An Introduction to the Genetics and Molecular Biology of the Yeast *Saccharomyces cerevisiae*

FRED SHERMAN

Department of Biochemistry and Biophysics

University of Rochester Medical School, Rochester, NY 14642

• 1998 •

The yeast *Saccharomyces cerevisiae* is clearly the most ideal eukaryotic microorganism for biological studies. The "awesome power of yeast genetics" has become legendary and is the envy of those who work with higher eukaryotes. The complete sequence of its genome has proved to be extremely useful as a reference towards the sequences of human and other higher eukaryotic genes. Furthermore, the ease of genetic manipulation of yeast allows its use for conveniently analyzing and functionally dissecting gene products from other eukaryotes.

Key Words

Ascus,
  (plual asci) is a sac-like structure containing a tetrad of four spores (or ascospores).

Heterothallic,
  strains of yeast have cross-compatible mating types and are stable both as haploids and diploids.

Homothallic,
  strains of yeast give rise to tetrads containing four potentially self-fertile members, because the transient haploid cells switch their mating types, and thus have only a stable diplophase.

Plasmid shuffle,
  is a procedure for screening of mutations, derived from a mutagenized plasmid, requiring the loss of a second plasmid to assay for the recessive mutations.

Shuttle vectors,
  are vectors that can be propagated in both yeast and *E. coli.*

Tetrad,
  is the four products of meiosis.

Two-hybrid system,
  is a genetic assay used in yeast for detection of protein-protein interactions.

1 Yeast is a Model Eukaryote

This chapter deals only with the yeast *S. cerevisiae*, and related interbreeding species. The fission yeast *Schizosaccharomyces pombe*, which is only distantly related to *S. cerevisiae*, has equally important features, but is not as well characterized. The general principles of the numerous classical and modern approaches for investigating *S. cerevisiae* are described, and the explanation of terms and nomenclature used in current yeast studies are emphasized . This article should be particularly useful to the uninitiated who are exposed for the first time to experimental studies of yeast. Detailed protocols are described in the primary literature and in a number of reviews in the books listed in the Bibliography. The original citations for the material covered in this chapter also can be found in these comprehensive reviews.

Although yeasts have greater genetic complexity than bacteria, containing 3½ times more DNA than *Escherichia coli* cells, they share many of the technical advantages that permitted rapid progress in the molecular genetics of prokaryotes and their viruses. Some of the properties that make yeast particularly suitable for biological studies include rapid growth, dispersed cells, the ease of replica plating and mutant isolation, a well-defined genetic system, and most important, a highly versatile DNA transformation system. Unlike many other microorganisms, *S. cerevisiae* is viable with numerous markers. Being nonpathogenic, yeast can be handled with little precautions. Large quantities of normal bakers' yeast are commercially available and can provide a cheap source for biochemical studies.

Unlike most other microorganisms, strains of *S. cerevisiae* have both a stable haploid and diploid state. Thus, recessive mutations can be conveniently isolated and manifested in haploid strains, and complementation tests can be carried out in diploid strains. The development of DNA transformation has made yeast particularly accessible to gene cloning and genetic engineering techniques. Structural genes corresponding to virtually any genetic trait can be identified by complementation from plasmid libraries. Plasmids can be introduced into yeast cells either as replicating molecules or by integration into the genome. In contrast to most other organisms, integrative recombination of transforming DNA in yeast proceeds exclusively via homologous recombination. Exogenous DNA with at least partial homologous segments can therefore be directed at will to specific locations in the genome. Also, homologous recombination, coupled with yeasts' high levels of gene conversion, has led to the development of techniques for the direct replacement of genetically engineered DNA sequences into their normal chromosome locations. Thus, normal wild-type genes, even those having no previously known mutations, can be conveniently replaced with altered and disrupted alleles. The phenotypes arising after disruption of yeast genes has contributed significantly toward understanding of the function of certain proteins *in vivo*. Many investigators have been shocked to find viable mutants with little of no detrimental phenotypes after disrupting genes that were previously assumed to be essential. Also unique to yeast, transformation can be carried out directly with synthetic oligonucleotides, permitting the convenient productions of numerous altered forms of proteins. These techniques have been extensively exploited in the analysis of gene regulation, structure-function relationships of proteins, chromosome structure, and other general questions in cell biology. The overriding virtues of yeast are illustrated by the fact that mammalian genes are being introduced into yeast for systematic analyses of the functions of the corresponding gene products.

In addition, yeast has proved to be valuable for studies of other organisms, including the use of the two-hybrid screening system for the general detection of protein-protein interactions, the

use of YACs for cloning large fragments of DNA, and expression systems for the laboratory and commercial preparation of heterologous proteins. Many of these techniques are described herein.

During the last two decades, an ever-increasing number of molecular biologists have taken up yeast as their primary research system, resulting in a virtually autocatalytic stimulus for continuing investigations of all aspects of molecular and cell biology. Most significantly, a knowledge of the DNA sequence of the complete genome, which was completed in 1996, has altered the way molecular and cell biologist approach and carry out their studies. In addition, plans are under way to systematically investigate the possible functions of all yeast genes by examining the phenotypes of strains having disrupted genes.

2 Information on Yeast

A general introduction to a few selected topics on yeast can be found in the book chapters "Yeast as the *E. coli* of Eucaryotic Cells" and "Recombinant DNA at Work" (1). Comprehensive and excellent reviews of the genetics and molecular biology of *S. cerevisiae* are contained in three volumes entitled "Molecular Biology of the Yeast *Saccharomyces*" (2-4). An important source for methods used in genetics and molecular biology of yeast is contained in the book edited by Guthrie and Fink (5). Overviews of numerous subjects are also covered in other sources (6, 7, 8, 9), including protocols applicable to yeasts (10) and introductory material (11). A more comprehensive listing of earlier reviews can be found in Sherman (12). Interesting and amusing accounts of developments in the field are covered in *The Early Days of Yeast Genetics* (13). The journal *Yeast* publishes original research articles, reviews, short communications, sequencing reports, and selective lists of current articles on all aspects of *Saccharomyces* and other yeast genera.

Current and frequently-updated information and databases on yeast can be conveniently retrieved on the Internet through World Wide Web, including the "Saccharomyces Genomic Information Resource" (http://genome-www.stanford.edu/Saccharomyces/) and linked files containing DNA sequences, lists of genes, home pages of yeast workers, and other useful information concerning yeast.

3 Yeast Strains

Although genetic analyses and transformation can be performed with a number of taxonomically distinct varieties of yeast, extensive studies have been limited primarily to the many freely interbreeding species of the budding yeast *Saccharomyces* and to the fission yeast *Schizosaccharomyces pombe*. Although "*Saccharomyces cerevisiae*" is commonly used to designate many of the laboratory stocks of *Saccharomyces* used throughout the world, it should be pointed out that most of these strains originated from the interbred stocks of Winge, Lindegren, and others who employed fermentation markers not only from *S. cerevisiae* but also from *S. bayanus, S. carlsbergensis, S. chevalieri, S. chodati, S. diastaticus,* etc. Nevertheless, it is still recommended that the interbreeding laboratory stocks of *Saccharomyces* be denoted as *S. cerevisiae*, in order to conveniently distinguish them from the more distantly related species of *Saccharomyces*.

Care should be taken in choosing strains for genetic and biochemical studies. Unfortunately there are no truly wild-type *Saccharomyces* strains that are commonly employed in genetic studies. Also, most domesticated strains of brewers' yeast and probably many strains of bakers' yeast and true wild-type strains of *S. cerevisiae* are not genetically compatible with laboratory

stocks. It is usually not appreciated that many "normal" laboratory strains contain mutant characters, a fact not too surprising since they were derived from pedigrees involving mutagenized strains. The haploid strain S288C is often used as a normal standard because it gives rise to well-dispersed cells, it is widely used, and because many isogenic mutant derivatives are available. However, S288C contains mutations making it undesirable to use in mitochondrial studies. An other strain, D273-10B, has been extensively used as a typical normal yeast, especially for mitochondrial studies. (These laboratory strains should be denoted as "normal" or "standard", but not "wild-type".) One should examine the specific characters of interest before initiating a study with any strain.

Many strains containing characterized auxotrophic, temperature-sensitive, and other markers can be obtained from the Yeast Genetics Stock Culture Center (Department of Molecular and Cell Biology, 229 Stanley Hall, University of California, Berkeley, CA 94720-3206; (510) 642-0815; Fax (510) 642-8589; E-mail ygsc@violet.berkeley.edu). Other sources of yeast strains include: American Type Culture Collection (12301 Parklawn Drive, Rockville, MD 20852; (301) 881-2600; (800) 638-6597; E-mail request@atcc.org; http://www.atcc.org/); National Collection of Yeast Cultures (Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.); Centraalbureau voor Schimmelcultures (Yeast Division, Julianalaan 67a, 2628 BC Delft, Netherlands); Slovak Collection of Yeasts (Institute of Chemistry, Slovak Academy of Sciences, Dubravaska cesta, 809 33 Bratislava, Slovak Republic).

4 Growth and Life Cycles

Vegetative cell division of yeast characteristically occurs by budding, in which a daughter is initiated as an out growth from the mother cell, followed by nuclear division, cell-wall formation, and finally cell separation. The sizes of haploid and diploid cells vary with the phase of growth and from strain to strain. Typically, diploid cells are 5 x 6 μm ellipsoids and haploid cells are 4 μm diameter spheroids. The volumes and gross composition of yeast cells are listed in Table 1. During exponential growth, haploid cultures tend to have higher numbers of cells per cluster compared to diploid cultures. Also haploid cells have buds that appear adjacent to the previous one; whereas diploid cells have buds that appear at the opposite pole. Each mother cell usually forms no more than 20-30 buds, and it age can be determined by the number of bud scars left on the cell wall.

In addition, certain diploid strains of *S. cerevisiae* can assume a markedly different cell and colony morphology, denoted pseudohyphae, when grown on agar medium limiting for nitrogen sources. These pseudohyphal cells are significantly elongated, and mother-daughter pairs remain attached to each other. This characteristic pseudohyphal growth causes extended growth of

Table 1. Size and composition of yeast cells		
Characteristic	Haploid cell	Diploid cell
Volume (μm^3)	70	120
Composition (10^{-12} g)		
Wet weight	60	80
Dry weight	15	20
DNA	0.017	0.034
RNA	1.2	1.9
Protein	6	8

branched chains outward from the center of the colony, and invasive growth under the surface of agar medium.

"Normal" laboratory haploid strains have a doubling time of approximately 90 min. in complete YPD (1% yeast extract, 2% peptone, and 2% glucose) medium and approximately 140 min. in synthetic media during the exponential phase of growth at the optimum temperature of 30°C. However, strains with greatly reduced growth rates in synthetic media are often encountered. Usually strains reach a maximum density of 2×10^8 cells/ml in YPD medium. Titers 10 times this value can be achieved with special conditions, such as pH control, continuous additions of balanced nutrients, filtered-sterilized media and extreme aeration that can be delivered in fermenters.

S. cerevisiae can be stably maintained as either heterothallic or homothallic strains, as illustrated in Figure 1. Both heterothallic and homothallic diploid strains sporulate under conditions of nutrient deficiency, and especially in special media, such as potassium acetate medium. During sporulation, the diploid cell undergoes meiosis yielding four progeny haploid cells, which become encapsulated as spores (or ascospores) within a sac-like structure called an ascus (plural asci). The percent sporulation varies with the particular strain, ranging from no or little sporulation to nearly 100%. Many laboratory strains sporulate to over 50%. The majority of asci contains four haploid ascospores, although varying proportions asci with three or less spores are also observed.

Because the **a** and α mating types are under control of a pair of *MAT***a**/*MAT*α heterozygous alleles, each ascus contains two *MAT***a** and two *MAT*α haploid cells. Upon exposure to nutrient condition, the spores germinate, vegetative growth commences and mating of the *MAT***a** and *MAT*α can occur. However, if the haploid spores are mechanically separated by micromanipulation, the haplophase of heterothallic strains can be stably maintained, thus allowing the preparation of haploid strains. In contrast, the presence of the *HO* allele in
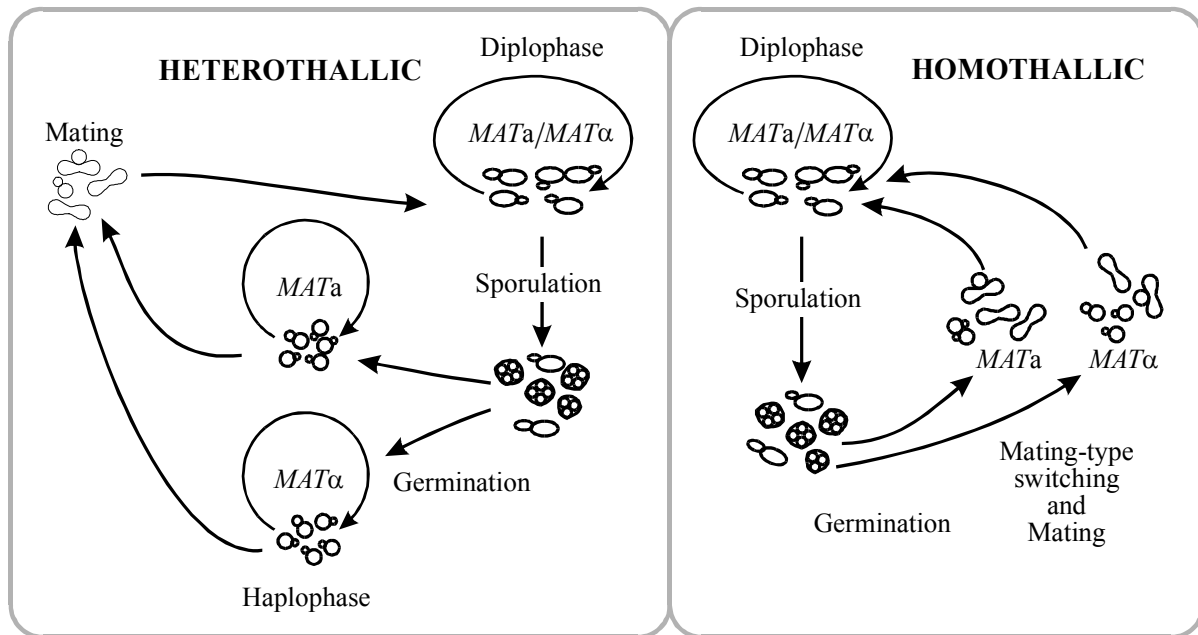

Figure 1. Life cycles of heterothallic and homothallic strains of *S. cerevisiae*. Heterothallic strains can be stably maintained as diploids and haploids, whereas homothallic strains are stable only as diploids, because the transient haploid cells switch their mating type, and mate.

homothallic strains causes switching of the mating type in growing haploid cells, such that *MAT**a*** cells produce *MAT*α buds and *MAT*α cells produce *MAT**a*** buds. As a consequence, mating occurs and there is only a transient haplophase in homothallic strains (Figure 1).

Controlled crosses of *MAT**a*** and *MAT*α haploid strains are simply carried out by mixing approximately equal amounts of each strain on a complete medium and incubating the mixture at 30°C for at least 6 hr. Prototrophic diploid colonies can then be selected on appropriate synthetic media if the haploid strains contain complementing auxotrophic markers. If the diploid strain cannot be selected, zygotes can be separated from the mating mixture with a micromanipulator. Zygotes are identified by a characteristic thick zygotic neck, and are best isolated 4 to 6 hr after incubating the mixture when the mating process has just been completed.

5 The Yeast Genome

S. cerevisiae contains a haploid set of 16 well-characterized chromosomes, ranging in size from 200 to 2,200 kb. The total sequence of chromosomal DNA, constituting 12,052 kb, was released in April, 1996. A total of 6,183 open-reading-frames (ORF) of over 100 amino acids long were reported, and approximately 5,800 of them were predicated to correspond to actual protein-coding genes. A larger number of ORFs were predicted by considering shorter proteins. In contrast to the genomes of multicellular organsims, the yeast genome is highly compact, with genes representing 72% of the total sequence. The average size of yeast genes is 1.45 kb, or 483 codons, with a range from 40 to 4,910 codons. A total of 3.8% of the ORF contain introns. Approximately 30% of the genes already have been characterized experimentally. Of the remaining 70% with unknown function, approximately one half either contain a motif of a characterized class of proteins or correspond to genes encoding proteins that are structurally related to functionally characterized gene products from yeast or from other organisms.

Ribosomal RNA is coded by approximately 120 copies of a single tandem array on chromosome XII. The DNA sequence revealed that yeast contains 262 tRNA genes, of which 80 have introns. In addition, chromosomes contain movable DNA elements, retrotransposons, that vary in number and position in different strains of *S. cerevisiae*, with most laboratory strains having approximately 30.

Other nucleic acid entities, presented in Figure 2, also can be considered part of the yeast genome. Mitochondrial DNA encodes components of the mitochondrial translational machinery and approximately 15% of the mitochondrial proteins. ρ^o mutants completely lack mitochondrial DNA and are deficient in the respiratory polypeptides synthesized on mitochondrial ribosomes, i.e., cytochrome *b* and subunits of cytochrome oxidase and ATPase complexes. Even though ρ^o mutants are respiratory deficient, they are viable and still retain mitochondria, although morphologically abnormal.

The 2-μm circle plasmids, present in most strains of *S. cerevisiae*, apparently function solely for their own replication. Generally cir^o strains, which lack 2-μm DNA, have no observable phenotype. However, a certain chromosomal mutation, *nib1*, causes a reduction in growth of cir^+ strains, due to an abnormally high copy number 2-μm DNA.

Similarly, almost all *S. cerevisiae* strains contain dsRNA viruses, that constitutes approximately 0.1% of total nucleic acid. RNA viruses include three families with dsRNA genomes, L-A, L-BC, and M. Two other families of dsRNA, T and W, replicate in yeast but so far have not been shown to be viral. M dsRNA encodes a toxin, and L-A encodes the major coat protein and components required for the viral replication and maintenance of M. The two

Inheritance	├── Mendelian ──┤	├──────────────── Non-Mendelian ──────────────┤				
Nucleic acid	├───── Double-stranded DNA ─────┤		├──── Double stranded RNA ───┤			
Location	├──── Nucleus ────┤		├─────────── Cytoplasm ───────────┤			

Genetic determinant	Chromosomes	2-μm plasmid	Mitochondrial DNA	RNA Viruses				
				L-A	M	L-BC	T	W
Relative amount	85%	5%	10%	80%	10%	9%	0.5%	0.5%
Number of copies	2 sets of 16	60-100	~50 (8-130)	103	170	150	10	10
Size (kb)	13,500 (200-2,200)	6.318	70-76	4.576	1.8	4.6	2.7	2.25
Deficiencies in mutants	All kinds	None	Cytochromes $a \cdot a_3$ & b	Killer toxin		None		
Wild-type	$YFG1^+$	cir$^+$	ρ^+	KIL-k$_1$				
Mutant or variant	$yfg1\text{-}1$	cir$^\circ$	ρ^-	KIL-o				

Figure 2. The genome of a diploid cell of *S. cerevisiae* (see the text). A wild-type chromosomal gene is depicted as $YFG1^+$ (Your Favorite Gene) and the mutation as *yfg1-1*.

dsRNA, M and L-A, are packaged separately with the common capsid protein encoded by L-A, resulting in virus-like particles that are transmitted cytoplasmically during vegetative growth and conjugation. L-B and L-C (collectively denoted L-BC), similar to L-A, have a RNA-dependent RNA polymerase and are present in intracellular particles. *KIL*-o mutants, lacking M dsRNA and consequently the killer toxin, are readily induced by growth at elevated temperatures, and chemical and physical agents.

Yeast also contains a 20S circular single-stranded RNA (not shown in Figure 2) that appears to encode an RNA-dependent RNA polymerase, that acts as an independent replicon, and that is inherited as a non-Mendelian genetic element.

Only mutations of chromosomal genes exhibit Mendelian 2:2 segregation in tetrads after sporulation of heterozygous diploids; this property is dependent on the disjunction of chromosomal centromeres. In contrast, non-Mendelian inheritance is observed for the phenotypes associated with the absence or alteration of other nucleic acids described in Figure 1.

6 Genetic Nomenclature

6.1 Chromosomal Genes

The genetic nomenclature for chromosomal genes of the yeast *S. cerevisiae* is now more-or-less universally accepted, as illustrated in Table 2, using *ARG2* as an example. Whenever possible, each gene, allele, or locus is designated by three italicized letters, e.g., *ARG*, which is usually a describer, followed by a number, e.g., *ARG2*. Unlike most other systems of genetic nomenclature, dominant alleles are denoted by using uppercase italics for all letters of the gene symbol, e.g., *ARG2*, whereas lowercase letters denote the recessive allele, e.g., the auxotrophic marker *arg2*. Wild-type genes are designated with a superscript "plus" (*sup6*$^+$ or *ARG2*$^+$). Alleles are designated by a number separated from the locus number by a hyphen, e.g., *arg2-9*. The symbol Δ can denote complete or partial deletions, e.g., *arg2-Δ1*. Insertion of genes follow the bacterial nomenclature by using the symbol :: . For example, *arg2::LEU2* denotes the insertion of the *LEU2* gene at the *ARG2* locus, in which *LEU2* is dominant (and functional), and *arg2* is recessive (and defective).

Table 2. Genetic nomenclature, using *ARG2* as an example

Gene symbol	Definition
ARG^+	All wild-type alleles controlling arginine requirement
ARG2	A locus or dominant allele
arg2	A locus or recessive allele confering an arginine requirement
$arg2^-$	Any *arg2* allele confering an arginine requirement
$ARG2^+$	The wild-type allele
arg2-9	A specific allele or mutation
Arg^+	A strain not requiring arginine
Arg^-	A strain requiring arginine
Arg2p	The protein encoded by *ARG2*
Arg2 protein	The protein encoded by *ARG2*
ARG2 mRNA	The mRNA transcribed from *ARG2*
arg2-Δ1	A specific complete or partial deletion of *ARG2*
ARG2::LEU2	Insertion of the functional *LEU2* gene at the *ARG2* locus, and *ARG2* remains functional and dominant
arg2::LEU2	Insertion of the functional *LEU2* gene at the *ARG2* locus, and *arg2* is or became nonfunctional
arg2-10::LEU2	Insertion of the functional *LEU2* gene at the *ARG2* locus, and the specified *arg2-10* allele which is nonfunctional
cyc1-arg2	A fusion between the *CYC1* and *ARG2* genes, where both are nonfunctional
P_{CYC1}-*ARG2*	A fusion between the *CYC1* promoter and *ARG2*, where the *ARG2* gene is functional

Phenotypes are sometimes denoted by cognate symbols in roman type and by the superscripts + and -. For example, the independence and requirement for arginine can be denoted by Arg^+ and Arg^-, respectively. Proteins encoded by *ARG2*, for example, can be denoted Arg2p, or simply Arg2 protein. However, gene symbols are generally used as adjectives for other nouns, for example, *ARG2* mRNA, *ARG2* strains, etc.

Although most alleles can be unambiguously assigned as dominant or recessive by examining the phenotype of the heterozygous diploid crosses, dominant and recessive traits are defined only with pairs, and a single allele can be both dominant and recessive. For example, because the alleles $CYC1^+$, *cyc1-717* and *cyc1-Δ1* produce, respectively, 100%, 5% and 0% of the gene product, the *cyc1-717* allele can be considered recessive in the *cyc1-717*/$CYC1^+$ cross and dominant in the *CYC1-717*/*cyc1-Δ1* cross. Thus, sometimes it is less confusing to denote all mutant alleles in lower case letters, especially when considering a series of mutations having a range of activities.

Although superscript letters should be avoided, it is sometimes expedient to distinguish genes conferring resistance and sensitivity by superscript R and S, respectively. For example, the genes controlling resistance to canavanine sulphate (*can1*) and copper sulphate (*CUP1*) and their sensitive alleles could be denoted, respectively, as *can*^{R}*1*, *CUP*^{R}*1*, *CAN*^{S}*1*, and *cup*^{S}*1*.

Wild-type and mutant alleles of the mating-type locus and related loci do not follow the standard rules. The two wild-type alleles of the mating-type locus are designated *MAT***a** and *MAT*α. The wild-type homothallic alleles at the *HMR* and *HML* loci are denoted, *HMR***a**, *HMR*α,

*HML***a** and *HML*α. The mating phenotypes of *MAT***a** and *MAT*α cells are denoted simply **a** and α, respectively. The two letters *HO* denotes the gene encoding the endonuclease required for homothallic switching.

Dominant and recessive suppressors should be denoted, respectively, by three uppercase or three lowercase letters, followed by a locus designation, e.g., *SUP4, SUF1, sup35, suf11,* etc. In some instances UAA ochre suppressors and UAG amber suppressors are further designated, respectively, **o** and **a** following the locus. For example, *SUP4*-**o** refers to suppressors of the *SUP4* locus that insert tyrosine residues at UAA sites; *SUP4*-**a** refers to suppressors of the same *SUP4* locus that insert tyrosine residues at UAG sites. The corresponding wild-type locus that encodes the normal tyrosine tRNA and that lacks suppressor activity can be referred to as *sup4$^+$*. Intragenic mutations that inactivate suppressors can denoted, for example, *sup4$^-$* or *sup4-**o**-1*. Frameshift suppressors are denoted as *suf* (or *SUF*), whereas metabolic suppressors are denoted with a variety of specialized symbols, such as *ssn* (suppressor of *snf1*), *srn* (suppressor of *rna1-1*), and *suh* (suppressor of *his2-1*).

Capital letters are also used to designate certain DNA segments whose locations have been determined by a combination of recombinant DNA techniques and classical mapping procedures, e.g., *RDN1*, the segment encoding ribosomal RNA.

The general form YCRXXw is now used to designate genes uncovered by systematically sequencing the yeast genome, where Y designates yeast; C (or A, B, *etc.*) designates the chromosome III (or I, II, *etc.*); R (or L) designates the right (or left) arm of the chromosome; XX designates the relative position of the start of the open-reading frame from the centromere; and w (or c) designates the Watson (or Crick) strand. For example, YCR5c denotes *CIT2*, a previously known but unmapped gene situated on the right arm of chromosome III, fifth open reading-frame from the centromere on the Crick strand.

E. coli genes inserted into yeast are usually denoted by the prokaryotic nomenclature, e. g., *lacZ*.

A list of gene symbols are tabulated in the book edited by Wheals *et al.* (6), whereas a current list can be found in the Internet file http://genome-www.stanford.edu/cgi-bin/dbrun/SacchDB?find+locus

6.2 Mitochondrial Genes

Special consideration should be made of the nomenclature describing mutations of mitochondrial components and function that are determined by both nuclear and mitochondrial DNA genes. The growth on media containing nonfermentable substrates (Nfs) as the sole energy and carbon source (such as glycerol or ethanol) is the most convenient operational procedure for testing mitochondrial function. Lack of growth on nonfermentable media (Nfs$^-$ mutants), as well as other mitochondrial alterations, can be due to either nuclear or mitochondrial mutations as outlined in Table 3. Nfs$^-$ nuclear mutations are generally denote by the symbol *pet*; however, more specific designations have been used instead of *pet* when the gene products were known, such as *cox4*, *hem1*, etc.

The complexity of nomenclatures for mitochondrial DNA genes, outlined in Table 3, is due in part to complexity of the system, polymorphic differences of mitochondrial DNA, complementation between exon and intron mutations, the presence of intron-encoded maturases, diversed phenotypes of mutations within the same gene, and the lack of agreement between various workers. Unfortunately, the nomenclature for most mitochondrial mutations do not follow the rules outline for nuclear mutations. Furthermore, confusion can occur between

Table 3. Mitochondrial genes and mutations with examples

Wild-type	Mutation (with examples)	Mutant phenotype or gene product
Nuclear genes		
PET^+	pet^-	Nfs$^-$
	$pet1$	Unknown function
	$cox4$	Cytochrome c oxidase subunit IV
	$hem1$	δ-Aminolevulinate synthase
	$cyc3$	Cytochrome c heme lyase
Mitochondrial DNA		
	Gross aberrations	
ρ^+	ρ^-	Nfs$^-$
	ρ^o	ρ^- mutants lacking mitochondrial DNA
	Single-site mutations	
ρ^+	mit^-	Nfs$^-$, but capable of mitochondrial translation
[$COX1$]	[$cox1$]	Cytochrome c oxidase subunit I
[$COX2$]	[$cox2$]	Cytochrome c oxidase subunit II
[$COX3$]	[$cox3$]	Cytochrome c oxidase subunit III
[$COB1$]	[$cob1$] or [box]	Cytochrome b
[$ATP6$]	[$atp6$]	ATPase subunit 6
[$ATP8$]	[$atp8$]	ATPase subunit 8
[$ATP9$]	[$atp9$] or [$pho2$]	ATPase subunit 9
[$VAR1$]		Mitochondrial ribosomal subunit
ρ^+	syn^-	Nfs$^-$, deficient in mitochondrial translation
	tRNA^{Asp} or M7-37	Mitochondrial tRNA^{Asp} (CUG)
	ant^R	Resistant to inhibitors
[ery^S]	ery^R or [$rib1$]	Resistant to erythromycin, 21S rRNA
[cap^S]	cap^R or [$rib3$]	Resistant to chloramphenical, 21S rRNA
[par^S]	par^R or [$par1$]	Resistant to paromomycin, 16S rRNA
[oli^S]	oli^R or [$oli1$]	Resistant to oligomycin, ATPase subunit 9

Nfs$^-$ denotes lack of growth on nonfermentable substrates.

phenotypic designations, mutant isolation number, allelic designations, loci, and cistrons (complementation groups).

6.3 Non-Mendelian Determinants

In addition to the non-Mendelian determinants described in Figure 2 (2 μm plasmid, mitochondrial genes, and RNA viruses) and discussed in Section 5 (The Yeast Genome), yeast contains elements that have been proposed to be prions, i.e., infectious proteins, on the bases of their genetic properties. The nomenclature of these putative prions, representing alternative protein states, are presented in Table 4.

Table 4. Nomenclature of presumptive prions exhibiting non-Mendelian inheritance

Prion state		Putative gene product	Phenotype of negative state
Positive	Negative		
ψ^+	ψ^-	Sup35p	Decreased efficiency of certain suppression
ξ^+	ξ^-	Sup35p	Decreased efficiency of certain suppression
[*URE3*]	[*ure3*$^-$]	Ure2p	Deficiency in ureidosuccinate utilization

7 Genetic Analyses

7.1 Overviews with Examples

There are numerous approaches for the isolation and characterization of mutations in yeast. Generally, a haploid strain is treated with a mutagen, such as ethylmethanesulfonate, and the desired mutants are detected by any one of a number of procedures. For example, if Yfg$^-$ (Your Favorite Gene) represents an auxotrophic requirement, such as arginine, or temperature-sensitive mutants unable to grow at 37°C, the mutants could be scored by replica plating. Once identified, the Yfg$^-$ mutants could be analyzed by a variety of genetic and molecular methods. Three major methods, complementation, meiotic analysis and molecular cloning are illustrated in Figure 3.

Genetic complementation is carried out by crossing the Yfg$^-$ *MAT***a** mutant to each of the tester strains *MAT*α *yfg1*, *MAT*α *yfg2*, etc., as well as the normal control strain *MAT*α. These *yfg1*, *yfg2*, etc., are previously defined mutations causing the same phenotype. The diploid crosses are isolated and the Yfg trait is scored. The Yfg$^+$ phenotype in the heterozygous control cross establishes that the Yfg$^-$ mutation is recessive. The Yfg$^-$ phenotype in *MAT*α *yfg1* cross, and the Yfg$^+$ phenotype in the *MAT*α *yfg2*, *MAT*α *yfg3*, etc., crosses reveals that the original Yfg$^-$ mutant contains a *yfg1* mutation.

Meiotic analysis can be used to determine if a mutation is an alteration at a single genetic locus and to determine genetic linkage of the mutation both to its centromere and to other markers in the cross. As illustrated in Figure 3, the *MAT***a** *yfg1* mutant is crossed to a normal *MAT*α strain. The diploid is isolated and sporulated. Typically, sporulated cultures contain the desired asci with four spores, as well as unsporulated diploid cells and rare asci with less than four spores. The sporulated culture is treated with snail extract which contains an enzyme that dissolves the ascus sac, but leaves the four spores of each tetrad adhering to each other. A portion of the treated sporulated culture is gently transferred to the surface of a petri plate or an agar slab. The four spores of each cluster are separated with a microneedle controlled by a micromanipulator. After separation of the desired number of tetrads, the ascospores are allowed to germinate and form colonies on complete medium. The haploid segregants can then be scored for the Yfg$^+$ and Yfg$^-$ phenotypes. Because the four spores from each tetrad are the product of a single meiotic event, a 2:2 segregation of the Yfg$^+$:Yfg$^-$ phenotypes is indicative of a single gene. If other markers are present in the cross, genetic linkage of the *yfg1* mutation to the other markers or to the centromere of its chromosome could be revealed from the segregation patterns.

The molecular characterization of the *yfg1* mutation can be carried out by cloning the wild-type *YFG1*$^+$ gene by complementation, as illustrated in Figure 3 and described below (Section 11.1 Cloning by Complementation).
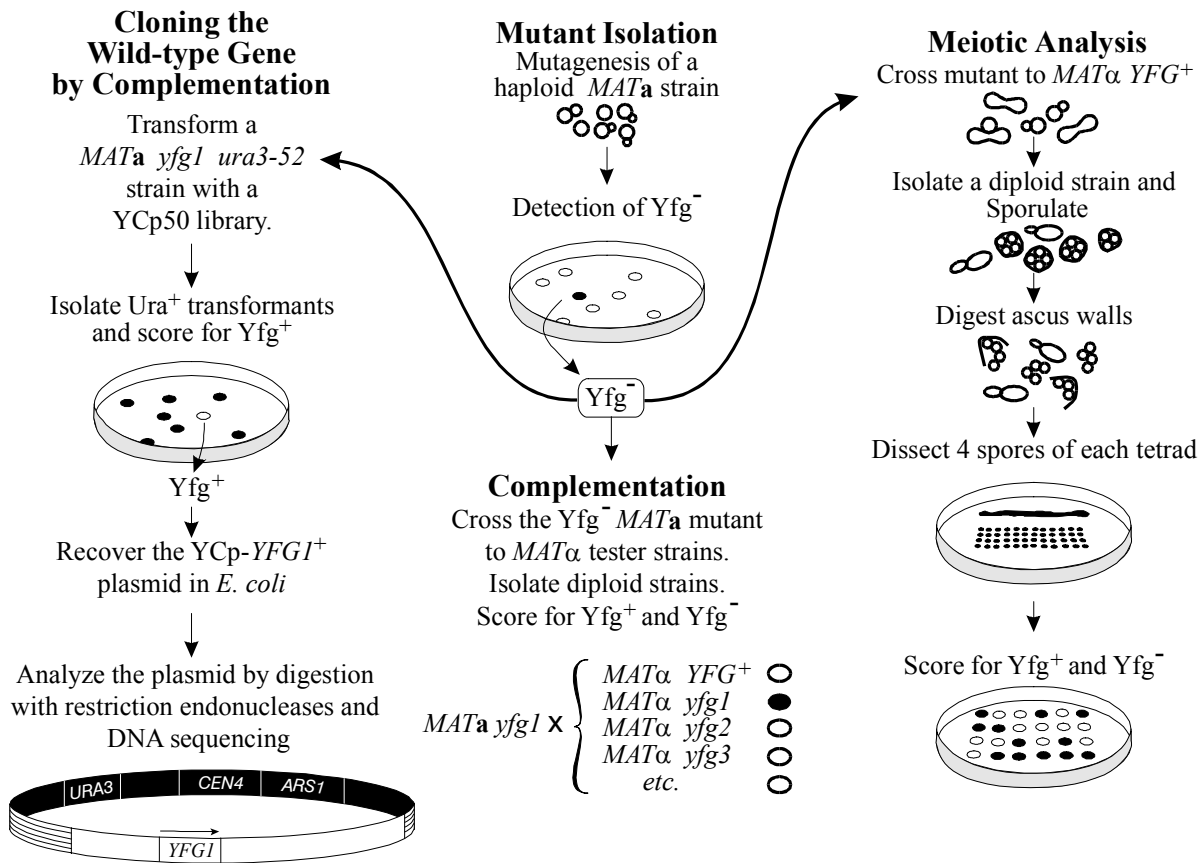
**Cloning the
Wild-type Gene
by Complementation**

Transform a
*MAT***a** *yfg1 ura3-52*
strain with a
YCp50 library.

Isolate Ura$^+$ transformants
and score for Yfg$^+$

Yfg$^+$

Recover the YCp-*YFG1$^+$*
plasmid in *E. coli*

Analyze the plasmid by digestion
with restriction endonucleases and
DNA sequencing

| URA3 | CEN4 | ARS1 |

YFG1

Mutant Isolation
Mutagenesis of a
haploid *MAT***a** strain

Detection of Yfg$^-$

Yfg$^-$

Complementation
Cross the Yfg$^-$ *MAT***a** mutant
to *MAT*α tester strains.
Isolate diploid strains.
Score for Yfg$^+$ and Yfg$^-$

*MAT***a** *yfg1* ✕
$\begin{cases} MAT\alpha \;\; YFG^+ & \bigcirc \\ MAT\alpha \;\; yfg1 & \bullet \\ MAT\alpha \;\; yfg2 & \bigcirc \\ MAT\alpha \;\; yfg3 & \bigcirc \\ \quad etc. & \bigcirc \end{cases}$

Meiotic Analysis
Cross mutant to *MAT*α *YFG$^+$*

Isolate a diploid strain and
Sporulate

Digest ascus walls

Dissect 4 spores of each tetrad

Score for Yfg$^+$ and Yfg$^-$

Figure 3. General approaches for genetic analysis. As an example, a *MAT***a** strain is mutagenized and a hypothetical trait, Yfg$^-$ (Your Favorite Gene) is detected. The Yfg$^-$ mutant is analyzed by three methods, complementation, meiotic analysis and molecular cloning (see the text).

7.3 Tetrad analysis

Meiotic analysis is the traditional method for genetically determining the order and distances between genes of organisms having well-defined genetics systems. Yeast is especially suited for meiotic mapping because the four spores in an ascus are the products of a single meiotic event, and the genetic analysis of these tetrads provides a sensitive means for determining linkage relationships of genes present in the heterozygous condition. It is also possible to map a gene relative to its centromere if known centromere-linked genes are present in the cross. Although the isolation of the four spores from an ascus is one of the more difficult techniques in yeast genetics, requiring a micromanipulator and practice, tetrad analysis is routinely carried out in most laboratories working primarily with yeast. Even though linkage relationships are no longer required for most studies, tetrad analysis is necessary for determining a mutation corresponds to an alteration at a single locus, for constructing strains with new arrays of markers, and for investigating the interaction of genes.
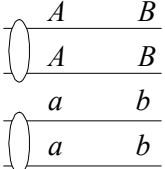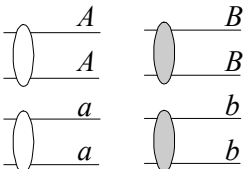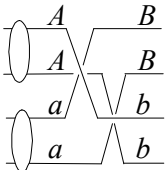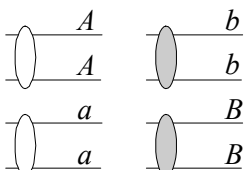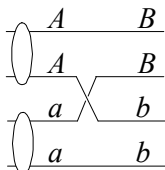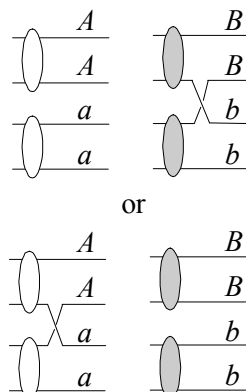
Tetrad type	Genes on homologous chromosomes	Genes on nonhomologous chromosomes
Parental ditype (PD) A B A B a b a b	No crossover 	No crossover
Non-parental ditype (NPD) A b A b a B a B	Double crossover 	No crossover
Tetratype (T) A B A b a B a b	Single crossover 	Single crossovers or

Figure 4. Origin of different tetrad types. Different tetrad types (left) are produced with genes on homologous (center) or nonhomologous (right) chromosomes from the cross *AB* x *ab*. When PD > NPD, then the genes are on homologous chromosomes, because of the rarity of NPD, which arise from four strand double crossovers. The tetratype (T) tetrads arise from single crossovers. See the text for the method of converting the %T and %NPD tetrads to map distances when genes are on homologous chromosomes. If gene are on nonhomologous chromosomes, or if they greatly separated on the same chromosome, then PD = NPD, because of independent assortment, or multiple crossovers. Tetratype tetrads of genes on nonhomologous chromosomes arise by crossovers between either of the genes and their centromere, as shown in the lower right of the figure. The %T can be used to determine centromere distances if it is known for one of the genes (see the text).

There are three classes of tetrads from a hybrid which is heterozygous for two markers, *AB* x *ab*: PD (parental ditype), NPD (non-parental ditype) and T (tetratype) as shown in Figure 4. The following ratios of these tetrads can be used to deduce gene and centromere linkage:

	PD	NPD	T
	AB	*aB*	*AB*
	AB	*aB*	*Ab*
	ab	*Ab*	*ab*
	ab	*Ab*	*aB*
Random assortment	1 :	1 :	4
Linkage	>1 :	<1	
Centromere linkage	1 :	1 :	<4

There is an excess of PD to NPD asci if two genes are linked. If two genes are on different chromosomes and are linked to their respective centromeres, there is a reduction of the

proportion of T asci. If two genes are on different chromosomes and at least one gene is not centromere-linked, or if two genes are widely separated on the same chromosome, there is independent assortment and the PD : NPD : T ratio is 1 : 1 : 4. The origin of different tetrad types are illustrated in Figure 4.

The frequencies of PD, NPD, and T tetrads can be used to determine the map distance in cM (centimorgans) between two genes if there are two or lesser exchanges within the interval:

$$cM = \frac{100}{2}\left[\frac{T + 6NPD}{PD + NPD + T}\right]$$

The equation for deducing map distances, cM, is accurate for distances up to approximately 35 cM. For larger distances up to approximately 75 cM, the value can be corrected by the following empirically-derived equation:

$$cM \text{ (corrected)} = \frac{(80.7)(cM) - (0.883)(cM)^2}{83.3 - cM}$$

Similarly, the distance between a marker and its centromere cM', can be approximated from the percentage of T tetrads with a tightly-linked centromere marker, such as *trp1*:

$$cM' = \frac{100}{2}\left[\frac{T}{PD + NPD + T}\right]$$

7.3   Non-Mendelian Inheritance

The inheritance of non-Mendelian elements can be revealed by tetrad analysis. For example, a cross of ρ^+ *MAT*a and ρ^- *MAT*α haploid strains would result in ρ^+ *MAT*a/*MAT*α and ρ^- *MAT*a/*MAT*α diploid strains, the proportion of which would depend on the particular ρ^- strain. Each ascus from a ρ^+ diploid strain contains four ρ^+ segregants or a ratio of 4:0 for ρ^+:ρ^-. In contrast, a cross of *pet1 MAT*a and *PET1*$^+$ *MAT*α strains would result in a *PET1*$^+$/*pet1 MAT*α/*MAT*a diploid, which would yield a 2:2 segregation of *PET1*$^+$/*pet1*. Similar, the other non-Mendelian determinants also produce primarily 4:0 or 0:4 segregations after meiosis.

Another means for analyzing non-Mendelian elements is cytoduction, which is based on the segregation of haploid cells, either *MAT*a or *MAT*α, from zygotes. Haploid cells arise from zygotes at frequencies of approximately 10^{-3} with normal strains, and nearly 80% with *kar1* crosses, such as, for example, *kar1 MAT*a x *KAR1*$^+$ *MAT*α. While the haploid segregants from a *kar1* cross generally retains all of the chromosomal markers from either the *MAT*a or *MAT*α parental strain, the non-Mendelian elements can be reassorted. For example, a *MAT*a *can*^{R}*1* *kar1* [ρ^- ψ$^-$ *kil*-o] x *MAT*α *CAN*^{S}*1* [ρ^+ ψ$^+$ *kil*-k] cross can yield *MAT*a *can*^{R}*1 kar1* haploid segregants that are [ρ^+ ψ$^+$ *kil*-k], [ρ^- ψ$^+$ *kil*-k], etc. In addition, high frequencies of 2 μm plasmids and low frequencies of chromosome can leak from one nucleus to another.

Also, the mating of two cells with different mitochondrial DNAs results in a heteroplasmic zygote containing both mitochondrial genomes. Mitotic growth of the zygote usually is accompanied by rapid segregation of homoplasmic cells containing either one of the parental mitochondrial DNAs or a recombinant product. The frequent recombination and rapid mitotic

segregation of mitochondrial DNAs can be seen, for example, by mating two different mit^- strains, and observing both Nfs⁻ parental types as well as the Nfs⁺ recombinant (see Table 3).

8 Transformation

8.1 Yeast Vector and DNA Fragments

In general, transformation is the introduction into cells of exogenously added DNA and the subsequent inheritance and expression of that DNA. The most important advances in the molecular characterization and controlled modification of yeast genes have relied on the use of shuttle vectors which can be used to transform both yeast and *E. coli*.

The following three main methods are currently used to transform yeast: (*i*) those using spheroplasts; or (*ii*) cells treated with lithium salts; and (*iii*) the use of electroporation.

Spheroplasts for transformations are prepared by the action of hydrolytic enzymes to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. Cell-wall digestion is carried out either with a snail-gut extract, usually denoted Glusulase, or with Zymolyase, an enzyme from *Arthrobacter luteus*. DNA is added to the spheroplasts, and the mixtures is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol. Although this protocol is particularly tedious, and efficiency of transformation can vary by over four orders of magnitude with different strains, very high frequencies of transformation, over 10^4 transformants/µg DNA, can be obtained with certain strains.

Most investigators use cells treated with lithium salts for transformation. After treating the cells with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents.

A commonly-used method for transforming a wide range of different species of cells is based on the induced permeability to DNA by exposure to electrical fields. The interaction of an external electric field with the lipid dipoles of a pore configuration is believed to induce and stabilize the permeation sites, resulting in cross membrane transport. Freshly-grown yeast cultures are washed, suspended in an osmotic protectant, such as sorbitol, DNA is added, and the cell suspension is pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Although electroporation procedures are simple, the specialized equipment and the required cuvettes are costly.

8.2 Synthetic Oligonucleotides

A convenient procedure has been described for producing specific alterations of chromosomal genes by transforming yeast directly with synthetic oligonucleotides. This procedure is easily carried out by transforming a defective mutant and selecting for at least partially functional revertants. Transformation of yeast directly with synthetic oligonucleotides is thus ideally suited for producing a large number of specific alterations that change a completely nonfunctional allele to at least a partially functional form. The oligonucleotide

should contain a sequence that would correct the defect and produce the desired additional alterations at nearly sites. The method is apparently applicable to all mutant alleles whose functional forms can be selected. Although it is a general procedure, so far it has been extensively used only with mutations of *CYC1*, that encodes iso-1-cytochrome *c*, and *CYT1* that encodes cytochrome c_1. The transformation is carried out by the usual lithium acetate procedure, using approximately 50 μg of oligonucleotides that are approximately 40 nucleotides long.

8.3 Mitochondrial Transformation

Standard methods for transformation of nuclear genes are ineffective for mitochondrial DNA genes. However, DNA can be delivered to the mitochondrial matrix by high-velocity bombardment of yeast cells with tungsten microprojectiles carrying mitochondrial DNA. Several high-velocity microprojectile bombardment devices are commercially available, and these are powered by gunpowder charge or compressed gas.

This method was used to demonstrated that ρ^o strains can be converted to stable "synthetic ρ^-" strains by transformation with bacterial plasmids carrying mitochondrial genes (see Table 3). Similar to natural ρ^- mitochondrial DNA, the synthetic ρ^- mitochondrial DNA can recombine with ρ^+ mitochondrial DNA, thus providing means to replace ρ^+ wild-type genes with mutations generated *in vitro*.

Synthetic ρ^- strains are isolated by bombarding a lawn of ρ^o cells on the surface of a petri plate with YEp or YCp plasmids carrying both a selectable marker, such as *URA3*, and the mitochondrial gene of interest. The nuclear and mitochondrial genes may either be on separate or the same plasmid. Ura^+ colonies, for example, are then screen for the presence of the mitochondrial gene by crossing the colonies to an appropriate *mit⁻* tester strain and scoring the diploids for Nfs^+ (see Table 3). The efficiency of mitochondrial transformation varies from experiment to experiment, and can be from 2 x 10^{-3} to less than 10^{-4} mitochondrial transformants per nuclear transformant.

9 Yeast Vectors

A wide range of vectors are available to meet various requirements for insertion, deletion alteration and expression of genes in yeast. Most plasmids used for yeast studies are shuttle vectors, which contain sequences permitting them to be selected and propagated in *E. coli*, thus allowing for convenient amplification and subsequent alteration *in vitro*. The most common yeast vectors originated from pBR322 and contain an origin of replication (*ori*), promoting high copy-number maintenance in *E. coli*, and the selectable antibiotic markers, the β-lactamase gene, *bla* (or *Amp^R*), and sometime to tetracycline-resistance gene, *tet* or (*Tet^R*), conferring resistance to, respectively, ampicillin and tetracycline.

In addition, all yeast vectors contain markers that allow selection of transformants containing the desired plasmid. The most commonly used yeast markers include *URA3*, *HIS3*, *LEU2*, *TRP1* and *LYS2*, which complement specific auxotrophic mutations in yeast, such as *ura3-52*, *his3-Δ1*, *leu2-Δ1*, *trp1-Δ1* and *lys2-201*. These complementable yeast mutations have been chosen because of their low-reversion rate. Also, the *URA3*, *HIS3*, *LEU2* and *TRP1* yeast markers can complement specific *E. coli* auxotrophic mutations.

The *URA3* and *LYS2* yeast genes have an additional advantage because both positive and negative selections are possible, as discussed below (Section 10.1, *URA3* and *LYS2*).

Table 5. Components of common yeast plasmid vectors

	YIp	YEp	YRp	YCp
Plasmid				
E. coli genes or segments				
ori, *bla*; *tet*	+	+	+	+
Yeast genes or segments				
URA3; *HIS3*; *LEU2*; *TRP1*; *LYS2*; *etc.*	+	+	+	+
leu2-d	0	+	+	0
2 μm; 2 μm-*ori REP3*;	0	+	0	0
ARS1; *ARS2*; *ARS3*; *etc.*	0	0	+	+
CEN3; *CEN4*; *CEN11*; *etc.*	0	0	0	+
Host (yeast) markers				
ura3-52; *his3-Δ1*; *leu2-Δ1*; *trp1-Δ1*; *lys2-201*; *etc.*	+	+	+	+
Stability	++	+	±	+

    Although there are numerous kinds of yeast shuttle vectors, those used currently can be broadly classified in either of following three types as summarized in Table 5: integrative vectors, YIp; autonomously replicating high copy-number vectors, YEp; or autonomously replicating low copy-number vectors, YCp. Another type of vector, YACs, for cloning large fragments are discussed in Section 13.2 (Yeast Artificial Chromosomes).

9.1 YIp Vectors

    The YpI integrative vectors do not replicate autonomously, but integrate into the genome at low frequencies by homologous recombination. Integration of circular plasmid DNA by homologous recombination leads to a copy of the vector sequence flanked by two direct copies of the yeast sequence as illustrated in the top of Figure 5. The site of integration can be targeted by cutting the yeast segment in the YIp plasmid with a restriction endonuclease and transforming the yeast strain with the linearized plasmid. The linear ends are recombinogenic and direct integration to the site in the genome that is homologous to these ends. In addition, linearization increases the efficiency of integrative transformation from 10- to 50-fold.

    The YIp vectors typically integrate as a single copy. However multiple integration do occur at low frequencies, a property that can be used to construct stable strains overexpressing specific genes. YIp plasmids with two yeast segments, such as *YFG1* and *URA3* marker, have the potential to integrate at either of the genomic loci, whereas vectors containing repetitive DNA sequences, such as Ty elements or rDNA, can integrate at any of the multiple sites within genome. Strains constructed with YIp plasmids should be examined by PCR analysis, or other methods, to confirm the site of integration.

    Strains transformed with YIp plasmids are extremely stable, even in the absence of selective pressure. However, plasmid loss can occur at approximately 10^{-3} to 10^{-4} frequencies by homologous recombination between tandemly repeated DNA, leading to looping out of the vector sequence and one copy of the duplicated sequence as illustrated in Figure 5 and discussed below in Section 11.2 (Two-Step Gene Replacement).
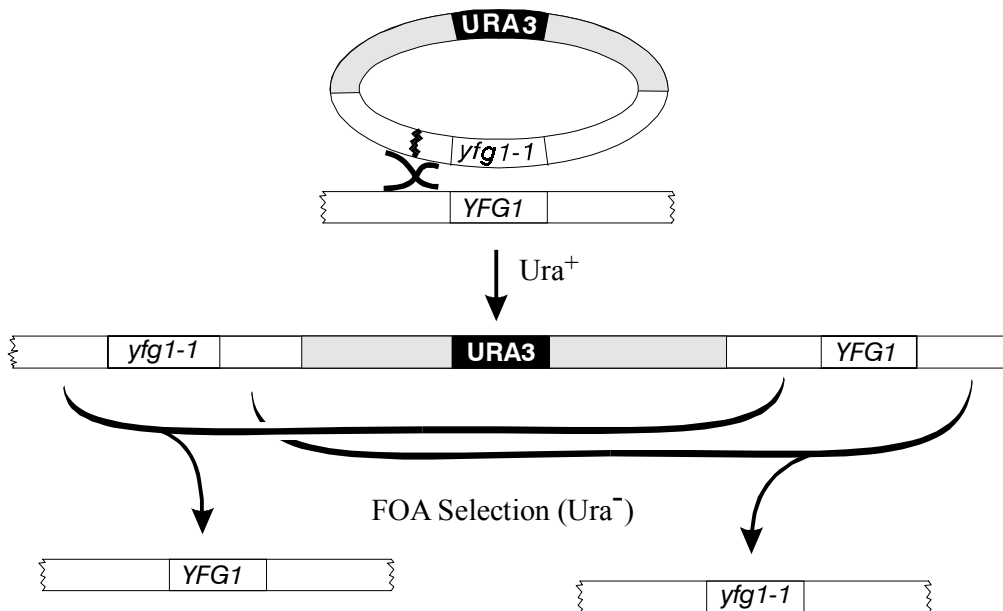
Figure 5. Two-step gene replacement. The wild-type chromosomal $YFG1^+$ allele can be replaced by the mutant *yfg1-1* allele from a YIp integrating plasmid. The plasmid is first integrated in the chromosome corresponding to the site on the plasmid that was cleaved by a restriction endonuclease. Strains that have excised the *URA3* marker *in vivo* by homologous recombination are selected on FOA medium. Either the original $YFG1^+$ allele, or the *yfg1-1* allele remains in the chromosome, depending on the site of the cross-over.

9.2 YEp Vectors

The YEp yeast episomal plasmid vectors replicate autonomously because of the presence of a segment of the yeast 2 μm plasmid that serves as an origin of replication (2 μm *ori*). The 2 μm *ori* is responsible for the high copy-number and high frequency of transformation of YEp vectors.

YEp vectors contain either a full copy of the 2 μm plasmid, or, as with most of these kinds of vectors, a region which encompasses the *ori* and the *REP3* gene. The *REP3* gene is required in *cis* to the *ori* for mediating the action of the *trans*-acting *REP1* and *REP2* genes which encode products that promote partitioning of the plasmid between cells at division. Therefore, the YEp plasmids containing the region encompassing only *ori* and *REP3* must be propagated in cir$^+$ hosts containing the native 2 μm plasmid (Figure 2).

Most YEp plasmids are relatively unstable, being lost in approximately 10^{-2} or more cells after each generation. Even under conditions of selective growth, only 60% to 95% of the cells retain the YEp plasmid.

The copy number of most YEp plasmids ranges from 10-40 per cell of cir$^+$ hosts. However, the plasmids are not equally distributed among the cells, and there is a high variance in the copy number per cell in populations.

Several systems have been developed for producing very high copy-numbers of YEp plasmids per cell, including the use of the partially defective mutation *leu2-d*, whose expression is several orders of magnitude less than the wild-type $LEU2^+$ allele. The copy number per cell

of such YEp *leu2-d* vectors range from 200-300, and the high copy-number persists for many generations after growth in leucine-containing media without selective pressure. The YEp *leu2-d* vectors are useful in large-scale cultures with complete media where plasmid selection is not possible. The most common use for YEp plasmid vectors is to overproduce gene products in yeast.

9.3 YCp Vectors

The YCp yeast centromere plasmid vectors are autonomously replicating vectors containing centromere sequences, CEN, and autonomously replicating sequences, ARS. The YCp vectors are typically present at very low copy numbers, from 1 to 3 per cell, and possibly more, and are lost in approximately 10^{-2} cells per generation without selective pressure. In many instances, the YCp vectors segregate to two of the four ascospore from an ascus, indicating that they mimic the behavior of chromosomes during meiosis, as well as during mitosis. The *ARS* sequences are believed to correspond to the natural replication origins of yeast chromosomes, and all of them contain a specific consensus sequence. The *CEN* function is dependent on three conserved domains, designated I, II, and III; all three of these elements are required for mitotic stabilization of YCp vectors. YRp vectors, containing *ARS* but lacking functional *CEN* elements, transform yeast at high frequencies, but are lost at too high a frequency, over 10% per generation, making them undesirable for general vectors.

The stability and low copy-number of YCp vectors make them the ideal choice for cloning vectors, for construction of yeast genomic DNA libraries, and for investigating the function of genes altered *in vivo*. *ARS1*, which is in close proximity to *TRP1*, is the most commonly used *ARS* element for YCp vectors, although others have been used. *CEN3*, *CEN4* and *CEN11* are commonly used centromeres that can be conveniently manipulated. For example, the vector YCp50 contains *CEN4* and *ARS1*.

10 Genes Important for Genetic Studies

Several genes and promoters are commonly employed for genetic manipulations and studies with yeast. Some of these genes have special properties, whereas others were originally choosen arbitarily and are conveniently available in strains and plasmids. Several of the genes most commonly used for a variety of purposes are described below.

10.1 *URA3* and *LYS2*

The *URA3* and *LYS2* yeast genes have a marked advantage because both positive and negative selections are possible. Positive selection is carried out by auxotrophic complementation of the *ura3* and *lys2* mutations, whereas negative selection is based on specific inhibitors, 5-fluoro-orotic acid (FOA) and α-aminoadipic acid (αAA), respectively, that prevent growth of the prototrophic strains but allows growth of the *ura3* and *lys2* mutants, respectively.

URA3 encodes orotidine-5′phosphate decarboxylase, an enzyme which is required for the biosynthesis of uracil. *Ura3⁻* (or *ura5⁻*) cells can be selected on media containing FOA. The *URA3⁺* cells are killed because FOA appears to be converted to the toxic compound 5-fluorouracil by the action of decarboxylase, whereas *ura3⁻* cells are resistant. The negative selection on FOA media is highly discriminating, and usually less than 10^{-2} FOA-resistant colonies are Ura⁺. The FOA selection procedure can be used to produce *ura3* markers in haploid strains by mutation, and, more importantly, for expelling *URA3*-containing plasmids, including YCp and YEp replicating plasmids, and integated YIp plasmids, as discussed below

for a number of genetic strategies (Section 11). Because of this negative selection and its small size, *URA3* is the most widely used yeast marker in yeast vectors. The specfic allele, *ura3-52*, which is the most commonly used host marker, contains a Ty1 insertion, is not revertible, and does not allow integation of YIp-*URA3* plasmids at the *URA3* chromosomal locus in most, but not all strains.

LYS2 encodes α-aminoadipate reductase, an enzyme which is required for the biosynthesis of lysine. *Lys2⁻* and *lys5⁻* mutants, but not normal strains grow on a medium lacking the normal nitrogen source, but containing lysine and αAA. Apparently, *lys2* and *lys5* mutations cause the accumulation of a toxic intermediate of lysine biosynthesis that is formed by high levels of αAA, but these mutants still can use αAA as a nitrogen source. Numerous *lys2* mutants and low frequencies of *lys5* mutants can be conveniently obtained by simply plating high densities of normal cells on αAA medium. Similar with the FOA selection procedure, *LYS2*-containing plasmids can be conveniently expelled from *lys2* hosts. Because of the large size of the *LYS2* gene and the presence of numerous restriction sites, the FOA selection procedure with *URA3* plasmids are more commonly used.

10.2 *ADE1* and *ADE2*

The *ADE1* and *ADE2* yeast genes encode phosphoribosylamino-imidazole-succinocarbozamide synthetase and phosphoribosylamino-imidazole-carboxylase, respectively, two enzymes in the biosynthetic pathway of adenine. *Ade1* and *ade2* mutants, but no other *ade⁻* mutants, produce a red pigment that is apparently derived from the polymerization of the intermediate phosphoribosylamino-imidazole (denoted AIR). Furthermore, the formation of AIR is prevented by blocks preceding the *ADE1* and *ADE2* steps. For example *ade2* strains are red, whereas *ade3* and the double mutant *ade2 ade3* are both white, similar to the color of normal strains. Red colonies and red-white sectored colonies are easily detected by visual inspection.

The *ade1* and *ade2* red pigmentation, and the prevention of the coloration by *ade3* or other *ade⁻* mutation has been incorporated as a detection scheme in a number of diversed genetic screens. Also, the *ade2-1* UAA mutation, and the suppression of formation of the red pigment by the *SUP4*-**o** suppressor has been used in a variety of genetic screens. Most of the screens are based on the preferential loss, or the required retention of a plasmid containing a component involved in the formation of the red pigment.

Examples of *ade⁻* red genetic screens include the detection of conditional mutations (Section 11.5, Plasmid Shuffle), isolation of synthetic lethal mutations (Section 12.5, Synthetic Enhancement and Epistatic Relationships), detection of YAC transformants (Section 13.2, Yeast Artificial Chromosomes [YACs]), and the isolation of mutations that effect plasmid stability.

10.3 *GAL1* Promoter

Cloned genes can be expressed with constitutive or regulatable promoters. The most commonly-used regulated promoter for yeast studies is P_{GAL1}.

There are two regulatory proteins, Gal4p and Gal80p, which effect the transcription of the following structural genes: *GAL1*, a kinase; *GAL2*, a permease; *GAL7*, a transferase; *GAL10*, an epimerase; and *MEL1*, a galactosidase. Gal3p appears to be required for the production of the intracellular inducer from galactose. In presence of the inducer, Gal4p binds to sites in the UAS (upstream activation sequence), and activates transcription. In the absence of the inducer, such as when cells are grown in media containing nonfermentable carbon sources, Gal80p binds to the carboxyl terminal region of Gal4p, masking the activation domain. Expression is repressed in

cells exposed to glucose-containing media for several reasons in addition to the absence of the inducer, including the action of repressors at sites between the UAS and the TATA box and the inhibition of galactose uptake. Therefore, the addition of glucose to cells growing in galactose meduim causes an immediate repression of tramscription. The UAS of galactose structural genes contains one or more 17 base-pair palidromic sequences to which Gal4p binds, with the different levels of transcription determined by the number and combinations of the palidromes.

The UAS of the divergently transcribed *GAL1* and *GAL10* is contained within a 365-bp fragment, denoted P_{GAL1}, that is sufficient for maximal galactose induction and thorough glucose repression. P_{GAL1} can rapidly induce the expression of downstream fused-genes over 1000-fold after the addition of galactose to cells growing in media with a nonfermentable carbon source. Furthermore, P_{GAL1} can be turned off by the addition of glucose to the galactose containing medium.

P_{GAL1} has been used in a wide range of studies with yeast, including the overproduction of yeast proteins as well as heterologous proteins (Section 13.3). Most importantly, the strong glucose repression of P_{GAL1} has been used to investigate the terminal phenotype of essential genes, in much the same way that temperature shifts are used to control the activity of temperature-sensitive mutations (see Section 11.2). Also, the P_{GAL1} system has been used to investigate suppression (Section 12.4) and growth inhibition by over expressed normal or mutant genes (dominant-negative mutations, Section 12.1). P_{GAL1} is also an important component of one of the two-hybrid systems (Section 13.1).

10.4 *lacZ* and Other Reporters

Activities of promoters, and protein-protein and protein-DNA interactions involving promoter regions can be readily converted into selectable and quantifiable traits by fusing the promoter regions to reporter genes. Reporter genes can be used to determine the levels of transcription, or the levels of translation of the transcript, under various physiological conditions. The most common use of reporter genes has been to identify elements required for transcription by systematically examining series of mutations in promoter regions. Similarly, reporter genes have been used to identify trans acting factors that modulate expression by transcription or translation.

The *Escherichia coli lacZ* gene, which encodes β-galactosidase, is the most commonly used reporter with yeast and other systems, because its activity can assayed semiquantitatively on plates and fully quantitatively by enzyme assay of liquid cultures. Rare events can be detected by the differential staining of colonies using X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside).

For positive selection, the reporter gene could include, for example, the translated region of the *HIS3* gene, lacking a UAS (upstream-activating sequence). His$^+$ colonies arise when active promoters are formed, such as in the cloning of heterologous components required for the activation of a defined DNA segment. Combining the *HIS3* selection with a *lacZ* screen is a commmoly used strategy; this approch of using two different reporters in parallel with the same promoter region is an efficient means for identifying trans-acting factors.

11 Manipulating the Genome In Vitro with Plasmids

The greatest virtues of using yeast has been the ease with which genes can be retrieved, deleted, inserted and modified in a controlled manner. These methods rely on the combined use of recombinant DNA techniques, transformation and classical yeast genetics procedures. Overviews of some of these major approaches are described in the following sections.

11.1 Cloning by Complementation

Molecular cloning and DNA analysis is the most definitive way of characterizing a gene that corresponds to a mutation. Cloning by complementation is usually carried out with a library of a YCp vector containing inserts of a more-or-less random series of genomic fragments, as illustrated in Figure 3 with the hypothetical *yfg1* mutation.

The *yfg1* strain is transformed with the YCp library, and the transformants are examined for the Yfg$^+$ trait. YCp vectors are generally used because each transformant contains a single or only few plasmids per cell. The method of screening transformants for complementation varies according to the specific phenotype that is to be scored. Direct selection can be used in some instances. However, if the original mutation reverts, as is often the case, a high frequency of false positives occurs among the transformants. Thus, an alternative method of indirect selection by replica-plating is preferred. Thus, by this method, the transformant containing the desired YCp-*YFG1*$^+$ plasmids appear as homogeneous Yfg$^+$ colonies, whereas the colonies containing *yfg1* revertants appear as heterogeneous Yfg$^+$ and Yfg$^-$ mixtures after replica-plating. Most importantly, the true transformants will be dependent on the YCp-*YFG1*$^+$ plasmid for the Yfg$^+$ phenotype. In almost all studies, plasmid dependency is conveniently determined with the *ura3* system and usually with the non-reverting allele *ura3-52*. Because *ura3* mutants can be selected on FOA (5-fluoro-orotic acid) medium, plasmid-free strains therefore can be recovered and subsequently tested for the loss of complementation. For example, the *yfg1 ura3* YCp-*YFG1*$^+$ strain would be Yfg$^+$ Ura$^+$, while the *yfg1 ura3* strain, lacking the plasmid, would be Yfg$^-$ Ura$^-$. Furthermore, the authenticity of the plasmid can be confirmed by first recovering the plasmid in *E. coli* and retransforming the *yfg1* strain.

It is also desirable to confirm that the cloned segment truly encompasses the *YFG1*$^+$ gene. Even though the transformants may contain only a single copy of the putative gene, there are examples in which two wild-type copies of a gene, one on the chromosome and the other on the plasmid, may suppress a mutation situated at a different locus. An independent test, based on homologous recombination, relies on YIp vector containing the insert. If the insert contains a unique restriction site, cleavage at this site will enhance integration of the plasmid at the homologous chromosomal site. Without cleavage, the plasmid could integrate at the site of other yeast markers on the plasmid, as well as at the *YFG1*$^+$ locus. After the integrant has been obtained, the site of integration can be investigated by meiotic analysis. For example, integration of the p[*YFG1*$^+$ *URA3*$^+$] plasmid at the site of *YFG1*$^+$ locus would result in a *YFG1*$^+$::[*YFG1*$^+$ *URA3*$^+$] *ura3* strain. After crossing to a *yfg1 ura3* strain and carrying out a meiotic analysis, the segregants should show a 2:2 segregation for both Yfg$^+$/Yfg$^-$ and Ura$^+$/Ura$^-$ and both markers would segregate as parental ditypes. On the other hand, if the plasmid integrated at a site other than the *YFG1* locus, an excessive number of Yfg$^+$ segregants would be recovered, indicating that the normal chromosomal *YFG1*$^+$ allele and the integrated plasmid *YFG1*$^+$ allele were not linked, or were at least not in close proximity.

If the sequence of the *YFG1* gene and flanking regions are known, the site of integration could be confirmed solely by PCR analysis.

After the desired plasmid has been demonstrated to encompass the *YFG1* gene, restriction fragments can be analyzed to narrowed down the region of the locus, which can be subsequently sequenced and studied by a variety of other methods.

11.2 Mutagenesis *In Vitro*

Two common experimental goals are to produce either specific or "random" mutations within a gene. DNA alteration are required for investigating, for example, structure-function relationships and essential regions of proteins, and for producing conditional mutations, such as temperature-sensitive mutation. Specific alterations are carried out by the general procedure of oligonucleotide-directed mutagenesis that is applicable to any cloned DNA segment, including those used for yeast studies. Oligonucleotide-directed mutagenesis has been used to systemically replace amino acids within proteins, especially the replacement of charged amino acids with alanine residues. Such alanine replacements have resulted in a multitude of effects, including proteins that were unaffected, inactive and temperature sensitive.

Also, numerous general procedures for producing "random" point mutations are available, including treating plasmid DNA with hydroxylamine and misincorporation by PCR mutagenesis. Most importantly, a simple procedure has been developed for the localized mutagenesis of yeast genes, as illustrated in Figure 6B. The region to be altered is first amplified under mutagenic PCR conditions, resulting in the generation of fragments containing "random" *yfg1-x* mutations. A *yfg1-Δ* strain is then cotransformed with these PCR products and with a gapped YCp plasmid containing homology to both ends of the PCR products. Repair of the gap with the PCR products (see Section 11.6) results in a series of strains with YCp plasmids containing the altered *yfg1-x* alleles. The yeast strains containing the *yfg1-Δ* chromosomal deletion can then be individually scored for the phenotype of each of the *yfg1-x* mutants. This procedure is particularly effective for targeting "random" mutations in specific regions, and does not require subcloning steps in *E. coli*.

11.3 Two-step Gene Replacement

After a gene has been cloned, the most efficient means for obtaining mutations in the gene is by mutagenesis *in vitro* of the cloned DNA segment as described above. The effects of the mutations can then be examined *in vivo* by introducing the altered gene in yeast by transformation. A simple and the most common procedure is to transform a yeast strain, which lacks a functional copy of the chromosomal gene, with a YCp plasmid, which contains the altered gene. This can be accomplished directly in a single step if the gene in question is not essential. However, the best procedure, eliminating the problems of copy number and vector sequences, is to replace the chromosomal copy of the gene with the altered plasmid copy. This can be accomplished by the two-step gene replacement procedure illustrated in Figure 5. A YIp plasmid, containing the altered *yfg1-1* gene is integrated in the chromosome in the region containing the $YFG1^+$ normal gene. Homologous recombination results in two copies of the gene, *yfg1-1* and $YFG1^+$, separated by the plasmid sequences. The second step involves homologous crossing over in the repeated DNA segment to loop-out the plasmid, along with the *URA3* gene. Such desired Ura⁻ strain can be selected on FOA medium. The resulting plasmid is lost during growth of the cells because the plasmid lacks an origin of replication.
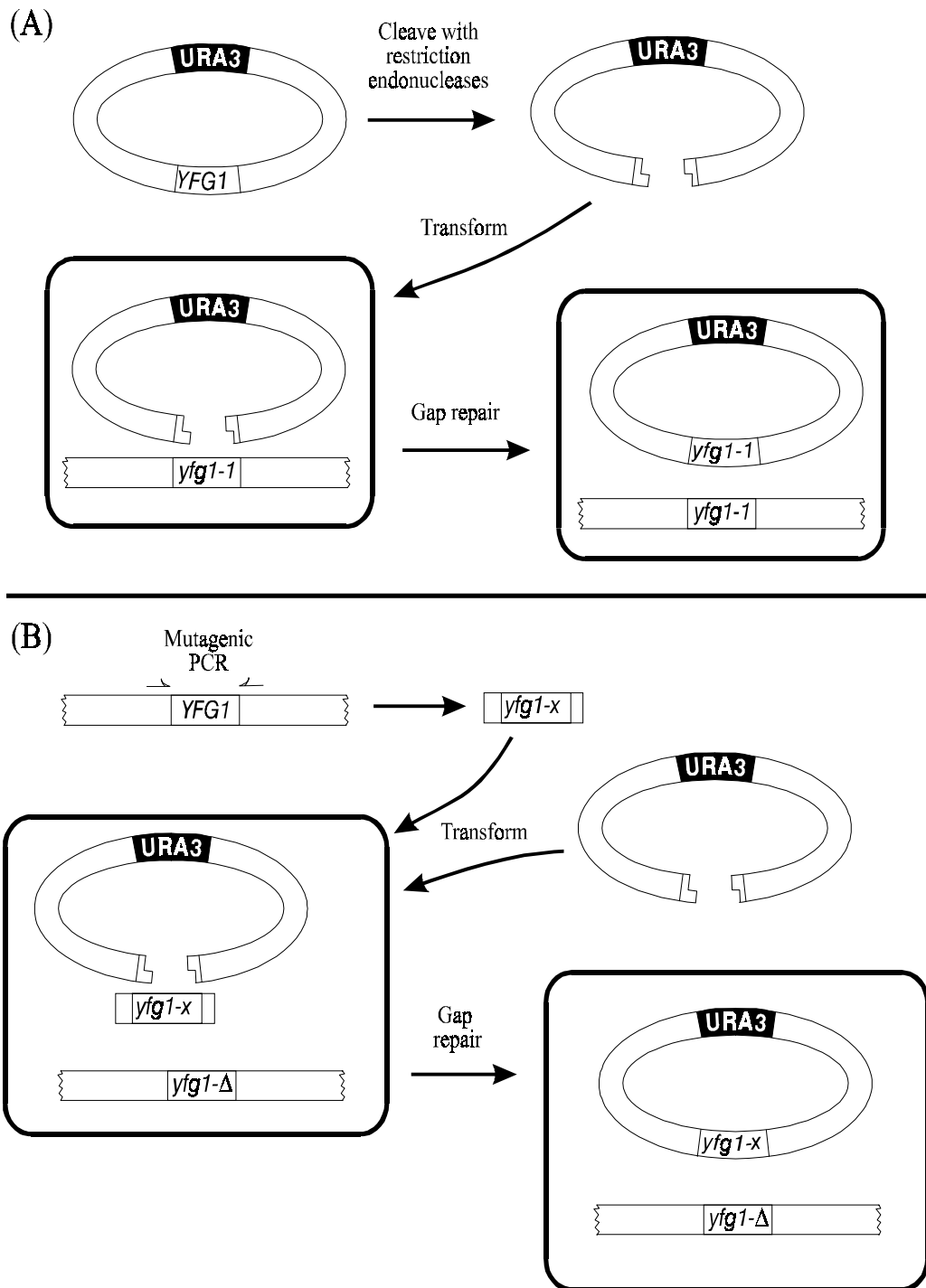
Figure 6. (A) Retrieval of a chromosomal *yfg1-1* mutation by transforming a mutant with a gapped plasmid. Only the repaired circular plasmid containing the mutation is stably maintained in yeast. (B) Generation of a series of *yfg1-x* mutations by PCR mutagenesis and gap repair. The *yfg1-Δ* mutant is cotransformed with the mutagenized PCR fragments and the gapped plasmid. The phenotypes of the *yfg1-x* mutants can be directly assessed in the strain containing the *yfg1-Δ* deletion.

The second cross-over can occur in either of two regions as depicted in Figure 6, the region either to the left or to the right of the *yfg1-1* alteration. The cross-over at the left results in the regeneration of normal *YFG1* allele, whereas a cross-over at the right results in the introduction of only the desired *yfg1-1* mutation.

The position of the cross-over in the second step is approximately random, resulting in recovery of both YFG1 and *yfg1-1* strains. However, the relative frequencies of cross-overs in the two regions are probably related to their lengths. In order to recover the altered *yfg1-1* allele, the second cross-over must occur at the opposite side of the site of integration. Therefore, it is desirable to force the initial integration at the smaller region by cutting the plasmid in this region with a restriction endonuclease.

In addition to the *URA3* marker, the *LYS2* can be also used for both positive and negative selection (see Section 10.1, *URA3* and *LYS2*). However, if neither *URA3* or *LYS2* can be used, loop-out recombination is often sufficiently high, 10^{-3} to 10^{-4}, making it possible to detect the loss of the marker by replica-plating. If a sufficiently large number of altered replacements are contemplated, an additional marker could be introduced into the $YFG1^+$ locus, allowing for the convenient scoring of the desired loop-out.

11.4 Gene Disruption and One-step Gene Replacement

One of the most important and widely used methods to characterize yeast genes is gene disruption. The complete disruption of a gene unambiguously reveals its function and can be helpful for generating additional mutations. Several methods can be used to produce deletions and null mutations, including the two-step gene replacement described above.

The one-step gene disruption procedure is usually preferred because of its simplicity. This procedure is based on the use of a linear fragment of DNA containing a selectable marker flanked by 5′ and 3′ homologous regions as illustrated in Figure 7A. The free ends of the fragment, prepared by digestion with restriction endonucleases, are recombinogenic, resulting in the integration of *URA3* marker and the loss of wild-type *YFG1* allele.

It should be noted that the transformation must be carried out in a diploid strain if the gene encodes an essential function. Also, the disruption of the desired genes should be verified by PCR or Southern blot analysis. The fragment required for single step disruptions can be also conveniently generated by PCR, alleviating the need to clone the *YFG1* gene.

Because the one-step gene disruption procedure results in a $URA3^+$ strain, the method has been modified as illustrated in Figure 7B. In this method, *URA3* is flanked by identical copies of the bacterial *hisG* gene (or any non-yeast DNA segment). The *hisG-URA3-hisG* is first used to produce the gene disruption; subsequently, recombination between the direct repeats and selection on FOA produces a single copy of *hisG* at the site of the disruption. Thus, multiple rounds of gene disruption can be carried out in the same strain.

A similar procedure has also been developed for conveniently replacing mutant alleles in a single step, as illustrated in Figure 7C. The *YFG1* gene is first disrupted with the *URA3* gene as described above. Replacements of the disrupted *YFG1* by altered alleles can be selected on FOA medium among transformants or after minimal growth of transformants on complete medium. This and related procedures are particularly useful when large numbers of replacements are required.

Another method for producing gene disruptions, as well as simultaneously testing for the promoter activity, have been based on a dominant resistant module consisting almost entirely of heterologous DNA. Transformants resistant to geneticin (G418) are selected and examined for
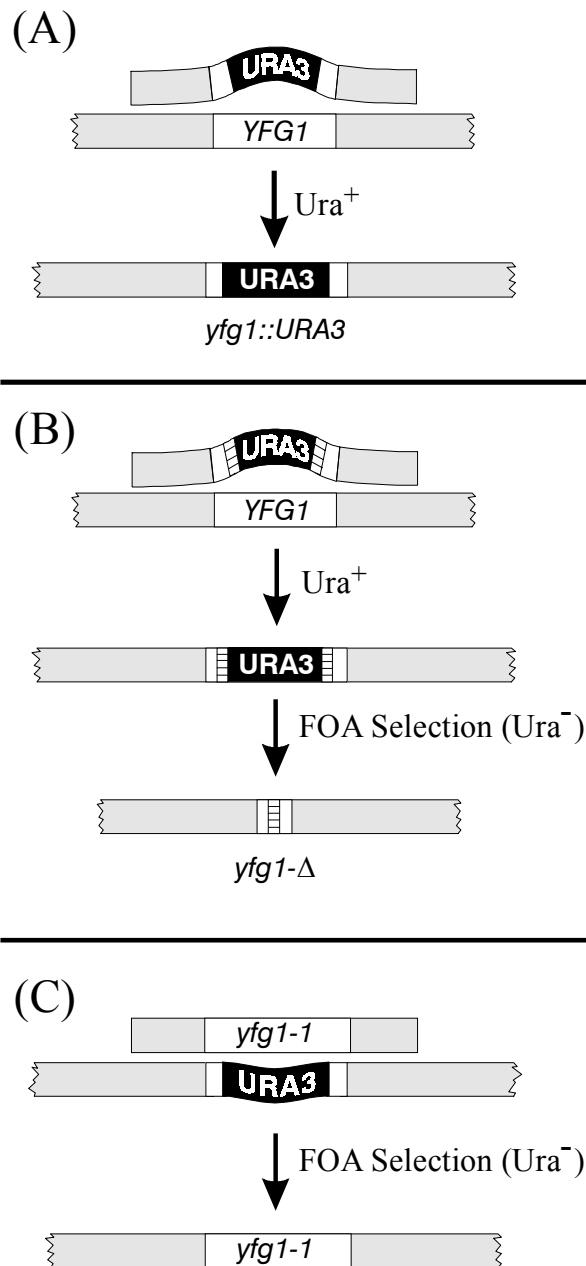
(A)

Ura⁺

yfg1::URA3

(B)

Ura⁺

FOA Selection (Ura⁻)

yfg1-Δ

(C)

FOA Selection (Ura⁻)

yfg1-1

Figure 7. Gene disruption and single-step gene replacement. (A) The *YFG1*$^+$ gene is disrupted by transforming the strain with a linear fragment containing a *URA3* selectable marker flanked by homologous sequences. The chromosomal segment is replaced by this *URA3* containing fragment after integration by homologous recombination at the two ends. (B) The *URA3* marker introduced in the *YFG1* locus, can be excised if *URA3* is also flanked by direct repeats of DNA, preferably not originating from yeast. Homologous recombinants, selection on FOA medium, lack the *URA3* marker and retain a single copy of the repeated DNA. (C) Single-step gene replacement of mutant alleles, such as *yfg1-1*, can be carried out by first replacing the *YFG1* gene by *URA3*, transforming the strain with linear fragment encompassing the *yfg1-1* mutation, and selecting transformants on FOA medium, in which *URA3* is replaced by *yfg1-1*.

lacZ activity. To allow for repeated use of the G418 selection, the module is flanked by short direct repeats, promoting excision *in vivo*.

11.5 Plasmid Shuffle

As mentioned above (Section 11.2), the most common procedure for isolating and characterizing a series of altered alleles, *yfg1-x*, is simply to transform a strain lacking the gene, *yfg1-Δ*, with YCp plasmids containing the altered forms and to examine the phenotype of the transformants, *yfg1-Δ* p[*yfg1-x*]. However, the characterization of mutations of an essential gene poses an additional technical difficulty because of the inviability of strains containing the *yfg1-Δ* null mutation, as well as of those containing nonfunctional *yfg1-x* alleles.
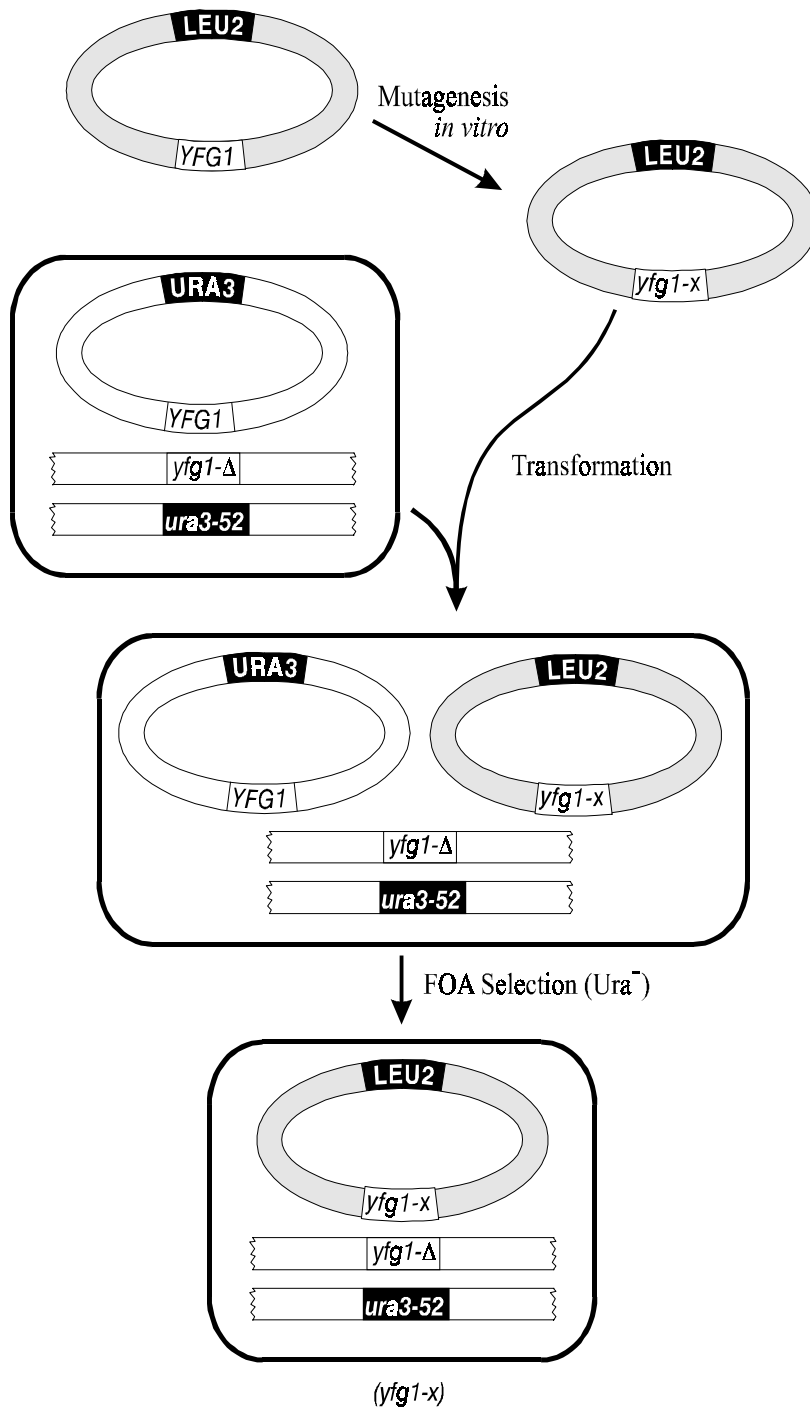
Figure 8. Plasmid shuffle. The chromosomal copy of *YFG1* is replaced by the *yfg1-Δ* deletion, but the Yfg$^+$ phenotype is maintained by the YCp plasmid containing *YFG1* and *URA3*. The strain is transformed with a mutagenized *LEU2* plasmids having the *YFG1* gene. Recessive *yfg1-x* mutations are manifested by selecting for strains on FOA medium. The strain will not grow on FOA medium if *YFG1* is an essential gene and if the *yfg1-x* mutation is not functional.

Although the two-step gene replacement procedure (Section 11.3) could be used to generate condition *yfg1-x* mutations of an essential gene, a method that more clearly reveals the nature of the *yfg1-x* mutation has been developed, the so-called "plasmid shuffle" procedure as illustrated in Figure 8.

A haploid strain is first prepared, which contains a YCp plasmid with the *URA3$^+$* and *YFG1$^+$* gene and in which the chromosomal copy of the *YFG1$^+$* gene has been deleted or disrupted (*yfg1-Δ*). Such a strain could be prepared by transforming a diploid strain that is hemizygous (*YFG1$^+$/yfg1-Δ*) and choosing the appropriate meiotic segregant with the plasmid.

The *YFG1$^+$* gene on YCp-*LEU2* plasmid, for example, is mutagenized and the resulting p[*yfg1-x LEU2*] plasmids containing the *yfg1-x* alterations are introduced into the *yfg1-Δ* p[*YFG1 URA3*] strain by transformation. Test on FOA medium then can be used to determine the nature of the *yfg1-x* mutation. If the *yfg1-x* allele was not altered or was completely functional, the p(*YFG1 URA3*) plasmid can be lost without preventing growth of the strain, which would appear to be FOA resistant (see Section 10.1, *URA3* and *LYS2*). On the other hand, if the *yfg1-x* allele

was completely nonfunctional, the strain will not grow on FOA medium. Furthermore, if the *yfg1-x* allele was conditional, the growth on FOA medium would correspond to the condition. For example, a temperature-sensitive mutation would be revealed by growth of the replica-plated colony on FOA medium at 22°C but not 37°C; whereas the strain would grow at both temperatures on complete medium lacking FOA.

The major disadvantage of the plasmid shuffle procedure, and other procedures using YCp plasmid, is that the copy number varies from one to three or possibly more copies per cell. Thus, different phenotypes may arise because of the different levels of expression of the altered gene product. Overexpression of some altered alleles can produce a nearly wild-type phenotype although a single copy produces a mutant phenotype. A more exact evaluation of a mutant allele may require the integration of a single copy with a YIp plasmid.

Several variations of the plasmid shuffle procedure have been developed, and these rely on the production of a red pigment by certain adenine mutations. Strains having mutations in the *ADE2* gene (for example, *ade2-1*) accumulate a red pigment and form red colonies. However, if the strain also contains the *ade3-Δ* deletion, then the adenine biosynthetic pathway is blocked before the step encoded by *ADE2* and the strain is white (see Section 10.2, *ADE1* and *ADE2*). If *ADE3* is carried by a *YFG1*⁺ plasmid, then the *ade2-1 ade3-Δ* strains are red, but produce white sectors when the *ADE3* plasmid is lost. The procedure does not require replica-plating and is useful for detecting rare events. However, there are numerous other mutations that also produce white colonies, including ρ⁻ mutations, resulting in relatively high numbers of false positives.

11.6 Recovering Mutant Alleles

A convenient method for recovering chromosomal mutations involves transformation with gapped YCp plasmids, as illustrated in Figure 5A. A double-stranded gap is produced by cleavage at two restriction sites within the cloned segment. The gapped plasmid is then used to transform a strain containing the desired mutation that is encompassed in the chromosomal region corresponding to the gap. The gapped plasmid is repaired with the homologous chromosomal region, resulting in the capture of the *yfg1-1* mutation by the plasmid. The gapped plasmid is preferentially maintained, because only the circular form is replicated. The plasmid with the *yfg1-1* mutation can be subsequently recovered by transforming *E. coli* with DNA from the yeast strain. As little as 100 base-pairs of homology on either side of the gap is sufficient to allow gap repair, although larger regions increase the efficiency of the process.

12   Interaction of Genes

Yeast genetics has been particularly amenable for identifying and characterizing gene products that directly or indirectly interact with each other, especially when two mutations alleviate or enhance each otherïs defects. Information on a protein sometimes can be inferred simply by examining the phenotypes of haploid and diploid strains containing two or more mutations. In addition, these genetic properties can be used for isolating novel genes whose products interact. The genetic terms used to denote the interaction of genes are summarized in Table 6, using *YFG1*⁺, etc., as hypothetical examples.

12.1 Heterozygosity and Dominant-negative Mutations

When two recessive mutants are crossed in a standard complementation test, the phenotype of the resulting diploid strain usually reveals if the two mutations are allelic and encode the same gene product. For example, if the *yfg1-1* and *yfg1-2* mutations produce inactive Yfg1 proteins,

Table 6. Interactions of *YFG* (**Y**our **F**avorite **G**ene) genes

Genotype	Ploidy	Pheno-type	Description
YFG1$^+$	1n	Yfg$^+$	Wild-type dominant allele
yfg1-1	1n	Yfg$^-$	Nonfunctional, recessive mutation
$\dfrac{YFG1^+}{yfg1\text{-}1}$	2n	Yfg$^+$	Heterozygous diploid
$\dfrac{yfg1\text{-}1}{yfg1\text{-}2}$	2n	Yfg$^-$	Heteroallelic diploid
$\dfrac{yfg1\text{-}1}{yfg1\text{-}\Delta1}$	2n	Yfg$^-$	Hemizygous diploid
$\dfrac{YFG1^+}{yfg1\text{-}4}$	2n	Yfg$^\pm$	Dominant-negative *yfg1-4* mutation
YFG1$^+$ p[*yfg1-4*]$_N$	1n	Yfg$^-$	Dominant-negative overexpressed *yfg1-4* mutation
$\dfrac{yfg1\text{-}1}{yfg1\text{-}3}$	2n	Yfg$^\pm$	Intragenic complementation
$\dfrac{YFG1^+ \quad yfg2\text{-}1}{yfg1\text{-}1 \quad YFG2^+}$	2n	Yfg$^+$	Double heterozygous diploid
$\dfrac{YFG1^+ \quad yfg3\text{-}1}{yfg1\text{-}1 \quad YFG3^+}$	2n	Yfg$^\pm$	Non-complementation of a double heterozygous diploid
suy1-1	1n	Yfg$^+$	Suppressor of *yfg1-1*
yfg1-1 suy1-1	1n	Yfg$^+$	Suppression of *yfg1-1* by *suy1-1*
yfg1-1 p[*YFG2*$^+$]$_N$	1n	Yfg$^+$	Suppression of *yfg1-1* by overexpression of *YFG2*$^+$
yfg1-1 P$_{GAL1}$-*YFG2*$^+$	1n	Yfg$^+$	Suppression of *yfg1-1* by overexpression of *YFG2*$^+$
yfg1-4	1n	Yfg$^\pm$	Partially functional mutation of *YFG1*$^+$
yfg2-2	1n	Yfg$^\pm$	Partially functional mutation of *YFG2*$^+$
yfg1-4 yfg2-2	1n	Yfg$^-$	Synthetic enhancement

the diploid cross will be Yfg$^-$. On the other hand, if the two recessive mutations, *yfg1-1* and *yfg2-1*, are in two different genes, encoding two different polypeptide chains, then the diploid cross, *yfg1-1 YFG2*$^+$ x *YFG1*$^+$ *yfg2-1* would be Yfg$^+$, because both Yfg1p and Yfg2p are produced by the wild-type alleles in the doubly heterozygous diploid strain.

As expected, mutations that inactivate a function are usually recessive. However, rare nonfunctional mutations can be dominant. Such dominant-negative mutations are particularly important because they can be used to identify nonfunctional forms of the protein that retain their proper structure and associate with other cellular components.

As illustrated in Table 6, dominant-negative mutations can be revealed either by overexpressing the mutation in a haploid (or diploid) strain, such as *YFG1$^+$* p[*yfg1-4*]$_N$, or by a single copy in heterozygous strains, such as *YFG1$^+$/yfg1-4*. Most studies use multicopy YEp plasmids for overexpressing mutations to uncover dominant-negative mutations. Similarly, the P$_{GAL1}$ promoter fused to mutant alleles, P$_{GAL1}$-*yfg1-4*, could be used for the controlled overexpression in tests for dominant-negative mutations (see Section 10.3, *GAL1* promoter).

Dominant-negative mutations act by a variety of mechanisms. For example, a mutationally-altered transcriptional activator that retains DNA-binding activity, but lacks the ability to transactivate, could complex with the DNA-binding sites and displace the wild-type protein. While most recessive missense mutations produce an overall misfolding of proteins, dominant-negative mutations retain at least portions of the structure, thus revealing specific critical regions.

Dominant-negative mutations can also act in heterozygous diploid strains with one copy of each allele. Such mutant proteins generally have a higher than normal affinity for a cellular component, and displace the wild-type protein. For example, numerous nonfunctional *CYC7* mutations were at least partially dominant because the altered forms of cytochrome *c* were arrested at one of the steps in mitochondrial import or heme attachment, and prevented entry of the normal form.

12.2 Intragenic Complementation

One common exception in which heteroallelic diploid have a wild-type or near wild type phenotype is intragenic complementation (also denoted allelic or intracistronic complementation) (Table 6). If large numbers of pairwise crosses of independent mutations of a gene are analyzed, complex complementation patterns are often encountered, with some alleles showing complementation while others do not. For example, a *yfg1-Δ* deletion would not show intragenic complementation with other *yfg1-1* and *yfg1-3*, although the *yfg1-1/yfg1-3* cross could. Also, intragenic complementation does not always restore the activity to the normal level and heteroallelic diploid strains even can have conditional phenotypes.

There are at least two mechanisms for intragenic complementation, one involving proteins with two or more functional domains, and the other involving proteins composed of two or more identical polypeptide chains.

If a protein has two or more functional domains that act independently, then a missense mutation (an amino acid replacement) could inactivate one domain without greatly effecting the others. Thus, partial or complete restoration could be observed in crosses of two such mutations, when each affected a different functional domain. As expected, intragenic complementation of this type is often observed with missense mutations, not with deletions, and only with special subsets of nonsense mutations. Numerous examples of intragenic complementation of this type occurs with genes encoding amino acid biosynthetic enzymes, including *HIS4*, which has three functional domains.

Intragenic complementation is also observed with genes encoding proteins that are composed of two or more identical polypeptide chains. A haploid mutant could produce an abnormal polypeptide that assembles into an inactive homomultimeric protein. On the other

hand, if two different abnormal polypeptide chains are produced in a heteroallelic diploid strain, the abnormal peptides could assemble in certain combination to produce a catalytically active multimeric protein. In such cases, the abnormal polypeptide chains in some way mutually compensates for each otherïs defect. Intragenic complementation of genes encoding multimeric proteins is surprisingly frequent among missense mutations of certain genes. For example, 5 out of 10 genes controlling histidine biosynthesis show extensive complementation, and the nature of the genetic complementation maps suggest that multimeric proteins may be involved.

12.3 Non-allelic Non-complementation

There are rare exceptions to the complementation of non-allelic genes, and these exceptions are denoted "non-allelic non-complementation" or "unlinked non-complementation". As illustrated in Table 6, certain recessive mutations of two different genes, *yfg1-1* and *yfg3-1* fail to complement in the doubly heterozygous diploid, $YFG1^+/yfg1-1$ $yfg3-1/YFG3^+$, even though they are clearly functional in each of the singly heterozygous diploid strains $YFG1^+/yfg1-1$ and $YFG3^+/yfg3-1$. Non-complementation of recessive non-allelic mutations is only rarely encountered and may be the property of only certain proteins.

Several explanations can account for non-allelic non-complementation. One can consider two proteins Yfg1p and Yfg3p, that carrying out related functions, and that are at near limiting concentration in the cell. In normal strains, both Yfg1p and Yfg2p are, by definition, at the normal 100% level, producing 100 units of each of the hypothetical proteins, with a total of 200 units; in singly heterozygous strains, the total would be 150 units, whereas, in doubly heterozygous strains the total would be 100 units. If the total level of 100 units for both proteins is below a critical threshold, than a mutant phenotype would be manifested in the doubly heterozygous strains. Non-complementation of non-allelic gene has been observed for mutant gene encoding cytoskeletal proteins whose cellular concentrations are critical for normal growth.

The lack of complementation of non-allelic genes has also been explained by the formation of inactive heteroligomers or protein complexes, in which, for example, both altered Yfg1p and Yfg3p independently assemble and inactivate the same protein complex, reducing its level below a critical concentration. Also, although recessive by themselves, these mutations may act like dominant-negative mutation when in the doubly heterozygous condition, and encode abnormal proteins that actively compete or replace wild-type subunits.

12.4 Suppressors

A suppressor is generally defined as a mutation that completely or partially restores the mutant phenotype of another mutation. In the example given in the bottom of Table 6, the Yfg⁻ phenotype of *yfg1-1* mutation is restored by the suppressor *suy1-1*. Suppressors can either have or not have a phenotype by themselves.

Suppressors an be broadly assigned to two major groups, informational suppressors and metabolic suppressors. Informational suppressors encode either altered tRNAs or other components of the translational machinery, and act by misreading mRNAs. For example, the nonsense suppressor, *SUP4*-**o** encodes an altered tyrosine tRNA that inefficiently inserts tyrosine residues at UAA chain terminating codons because of an altered anticodon. Another class of informational suppressors are the so-called omnipotent suppressors which cause ribosomal misreading because of alterations in any one of a number of proteins of the translational apparatus, including ribosomal proteins, elongation factors and release factors. Informational

suppressors characteristically act on certain mutations of most, if not all genes, i.e., they are allele specific but not gene specific.

On the other hand, metabolic suppressors usually act on genes common to the same pathway or to a single metabolic function. While there are many possible mechanisms for metabolic suppression, these suppressors are now routinely isolated and investigated as a means for identifying novel gene products of a pathway and for identifying proteins that directly interact with each other.

The suppressors that act by direct physical interactions between two mutant proteins are expected to be both gene and allele specific, i.e., the suppressors should act on only one or a few genes, and on only a restricted subclass of mutants. For example, a temperature-sensitive mutation of actin gene was used to identify a suppressor that was subsequently shown to encode an actin-binding protein.

In contrast, the suppressors that act indirectly on the same pathway are expected to be gene specific, but allele non-specific, i.e., the suppressor could act on any allele, including null mutations, of a specific group of genes. Some of these are denoted as "bypass" suppressor if they replace the function of the initial mutation. For example, $cyc1$-Δ mutations, which cause the deficiency of the major isocytochrome c, can be replaced by $CYC7$-H mutations, which act as suppressors by overproducing the minor isocytochrome c, and allows growth on lactate medium. Another mechanism for pathway suppression involves the loss of regulation of successive steps, such as is encountered, for example, in amino acid biosynthesis or the signal transduction pathways. If an upstream gene, that normally causes the activation of a downstream gene, is destroyed by mutation, then mutant forms of the downstream gene can act as suppressors by rendering it independent of signalling from the upstream gene.

Another often-used means for uncovering interacting components of common functions is suppression by overexpression of wild-type alleles. As in the case for suppression caused by mutation, suppression caused by overexpression can occur by many different mechanisms. Overexpression can be brought about with YEp multicopy plasmids, denoted as $yfg1$-1 p[$YFG2^+$]$_N$ in Table 6; or by fusion to a strong inducible promoter such as P$_{GAL1}$, denoted as $yfg1$-1 P$_{GAL1}$-$YFG2^+$ in Table 6. Furthermore, P$_{GAL1}$ has the advantage that the activity can be conveniently turned off by the addition of glucose to the galactose containing medium, resulting in the loss of suppression (see Section 10.3, $GAL1$ Promoter).

12.5 Synthetic Enhancement and Epistatic Relationships

In some instances, the combination of two different mutant genes in a haploid strain can enhance the severity of the phenotype more than when either of the mutant genes are by themselves. This exacerbation by the combination of two genes, illustrated in the bottom of Table 6 with the $yfg1$-4 and $yfg2$-2 alleles, have been denoted by a variety of terms in the early literature of genetics; currently, this phenomenon is denoted "synthetic lethality" when cell growth is involved, or more generally, as "synthetic enhancement" or "synthetic phenotypes". For example, two different mutant genes individually could cause temperature-sensitive growth, but the double mutant would be inviable at any temperature.

Synthetic enhancement is conceptually identical to epistatic relationships. If two mutant genes in a haploid condition confers a phenotype that is quantitatively identical as that conferred by each of the single mutant genes alone, the two genes are defined as being epistatic with respect to one another. If, however, the doubly mutant strain has an enhanced phenotype, the two genes are defined as being in two separate epistasis groups. For example, detailed epistatic

relationships of over thirty UV or ionizing radiation sensitive mutants has revealed three non-overlapping epistasis groups.
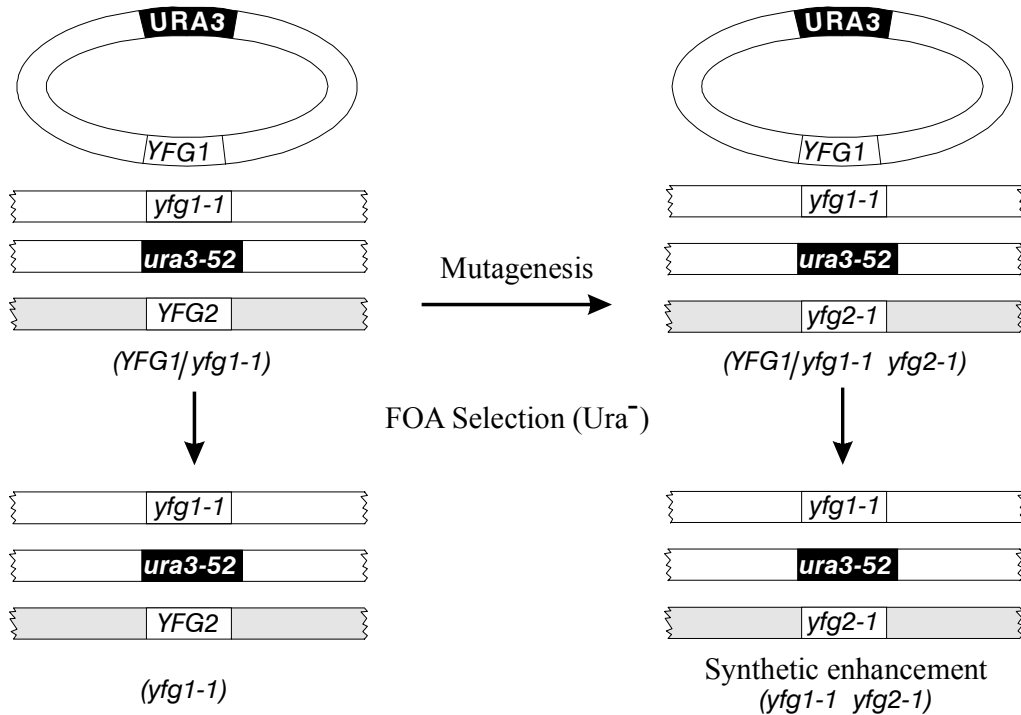

Figure 9. Synthetic enhancement and synthetic lethality. Novel chromosomal genes mutations, *yfg2-1*, etc., that enhance the phenotype of a *yfg1-1* mutations can be uncovered by mutagenizing a *yfg1-1* strain containing a YCp plasmid p[*YFG1 URA3*], and subsequently selecting for the loss of the plasmid. If *yfg1-1 yfg2-1* strain are inviable, they will not grow on FOA medium.

As with suppression, synthetic enhancement can be caused by a number of mechanisms, but often the genes are associated with parallel or related pathways controlling the same function, and some encode proteins that physically interact with each other.

There are numerous examples when synthetic lethality arise when two proteins are functionally redundant, such that neither nonfunctional form cause inviability, but both together are inviable. Synthetic enhancement of redundant genes is illustrated with strains having deletions of either of the *CYC1* or *CYC7* genes encoding, respectively, the two isocytochromes *c*; single deletion strains are respiratory competent, while strains carrying both deletions are respiratory deficient. Synthetic enhancement can be indirect, as exemplified with mutations of *ARG4*, encoding an arginine biosynthetic enzyme and *CAN1*, encoding the arginine permease; the *arg4 can1* double mutant does not grow on synthetic medium containing arginine, while each of the *arg4* and *can1* single mutants do grow.

As with suppressors, genes causing synthetic enhancement are commonly isolated as means to identify genes in the same or related pathways and ones encoding interacting proteins. Several methods have been devised for this aim, and one method is illustrated in Figure 9. A *yfg1-1 ura3-52* strain, carrying a YCp plasmid, p[*YFG1*$^+$ *URA3*$^+$], is mutagenized and tested on FOA medium. A mutation, *yfg2-1*, causing synthetic enhancement is manifested in the Ura⁻ strain because of the loss of the *YFG1*$^+$ allele; in contrast, the other Ura⁻ strains would show only the *yfg1-1* phenotype. If the *yfg2-1* mutation causes synthetic lethality and the *yfg1-1 yfg2-*

1 double mutant is inviable, the colony would not grow on FOA medium (see Section 10.1, *URA3* and *LYS2*).

Other methods have been developed for isolating mutations that confer synthetic lethality, similar to the detection schemes used with the plasmid shuffle procedures (Section 11.5). One of these other methods relies on the *ade2 ade3* host mutations and a *YFG1 ADE3* plasmid, that results in red sectoring when lost (see Section 10.2, *ADE1* and *ADE2*). Another method is based on fusion of the *GAL1* promoter, and screening P$_{GAL1}$-*YFG1* colonies for those that grow on galactose but not on glucose medium, presumably because of the presence of the desired *yfg2-1* mutations (see Section 10.3, *GAL1* Promoter).

13   Genomic analysis

Many diverse studies require the determination of the abundance of large numbers of specific DNA or RNA molecules in complex mixtures, including, for example, the determination of the changes in mRNA levels of many genes. While a number of techniques have been used to estimate the relative abundance of two or more sets of mRNA, such as differential screening of cDNA libraries, subtractive hybridization, and differential display, far more superior methods have been recently developed that are particularly amenable to organisms whose entire genome sequences are known, such as *S. cerevisiae*. It is now practicable to investigate changes of mRNA levels of all yeast ORFs in one experiment.

Numerous companies and academic groups have developed novel approaches to DNA sample preparation, probe synthesis, target labeling and readout of arrays. The following procedures have been successfully used for determining mRNA levels in yeast: (*i*) the DNA Microarray System; (*ii*) the Oligonucleotide Microarray System; (*iii*) the Low-density DNA Array System; and (*iv*) the kRT-PCR System.

The DNA Microarray System. As a general means to address such problems as the differential expression of an entire genome, Brown (14) and his colleagues developed a system for making microarrays of DNA samples on glass slides, probing the DNA micro-spots by hybridization with fluorescent-labeled probes, and using a laser-scanning microscope to detect the fluorescent signals corresponding to hybridization. Fluorescent labeling allows for simultaneous hybridization and separate detection of the hybridization signal from two probes, thus allowing accurate determinations of the relative abundance of specific sequences in two complex samples. For example, with the DNA Microarray System, the entire 6,400 ORFs of the yeast *S. cerevisiae* can be placed on one slide. The mRNA levels of all ORFs (open reading frames) can be determined after, for example, metabolic shifts or in strains deleted for a single gene (15).

The use of DNA Microarray System requires the following basic steps for investigating gene expression of two related cell types or conditions: (*i*) preparation of the large set of DNA elements, usually consisting of ORFs amplified by PCR with sets of primer pairs specific for each ORF; (*ii*) preparation of DNA microarrays consisting of these ORFs spotted on glass slides by a robotic printing device, the Arrayer; (*iii*) preparation of two related mRNAs derived from cells that differ in the trait that is to be investigated; (*iv*) preparation from the mRNAs of fluorescently labeled cDNA by reverse transcription in the presence of Cy3 (green) or Cy5 (red) labeled dUTP; (*v*) hybridization of the fluorescently labeled cDNAs to the ORFs of the DNA microarrays printed on the glass slide; and (*vi*) quantitative analysis of the relative abundance of the mRNAs from the degree of hybridization, using the Scanner.

The DNA Microarray System requires two machines, the Arrayer and the Scanner. The Arrayer prints DNA samples robotically onto a glass slide. After hybridization, the Scanner analyzes the two colored fluorescence of the array with a specially designed scanning confocal microscope.

The Oligonucleotide Microarray System. Modern photolithographic techniques are being used to generate miniaturized arrays of densely packed oligonucleotide probes. These oligonucleotide microarrays, or DNA chips, can then be used, for example, for comparing two sets of cDNAs, similar to the DNA Microarray System described above. Also, as with the DNA Microarray System, the hybridization pattern of fluorescently labeled cDNAs is detected by epifluorescence microscopy. This technology was developed by the Affymax Research Institute (16), and the oligonucleotide microarrays can be produced only by the company.

Using this technology, approximately 6,200 or almost all ORFs of the entire *S. cerevisiae* genome were probed for differential mRNA expression in two growth conditions (16). Approximately 12% of the ORFs, or approximately 750, showed appreciable differential expression rates in the two growth conditions.

The Low-density DNA Array System. In addition to dotting on glass at high densities, as described above for the DNA Microarray System, sets of PCR-generated ORFs can be spotted on nylon membranes, and hybridization to the dot blots can be carried out with ^{32}P-labeled cDNA for detection of the relative levels of mRNAs. In some procedures, 1,536 dots were prepared on a single 9 x 13 cm positively charged nylon membrane. The lower density arrays on nylon membranes can be achieved by automatic robotic and hand-held devices. No special equipment is required to process the membranes once they are prepared.

The kRT-PCR system. In addition to screening for the differential expression of genes by the methods described above, many investigations also require more accurate determinations of large number of mRNA levels. While in the past, Northern blot analysis has been by far the most common means to quantitate mRNA levels, new methods have been developed that are capable of producing more accurate estimations of large number of samples. Some of these are based on the automatic monitoring of reverse transcriptase initiated PCR (or kinetic monitored reverse transcriptase initiated PCR, "kRT-PCR") (17). Several instruments are commercial-available, such as the LightCycler™ and the ABI Prism™ 7700 Sequence Detection System, and some are under development.

Kang *et al*. (17) have developed and evaluated kinetically monitored, reverse transcriptase initiated PCR (kRT-PCR) as a method for transcript level determinations in total unfractionated yeast cellular RNA. This kRT-PCR method supports fully automated quantitation of transcript abundance using total cellular RNA as template and transcript-specific primer pairs. The sensitivity, accuracy, and reproducibility of kRT-PCR are remarkably high. The results show that the kRT-PCR assay quantitates transcript level differences between two physiological or genetic states within a factor of 20%. Absolute mRNA levels are quantitated by kRT-PCR assay within a factor of two. The results show that kRT-PCR assays readily quantitate *S. cerevisiae* transcripts ranging from 367 to 0.00075 copies per cell, providing sufficiently accurate measurements to investigate subtle genetic interactions.

14 Analyses with Yeast Systems

The accessibility of the yeast genome for genetic manipulation and the available techniques to introduce exogenous DNA into yeast cells has led to the development of methods for analyzing and preparing DNA and proteins not only from yeast itself, but also from other

organisms. For example, many mammalian homologs of yeast genes have been cloned by using heterologous cDNA expression libraries in yeast expression vectors. Also, yeast is being used to investigate the detailed functions of heterologous proteins, such as mammalian transcription factors and nuclear hormone receptor. In fact, like *E. coli*, yeast has become a standard microorganism for carrying out special tasks, some of which are described in this section.

14.1 Two-hybrid Systems

   Powerful methods, denoted two-hybrid systems, have been designed for screening and investigating interacting proteins. Because of the ease of the assay, exploratory two-hybrid screens are usually the first method of choice when information of interacting proteins are desired.
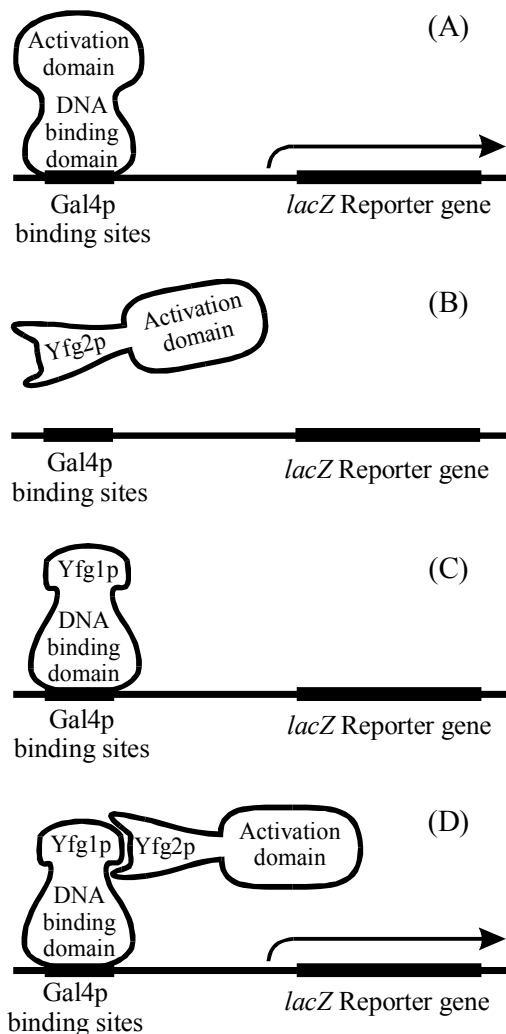

Figure 10. The two-hybrid system. (A) Normally, the Gal4 transcription activator binds to DNA at the Gal4p binding sites and activates transcription of the *lacZ* reporter gene. (B) A hybrid of the Gal4 activation domain with the Yfg2 protein does not activate transcription because it does not localize at the Gal4 binding site. (C) A hybrid of the Gal4 DNA-binding site domain with the Yfg1 protein does not activate transcription of the reporter gene because of the lack of the transcriptional activation domain. (D) Protein-protein interaction between Yfg1p and Yfg2p reconstitutes Gal4p function and activates transcription of the reporter gene.

   Some of these two-hybrid systems are based on the properties of certain eukaryotic transcription factors, usually Gal4p, that have two separate domains, one for DNA binding and the other for transcriptional activation. While the two domains are normally on the same polypeptide chain, the transcription factor also functions if these two domains are brought together by noncovalent protein-protein interactions. In practice, gene fusions are constructed so that the DNA-binding domain is linked to one protein, Yfg1p, and the activation domain is linked to another protein, Yfg2p, as illustrated in Figure 10. Interactions of Yfg1p and Yfg2p brings the DNA-binding and activation domains close together, leading to the expression of a reporter gene that is regulated by the transcription factor.

Another two-hybrid system is based on the use of the *lexA* repressor protein and the *lexA* operator sequences from *E. coli*. These assays are almost always carried out in yeast, although mammalian cells have been used.

   Yeast plasmid vectors are available, in which the *GAL4* DNA-binding domain and the *GAL4* activation domain are on separate plasmids with convenient restrictions sites and with selectable yeast markers. These plasmids are used in conjunction with reporter yeast strains, in which upstream activation sequences from the *GAL1-GAL10* region are used to promote transcription

of the *E. coli lacZ* gene (the P$_{GAL1}$-*lacZ* reporter gene) (see Section 10.3, *GAL1* Promoter, and Section 10.4, *lacZ* and Other Reporters). Complete or partial genes are fused in frame with the *GAL4* DNA-binding domain and the *GAL4* transcription activation domains. If these two hybrid proteins interact, then the *lacZ* reporter gene is transcribed, leading to the blue color of the strain on medium contain the chromogenic substrate X-gal (Figure 10). In addition, yeast strains having not only the P$_{GAL1}$-*lacZ* but also the P$_{GAL1}$-*HIS3* reporter genes are also available. It is advantageous to select directly for expression of the P$_{GAL1}$-*HIS3* reported gene, followed by a screen for P$_{GAL1}$-*lacZ* expression.

Another version of the two-hybrid system uses the *lexA* operator sequence and the DNA-binding domain from the *E. coli lexA* repressor protein. In this system, the activator domain is a segment of *E. coli* DNA that expresses an acidic peptide, which acts as a transcriptional activator in yeast when fused to a DNA-binding domain. As with the *GAL4* system, *lexA* transcriptional activator also contains a nuclear localization signal that directs the protein into the nucleus. Yeast strains having *lexA* operators upstream of both the *E. coli lacZ* and yeast *LEU2* gene have served as reporter genes.

In addition, epitope tags have been built into the constructs of both the *GAL4* and *lexA* systems, allowing for the detection of expressed hybrid proteins. Although false-positive and false-negative results can be obtained, a substantial number of protein combinations have proved to be successfully uncovered with the two-hybrid system and its use has become widely accepted.

Because of its sensitivity, relatively low-affinity interactions can be detected. Also, the cloned genes encoding proteins that interact with the target protein becomes immediately available when the two-hybrid system is used in a screen with libraries of fused genes.

The two hybrid-system has been mainly used for the following three applications: testing proteins that are believed to interact on the basis of other criteria; defining domains or amino amino acids critical for interactions of proteins that are already known to interact; and screening libraries for proteins that interact with a specfic protein. The two hybrid-system has been successfully used to identify diversed sets of interacting proteins in yeast and mammalian cells, and it has been particularly successful in studies of oncogenes, tumor suppressors, protein kinases, and cell-cycle control. Some examples of interacting proteins uncovered with the two hybrid-system in mammalian cells include Jun and Fos; Ras and the protein kinase Raf; the retinoblastoma protein or p53 and the SV40 large T antigen; and other oncoproteins.

14.2 Yeast Artificial Chromosomes (YACs)

The initial step in the molecular characterization of eukaryotic genomes generally requires cloning of large chromosomal fragments, which is usually carried out by digestion with restriction endonucleases and ligation to specially developed cloning vectors. Usually 200 to 800 kb fragments are cloned as Yeast Artificial Chromosomes (YACs), and 100-200 kb fragments are cloned as Bacterial Artificial Chromosomes (BACs or PACs). The importance of YAC technology has been heightened by the recently developed methods for transferring YACs to cultured cells and to the germline of experimental animals.

YAC cloning systems are based on yeast linear plasmids, denoted YLp, containing homologous or heterologous DNA sequences that function as telomeres (*TEL*) *in vivo*, as well as containing yeast *ARS* (origins of replication) and *CEN* (centromeres) segments. Manipulating YLp linear plasmids *in vitro* is complicated by their inability to be propagated in *E. coli*.

However, specially developed circular YAC vectors have been developed for amplification in *E. coli*. For example, a circular YCp vector, containing a head-to-head dimer of *Tetrahymena* or
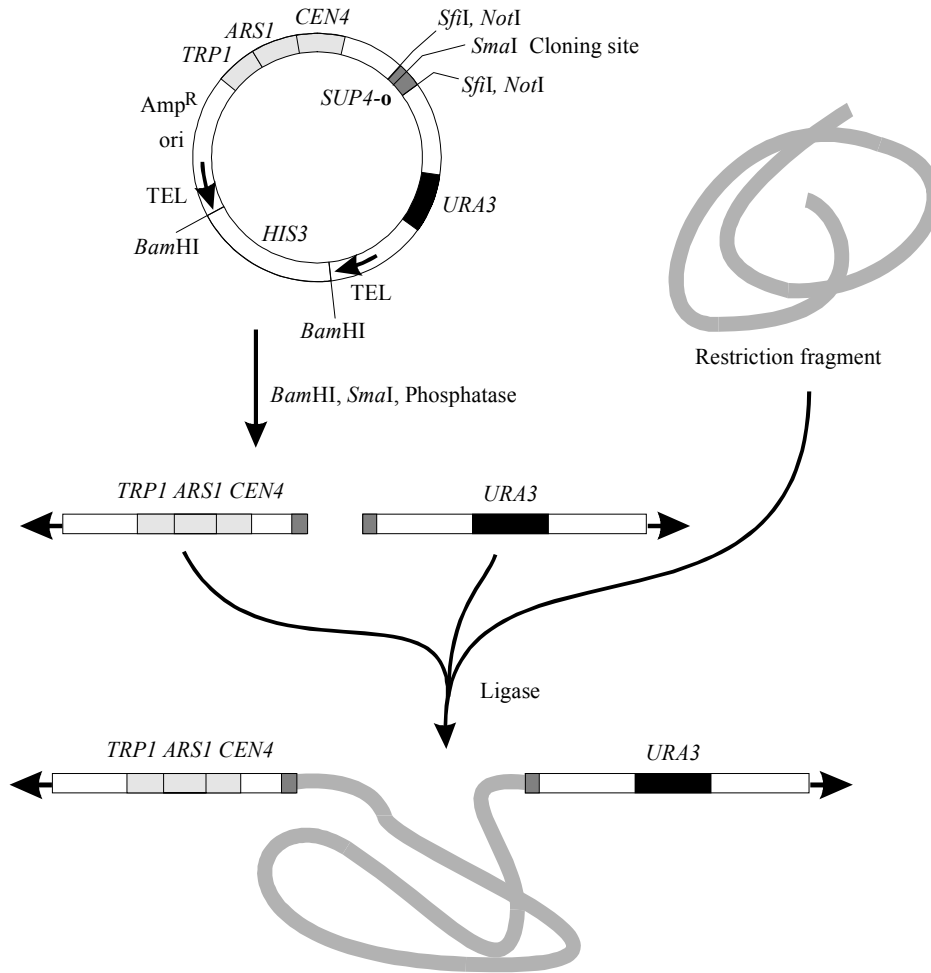

Figure 11. A yeast artificial chromosome (YAC) cloning system. The YAC vector contains telomeric ends that are denoted by black arrow heads. The vector also contains a unique *Sma*I cloning site flanked by *Sfi*I and *Not*I 8-base-pair restriction sites. The vector can be used to clone 50 to 500 kb restriction fragments (see the text).

yeast telomeres, is resolved *in vivo* after yeast transformation into linear molecules with the free ends terminated by functional telomeres. One common type of YAC vector that can be propagated in *E. coli*, contains telomeric sequences in inverted orientation, which flank a DNA cassette containing the *HIS3* gene (Figure 11). After amplification in *E. coli* and before transforming yeast the plasmid is digested with a restriction endonuclease, usually *Bam*HI, which excises the *HIS3* cassette and generates a linear form *in vitro*. Yeast are transformed by this linear structure at high frequencies, although the transformants are unstable. Despite the presence of a *CEN* sequence, the YLp is present at high copy numbers and is lost at high frequency because of its small size. Increasing the size of the YLp by homologous integration *in vivo* or by ligation *in vitro* increases the stability of the plasmid and reduces the copy number to approximately one per cell.

The developed highly-efficient YAC cloning vectors also contain *TRP1* and *URA3* markers and a *SUP4*-**o** gene flanked by the *Not*I and *Sfi*I rare restriction sites as shown in Figure 11. The *SUP4*-**o** suppressor also harbors a naturally occurring *Sma*I site. The host strain contains the *ade2-1* UAA mutation, causing the formation of a red pigment, unless the mutation is suppressed by *SUP4*-**o** (see Section 10.2, *ADE1* and *ADE2*). The YAC vector is cleaved with *Bam*HI and

*Sma*I, treated with alkaline phosphatase and the two arms are ligated to the exogenous DNA fragments desired to be cloned. A *ade2-1 ura3-52 trp1-Δ* host strain is transformed with the
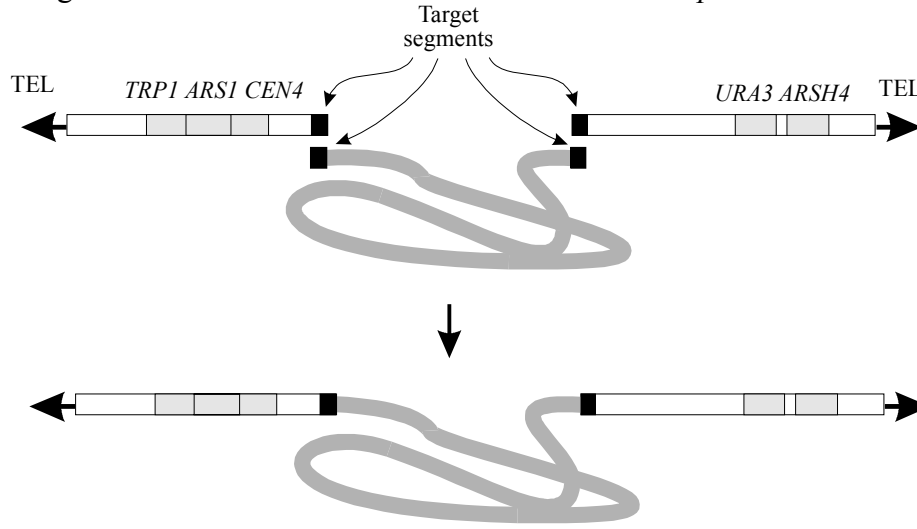

Figure 12. Recombinational targeted cloning with YAC vectors. A yeast strain is transformed with a mixture of the two YAC vector arms and large fragments of DNA. Recombination *in vivo* results in the formation of a specific YAC clone. The two YAC vector arms are derived from linearized plasmids that contain targeting segments that are homologous to the termini of the DNA segment that is to be cloned.

ligated mixture. Both arms are anticipated to be present in Ura$^+$ and Trp$^+$ transformants and inserts should be present in the Ura$^+$ Trp$^+$ Ade$^-$ (red) transformants. YACs present in these transformants are then subjected to pulse-field electrophoresis in order to estimate the size of the inserts.

The potential to use YAC cloning technology has been enhanced by the ability to use homologous recombination for manipulating exogenous DNA in the yeast host. In recombinationally-targeted YAC cloning, YACs are assembled *in vivo*, by recombination, and not by ligation *in vitro*. Recombination takes place between a target segment of the exogenous DNA, and the YAC vector that contain sequences homologous to these targets as illustrated in Figure 12. Transformation of the two YAC vectors arms and the exogenous segment, flanked by the target segments, followed by recombination, results in the formation of the desired stable YACs. The specific target DNA segments for the YAC vector can be obtained from the exogenous DNA as restriction fragments or PCR products.

Also YACs can be modified after cloning by "retrofitting", using homologous recombination with yeast plasmids having targeting sequences. For example, a neomycin resistant gene has to be incorporated into a YAC that will be transferred to mammalian cells using selection, as was done for transfering the entire human β-globin gene in embryonic stem cells. Also, overlapping YAC clones can be recombined, resulting in larger clones encompassing more extensive regions. Furthermore, special YAC vectors have designed for generating terminal and internal deletions of cloned YAC inserts.

YACs have been useful for not only cloning genes but also mammalian telomeric and centromeric regions, and chromosomal origins of replication.

14.3 Expression of Heterologous Proteins in Yeast

Although *E. coli* is still the first choice for the producer of heterologous proteins, the yeast *S. cerevisiae* has some attractive features. Proteins produced in yeast, unlike those produced in *E. coli*, lack endotoxins. In certain special cases, such as hepatitis B core antigen, the products

produced in yeast have higher activity than those produced in *E. coli*. In contrast with using *E. coli*, several posttranslational processing mechanisms available in yeast have allowed the expression of several human or human pathogen-associated proteins with appropriate authentic modifications. Such posttranslational modifications include particle assembly, amino terminal acetylation, myristylation and proteolytic processing. In addition, heterologous proteins secreted from specially engineering strains are correctly cleaved and folding and are easily harvested from yeast culture media. The use of either homologous or heterologous signal peptides has allowed authentic maturation of secreted products by the endogenous yeast apparatus.

The importance of yeast for production of protein products by recombinant DNA methods is illustrated by the fact that the first approved human vaccine, hepatitis B core antigen, and the first food product, rennin, were produced in yeast.

The cloning of specific cDNAs from other organisms and the study of their function using yeast as a surrogate does not necessarily require high-level expression of the foreign protein. In these instances, the aim is just to produce physiological quantities of the protein in a form that is correctly modified and localized in the cell such that the activity accurately reflects the activity in the original organism. However, commercial and laboratory preparations of proteins generally require high expression vectors.

There are numerous varieties of expression vectors currently available for producing heterologous proteins in yeast, and these are derivatives of the YIp, YEp and YCp plasmids described above in Section 9. The cDNA, synthetic DNA or genomic DNA lacking introns are inserted in the vector. Promoters used in expression vectors includes a transcription initiation site and variable amounts of DNA encoding the 5′ untranslated region. Because most of the yeast expression vectors do not contain an ATG in the transcribed region of the promoter, the heterologous gene must provide an ATG that establishes the correct reading frame corresponding to the amino-terminus of the protein. It is essential that this ATG corresponds to the first AUG of the mRNA, because translation almost always initiates at the first AUG on mRNAs from yeast as well as from other eukaryotes. The 5′ untranslated region of the vector also should be similar to the naturally-occurring leader region of abundant mRNAs by lacking secondary structures and being A-rich, and G-deficient, and by having an A at position -3 relative to the ATG translational initiator codon.

Many of the expression vectors include a known signal for 3′ end formation of yeast mRNA, although vectors lacking such defined signals synthesize transcripts until encountering a 3′-end forming signal from another gene or a fortuitous signal on the plasmid.

Numerous normal and altered yeast promoters have been used, and these are chosen because of their high activity and some times because of their regulatory properties. Some of the promoters have been derived from genes encoding alcohol dehydrogenase I, enolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, triose phosphate isomerase, galacokinase (P_{GAL1}, see Section 10.3, *GAL1* Promoter), repressible acid phosphatase, α mating factor, etc. These promoters almost always produce high levels of transcription of heterologous gene, but there is a wide variation in the amount of the corresponding proteins that is finally produced in the yeast strain, depending on the specific heterologous gene.

Bibliography

1. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M.: *Molecular Biology of the Gene*, Chapters 18 and 19, Benjamin/Cummings Pub., Menlo Park, CA (1987).

2. Broach, J. R., Jones, E. W. and Pringle, J. R. (Eds.): *The Molecular and Cellular Biology of the Yeast* Saccharomyces*, Vol. 1. Genome Dynamics, Protein Synthesis, and Energetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1991).

3. Jones, E. W., Pringle, J. R. and Broach, J. R. (Eds.): *The Molecular and Cellular Biology of the Yeast* Saccharomyces, *Vol. 2. Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1992).

4. Pringle, J. R., Broach, J. R. and Jones, E. W. (Eds.): *The Molecular and Cellular Biology of the Yeast* Saccharomyces*, Vol. 3. Cell cycle and Cell Biology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1997).

5. Guthrie, C. and Fink, G. R. (Eds.): *Methods in Enzymology, Vol. 194, Guide to Yeast Genetics and Molecular Biology*. Acad. Press, NY, (1991).

6. Wheals, A. E., Rose, A. H. and Harrison, J. S. (Eds.): *The Yeast, Volume 6, Second edition, Yeast Genetics*. Academic Press, London (1995).

7. Johnston, J. R. (Ed.): *Molecular Genetics of Yeast, a Practical Approach*. IRL Press, Oxford (1994).

8. Fields, S. and Johnson, M. (Eds.): *Yeast Methods for Analyzing Proteins from Other Organisms*. Methods **5**, 77-187 (1993).

9. Heslot, H. and Gaillardin, C.: *Molecular Biology and Genetic Engineering of Yeast*. CRC Press, Boca Raton, FL (1992).

10. Evans, I. H. (Ed.): Yeast Protocols, Methods in Cell and Molecular Biology. Humana Press, Totowa, NJ (1996)

11. Walker, G. M.: Yeast Physiology and Biotechnology. John Wiley and Son, Chichester (1998)

12. Sherman, F.: Getting started with yeast, pp. 3-21. In: Guthrie, C. and Fink, G. R. (Eds.). *Methods in Enzymology, Vol. 194, Guide to Yeast Genetics and Molecular Biology*. Acad. Press, NY, (1991).

13. Hall, M. N. and Linder, P. (Eds.): *The Early Days of Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1992).

14. Brown, P. O. http://cmgm.stanford.edu/pbrown/array.html (1998)

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

16. Wodicka, L., H. Dong, M. Mittmann, M. H. Ho and D. J. Lockhart. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nature Biotechnol. **15**, 1359-1367 (1997).

17. Kang, J. J., R. M. Watson, M. E. Fisher, R. Higuchi, D. H. Gelfand and M. J. Holland Transcript Quantitation in total yeast cellular RNA using kinetic PCR.. Proc. Natl. Acad. Sci. (in press) (1998).

Brown, A. J. P., and M. F. Tuite (Eds): *Yeast Gene Analysis*. Methods in Microbiology, Vol. 26. Acad. Press, NY, (1998).