REGULATION OF CATALASES IN ARABIDOPSIS

C. ROBERTSON MCCCLUNG

Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA

Abstract—The catalase multi-gene family in Arabidopsis includes three genes encoding individual subunits which associate to form at least six isozymes that are readily resolved by non-denaturing gel electrophoresis. CAT1 and CAT3 map to chromosome 1, and CAT2 maps to chromosome 4. The nucleotide and deduced amino acids sequences of the three coding regions are highly related to each other and to other catalases. Both the individual isozymes and the individual subunit mRNAs show distinct patterns of spatial (organ-specific) expression. Six isozymes are detected in flowers and leaves and two are seen in roots. All three mRNAs are highly expressed in inflorescences, and CAT2 and CAT3 are highly expressed in leaves. All three mRNAs are detectable in freshly imbibed seeds, although the pattern of mRNA relative abundance varies among the three genes during early germination. CAT1 and CAT2 mRNA abundance is induced by light. In contrast, CAT3 is negatively light-responsive. CAT2 and CAT3 mRNA abundance is controlled by the circadian clock. Interestingly, the peak in CAT3 mRNA abundance occurs in the subjective evening, which is out of phase with expression of the Arabidopsis CAT2 catalase gene that shows clock-regulated expression gated to the subjective early morning. CAT1 mRNA abundance is not clock-regulated. © 1997 Elsevier Science Inc.

Keywords—Arabidopsis thaliana, Catalase, Circadian rhythm, Free radical, Oxidative stress, Salicylic acid

INTRODUCTION

Catalase (H₂O₂:H₂O₂ oxidoreductase; EC 1.11.1.6) is a tetrameric iron porphyrin protein that catalyzes the dismutation of H₂O₂ to water and oxygen. No multicellular organism has been found which does not possess at least some catalase activity.¹ Catalase is one of several cellular antioxidant defenses [others include superoxide dismutase, peroxidases, glutathione] that provide a defense system for the scavenging of active oxygen species (AOS).² Although all aerobic organisms need a defense system to scavenge oxygen radicals generated in mitochondrial electron transport and the β-oxidation of fatty acids, that need can be magnified in plants. Oilseed plants, such as Arabidopsis, store the majority of their seed carbon and energy reserves as lipids and, during germination, fatty acids are converted into carbohydrates. This process begins with the hydrolysis of stored triglycerides, followed by the β-oxidation of fatty acids in specialized peroxisomes called glyoxysomes.³

Glyoxysomal catalase plays an essential role in breaking down the H₂O₂ generated during β-oxidation. A second major metabolic role for catalase emerges after the development of photosynthetic competence, when peroxisomal catalase degrades H₂O₂ produced during the photorespiratory oxidation of glycolate by another peroxisomal enzyme, glycolate oxidase.³

A growing body of evidence indicates that catalase plays multiple roles in a variety of plant tissues at various developmental stages. Unlike animals, which generally have a single catalase gene [an exception is the nematode Caenorhabditis elegans, in which two different cDNAs encode two distinct catalases⁴], many plants contain multiple catalase isozymes that are encoded by gene families. Small families of catalase genes have been described for castor bean,⁵ N. plumbaginifolia,⁶,⁷ and maize.⁸ This review will summarize recent work characterizing the Arabidopsis catalases, a family of three genes that encode at least six catalase isoforms.⁹ In particular, my lab has focused on the role of the circadian clock in regulating the expression of the Arabidopsis CAT genes.
ARABIDOPSIS AS A SYSTEM FOR GENETIC AND MOLECULAR GENETIC ANALYSIS

Arabidopsis thaliana (L.) Heynh, has emerged as a powerful model system for plant genetic and molecular biological studies. The plant, a self-fertilizing diploid member of the mustard family, is small and hardy, with a rapid generation time (5–6 weeks). The genome is small (~100 Mb) and simple, with small gene families, few and short introns and little repetitive DNA, offering advantages over other plant species for molecular genetic studies. The map position has been assigned for more than 500 of the estimated 20,000 genes. The genetic map of the 5 chromosomes has been integrated with the physical map (see AtDb, Stanford Genomic Resources, http://genome-www.stanford.edu/Arabidopsis), and the completion of the physical map is anticipated within two years. Physical and genetic maps, together with ordered cosmid, bacterial artificial chromosome (BAC), and yeast artificial chromosome (YAC) libraries generated in a number of laboratories provide the necessary tools for chromosomal walking to clone genes identified through mutation. The large insert sizes of the BAC and YAC libraries are particularly useful in the assembly of “contigs,” sets of overlapping clones corresponding to large regions of genomic DNA. At present, a large collection (>28,000 as of July 18, 1996) of Expressed Sequence Tags (ESTs) is available. GenBank (as of 6-11-96) includes 29,689 entries (1,609 are non-EST) of Arabidopsis genes, totalling 14.4 Mb of DNA. The generation of large numbers of transgenic plants by Agrobacterium-mediated transformation is routine.

Arabidopsis offers considerable potential for the study of basic plant processes, including the role of the circadian biological clock in the regulation of gene expression. Initial studies of plant circadian gene expression focused on genes with roles in photosynthesis, most notably the CAB genes, which encode chlorophyll a/b binding proteins, the RBCS genes, which encode the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), and the RCA gene, which encodes Rubisco activase. Because Rubisco is a bifunctional enzyme that initiates the photorespiratory pathway as well as the photosynthetic carbon reduction pathway, we wished to determine whether photorespiratory genes were regulated by the circadian clock.

THE ARABIDOPSIS CATALASE GENE FAMILY

The Arabidopsis catalase gene family includes at least three members. A CAT1 cDNA (GenBank Accession No. U43340) was isolated by differential display analysis of ecotype Landsberg erecta (Ler) leaf and floral tissue. CAT2 (GenBank Accession No. X64271) was isolated from an ecotype Columbia (Col) cDNA expression library using a polyclonal antibody directed against pumpkin catalase. The genomic sequence for CAT2 of ecotype Ler (GenBank Accession No. X94447) has been determined. CAT3 (originally named CAT1) was identified as a randomly sequenced ecotype Col EST and both cDNA and genomic sequences (GenBank Accession No. U43147) have been determined. The nucleotide sequences of the three coding regions are 70–72% identical. The amino acid sequences of the three catalase subunits are 75–84% identical and 87–94% similar. CAT2 maps to chromosome 4 and CAT1 and CAT3 map to chromosome 1. CAT1 and CAT3 are tightly linked (within 40 kb) on a single BAC clone.

At least six catalase isoforms are readily resolved by non-denaturating gel electrophoresis (Figure 1), although we may not yet have resolved all the existing isoforms. One possibility is that this reflects post-translational modifications, such as phosphorylation, of homotetramers of the three CAT subunits. Alternatively, if the three CAT subunits can form heterotetramers, the three Arabidopsis CAT genes could encode 15 distinct isoforms. That fewer than the possible number of catalase heterotetramers are detected could indicate some specificity of subunit interaction. Although each of the CAT genes are expressed in leaves and inflorescences, it is not yet known whether each gene is co-expressed within a single chromosome.

**Fig. 1.** Zymogram analysis of catalase activity in different organs of five week old Arabidopsis plants. Protein extracts (75 μg per lane) were resolved in a 7.5% non-denaturing polyacrylamide gel and stained for catalase activity. The gel was loaded as follows: (R) roots, (L) leaves, (B) bolts, (S) siliques and (F) flowers. Horizontal lines indicate the six catalase isoforms resolved. Reprinted from ref. 9, with permission, American Society of Plant Physiologists.

---

1 J. A. Frugoli; T. L. Thomas; C. R. McClung, unpublished data.
cell type in those organs. Furthermore, it is not yet clear that all three gene products are co-localized in the same sub-cellular compartments (see below). Efforts to correlate individual *Arabidopsis* CAT genes with catalase isozymes are ongoing. We are currently expressing the carboxy-terminal portions (~100 amino acids), which are the most diverged portions of the three CAT proteins, as glutathione S-transferase fusions in *E. coli*, and hope to use the purified fusion proteins as immunogens to raise subunit-specific antibodies. These antibodies would greatly enhance our efforts to examine cell type and subcellular expression patterns of the three CAT subunits, and would allow us to follow the expression of individual CAT subunits.

**CATALASE EXPRESSION**

*Subcellular localization of catalase*

The putative carboxy-terminal peroxisomal targeting sequence (PTS1) Ser/Glu/Cys-Lys/Arg/His-Leu21,22 is located 9 residues from the carboxy terminus in both *CAT2* and *CAT3*. In *CAT1*, this sequence is replaced by Thr-Arg-Leu. This Ser to Thr substitution, although conservative, is sufficient to abolish peroxisomal targeting of firefly luciferase,21,22 suggesting that *CAT1* could be a non-peroxisomal catalase and thus prevented from heteromerization with *CAT2* and *CAT3* by spatial separation. Consistent with this possibility, both maize *Cat1* and *Cat2* sequences contain the conserved Ser-Arg-Leu and are localized to peroxisomes, but maize *Cat3*, in which the Ser is replaced by Thr, is mitochondrial.23 However, each of the *Arabidopsis* CATs also has a sequence (355 Arg-Leu-X5-Gln-Leu363) that matches the conserved consensus amino-terminal peroxisomal targeting sequence (PTS2). Arg-Leu/Ile/Gln-X5-Gln/His-Leu21,22 PTS2 can be cleaved during import but, in yeast, is retained in the mature *Saccharomyces cerevisiae* thiolase and *Hansenula polymorpha* malate dehydrogenase.21,22 It is not known if PTS2 can target a protein into the peroxisome when it is located in a carboxy-terminal position, as it is in the *Arabidopsis* CATs. In addition to PTS1 and PTS2, internal targeting information is used to target *Saccharomyces cerevisiae* catalase A into the peroxisome.21,22

*Organ specificity*

The individual catalase isozymes and the individual subunit mRNAs show distinct spatial (organ-specific) expression patterns. Six isozymes are detected in inflorescences and leaves and two are seen in roots (Fig. 1). mRNA abundance of the three genes varies among organs. Each is highly expressed in inflorescences, and *CAT2* and *CAT3* are highly expressed in leaves.9 Three isozymes are detected in freshly imbibed seeds, and two more isozymes become evident in two-day-old seedlings.5 All three mRNAs are detectable in freshly imbibed seeds, although the pattern of mRNA abundance varies among the three genes during early germination.20,24

*Light-responsiveness*

*CAT1* and *CAT2* mRNA abundance increases upon illumination, although the induction is stronger for *CAT2* than for *CAT1*.5,24 In contrast, *CAT3* mRNA shows a rapid and transient decline in response to illumination of etiolated seedlings.25 These responses are seen with white or with red light, indicating that phytochrome, the red and far red light-responsive photopigment, mediates, at least in part, the observed effects.24,25 Both barley and maize contain light-inducible and light-repressible catalases.23,26 The light-repressible catalases, barley *Cat1* and maize *Cat3*, encode catalases with enhanced peroxidatic activity.26,27 The catalase encoded by *Arabidopsis CAT3* has not been characterized biochemically.

*Regulation by the circadian clock*

Our initial goal in characterizing the *Arabidopsis* CAT gene family was to determine whether the photosynthetic *CAT* gene(s) was clock-regulated, as might be predicted based on the periodic flux through the photorespiratory pathway and the clock-regulation of many genes encoding products involved in photosynthetic light-harvesting and carbon fixation.15 Accordingly, we examined *CAT* expression in plants grown for several weeks in a light–dark cycle and observed robust diurnal oscillations in mRNA abundance of *CAT2* and *CAT3*, but not of *CAT1*.9,20,24 These oscillations in *CAT2* and in *CAT3* mRNA abundance persisted for multiple cycles when plants were released into continuous light (Fig. 2). That robust oscillations in mRNA abundance persist with circadian (~24 h) period for multiple cycles in plants deprived of external time cues demonstrates that both *CAT2* and *CAT3* are regulated by an endogenous circadian clock.20,24 Interestingly, the phases of mRNA abundance for *CAT2* and *CAT3* are distinct. *CAT2* mRNA is most abundant in the early

---

1. J. Frugoli; P. McCourt; C. R. McClung, unpublished observations.
2. E. Connolly; M. Learned; C. R. McClung, unpublished observations.
morning,\textsuperscript{24} whereas CAT3 mRNA peaks in the evening.\textsuperscript{20} To date, two other catalase genes, maize Cat3\textsuperscript{28,29} and N. plumbaginifolia Cat1,\textsuperscript{6} have been shown to be regulated by the circadian clock at the level of mRNA abundance, while four other catalase genes, maize Cat1 and Cat2\textsuperscript{28} and N. plumbaginifolia Cat2 and Cat3\textsuperscript{29} are not clock regulated. \textit{Arabidopsis} CAT2 mRNA abundance increases in anticipation of dawn and is maximally abundant in the early morning and, in this respect, is quite similar in temporal expression pattern to photosynthetic genes such as \textit{CAB}, \textit{RBCS} and \textit{RCA}.\textsuperscript{15,30} This is consistent with a photorespiratory role for \textit{CAT2}. \textit{Arabidopsis} CAT3 and maize Cat3 both show maximal mRNA abundance late in the day, as has been seen with a small number of other plant genes.\textsuperscript{31–33} However, the physiological significance of expression at this circadian phase remains uncertain. Considerable progress has been made in elucidating the cis-acting elements necessary for circadian transcription of the \textit{Arabidopsis} \textit{CAB}\textsubscript{234} and \textit{RCA}\textsubscript{35} genes. Because \textit{CAT3} transcription is gated to a different circadian phase, at least some of the elements of the output pathway from the clock to control \textit{CAT3} transcription must be novel, and it will be interesting to compare the circadian regulation of \textit{CAT3} transcription with that of the morning-specific genes.

What signals are required to initiate circadian rhythmicity? Circadian oscillations in \textit{CAT2} and \textit{CAT3} mRNA are detected in seedlings germinated and grown in continuous light\textsuperscript{23} but not in etiolated (dark-grown) seedlings, although both mRNAs are abundant.\textsuperscript{24,25} In an acute response to illumination, \textit{CAT2} mRNA rapidly accumulates to an initial peak in abundance 4 h after the onset of illumination. Subsequently, circadian oscillations in \textit{CAT2} mRNA abundance are detected, with peaks 24, 48 and 72 h after the onset of illumination.\textsuperscript{24} Similar results were observed with \textit{CAT3} with two important exceptions. The acute response of \textit{CAT3} to illumination is negative; \textit{CAT3} mRNA abundance exhibits a transient decrease immediately following illumination. In addition, the circadian peaks in \textit{CAT3} mRNA occur at a distinct circadian phase about 12 h later than the peaks in \textit{CAT2} mRNA.\textsuperscript{25} Two hypotheses can be proposed to explain the induction of circadian oscillations in \textit{CAT2} and \textit{CAT3} mRNA abundance following illumination of etiolated seedlings. One is that illumination either initiates clock function or synchronizes multiple out of phase oscillators. Alternatively, the clock may be running but either is disconnected from \textit{CAT2} and \textit{CAT3} expression or else the effects of the clock on \textit{CAT2} and \textit{CAT3} expression are masked in extended dark. We are experimentally testing these hypotheses.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{circadian_oscillations.png}
\caption{Circadian oscillations in \textit{CAT2} and \textit{CAT3} mRNA abundance persist for 5 days in continuous light. \textit{Arabidopsis} plants were grown in a 14:10 light-dark cycle for 5 weeks and then transferred into continuous light. Slot blots loaded with total RNA (1 \&mu;g) were hybridized with gene-specific \textit{CAT} probes. For each gene, \textit{CAT} mRNA levels are expressed relative to the highest value; comparison of absolute abundances between genes is not valid. Filled squares indicate \textit{CAT} mRNA in plants grown in a 14:10 light-dark cycle (indicated by the alternating open and filled bars, respectively, beneath the graph). Open circles indicate \textit{CAT} mRNA in plants grown in a 14:10 light-dark cycle and transferred into continuous light (indicated by the open bar beneath the graph). The time since the onset of light in the continuous light treatment is indicated. Data are replotted from 20,24.}
\end{figure}

Identification of photoreceptors required for initiation of circadian rhythmicity

In plants transferred from a light-dark cycle into extended dark, \textit{CAT2} oscillations damp to low-level expression over two or three circadian cycles, as has been seen with \textit{CAB} and other genes.\textsuperscript{15} In striking contrast, \textit{CAT3} mRNA damps to maximal abundance and the oscillation in abundance is obscured.\textsuperscript{25} We asked which photoreceptors were required for this damping to maximal mRNA abundance. Short pulses of white or of red light administered at subjective dawn restore one circadian oscillation, indicating that phytochrome is mediating the response, although other photoreceptors may also be involved. Several mutations (\textit{hy1,hy4}) that disrupt light perception or signal transduction prevent \textit{CAT3} damping and reveal strong circadian oscillations.
in CAT3 mRNA that do not damp over at least three circadian cycles in extended dark.\textsuperscript{25} hyj is impaired in chromophore synthesis and is depleted in spectrally-active phytochrome.\textsuperscript{36,37} hy4 lacks the CRY1 blue-light receptor, cryptochrome.\textsuperscript{37,38} Our results establish that both phytochrome and the CRY1 blue light receptor are required for the accumulation of CAT3 mRNA in extended dark.

**Responses to oxidative stress**

Neither steady state *Arabidopsis* CAT2 mRNA abundance\textsuperscript{39} nor catalase activity\textsuperscript{40,41} increases in response to O\textsubscript{3}-induced oxidative stress. However, neither CAT1 nor CAT3 mRNAs have been measured following oxidative stresses. Furthermore, measures of total catalase activity would be unable to distinguish effects specific to individual isoforms. Therefore, the role for individual *Arabidopsis* catalases in response to oxidative stresses remains unclear. The three catalases of *Nicotiana plumbaginifolia* each respond to exposure to O\textsubscript{3} or SO\textsubscript{2}, and to irradiation with UV-B, with Cat2 showing the most pronounced response.\textsuperscript{42} In *Arabidopsis*, CAT2 mRNA abundance showed a rapid but transient decrease in response to either heat shock (37°C) or to cold treatment (0°C), but CAT3 mRNA abundance was unaffected by either treatment.\textsuperscript{25} Individual catalase mRNAs respond differently to temperature stress (heat or cold) in *N. plumbaginifolia*.\textsuperscript{6} Transient accumulation of H\textsubscript{2}O\textsubscript{2} during cold acclimation of maize seedlings induces antioxidant enzymes, including Cat3, and is correlated with protection against subsequent chilling stress.\textsuperscript{43,44}

**Pathogenesis and systemic acquired resistance**

Increases in H\textsubscript{2}O\textsubscript{2} and other AOS are early events in plant-pathogen interactions and may be important in plant defense responses.\textsuperscript{45,46} Disease resistance can be conferred by elevated levels of H\textsubscript{2}O\textsubscript{2} and this resistance is counteracted by exogenous catalase activity.\textsuperscript{47} Compounds that induce pathogenesis-related (PR) gene expression and resistance directly or indirectly target catalase.\textsuperscript{48} Salicylic acid (SA) plays a key role in signal transduction leading to the development of systemic acquired resistance (SAR) and to the hypersensitive response (HR).\textsuperscript{49–52} One model proposes that salicylic acid (SA) binds to and inactivates catalase, raising the level of H\textsubscript{2}O\textsubscript{2} and other AOS and triggering SAR.\textsuperscript{53} This is supported by the correlation of the *in vivo* ability of analogs of SA to induce PR gene expression, a marker of SAR, with their *in vitro* ability to inhibit catalase activity.\textsuperscript{54} However, others have failed to detect increased H\textsubscript{2}O\textsubscript{2} in tissues expressing SAR and have shown that H\textsubscript{2}O\textsubscript{2} induction of PR genes is mediated through increased SA accumulation.\textsuperscript{55–57} Consistent with this interpretation, H\textsubscript{2}O\textsubscript{2} stimulates SA production in tobacco; the generation of molecular oxygen from catalase dismutation of H\textsubscript{2}O\textsubscript{2} produced during the initial oxidative burst of pathogen response may fuel the production of SA from benzoic acid.\textsuperscript{58} In some plants, catalase gene expression has been shown to respond to SA. Treatment of maize during late embryogenesis with SA induces scutellar accumulation of Catalase\textsubscript{2} transcript and protein\textsuperscript{59} and application of SA to *Solanum tuberosum* induces Cat2St.\textsuperscript{60} In contrast, the catalase genes of *Nicotiana plumbaginifolia* are not induced by SA.\textsuperscript{6} However, response to SA is complex and includes organ-specific and possibly tissue-specific components, as the SA-mediated induction of *S. tuberosum* Cat2St is strong in tuber and stem but weak in roots.\textsuperscript{60} Much remains to be learned about the potentially important roles of catalase and H\textsubscript{2}O\textsubscript{2} in the defense response.

**Catalase Mutants**

Because catalase plays multiple roles in a variety of plant tissues at various developmental stages, and because the multiple catalase isoforms are encoded by members of a multi-gene family, assignment of specific roles to the individuals genes and isoforms is difficult, yet important, and would be facilitated by the characterization of mutations in individual CAT genes. Most of the genes of the photorepiratory pathway of *Arabidopsis* have been identified genetically by conditionally-lethal mutations; mutants die under photorepiratory conditions but survive at elevated CO\textsubscript{2} concentrations under which Rubisco oxygenase activity and, hence, flux through the photosynthetic pathway, are reduced.\textsuperscript{61} Barley mutants lacking catalase exhibit photorepiratory defects and tissue damage consistent with oxidative damage resulting from accumulated H\textsubscript{2}O\textsubscript{2}.\textsuperscript{52,63} However, no *Arabidopsis* mutants lacking catalase have been identified.\textsuperscript{61} Possibly genetic redundancy among the three *Arabidopsis* CAT genes renders the mutational loss of a single gene undetectable by screens employed to date. Additionally, if detoxifying H\textsubscript{2}O\textsubscript{2} produced during germination by fatty acid β-oxidation in oilseed species, such as *Arabidopsis*, is essential, then catalase mutants would die during germination.

A number of new approaches are now available to screen for *Arabidopsis* catalase mutants. The development of a routine system to resolve catalase isozymes and visualize the isozymes by activity staining\textsuperscript{9} puts us in the position to use a brute-force screen.
to examine mutagenized populations for alterations in the wild type pattern of isoynzyme expression. Insertional mutagenesis may allow the isolation of an Arabidopsis CAT mutant. The relative ease and efficiency of Arabidopsis transformation has allowed the generation of more than 25,000 mutant lines by T-DNA insertion, and we are using a PCR-based approach to screen pools of these lines for T-DNA insertions into each of the 3 CAT genes. The maize Ac/Ds transposable element system has emerged as a workable system for insertional mutagenesis. That Ac/Ds tends to transpose to linked sites in the genome allows targeted mutagenesis of any gene linked to one of the large numbers of mapped Ac insertions. Should we not find a T-DNA insertion into each of the CAT genes, we will attempt to isolate CAT insertions by the induction of transposition of a linked Ac insertion. Finally, we are also attempting to either increase or decrease CAT expression through the use of antisense and sense expression constructs. Such efforts in tobacco have recently yielded the intriguing observation that antisense inhibition of Cat1, but not of Cat2, results in conditional lethality in photosynthetic tissues. In: Simic, M. G.; Taylor, K. A.; Ward, J. F.; Von Sonntag, C., eds. Oxygen Radicals in Biology and Medicine. New York: Plenum Press; 1987:651–661.


Chevalier, C.; Yamaguchi, J.; McCourt, P. Nucleotide sequence search Award from the American Cancer Society, Inc., by grants from the United States Department of Agriculture and the National Science Foundation, and by an institutional grant from the American Cancer Society, Inc., administered through the Norris Cotton Cancer Center at Dartmouth.


**ABBREVIATIONS**

AOS — active oxygen species  
AtDb — *Arabidopsis thaliana* data base  
BAC — bacterial artificial chromosome  
CAB — genes which encode chlorophyll a/b binding proteins  
Col — columbia ecotype of *Arabidopsis*  
CRY1 — the cryptochrome blue-light receptor encoded by the *HY4* gene  
EST — expressed sequence tag  
hy — long hypocotyl mutant  
PR — pathogenesis-related  
PTS — peroxisomal targeting sequence  
RBCS — genes which encode the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase  
RCA — the gene which encodes rubisco activase  
Rubisco — ribulose-1,5-bisphosphate carboxylase/oxygenase  
SA — salicylic acid  
SAR — systemic acquired resistance  
T-DNA — transferred DNA, that portion of the *Agrobacterium* tumor-inducing plasmid that is transferred into the plant during infection  
YAC — yeast artificial chromosome