LETTER

The Complete Set of Predicted Genes from *Saccharomyces cerevisiae* in a Readily Usable Form

James R. Hudson, Jr.,[1] Elliott P. Dawson,[2] Kimberly L. Rushing,[1] Cynthia H. Jackson,[1] Daniel Lockshon,[3] Diana Conover,[3] Christian Lanciault,[3] James R. Harris,[2] Steven J. Simmons,[2] Rodney Rothstein,[4] and Stanley Fields[3,5]

[1]Research Genetics Inc., Huntsville, Alabama 35801; [2]BioVentures, Inc., Mufreesboro, Tennessee 37129; [3]Departments of Genetics and Medicine, University of Washington, Seattle, Washington 98195; [4]Department of Genetics and Development, Columbia University, New York, New York 10032

Nearly all of the open reading frames (ORFs) of the yeast *Saccharomyces cerevisiae* have been synthesized by PCR using a set of ~6000 primer pairs. Each of the forward primers has a common 22-base sequence at its 5′ end, and each of the back primers has a common 20-base sequence at its 5′ end. These common termini allow reamplification of the entire set of original PCR products using a single pair of longer primers—in our case, 70 bases. The resulting 70-base elements that flank each ORF can be used for rapid and efficient cloning into a linearized yeast vector that contains these same elements at its termini. This cloning by genetic recombination obviates the need for ligations or bacterial manipulations and should permit convenient global approaches to gene function that require the assay of each putative yeast gene.

Knowledge of the complete genome sequence of the yeast *Saccharomyces cerevesiae* is enabling global approaches for the analysis of gene function (see e.g., Oliver 1996; Johnston 1996). In particular, it permits experiments in which each gene is systematically analyzed for activity in contrast to those that rely on random screens. Such comprehensive efforts require efficient strategies to handle the ~6000 open reading frames (ORFs) predicted from the sequence. We describe a simple PCR-based approach that has generated a nearly complete set of the yeast genes in a form that allows multiple uses, including the construction of DNA arrays, epitope-tagged or hybrid proteins, and regulated versions of the genes. As an example, we demonstrate how these reagents are being used to produce fusions of the Gal4p activation domain to each yeast protein for large-scale two-hybrid analysis (see Finley and Brent 1994; Bartel et al. 1996).

RESULTS AND DISCUSSION

The strategy outlined in Figure 1 uses a set of ~6000 PCR primer pairs to amplify individually each of the yeast ORFs. The primers were designed from a list of the first 50 bases (corresponding to the predicted ATG and succeeding 3′ sequences) and last 50 bases (corresponding to the sequences ending with and including the predicted termination codon) for 6102 ORFs in the Saccharomyces Genome Database, kindly provided by Michael Cherry (Stanford University, Palo Alto, CA). Each forward primer contains both a unique sequence that allows priming at the start of one of the yeast ORFs and a 22-nucleotide-long sequence at its 5′ end, which is shared by all forward primers (see Fig. 1). The unique sequence begins with the codon after the initiator ATG and is followed by an additional 17–29 bases of ORF sequence to allow annealing of each primer to its target sequence with a uniform T_m of 68°C–72°C. Each back primer also contains both a unique sequence, which allows priming at the terminus of an ORF, and a 20-nucleotide-long tail at its 5′ end, which is shared by all the back primers. The unique sequence of the back primer corresponds to the reverse complement of the termination codon followed by 17–29 bases that are the reverse complement of the last ~6–10 codons of the ORF.

The ~20-base tails serve as priming sites for a second round of PCR (Fig. 1; rePCR), which allows reamplification of the entire set of genes with a

[5]Corresponding author.
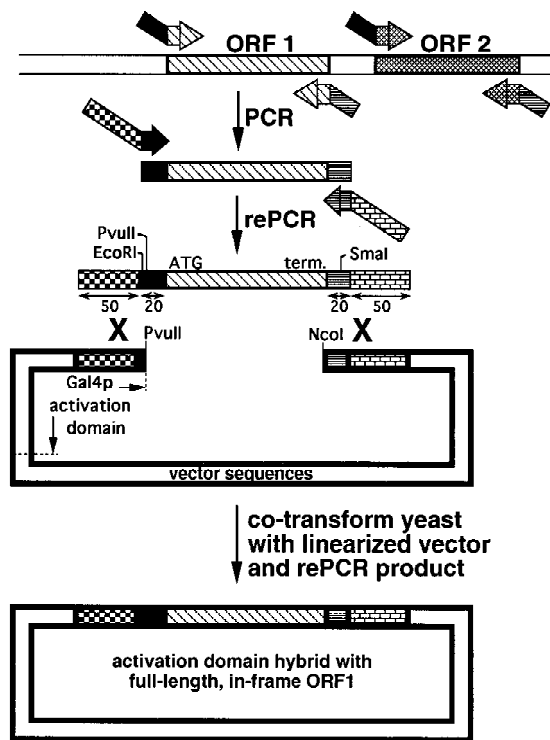E-MAIL fields@genetics.washington.edu; FAX (206) 543-0754.

Figure 1 Strategy for cloning each of the yeast ORFs. A segment of the yeast genome containing illustrative ORFs 1 and 2, both of which would be transcribed left to right, is shown at the *top.* Each ORF is flanked by two bent arrows representing the unique pairs of primers, with the solid black fill (*rightward*-pointing arrows) indicating the common 5′ termini of the forward primers and the horizontal-lined fill (*leftward* pointing arrows) indicating the common 5′ termini of the back primers. The sequence of each forward primer is 5′-GGAATTCCA-GCTGACCACCATGN$_{17-29}$-3′ corresponding to 19 bases of non-yeast sequence, the initiator ATG of the ORF, and 17–29 subsequent bases of the ORF. The sequence of each back primer is 5′-GATCCCCGGGAATTGCCAT-GENDN$_{17-29}$-3′ corresponding to 20 bases of non-yeast sequence, followed by the reverse complement of a stop codon (noted above as ''END''), followed by the reverse complement of 17–29 bases found at the end of the ORF. The product of the PCR of ORF1, shown as a box with diagonal-lined fill flanked by the 19 and 20 bp of non-yeast sequences, is template for the rePCR. The 70-base sequences of the rePCR primers, shown as bent arrows with checkered and ''brick'' 5′ termini, and 3′ termini matching the sequences in the first set of PCR primers, are provided in Methods. The product of the rePCR is ORF1 flanked by 70-base elements that are identical to those in the two-hybrid vector pOAD. The positions of the translation initiation codon (ATG) and termination codon (term.) of the ORF1 rePCR product are shown. Digestion of the vector with *Nco*I and *Pvu*II and cotransformation with the rePCR product results in two recombination events that precisely insert the ORF in-frame with the activation domain of Gal4p.

single pair of primers that contain the common sequences at their 3′ ends. Reamplification products are thus the ORFs flanked by a common pair of longer sequences corresponding to the rePCR primers. These longer flanking regions are designed to be homologous to sequences in a yeast vector that surround a unique *Nco*I restriction site. Cotransformation of a rePCR product with linearized vector exploits the efficient homologous recombination machinery of yeast to insert the reamplification product into the vector with precise fusion joints at either end (Ma et al. 1987; Smith et al. 1995; Oldenburg et al. 1997). The common tails of the forward and back primers of the first PCRs (described in the legend to Fig. 1) were designed to meet the following criteria: (1) no significant match to the yeast genome; (2) a Kozak consensus sequence CCACC immediately 5′ to the initiator ATG in the forward primer; (3) restriction sites *Eco*RI and *Pvu*II in the forward primer; (4) restriction site *Sma*I in the back primer; (5) sequences in the forward and back primers that correspond to 5 of 6 bases of an *Nco*I site present in the vector; and (6) no stop codons or other ATG in the forward primer.

PCR was carried out on yeast genomic DNA of a version of strain S288C, the predominant strain used in the Genomic Sequencing Project. With the initial PCR conditions, 99.1% (6046 ORFs) of the primer pairs yielded a band of the predicted size [including intron sequence, which is found in ~4% of yeast genes (Dujon 1996)]. An additional 14 genes yielded a band of an incorrect size, and 56 genes yielded no discrete product or a doublet, both bands of which were not the predicted size. The rePCRs of the ORFs were primed by a single pair of 70-nucleotide-long primers (see Methods) that result in flanking sequences for recombination into a two-hybrid Gal4p activation domain vector carrying the *LEU2* gene for selection. This vector, pOAD, was constructed to carry a unique *Nco*I site flanked by the two 70-bp sequences corresponding to the reamplification primers. This second round of PCR yielded products with an efficiency of ~99% (Fig. 2). *Alu*I digestion of a sample of 16 rePCR products produced the restriction pattern predicted by the genome sequence (data not shown). Both the initial PCR products and the products derived from the reamplification showed a low level of DNA fragments that migrated at approximately twice the size of the predicted bands; these extra fragments appear to be attributable to annealing to each of the ~20-nucleotide long priming sites that flank the ORFs to generate dimers of the ORFs. In addition, a small fraction (~8%) of the rePCR products contained an
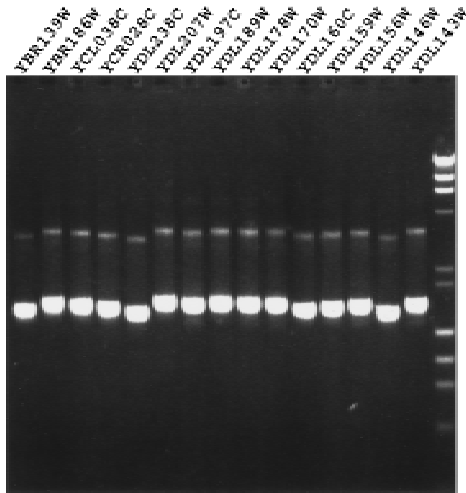
Figure 2 Products of a sample of rePCRs. Fifteen ORFs were reamplified using the 70-base activation domain rePCR primers. The sizes of these ORFs from the genome sequence are YBR139W (1527 bp), YBR186w (1611 bp), YCL038c (1587 bp), YCR028c (1539 bp), YDL238c (1470 bp), YDL207w (1617 bp), YDL197c (1578 bp), YDL189w (1612 bp), YDL178w (1593 bp), YDL170w (1587 bp), YDL160c (1521 bp), YDL159w (1548 bp), YDL156w (1569 bp), YDL146w (1476 bp), and YDL143w (1587 bp). The DNA size marker in the *right* lane has fragments of 23, 9.4, 6.6, 4.4, 2.3, 2.0, 1.4, 1.1, 0.9, and 0.6 kb. The rePCR products all migrate slightly slower than the predicted sizes of their ORFs because of the 70 bp of extra sequence at each end.

~140 bp band that appears to consist of a dimer of the 70-nucleotide-long rePCR primers.

We cotransformed into yeast the linearized vector and a pilot set of 11 of the reamplified ORFs that ranged in size from 0.86 to 2.6 kb, chosen to include six rePCR products with undetectable amounts of the presumed primer dimer and five with variable amounts of the presumed dimer to assay the efficiency of transformation and the effect of the ~140-bp fragment. These transformations were carried out using a 96-well protocol and yielded ~50-fold more Leu+ transformants than when vector alone was added. To assess whether the transformants contained the appropriate inserts, we analyzed plasmids present in the transformants by restriction digestion. When the rePCR products containing no detectable primer dimer were used, all individual yeast colonies (23 of 23) derived from these six transformation plates bore plasmid that contained an insert of the correct size. The other five rePCR products, containing the presumed primer dimers, yielded only 5 of 18 colonies with plasmid that con-

tained inserts of ORF size. The remainder likely contained a primer dimer insert. We conclude from this experiment that the overall protocol of two rounds of PCR followed by recombination-mediated transformation is an effective strategy for cloning large numbers of different inserts into a yeast plasmid, although the efficiency of obtaining the correct insert is dependent on the quality of the rePCR product.

Another concern, given the extensive use of PCR in generating the cloned genes in this protocol, is the possibility of mutations introduced by PCR. To assess the percentage of transformants that express a functional protein, we carried out a two-hybrid experiment using a known pair of proteins, Rad17p and Mec3p, which interact in this assay (B. Drees and S. Fields, unpubl.). The linearized pOAD vector was cotransformed along with the rePCR product of the *RAD17* gene into yeast strain CBY14**a** (Bendixen et al. 1994) already containing pOBD–Mec3, a *TRP1*-containing plasmid that expresses Mec3p fused to the Gal4p DNA-binding domain. Of 101 Trp+ Leu+ transformants that contain both plasmids tested, **88 (87%)** produced a blue color in a filter assay for β-galactosidase, indicating expression of the *GAL1–lacZ* two-hybrid reporter gene. Thus, while PCR errors, as well as cases in which the *RAD17* gene failed to recombine into the vector, may be responsible for those transformants that did not yield a two-hybrid signal, the overwhelming majority of the rePCR products produced a protein that was active in the two-hybrid assay. For some genomic analyses, pools of yeast transformants can be used to avoid the problem of dealing with single transformants that may contain a product with a PCR error. For example, global two-hybrid analysis can be carried out by mating a pool of transformants that carry an activation domain plasmid with one carrying a DNA-binding domain plasmid.

The approach presented here allowed the synthesis of the nearly complete set of yeast genes as discrete PCR products that exactly correspond to the predicted ORFs. This procedure required that the ~6000 primer pairs be generated once and similarly that the ~6000 PCRs on genomic DNA needed to be performed once. However, now that this set of ORFs has been generated, one can use these PCR fragments for other applications by designing single pairs of primers for the reamplification step and corresponding yeast vectors if appropriate. For example, reamplification using primers that contain only the ~20-nucleotide-long tails permits abundant synthesis of every gene for creating DNA arrays to monitor expression (see, e.g., DeRisi et al. 1996).

Reamplification with primers carrying sequences for transcription by T7 RNA polymerase can generate the set of genes for in vitro transcription and translation. A variety of vectors can be constructed that contain sequences flanking a unique restriction site that match sequences in a set of primers used for reamplification. These vectors can be used for such purposes as the inducible expression of genes (e.g., from the *GAL1* promoter) or the generation of amino-terminal fusion proteins containing epitope tags, protein localization sequences, or specific protein domains. Although carboxy-terminal protein fusions are not easily constructed because of the inclusion of the natural stop codons in the set of back primers used in the initial amplifications, such fusions may be possible using a set of three back primers for reamplification containing single mismatches that alter the terminators. Alternatively, a new set of back primers can be synthesized that do not include the terminators and that permit such fusions.

The complete set of yeast PCR products, or yeast transformants containing them cloned into a vector of choice, can be arrayed in merely 16 microtiter dishes of 384 wells each. The highly efficient steps of reamplification and yeast transformation by recombination should enable this set of products, or any subset of genes, to be used for additional genomic strategies. In particular, the absence of any ligations or bacterial protocols prior to analyzing a set of constructions in yeast should make this a convenient means of genomic analysis. Finally, it may be feasible to adopt similar strategies for other genomes, including those more complex than yeast, as their sequences become available.

METHODS

Primer Synthesis

Primer pairs for each of the 6102 annotated ORFs in the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces/), provided by Michael Cherry (Stanford University, Palo Alto, CA), were synthesized at a 0.4 μM scale using standard phosphoramidite chemistry. The primers and PCR products for all of the yeast genes are available from Research Genetics, Inc.

The rePCR primers for the product compatible with the activation domain vector pOAD were synthesized with the following sequences. Forward primer, 5′-CTATCTATTCGAT-GATGAAGATACCCCACCAAACCCAAAAAAAGAG-ATCGAATTCCAGCTGACCACCATG-3′; back primer, 5′-CTTGCGGGGGTTTTTCAGTATCTACGATTCATAG-ATCTCTGCAGGTCGACGGATCCCCGGGAATTGCCATG-3′. For the rePCR primers for the product compatible with the DNA-binding domain vector pOBD, the forward primer was

5′-ATCGGAAGAGAGTAGTAACAAAGGTCAAAG-ACAGTTGACTGTATCGCCGGAATTCCAGCTGAC-CACCATG-3′ and the back primer was 5′-TCATAAATCATA-AGAAATTCGCCCGGAATTAGCTTGGCTGCAGGTC-GACGGATCCCCGGGAATTGCCATG-3′.

PCR

Template DNA was isolated from strain S288Ctrp1° (kindly provided by Maynard Olson, University of Washington, Seattle), which differs from S288C (Mortimer and Johnston 1986) in having a spontaneous *trp1* allele and lacking mitochondrial DNA, as described (Polaina and Adam 1991). Individual PCRs were performed on 0.2-ml reactions containing 30 ng of genomic DNA template, 20 pmoles each primer, 5 units of *Taq* polymerase (Perkin Elmer), 0.02 units of *Pfu* polymerase (Stratagene), and 1.5 mM Mg^{2+}. Amplifications were performed using a ''hot start'' method and a Tetrad Model PTC 225 Thermocycler from MJ Research. The first PCR cycle was 95°C for 3 min, followed by 36 cycles of 50°C for 45 sec, 72°C for 210 sec, 95°C for 60 sec, and a final cycle of 72°C for 8 min. The first attempt of the 6102 PCRs produced 5868 products of the correct size. 234 primer pairs were resynthesized when the initial PCR failed or produced unexpected results. In addition, four other conditions were tried on the failures, which involved increasing the extension time and/or raising or lowering the annealing temperature. Of the failures, 178 yielded correct product and 56 failed or gave the same but unexpected results.

Reaction conditions for rePCRs were 5 ng of template, 2.5 pmoles each primer, 1.5 M Mg^{2+}, 0.6 units of *Taq* polymerase and 0.003 units of *Pfu* polymerase in 25 μl. Reactions were carried out at 94°C for 45 sec, then cycled through 35, two-temperature cycles of 94°C (15 sec.) and 72°C (2–12 min, depending on ORF size). A portion (2 μl) of each reaction was analyzed on 0.7%–1.4% agarose gels.

Yeast Transformation and Two-Hybrid Assay

ORFs were cloned into the vector pOAD (ORFs fusable to Gal4p activation domain), which was constructed from pGAD424.C (Bartel et al. 1996) by ligation of a pair of 22-base oligonucleotides into the unique *Eco*RI site to recreate the *Eco*RI site at the end of this insert closer to the Gal4p activation domain-coding region. The sequences of these oligonucleotides are 5′-AATTCCAGCTGACCACCATGGC-3′ and 5′-AATTGCCATGGTGGTCAGCTGG-3′. pOAD was prepared for recombination-mediated cloning of the ORFs by digestion with *Nco*I and *Pvu*II to remove 8 bp of sequence, which reduced the background number of transformants significantly relative to that from *Nco*I digestion alone. pOBD was constructed from pGBT9.C (Bartel et al. 1996) by ligation of the same pair of 22-base oligonucleotides. pOBD–Mec3 was constructed by cotransformation of *Pvu*II- and *Nco*I-digested pOBD with the rePCR product of the *MEC3* gene derived from the 70-base primers corresponding to the pOBD vector.

Strain PJ69-4A, kindly provided by Philip James (James et al. 1996), was transformed to Leu+ by incubation with 2 ng of *Pvu*II–*Nco*I-cut pOAD and 10 μl of rePCR product in 0.25 ml of 35% PEG (Sigma P3640), 93 mM lithium acetate, 8 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% DMSO, 0.17 mg/ml of salmon sperm DNA in a 96-well microtiter dish for 30 min at room temperature, then for 30 min at 42°C. Yeast were pelleted, the

supernatant was aspirated and replaced with water, and the resuspended yeast were plated on 35-mm-diam. culture plates of yeast synthetic media lacking leucine. To analyze plasmid from transformants, we prepared total DNA from liquid cultures of each colony and transformed *Escherichia coli.* One ampicillin-resistant colony from each bacterial transformation was used for a miniplasmid preparation, which was digested with *Pvu*II and *Pst*I to release the inserts. The restriction products were analyzed on agarose gels.

For two-hybrid analysis, strain CBY14**a** (Bendixen et al. 1994) was first transformed with pOBD–Mec3, and then with linearized pOAD and the rePCR product of the *RAD17* gene. The filter assay for β-galactosidase was as described (Breeden and Nasmyth 1985).

ACKNOWLEDGMENTS

REFERENCES

Bartel, P.L., J.A. Roecklein, D. SenGupta, and S. Fields. 1996. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet.* **12:** 72–77.

Bendixen, C., S. Gangloff, and R. Rothstein. 1994. A yeast mating-selection scheme for detection of protein-protein interactions. *Nucleic Acids Res.* **22:** 1778–1779.

Breeden, L. and K. Nasmyth. 1985. Regulation of the yeast *HO* gene. *Cold Spring Harbor Symp. Quant. Biol.* **50:** 643–650.

DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14:** 457–460.

Dujon, B. 1996. The yeast genome project: What did we learn? *Trends Genet.* **12:** 263–270.

Finley, R.L.J. and R. Brent. 1994. Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci.* **91:** 12980–12984.

James, P., J. Halladay, and E.A. Craig. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144:** 1425–1436.

Johnston, M. 1996. Towards a complete understanding of how a simple eukaryotic cell works. *Trends Genet.* **12:** 242–243.

Ma, H., S. Kunes, P.J. Schatz, and D. Botstein. 1987. Plasmid construction by homologous recombination in yeast. *Gene* **58:** 201–216.

Mortimer, R.K. and J.R. Johnston. 1986. Geneology of principal strains of the Yeast Genetic Stock Center. *Genetics* **113:** 35–43.

Polaina, J. and A.C. Adam. 1991. A fast procedure for yeast DNA purification. *Nucleic Acids. Res.* **19:** 5443.

Oldenburg, K.R., K.T. Vo, S. Michaelis, and C. Paddon. 1997. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25:** 451–452.

Oliver, S. 1996. A network approach to the systematic analysis of yeast gene function. *Trends Genet.* **12:** 241–242.

Smith, J., H. Zou, and R. Rothstein. 1995. PCR-based gene disruption in *Saccharomyces cerevisiae. Methods Mol. Cell. Biol.* **5:** 270–277.