharacterized on the *Saccharomyces* Genome Database, 10 years after sequencing the genome of this powerful model organism? Examination of the currently uncharacterized gene set suggests that while some are small or newly discovered, the vast majority were evident from the initial genome sequence. Most are present in multiple genomics data sets, which may provide clues to function. In addition, roughly half contain recognizable protein domains, and many of these suggest specific metabolic activities. Notably, the uncharacterized gene set is highly enriched for genes whose only homologs are in other fungi. Achieving a full catalog of yeast gene functions may require a greater focus on the life of yeast outside the laboratory.

SEVERAL years ago we projected that, judging from the steady and predictable increase in database entries, all yeast genes would have “known” functions by mid-2007 (HUGHES *et al.* 2004). This was an optimistic estimate, since what was known at the time about many of the “known” genes was unsatisfying (or even conflicting) in terms of actual understanding of physiological purpose or molecular role. Nonetheless, it seemed reasonable, and still does, that yeast will be the first organism for which the functions of all the genes are characterized to a degree that would satisfy most molecular biologists.

Characterizing the functions of all of the yeast genes is not a senseless academic pursuit. Yeast is a part of everyday life, being critical to baking and brewing. Yeasts can also be pathogenic. Much of what is known about how all eukaryotic cells work has come from studying yeast. And the possibility that something would be known about virtually all yeast genes only a decade or so after the initial sequencing [at which time only ~1000 genes appeared in the literature, and roughly half could be ascribed some function on the basis of sequence features (GOFFEAU *et al.* 1996)] could be taken as a harbinger of what might eventually be anticipated from similar efforts in more complicated organisms with larger genomes and more genes, such as humans. Examination of the path taken to such an achievement, including any hurdles or missteps, should provide a level of gui-

dance. Although many organisms now have a genome sequence and a set of useful genetic tools, yeast has been the major proving ground for large-scale application of new technologies, and in many respects remains the advance guard of functional genomics, proteomics, and systems biology, as outlined in other recent reviews (BADER *et al.* 2003; MUSTACCHI *et al.* 2006; SUTER *et al.* 2006). Knowing what all the parts do is important if you want to know how a machine works.

It is now mid-2007. Are the functions of all 6000 yeast genes “known”? Given a very liberal definition of known function, this goal has nearly been reached. The commercial database used in our initial time line (HUGHES *et al.* 2004) is no longer freely available. We therefore examined information present in the *Saccharomyces* Genome Database (SGD) (NASH *et al.* 2007), the public database that curates the yeast genes and also compiles Gene Ontology (GO) annotations, publications, interactions, and a host of additional information. As of March 20, 2007, there are only 38 genes with no information available whatsoever, and only 566 lack any annotations in any of the three major branches of Gene Ontology (biological process, molecular function, or cellular component), which is a somewhat fluid categorical data structure that incorporates information with a variety of evidence levels (ASHBURNER *et al.* 2000). This is, to be sure, a phenomenal achievement, having occurred mostly in a single decade.

Unfortunately, much of this annotation is derived from large-scale studies, and depth is lacking for many genes. In the last 3 years, the overall proportion of genes

¹Corresponding author: Banting and Best Department of Medical Research, 160 College St., Room 1302, University of Toronto, Toronto, ON M5S 3E1, Canada. E-mail: t.hughes@utoronto.

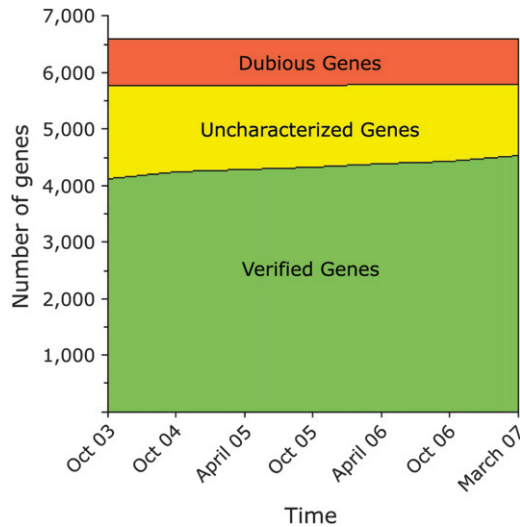


FIGURE 1.—Distribution of ORFs in “dubious,” “uncharacterized,” and “verified” as classified by SGD since October 2003.

that SGD classifies as “uncharacterized ORF” has not changed a great deal, decreasing by only 285 genes (Figure 1). To be certain, gene characterization is not a binary attribute, and not everyone will agree on the definition or the individual assignments. Nonetheless, the list of “uncharacterized ORF” classifications confirms that the vast majority of SGD decisions are in agreement with our own sense of gene characterization. Although expectation differs depending on the type of protein and its proposed role, in our view gene characterization would typically be expected to include multiple lines of specific and consistent evidence, often encompassing phenotype, cell biology, and biochemistry. By manual analysis, we found only 53 (<5%) genes classified as uncharacterized by SGD that we considered characterized by conventional standards. For many of the uncharacterized ORFs, there is little or no experimental evidence to support any specific physiological function or even functionality—although there are many instances such as that of YDR185C: “Mitochondrial protein of unknown function; has similarity to Ups1p, which is involved in regulation of alternative topogenesis of the dynamin-related GTPase Mgm1p,” or YKL098W, which encodes a 357 amino acid “Putative protein of unknown function,” and interacts genetically with both *CDC8* and *SKP1*, which are involved in DNA synthesis and mitosis, respectively.

As of March 20, 2007, there are still 1253 uncharacterized yeast genes listed on SGD—21% of all genes. Why so many? What might their functions be? How will we go about characterizing them? Previous analyses have outlined the relative strengths of specific approaches in functional genomics and proteomics, usually using cross-validation (*i.e.*, sensitivity and specificity at reproducing the genes whose functions are already

known in blinded tests) (VON MERING *et al.* 2002; WONG *et al.* 2005; MYERS *et al.* 2006). All methods work in at least some instances and, in selected cases, have been extremely successful. RNA and ribosome biogenesis categories, for example, which include hundreds of genes, often dominate functional predictions, since these genes and their encoded proteins share regulatory patterns, exist in relatively abundant protein complexes, and often contain conserved sequence features. Consequently, investigators have capitalized on large-scale data sets to study new RNA and ribosome biogenesis genes and their products (*e.g.*, GRANDI *et al.* 2002; PENG *et al.* 2003).

It is less clear how useful predictions from large-scale data are in all classes of gene function, however. Our previous analysis (HUGHES *et al.* 2004) included predicted GO annotations for 122 proteins that were uncharacterized at the time, but whose GO assignments were supported by three or more large-scale data sets. While it is encouraging that 82 of these have since been characterized, only 23 were assigned to one of the precise predicted annotation categories, and 12 of these were in RNA processing or ribosome biogenesis. This is certainly higher than random, but cautions that predictions might best be used as a rough guide to areas for further exploration.

Thus, rather than evaluating the relative merits of the methods, here we instead consider the uncharacterized genes themselves, with an eye to why they are not yet known, how their functions might be understood, and what lessons might hold for similar efforts in other genomes.

ARE ALL OF THE UNCHARACTERIZED GENES REAL?

A trivial explanation for difficulty in associating uncharacterized genes with specific functions would be that they are not real genes. The yeast gene catalog changes occasionally as gene boundaries are redefined and small expressed ORFs are discovered (FISK *et al.* 2006). Expression or even localization (discussed below), which are often taken as evidence that a gene is real, is not on its own conclusive proof of function. Even mutant phenotypes are not foolproof, as they can be caused by disruption of regulatory sequence. Anecdotally, microarray analysis of yeast ORF deletion strains occasionally reveals misregulation of an adjacent gene on the chromosome (T. R. HUGHES, unpublished observation), which could, in principle, have a phenotypic consequence that is unrelated to the function of the deleted gene. These caveats partially underlie the general notion that gene characterization should be based on more than one independent line of evidence.

Perhaps the best evidence that most of the uncharacterized yeast genes are *bona fide* genes comes from sequence attributes. The vast majority are conserved in

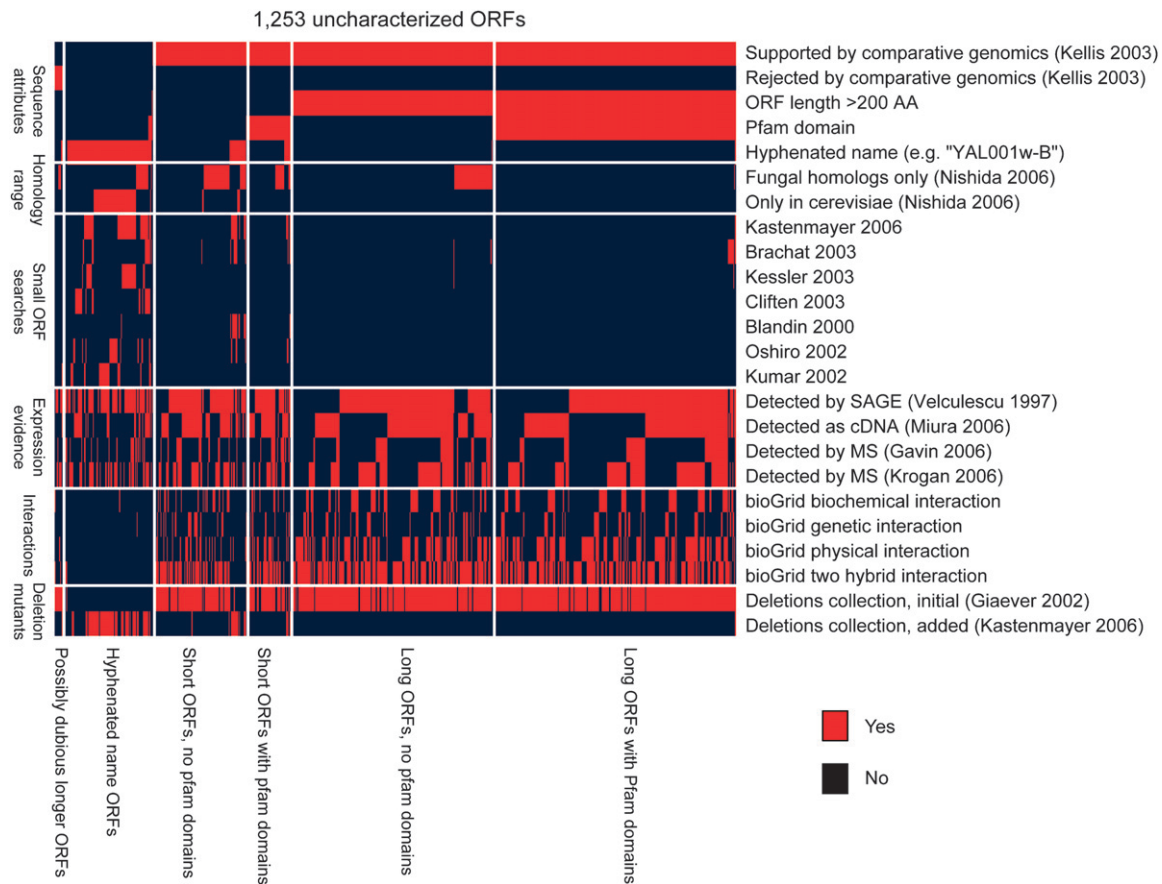


FIGURE 2.—Properties of uncharacterized yeast genes. Information from supplemental data of the articles indicated (VELCULESCU *et al.* 1997; BLANDIN *et al.* 2000; GIAEVER *et al.* 2002; KUMAR *et al.* 2002; OSHIRO *et al.* 2002; BRACHAT *et al.* 2003; CLIFTEN *et al.* 2003; KELLIS *et al.* 2003; KESSLER *et al.* 2003; GAVIN *et al.* 2006; KASTENMAYER *et al.* 2006; KROGAN *et al.* 2006; MIURA *et al.* 2006; NISHIDA 2006; STARK *et al.* 2006).

syntenic positions in related yeasts, indicating selective pressure for retention (CLIFTEN *et al.* 2003; KELLIS *et al.* 2003) (Figure 2). For nearly half (538), other sequence features include a domain structure cataloged on Pfam (FINN *et al.* 2006) (Figure 2). In addition, although the length of uncharacterized genes is shorter than those of characterized genes, on the whole (median 281 aa *vs.* 379 aa for all yeast proteins), nearly two-thirds (822/1253) are >200 amino acids long (Figure 2). ORFs of such length would occur at most a few dozen times in a random-sequence version of the yeast genome. Therefore, the majority of uncharacterized yeast genes are probably real genes, albeit not all.

ARE THE UNCHARACTERIZED GENES TOO NEW TO HAVE BEEN STUDIED?

Another explanation for lack of functional knowledge on many yeast genes is that they have only recently been discovered. Systematic yeast gene names (*e.g.*, YAL001W) added after the initial sequence assembly (GOFFEAU *et al.* 1996) typically carry a suffix (*e.g.*,

YAL001W-A) (FISK *et al.* 2006). A total of 204 of the uncharacterized genes carry such a suffix, while only 285 genes overall carry such a suffix [not including “dubious ORFs,” which are no longer believed to be genes (FISK *et al.* 2006)]. Not being in the database clearly puts a gene at a disadvantage for characterization: 188 of the 204 uncharacterized genes carrying a suffix were not in the initial deletions consortium collection (GIAEVER *et al.* 2002), which has formed the basis of hundreds of genetic screens over the last 5 years; similarly, 156 are not in the TAP/GFP collections (GHAEMMAGHAMI *et al.* 2003; HUH *et al.* 2003). However, on the whole, most of the uncharacterized genes have been in databases for over a decade. The majority (982) are present in the initial deletions consortium collection and many others have been added subsequently (KASTENMAYER *et al.* 2006) (Figure 2). In any case, lack of ready genetic reagents does not prevent proteins from appearing in pulldowns or emerging from traditional genetic screens. Therefore, although some of the uncharacterized genes are newer than other yeast genes, they are a fairly small minority (Figure 2).

DO UNCHARACTERIZED GENES HAVE ANY DISTINGUISHING CHARACTERISTICS IN LARGE-SCALE ANALYSES?

Uncharacterized genes tend to be underrepresented in virtually all types of large-scale data (HUGHES *et al.* 2004), and this may partially explain why they are not studied as frequently; however, we found no criteria that completely separate currently uncharacterized from characterized genes. One of the greatest differences in distribution of properties is requirement for viability: <4% of uncharacterized genes (40) are listed as essential, in comparison to ~19% of genes on the whole (FISK *et al.* 2006). In the GFP localization database (HUH *et al.* 2003), there are no striking differences in distribution among subcellular categories, and while 72% of “verified” ORFs are localized (3491/4505), in comparison to 57% of uncharacterized ORFs (629/1094), this is at least partially due to the use of localization information in gene characterization since 2003. The difference in the distribution of expression levels between characterized and uncharacterized is significant, although very far from a bimodal distribution (Figure 3). Uncharacterized genes are also represented in interaction databases, although again less frequently than characterized genes. In the current BioGrid catalog (STARK *et al.* 2006), 683 uncharacterized genes have interactions by two hybrid, 243 have interactions by biochemical activity (*e.g.*, phosphorylation), 457 have protein–protein interactions (including pull-downs but not two hybrid), and 197 participate in genetic interactions. Although the meaning of all of these interactions remains to be seen, it is certainly possible to draw hypotheses. Altogether, the number of uncharacterized genes cannot be generally attributed to their absence from large-scale data sets.

ARE THE UNCHARACTERIZED GENES NEEDED ONLY UNDER CERTAIN CONDITIONS?

Accepting that the current set of uncharacterized genes are in large part real genes, that most of them are present in current reagent sets (including the deletions collection), and that many are expressed under standard laboratory conditions and participate somehow in the life of the cell, we are faced with the possibility that there is something about their biology that makes them difficult to characterize. That they tend to lack strong phenotypes indicates that many of them are not needed under standard laboratory conditions. Two general explanations are (1) that the uncharacterized genes are redundant with other genes or with each other or (2) that they are important only under conditions that are not usually assayed in the laboratory. These explanations are related, are not mutually exclusive, and one or both may apply to only a subset of genes.

Are these explanations supported by data? Redundancy carries several expectations. First, since one mech-

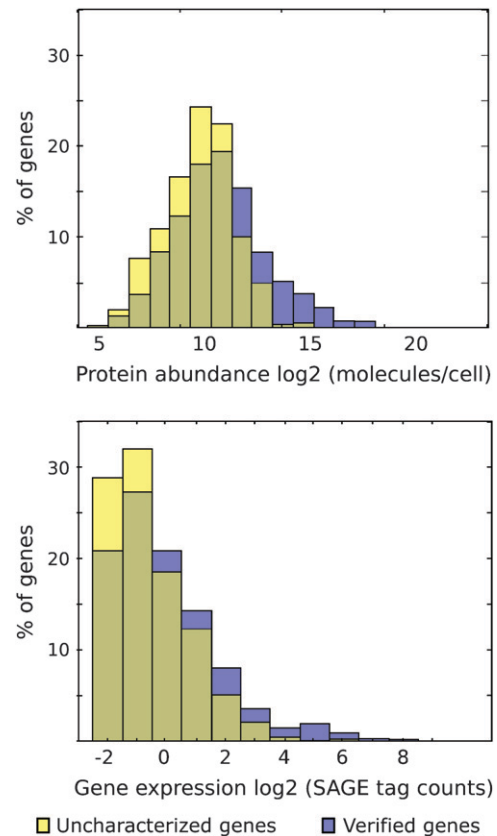


FIGURE 3.—Uncharacterized *vs.* verified ORFs expression as mRNA (VELCULESCU *et al.* 1997) and protein (GHAEMMAGHAMI *et al.* 2003).

anism for achieving redundancy is gene duplication, we might expect the uncharacterized ORFs to contain many duplicated genes. In fact, 161 uncharacterized proteins have sequences at least 50% identical to the sequences of another uncharacterized protein (median blast *E*-value 10^{-51}), and these 161 proteins cluster into 54 groups (Figure 4). These clusters all contain between two and five proteins, with the exception of 26 *Y'* helicase-like elements and 15 seripauperin (PAU) family genes, which may be cell-wall mannoproteins (ABRAMOVA *et al.* 2001). *Y'* and PAU genes are found near telomeres, and in fact most of the duplicated genes overlap or are adjacent to telomeres, tRNAs, or transposon-like sequences (Figure 4), loci that are prone to duplication, recombination, and rearrangement. It is conceivable that at least some of these ORFs are parasitic, and/or react to unusual circumstances that affect the loci they occupy (*e.g.*, telomeres). Their role could benefit from a large number of family members, for example, to achieve high copy number, rapid diversification, and/or redundancy, making them complicated to study by genetics and harder to identify in genetic screens. The uncharacterized gene set also contains several groups of proteins with similar domain structures, but more distant primary sequences; these could also be genetically redundant.

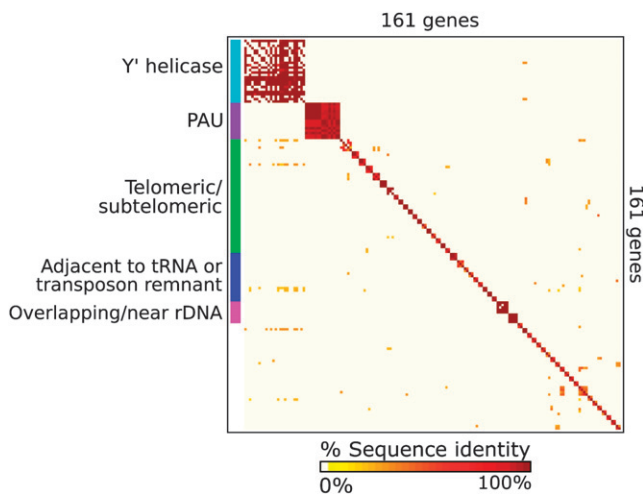


FIGURE 4.—Redundant uncharacterized ORFs. The heat map shows the percentage of identical sequence among 161 uncharacterized ORFs (the same ordering is used on both axes).

Second, redundancy might result in a substantial number of genetic interactions with and among uncharacterized genes. Genetic redundancy with >100 genes has been measured systematically across all strains in the deletion collection (TONG *et al.* 2004). In the BioGrid database (STARK *et al.* 2006), only 197 (or 18%) of uncharacterized genes have at least one genetic interaction, in comparison to 3275 (or 72%) of verified ORFs. This is far from conclusive evidence, as only a small proportion of the full genetic interaction network has been mapped (even under a single growth condition) and relatively few uncharacterized genes have been crossed to all other deletions. In general, it is more difficult to understand the mechanistic basis of genetic interactions than it is for physical interactions, and there is still some debate regarding expectation and scoring of double-mutant crosses. Nonetheless, this result does raise the possibility that at least some of the uncharacterized genes exist solely as genetic buffers, but also suggests that high genetic redundancy does not distinguish them from characterized genes.

If the uncharacterized genes are important only under specific conditions—particular situations that are not normally tested in the laboratory—then we would expect mutation of these genes to have little or no phenotype in general, although they may be constitutively expressed. Results obtained to date do roughly match this expectation. Another expectation of genes required to survive specific situations in nature is that their sequence features may be indicative of activities related to interaction with the environment. Many of them do contain sequence features indicating that they are biosynthetic enzymes (147) or transporters/permeases (40), consistent with a role in dealing with biochemistry in the wild. Consistent with this notion, the uncharacterized gene set contains an extremely significant number of

genes (177 of 405) that have a homolog in one or more other fungi, but not in any other organisms (NISHIDA 2006). The set of 228 characterized genes with exclusively fungal homologs is enriched for yeast-specific functions, including genes involved in cell-wall organization and biogenesis, mating, and cell-cell adhesion (ROBINSON *et al.* 2002).

HOW WILL WE COMPLETE THE ENCYCLOPEDIA OF THE CELL?

To summarize the evidence, a variety of factors are likely to contribute to the relatively large number of as-yet-uncharacterized yeast genes, including genetic redundancy, lack of strong growth phenotype, and the possibility that not all of them are real genes. In addition, many of the remaining genes may be involved in environmental and metabolic responses or growth modes not normally queried in the laboratory. Thus, despite the success of high-throughput methods in many instances, characterization of these remaining refractory genes may require a different tack.

Certainly there is no shortage of researchers and enthusiasm. On the basis of first-author names on articles cataloged at SGD, there were at least 9447 yeast researchers active between 2003 and 2007. Yeast molecular biologists typically jump at the chance to describe a new gene function, especially if it involves a topic of current interest; for example, *RTT109*, although previously known to influence Ty transposition (SCHOLES *et al.* 2001), was simultaneously described as encoding a histone H3-K56 acetyltransferase by at least five different groups in late 2006 and early 2007 (SCHNEIDER *et al.* 2006; COLLINS *et al.* 2007; DRISCOLL *et al.* 2007; HAN *et al.* 2007; TSUBOTA *et al.* 2007).

Indeed, it does appear that present approaches to gene characterization result in a gradual increase in the proportion of genes with known functions (Figure 1). What is of concern is that if the trend in Figure 1 continues at its present course, characterization of all yeast genes (that would be likely to pass peer review) will be achieved in the neighborhood of the year 2020 (and likely later, assuming an asymptotic approach), which is much longer than many yeast enthusiasts would have hoped for. How might the conquering of uncharacterized yeast genes be accelerated? We can propose several concrete experimental possibilities, which might also be accompanied by some modification of how we think about the enterprise of gene characterization.

First, a great deal might be accomplished simply by having seasoned yeast researchers systematically peruse the list of uncharacterized genes and their attributes—including what is known or predicted about them from sequence features or large-scale surveys—and encouraging those engaged in large-scale efforts to focus on the uncharacterized genes. Figure 5 gives an overview of information regarding the function or functionality of

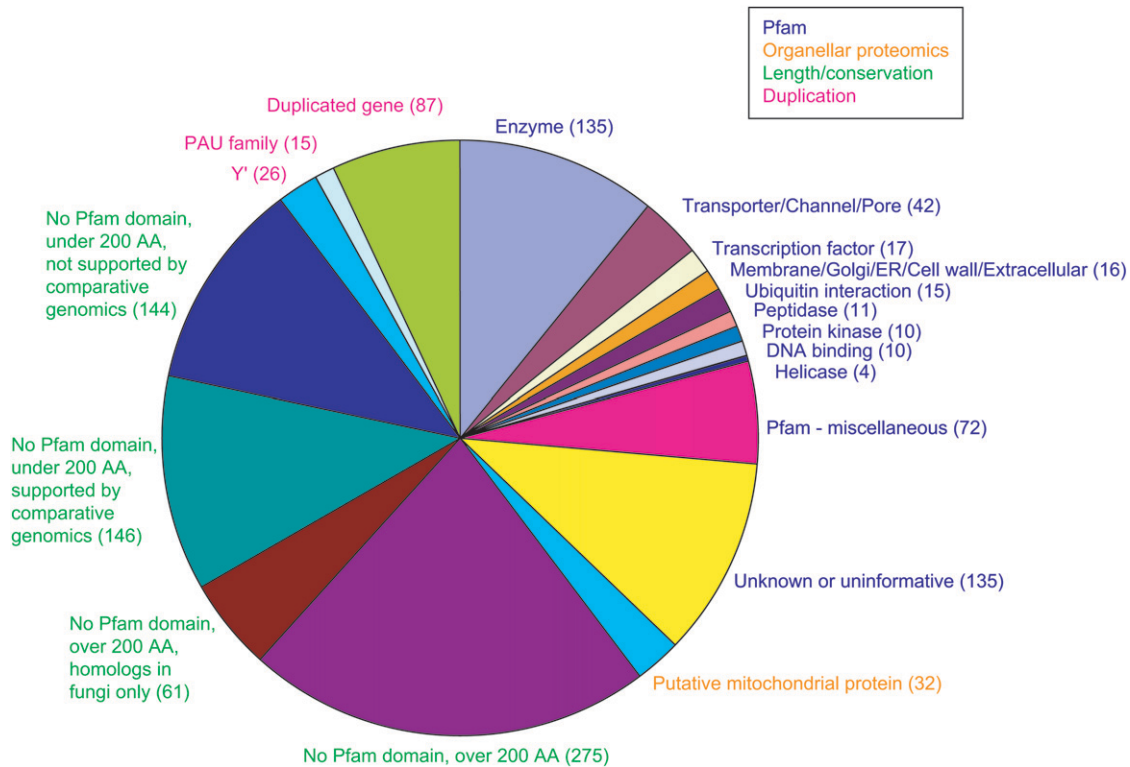


FIGURE 5.—Potential function or functionality for 1253 uncharacterized yeast genes.

the 1253 currently uncharacterized genes, and the supplemental data at <http://hugheslab.med.utoronto.ca/geneticsReview07/> contains a spreadsheet summarizing the information in Figures 2 and 5. Engaging individual researchers is often a major hurdle in proceeding from categorical predictions (which are what typically emerge from large-scale and computational analyses) to biological insight. Admittedly, individual labs may have already attempted to characterize these genes without much success (or alternatively, the articles are all in preparation). Assuming not, many of these genes may be approachable by the conventional one-gene-at-a-time hypothesis-driven approach or by devising assays that probe a specific pathway or activity. For example, the uncharacterized genes include 10 protein kinases; both genetic and biochemical array-based methods have been developed for the identification of kinase targets (PTACEK *et al.* 2005), thus providing clues for physiological roles. The uncharacterized gene list also includes several potential RNA-binding proteins [YLR271W (G-patch domain), YPL184C/MRNI (RRM domain), YJL010C/NOP9 (PUF domain), YGR250C (RRM domain), and YFR032C (RRM domain)], whose ligands might be explored using microarrays (GERBER *et al.* 2004).

Given the large number of apparent biosynthetic enzymes and mitochondrial proteins, a global metabolite profiling system would be extremely helpful. Such a system has not yet entered into the mainstream, although the technology exists (GRIFFIN 2006). Nonetheless, biochemical genomics (MARTZEN *et al.* 1999), in which

genes are characterized by the activities of their purified products, and the related protein chip approach that deposits the same proteins on a microscope slide for parallel analysis (ZHU *et al.* 2001) appear to be partly filling the same niche. A surprising number of tRNA-modifying activities have been identified using biochemical genomics, including that encoded by *TRM13*, which is still listed on SGD as uncharacterized but was published this month as a tRNA 2'-O-methyltransferase (WILKINSON *et al.* 2007). Structural genomics is a form of “reverse” biochemical genomics: *SDO1*, which was changed from “uncharacterized” to “verified” by SGD in March 2007, was found by structural genomics to have an OB fold (a common oligonucleotide-binding motif), suggesting interaction with RNA, consistent with its coregulation with other RNA biogenesis factors (SAVCHENKO *et al.* 2005). Data published this month show that *SDO1* specifically controls maturation and translational activation of ribosomes (MENNE *et al.* 2007). Chemical genetic profiling (GIAEVER *et al.* 2002; PARSONS *et al.* 2006), which simultaneously probes the sensitivity of all of the deletion strains to a small molecule—often a metabolic inhibitor or an environmental chemical stress—also qualifies as a form of biochemical genomics and has the advantage of inherently revealing phenotypes.

While these examples highlight fundamental aspects of cell biology, which are amenable to laboratory study, it appears as if more exploration of the life of yeast in the wild would also be productive, on the basis of the fact that half of the fungal-specific genes remain

uncharacterized. Examination of the lifestyle and genetics of yeast in its ecological niche is still in its infancy. In nature, *Saccharomyces cerevisiae* appears to be one of many opportunistic microorganisms found in rotting fruit (MORTIMER and POLSINELLI 1999; FLEET 2003). The cells presumably are dormant much of the time. Fruit flies prey on yeast, but they and other insects are also critical to the yeast life cycle, as they are believed to be the primary vector for transport of yeast (MORTIMER and POLSINELLI 1999). Thus, ability to stick to insects through dormancy, without being eaten, might be a requisite for propagation of wild yeast. Chemical defenses and competitive nutrient utilization strategies are also undoubtedly essential for wild yeast; even human-made fermentations contain a wide variety of interacting microbes, including yeasts, other fungi, and bacteria (FLEET 2003). Indeed, in addition to the biosynthetic enzymes and transporters listed above, the uncharacterized genes include nine zinc-cluster transcription factors [most characterized members of this class of proteins regulate metabolic pathways or responses to environmental stress (MACPHERSON *et al.* 2006)] and a smattering of other proteins that appear likely to be involved in environmental adaptation. YNL234W, for example, is a direct target of Rgt1 (a glucose-responsive transcription factor) and encodes a functional globin (SARTORI *et al.* 1999; KANIAK *et al.* 2004); however, its physiological purpose remains elusive. Other uncharacterized proteins may be related to the cell surface or cell shape; YOR111W, for instance, consists almost entirely of a Maf domain, named for a protein that influences septum formation in bacteria by an unknown mechanism (BUTLER *et al.* 1993). Some environmental behaviors, such as filamentous growth (KRON 1997), might be better modeled in strains in addition to S288C; for instance, the filamenting strain Sigma 1278B is amenable to genetic study in the laboratory (*e.g.*, LORENZ and HEITMAN 1998).

Finally, it may be time to reconsider the argument that gene functions need not manifest as cut-and-dried laboratory single-mutant phenotypes to be selected over evolutionary time. Marginal but reproducible fitness contributions have previously been noted for many single mutants in nonessential yeast genes (THATCHER *et al.* 1998), and more recent efforts to compile synthetic genetic interactions have indicated that the full yeast genetic network may encompass as many as 200,000 pairwise interactions (DAVIERWALA *et al.* 2005). The effects of loss of tRNA modifications appear to be synthetic in at least some cases (ALEXANDROV *et al.* 2006), and these modifying enzymes seem to be among the more refractory genes with respect to rapid characterization. Similar genetic behavior might be expected of other currently uncharacterized genes.

At the time of the original sequencing effort, only an estimated 1000 of 6000 apparent genes had been previously described in a research article (GOFFEAU *et al.*

1996); a decade later, the running total is 4687, counting only genes that appear in focused articles (those that mention 10 or fewer yeast genes) (FISK *et al.* 2006). By any standard, this is a phenomenal accomplishment, giving us confidence that the full gene complement can be characterized and optimism that similar accomplishments will be made during our lifetimes in other organisms, perhaps even mammals. Current technology appears to be sufficient to characterize most gene products at a cellular level, provided they are active and nonredundant under the conditions examined. It is possible to generate hypotheses regarding functions and associate them with confidence levels, but someone still has to follow them up. There is clearly a need for human inference and domain knowledge in the creation of new approaches for specific problems, including the characterization of individual genes and their roles in nature.

We are grateful to Mike Cherry, Dianna Fisk, Owen Ryan, Charlie Boone, Brenda Andrews, Guri Giaever, Allan Spradling, Mark Johnston, and Linda Bisson for helpful discussions and feedback on the manuscript.

LITERATURE CITED

- ABRAMOVA, N., O. SERTIL, S. MEHTA and C. V. LOWRY, 2001 Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* **183**: 2881–2887.
- ALEXANDROV, A., I. CHERNYAKOV, W. GU, S. L. HILEY, T. R. HUGHES *et al.*, 2006 Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* **21**: 87–96.
- ASHBURNER, M., C. A. BALL, J. A. BLAKE, D. BOTSTEIN, H. BUTLER *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**: 25–29.
- BADER, G. D., A. HEILBUT, B. ANDREWS, M. TYERS, T. HUGHES *et al.*, 2003 Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends Cell Biol.* **13**: 344–356.
- BLANDIN, G., P. DURRENS, F. TEKAIA, M. AIGLE, M. BOLOTIN-FUKUHARA *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett.* **487**: 31–36.
- BRACHAT, S., F. S. DIETRICH, S. VOEGELI, Z. ZHANG, L. STUART *et al.*, 2003 Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol.* **4**: R45.
- BUTLER, Y. X., Y. ABHAYAWARDHANE and G. C. STEWART, 1993 Amplification of the *Bacillus subtilis* maf gene results in arrested septum formation. *J. Bacteriol.* **175**: 3139–3145.
- CLIFTON, P., P. SUDARSANAM, A. DESIKAN, L. FULTON, B. FULTON *et al.*, 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**: 71–76.
- COLLINS, S. R., K. M. MILLER, N. L. MAAS, A. ROGUEV, J. FILLINGHAM *et al.*, 2007 Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**: 806–810.
- DAVIERWALA, A. P., J. HAYNES, Z. LI, R. L. BROST, M. D. ROBINSON *et al.*, 2005 The synthetic genetic interaction spectrum of essential genes. *Nat. Genet.* **37**: 1147–1152.
- DRISCOLL, R., A. HUDSON and S. P. JACKSON, 2007 Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. *Science* **315**: 649–652.
- FINN, R. D., J. MISTRY, B. SCHUSTER-BOCKLER, S. GRIFFITHS-JONES, V. HOLLICH *et al.*, 2006 Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**: D247–D251.
- FISK, D. G., C. A. BALL, K. DOLINSKI, S. R. ENGEL, E. L. HONG *et al.*, 2006 *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast* **23**: 857–865.

- FLEET, G. H., 2003 Yeast interactions and wine flavour. *Int. J. Food Microbiol.* **86**: 11–22.
- GAVIN, A. C., P. ALOY, P. GRANDI, R. KRAUSE, M. BOESCHE *et al.*, 2006 Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 631–636.
- GERBER, A. P., D. HERSCHLAG and P. O. BROWN, 2004 Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2**: E79.
- 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.
- GOFFEAU, A., B. G. BARRELL, H. BUSSEY, R. W. DAVIS, B. DUJON *et al.*, 1996 Life with 6000 genes. *Science* **274**: 546, 563–547.
- GRANDI, P., V. RYBIN, J. BASSLER, E. PETFALSKI, D. STRAUSS *et al.*, 2002 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* **10**: 105–115.
- GRIFFIN, J. L., 2006 The Cinderella story of metabolic profiling: Does metabolomics get to go to the functional genomics ball? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**: 147–161.
- HAN, J., H. ZHOU, B. HORAZDOVSKY, K. ZHANG, R. M. XU *et al.*, 2007 Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. *Science* **315**: 653–655.
- HUGHES, T. R., M. D. ROBINSON, N. MITSAKAKIS and M. JOHNSTON, 2004 The promise of functional genomics: completing the encyclopedia of a cell. *Curr. Opin. Microbiol.* **7**: 546–554.
- 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.
- KANIAK, A., Z. XUE, D. MACOOL, J. H. KIM and M. JOHNSTON, 2004 Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **3**: 221–231.
- KASTENMAYER, J. P., L. NI, A. CHU, L. E. KITCHEN, W. C. AU *et al.*, 2006 Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* **16**: 365–373.
- KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- KESSLER, M. M., Q. ZENG, S. HOGAN, R. COOK, A. J. MORALES *et al.*, 2003 Systematic discovery of new genes in the *Saccharomyces cerevisiae* genome. *Genome Res.* **13**: 264–271.
- 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.
- KRON, S. J., 1997 Filamentous growth in budding yeast. *Trends Microbiol.* **5**: 450–454.
- KUMAR, A., P. M. HARRISON, K. H. CHEUNG, N. LAN, N. ECHOLS *et al.*, 2002 An integrated approach for finding overlooked genes in yeast. *Nat. Biotechnol.* **20**: 58–63.
- LORENZ, M. C., and J. HEITMAN, 1998 Regulators of pseudohyphal differentiation in *Saccharomyces cerevisiae* identified through multicopy suppressor analysis in ammonium permease mutant strains. *Genetics* **150**: 1443–1457.
- MACPHERSON, S., M. LAROCHELLE and B. TURCOTTE, 2006 A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiol. Mol. Biol. Rev.* **70**: 583–604.
- MARTZEN, M. R., S. M. MCCRAITH, S. L. SPINELLI, F. M. TORRES, S. FIELDS *et al.*, 1999 A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**: 1153–1155.
- MENNE, T. F., B. GOYENECHEA, N. SANCHEZ-PUIG, C. C. WONG, L. M. TONKIN *et al.*, 2007 The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat. Genet.* **11**: 11.
- MIURA, F., N. KAWAGUCHI, J. SESE, A. TOYODA, M. HATTORI *et al.*, 2006 A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl. Acad. Sci. USA* **103**: 17846–17851.
- MORTIMER, R., and M. POLSINELLI, 1999 On the origins of wine yeast. *Res. Microbiol.* **150**: 199–204.
- MUSTACCHI, R., S. HOHMANN and J. NIELSEN, 2006 Yeast systems biology to unravel the network of life. *Yeast* **23**: 227–238.
- MYERS, C. L., D. R. BARRETT, M. A. HIBBS, C. HUTTENHOWER and O. G. TROYANSKAYA, 2006 Finding function: evaluation methods for functional genomic data. *BMC Genomics* **7**: 187.
- NASH, R., S. WENG, B. HITZ, R. BALAKRISHNAN, K. R. CHRISTIE *et al.*, 2007 Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.* **35**: D468–D471.
- NISHIDA, H., 2006 Detection and characterization of fungal-specific proteins in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* **70**: 2646–2652.
- OSHIRO, G., L. M. WODICKA, M. P. WASHBURN, J. R. YATES, III, D. J. LOCKHART *et al.*, 2002 Parallel identification of new genes in *Saccharomyces cerevisiae*. *Genome Res.* **12**: 1210–1220.
- PARSONS, A. B., A. LOPEZ, I. E. GIVONI, D. E. WILLIAMS, C. A. GRAY *et al.*, 2006 Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* **126**: 611–625.
- PENG, W. T., M. D. ROBINSON, S. MNAIMNEH, N. J. KROGAN, G. CAGNEY *et al.*, 2003 A panoramic view of yeast noncoding RNA processing. *Cell* **113**: 919–933.
- PTACEK, J., G. DEVGAN, G. MICHAUD, H. ZHU, X. ZHU *et al.*, 2005 Global analysis of protein phosphorylation in yeast. *Nature* **438**: 679–684.
- ROBINSON, M. D., J. GRIGULL, N. MOHAMMAD and T. R. HUGHES, 2002 FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics* **3**: 35.
- SARTORI, G., L. ALDEGHERI, G. MAZZOTTA, G. LANFRANCHI, H. TOURNU *et al.*, 1999 Characterization of a new hemoprotein in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**: 5032–5037.
- SAVCHENKO, A., N. KROGAN, J. R. CORT, E. EVDOKIMOVA, J. M. LEW *et al.*, 2005 The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. *J. Biol. Chem.* **280**: 19213–19220.
- SCHNEIDER, J., P. BAJWA, F. C. JOHNSON, S. R. BHAUMIK and A. SHILATIFARD, 2006 Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* **281**: 37270–37274.
- SCHOLLES, D. T., M. BANERJEE, B. BOWEN and M. J. CURCIO, 2001 Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics* **159**: 1449–1465.
- STARK, C., B. J. BREITKREUTZ, T. REGULY, L. BOUCHER, A. BREITKREUTZ *et al.*, 2006 BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**: D535–D539.
- SUTER, B., D. AUERBACH and I. STAGLJAR, 2006 Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *BioTechniques* **40**: 625–644.
- THATCHER, J. W., J. M. SHAW and W. J. DICKINSON, 1998 Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. USA* **95**: 253–257.
- TONG, A. H., G. LESAGE, G. D. BADER, H. DING, H. XU *et al.*, 2004 Global mapping of the yeast genetic interaction network. *Science* **303**: 808–813.
- TSUBOTA, T., C. E. BERNDSEN, J. A. ERKMANN, C. L. SMITH, L. YANG *et al.*, 2007 Histone H3–K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Mol. Cell* **25**: 703–712.
- VELCULESCU, V. E., L. ZHANG, W. ZHOU, J. VOGELSTEIN, M. A. BASRAI *et al.*, 1997 Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- VON MERING, C., R. KRAUSE, B. SNEL, M. CORNELL, S. G. OLIVER *et al.*, 2002 Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**: 399–403.
- WILKINSON, M. L., S. M. CRARY, J. E. JACKMAN, E. J. GRAYHACK and E. M. PHIZICKY, 2007 The 2'-O-methyltransferase responsible for modification of yeast tRNA at position 4. *RNA* **13**: 404–413.
- WONG, S. L., L. V. ZHANG and F. P. ROTH, 2005 Discovering functional relationships: biochemistry versus genetics. *Trends Genet.* **21**: 424–427.
- ZHU, H., M. BILGIN, R. BANGHAM, D. HALL, A. CASAMAYOR *et al.*, 2001 Global analysis of protein activities using proteome chips. *Science* **293**: 2101–2105.