PATTERNS OF GENE EXPRESSION IN MAIZE ENDOSPERM:

CHARACTERIZATION OF THE eEF1A GENE FAMILY

by

Newton Portilho Carneiro

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PLANT SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 8

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

PATTERNS OF GENE EXPRESSION IN MAIZE ENDOSPERM:

CHARACTERIZATION OF THE eEF1A GENE FAMILY

by

Newton Portilho Carneiro

_____

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PLANT SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 8

UMI Number: 9829346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have

read the dissertation prepared by___Newton Portilho Carneiro_____

entitled _____Patterns of gene expression in maize endosperm:_____

_____Characterization of the eEF1A gene family_____

_____

_____

_____

and recommend that it be accepted as fulfilling the dissertation

requirement for the Degree of _____Doctor of Philosophy_____

Brian A. Larkins	4-12-98
Brian A. Larkins	Date
_____	4-12-98
Kenneth A. Feldmann	Date
Gary A. Thompson	4-17-98
Gary A. Thompson	Date
Robert T. Leonard	4-17-98
Robert T. Leonard	Date
Kathryn C. Taylor	4-12-98
Kathryn C. Taylor	Date

Final approval and acceptance of this dissertation is contingent upon
the candidate's submission of the final copy of the dissertation to the
Graduate College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

Brian A. Larkins	4-12-98
Dissertation Director	Date

STATEMENT BY THE AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interest of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _Nestor Portillo Ramirez_

ACKNOWLEDGMENTS

To Andrea and Sarah

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF FIGURES

LIST OF FIGURES - Continued

LIST OF TABLES

ABSTRACT

One of the major challenges for maize improvement is to enhance the protein quality of the endosperm. We do not know what genes encode the proteins that contribute most to the protein quality, but lysine is the most limiting amino acid.

One way to determine the proteins made in the endosperm is by isolating the genes expressed in this tissue and identifying their function. The results of the maize genome project with etiolated seedling and endosperm cDNA libraries support the basis of this strategy, in that the putative functions of a large number of cDNA sequences were identified through sequence similarity comparison and genomic Southern analyses.

One of the genes I selected for further analyses is the elongation factor 1 alpha (eEF1A) gene family. eEF1A interacts with many cellular components and is therefore classified as a multifunction protein. In addition, its level is highly correlated with the amount of protein-bound lysine in maize endosperm. Even though many proteins are known to interact with eEF1A, the basis for the relationship between eEF1A and endosperm lysine content is not known.

Transcript level and sequence of different members of the maize eEF1A gene family were analyzed in this work. All the eEF1A maize genes examined had GTP, aminoacyl tRNA, eEF1B and actin binding domains. The substitution of glutamic acid for aspartic acid in a region that has been shown to be important for actin binding in eEF1Aa, eEF1Ae and eEF1Af suggest that there might be two groups of eEF1A genes that bind differently to actin. Analysis of transcript levels demonstrated that different members of the eEF1A gene family are expressed at different levels in the tissues examined.

The results from the analyses of transcript levels demonstrated that most of the maize eEF1A gene family members are expressed and their mRNAs vary in different tissues and in developing endosperm.

Physiological differences were also determined for the two most abundant eEF1As members by the yeast two-hybrid system. This research provides a significant step toward understanding eEF1A functions in maize and why this protein correlates with the lysine content of the endosperm.

CHAPTER 1

PATTERNS OF GENE EXPRESSION IN MAIZE ENDOSPERM:
CHARACTERIZATION OF THE eEF1A GENE FAMILY

INTRODUCTION

Maize is an important crop that provides a significant amount of protein and energy for human and livestock nutrition. Maize can be grown in diverse environments because of its great genetic variability. Understanding basic developmental, genetic and metabolic processes in maize will increase our ability to genetically manipulate this plant for agronomic improvement. The two major structures of the kernel, the endosperm and the germ (embryo), constitute approximately 80 and 10% of the mature kernel dry weight, respectively (Boyer and Hannah, 1994). One major focus in maize research is to understand the starch and protein accumulation in the endosperm.

The endosperm proteins of maize and most other cereals have a low percentage of several amino acids that are essential for nutrition of humans and other monogastric animals. These amino acids include lysine, threonine, and tryptophan. Lysine is the most limiting amino acid in maize endosperm. However, we know very little about the genetic regulation of storage protein in developing endosperm.

One way to investigate the regulation of development and biosynthesis of starch and protein in maize endosperm is by isolating and analyzing genes expressed in this tissue. In the maize genome project, genes from endosperm were isolated and sequenced, and their sequences compared with known genes from other organisms. Putative functions were assigned to these sequences, and they were made available to researchers as a starting point for studying their roles in maize development and homeostasis. The Maize Genome Database (MaizeDB) can be accessed via http://www.agron.missouri.edu.

One gene family that was of particular interest to us encodes the eukaryotic translation elongation factor 1 alpha (eEF1A). The level of this protein had been shown to correlate with the protein-bound lysine content of maize endosperm (Habben *et al.*, 1995). Consequently, eEF1A makes an excellent marker to develop high lysine maize genotypes in breeding programs. I characterized the eEF1A gene family with the objective of understanding more about its biology and its relationship to the nutritional quality of maize endosperm.

THE MAIZE ENDOSPERM

The maize endosperm is the site of many important biological processes, such as storage protein and starch synthesis. This tissue accounts for 98% of the starch, 80% of the protein and 15% of the lipid in the total kernel (Glover and Mertz, 1987). The endosperm is a triploid tissue which forms as a result of the fusion of the pollen nucleus and the two female polar nuclei of the central cell (Wolf *et al.*, 1952). Maize endosperm development starts with mitotic division of the primary triploid nucleus a few hours after fertilization. Free nuclear division continues until the third day, when the endosperm becomes a cell with 250 nuclei. After this point, cytokinesis starts, followed by mitotic division until 12 days after pollination. Between 12 and 20 days after pollination mitotic division ceases, and the outermost layer surrounding the endosperm becomes the aleurone, except in the area adjacent to the embryo. The basal endosperm cells differentiate into several layers of transfer cells with the function of nutrient conducting cells. By the end of differentiation, this tissue consists of the starchy endosperm, aleurone and the basal transfer cells (Knowles and Phillips, 1988). Despite all the knowledge of endosperm morphology and histo-differentiation, little is known about the mechanisms that control endosperm development. Many developmental aspects, such as the ones leading to proliferation of cell lineage and the definition of cell identity and specialization are not understood.

To understand more about the biology of maize endosperm, many endosperm mutants have been isolated, including those that affect starch and storage protein synthesis. The mature corn kernel is composed of over 70% starch. By far, most (80 to 90%) of this starch is found in the endosperm. Starch is a homopolymer of glucose that serves as a carbohydrate reserve. The molecular structure involves only two linkages of glucose molecules: α-1,4 and α-1,6. Two distinct types of polymers are found in most starches: amylose and amylopectin. Amylose is essentially a linear molecule of α-1,4 linked glucose. Both linear and slightly branched amylose molecules are found in starch. In the branched amylose molecules, the long stretches of linearly linked glucoses are interrupted with a branch (α-1,6 linkage) about every 200 glucose units. In general, amylose molecules are highly dispersed in molecular weight. They range in size from 100 to 1000 glucose subunits, with the larger molecules usually containing a few α-1,6 branch points. Amylopectin molecules are much larger than amylose and contain up to 200,000 glucose subunits. About 4 to 5% of the total glucose units of amylopectin are linked α-1,6.

Although starch looks like a simple polymer, its synthesis is still a subject of intense study. In starch biosynthesis, the enzyme sucrose synthase (UDP-glucose:D-frutose-2-glucosyltransferase) (*Sh1* and *Sus*) reversibly catalyses the conversion of sucrose to fructose and glucose. Both hexoses are then converted to glucose-1-phosphate by UDP-glucose pyrophosphorylase and enzymes of the glycolytic pathway. Glucose-1-phosphate reacts with ATP via ADP-glucose pyrophosphorylase, whose subunits are encoded by the *Shrunken-2* (*Sh2*) and *Brittle-2* (*Bt2*) genes to produce ADP-glucose. ADP-glucose is converted to amylopectin by the starch branching enzyme or to amylose by the starch-granule-bound starch synthase (*waxy*) (Boyer and Hannah, 1994). The *Bt1* locus encodes a protein of unknown function, although it structurally resembles an adenylate carrier protein. The *sh* mutation in maize leads to a shrunken or collapsed endosperm, due to a reduced amount of starch (Chourey and Nelson, 1976). The phenotype of the *bt1* mutant is similar to that of *sh2* and *bt2*, which encode the subunits

of ADP-glucose pyrophosphorylase (Hannah *et al.*, 1993). The waxy gene only reduces the activity of the major starch-granule-bound enzyme.

Protein in the maize endosperm can be divided into four solubility fractions: albumins extracted with water, globulins extracted with low salt solution, prolamin extracted with aqueous alcohol and glutenins extracted with diluted alkali or acid (Osborne and Mendel, 1914). The prolamins in maize are called zeins. Zeins have no known enzymatic or structural function. They can be resolved by SDS-PAGE into four major classes of polypeptides: 19- and 22-kD alpha (α-), 14-kD beta (β-), 16- and 27-kD gamma (γ-) and 10-kD delta (δ-) zeins (Larkins *et al.*, 1989). The α-zeins are the major component of the storage protein (60%), followed by the γ-zeins (25%), β-zeins (5 to 10%) and the δ-zeins (5%). Genes encoding zein proteins have been characterized (reviewed by Feix and Quayle, 1993). The α-zeins are encoded by a multigene family with 50 to 100 members (Heidecker *et al.*, 1991). It is not known whether all the α-zeins genes are active in the synthesis of the zein protein nor if different genes may be expressed at different levels. Three major clusters of α-zein genes were genetically mapped to chromosomes 4, 7 and 10 (Soave and Salamine, 1984). In contrast, β-, γ- and δ-zeins are encoded by only one or two genes located on the long arm of chromosome 6, 7 and 9, respectively (Murray *et al.*, 1988; Benner *et al.*, 1989). Zeins are very low in several essential amino acids such as lysine, tryptophan and threonine (Shotwell and Larkins, 1989). Lysine is the most limiting amino acid (Nelson, 1969). One of the major challenges for maize improvement is to understand the regulation of the storage protein genes and to enhance the protein quality.

A number of loci affecting the expression and accumulation of zein proteins have been identified (Coleman *et al.*, 1996). One mutant with altered zein levels is called *opaque2* (*o2*). *o2* corresponds to a defective leucine zipper-type transcription factor (Hartings *et al.*, 1989; Schimdt *et al.*, 1990) which regulates expression of 22-kD α-zeins (Schimdt *et al.*, 1990). Another mutant, *floury2* (*fl2*), is a semidominant mutation caused by a defective α-zein in which the signal peptide is not cleaved (Lopes *et al.*, 1994;

Coleman *et al.*, 1995). Both *o2* and *fl2* kernels have a soft, starchy endosperm that reduces transmission of light. The *o2* and *fl2* mutants typically have a significantly greater lysine content than the wild type in the endosperm (Mertz *et al.*, 1964; Nelson *et al.*, 1965), but their pleiotropic effects, such as a soft endosperm that results in damaged kernels, increased susceptibility to insects and fungal pests, inferior food processing and generally reduced yield, are not easily overcome (Glover and Mertz, 1987).

A possible way to overcome the negative effects of the *o2* mutation came with the identification of *o2*-modifier genes. *o2* mutants expressing the modifier genes are hard and vitreous, and have the enhanced nutritional quality (lysine content) of *o2* genotypes. The modified *o2* genotypes are designated Quality Protein Maize or QPM (Villegas *et al.*, 1992; Gevers and Lake, 1992).

Transfer of *o2*-modifier genes to *o2* elite genotypes has not been easy. The development of QPM has been difficult due to the necessity of manipulating at least three major components: the *o2* mutation, modifier genes and lysine content. The first component, the *o2* gene, can be easily followed by DNA marker-assisted breeding. The second component, the modifiers genes, have been shown to be associated with the level of 27-kD γ-zein. One of these loci is tightly linked to the duplicated γ-zein gene (Lopes *et al.*, 1995).

Although many studies have been done to understand and modify the levels of some proteins in the maize endosperm to increase its nutritional value, little is known about the genes involved in this process. One way to identify and characterize genes and determine their putative function is by reverse genetics.

FORWARD AND REVERSE GENETICS

Forward genetics is a process that involves the characterization of a particular mutant organism's properties, starting with the phenotype and ending with the genotype. It usually involves the molecular analysis of the mutant genes DNA, RNA and protein

expression. Genes have been isolated by positional cloning (Goodman *et al.*, 1995), insertional mutagenesis with transposons (Aarts *et al.*, 1993) or T-DNA vectors (Feldmann, 1991), library screenings and differential display techniques (Gibson and Somerville, 1993). Now, with the ability to clone, modify and examine the biological activities of DNA segments, this process can be done in reverse, an approach referred to as "reverse genetics." This new approach begins with a gene whose molecular structure is known, and it proceeds to explore the gene's contribution to the organism's phenotype. Thus, the experimental path is from the gene as a nucleotide sequence to the corresponding fuction (Berg, 1993).

Once the sequence of a specific gene is known, its function can be revealed by inactivating the gene by homologous recombination with a defective sequence or down regulation with antisense or cosupression. Homologous recombination often includes incorporation of a marker DNA sequence into the transcription unit of the target gene, allowing detection and survival of the cell following selection. Methods for disrupting genes via homologous recombination have been described for yeast and mice (Melton, 1994; Goffeau *et al.*, 1996). Whereas gene replacement through homologous recombination is now a standard procedure in some organisms, in plants the rates of homologous recombination are low. Homologous recombination has been demonstrated in plants by the disruption of the AGL MADS-box gene in *Arabidopsis* (Kempim *et al.*, 1997).

The most common approach to create loss-of-function alleles for plant genes has been through suppression of gene expression by antisense or sense RNA (cosupression). The process called "antisense RNA regulation" involves introduction into cells of an RNA or single-strand DNA molecule that is complementary to the mRNA of the target gene. One mechanism to explain the mechanism of antisense knockout genes is that the antisense RNA molecule can base pair with the mRNA and prevent its translation into protein. Antisense RNA has been used in plants to inhibit transient expression in electroporated protoplasts (Ecker and Davis, 1986) and to reduce the expression of an

Agrobacterium gene in transformed plants (Rothstein *et al.,* 1987). The constitutive expression of an antisense chalcone synthase (*chs*) gene has been shown to alter flower pigmentation in transgenic petunia and tobacco (van der Krol *et al.,* 1988). Sheehy *et al.* (1988) used antisense RNA to inhibit polygalacturonase during fruit development. Cosupression, on the other hand, appears to be caused by overexpression of an endogenous gene as a consequence of introducing extra copies of an endogenous gene or by introducing a copy of an endogenous gene under the regulation of a strong promoter. The first report of cosupression resulted from studies on the variation of petunia flower coloration after the introduction of a *chs* transgene under the control of a strong 35S promoter (Napoli *et al.,* 1990). The process of gene knock out by cosupression is not fully understood.

Foreign DNA has been used to transform cells, in order to modify an organism's phenotype by either antisense gene expression or gene knock out by homologous recombination. In plants, the *Agrobacterium*-Ti plasmid vector system is very effective for introducing DNA into dicotyledonous plants, but there have been some restrictions for monocots. Numerous reviews of *Agrobacterium*-mediated transformation systems have been published (Fraley *et al.,* 1986; Hooykaas, 1989). The T-DNA is a segment of the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* delimited by short imperfect-repeat border sequences. The T-DNA, including any sequence inserted between the borders, can be transferred by *Agrobacterium* to the plant cell, where it integrates into the genome. In the wild, the T-DNA converts the plant into a factory of producing substances used by *Agrobacterium* for growth. Engineered T-DNAs have been disarmed of the tumor-promoting and opine biosynthetic genes, which are typically replaced with an antibiotic resistance marker to allow the selection of the transformed plant or tissue (Azpiroz-Leehan and Feldmann, 1997).

One way to sidestep the species restriction imposed by *Agrobacterium*'s host range is to introduce DNA directly into cells using physical rather than biological means. DNA uptake can be promoted by various treatments including polyethylene glycol and/or

electroporation (Fromm and Walbot, 1987; Negrutiu *et al.*, 1987; Shillito *et al.*, 1985). Integrative transformation can be very efficient (Negrutiu *et al.*, 1987), and it leads to stable inheritance of the foreign gene (Potrykus *et al.*, 1985). Virtually every plant protoplast system has proven transformable. though with different efficiencies. Unfortunately, there can be severe problems regenerating transgenic plants from protoplasts (Roest and Gilissen. 1989; Potrykus and Shillito, 1986). Transgenic plants have been regenerated from protoplasts of important crop species that can not be transformed with *Agrobacterium*. Several laboratories have reported the transformation of Japonica-type rice (Shimamoto *et al.*, 1989; Toriyama *et al.*, 1988; Zhang *et al.*, 1988), Indica-type rice (Datta *et al.*, 1990) and maize plants (Wang *et al.*, 1996). Rhodes *et al.* (1988) were the first to describe regeneration of maize plants derived from protoplasts into which the neomycin phosphotransferase II gene was introduced via electroporation.

Another alternative for plant transformation is the gene gun, which is a device that delivers minute metal particles coated with DNA directly into cells. The first transgenic soybean plants were created via *Agrobacterium*-mediated transformation and biolistic methods (McCabe *et al.*, 1988). A real breakthrough for biolistics came with the recovery of fertile transgenic maize. Gordon-Kamm *et al.* (1990) transformed embryonic maize suspension cultures with the bacterial "bar" gene, which encodes phosphinothricin acetyltransferase (PAT) and protects against the herbicide phosphinothricin (PPT). The DNA coated particles were shot into embryogenic cells. and the transformed cells were selected by culturing in a medium containing the herbicide PPT. Whole plants regenerated from these cells were resistant to the herbicide when it was applied directly to the leaves.

Transposons and T-DNA tagging can be used for either forward or reverse genetics. Many genes have been isolated using transposons as molecular tags without information on the biochemical properties or expression pattern of genes they affect. The first cloning of a plant gene by transposon tagging was the maize *bronze* gene, which encodes UDP-glucose:flavonoid 3-*O*-glucosyltransferase, a key enzyme of anthocyanin

synthesis (Fedoroff *et al.*, 1984). Afterwards, a large number of genes were isolated using endogenous transposons as tags (Walbot, 1992). Baker *et al.* (1986) demonstrated active transposition of the maize *Ac/Ds* elements in transgenic tobacco. Since then, transposition of maize elements has been demonstrated in other species including *Arabidopsis*, tomato, petunia, flax, carrot, potato, soybean and rice (Izawa *et al.*, 1991). Subsequently, genes affecting leaf morphology and albinism were isolated from *Arabidopsis* using the *Ac/Ds* system (Bancroft *et al.* 1993; Long *et al.*, 1993) and an *Arabidopsis* gene involved in male fertility was cloned by the use of another maize transposon system, *En/Spm* (Aarts *et al.*, 1993).

Reverse genetics using transposon insertion mutants was first described in the fruit fly *Drosophila melanogaster* (Ballinger and Benzer, 1989; O'Hare, 1990). It depends on polymerase chain reaction (PCR), using one DNA primer that corresponds to the end of the transposon (Rosenzweig *et al.*, 1983; Emmons and Yesner, 1984) and another that corresponds to the gene of interest. PCR products are obtained if a transposon is inserted into the gene of interest. Mutant genes are recovered following PCR amplification of DNA from the mutant (Zwaal *et al.*, 1993). For *Drosophila*, this strategy is unfortunately limited by the non-randomness of *P*-element insertion and the difficulty of maintaining large stocks of live flies (Azpiroz-Leehan and Feldmann, 1997). Later, this procedure was applied to *Caenorhabditis elegans* (Rushforth *et al.*, 1993; Plasterk, 1993) and maize (Bensen *et al.*, 1995).

The Trait Utility System for Corn (TUSC) uses the *Mutator* (*Mu*) transposon (Bensen *et al.*, 1995). The *Mu* transposable element family has both a high forward mutation rate and a high degree of conservation for its terminal-inverted-repeat (TIR) sequences (Walbot, 1992). These two features were exploited to develop a resource from which gene-specific *Mu* insertion alleles can be selected for any maize gene whose sequence is known. *Mu* insertion alleles for a selected gene are identified by pairing a gene-specific PCR primer with a bi-directional outward primer specific for the *Mu* TIR. Geometric amplification of the targeted sequence by PCR is indicative of a *Mu* insertion

within the gene of interest. Specific *Mu* insertion alleles can be recovered and propagated from F2 seed that has been archived from over 42,000 mutagenized plants. The consequences of *Mu* insertion on the target gene's expression can be analyzed in the F2 and subsequent generations.

Bensen *et al.* (1995) characterized the *Anther ear1* (*An1*) mutant using the TUSC screen. The *An1* gene product is involved in the synthesis of *ent*-kaurene, the first tetracyclic intermediate in the gibberellin (GA) biosynthetic pathway. Mutation of *An1* results in a GA-responsive phenotype that includes reduced plant height, delayed maturity, and development of perfect flowers on normally pistillate ears. The identity of the cloned gene as *An1* was confirmed by using a reverse genetics screen for maize families that contain a *Mu* element inserted into the cloned gene and by demonstrating that the insertion causes an *an1* phenotype. The predicted amino acid sequence of the *An1* cDNA has homology with plant cyclases and contains a basic N-terminal sequence that may target the *An1* gene product to the chloroplast.

Mena *et al.* (1996) isolated a transposon-induced mutation in *ZAG1*, the maize homolog of *Agamous*. The *Arabidopsis* gene *Agamous* is required for male and female reproductive organ development and for floral determinacy. *ZAG1* mutants exhibited a loss of determinacy, but the identity of reproductive organs was largely unaffected (Mena *et al.*, 1996)

The *Agrobacterium-Ti* vector has been used to tag and identify genes influencing plant morphology, such as height, flowering, morphology and trichome formation (Hooykaas and Schilperoort, 1992). This approach was especially successful for *Arabidopsis thaliana*, where Feldmann and Marks (1987) developed a seed transformation protocol that allowed large numbers of independent T-DNA tagged mutants to be obtained. Using a T-DNA-tagged homeotic mutant, the *Arabidopsis agamous* gene was identified, and found to encode a transcriptional regulator necessary for flower development (Yanofsky *et al.*, 1990)

In *Arabidopsis*, a PCR-based reverse genetics strategy has also been used to identify T-DNA insertion mutants. A population of 20,000 *Arabidopsis* transformants containing 30,000 independent inserts has a 65% probability of finding a mutant in a gene of average length. The first application of this method to *Arabidopsis* was described by McKinney *et al.* (1995). Around 5,300 transformed lines were screened in pools of 100 with primers from the T-DNA borders and degenerate actin sequence primers. McKinney *et al.* (1995) estimated that they would find mutations in two or three of the 10 actin genes (average size of 3 kb for actin genes) in *Arabidopsis* among of 8,000 T-DNA insertions in this tester population. Indeed, McKinney *et al.* (1995) identified null mutants in *ACT2* and *ACT4*. Krysan *et al.* (1996) developed a more general gene tagging procedure based on gene-specific, non-degenerate PCR primers. They directly tested the generality and sensitivity of the method by searching for T-DNA inserts within 63 different genes. Their results demonstrated the method was capable of detecting most, if not all, of the T-DNA inserts expected to be present in the available T-DNA transformed population. In a collection of 9,100 transformed lines, a total of 17 T-DNA insertions within 63 genes were identified (Krysan *et al.*, 1996).

Transposon and T-DNA tagging have also been used in another reverse genetic approach called "gene trapping." Gene trapping relies on the random insertion of a promoterless reporter gene into the genome of the host cell. Clones of cells in which the expression of the reporter is driven by constitutive regulatory elements can easily be recovered by applying the appropriate selection (Evan *et al.*, 1997). Twelve genes have been isolated from *Arabidopsis* using a T-DNA containing a bacterial reporter gene coding for neomycin phosphotransferase II (*npt II*). The DNA sequences of the genes were obtained by analysis of the product of the Rapid Amplification of the CDNA Ends (RACE) PCR with comparison to genes and protein of known function (Babiychuk *et al.*, 1997).

THE USE OF EXPRESSED SEQUENCES TAGS (ESTS) TO STUDY GENE EXPRESSION

Reverse genetics is becoming increasingly important due to the expressed sequence tags (ESTs) that have been described in the EST databases (dbESTs). Several viral and organellar genomes have been completely sequenced. The bacteriophage ΦX174 genome [5386 base pairs (bp)] was the first to be sequenced (Sanger *et al.*, 1977). Sanger *et al.* (1977) were also the first to complete the genome sequence of the λ bacteriophage (48,502 bp). Subsequently, the genome of cytomegalovirus (CMV) (229 Kb) (Bankier *et al.*, 1991), vaccinia (192 kb) (Goebel *et al.*, 1990), the mitochondrial (187 kb) and chloroplast (121 kb) genomes of *Marchantia polymorpha* (Oda *et al.*, 1992), variola (smallpox) (187 kb) (Massung *et al.*, 1993), *Escherichia coli* (Sofia *et al.*, 1994), *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996), *Mycoplasma genitalium* (Fraser *et al.*, 1995) and *Methanococcus jannaschii* (Bult *et al.*, 1996) were sequenced. Gene maps of 141 viruses, 51 organelles, two eubacteria, one archeon and one eukaryote are available (http://www.ncbi.nlm.nih.gov/) (Schultes *et al.*, 1996).

In plants, accumulation of DNA sequence information has been slow compared with other organisms. By mid-1992, DNA sequences were available for less than 500 proteins in higher plants, whereas the number of genes expressed in the lifetime of a plant is estimated to be between 15,000 and 60,000 (Goldberg *et al.*, 1978). Analysis of 830 rice cDNA sequences has been published (Uchimiya *et al.*, 1992). The strategy for this project was based on the construction of cDNA libraries using a bacterial plasmid vector and the generation of double stranded plasmid DNA for dideoxy DNA sequencing. In August 1996, the Rice Genome Research Program (RGP) had isolated and partially sequenced more than 29,000 cDNA clones from various rice tissues, including cell culture. Currently, the numbers of characterized clones from various libraries are as follows: around 2,000 cDNAs from root, 4,900 from green shoot, 4,500 from etiolated shoot, 1,700 from the panicle at flowering stage, 1,700 from panicle at ripening stage,

2,500 from growth-phase callus, 3.000 from gibberelic acid (GA3)-treated callus. About 25% of the clones have significant similarity to known proteins. Some of the identified clones show library-specific distribution, indicating that the composition of the sequences in each library reflects, to some extent, the regulation of gene expression specific to differentiation, growth conditions or environmental stress. Including the unknown clones, nucleotide sequence similarities of 24,728 clones were analyzed, and the clones classified into around 10,000 independent groups (Yamamoto and Sasaki, 1997).

In *Arabidopsis* 1,152 sequences were described from five cDNA libraries prepared from mRNAs expressed during floral developmental, embryogenesis, seed maturation and development of etiolated plants (Hofte *et al.*, 1993). Clones from the oriented libraries were sequenced from the 5' end to target preferentially the cDNA coding region. Sequences were translated in all six reading frames and compared with protein sequence databases using the program FASTA (Pearson and Lipman, 1988) or BLASTX (Altschul *et al.*, 1990), except for a number of cases in which nucleotide sequences were directly compared with DNA data banks. Their analysis allowed identification of 895 non-redundant ESTs of 250-350 bp, 32% of which match with known genes in *Arabidopsis* or other organisms. Many of these correspond to genes implicated in general cellular metabolism which had not yet been identified in plants (Hofte *et al.*, 1993).

In maize, results from a pilot genome project were reported by Keith *et al.* (1993) to determine the percentage of cDNAs that can be functionally identified from a plant cDNA library using partial DNA sequence analysis. From a total of 130 cDNAs sequenced, 7 correspond to previously identified maize genes, 18 had a high degree of sequence similarity to related maize genes or to genes from species other than maize. The remaining 105 clones showed little or no similarity to genes in the databases and were reported as novel genes.

In our maize genome project, we combined sequence comparison, map location and gene expression to improve the identification of putative gene function. The

measurements of mRNA expression was done by hybridizing the library with a probe prepared from total mRNA and categorizing the cDNAs as corresponding to abundantly or rarely expressed mRNAs that are either constitutive or tissue-specific. A total of 313 clones from an endosperm and seedling cDNA libraries were used for 'single-pass' sequencing from the presumed 5' end of the mRNA, and the nucleotide sequence was compared with the GenBank database. The chromosomal location of more than 300 clones was determined by RFLP mapping using a standard maize population.

One gene family identified in the maize genome project that was interesting for further characterization encodes elongation factor 1-alpha (eEF1A). The interesting aspect of this protein is that it interacts with many cellular components and is therefore classified as a multifunction protein. In addition, its level highly correlates with the amount of protein-bound lysine in maize endosperm (Habben *et al.*, 1995). Even though many cellular components have been shown to be related to eEF1A, the basis for the relationship between this protein and lysine content in endosperm is not known.

ELONGATION FACTOR 1 ALPHA (eEF1A)

To identify lysine-rich proteins that are increased in *o2* mutants, a complex polyclonal antiserum was developed against the non-zein endosperm proteins and used to screen an endosperm cDNA expression library (Habben *et al.*, 1993). To determine if a specific mRNA sequence was up-regulated in *o2*. RNA blots of mutant and wild type were probed with cDNAs found in the screen. One of the mRNAs found to be significantly increased in W64A*o2* compared to the wild type is eEF1A (Habben *et al.*, 1993). Subsequently, a 2 to 3-fold increase of this protein was found in *fl2* mutants as well (Sun *et al.*, 1997).

Habben *et al.* (1995) developed an ELISA using an antiserum against a bacterially expressed maize eEF1A. The analysis revealed a 2- to 3-fold increase in eEF1A in *opaque2* genoytpes relative to their normal counterparts. These genotypes showed a 4-

fold range of variation in lysine content and a highly significant relationship ($r^2 = 0.91$) between lysine and eEF1A concentration. Comparison of the concentration of three other proteins with lysine content yielded a much poorer correlation. The concentration of EF-2, another abundant protein involved in polypeptide chain elongation, showed no correlation with lysine ($r^2 = 0.07$). Sucrose synthase (*Sh1*), a major endosperm protein involved in carbohydrate biosynthesis, had a higher correlation with lysine content ($r^2 = 0.57$), but there was a poor correlation ($r^2 = 0.21$) between lysine content and the concentration of ADP glucose pyrophosphorylase.

The relationship between eEF1A and lysine content was found to occur in other cereals, including barley and sorghum, which also have considerable genetic variability in lysine content. For barley, the relationship between eEF1A and lysine content was $r^2 = 0.60$, and with sorghum it was 0.96.

eEF1A itself only accounts for 1 to 3% to the total lysine in the endosperm (Sun *et al.*, 1997). Because the concentration of EF-2 is not correlated with the lysine content, it appears this relationship between eEF1A concentration and lysine content does not generally apply to protein synthesis factors. Besides binding to aminoacyl-tRNAs, eEF1A has been shown to associate with the cytoskeleton.

Clore *et al.* (1996) used indirect immunofluorescence and confocal microscopy to visualize actin, tubulin, eEF1A and protein bodies in intact endosperm cells and determined their spatial relationships *in situ*. These components were analyzed in interphase cells at two developmental stages (early stage III and mid-stage III to early stage IV). Clore *et al.* (1996) demonstrated that protein bodies are juxtaposed to a reticulate array of microtubules and are surrounded by eEF1A and actin. The two latter components appeared to exist in a complex, because treatment of the endosperm with cytochalasin D resulted in a redistribution of eEF1A within the cells. eEF1A has been found to be associated with protein bodies after sucrose gradient centrifugation of endosperm homogenates (Habben *et al.*, 1993). According to previous studies eEF1A binds to actin and facilitates its polymerization *in vitro* (Yang *et al.*, 1993). Clore *et al.*

(1996) speculated that actin surrounding protein bodies may be polymerized or cross-linked by eEF1A. It is also possible, given the reports that eEF1A bundles microtubules (Durso and Cyr, 1994) and binds to ER membranes (Hayashi et al., 1989), that this protein somehow connects the actin with microtubules or attaches actin to the ER membrane. Another possibility is that eEF1A colocalizes with actin around protein bodies to facilitates zein peptide elongation. The association of eEF1A with the cytoskeleton surrounding protein bodies may help explain the basis for the high correlation between lysine content and the concentration of eEF1A.

STRUCTURE AND FUNCTION OF eEF1A

eEF1 is the most abundant eukaryotic translational factor. It catalyzes, together with the translation factor eEF2, the elongation cycle in eukaryotic protein synthesis (Moldave, 1985; Hershey, 1991). The eEF complex is composed of four subunits: eEF1A, eEF1Bα, eEF1Bβ and eEF1Bγ with molecular weight of 52-, 48-, 36- and 34-kDa, respectively. The concentration of eEF1A is high (1 -10% of crude protein) and far exceeds the levels of other protein synthesis components, including eEF1B, tRNAs, initiation factors, EF-2 and the ribosomes themselves (Slobin, 1980; Thiele et al., 1985; Browing et al., 1990). Studies indicate that the ratio of each subunit in the eEF1 complex is 1:1:1:1 (Ejiri et al., 1994). The ratio of these subunits in Artemia was found to be 2:1:1:1 (Janssen et al., 1994). Janssen et al. (1994) demonstrated that one molecule of eEF1A can dissociate easily from the eEF1 complex under the influence of aminoacyl-tRNA and GTP, while the second molecule of eEF1A was found to remain firmly attached. eEF1A is one of the most abundant proteins in eukaryotic cells. In wheat embryos it represents 5% of the soluble protein (Bronwing et al., 1990).

The process of peptide elongation starts with the formation of a ternary complex between eEF1A, aminoacyl tRNA and GTP. Subsequently, aminoacyl tRNA is attached to the mRNA-ribosome complex. This later step requires hydrolysis of GTP (Hershey,

1991). The elongation rate is increased by eEF1B, which accelerates the exchange of GDP bound to eEF1A, for GTP (van Damme *et al.*, 1990).

Another potential function for eEF1A is related to the cytoskeleton. Yang *et al.*, (1990) demonstrated that ABP-50 from *Dictyostelium*, later identified as eEF1A, binds to monomeric actin in the cytosol of unstimulated cells, and the association of ABP-50 with the actin cytoskeleton is regulated during chemotaxis. Yang *et al.* (1993) purified and characterized eEF1A as a new phosphatidylinositol kinase (PI4 kinase) activator. PI4-kinase catalyzes the phosphorylation of position 4 of the inositol ring of PI to form phosphatidylnositol-4-monophosphate. PI4-P serves as the precursor of phosphatidylinositol 4,5-bisphosphate, which is the source of the second messengers, inositol 1,4,5-triphosphate and diacylglycerol (Berridge, 1987; Majerus *et al.*, 1986). In addition, PI4-P can enhace the polymerization of actin by binding actin-severing proteins (Lassing and Lindberg, 1985; Janmey and Stossel, 1989; Yonezawa *et al.*, 1990). PI4-kinase activity is present in both the membrane and soluble fraction of plant cells (Chen *et al.*, 1991). Studies of cytoskeleton-associated proteins in carrot and A431 cells have shown that PI4-kinase activity is associated with the cytoskeleton, specifically with the actin-rich fraction (Payrastre *et al.*, 1991).

eEF1A is also a member of the GTPase superfamily of proteins (reviewed by Bourne *et al.*, 1991). The GTPases share a similar cycle: they release GDP with help of additional factors such as EF-Ts for prokaryotes and eEF1B for eukaryotes and assume an 'inactive' conformation. Upon (preferential) binding of GTP over GDP in the guanine nucleotide binding site, the GTPase assumes an 'active' conformation. The GTPase superfamily shares many structural features that contribute to their role as "molecular switches" in signal transduction of cells (Stankovic *et al.*, 1993).

LOCALIZATION OF FUNCTIONAL DOMAINS IN eEF1A

Four structural domains have been determined for eEF1A. One binds to GTP, another to aminoacyl tRNA, another to eEF1B and another to actin. Some of the eEF1A domains have been determined by comparison with *E. coli* EF-Tu. Prokaryotic polypeptide chain elongation factor Tu (EF-Tu) from *E. coli* has 393 amino acid residues (Arai *et al.*, 1980) and is encoded by two nearly identical but unlinked genes in the *E. coli* chromosome: *tufA* (73 min.) and *tufB* (89 min.) (Jaskunas *et al.*, 1975). Both *tufA* (Shibuya *et al.*, 1979) and *tufB* (Miyajima *et al.*, 1979) have been cloned and their sequences determined (Yokota *et al.*, 1980; An and Friesen, 1980). EF-Tu is arranged in three distinct domains (Clark *et al.*, 1990). Domain I (residues 1-209) contains the amino-terminal situated guanine-nucleotide-binding pocket. Domain II (residues 210-300) includes the second conserved region between EF-Tu and eEF1A (Riis *et al.*, 1990), whereas domain III includes the carboxy-terminal residues 301-363 (Clark *et al.*, 1990). Biochemical data from different reports indicate that particular amino acid residues in domain I and II from EF-Tu are close to the binding site for aminoacyl tRNA (Jonak *et al.*, 1984). The portion of eEF1A spanning residues 1-287 shows a stimulatory effect on the transport of aminoacyl tRNA to the ribosome. Removal of the region spanning residues 37-68 and 37-128 completely abolished the transport of aminoacyl tRNA to the ribosome (van Damme *et al.*, 1992). Two other reports showed that the removal of residues 37-68 in eEF1A reduced the aminoacyl tRNA binding to only 30 and 60%, respectively (Slobin *et al.*, 1981). The prokaryotic fragment of EF-Tu containing domain II and III binds EF-Ts in *Thermus thermophilus* (Peter *et al.*, 1990).

van Damme *et al.* (1992) demonstrated that the carboxy-terminus of eEF1A, together with the attached polypeptide stretch 1-36, were sufficient to associate with eEF1B, indicating that an intact binding site is present in these fragments. Removal of the last 174 amino acid residues of eEF1A completely changes the native conformation of the remaining 30-kDa fragment, so that a strong reduction in affinity toward eEF1B takes

place. However, the 30-kDa fragment by itself showed a clear activity toward GDP binding and transport of aminoacyl tRNA to the ribosome, indicating that the native conformation, at least for these two functions, remains intact. On the other hand, a secondary binding site for eEF1B in the amino-terminal part of eEF1A was not excluded.

The actual differences in spatial structure between prokaryotic EF-Tu and eukaryotic eEF1A manifest themselves as extra 'looping outs' in the guanine-nucleotide-binding region of the polypeptide chain of eEF1A (van Damme *et al.*, 1992). One could speculate that these extra sites are important for the interaction with other eukaryotic cellular components. The first loop spans the residues 24-38, which has a proteolytic site at Arg36. This extension in the GTP-binding domain is not only found in eEF1A, but in all other eukaryotic GTP-binding proteins as well (Gilman, 1987). The second region spans the residues 121-130, which includes a sensitive proteolytic cleavage site at Lys128.

A comparison between EF-Tu and other GTP binding proteins, including eEF1A, shows three GTP binding consensus sequences. The first domain is GXXXXGK, corresponding to residues 18-24 in EF-Tu and to residues 13-19 in eEF1A; the second is DXXG corresponding to residues 80-84 in EF-Tu and to residues 90-93 in eEF1A; and the third is NKXD, corresponding to residues 135-138 in EF-Tu and to residues 152-155 in eEF1A (Gilman, 1987). The first two elements are involved in the interaction with the phosphate portion of the GTP molecule and the last element is involved in nucleotide specificity (McCormick *et al.*, 1985).

Comparison of eEF1A from *Dictyostelium* with other actin-binding proteins showed that the region corresponding residues 164-184 is homologous to the actin-binding domain of depactin (residues 1-20) (Yang *et al.*, 1990). A region with homology to the actin-binding domain of depactin exists in *Arabidopsis* (Axelos *et al.*, 1989) and tomato (Pokalsky *et al.*, 1989) eEF1A. Five of seven residues implicated in actin-binding by cross-linking are conserved. A capacity to bind actin would be expected in plants (Collings *et al.*, 1994).

eEF1A POST-TRANSLATIONAL MODIFICATIONS

eEF1A post-translational modifications have been determined in many organisms and might be related to different functions of eEF1A (Venema *et al.*, 1991). In *Mucor racemosus*, eEF1A is post-translationally methylated with the formation of mono-, di- and tri-methyllysine at as many as 16 sites. Nearly 20% of the 44 lysine residues of eEF1A from mycellia are modified while those from sporangiospores are virtually unmethylated (Hiatt *et al.*, 1982). Fonzi *et al.* (1985) proposed the methylation is important to regulate the activity of eEF1A and is not involved in transcriptional regulation. During the course of spore germination, the specific activity of the factor in crude extracts increases six-fold. This increase in activity is accompanied by a constant level of eEF1A-specific mRNA and a constant level of eEF1A protein. Methylation of the protein, however, accelerates during the germination process, in parallel with the increase in specific activity of the factor (Fonzi *et al.*, 1985). Sherman and Sypherd (1989) suggested that hypomethylated and fully methylated eEF1A has equal affinities for GTP, aminoacyl-tRNA and ribosomes. Also, methylation did not appear to affect the accuracy of translation in an *in vitro* system. However, the experiments suggested methylation may affect the ability of the eEF1A to form complexes with other subunits (EF-1βγ), which are known to enhance the overall rate of protein synthesis. The role of these modifications in the activity and stability of eEF1A is not fully understood at present. Cavallius *et al.* (1997) used site directed mutagenesis to study the post-translational methylation of lysine residues in eEF1A in an attempt to separate possible roles of eEF1A in translation versus other biologic phenomena. No phenotypic differences were found even if all four methylated lysines were mutated to arginines. Growth rates and *in vitro* activities were unchanged.

eEF1A LEVELS. ACTIVITY AND GENE EXPRESSION

There is evidence to suggest that eEF1A activity affects cellular protein synthesis rates and possibly development. In fungal sporangiospores, in which protein synthesis rates are markedly higher than in hyphae, levels of mRNA corresponding to one eEF1A gene family member increase two-fold (Linz and Sypherd, 1987). In plants, steady-state levels of eEF1A mRNA are higher in meristematic regions of the plant, which are characterized by higher rates of protein synthesis than in older, less metabolically active tissues (Pokalsky *et al.*, 1989). The most convincing evidence, however, for the role of eEF1A in regulating development comes from studies on the aging process, where decreasing eEF1A activity has been shown to be an important cause in the characteristic decline in protein synthesis in cultured mammalian cell lines and in some invertebrate and vertebrate species (Cavallius *et al.*, 1986; Shepherd *et al.*, 1989). In cultured human fibroblasts, eEF1A levels remain constant during the first 80 to 85% of the culture life span, but beyond this point, eEF1A levels and activity decline dramatically. eEF1A activity also decreases after low-serum-associated arrest in G1 of the cell cycle. In contrast, immortalized (transformed) cell lines exhibit no decrease in eEF1A levels during culture or drop in activity after arrest in G1 (Cavallius *et al.*, 1986). Similarly, in *Drosophila* it was demonstrated that levels of translatable eEF1A mRNA decline in aging flies and that overexpression of eEF1A slowed the aging process (Shepherd *et al.*, 1989). *Agrobacterium*-mediated integration of a tomato eEF1A promoter-β-glucuronidase (GUS) fusion into the tobacco genome demonstrated that eEF1A is regulated at the level of either transcription or mRNA stability during plant development and that this regulation is correlated with specific changes in patterns of growth and development (Ursin *et al.*, 1991).

eEF1A genes undergo strict cell type-specific or stage-specific regulation of expression as reported for *Saccharomyces cerevisiae* (Schirmaier *et al.*, 1984), *Mucor racemosus* (Linz and Sypherd, 1987), *Artemia salina* (Lenstra *et al.*, 1986), *Drosophila*

melanogaster (Hovemann *et al.*, 1988) and *Xenopus laevis* (Krieg *et al.*, 1989; Dje *et al.*, 1990; Frydenberg *et al.*, 1991). It is in the latter two species that most of the developmental work has been performed. *Drosophila* has two copies of the eEF1A gene (F1 and F2). Both genes are differentially expressed during early development, but only F1 mRNA can be detected in adult tissues (Hovemann *et al.*, 1988). The situation is similar in *X. laevis*, where three active eEF1A genes have been characterized (42Sp50, EF-1αO and EF-1αS); all of the genes are active in early development, while only EF-1aS is expressed after neurolation and in the adult (Krieg *et la.*, 1989; Dje *et al.*, 1990). In rats, two eEF1A have been reported. The first eEF1A identified, referred as EF-1α by Lee *et al.* (1993), can be found in all tissues, while the second one, S1 (Ann *et al.*, 1991), can only be detected in brain, heart and muscle (Ann *et al.*, 1992; Lee *et al.*, 1992). Linz and Sypherd (1987) demonstrated that there is differential expression of the genes encoding eEF1A in *Mucor racemosus*. At least one gene, TEF-3, showed a morphology-specific pattern of transcript accumulation. The accumulation of mRNA levels for each of the three genes encoding eEF1A in *M. racemosus* varied several fold, so that TEF-1 > TEF-3 > TEF-2. In addition, mRNA levels of TEF-3 showed a morphology-specific pattern of accumulation (Linz and Sypherd, 1987).

In plants, eEF1A gene expression changes in response to stress (e.g. low temperature, hypoxia, fungal infection and wounding) or to plant hormones (e.g. auxin). Ursin *et al.* (1991) conducted an extensive analysis of the expression of a tomato eEF1A-GUS construct in transgenic tobacco plants during stages of growth. GUS activity was found to be highest in tissues undergoing rapid growth, such as meristems, root tips, young leaves and developing ovules (Ursin *et al.*, 1991). GUS staining was prominent in the root tips, but was found at regions of auxin-induced curvature in the hypocotyl of seedlings treated with auxin (Ursin *et al.*, 1991). eEF1A was found to be expressed at high levels during the globular stage of carrot embryogenesis (Kawahara *et al.*, 1992), during cold response in barley (Dunn *et al.*, 1993), in response to light in soybean seedlings (Aguilar *et al.*, 1991) and following wounding in potato tubers (Morelli *et al.*,

1994). On the other hand, transcript levels of bean eEF1A dropped during fungal infection (Mahe *et al.*, 1992). Taken together all these observations suggest a rapid response in the expression of eEF1A to changes in the protein synthesis needs of the plant, either due to environmental conditions or to internal signals for development.

SUMMARY AND OBJECTIVES

Significant progress has been made in sequencing the genome of several model organisms, such as *Escherichia coli* and *Bacillus subtilis* and the budding yeast, *Saccharomyces cerevisiae*. Genome projects have also been undertaken in many animals and plants. The general goal of the plant genome projects is to understand more about the function of genes that have not been characterized in other systems in order to help solve problems related to disease and crop production.

In the maize genome project, we combined sequencing, mapping and measurement of transcript levels of cDNAs with the objective of eventually deducing their putative functions. We reported the analysis of 490 cDNAs from etiolated seedlings and 576 cDNAs from membrane-free endosperm polysomes. At the time we presented the data, 61% of the clones showed no significant similarity within those in Genbank. Among those clones with significant matches, 6.7% were considered identical to known maize genes and 7% represented potentially new maize gene family members. With the increased amount of data that has been deposited in the GenBank since the project started, many more of these ESTs can be assigned functions.

One gene identified multiple times in both endosperm and seedling cDNA libraries that was interesting for further characterization was eEF1A. According to previous reports, eEF1A is a multifunctional protein that might be involved in the cytoskeleton network and protein body formation in the endosperm. eEF1A was also shown to be remarkably correlated with the level of lysine in maize endosperm and has the potential to be used as an indicator of nutritional value in maize breeding programs.

The mRNA levels in different tissues and different developmental stages of the *o2* mutant and wild type endosperm were used to investigate the expression of eEF1A gene family members. The yeast two-hybrid system was used for the identification of proteins that interact with eEF1A to help understand more about its many functions.

CHAPTER 2

PARTIAL SEQUENCING AND MAPPING OF CLONES FROM TWO MAIZE cDNA
LIBRARIES

INTRODUCTION

One aspect of many modern biological studies is the isolation of clones for
desired genes, and with the development of improved mapping methods, cloning through
the use of positional information has been widely explored as a means to this end. One
technique, chromosome walking has been restricted to organisms with large genomes,
due to the large fraction of repetitive sequences and the increased physical distance
between markers (Clark and Carbon, 1980). Consequently, there is a real need for
alternative approaches which take advantage of positional information. We have begun
to explore one such approach in maize, based solely upon the analysis of expressed
sequences (i.e. cDNA clones), as a means to identify function through coincident
mapping with phenotypic mutants. By analyzing cDNA clones and not genomic
sequences, this strategy is essentially neutral to genome size and ignores the vast majority
of non-expressed sequences. The choice of maize as the model organism was based upon
several aspects of its biology. Maize has a long history of Mendelian, cytological, and
molecular genetics and well-developed genetic and RFLP maps (Coe *et al.*, 1988;
Helentjaris, 1987). The wealth of mutants, many of which have already been mapped and
extensively studied, increases the chances of actually correlating mapped cDNA clones
with phenotypic disruptions (Coe *et al.*, 1988). Moreover maize possesses one of the
highest sequence polymorphism rates of any species (Shattuck-Eidens *et al.*, 1990),
therefore it is relatively trivial to detect sequence polymorphisms for most clones and
then assign those cDNA clones to a genomic location through RFLP analysis.

Recently, another strategy has been developed which also is based upon the analysis of expressed sequences in order to identify new genes. The single-pass sequencing of anonymous cDNAs, producing expressed sequence tags or ESTs, has shown great promise in the identification of clones through similarity with prior entries in the genome databases (Adams *et al.*, 1991; Adams *et al.*, 1992). Since the initial studies with human sequences, this approach has been applied to other species, including many plants such as *Arabidopsis thaliana* (Hofte *et al.*, 1993), *Brassica napus* (Park *et al.*, 1993), *Oryza sativa* (Uchimiya *et al.*, 1992) and *Zea mays* (Keith *et al.*, 1993). While one is often able to identify a sequence similarity with a cDNA by this method, the actual function may remain obscure. For instance, identifying the sequence for a protein kinase will not by itself identify the role of this sequence in organismal growth and development. Expectations that EST analysis can completely circumvent this problem are probably unjustified and in many, if not most cases, the final determination of the function of any candidate sequence must be verified through linkage to a disrupted phenotype.

We have used a combination of approaches in an attempt to improve the process of identification of putative functions of cDNA clones by comparing the sequence with the entry in GenBank database and coupling that with analysis of map locations and gene expression. Our initial results with etiolated seedling and endosperm libraries support the basis of this strategy, in that the putative functions of a large number of cDNA sequences were identified through sequence similarity comparison and genomic Southern analysis. In some cases we provided coincident mapping between specific clones and mutant loci with predictable phenotypes. We believe that this approach will lead to the identification of many cloned genes with functional roles in organismal growth and development. Furthermore, the sequencing and mapping data are useful in addressing basic questions in genome structure and organization, such as gene clustering, duplication and synteny with other plant species.

MATERIALS AND METHODS

Plant material, total and mRNA isolation

Maize (*Zea mays* L. cv W64A) RNA was isolated from membrane-free polysomes of endosperm harvested at 18 to 22 days after pollination (Habben *et al.*, 1993). Etiolated seedlings of maize (cv B73) were grown in a growth chamber for 12 days (30°C) and the total RNA isolated (Chomczynski and Sacchi, 1987). The mRNAs from both tissues were subsequently isolated using the PolyATract mRNA Isolation System, according to the manufacturer's instructions (Promega).

Construction of cDNA libraries and screening of clones

cDNA libraries were constructed in the ZipLox system, using directional cloning (5'-*SalI*-3'-*NotI*) according to the manufacturer's instructions (GIBCO-BRL). The excision of cDNA clones was performed in bulk by combining 10^3 pfu of phage with *E. coli* DH10B (200 µl of 10^9 cell/ml) and selecting in LB medium (with ampicillin, X-Gal and IPTG). White colonies (i.e. those containing presumptive cDNA inserts) from each library were selected and distributed into 96-well-plates.

The level of mRNA expression corresponding to specific clones was estimated by hybridizing radioactively-labeled first strand cDNA, prepared from endosperm and seedling mRNA, to colony lifts prepared from each library (Hodge *et al.*, 1992). Colony lifts were prepared by transferring bacteria from 96 well plates to growth on solid LB medium and then subsequently lifting the resulting colonies onto nylon membranes. These membranes were placed upon paper towels soaked in a solution of 0.5M NaOH, 1.5 M NaCl, for 2 min, transferred onto fresh blotting paper containing a solution of 0.5M Tris-HCl, 1.5M NaCl, for 5 min, and finally equilibrated in 2X SSC for 5 minutes.

The membranes were dried at room temperature and the DNA fixed by uv crosslinking and baking for 2 hours at 80°C. cDNA probes were prepared from the mRNA fraction of membrane-free polysomes from endosperm and etiolated seedlings according to the manufacturer's instructions for first strand synthesis (GIBCO-BRL) with the addition of 50 μCi ^{32}P-dCTP. Prehybridization and hybridization solutions contained 5X SSC, 50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 0.2% SDS, 2X Denhardt's Reagent, 100 μg/ml of denatured salmon sperm DNA, and 1 μg/ml oligo-dT. The hybridizations were carried out for 18 hr at 65°C. Post-hybridization washes were performed in 0.1X SSC; 0.1% SDS two times at room temperature for 5 min each and two times at 65°C for 30 min each. The washed filters were exposed to X-ray film overnight (-80°C) and the films were then developed. Individual clones were classified as corresponding to either abundant or rare mRNAs according to the relative intensity of the hybridization signal.

Inserts from cDNA clones derived from abundantly-expressed mRNAs were purified from an agarose gel using Geneclean (BIO 101), ^{32}P-labeled using random primers (Sambrook *et al.*, 1989), and hybridized to colony lift filters. This rehybridization facilitated the classification of many of the abundantly-expressed clones into related families, obviating the repeated analysis of identical clones.

Nucleotide sequence

Plasmids from the endosperm library were isolated using the Magic Minipreps DNA purification sample system (Promega) and the inserts sequenced using an Applied BioSystems DyePrimer cycle sequencing protocol and the 373a automated sequencer. Plasmids from the seedling library were purified on Qiawell Plus columns (Qiagen) and the inserts were sequenced using the Dideoxy manual sequencing method (Sanger *et al.*, 1977) for double-stranded DNA and ^{35}S-dATP as per the manufacturer's instructions (USB). Sequencing, which was from the presumed 5' end of the mRNA (using the T7

primer) yielded an average of 200+ bases for the seedling library and 400+ bases for the endosperm library. Nucleotide sequence information from the presumed 3' end of the mRNA (using the F21 primer) also was performed in some cases. After sequences were obtained, the vector segments were trimmed and the insert sequence was compared to GenBank entries, first at the amino acid level with the BlastX subroutine and, if necessary, at the nucleotide level by the BlastN subroutine (Altschul *et al.*, 1990). Similarity was considered significant if the percentage of identity was higher than 40% at the amino acid level or if the BlastX score was greater than 80. All sequences with appropriate annotations were subsequently submitted for entry into the dbEST database section of GenBank (Boguski *et al.*, 1993).

Mapping

Each cloned insert was analyzed by Southern blot hybridization to maize genomic DNA, using digoxiginin-labeled probes and non-radioactive detection. Labeling was accomplished by PCR amplification of the insert via flanking primer sequences in the vector with incorporation of a modified digoxiginin-dUTP in the reaction (McCreery and Helentjaris, 1994a). DNA from four inbred lines, Cm37, T232, Tx303, and Co159, which are the parental lines of the Brookhaven recombinant inbred mapping populations (Burr *et al.*, 1988), were digested with three different enzymes (*HindIII*, *EcoRV* and *EcoRI*), transferred to nylon membranes by Southern blotting and hybridized to the dig-labeled probes (McCreery and Helentjaris, 1994b) to determine informative probe-cross-enzyme combinations for RFLP mapping. The probes were subsequently applied to blots containing the Brookhaven RI progeny for the appropriate cross (the CmT progeny included 47 individuals and the CoTx population included 42 individuals). Genotypes of individual progeny were noted and mapping assignments were initially evaluated using RI Plant Manager. All of the segregation data were subsequently sent to B. Burr at Brookhaven for incorporation into the Maize Genome Database and Map.

RESULTS

cDNA libraries and expression pattern of clones

The endosperm and seedling cDNA libraries in this study were superior to prior libraries we examined in several aspects. Size selection of cDNAs and directional-cloning both improved the percentage of clones with large inserts and the identification of homologies from the presumed 5' terminus of the mRNA. The percentage of clones with inserts in these libraries is approximately 95%, and the average size of cDNAs is greater than 600bp. Contamination of the cDNA libraries with organellar sequences was detectable but was less than 2% as determined by sequencing and Southern analysis. Approximately 6% of the clones contain inserts originating from rearranged vector as determined by DNA sequencing. Furthermore, RNA prepared from membrane-free endosperm polysomes was utilized in one of these libraries to reduce the large number of zein (seed storage proteins) clones that were found to be very abundant in an earlier library prepared from total endosperm mRNA. The percentage of storage protein sequences was reduced from about 40% to less than 5% by this step. Both libraries were prepared with a modified vector that contains a genetic system for automatic subcloning *in vitro*, via the highly site-specific Cre-*lox* recombinational machinery of bacteriophage P1 (GIBCO-BRL). This facilitated the easy mass isolation of inserts from the plasmid products of excised phage.

Besides seed storage mRNAs which are abundant in the endosperm library, other abundant sequences, such as ribulose bisphosphate carboxylase in seedlings, also increase the difficulty of identifying unique sequences. In order to more efficiently analyze unique entries in our libraries, we used a variation of the "Cold Plaque" method (Hodge *et al.*, 1992) to identify mRNA sequences in the library that are rarely expressed. Colony lifts were hybridized to ^{32}P-labeled cDNA prepared from endosperm or seedling mRNA. Abundant cDNAs should produce a stronger hybridization signal than rarely expressed

sequences that represent a smaller percentage of probe. By also testing each library with a total cDNA probe prepared from the other tissue, it is possible to identify highly-expressed sequences as being either constitutive (i.e. abundant in both tissues) or tissue-specific (i.e. found in only one of the two tissues). Hence, we were not only able to improve the overall efficiency of this approach but also to develop data regarding expression of each clone, which might prove useful in later attempts to understand its function.

As is illustrated in figure 2.1, these expectations are born out to the extend that some clones exhibited a strong hybridization signal, while many signals were much weaker. Based upon signal strength, the clones were first classified either as abundantly-expressed or as rarely-expressed; the former group was subsequently divided into constitutive and tissue-specific members (Fig. 2.2). For the seedling cDNA library, the rare group also contained about three-fourths of the clones; with the remaining segment being approximately equally divided among the other three classes. Interestingly, in each library a small group of clones distinguished by this method were more abundantly expressed in the alternative tissue than the source tissue. This method of classification proved to be reasonably effective in that many of these clones determined to be abundantly expressed were subsequently identified by sequence homology as genes that are known to be highly expressed, including ribulose bisphosphate carboxylase, sucrose synthase, catalase, ubiquitin, and zeins. On the other hand, the delineation of sequences as either abundant or rarely expressed is somewhat subjective and limited by the sensitivity of this method. For instance, most of ribosomal protein genes identified by sequence similarity were classified by this approach as rarely-expressed, although many researchers might consider them to be abundant. We have since found that greater discrimination can be achieved by screening the original plaques prior to excision. The percentage of abundantly-expressed endosperm sequences rises to nearly 40% using this method (data not shown). This increase in screening efficiency is achieved not through more intense hybridization of the plaques, but rather to the lower background (i.e. higher
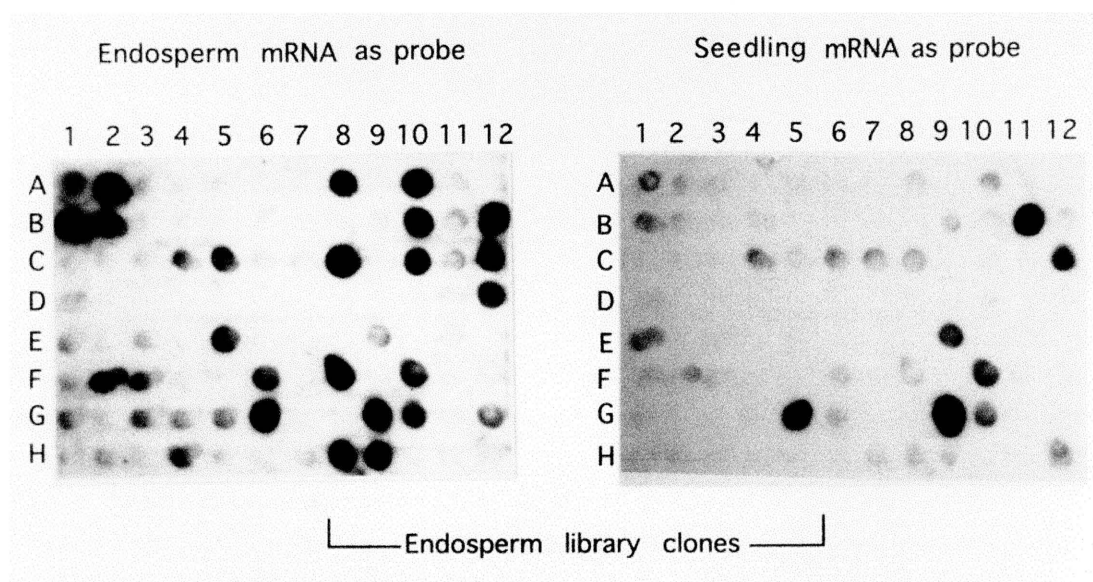
Figure 2.1 - Example of colony screening of cDNAs.

Duplicate colony membranes prepared from the endosperm library were hybridized with ^{32}P-labeled first strand cDNA probes prepared from either endosperm mRNA or seedling mRNA. Based upon signal strength, the clones could be classified into an abundantly-expressed group and a rarely-expressed group. For instance, A-2 and A-10 appear to be abundantly expressed only in endosperm, C-12 and G-9 are abundantly expressed both in endosperm and seedling, and B-11 is abundantly expressed only in seedling

Seedling cDNA library

7%
8%
8%
77%

- ☐ Abundant expression only in seedling
- ▨ Abundant expression both in endosperm and seedling
- ■ Abundant expression only in endosperm
- ▨ Rare expression in seedling

Endosperm cDNA library

2% 5%
20%
73%

- ☐ Abundant expression only in seedling
- ▨ Abundant expression both in endosperm and seedling
- ■ Abundant expression only in endosperm
- ▨ Rare expression in seedling

Figure 2.2 - Expression pattern of cDNA clones in endosperm and seedling. Based upon the results obtained by colony hybridization with total mRNA probes, all of the clones analyzed in this study were classified as abundantly expressed and tissue-specific, abundantly and constitutively-expressed, or rarely expressed.

signal:noise), which allow a more reliable selection.

Given the identification of a large number of rare sequences, it was practical to proceed directly with analyzing this group by both mapping and sequencing with little concern that identical sequences would routinely be found. On the other hand, the abundant group posed a problem in that many clones presumably represented the same sequences. To address this problem, a series of colony screens were performed by selecting clones and using them as probes. Six to twelve cDNA inserts were excised from the plasmid vector, labeled, and simultaneously used as probes against these same colony lifts as before. A number of additional colonies, besides the original source plasmids, were identified as being similar to this select group. Those remaining colonies originally detected as abundantly-expressed with the total mRNA probe but not similar to this pool were then isolated in subsequent pools and tested by the same method. Most of the abundantly-expressed clones could be reduced to a much smaller number of families; for instance with the endosperm library, the abundantly-expressed group was reduced to about 30 different families. The representation of each clone in the library ranged from 5 (sucrose synthase) to 27 members (α-zein, 19KD). Representative clones from each of these families were selected for analysis by both sequencing and mapping. Consequently comparison of the expression patterns of endosperm and seedling cDNA libraries has allowed us to develop a global representation of tissue distribution of abundantly and rarely expressed gene products in these two tissues.

Characterization of cDNA clones

A total of 313 single pass sequences from the two libraries were determined by either automated or manual sequencing. The clones were all sequenced from 5' end with a small subset also being sequenced from the presumed 3' terminus. The 5' nontranslated regions of many plant genes are relatively short when compared to the 3' nontranslated regions (Joshi, 1987) and both of the 3' and 5' nontranslated regions are much less

conserved than the coding regions. Not surprisingly, sequencing from the 3' end provided much less opportunity for detecting homologies by comparison with GenBank, as these sequencing passes rarely entered the conserved coding region of the genes. Only 10 of 52 3'-sequences produced any significant homologies, and almost all of these were to maize genes at the nucleic acid level. However, the 3'-sequences were useful in detecting sequence variations between members of a homologous family. For example three clones that matched to a 19kD α-zein by amino acid homology had virtually identical 5'-sequences but their 3'-sequences only shared 70% identity (data not shown).

All sequences were automatically translated in the six open reading frames and compared with the protein sequence database in GenBank using the subroutine BLASTX. If no significant homology was found, the sequences were subsequently compared at the nucleotide level using BLASTN. We found that some similarities were not detected by BLASTX even when the two clones were derived from homologous genes. This occurred when the sequence was located primarily in a nontranslated region or when the reading frame was shifted due to sequencing error. Consequently a small fraction of the time, BLASTN analysis was found to be useful in detecting additional high similarities missed by the BLASTX. Because BLASTX does not search other dbEST entries, we found the BLASTN subroutine useful for detecting similarities to randomly-sequenced cDNAs from other organisms. Comparison of the BLAST program results with FASTA revealed no significant advantage in detecting similarity.

As shown in Table 2.1, 39.3% of the rare group exhibited significant Genbank similarities and the percentage of significant matches in endosperm and seedling libraries was comparable despite the fact that all of the endosperm cDNA sequencing was performed with automated cDNA sequencing and the seedling sequences were primarily obtained by manual sequencing. Given that the data obtained with these two methods differ considerably in the number of bases sequenced (400+ for automated sequencing vs. 200+ for manual sequencing) and assuming that the libraries are otherwise equivalent, one might have expected to see a significant difference in the fraction of identified clones,

but this was not the case. The percentage of significant matches in the abundantly-expressed group of genes was much higher than that in rarely-expressed group. In addition, most of the clones from the abundant group were similar to previously-identified maize genes, while in the rare group, a relatively higher percentage was found to match to non-maize genes in GenBank.

Table 2.2 lists all clones identified in this study that exhibited significant similarity with any entry in the public databases. In some cases where multiple clones with similarity to the same gene were identified, only a single representative clone was listed. There were other instances where different clones matched the same GenBank entry, yet their amino acid and nucleotide sequences exhibited significant differences and these clones obviously represented different members of a multigene family. For instance, two clones both matched the same calcium-dependent protein kinase entry, but they exhibited 30% divergence in their 5'-sequences and virtually complete divergence in their 3'-sequences.

Among the sequences with significant similarity to sequences in the public database, 21 clones (6.7% of total) were considered to be identical to previously-sequenced maize genes, due to their very high similarity (>95%). Examples include ribosomal-inactivating protein, sucrose synthase and pyruvate, orthophosphate dikinase. Another 22 clones (7.0%) were believed to represent related genes of previously-identified maize genes, due to their lower similarity to known maize genes (40-90%) Examples include Ac transposase, actin1 and trypsin/factor XII inhibitor. An additional 80 clones (25.6%) exhibited some similarity to gene sequences from other plants, yeast, animals, and even *E. coli*. Examples include a calcium-dependent protein kinase, a proteasome component, and alanine transaminase.

In looking through Table 2.2, it can be seen that a very diverse group of gene types have been identified, despite the somewhat lower number of clones actually sequenced. Many clones were identified as "housekeeping" genes, i.e. genes that might be expected to be found in any cell type, with representatives from various metabolic

Table 2.1

Composition of maize cDNA libraries

	Endosperm library	Seedling library	Abundant group	Rare group
No database match	135 (57.9%)	55 (68.7%)	4 (11.8%)	186 (66.7%)
Database match-maize gene	33 (14.2%)	10 (12.5%)	23 (67.6%)	20 (7.2%)
Database match-other species	65 (27.9%)	15 (18.8%)	7 (20.6%)	73 (26.1%)

For each column, the number of actual clones in each category is indicated followed by the relative percentage in parentheses.

Table 2.2

Putative identification of maize cDNA clones

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
5C05F05	1,4-α-glucan branching enzyme	JT0968	118(94%)	complex	UAZ 265	2C,6L	Rare
6C02E06	10KD chaperonin, chloroplast precursor	Q02073	65 (80%)	simple	UAZ 222	4L	Rare
6C02A09	2-oxoglutarate dehydrogenase	P20967	63 (64%)	simple	UAZ 215	5L,5S	Rare
5C06D11	2-oxoglutarate dehydrogenase	Q02218	63 (47%)	simple			Rare
6C02G01	ABA-induced glycine-rich protein	P10979	45 (100%)	complex			Abund/E, S
5C02G05	ABA-induced glycine-rich protein	P10979	26 (96%)				Abund/E
5C04F09	Acidic ribosomal protein	P15826	50 (38%)	complex	UAZ 198	3L	Rare
5C05D03	Acidic ribosomal protein 60S, P2	P08094	76 (51%)	simple			Rare
6C02F11	Actin 1	J01238	150 (78%) [n]	complex	UAZ 233	6C,7L,8C	Abund/E
5C01H03	Actin 97	P30171	127 (96%)	complex			Rare
2C01H08	Acyl carrier protein reductase	S19832	61 (77%)	simple	UAZ 099	10C	
5C06E05	ADPG pyrophosphorylase	S18238	143 (86%)				Rare
5C04B05	Alanine transaminase	S28429	139 (93%)	simple	UAZ 158	5C	Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
5C06G10	Annexin VII	P24639	39 (58%)				Rare
5C04B11	Antifreeze protein precursor (Ala-rich protein)	P07835	54 (40%)	complex	UAZ 159	5C	Rare
RSP13	Aspartate aminotransferase	D14673	43 (90%)				
5C01B12	Aspartyl-tRNA synthetase α-2 subunit	P15178	120 (47%)	simple	UAZ 131	5L	Rare
5C04E07	ATP synthase bate chain	P19023	133 (96%)	complex	UAZ 243	3L,6L,8S	Rare
5C02D08	ATP-dependent protease protein precursor	P31542	129 (83%)	simple	UAZ 242	10L	Rare
5C04A11	Brittle-1 protein precursor	P29518	39 (92%)	complex	UAZ 155	10C	Rare
5C04A03	Calcium-dependent protein kinase	P28582	54 (43%)	complex	UAZ 130	1L,4L,5S	Rare
5C04G11	Calcium-dependent protein kinase	L14771	90 (54%)	complex	UAZ 197	6C,6C	Rare
5C06B08	Caltractin	X69220	125 (66%)	simple			Rare
6C02C04	Catalase 1	GB-M33104	240 (94%) [n]	simple	UAZ 226	5S	Abund/E
6C02C06	Catalase 3	GB-X12539	187 (98%) [n]	simple		mitochondria?	Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expressio n
5C02D11	Chloroplast small heat shock protein	P11890	48 (70%)	simple	UAZ 171	4L	Rare
5C04A08	Chymotrypsin inhibitor 2A precursor	A29537	51 (56%)	simple	UAZ 232	2L	Rare
6C02A04	CP29, Chl A/B binding protein	X71878	61 (69%)	simple	UAZ 200	7L	Rare
6C06B11	Cytochrome P450 hydroxylase	X70981	54 (46%)	simple			Abund/S
5C04A12	Early nodulin	D13506	82 (56%)	complex	UAZ 227	6C	Rare
6C06E02	Early nodulin 8	L18899	27 (55%)				Rare
5C04H09	Elongation factor 1-α	X56856	138 (90%)	complex	UAZ 220	6L	Rare
5C04C04	Elongation factor 1-γ	L17307	127 (48%)	complex	UAZ 161	6L,9S,3L ,4L	Rare
5C01C07	Enolase	P26301	63 (100%)	simple	CSU15 8	9S	Rare
6C06C04	Enoyl-CoA hydratase	P07896	47 (40%)				Abund/S
5C04C02	Floral homeotic protein	P29383	42 (76%)	complex	UAZ 231	9L	Rare
6C06C01	Gene MS2 protein	S33804	55 (50%)				Rare
RSP 33	α-glucan phosphorylase	S15531	73 (86%)				
5C01G05	Glucosamine-fructose-6-phosphate	P14742	131 (62%)	complex			Rare
6C06C10	b-glucosidase	X74217	122(99%)	simple			Abund/S

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
5C03G12	GTP-binding protein	Q01474	45 (80%)	complex	UAZ 151	1L	Rare
5C04D03	GTP-binding protein	D10715	71 (77%)	complex	UAZ 245	7L	Rare
5C06F05	H ß-58 homolog	B44882	42 (57%)				Rare
5C04F02	Heat shock protein 17.6KD	P24631	28 (82%)	simple	UAZ 210	3S	Abund/E
5C04D01	Heat shock protein 70KD,mitochondria	Q01899	142 (91%)	simple	UAZ 205	5S,1L	Rare
6C01E09	Heat shock protein 70KD	P26791	37 (81%)	complex			Abund/S
5C04H04	Heat shock protein 80KD	M9956 5	124 (88%)	complex	UAZ 219	5S	Rare
6C02C08	Histone H2A3	P25470	32 (59%)	complex	UAZ 221	7L	Abund/E
5C04D12	Histone H2B	P30756	69 (86%)	complex	UAZ 228	2L,10C,4 L,1L	Rare
5C03H09	Histone H3	S24346	102 (97%)	complex	UAZ 248	1S,5L	Abund/E, S
5C02C03	Histone H4	A25642	60 (88%)				Rare
5C05C09	Hypothetical 42.8KD protein	P32615	74 (52%)	complex			Rare
5C01G12	Hypothetical protein YCR072C	S19487	42 (52%)	simple			Rare
5C06C07	Initiation factor 5A (EIF-5A)	P24922	68 (79%)	complex			Rare
5C02E08	Inorganic pyrophosphatase	P31414	87 (59%)	complex	UAZ 280	4S,9L	Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expressio n
5C06G04	IR4 protein	M7672 6	39 (43%)				Rare
2C02A04	Iron-deficiency induced gene	Blyids3	106 (81%) ^{n}	simple	UAZ 080	6S	
5C05H04	Ketol-acid reductoisomerae	X68150	117(88%)	complex	UAZ 269	2L,6S,6L ,8C	Rare
6C02E02	male sterility 2/MS2 gene	X73652	36 (50%)	simple	UAZ 195	4C	Rare
5C05E10	Metallothionein	P30564	45 (62%)				Rare
5C02F05	Mitochondria carrier protein YMC1	P32331	35 (62%)	complex	UAZ 282	1L	Rare
5C02A01	Nonspecific lipid-transfer protein precursor	P19656	52 (92%)				Rare
SPF4	Nucleoside diphosphate kinase I	S24165	47 (65%)	simple	UAZ 091	7C	
2C01C07	Oligopeptidase A	A42298	42 (45%)	simple	UAZ 100	10S	
5C04D09	p23=tumor-specific transplantation antigen	A44367	103 (66%)	simple	UAZ 208	1L	Rare
6C06B11	P450 hydroxylase	X70981	54 (46%)				Rare
5C02C04	Peptidyl-prolyl cis-trans isomerase	P21569	116 (90%)	complex	UAZ 238	5L	Abund/E
6C02D10	Peroxidase precursor	S22087	53 (47%)	simple	UAZ 235	2L	Rare
6C06D10	pG1 protein	A44803	46 (71%)				Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
5C04C05	Phosphoprotein phosphatase 1	S29396	65 (61%)	complex	UAZ 244	6L,8C	Rare
5C06E01	Porin	S34146	102 (66%)				Rare
5C06B09	Profilin	P35081	80 (78%)	simple			Rare
5C02A05	Proteasome component	P25789	66 (51%)	complex	UAZ 237	9S,6C	Rare
5C05G07	Protein disulfide-isomerase	A41440	121(53%)				Rare
5C02A07	Protein kinase	L07248	76 (57%)	complex	UAZ 252	4L,8L	Rare
5C06D06	Pyruvate dehydrogenase	S13825	109 (55%)				Rare
5C04A04	Pyruvate, orthophosphate dikinase	M5865 6	105 (95%)	simple	UAZ 127	8L,6L	Abund/E
6C02C02	Retinoblastoma-associated protein-like homolog	P28749	80 (32%)	simple	UAZ 191	2L	Abund/S, E
5C06C03	Ribosomal protein P2	X77253	76 (56%)	simple			Rare
5C06A11	Ribosomal protein 40S, S5	P24050	121 (80%)	simple			Rare
5C06H03	Ribosomal protein S8	Z26879	42 (66%)				Rare
5C02F12	Ribosomal protein 40S, S11	P25460	61 (98%)	complex	UAZ 251	2S,3L,6L ,8C, 10L	Rare
5C01H07	Ribosomal protein 40S, S14	P19951	107 (81%)				Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
5C01E10	Ribosomal protein 40S, S24(S19)	P16632	33 (54%)	simple			Rare
5C06G08	Ribosomal protein 40S, S24E	P14249	54 (62%)				Rare
5C01A05	Ribosomal protein 40S, S28	P25112	61 (75%)	complex	UAZ 146	1S, 1S	Rare
5C01A09	Ribosomal protein 40S, S8	P09058	74 (39%)	simple	UAZ 115	4L	Abund/E
5C04G09	Ribosomal protein 40S, S8	P09058	74 (64%)	simple			Abund/E
RSP 81	Ribosomal protein 60S, L7	A37055	107 (44%)			2, 10	
5C05D11	Ribosomal protein 60S, L12	P23358	50 (64%)				Rare
5C04B03	Ribosomal protein 60S, L19	P14118	140 (64%)	simple	UAZ 157	4L	Rare
5C04D11	Ribosomal protein 60S, L5A	P15125	127 (63%)	simple	UAZ 189	3C	Rare
5C05G01	Ribosomal protein L17-1	S35101	72 (80%)	simple			Rare
5C01A12	Ribosomal protein S6	S25550	23 (86%)	complex	UAZ 119	7L,8L,9 L	Rare
5C01B01	Ribosomal RNA gene, 25S	M11585	308 (80%) n	simple			Rare
5C04F01	Ribosomal-inactivating protein	P25898	113 (98%)	simple	UAZ 193	8C	Abund/E

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
6C02F07	Rice lipoxygenase L-2	P29250	51 (58%)	complex			Rare
5C03H11	Rice partial cDNA	D1561 9	99 (88%) n	simple	UAZ 186	5L	Abund/E, S
6C02C05	RuBPcase large subunit	GB-V0017 1	222 (96%) n			chloropl ast	Abund/E
5C01H11	S-adenosyl-L-homocystein hydrolase	D1613 8	110 (91%)	simple	UAZ 145	4S	Rare
5C01G10	Salt-stress induced hydrophobic protein	U0096 6	51 (64%)	simple	UAZ 250	10L	Abund/E
6C02G11	serine-tRNA ligase	P07284	43 (48%)	complex	UAZ 236	2S,9L	Rare
5C04H06	*Sh2* gene, ADPG pyrophosphorylase	S48563	110 (91%)			1,3L,4,5	Abund/E
2C07F04	Signal recognition particle receptor α-subunit	A2457 0	64 (53%)	simple	UAZ 008	3L	
5C04A01	sorbitol dehydrogenase	Q0079 6	37 (62%)	simple	UAZ 152	9L	Rare
5C03G08	Starch branching enzyme II	L08065	250 (97%) n	simple			Rare
5C04B10	Starch synthase precursor	P19395	64 (38%)	complex	UAZ 218	4C,3L	Rare
5C04A07	Sucrose synthase	P04712	108 (93%)	simple	UAZ 154	9S	Abund/E, S
5C05H07	TATA-binding protein	S32622	44 (100%)				Rare

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expression
6C02E11	Translation initiation factor eif-2 subunit	L19161	41 (80%)	simple	UAZ 229	7L	Rare
5C04E06	Transposable element Ac	X0138 0	68 (46%)	comple x			Rare
6C02G05	Triose phosphate isomerase	A2550 1	45 (100%)	simple	UAZ 093	8L	Rare
5C05F10	Triose phosphate isomerase	P12863	107(106%)				Rare
5C03B04	Trypsin/factor XII A inhibitor	P01088	39 (74%)	comple x	UAZ 184	4S	Abund/E
6C02D05	Tubulin α-3	GB-X6317 6	241(100%) n	comple x	UAZ 201	5S	Rare
6C06B09	Tubulin α-6 chain	P33627	109(100%)				Rare
5C01E01	Ubiquitin precursor	S04863	63 (80%)	comple x	UAZ 247	4L	Abund/E,S
5C04D06	Ubiquitin-conjugating enzyme E2	P16577	93 (86%)	simple	UAZ 206	6S	Rare
6C02D08	UDP-glucose pyrophosphorylase	P19595	36 (69%)	simple	UAZ 194	2L	Rare
5C06D10	Uromodulin	P27590	33 (57%)	simple			Rare
5C02H07	UTP-glucose-1-phosphate Uridyltransferase	P19595	118 (83%)	comple x	UAZ 194	2L	Rare
6C02E07	vacuolar ATPase β subunit	L11862	66 (96%)	comple x	UAZ 223	9C	Rare
5C01C06	Vegetative specific protein H7	P14327	61 (39%)	simple	UAZ 246	4C	Abund/E
2C01B06	Zein-α precursor 19KD, 19A30	S21970	45 (97%)	comple x	UAZ 049	4S	Abund/E

Table 2.2 (cont.)

Clone Desig.	Putative identification (GenBank entry)	GB Access	Similarity	Gen Org	Locus	Map Location	Expressio n
2C06H03	Zein-α precursor 19KD, 19A20	Zizm2	117 (89%)	complex	UAZ 068	4S	Abund/E
5C02A08	Zein-α precursor 19KD, 19D1	P06678	128 (99%)	complex	UAZ 272	1C	Abund/E
5C03G02	Zein-α precursor 19KD, PMS2	P24450	103 (91%)	complex	UAZ 149	4S	Abund/E
5C01H08	Zein-α precursor19KD, 19C2	P06677	132 (96%)	complex			Abund/E
5C03B06	Zein-α precursor 22KD, PZ22.1	P04700	66 (93%)	complex	UAZ 185	4S	Abund/E
5C02C05	Zein-α precursor 22KD, PZ22.3	P04698	108 (100%)				Abund/E
5C04E12	Zein-β precursor 16KD, 15A3	P06673	45 (71%)				Abund/E
5C02A03	Zein-β precursor16KD, 15A3	P06673	51 (96%)				Rare
RSP80	Zein-γ	Zmzm19	47 (78%)	simple		7S	Abund/E

The columns refer respectively to :

(1) the lab designations for each clone; those entries with a designation of RSP, SPF, or first number of "5" or "2" are derived from endosperm library while those beginning with "6" are from a etiolated seedling library,

(2) the GenBank entries with highest similarity score.

(3) the accession number of the GenBank entry,

(4) overlap amino acid or nucleotide number; the % of sequence identity in parentheses, n is nucleotide CLARIFY

(5) "complex" indicates more than three significant hybridizing fragments on a genomic Southern blot with more than one enzyme, while "simple" means three or less,

(6) the map designation for any loci detected by this clone,

(7) chromosome number with L = long arm of chromosome, S = short arm, C = centromeric, and

(8) abundantly-expressed or rarely-expressed are evaluated based upon signal strength in a colony hybridization with 1st strand cDNA made from the same mRNA used for construction of libraries.

pathways such as glycolysis (triose phosphate isomerase, enolase, glyceraldyhyde 3-phosphate dehydrogenase, pyruvate dehydrogenase), carbohydrate metabolism (sucrose synthase, sorbitol dehydrogenease, ADPG pyrophosphorylase, starch synthase, starch branching enzymes, UTP-glucose-1-phosphate uridyltransferase), transcription (TATA-binding protein, TAT-binding protein-1, CCAAT-binding factor), translation (two elongation factors, initiation factors, serine-tRNA ligase, 16 different ribosomal proteins) and protein degradation (ubiquitin, ubiquitin-conjugating enzyme and proteasome components). In addition, a number of genes were identified that may be involved in signal transduction (three GTP-binding proteins and four protein kinases). Other clones exhibited high similarity to genes which could be tissue or situation-specific (i.e. salt-stress-induced proteins and numerous heat shock genes). Many of these would be expected to be rarely expressed. Identification of such clones by this random-selection approach indicates that our screening protocol is successful and perhaps that expression of these genes may not be as rare as expected.

Mapping of cDNA clones to maize chromosome

Unique representatives of each clone family were hybridized to Southern blots containing DNA from the Brookhaven parents. Approximately half of the clones presented simple RFLP patterns (i.e. three or fewer hybridization fragments) while the other fraction produced more complex RFLP patterns (Fig. 2.3). All clones with simple hybridization patterns and any others with interesting similarities to GenBank entries were mapped in the Brookhaven progeny. More than 300 cDNA clones have been analyzed to date and their loci assigned to the ten maize chromosomes. One chromosomal map is presented to illustrate the type of data accumulated by this approach (Fig. 2.4). On this linkage map we eliminated most of the prior loci, leaving only an informative subset as landmarks. All of the loci shown to the right were mapped as part of this project. Those with CSU designations were analyzed earlier from a mature
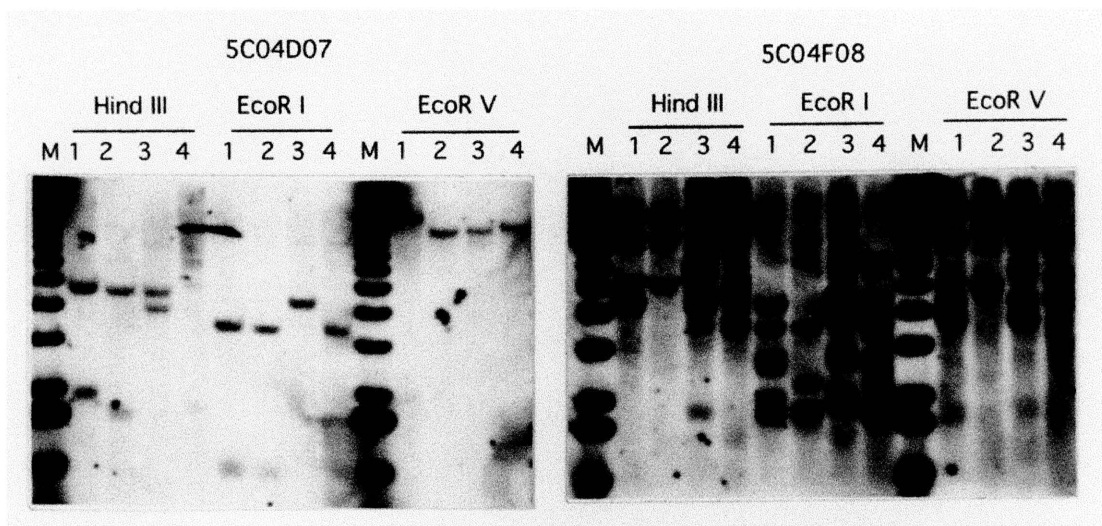
Figure 2.3 - Example of Southern blot analysis of two randomly-selected
cDNA clones. This membrane contains the following genomic DNAs: Cm37
(lanes 1), T232 (lanes 2), Tx303 (lanes 3), Co159 (lanes 4) digested with
either *HindIII, EcoRI or EcoRV*. Lanes M contain the 1kb molecular weight
markers from GIBCO-BRL. The clone 5C04D07 reveals a relatively simple
 hybridization pattern while the clone 5C04F08 represents a more complex
gene family.

Figure 2.4 - Genetic map of chromosome 5.

All of the loci defined on chromosome 5 by this study are shown along with several landmark loci from the Brookhaven database. Most of the other Brookhaven loci have been stripped off from this particular map for the purposes of illustration. The short arm of this chromosome is shown at the top and the long arm at the bottom, with the centromere being defined approximately by the *BT1* locus. The total length of this chromosome is 148.4 cM. This map was constructed using the program, RI Plant Manager and subsequently edited.

66

Core and Landmark Loci	Bins	Newly Mapped cDNA Loci		
BNL(TAS1N)				
		UAZ075(gfu)	UAZ214(Gly)	UAZ259(GaPp)
		CSU149(gfu?)		
		UAZ076A(ChS?)		
BNL6.25		UAZ134(gfu)	CSU033(gfu?)	
UMC147A	5.01	UAZ259(gfu)	UAZ201(ATub)	UAZ166B(gfu)
		UAZ163(gfu)		
		CSU137A(MADS)	UAZ130C(PKas)	CSU010(gfu?)
		UAZ211(gfu?)	UAZ215B(AKgD)	UAZ219(Heat) UAZ167B(gfu)
PGM2	5.02/ 5.03	CSU048(gfu?)		
		CSU150B(RPI2) UAZ 025E(MCP?)	UAZ159(gfu?)	UAZ205A(Heat)
UMC27A MDH5 UMC43	5.04	UAZ111(gfu)		
	5.05	CSU168A(gfu?) UAZ158(ATas) UAZ299B(PKas) UAZ213A(L19) UAZ275(Hest)		
BNL6.22		UAZ186 (Rest) UAZ226(Cat1)		
	5.06	UAZ111(gfu) UAZ077(gfu?) UAZ131(gfu)	UAZ285D(ACTr)	
AMP3		UAZ132(gfu)	UAZ070A(gfu)	UAZ295B(YEST)
BT1 ALS2 BNL10.12		UAZ279B(gfu?) UAZ161C(EF1G)	UAZ238(PpCI)	UAZ168(gfu)
	5.07	UAZ079(gfu) UAZ261Ae	UAZ190(GaPD)	
		UAZ248B(H3) UAZ164A(gfu)	UAZ078(gfu) UAZ138C(gfu)	
UMC126A UMC51	5.08	CSU026 (ATPT)	UAZ204(gfu?)	
UMC108	5.09	UAZ215A (AKgD)	UAZ254(gfu)	UAZ144B(ATPS)
	5.10			
UMC68	5.11/ 5.12	UAZ071B(gfu)	UAZ240A(RNAB)	
GOT2 UMC104B PIO100017	5.13	UAZ268(gfu)		
	5.14			

Figure 2.4

vegetative tissue library (Keith *et al.*, 1993), but mapped during this study. All of these loci represent cDNA clones, as evidenced by either a "gfu" designation (gene function unknown) when sequence data indicated no significant similarities in GenBank, or by a mnemonic when a significant sequence similarity was identified.

It can be seen that our loci are distributed relatively evenly across chromosome 5. There are apparently no significant clusters of expressed genes when compared to the other mapped loci, most of which represent anonymous genomic sequences. On chromosome 5, there are a preponderance of loci mapped to the short arm, but this is also true of the distribution of prior-mapped clones. In fact, we saw little evidence for any distinct differences in the distribution of expressed sequences when compared to either those clones mapped previously or the physical distances of each of these linkage groups as shown in Table 2.3. We compared the percentage of the clones mapped to a particular chromosome and arm with the percentage of the physical map that this segment represents, and there is very close agreement with one exception, chromosome 4. The apparent explanation for this is that most of the α-zein genes are located on this chromosome, and they were initially over-represented in our study before we incorporated our colony prescreening and sequencing efforts. This hypothesis is further verified by noting that of the two arms of this chromosome, 4s is the only one with significant over- representation of expressed genes, again the primary location of the seed storage genes. Examination of other chromosome arms reveals reasonably good agreement between the number of expressed sequences and the physical length of that chromosomal segment. We have also examined our mapping data for the clustering of sequences related by function and find no convincing evidence to date for such a pattern. Evaluation of the map locations of individual carbohydrate metabolic genes, ribosomal protein genes, or heat shock genes has revealed few cases where members are found to map closely together. The zein gene families, notably the α-zeins on chromosome 4 are the lone exception to this pattern.

Table 2.3

Distribution of mapped loci for randomly-selected cDNAs

Chrom	#Clones	%Total	%Phys.	Short Arm		Long Arm	
		Clones	Genome	%Clone	%Phys.	%Clone	%Phy.
1	41	11.8	14.5	51	45	49	55
2	41	11.8	12.4	29	41	71	59
3	28	8.1	11.3	29	33	71	67
4	71	20.4	11.1	70	38	30	62
5	44	12.7	11.1	73	48	27	52
6	28	8.1	7.7	18	13	82	87
7	36	10.4	8.9	33	28	67	72
8	23	6.6	8.9	4	25	96	75
9	18	5.2	7.7	28	33	72	66
10	17	4.9	6.3	29	28	71	72

The locations of all of the mapped loci have been summarized by both genetic and physical methods. The columns refer respectively to :

(1) the ten chromosomes of maize,

(2) the number of mapped loci corresponding to each chromosome,

(3) a value as a percentage of all of the mapped loci,

(4) the length of each chromosome as its fraction of the total physical length of the maize genome as figured from (21), and

(5-8) the genetic and physical comparisons repeated for each chromosome arm.

DISCUSSION

Based upon similarity searches, 39.3% of the clones in our study were significantly similar to entries from the public databases. This percentage is comparable with that observed for the 1517 EST's from *C. elegans* (31%) (Waterston *et al.*, 1992) and 1152 EST's from *A. thaliana* (32%) (Hofte *et al.*, 1993) but much higher than that observed for 830 EST's from rice (8%) (Uchimiya *et al.*, 1992), 2375 EST's from human (17%) (Adams *et al*, 1991; Adams *et al.*, 1992), 130 EST's from maize (19%) (Keith *et al.*, 1993) and 197 EST's from *Brassica napus* (8.9%) (Park *et al.*, 1993). There are several reasons for a apparent differences among these studies. Our results indicated that the percentage of EST's with high similarity to known proteins in the abundantly-expressed group (88%) is much higher than that of the rare group (33%). This is not surprising, in that the abundantly-expressed genes are generally easier to isolate and more likely to have been studied previously in maize. Similarly in mouse, comparison of EST's to the database found that 90% of abundantly-expressed sequences showed a significant similarity, while only 25% of sequences in the rarely-expressed group yielded such a result (Hoog, 1991). Therefore, one might expect our fraction of identified sequences to decrease once all of the abundantly expressed group of clones have been identified and sequenced. We have also shown that sequencing from the 5' terminus of the mRNA instead of the 3' terminus is significantly more informative. Studies which do not utilize directionally-cloned libraries, such as the *Brassica* study (Park *et al.*, 1993), will likely not produce as many significant matches. The standards adopted by each group to define significant matches could also vary, as the research community has yet to define universal values. Whether one was rigorous in defining a "match" as a detection of the homologous gene requiring very high scores or whether a much lower score indicating similarity of function was acceptable could also in part explain the differences among studies.

Among those clones with significant matches in this study, 6.7% were considered identical to known maize genes, 7.0 % represented potentially new maize gene family members and 25.6 % were similar to genes from other species. Each of these groups include examples of both "housekeeping" genes and specifically-induced genes. Several significant matches were somewhat unexpected and may raise new questions in plant biology. For instance, 5C04A12 and 6C06E02 exhibited high similarity to plant genes involved in nodulation, a process considered not to take place in maize (Kouchi and Hata, 1993). Similarly, two A. *thaliana* clones have been identified that show significant similarity to hemoglobinase and urease (Hofte *et al.*, 1993), which are genes usually thought to be specific to legume species. Another clone 5C01C06 exhibited 40% similarity with a vegetative-specific protein from slime molds, which is expressed only in growing cells and deactivated upon the initiation of development in slime mold (Singleton *et al.*, 1991). Genes with specific functions in some species may have been "borrowed" through evolution to form new genes with quite different functions or simply share some common functional domain.

The identification of highly homologous clones to genes previously identified in maize is not the primary goal of this program, but we wished to determine how efficient it was to find related clones that could not be identified by nucleic acid hybridization or to find clones that are similar enough in their sequence to suggest a possible function. This is a more complicated task to accomplish, given the difficulty in distinguishing between weak homologies and "noise" with existing subroutines designed for mass analysis of sequences. We have, however, identified a few interesting clones which are being further analyzed. For instance, a sequence with some homology to the maize Ac transposase was identified, but because of its weak homology (BlastX scores of 87 and 74 for two different regions), it might have been overlooked by a superficial examination of the reports of large numbers of sequences. Interestingly, this entry has a repeated segment of Ser-Ala near the 5' terminus of its sequence in exactly the same position as a Pro-Glu repeat in the original Ac sequence. These genes are obviously distinct enough that our

clone would never have been detected through nucleic acid hybridization and in fact it may represent not only a new gene but a cellular homolog for the transposase. Likewise clones with similarity to several ribosomal proteins from mammals and yeast, an annexin from *Dictyostelium*, signal recognition particle receptor from dog, two proteasome components from human, ribonuclease from *E. coli*, various protein kinases with homologies to yeast, even an essential embryonic gene from chicken were all detected by this approach while nucleic acid similarities are far below what would be required for detection through cross hybridization. We also found new representative clones for previously identified families, such as 5C04C02, which is similar to the MADS family of floral homeotic genes, many of which have been shown to be involved in floral development in *Arabidopsis thaliana* (Ma *et al.*, 1991). Other new representatives for genes previously identified in maize include actin, β-glucosidases, a short chain alcohol dehydrogenase, malate dehydrogenase, and a ubiquitin precursor.

Of the 313 clones, 190 clones did not exibit significant similarity to any sequences in the GenBank database. Motif searching might detect some distantly related genes that exhibit weak amino acid identity by sharing the same functional protein domain. We found that one such program, PROSITE, was very ineffective, evenin analyzing even known sequences where an identification might have been expected. An alternative program, BLOCKS, was effective with known genes and even with some genes not identified by the BLAST routines. Of course, with the rapid increase of submissions to GenBank of genes isolated from other organisms, it is likely that the identifiable genes will substantially increase simply by database searches. In fact, we have found that routine resubmission of our "unknowns" usually resulted in several more identifications, simply due to new additions to the database in the interim. For instance, similarity of two maize cDNA clones to a male sterility gene from *Arabidopsis thaliana* was detected within weeks of its submission of publication (Aarts *et al.*, 1993).

Our expression data indicate that approximately 25% of the endosperm clones were abundantly expressed in this tissue. These clones can be classified into about 30

genes or families based upon hybridization or sequence similarity. Most of these were identified to known genes such as zeins, pyruvate orthophosphate dikinase, ribosomal-inactivating protein, 17.6KD heat shock protein, *Sh2* gene, ubiquitin, peptidyl-prolyl cis-trans isomerase, trypsin/factor XII A inhibitor, ABA-induced glycine-rich protein and sucrose synthase. On the other hand several abundantly-expressed sequences exhibited no significant match with any GenBank entries, which confirm that some major proteins in endosperm are yet to be identified, despite the intensive study of this tissue. About 50% of the clones in both the abundantly-expressed and the rarely-expressed group exhibit complex RFLP patterns (> 4 fragments in a Southern blot), which suggests that there is no significant relationship between expression level and gene copy number in the genome. Some unique-sequence genes were expressed abundantly while some multi-family genes were expressed rarely in endosperm. While such a result is not totally unexpected, this systematic study of abundantly-expressed genes from one tissue and one species shows that gene copy number is not the predominant method for controlling gene expression.

Due to the relatively high sequence variation found within maize (Shattuck-Eidens *et al.*, 1990), it was relatively easy to detect sequence polymorphisms with clones used in this study and hence to map their genomic origins. The primary difficulty in mapping all of these clones resulted from the fact that approximately 50% of them exhibited complex hybridization patterns, and most of these then represented multiple chromosomal origins. In relatively few cases, the multiple fragments detected by a clone mapped to a single chromosomal location. Since not all fragments are polymorphic with any enzyme, attempting to map all of them becomes an economically untenable proposition on such a large scale. Given similar programs in rice, maize and *Arabidopsis*, it may be possible to construct synteny gene maps through comparison of mapping data and sequence similarity of EST's from different species. This would greatly facilitate map-based cloning techniques among other species. If one is interested in cloning a mapped maize gene, the conservation of gene order and composition could allow one to

use a smaller genome, with its relatively low amount of repetitive sequences (e.g. rice) for chromosome walking.

An endosperm library was originally chosen for this study, as this maize tissue has probably been as well characterized as any other tissue from a physiological and biochemical perspective, and more endosperm mutants have probably been analyzed and mapped than any other type. We felt this would enhance our chances of correlating mapped cDNA loci with mutants, but in fact this has turned out to be more problematic than expected. One example that exemplifies this case is the identification of a new sequence with weak homologies to starch synthase (BlastX scores of 72 and 72 to two different regions). This particular entry even exhibits higher homology to a potato and rice sequence than to a maize *waxy* sequence, and it is obviously a newly-identified gene. Interestingly, this gene exhibits a moderately complex hybridization pattern, but one of the fragments maps to 4c near the *sugary* gene, an important gene involved in carbohydrate biosynthesis which had yet to be cloned. We had originally thought that this clone might represent the *su1* gene, but subsequent comparison with a *su1* candidate clone from another effort demonstrates that it does not (M. James, pers. communication). Obviously the use of small numbers of RI mapping progeny do not provide enough resolution to unequivocally correlate candidate cDNA clones with mutant loci. We will need to improve this aspect of the project. Additional cases have been identified, particularly some involved in gene regulation and signal transduction, which need to be followed up. We believe that with refinements, this will develop into powerful approach, particularly with the improvement of sequence homology search routines and better methods to define functional moieties of proteins.

CHAPTER 3

CHARACTERIZATION OF THE eEF1A GENE FAMILY IN MAIZE

INTRODUCTION

eEF1 is a protein synthesis factor composed of four subunits, namely A, Bα, Bβ and Bγ. Subunit eEF1A, also known as EF-1α, binds aminoacyl-tRNAs to the acceptor (A) site of ribosomes during peptide chain elongation (Browning, 1996). Studies on cDNAs or genomic clones for eEF1A have been done in a range of organisms, including yeast (Cottrelle *et al.*, 1985), fungi (Linz *et al.*, 1986), crustaceans (Lendstra *et al.*, 1986), insect (Hovemann *et al.*, 1988), mammals (Brands *et al.*, 1986; Rao and Slobin 1986) and plants. In plants, eEF1A clones were characterized from *Arabidopsis* (Axelos *et al.*, 1989), tomato (Pokalsky *et al.*, 1989), potato (Vayda *et al.*, 1995), soybean (Aguilar *et al.*, 1991), barley (Dunn *et al.*, 1993), wheat (Sacchi *et al.*, 1984) and maize (Berberich *et al.*, 1995).

It appears that eEF1A is involved in multiple cellular processes. It activates phosphatidylinositol 4-kinase (PI4-kinase) on the plasma membrane of carrot suspension cultured cells (Yang *et al.*, 1993), associates with the centromere (Kuriyama *et al.*, 1990) and mitotic apparatus (Ohta *et al.*, 1990) of sea urchin eggs, and binds the endoplasmic reticulum membranes in Chinese hamster fibroblast cells (Hayashi *et al.*, 1989). It also has been shown to interact with a number of proteins, such as the valyl-tRNA synthase complex (Motorin *et al.*, 1988), actin (Yang *et al.*, 1990, Yang *et al.*, 1993), tubulin (Durso and Cyr, 1994) and calmodulin (Kaur and Ruben, 1994). In *Dictyostelium*, bundling of actin by eEF1A was reported to be pH dependent in a physiological range that coincides with the well documented dependency of protein synthesis on pH (Edmonds *et al.*, 1995). Evidence is accumulating that suggests eEF1A is a

multifunctional protein, but the biological significance of many of its interactions is not fully understood.

eEF1A was found to co-localize with protein bodies in maize endosperm (Clore *et al.*, 1996), where its concentration predicts the lysine content of endosperm proteins (Habben *et al.*, 1995). This relationship is also true of other cereal grains. eEF1A was purified and shown to contribute 0.4% of the total protein in W64A+ endosperm and approximately 1% of the protein in W64A*o2* (Sun *et al.*, 1997). The 10% lysine in eEF1A accounts for 2.2% of the total lysine in W64A+ and 2.3% of the lysine in W64A*o2*. One explanation for the correlation of eEF1A concentration and lysine content in maize endosperm is that eEF1A is present in the protein body fraction where it is complexed with actin and other cytoskeleton proteins.

To understand more about the multifunctionality of eEF1A, we characterized the eEF1A gene family in maize. Southern blot reconstruction was performed to determine the number of genes in the maize genome. An endosperm and seedling cDNA library was screened with an eEF1A coding sequence to determined the number of expressed genes. The genes were isolated and classified according to the 3' non-coding sequences. I demonstrated that there are two major transcripts, eEF1Aa and eEF1Ab. Sequence analysis demonstrated that these genes differ by a single amino acid in a region that has been demonstrated to be important for actin binding.

Whether the multiple roles of eEF1A are attributable to expression of different members of the gene family, to isoforms generated by post-translational modifications of one member, or to a combination of the two mechanisms has yet to be determined. The possibility that the functional differences are due to differential gene expression and amino acid sequences were examined in this chapter.

MATERIALS AND METHODS

Plant material and chemicals

Maize (*Zea mays* L.) tissues were from the inbred lines W64A+ and W64A*o2*, which were grown in Tucson, Arizona, between 1995 and 1996. Developing kernels were harvested at 8, 11, 12, 15, 18, and 24 days after pollination (DAP), frozen in liquid nitrogen and stored at -80°C. For RNA extraction, the embryo and scutellum were removed by hand dissection.

Enzymes and reagents were purchased from Life Technologies (Rockville, MD). γ-^{32}P-dATP and α-^{35}S-dATP were purchased from DuPont (NEN - Boston, MA). MagnaGraph nylon membrane was purchased from Micron Separations Inc. (Westboro, MA)

Southern blots

Genomic DNA was isolated from lyophilized maize leaf tissue according to the methods of Taylor and Powell (1982). Digested genomic DNA (20 μg) was separated by electrophoresis overnight in 0.8% agarose, denatured with a solution of 0.5 M NaOH containing 1.5 M NaCl for 45 minutes, transferred to a solution of 0.5 M Tris-HCl containing 1.5 M NaCl for 30 min, and blotted overnight to nylon membrane. The membrane was washed in a solution of 2% SSC, dried at 80°C for 15 min and the DNA was cross-linked to the membrane. Probes were prepared by PCR amplification, using digoxiginin-dUTP in the reaction, with primers flanking the cDNA sequences in the vector (McCreery and Helentjaris, 1994a).

Pre-hybridization, hybridization and chemiluminescent detection of DNA were performed as previously described (McCreery and Helentjaris, 1994b).

Genomic reconstruction

The number of eEF1A genes in the maize genome was estimated by comparing the number and intensity of bands obtained from the hybridization of an eEF1A coding sequence to Southern blots containing 20 µg of maize DNA digested with different restriction enzymes. Samples of an eEF1A cDNA clone containing 5, 10 and 20 picograms of DNA were used to demonstrate the hybridization intensity of 1, 2 and 3 copies of an eEF1A genomic sequence.

Construction of cDNA libraries

An endosperm cDNA library was made with poly(A) RNA isolated from free polysomes of field grown maize endosperm at 14 to 26 DAP (Larkins and Hurkman, 1978). Another cDNA library was made from RNA isolated from 12 day-old-seedlings grown at 30°C. The poly(A) mRNA from both samples was isolated using oligo (dT) cellulose according to manufacture's instructions (Promega - Madison, WI).

The cDNA libraries were constructed with the vector ZAP II Express by directional cloning (5'-*EcoRI* - *XhoI* - 3') according to the manufacture's instructions (Stratagene - La Jolla, CA). The initial titer of the endosperm and seedling cDNA libraries was 3×10^7 pfus/µg and 6×10^6 pfus/µg of cDNA, respectively. The endosperm and seedling cDNA libraries contained 85% recombinant phages with an insert size of 0.5 to 1.5 kb. Both cDNA libraries were amplified once.

Characterization of eEF1A genes

To examine the complexity of eEF1A genes expressed in endosperm and seedling tissues, the endosperm and seedling cDNA libraries were screened with a probe corresponding to the highly conserved coding sequence of eEF1A. The hybridizing

clones were then isolated and their nucleotide sequences determined. Sequencing of the 3' end (Sanger *et al*, 1977) of the double stranded plasmids was with Sequenase and α-^{35}S-dATP. This region was used to distinguish members of the multigene family. The following primers were used for eEF1A sequencing (the numbers correspond to the order of nucleotides in the coding sequence starting at ATG; the "R" designates the reverse orientation of the primer: 1, GGCCGGATCCGCCACCATGGGTAAGGAGAAGAC; 320, GGTCATTGATGCCCCTGGACACC; 320 R, GGTGTCCAGGGGCATCAATGA CC; 480R, CCACCAGTGGTGGAGTC; 610, CTCAAGAAAGTTGGGTAC; 750, CTTGACCAGATACACCGAGCC; 750 R, GGCTCGGTGATCTGGTCAAG; 917, GCACCACGAGACACTTCAG; 1070, CCACCCTGGGCAGATTGGC, 1031R, CAGC CTCCTTGGCAGGGTC; 1250, CACCATACCAGCATCACCG; 1250, ATGATTC CCACCAAGCC; 1350, GGTGACCAAGGCTGCTGCC; 1351R, GGCAGCAGCCTT GGTCACC. Specificity of the 3' end as a probe was determined by hybridization to genomic DNA.

Northern blots

Poly (A) RNA was isolated as previously described from developing ears (8-12 cm), 8, 12, 18 and 24 DAP endosperms, mature leaves, young roots (5 cm), silks and tassel. About 10 g of tissue was frozen in liquid nitrogen, powdered in mortar and pestle and resuspended in 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS, 0.5M NaCl. The sample was extracted once with phenol:chloroform:isoamyl-alcohol (24:24:1) and once with chloroform:isoamyl-alcohol (24:1). RNA was precipitated with two volumes of ethanol, the pellet resuspended in diethylpyrocarbonate (DEPC)-treated water and LiCl added to a final concentration of 2.8 M. RNA was incubated on ice 30 min and concentrated by centrifugation at 10,000 X g. The pellet was washed with 3M LiCl and resuspended in DEPC-treated water. The poly (A) RNA was isolated by oligo (dT) cellulose chromatography according to manufacture's instructions (Promega - Madison,

WI). Specific eEF1A probes were synthesized from the 3' non-coding sequence of eEF1A cDNAs obtained from endosperm. The levels of mRNAs in developing endosperm were assessed by Northern blot analysis using RNA from W64A+ and W64Ao2 endosperm at different developmental stages and also with RNA from different tissues of these plants. The standard deviation of transcript levels was determined by comparing the results from three independent Northern blots. The level of ubiquitin conjugating enzyme mRNA was used to normalize the amount of eEF1A mRNA. The measurement of Northern blot signals was based on scanning various exposures of X-ray films.

Estimation of eEF1A transcript levels

The expression pattern of five eEF1A genes in various maize tissues was estimated by quantitative RT-PCR. All RNA samples were treated with DNAse I (1 unit /µl) for 30 min at 37°C to avoid any contamination of genomic template in the RT-PCR. After DNAse treatment, the samples were incubated for 10 min at 70°C to inactivate the DNAse, extracted once with phenol:chloroform:isoamyl-alcohol (24:24:1), once with chloroform:isoamyl-alcohol (24:1), precipitated with two volumes of ethanol, the pellet washed with 70% ethanol and resuspended in diethylpyrocarbonate (DEPC)-treated water. To confirm the efficiency of the DNAse treatment, a 30 cycle PCR reaction was performed with the DNAse treated RNA. The first strand cDNA was prepared from 20 µg of DNAse treated RNA primed with oligo(dT) according to manufacture's instructions (GIBCO-BRL - Gaithersburg, MD). Before the RT-PCR reaction, the first strand cDNA was treated with 10 µg/ml of RNAse A for 30 min at 37°C. The PCR reaction was for 15 cycles: 95°C, 30s; 60°C, 30s and 72°C, 30s, and included two sets of primers in the same reaction. One reaction compared the level of total eEF1A to the level of ubiquitin in all tissues, and the other reaction compared the level of the total eEF1A to each of the five members of the eEF1A gene family. The number of PCR cycles for the RT-PCR were

kept very low (under 15), because some amplifications could give preference to some fragments and the results be misinterpreted. The 3' end primers of the RT-PCR were: eEF1Aa, GAAGGTGAAATAGCAAGAAGG; eEF1Ab, GCCAGACGCTGCGCTACC; eEF1Ac, CAACTGCAAGGATGTCAAC; eEF1Ad, TGCAAGTCAACCATGTCTC and eEF1Af, GCACAGGGGCGAGTGGC. The 5' end general eEF1A primer (GATGATACCCACCAAGCC) was located at nucleotide 1063 in the coding sequence. PCR products of the right size were detected only with primers corresponding to the appropriate plasmids after 30 cycles of amplification. Primers corresponding to a conserved eEF1A sequence between nucleotides 611 (GTCCACCAACCTTGACTGG) and 955R (CTGCCTCCTTGGCAGGGTC) were used in the same reaction with those for the 3' end-specific region. The 5' and 3' primers for the ubiquitin gene (U29163) (CCCTTACTGGCAAGACCATG and CACGAAGATCTGCATTCCTC) and α-zein gene (G22531) (TTCGAACATCCAGC CGTGC and CACCATGGCTAGGTGGCTC) were used as controls to normalize first strand cDNA synthesis and confirm the genotypes, respectively. All the 5' primers for the PCR reaction were labeled with γ-^{32}P-ATP. RT-PCR reactions were done with W64A+ and W64Ao2 developing endosperms at 11, 15, 18, and 24 DAP and with developing ear (12 cm), mature leaf, young root (5 cm), silk and mature stem RNAs. The products of each PCR reaction were confirmed by Southern blot with gene-specific probes labeled with γ-^{32}P-ATP. The conditions for the Southern blots were: pre-hybridization and hybridization solution (5 X SSC, 5% dextran sulfate, 0.1% SDS, 0.1% N-lauryl sarcosine, 0.1% I-Block) at 50°C for 1 and 16 hours, respectively. Filters were washed twice (2X SSC; 0.2% SDS) at room temperature for 5 min and twice at 50°C for 15 min each. The products of the reactions were separated in TBE-denatured polyacrylamide gel (50% urea, 1.8%Tris, 0.65% boric acid, 0.37% EDTA, 5% acrylamide, 0.25% bisacrylamide), and the gel dried and exposed to X-ray film. The quantification of the radioactive signal was performed on the Storm 869 Phosphoimager and the analysis of the data by the ImagequaNT software (Molecular Dynamic, Sunnyvale, CA).

To standardize the PCR reactions and measure the coding and 3' end noncoding eEF1A sequences, plasmids containing each cDNA clone were transcribed *in vitro* with T3 RNA polymerase (GIBCO-BRL), and the RNA products were reverse transcribed to produce a cDNA. RT-PCR was performed with a serial dilution of the cDNA product to determine if the products of the RT-PCR were in a linear range of measurement.

RESULTS

Southern blot analysis was performed to determined if eEF1A in maize is encoded by one or multiple genes. The genomic reconstruction showed that in maize the eEF1A gene family is composed of 10 to 15 members (Fig. 3.1). At least 7 bands with different intensities were apparent after *HindIII* digestion and hybridization with the eEF1A coding sequence. The average size of the fragments generated after *BamHI* and *EcoRI* digestion was larger than that generated with *HindIII* digestion, and the fragments could not be used as easily to estimate the number of genes. The number of eEF1A genes was similar in different genetic backgrounds (data not shown).

To assess the eEF1A genes expressed in maize endosperm and seedlings, an eEF1A coding sequence identified in the maize genome project (5C04H09, see Chapter 2) was used as a probe to screen cDNA libraries from these tissues. It was assumed that the probe would identify all the eEF1A expressed genes. Approximately 144 eEF1A clones, corresponding to 1% of the total number of plaques, were isolated and the 3' end-sequence determined. The cDNAs were separated into ten groups according to the 3' end-sequence. The frequency of each cDNA clone is shown in table 3.1. The groups were named cEF1Aa-j. Based on sequencing and restriction endonuclease analysis, eEF1Ah, eEF1Ai and eEF1Aj are not full length clones (1.2, 1.3 and 0.9 kb, respectively).

In the endosperm and seedling cDNA libraries, eEF1Aa and eEF1Ab represented 80% of the total eEF1A mRNA (Table 3.1). eEF1Ad, eEF1Ai and eEF1Aj were found exclusively in the endosperm cDNA library, and eEF1Ae, eEF1Af, eEF1Ag and eEF1Ah were found exclusively in the seedling cDNA library.

The nucleotide and amino acid sequence of seven members of the eEF1A gene family are shown in the figures 3.2 A and B. respectively. The 5' non-coding sequence (~80 bp) of the full length eEF1A clones was found to be shorter and more conserved than the 3' non coding sequence (200 bp) from all the eEF1A gene members examined.
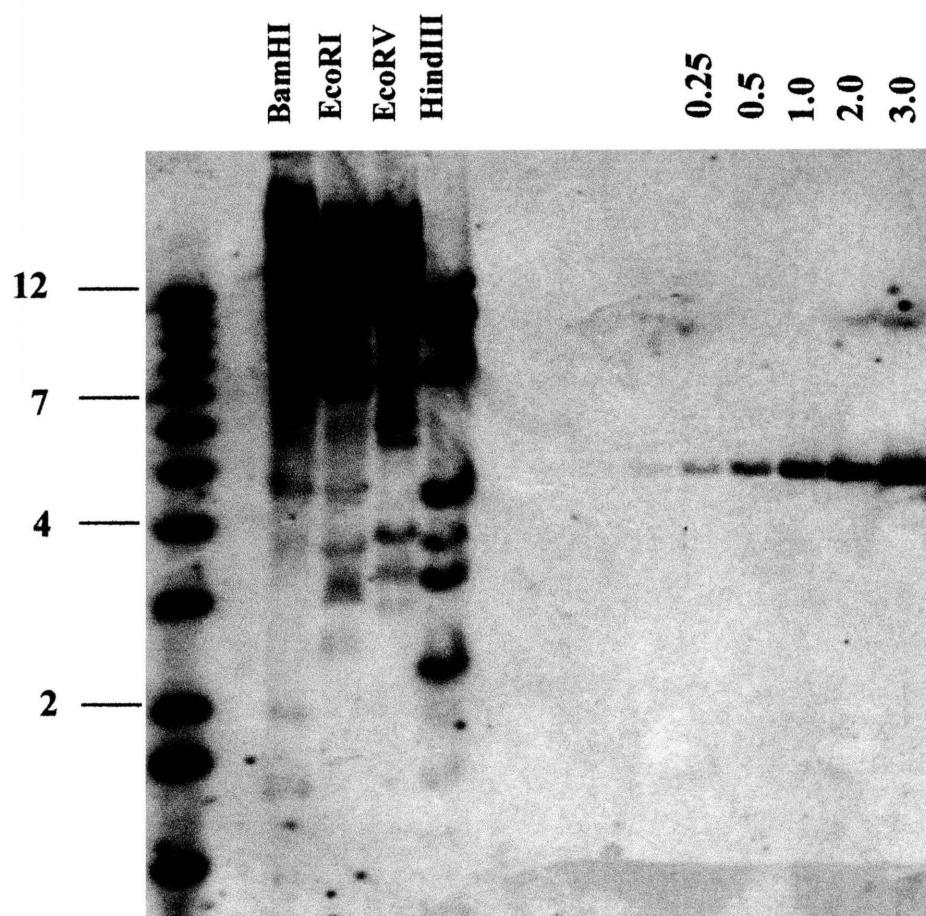
Figure 3.1 - Estimation of eEF1A gene copy number in W64A+. The number of eEF1A genes in the maize genome was estimated by comparing the number and intensity of bands obtained from the hybridization of an eEF1A coding sequence to Southern blots of 20 µg of W64A+ DNA digested with different restriction enzymes. Gene copy number was estimated by a serial dilution of an eEF1A cDNA sequence (1340 nucleotides). The hybridization signal was detected with a dig-labeled probe. The numbered lanes illustrate the hybridization intensity of 0.25, 0.5, 1, 2 and 3 copies of an eEF1A gene.

Table 3.1

Frequency of different eEF1A cDNA clones in endosperm and seedling cDNA libraries.

Clone	Frequency	
	Endosperm	Seedling
a	49	26
b	29	15
c	9	4
d	3	-
e	-	2
f	-	2
g	-	2
h	-	1
i	1	-
j	1	-

Figure 3.2 - Nucleotide (A) and amino acid (B) sequences comparison of 7 members of the maize eEF1A gene family. Nucleotide sequences were determined by dideoxy sequencing of the double stranded plasmid using Sequenase and α-^{35}S-dATP.

A

```
eEF1Aa                                            C GCCTCGCCGT CTGCCTTCTG CGCCTCCATT TCGGCCTCTG TCCCTTGCAA GTTCAATCTC ACCTCCAACC
eEF1Ab                                    TGCCTTCTGC GCCCTCCATT TCGGCCTCTG TCCCTTGCAA GTTCAATCTC ACCTCCAACC
eEF1Ac                          CGCTTC GCCGTCTGCC TTCTGGGCCT TCCTTTTTAA CGGCCTCTAT CCCTAGCAAG TTCATTCTCA CGGTCCAACC
eEF1Ad                                                       T GTAAGGTAAG ATTAAGCTTA ACCTCCAACC
eEF1Ae                   GGACGAGC TCTCGACCGC CTCCTCTCCC CTCCCCCGGC TCTGTCTGTT GACTTACCCC GTCAGCTTTC TAGTCAGCTG CTCCCGTTGC
eEF1Af                            CGTCGTCTC TGCGGTCTGC GTCTAGCCTC AATTGCCTCT CTCCCTGCAG TTAAGCTTCC ATCTTCAGCC
eEF1Ag  TGGGGCACGA GCAAGTCGCC GTCGTCTTCT GCGGTTCCTG CCCGTCTACG CCCTCCAATT TGGCCTCTCT CCCTTGCAAG TTTAAGCTTC ATCTTCAGCC

CONSENSUS  ATGGGTAAGG AGAAGTCCCA CATCAACATT GTGGTTATTG GCCATGTCGA CTCTGGCAAG TCGACCACCA CCGGCCACCT GATCTACAAG CTTGGAGGCA
eEF1Aa     ........A. .......... .......... .......... .......... .......... .A..A..... T......... ..........
eEF1Ab     ........A. .......... .......... ....C..T.. .......... .......... .A..A..... .......... ..........
eEF1Ac     ........C. .......... .......... ...C...... .......... .......... .A..A..... T......... ..........T
eEF1Ad     .....C.... .......... .......... .......... .......... ...A...... .......... .......... ..........
eEF1Ae     .....G.... .......... .......... .....C.... .......... ...C..... ..C..A.... .......... C......... ..........
eEF1Af     .......... .......... .......... .....C.... .......... .......... .......... .......... ..........
eEF1Ag     .......... .......... .......... .....C.... .......... .......... .......... .......... ..........

CONSENSUS  TTGACAAGCG TGTGATCGAG AGGTTCGAGA AGGAGGCTGC TGAAATGAAC AAGCGGTCCT TCAAGTACGC TTGGGTGCTC GACAAGCTCA AGGCTGAGCG
eEF1Aa     .......... .......... .......... .......... .......... .......... G......... .......... ..........
eEF1Ab     .......... .......... .......... .......... .......... .......... A......... .......... ..........
eEF1Ac     .......... .......... .......... .......... .......... .......... G......... .......... ..........
eEF1Ad     .......T.. .......... ....A..C.. .......A.. .......... .......... ......T... .......... ..........
eEF1Ae     .......... .......T.. .......... C..G..... ...A...... .......... ......C..T .......... ...C.....
eEF1Af     ....A..A.. ....T.... .......... .....G.... .......... .......T.. .......... .......... ..........
eEF1Ag     ....A..T.. ....T.... .......... .....G.... .......... .......T.. .......... .......... ..........

CONSENSUS  TGAGAGAGGT ATCACCATTG ATATTGCTCT GTGGAAGTTT GAGACCACCA AGTACTACTG CAC-GTCATT GATGCCCCTG GACACCGTGA CTTCATCAAG
eEF1Aa     .......... ......C... .......... .......... ...G...... .......... .......... ..........
eEF1Ab     .......... ......C... .......... .......... ...G...... .......... .......... ..........
eEF1Ac     ......T... ......C... .......... .......... ...G...... ...G.... ........C. ..........
eEF1Ad     .......... .......... .......... .......... ...T...... .......... ...T..C.. ..........
eEF1Ae     C.......C. .C.....C. .......... .......... ...C...... ...C..... ..........
eEF1Af     ....G... .......... .......... .......... ..T..... ...C...... .......... ..........
eEF1Ag     .......... .......... .......... .......... ..T..... ...C...... .......... ..........

CONSENSUS  AACATGATCA CTGGTACCTC CCAGGCTGAC TGTGCTGTCC TTATCATTGA CTCCACCACT GGTGGTTTTG AGGCTGGTAT CTCCAAGGAT GGCCAGACCC
eEF1Aa     .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ab     .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ac     .......... ..C....... .......... .......... ...A...... .......... .......... ..........
eEF1Ad     .......G.. ...T...... .......T.. .......... .......... .......C.. ...T...... ..........
eEF1Ae     .......... ...C...... .......T .......... .......... ...C...... .......... ...C......
eEF1Af     .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ag     .......... .......... .......... .......... .......... .......... .......... ..........

CONSENSUS  GTGAGCATGC TCTCCTTGCG TTCAC-CTTG GAGTGAAGCA GATGATTTGC TGCTGCAACA AGATGGATGC AACCACTCCC AAATACTCCA AGGCACGTTA
eEF1Aa     ....A..... .......... .....C.... .......... .......... .......... .......... ..........
eEF1Ab     ....A..... .......... .....C.... .......... .......... .......... .......... ..........
eEF1Ac     .......... ........T .....A..... .T........ .......... .......... T..T.A... ..G..T.... ...C.....
eEF1Ad     .......... ........T ...C..... .G..C..... .......... .......... .......T ... ...C.....
eEF1Ae     .......... ..C.....T .....T.... .T....G... .......C.. .......... C..T..C... ..G.......
eEF1Af     .......... .......... .....T.... .......A.. .......... .......... ..G....... ..........
eEF1Ag     .......... .......... .....T.... .......A.. .......... .......... ..G....... ....G.....

CONSENSUS  TGATGAGATT GTGAAGGAAG TCTCATCCTA CCTCAAGAAA GTTGGGTACA ACCCTGATAA GATTGCCTTT GTTCCCATTT CTGGTTTTGA GGGCGACAAC
eEF1Aa     ...A...... .......... .......... .......... .......... .......... .......... ..........
eEF1Ab     .......... .......... .......... .......... .......... .......C.. .......... ..........
eEF1Ac     .......... ....T..... .......... .......A.. .......... .......... .....A..C. ..........A..T.....
eEF1Ad     C......... .T........ ........G .....A.... .......C. ..C..T..C. .......... ...T......
eEF1Ae     ...A...... .......... ....T..... .......... .......... .......... ..........
eEF1Af     ...A...... .......C. .......... .......C.. .......... .......... ......A... ..........
eEF1Ag     .......... .......... .......... .......... .......... ..C.....C. .......... .....C....

CONSENSUS  ATGATTGAGA GGTCCACCAA CCTTGACTGG TACAAAGGCC CAACCCTGCT TGAGGCTCTT GACCAGATCA CCGAGCCCAA GAGGCCTTCA GACAAGCCCC
eEF1Aa     .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ab     .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ac     .......... .......... .....G.... .......A.. .......... ...A...... .......G.. .........C........
eEF1Ad     .......... .......... .......T. .......... .......... .......... .......... ..........
eEF1Ae     .......... .......... .....T.... .......... .......C.. .......... .......... .........A
eEF1Af     .......... .......... .......... .......... .......... ...A.... ......G... ..........G.
eEF1Ag     .......... .......... .......... .......... .......... .......... ......G.. ..........
```

Figure 3.2A (cont.)

```
CONSENSUS  TGCGTCTAGC CCTCCAGGAT GTGTACAAGA TTGGTGGTAT TGGAACTGTA CCGGTTGGTC GTGTGGAGAC TGGTGTCATC AAGCCTGGTA TGGTAGTCAC
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ac   .......... .......... .......C... .......... ...G...... .......... .......G.. .......... .......... ..........
  eEF1Ad   .......... .......... .......... .......... .......T..A .......... .......... .......... .......A.. ....T.....
  eEF1Ae   .......... .......... .......... .......... .......T.. .......... .......... .......... .......T.. ..........
  eEF1Af   .......... .T........ .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ag   .......CC. ...G...... .......... .......... .......... .......... .......... .......... .......... .....T....

CONSENSUS  CTTTGGTCCA ACTGGCCTGA CTACCGAGGT GAAGTCTGTT GAGATGCACC ACGAGGCACT TCAGGAGGCT CTTCCGGGTG ACAATGTTGG CTTCAACGTG
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   .......... .......... .......T... .......... .......T...G.. .......... .......... .......C.. .......... ..........
  eEF1Ac   G......... ...A...... .......... .......... .......G.. .......... .......... .......... .......... ..........
  eEF1Ad   ...C...... .......... ..C..T.... .......... .......... .......T.. C.........C ......T.... .......... ......T...
  eEF1Ae   .......... .......... .......... .......... .......... .......A.. .......... .......... .......... ......T..
  eEF1Af   .A........ .......... .......... .......... .......... .......... .......... ......A... .......... ..........
  eEF1Ag   ...C...... .......... ......T... .......... .......... .......... .......... .......... .......... ..........

CONSENSUS  AAGAATGTTG CTGTCAAGGA TCTCAAGCGT GGGTTTGTGG CCTCCAACTC CAAGGATGAC CCTGCCAAGG AGGCTGCCAG CTTCACCTCC CAGGTCATCA
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   .....C.... .......... .......... ......C... .......... .......... .......... .......... .......... ..........
  eEF1Ac   .......... .......... .......... .......... .......... ..G....... .......... .......G.. .......... .....A....
  eEF1Ad   .....C.... ...G...... ......T... .......... .......... .T........ .......... ......T... .......... .....A....
  eEF1Ae   .......... .......... .......... .......... .......... .T........ .......... ......T... .....G....
  eEF1Af   .......... .......... .......... .......... .......... ..C..... .......... .......... .......... ..........
  eEF1Ag   .....C.... .......... .......... ...A...... .......... .......... .......... .......G.. .......... ..........

CONSENSUS  TCATGAACCA CCCTGGGCAG ATTGGCAACG GCTATGCCCC TGTGCTGGAC TGCCACACCT CCCACATCGC TGTCAAGTTT GCTGAGCTCA TTACCAAGAT
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   .......... .......... .......... A......... .......... .......C.. .......... C......... .......... ..........
  eEF1Ac   .......... .......... G......... .......... .......... .......C.. .......... .......... .......... .......G...
  eEF1Ad   .......... .......... .......T.. A......... .......... .......... .......T.. .......... .......... ..........
  eEF1Ae   .......... ...G...... ..C....... ......C... .......... .......G.. .......... C..G...... .......... ..........
  eEF1Af   .....C.... .......... .......... .......G.. .......... .......... .......... .......... .......... ..........
  eEF1Ag   .......... .......... ..A.....T. .......... A......... .......... .......... .......... .......... ..........

CONSENSUS  CGACAGGCGC TCTGGCAAGG AGCTTGAGAA GGAGCCAAAG TTCCTGAAGA ACGGTGATGC TGGTATGGTG AAGATGATAC CCACCAAGCC TATGGTGGTG
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   ........A. .......... ...C...... .......C... .......... .......... .......... .......... ......T... ..........
  eEF1Ac   .......... .......... .A........ .......... .......... ......T... .......... .......... .A........ ..........
  eEF1Ad   .......... .......... .......... .......... .......T.. .......C.. .......... ......T... ...T....C .......T
  eEF1Ae   T...C....G ..G....... ...A.GAG.. .......... ..T..C.... .C........ C...T.C... ......G.C. ...G...... C.........
  eEF1Af   .......... .......... .......A.. .......... .......... .......... .......... .......... .......... ..........
  eEF1Ag   .......... .......... .......... .......... ......G... .......... .......... .......... .......... ..........

CONSENSUS  GAGACATTCT CCGCGTATCC TCCCCTGGGT AGGTTTGCCG TCCGCGACAT GAGGCAGACG GTTGCTGTTG GAGTCATCAA GAGTGTGGAG AAGAAGGACC
  eEF1Aa   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ab   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  eEF1Ac   .......... .......... ...T...... .......... ....T..... ......A... ...G....C. .......... .......... ..........
  eEF1Ad   .....C.... .TCA...... ...T...... C.T.....T. ...T...... .......... .......... ...C...... ..........
  eEF1Ae   ....T..... ....A..... ...A...... ......G... .......... .......... A..C. .C..G...... .......... ..........
  eEF1Af   .......... .......... .......... ......A... .......... .......... .......... .......... .......... ..........
  eEF1Ag   .......... .......T... .......T.. ........T. .......... .......A.. .......... .......... .......... ..........

CONSENSUS  CAACCGGCGC CAAGGTGACC AAGGCGGCCG CCAAGAAGAA A
  eEF1Aa   .......... .......... .......... .......... .TGATGCGAT CCCTGCGCCT GCTTTAGCAA TACCCTAGTT TCCATCATTA CAAGTTTGTT
  eEF1Ab   .......... .......... .......... .......... .TGATCGATC CGTGTCTGCT TTGCAGTGCC TGGATAACTT GCCACTAAAG TTGTGTATGC
  eEF1Ac   ....T..A.. ........G. .....T..T. .......... .TGATCTATC TGGCTGGTGC TGTGCTTGCT TGTTGGTTAC CTCGTGGAAG CTTAAAATAT
  eEF1Ad   ....T..T.. .......... .......... .T...AA TA....... .TGATTGTC TCCCATGTGC CCTCGTTTAT TTTG ACTTG TGCCCGTATA CTTCGATCTG
  eEF1Ae   .......... .......... .......... .......... .TGATGATGA TGATCTGAAT GTAATCCATG GTTCCCTGGT TTCTTTGCTG GTCGGAGTAT
  eEF1Af   .G........ ......T... .....A..T. ....A..... .TGATGCAAC CTGTCGTTAA TCCTGTATCT GCTTTGCAAT ACCTGACTTG TCACTCCATC
  eEF1Ag   .G........ .......... .......... .......... .TGATGGAGT TTGCATGCTG ATGCGGTTAG CTAGCTAGCT AGCTAGGCTA GCTAGCTACG

  eEF1Aa   TGTGGTCGTT GCTGTTATTG TGTGAACTGT TGAGCTCTGT TAGCCTGTGC ACTTTTATCT ATTTATTTGG TACCTTCTTG CTATTTCACC TTCTGCAATA
  eEF1Ab   GGTTGTTGCT GTTATTGCGT GGACTGTTTG CTGCGGTGTC TCCTGTTTGT TTGGGTAGCA CAGCGTCTGG CCTCTTCAGG TGTCTGTTGT CTGTTGTCTT
  eEF1Ac   CTCGTCTTGT TGACATCCTT GCAGTTGTCG CGTGAACAAG TAATCTATGA TGGCAATGTT TGCCTGGTCA TGAAGTATTA TTACTTTCGT TGAACCGTGA
  eEF1Ad   TTAAAATGTA TTGGTTGTGT GCCCGTGTCT GTTGTTTGGG TATCTATACT GTTATTTTGA GACATGGTTG ACTTGCATGG TCTCTGGATT GGGTCGCATT
  eEF1Ae   TTTGTTGGGT GGTGCTAGTA GTAGTTAACA AAAGGATAAT TTTCATGGGT TCATTGTCTA CCTTTGGACT GTTGTGTTCG ATGATTGCGT TTTTGCTTAC
  eEF1Af   CGAAGGTTTT GTATGCGGTG TACTGTTATT GTGTACTGTT AAGCTCTGCT GCGGCGTGCC CTGTTGTTTG GGTAGCGTCA ACGTGAGCG CCACTGCGGG
  eEF1Ag   ATCGATCAGC TACGTATTGT CATTGTTGGC GTATGTTTTG AAGCCCTACC TCACTGTTCA GTCCATAACG TTTTGGTCTA TGTCCTAATG TG

  eEF1Aa   CAAGAATGCT GTAAGAGCTA TATGTTAAC
  eEF1Ab   TACTATCAGT AATGCTCCTG AGACTTCTAT TTGTCCTTTT GGCG
  eEF1Ac   TGCTATTAAG ACTTATTTTT TATTGGAACA CTA
  eEF1Ad   TCTGAGTCAT TTGCTG
  eEF1Ae   CACACTATAT ATATTGGG
  eEF1Af   GTGTGCCTTT GTCTGCTACAG TAG CGG
```

Figure 3.2B (cont.)

B

```
eEF1Aa  MGKEKSHINI VVIGHVDSGK STTTGHLIYK LGGIDKRVIE RFEKEAAEMN KRSFKYAWVL DKLKAERERG ITIDIALWKF ETTKYYCTVI DAPGHRDFIK
eEF1Ab  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ac  ..N....... ....P..... .......... .......F.... ......P... .......... F......... .......... ..R...H...
eEF1Ad  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ae  .......... .......... .......... .......... .......... .......... .......... .......... .......... ....P.....
eEF1Af  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ag  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Aa  NMITGTSQAD CAVLIIDSTT GGFEAGISKD GQTREHALLA FTLGVKQMIC CCNKMDATTP KYSKARYEEI VKEVSSYLKK VGYNPDKIAF VPISGFEGDN
eEF1Ab  .......... .......... .......... .......... .......... .......... ......D... .......... .......... ..........
eEF1Ac  ....R..... .......... .......... .......... .......... .......... ......D... .......... .......... ..........
eEF1Ad  .......... .......... .......... .......... .......... .......... ......D... .......... .......... ..........
eEF1Ae  .......... .......... .......... .......... ....R.... .......... .......... .......... .......... ....Y....
eEF1Af  .......... .......... .......... .......... .......... .......... ..D....... .......... .......... ....Y....
eEF1Ag  .......... .......... .......... .......... .......... .......... ......D... .......... .......... .......A.
eEF1Aa  MIERSTNLDW YKGPTLLEAL DQITEPKRPS DKPLRLALQD VYKIGGIGTV PVGRVETGVI KPGMVVTFGP TGLTTEVKSV EMHHEALQEA LPGDNVGFNV
eEF1Ab  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ac  .......... .......... .K...A.... H......... ..Q....... ......A... .......... .......... ..Q....... ..........
eEF1Ad  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ae  .......... .N......P .......... ...M...... ......F... .......... ..V....... .......... ......I... .........L
eEF1Af  .......... .......... .......... ......F. .......... .......... .I........ .......... .......... ..S.......
eEF1Ag  .......... .......... .......... ......P.. .......... .......... .......... .......... .......... ..........
eEF1Aa  KNVAVKDLKR GFVASNSKDD PAKEAASFTS QVIIMNHPGQ IGNGYAPVLD CHTSHIAVKF AELITKIDRR SGKELEKEPK FLKNGDAGMV KMIPTKPMVV
eEF1Ab  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
eEF1Ac  .......... .......R.. .....G.... .......... V......... ..P....... .....R.... .......... ...Y...... ..........
eEF1Ad  .......N. .......... .......L... .......... .......... .......... .......... .......... .......... ..........
eEF1Ae  .......... .......L.. .......... .......... .......... .......... .......... ......GE... ........F. ..V.......
eEF1Af  .......... .......H. .......... ......T.... .....G.... .......... .......... .......... .......... ..........
eEF1Ag  .......... .......... .......... ......P. .......... .......... .......... .......... ...D...... ..........
eEF1Aa  ETFSAYPPLG RFAVRDMRQT VAVGVIKSVE KKDPTGAKVT KAAAKKK
eEF1Ab  .......... .......... .......... .......... .......
eEF1Ac  .......... .......... .......... .......... .......
eEF1Ad  ....Q..... ....L..... .......... .......... ...I...
eEF1Ae  ...E..... .......... .......... .......... .......
eEF1Af  .......... .L........ .......... .......... .......
eEF1Ag  .....F.... .......... .......... .......... .......
```

The identity on the amino acid and nucleotide levels are shown in the tables 3.2 and 3.3, respectively.

To determine if the eEF1A genes could be distinguished by hybridization, a Southern blot containing maize genomic DNA digested with *BamHI*, *EcoRI* and *HindIII* was hybridized with dig-labeled 3' end-specific probes made by PCR. The PCR was done using a conserved primer corresponding to the last six amino acids of the eEF1A coding sequence and the Bluescript Forward primer corresponding to a vector sequence located downstream of eEF1A clone. The hybridization was found to be specific for each eEF1A member tested (Fig. 3.3).

Levels of different eEF1A mRNAs were examined in developing W64A+ and W64A*o2* endosperms by Northern blot hybridization using specific 3' end non-coding sequences as probes (Fig. 3.4). Computer software was used to measure the hybridization intensity as the difference of pixels in the autoradiographs from the Northern blots (Fig. 3.5). No hybridization signal was detected by Northern analysis with eEF1Ac-h genes in any tissue. Statistical analysis using the t-test at 95% confidence demonstrated that total eEF1A transcript levels are different between W64A+ and W64A*o2* at 12, 18 and 24 DAP (Table 3.4). eEF1Aa transcript levels were different at 8 and 18 DAP in W64A+ and W64A*o2*. eEF1Ab transcript levels were not significantly different at any endosperm developmental stage. Total eEF1A, eEF1Aa and eEF1Ab transcripts were also compared in developing ear (12 cm), W64A+ and W64A*o2* endosperms at 18 DAP, mature leaf, young root (5 cm), silk and tassel by Northern blot analysis (Fig. 3.6). According to the results, eEF1Aa and eEF1Ab are expressed in all tissues examined. The expression of eEF1Aa and eEF1Ab were higher in endosperm, roots and silk.

Transcript levels of total eEF1A and different members of the eEF1A gene family were verified by RT-PCR in different tissues and in W64A+ and W64A*o2* developing endosperm. RT-PCR is a more sensitive method to detect low levels mRNA than Northern blots. Measurements of the transcript levels of each member of the

Table 3.2

Amino acid sequence identity of 7 members of the eEF1A gene family. The numbers represent the percentage of amino acid sequence identity among members of the eEF1A gene family.

	eEF1Aa	eEF1Ab	eEF1Ac	eEF1Ad	eEF1Ae	eEF1Af	eEF1Ag
eEF1Aa		99.78	95.14	98.87	96.56	97.96	98.87
eEF1Ab			95.38	99.10	96.33	97.73	99.10
eEF1Ac				94.42	91.46	92.96	94.66
eEF1Ad					95.62	96.80	98.19
eEF1Ae						94.42	95.38
eEF1Af							96.80
eEF1Ag							

Table 3.3

Nucleotide sequence identity of 7 member of the eEF1A gene family. The numbers represent the percentage nucleotide sequence identity among members of the eEF1A gene family.

	eEF1Aa	eEF1Ab	eEF1Ac	eEF1Ad	eEF1Ae	eEF1Af	eEF1Ag
eEF1Aa		98.65	95.70	93.62	93.62	97.11	96.80
eEF1Ab			94.43	93.30	92.72	95.94	96.25
eEF1Ac				91.40	90.56	93.38	93.21
eEF1Ad					89.53	92.06	93.30
eEF1Ae						92.72	92.39
eEF1Af							96.88
eEF1Ag							
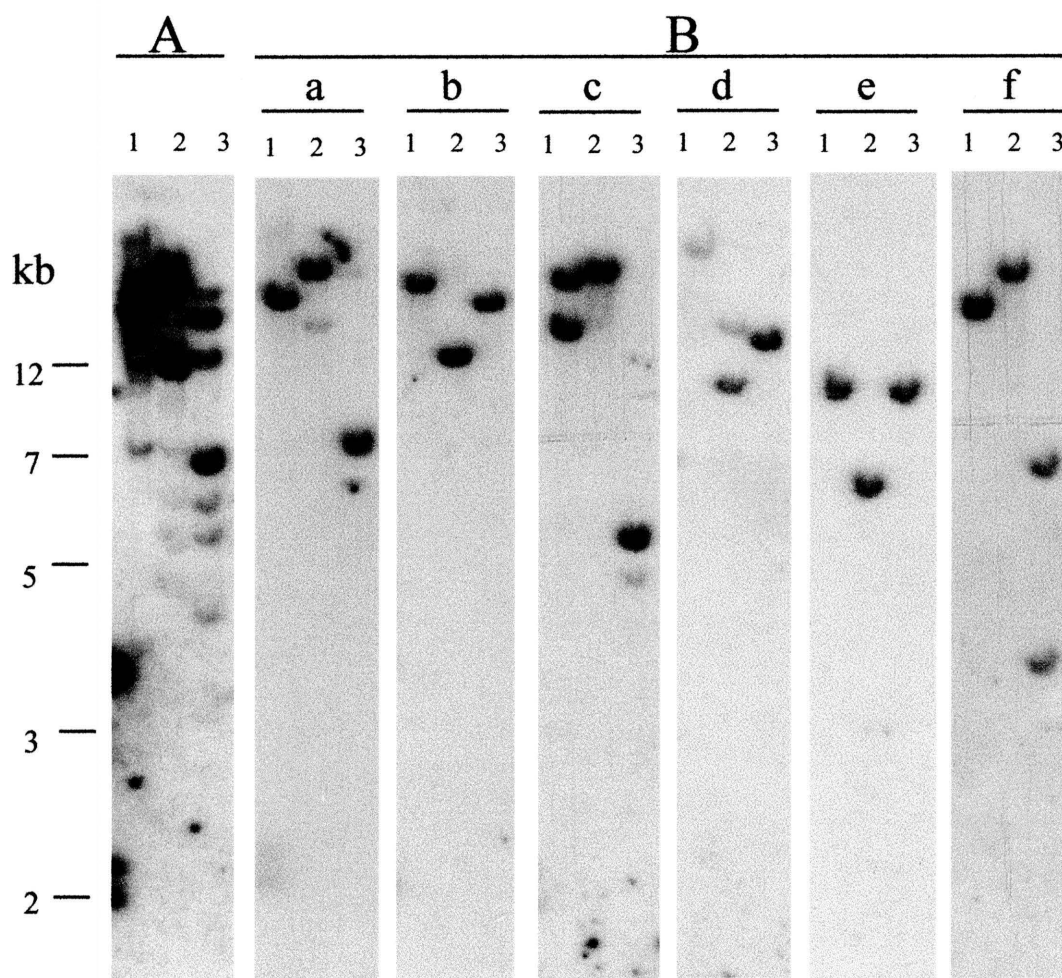
Figure 3.3 - Identification of eEF1A genes in W64A+. DNA from W64A+ was digested with *BamHI* (1), *EcoRI* (2) and *HindIII* (3) and separated by electrophoresis in 0.8% agarose. Following Southern blotting, the DNA was probed with an eEF1A coding sequence (A) or the 3' non-coding sequence of various eEF1A cDNA clones (B). Lower case letters in B correspond to cDNAs listed in Table 3.1.
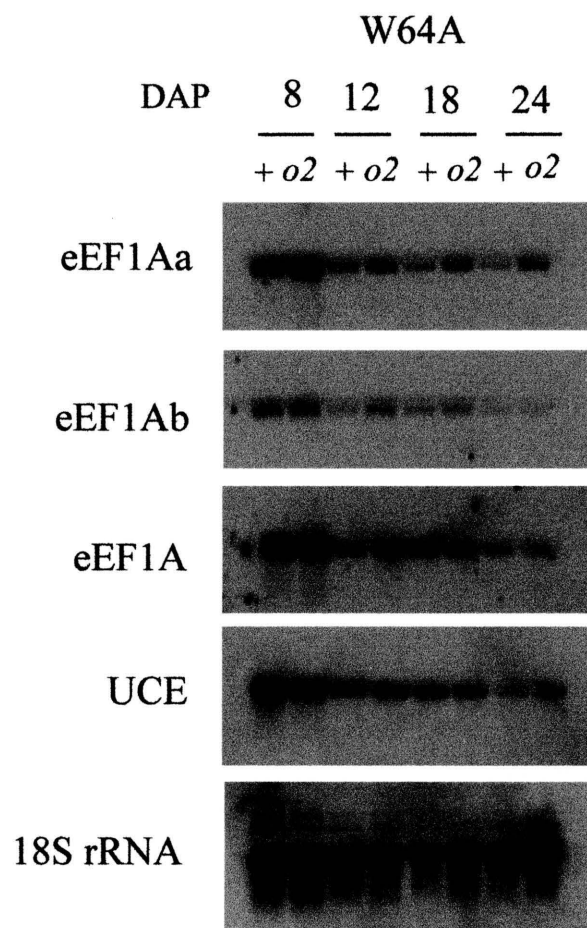
Figure 3.4 - Detection of eEF1A transcripts at various developmental stages of W64A+ (+) and W64Ao2 (o2) endosperm. Each lane contained 5 µg of poly(A) RNA which was separated by 1.2% agarose gel eletrophoresis and blotted onto nylon membrane. Probes corresponding to the 3' non-coding regions of eEF1Aa, eEF1Ab, the coding sequence of eEF1A, ubiquitin-conjugating enzyme (UCE) and 18S rRNA. DNAs were labeled with the modified base, dig-dUTP, by PCR and hybridized to the membranes as described in Material and Methods. The numbers (8, 12, 18 and 24) correspond to days after pollination

W64A

| A | B | C |
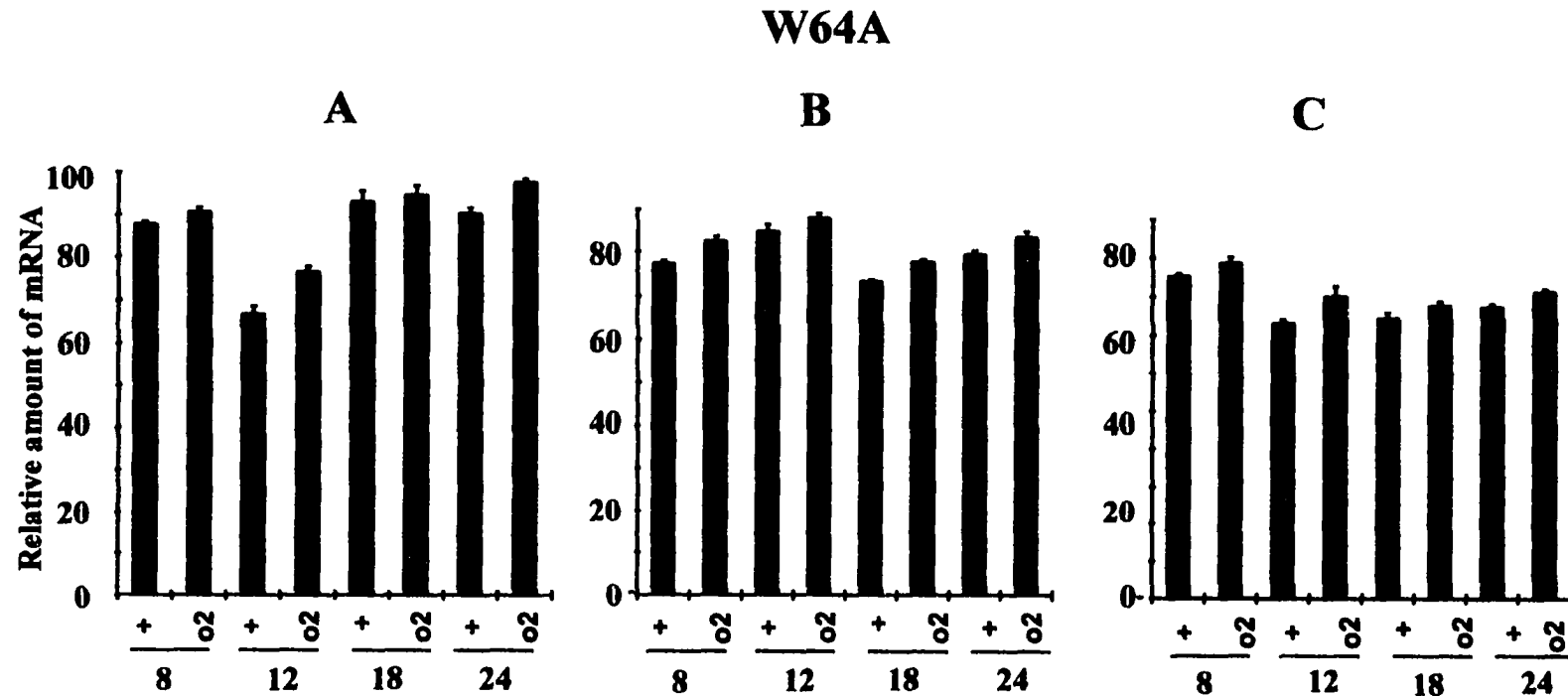
Figure 3.5 - Comparative levels of eEF1A, eEF1Aa and eEF1Ab transcripts by Northern blots of RNA from developing endosperm of W64A+ and W64Ao2. Transcripts were detected using the coding sequence of eEF1A (A), or the 3' non-coding sequence of eEF1Aa (B) and eEF1Ab (C). Each sample contained 5 µg of poly (A) RNA. Bars are the mean of three replicas normalized to the level of ubiquitin conjugating enzyme mRNA, and the error bars show the standard error. The number on the Y-axis represent the percentage of eEF1A transcripts related to ubiquitin conjugating enzyme transcript. The number on the X-axis represent endosperm DAP.

Table 3.4

Comparison of eEF1A (A), eEF1Aa (B) and eEF1Ab (C) transcript levels in W64A+ and W64Ao2 developing endosperm measured by Northern blot. The number on the second and third columns are the average of three measurements from Northern blot showed in figure 3.5. The t-test results are the chance that the difference observed between the means is due to random chance. Treatments are significantly different if the numbers are lower than 5%.

A

DAP	W64A+	W64Ao2	W64Ao2/W64A+	t-test
8	87.5	90.3	1.03	7
12	67.5	76.5	1.14	3.2
18	88	97	1.10	1
24	90.3	97.5	1.08	2.3

B

DAP	W64A+	W64Ao2	W64Ao2/W64A+	t-test
8	77.4	82.3	1.06	2.5
12	85.5	87.7	1.03	21
18	73	75.2	1.06	0.2
24	78.5	85.1	1.04	6.6

C

DAP	W64A+	W64Ao2	W64Ao2/W64A+	t-test
8	85.1	88.3	1.04	16.7
12	73.1	80.1	1.09	6.7
18	74.3	77.1	1.04	9.8
24	76.3	81	1.05	16.4

A B C D E F

eEF1A
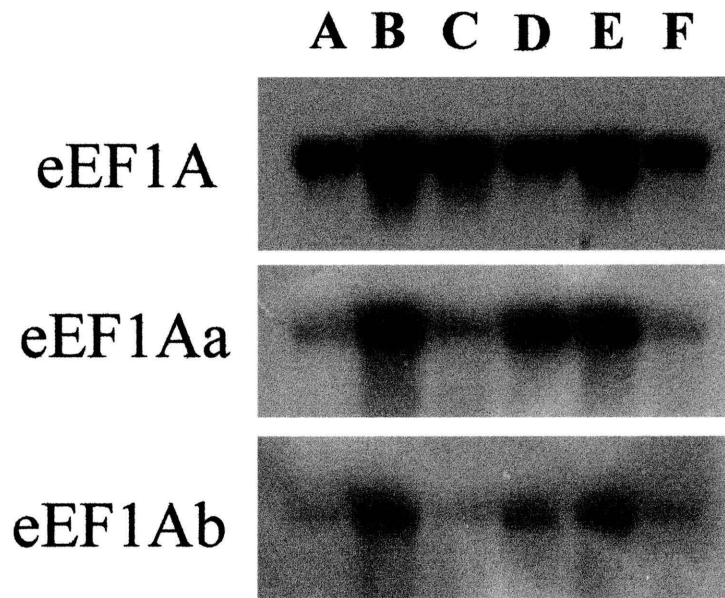
eEF1Aa

eEF1Ab

Figure 3.6 - Detection of eEF1A transcripts in various maize tissues and organs. Samples containing 5 µg of poly(A) RNA were separated by 1.2% agarose gel eletrophoresis and blotted onto nylon membrane. Probes corresponding to the coding sequence of eEF1A 3' non-coding regions of eEF1Aa and eEF1Ab were labeled with dig-dUTP and hybridized to the membranes as described in Material and Methods. A, developing ears (8 cm); B, W64A+ endosperm at 18 DAP; C, mature leaf; D, young root (5 cm); E, silk and F, tassel.

eEF1A gene family in different tissues were calculated as percentage of the total eEF1A (Fig. 3.7). The results showed that the average abundance of the each member of the eEF1A gene family were as follows: eEF1Aa (44%), eEF1Ab (25%), eEF1Ac (9%), eEF1Ad (13%) and eEF1Af (14%) (Fig. 3.7). eEF1Af transcript was higher in roots and silk compared to other tissues. In these tissues, there is also a decrease in eEF1Aa transcript compared to the other members of the eEF1A gene family.

The total eEF1A transcript level in different maize tissues determined by RT-PCR is shown in figure 3.8. Measurements of total eEF1A were based on ubiquitin transcript levels. The level of eEF1A transcript in different tissues was 2- to 8- times higher than the ubiquitin transcript. Statistical analysis using the t-test of the total eEF1A transcript level in different tissues is shown on table 3.5. Means should be considered different if the probability is less than 5%. The greatest variation in eEF1A transcript levels was in leaf and silk. The lowest variation was in stem and silk.

Transcript levels of different eEF1A gene family members and total eEF1A were measured by RT-PCR in W64A+ and W64Ao2 developing endosperm just once, because of the time constraints. Measurement of different eEF1A gene family members and total eEF1A transcripts were calculated as percentage of the total eEF1A and on ubiquitin transcript levels, respectively. According to this analysis, eEF1Aa is the most abundant mRNA in developing endosperm, but there is a change in transcript proportion among the members of the eEF1A gene family in normal and mutant endosperm by 24 DAP (Fig. 3.9A). eEF1Ac and eEF1Ad transcripts in W64A+ at 11 DAP and W64Ao2 at 24 DAP respectively, are not shown. Total eEF1A transcript varied between 4- to 7-times the level of ubiquitin transcripts in developing endosperm (Fig. 3.9B).

To determine if the two most abundant genes are regulated by cold stress, the levels of eEF1Aa and eEF1Ab transcripts were compared by Northern blot analysis in room temperature (RT) and cold treated maize roots and shoots. The results showed that eEF1Aa transcript level is higher in RNA from cold stressed shoots. There is no apparent change in the expression of eEF1Ab in response to cold stress (Fig. 3.10).
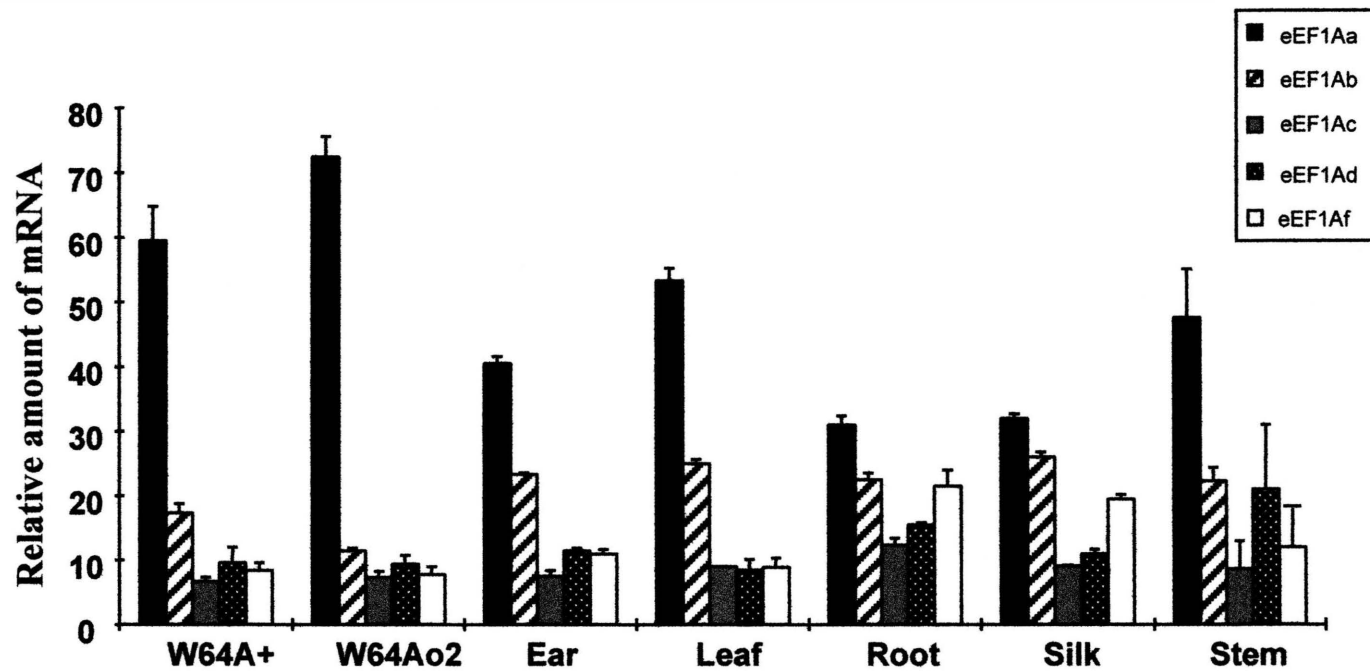
Figure 3.7 - Comparative levels of total eEF1Aa, eEF1Ab, eEF1Ac, eEF1Ad and eEF1Af transcripts based on RT-PCR of RNA from W64A+ and W64A*o2* endosperms at 18DAP, developing ear (12 cm), mature leaf, young root (5 cm), silk and mature stem. Bars are the mean of three replications. The error bars show the standard error. The relative amount of transcripts of each eEF1A gene was determined as the percentage relative to the total eEF1A.
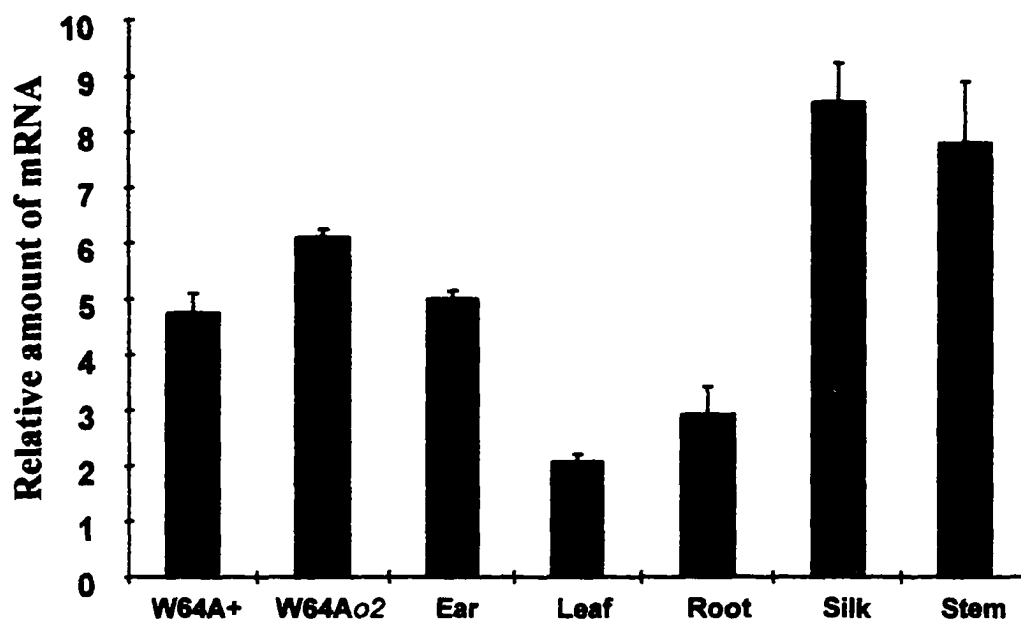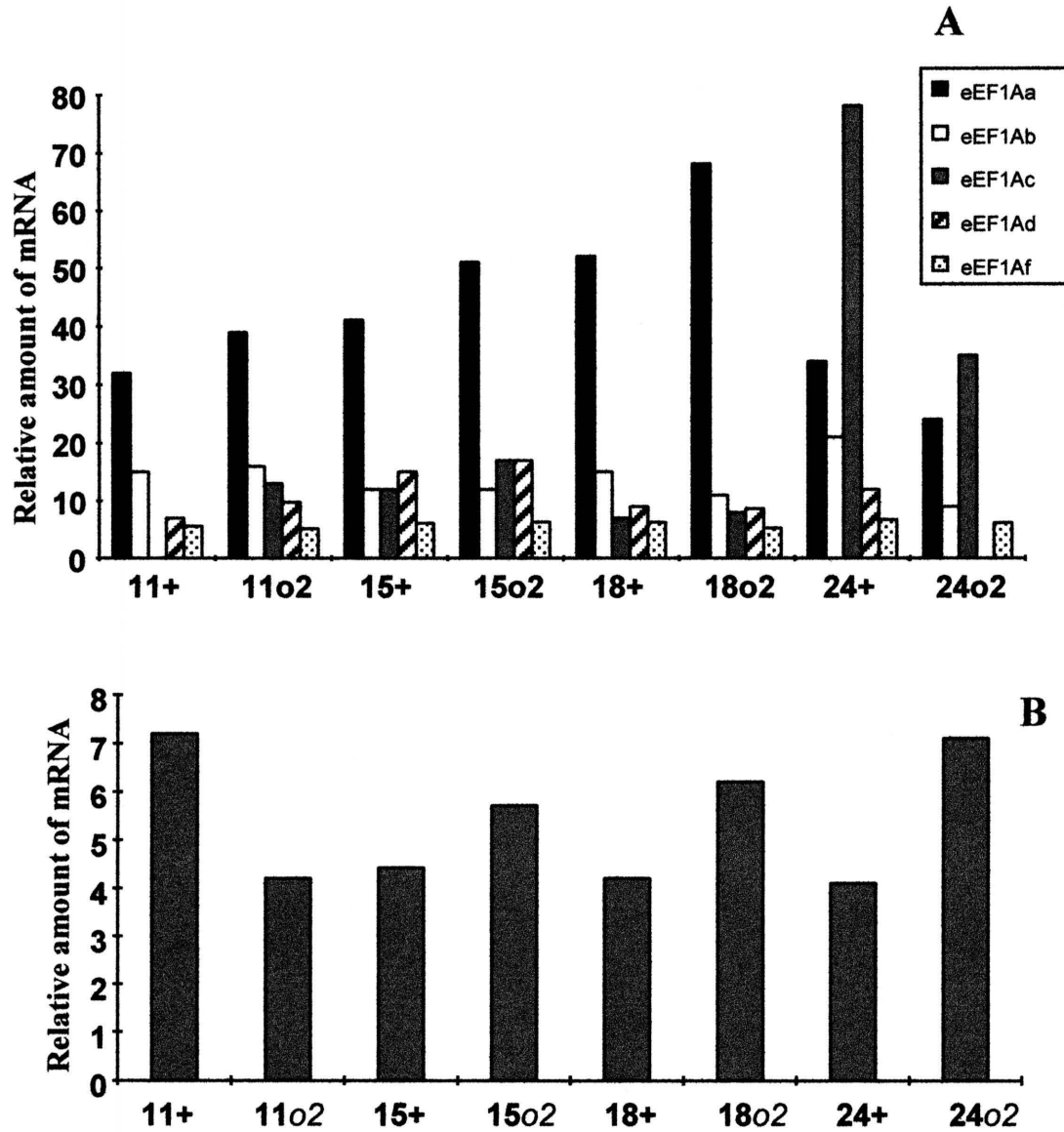
Figure 3.8 - Comparative levels of total eEF1A transcripts by RT-PCR of

RNA from W64A+ and W64Ao2 endosperms at 18DAP, developing ear (12 cm),

mature leaf, young root (5 cm), silk and mature stem. The numbers represent

the comparison of eEF1A to ubiquitin transcript levels. Bars are the mean of

three replications. The error bars show the standard error.

Table 3.5

Statistical analysis of transcript levels of eEF1A in W64A+ and W64Ao2 endosperm at 18 DAP, developing ear (12 cm), mature leaf, young root (5 cm), silk and mature stem by RT-PCR. The t-test result is the probability that the difference observed between the means is due to chance.

	eEF1A	t-test						
		W64A+	W64Ao2	Ear	Leaf	Root	Silk	Stem
W64A+	4.75		7	42	3	24	16	29
W64Ao2	6.1			6	0.5	12.4	23	45
Ear	5				0.9	19.6	15.8	30
Leaf	2					44.5	8.3	15.6
Root	2.9						5	14.2
Silk	8.5							72
Stem	7.8							

Figure 3.9 - Comparative levels of eEF1Aa, eEF1Ab, eEF1Ac, eEF1Ad and eEF1Af (A) and total eEF1A (B) transcripts by RT-PCR in W64A+ and W64A*o2* developing endosperm RNA. These are results from single sample measurement. The relative amount of transcripts for each member of the eEF1A gene family was determined as the percentage relative to the total eEF1A. The relative amount of total eEF1A transcript was determined by comparison to the ubiquitin transcripts.

A B C D

eEF1A
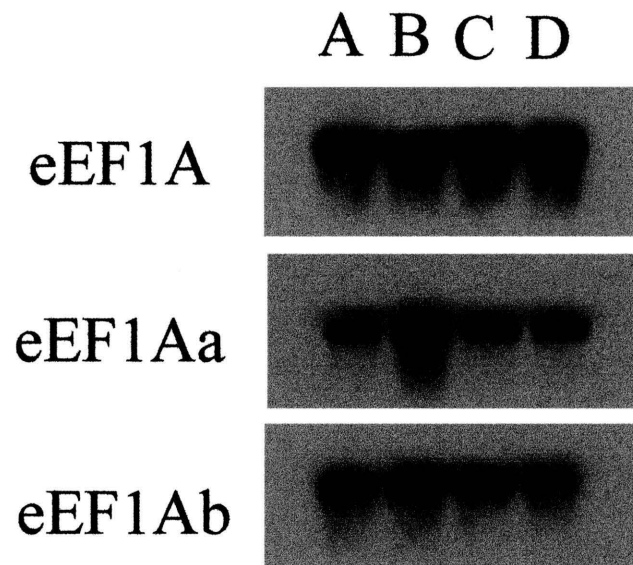
eEF1Aa

eEF1Ab

Figure 3.10 - Comparison of eEF1A transcripts in room temperature (RT) and cold treated maize roots and shoots. Samples containing 5 μg of poly(A) RNA were separated overnight by 1.2% agarose gel eletrophoresis and blotted onto nylon membrane. Probes corresponding to the 3' non-coding regions of eEF1Aa and eEF1Ab, or the coding sequences of eEF1A, were labeled with dig-dUTP and hybridized to the membranes as described in Material and Methods.
A, root temperature (RT) seedlings; B, cold treated seedlings; C, RT roots; D, cold treated roots.

DISCUSSION

Maize eEF1A is encoded by a multigene family

eEF1A, with its central role in protein synthesis, is an abundant protein in both eukaryotes and prokaryotes. It is therefore not surprising that it is often encoded by a multigene family. The number of eEF1A genes has been determined in many organisms. *Escherichia coli* and *Thermus thermophilus* contain two EF-Tu genes (Jaskunas *et al*, 1975, Seidler *et al.*, 1987); in yeast there are two eEF1A genes (Nagata *et al.*, 1984); *Mucor* contains three genes (Linz *et al.*, 1986) and *Artemia* has four genes (van Hermert *et al.*, 1983). In humans, more than eighteen genes were reported (Lund *et al.*, 1996). In tomato, there are at least six distinct eEF1A genes (Pokalsky *et al.*, 1989). The most extensive study of eEF1A in plants was performed in *Arabidopsis thaliana*, where this protein is encoded by four actively transcribed genes. Gene organization does not appear to vary among different varieties of *Arabidopsis*. Three of the four genes (A1, A2 and A3) are contained within a 10 kb fragment and possibly arise from a gene duplication event. The location of A4 is unknown (Axelos *et al.*, 1989).

The eEF1A gene family in maize was characterized in this chapter. The number of eEF1A gene copies in the maize genome was estimated by Southern blotting and genomic reconstruction. Even though a dot blot hybridization of serially diluted genomic DNA can give an estimation of the number of eEF1A genes, this would not allow one to characterize the organization of the eEF1A genes in the genome. There are between 10 and 15 copies of eEF1A in the maize genome, and the intensity of some hybridizing bands may represent clusters of eEF1A genes.

The strategy used here to identify eEF1A expressed genes was similar to that described for the characterization of the maize tubulin gene family (Villemur *et al.*, 1992). Six different α-tubulin genes were identified among 81 α-tubulin cDNA clones in libraries from maize seedling, shoot, endosperm and pollen tissues (Villemur *et al.*,

1992). To identify the expressed eEF1A genes, endosperm and seedling cDNA libraries were screened with eEF1A coding sequence. A total of 144 clones were isolated and sequenced from the 3' end, and they can be divided into 10 groups (Table 3.1). The seedling contains more tissue types than endosperm. Consequently, a seedling cDNA library will be more complex, and so it may contain a greater number of different eEF1A genes than the endosperm. The analysis here showed that there is a similar number of different eEF1A sequences in the endosperm and the seedling cDNA libraries. It is possible that there are eEF1A genes transcribed at very low levels or in specific cell types that could not be identified in the cDNA libraries described in this study.

Two members, eEF1Aa and eEF1Ab, were shown to be the major genes expressed in endosperm and seedling tissues. They represent 53 and 31%, respectively, of the eEF1A clones isolated. The fact that some sequences were found only in the endosperm or seedling libraries such as eEF1Ad and eEF1Ae, does not demonstrate tissue specific expression.

The structure of eEF1A is highly conserved

Comparison of amino acid sequences for eEF1A from several species showed that sequences among plants are highly conserved (>95%) (Browning, 1996). The similarity of the wheat eEF1A amino acid sequence to yeast and human eEF1A is 73 and 77%, respectively (Metz et al., 1992). Scattered differences in the amino acid sequences are observed, and most of the differences are present in more than one species, suggesting that the changes are not necessarily random. Some differences appear to be monocot and dicot specific, but some may be artifactual due to sequencing errors (Browning, 1996)

Sequence analysis was performed for seven of the maize eEF1A genes. The amino acid and nucleotide sequence identity of these genes is shown in Tables 3.2 and 3.3, respectively. The amino acid identity among the eEF1A genes ranged from 91.46 to 99.78%. The highest identity at the amino acid level was between eEF1Aa and eEF1Ab

(99.78%) and the lowest was between eEF1Ac and eEF1Ad (91.46%). The nucleotide identity among the eEF1A genes ranged from 89.53 to 98.65 %. The highest identity at nucleotide level is between eEF1Aa and eEF1Ab (98.65%) and the lowest is between eEF1Ad and eEF1Ae (89.53%). These results demonstrate that all maize eEF1A genes are highly conserved. Differences in the 3' end noncoding sequence are expected in a gene family. The relationship between the differences on the 3' end-sequences and the steady state of each member of the eEF1A gene family mRNA require further investigation.

The seven maize eEF1A genes encode three GTP-binding domains. The first, GXXXXGK, corresponds to residues 13-19, the second, DXXG corresponds to residues 90-93 and the third, NKXD, corresponds to residues 152-155 (Gilman, 1987). The first two elements are involved in the interaction with the phosphate portion of the GTP molecule, and the last element is involved in nucleotide specificity (McCormick et al., 1985).

It has been demonstrated that the removal of the eEF1A region spanning residues 37-68 and 37-128 completely abolishes the transport of aminoacyl tRNA to the ribosome (van Damme et al., 1992). Another report showed that the removal of residues 37-68 in eEF1A reduced the aminoacyl tRNA binding to only 30 to 60% (Slobin et al., 1981). van Damme et al. (1992) demonstrated that the carboxy-terminus of eEF1A, together with the attached polypeptide stretch 1-36, were sufficient to associate with eEF1B, indicating that an intact binding site is present in these fragments. Removal of the last 174 amino acid residues of eEF1A changes the native conformation of the remaining 30 kDa fragment, strongly reducing the affinity toward eEF1B. The regions for aminoacyl tRNA and eEF1B binding domains are highly conserved in all sequenced members of the maize eEF1A gene family. It is assumed that all members of the eEF1A gene family have the aminoacyl tRNA and the eEF1B binding domains.

It has been shown that eEF1A contains a region (amino acid 161 to 179) with 5 to 7 residues important for actin binding (Collings et al., 1994). In Dictyostelium, eEF1A

acts as an actin-bundling protein (Yang *et al.*, 1990). It might also perform this role in carrot (Yang *et al.*, 1993). A domain with homology to the actin-binding domain of starfish depactin (Sutoh and Mabuchi, 1989) exists in *Arabidopsis* (Axelos *et al.*, 1989) and tomato eEF1A (Pokalsky *et al.*, 1989). Based on cross linking studies, five of seven residues implicated in actin binding are conserved (Collings *et al.*, 1994). The eEF1A sequences corresponding to the amino acids 161 to 179 described in this work were compared to the actin binding region in *Dictyostelium* (Fig. 3.11). eEF1Aa, eEF1Ae and eEF1Af contain glutamic acid, while eEF1Ab, eEF1Ac, eEF1Ad and eEF1Ag contains aspartic acid at position 178. This is the only amino acid difference between eEF1Aa and eEF1Ab. eEF1Af is the only member that contain aspartic acid at the position 180. Even though both amino acids are biochemically similar, this result implies that there is a group of eEF1A proteins that may have altered interactions with actin.

Expression of members of the eEF1A gene family in maize

Studies of eEF1A expression in *Mucor racemosus* have shown that there are three genes expressed at different levels, and one gene had morphology-specific expression. In *Artemia*, it was found that an increase in eEF1A activity was observed when dormant cysts began to develop, and this could be correlated with an increase in eEF1A mRNA. In *Drosophila*, the sharp decline in protein synthesis observed with aging could be correlated with decreased levels of translatable polyA+ RNA for eEF1A (Webster and Webster, 1984). Studies in human melanoma cells showed that eEF1A RNA increased on treatment with phorbol ester, a response where transcription and translation of specific genes increases (Opdenakker *et al.*, 1987). In *Xenopus*, three eEF1A genes were characterized: 42Sp50, EF-1α0 and EF-1αS (Dje *et al.*, 1990). The gene 42Sp50 is expressed exclusively in oocytes, while EF-1α0 is active from fertilization until early onset of neurolation. The human eEF1A gene family consists of at least two actively transcribed genes, eEF1A and eEF1A2, and more than 18 homologous loci. eEF1A2 is

		*						*	*		*	*	*					*		
1	P	Q	S	G	T	A	L	D	E	N	V	K	E	E	I	R	A	F	K	M
2	K	Y	S	K	A	R	Y	D	E	I	V	K	E	V	S	S	Y	L	K	K
3	K	Y	S	K	A	R	Y	D	E	I	I	K	E	V	S	S	Y	L	K	K
4	N	Y	S	Q	A	R	Y	D	E	I	V	K	E	V	S	S	F	I	K	K
5	P	W	D	E	S	R	F	Q	E	I	V	K	E	T	S	N	F	I	K	K
6	K	Y	S	K	A	R	Y	E	E	I	V	K	E	V	S	S	Y	L	K	K
7	K	Y	S	K	A	R	Y	E	D	I	V	K	E	V	S	S	Y	L	K	K
8	K	Y	S	K	A	R	Y	E	D	I	V	K	E	V	S	S	Y	L	K	K
9	K	Y	S	K	A	R	Y	E	D	I	V	K	E	V	S	S	Y	L	K	K
10	K	Y	S	K	A	R	Y	E	E	I	V	K	E	V	S	S	Y	L	K	K
11	K	Y	S	K	A	R	Y	E	E	I	V	K	D	V	S	S	Y	L	K	K
12	K	Y	S	K	A	R	Y	E	D	I	V	K	E	V	S	S	Y	L	K	K

Figure 3.11 – Comparison of the actin-binding region of sea urchin depactin (Axelos *et al.*, 1989) to various eukaryotic eEF1A sequences. The presumed actin-binding domain of sea urchin depactin (Sutoh and Mabuchi, 1989) is homologous to conserved regions in various eEF1A sequences. Tomato (2) (Pokalsky *et al.*, 1989), *Arabidopsis* (3) (Axelos *et al.*, 1989), *Dictyostelium* ABP50 (4) (Yang *et al.*, 1990), yeast (5) (Nagata *et al.*, 1984), eEF1Aa (6), eEF1Ab (7), eEF1Ac (8), eEF1Ad (9), eEF1Ae (10), eEF1Af (11) and eEF1Ag (12). Depactin residues identified in actin-binding are marked with an asterisk (*)

expressed only in the terminally differentiated cells of brain, muscle and heart, whereas eEF1A is expressed ubiquitously. Both proteins can function in mRNA translation (Lund *et al.*, 1996).

The pattern of eEF1A expression in tomato has been examined at the transcriptional level by northern hybridization, *in vitro* translation and *in situ* hybridization (Pokalsky *et al.*, 1989). In all tissues examined, there was less eEF1A mRNA in the older, mature tissue than in young, developing tissue. The highest levels of eEF1A mRNA in both root tip and shoot apices were seen in the meristematic or rapidly dividing regions (Pokalsky *et al.*, 1989). Pea root tip cells which are actively dividing contain a higher percentage of polyribosomes than those whose division has been arrested.

Previous experiments comparing the steady state levels of eEF1A mRNAs of W64A+ and W64A*o2* developing kernels (16, 20 and 28 DAP) by RNA dot blot showed that eEF1A transcripts levels were higher in W64A*o2* at all developmental stages tested (Habben *et al.*, 1993). The greatest difference was at 20 DAP, with eEF1A transcripts levels 2.7 times higher in W64A*o2* than W64A+ (Habben *et. al.*, 1993).

Transcript levels of the eEF1A gene family members were examined in this work by three different approaches: frequency of eEF1A genes in the endosperm and a seedling cDNA library, Northern blot and quantitative RT-PCR. The first method was performed with the initial objective to identify as many expressed genes as possible. This method can not estimate precisely the abundance of each member of the gene family, and because it would require isolation of more clones and amplification of libraries, it can give preference to specific sequences. The limitation for the Northern blot is its lack of sensitivity to low levels of mRNA. The Northern blots in this work were performed with DNA dig-labeled probes, and the measurements were done by scanning the autoradiograph and analysis with computer software. These measurements are prone to error, since the autoradiograph has limited linear range. Transcript levels measured by RT-PCR depends on many factors, such as quantity and quality of cDNA template,

presence of RNA in the sample and sequence, specificity and labeling of the primers. To solve these problems a PCR was performed with a serial dilution of the cDNA, primers were checked for specificity by hybridization and by PCR amplification of other eEF1A cDNAs, RNA was removed from the reactions and the number of cycle were kept low (15 cycles). Due to time constraints, some of the RT-PCRs were not repeated. The RNA used for the Northern blot was the same as used for the RT-PCR, except the developing endosperm that was collected at two different seasons. The transcript levels were normalized with the ubiquitin-conjugating enzyme in the Northern blots and with ubiquitin in the RT-PCR reactions. It is assumed that ubiquitin-conjugating enzyme and ubiquitin are constitutively expressed and the transcript levels are similar in all tissues and developmental stages.

Northern blot and the RT-PCR analysis were performed with the total eEF1A, eEF1Aa and eEF1Ab at different developing endosperms and different tissues. In the Northern blot, the intensity of the signal was measured using densitometric analysis of the signal generated on the X-ray film. The experiments with developing endosperm were performed three times, and the results are the average of the comparison with the ubiquitin-conjugating enzyme transcripts. Statistical analysis (t-test at 5% probability) of the signal pattern of the total eEF1A obtained by the Northern blots indicates that its level is similar throughout the endosperm development (Table 3.4). The gradual drop in eEF1A and ubiquitin conjugating enzyme transcript levels detected during maize endosperm development might be due to the accentuated increase of storage protein synthesis related mRNAs.

The analysis of the eEF1Aa and eEF1Ab transcripts by Northern blot demonstrated that eEF1Aa is significantly increased in W64Ao2 endosperm at 8 and 18 DAP compared to W64A+. No significant difference between o2 and + genotypes was seen in eEF1Ab transcript level in any stage of developing endosperm. However, the greatest difference between eEF1Ab transcript occurred at 12 DAP, when the difference between genotypes was smaller for eEF1Aa. These results suggest that there may be a

change in expression of different members of the eEF1A gene family throughout endosperm development to maintain a given amount of the total eEF1A transcripts. eEF1Aa and eEF1Ab transcript were the only mRNAs detected by Northern blot analysis in the other tissues (Fig. 3.6). The hypothesis that other eEF1A genes were not detected by Northern blot analysis because they are expressed at low level or they are expressed in specific cell type, might not be correct. The isolation of cDNA clones in the endosperm and seedling cDNA libraries shows that the frequency of the eEF1Ac is 10% in the endosperm cDNA library (Table 3.1), and the RT-PCR results show that eEF1Af transcript levels are similar to eEF1Ab (Fig. 3.7).

The RT-PCR is a more sensitive technique than Northern blots, and it was possible to quantify the abundance of other members of the eEF1A gene family relative to the total eEF1A levels. The results obtained by RT-PCR demonstrated that eEF1Aa was the predominant transcript, followed by eEF1Ab, eEF1Af, eEF1Ad and eEF1Ac in all tissues examined. eEF1Af showed higher transcript levels in roots and silk. High levels of eEF1Af in these tissues suggest it may be regulated at the transcriptional levels at elongating tissues such as young roots and silk. A shift in the transcript levels of different members of the eEF1A gene family in different tissues was also observed. In roots and silk, there is a drop in the proportion of the eEF1Aa transcripts relative to the others. A shift in transcription of different members of the eEF1A gene family was also seen in developing endosperm by RT-PCR. These results suggest that eEF1A genes are differently regulated in different tissues and in developing endosperm.

The results show that the two most abundant genes are eEF1Aa and eEF1Ab in most tissues. Quantification of the transcript levels by Northern blots and quantitative RT-PCR demonstrated that there is an increase in eEF1A transcript between 1.04 to 1.28 times in W64Ao2 compared to W64A+ during endosperm development. Similar results were found with eEF1A protein levels at 10, 15 and 20 DAP (Sun *et al.*, 1997). However there is an increase in eEF1A protein levels in W64Ao2 at 25 DAP by more than two-fold. In mature seed the eEF1A protein levels are increased 1.9 times in W64Ao2

compared to W64A+ (Sun *et al.*, 1997). RT-PCR only demonstrated that the eEF1Af transcript level is higher in endosperm and silk, compared to other tissues. Transcript levels of the different members of the eEF1A gene family were not verified after 24 DAP developing endosperm.

eEF1A gene expression is affected by stress

Stress related eEF1A genes have been isolated in barley, potato and maize. A barley eEF1A gene was isolated from shoot meristem cDNA library prepared from RNA from plants were exposed to low temperature (Dunn *et al.*, 1993). In potato there is a stable association of eEF1A and tuber polysomes at the onset of hypoxia, coincident with a sharp rise in lactate and decrease in tissue pH. This aberrant association of eEF1A with polysomes also occurred when aerobic tuber extracts were acidified *in vitro* (Vayda *et al.*, 1995). eEF1A transcription is induced by hormone application (Ursin *et al.*, 1991), wounding (Morelli *et al.*, 1994) and cold stress (Dunn *et al.*, 1993). Conversely, the large decrease in translational activity that occurs upon starvation, fungal infection or senescence has been attributed to depletion of elongation factors, or correlated with decreased expression of eEF1A (Cavallius *et al.*, 1986, Jurss *et al.*, 1992, Mahe *et al.*, 1992, Ursin *et al.*, 1991). Berberich *et al.* (1995) demonstrated that maize eEF1A transcript is increased in leaves at low temperature, whereas time-course experiments over 24 h at 5°C showed that in roots the overall mRNA level of eEF1A transiently decreased. These results suggested that the expression of eEF1A is differentially regulated in leaves and roots under cold stress. These observations led us to question which member of the eEF1A gene family is involved in cold stress and in which tissue its expression is up regulated.

To answer this question, Northern blot analysis was performed with RNA from leaf and roots of seedling under cold stress. The results showed that eEF1Aa transcript was higher than eEF1Ab in seedlings exposed to cold temperatures (4°C - 6 h). This

indicates that eEF1Aa and eEF1Ab genes are differently regulated under stress. Both transcripts occur at the same levels under normal conditions, but under stress the level of eEF1Aa transcript increases in seedlings. The difference in expression of different members of the eEF1A gene family opens the possibility of isolating the promoter of eEF1Aa and searching for cold responsive elements in the promoter regions or using this promoter to induce other genes under cold stress.

CONCLUSION

In this study, eEF1A gene family characterization was performed to help understand more about the functions of this protein. We reported the isolation of cDNA clones representing 10 different maize eEF1A genes. The results of genomic Southern blot analysis suggest that the eEF1A sequence clones represent nearly all the eEF1A genes in maize

Transcript level and sequence of different members of the maize eEF1A were analyzed. All the eEF1A maize genes examined had GTP, aminoacyl tRNA, eEF1B and actin binding domains. eEF1Aa, eEF1Ae and eEF1Af contain a substitution of glutamic acid to aspartic acid in a region that has been shown to be important for actin binding. This suggests that there might be two groups of eEF1A genes that bind differently to actin. Analysis of transcript levels demonstrated that different members of the eEF1A gene family are expressed at different levels in the tissues examined. The eEF1A clones fall into two classes: two abundantly expressed genes, eEF1Aa and eEF1Ab, with 99.79% amino acid sequence identity and five genes expressed at low level, with amino acid identity that ranged between 91.46 to 99.10 %. Even though the presence of conserved domains in the eEF1A sequences suggests that the seven isoforms of maize eEF1A genes perform similar functions, the difference in their gene expression supports the conclusion that they may differ in some other aspects (e.g. elongation rate, specificity for different tRNA species or specificity for the different types of ribosomes that exist in different cell lines and tissues). Highly expressed genes may function as the protein-synthesis elongation factor in tissues, where other eEF1A genes are expressed at a low level and vice-versa.

As noted before, transcript levels of the eEF1A genes did not appear to be significantly increased in W64Ao2 compared to W64A+ in some endosperm development stages. Transcript levels for eEF1Aa were higher than those for eEF1Ab in seedlings under cold stress. The expression of eEF1Af is increased in roots and silks,

compared to other tissues and the other members of the eEF1A gene family. Regulation of the steady state mRNA levels of different members of the maize eEF1A gene family still needs to be determined.

CHAPTER 4

IDENTIFICATION OF PROTEINS IN MAIZE ENDOSPERM THAT INTERACT
WITH eEF1A

INTRODUCTION

The protein eEF1A is known to bind aminoacyl-tRNAs to the acceptor (A) site of ribosomes during peptide chain elongation (Hershey, 1991). Other functions have been attributed to eEF1A, such as activation of phosphatidylinositol 4-kinase (PI4-kinase) in carrot suspension cell cultures (Yang et al., 1993), bundling of actin (Yang et al., 1993), bundling and severing microtubules (Shiina et al.,1994) and ubiquitin-dependent proteolysis of N-acetylated proteins (Gonen et al., 1994). The biological significance of many of these interactions is not fully understood. However, the evidence suggests that eEF1A has many important functions in the cell, in addition to protein synthesis.

The level of eEF1A was shown to correlate with the content of lysine in maize endosperm (Habben et al., 1995). The basis of the relationship between eEF1A and the lysine-containing proteins is not known, but eEF1A was found to account for only 1 to 2% of the protein in maize endosperm (Sun et al., 1997). Therefore, eEF1A would account for only 2 to 3% of the total protein-bound lysine. This implies that there must be a relationship between eEF1A and the other major lysine-rich proteins in maize endosperm.

It is possible that the relationship between eEF1A and the lysine-rich protein in maize endosperm arises from the fact that eEF1A is a component of the cytoskeleton. Clore et al. (1995) used immunocytochemistry to investigate the subcellular localization of eEF1A in maize endosperm. They found that in older interphase cells actively synthesizing starch and storage protein, protein bodies are enmeshed in eEF1A and actin and are found juxtaposed with a multidirectional array of microtubules. Treatment with

cytochalasin D resulted in redistribution of the eEF1A that previously was colocalized with actin. A possible explanation for the high correlation between eEF1A concentration and the lysine content is a physical or structural association between eEF1A and the other lysine-rich proteins.

Since eEF1A is an actin bundling protein it is possible that eEF1A physically interacts with other components of the cytoskeleton. One way to identify proteins that interact *in vivo* is with the yeast two-hybrid system (Chien *et al.*, 1991). This technique is based on the ability of the GAL4 transcription factor to be divided into two separate but functional polypeptides: a DNA-binding domain and an activation domain.

In the yeast two-hybrid system, two types of fusion proteins are constructed: one generates a hybrid between sequences for the DNA-binding domain of the yeast transcription factor Gal4 (amino acids 1-147, Keegan *et al.* 1986) and the protein of interest (eEF1A in this case). A second expression plasmid contains sequences for the Gal4 activation domain (amino acids 768-881, Ma and Ptashne, 1987), fused to a cDNA library from the tissue of interest. As first demonstrated with the Gal4-Gal 80 interaction (Ma and Ptashne, 1988) and later generalized by Fields and Song (1989), if the two proteins expressed in yeast are able to interact, the resulting complex will regain the ability to activate transcription from promoters containing Gal 4-binding sites, upstream activating sequence from GAL1 (UASG).

The yeast two-hybrid system provides a versatile and powerful technique that is applicable to many proteins once their genes have been isolated. Protein-protein interactions are detected by reconstituting a functional transcriptional activator in yeast. This method not only allows identification of proteins that interact, but also can be used to define and/or test the domain/residues necessary for the interaction of two proteins. In the last few years, a large number of genes have been characterized using this method, including many cell cycle regulators. This approach has contributed significantly to our understanding of the eukaryotic cell cycle (Chien *et al.*, 1991; Vojtek *et al.*, 1993; Durfee *et al*, 1993).

To understand more about the biology of eEF1A, its relation to the cytoskeleton and the level of lysine in maize endosperm, eEF1A and actin cDNA clones were used to identify interacting proteins via the yeast two-hybrid system. The utility of this *in vivo* approach to screen an endosperm cDNA library for an interacting proteins and define and/or tested the domains/residues necessary for the interactions of two proteins was demonstrated. The test case was maize eEF1A and actin. Yeast was transformed with plasmids encoding the GAL4 DNA-binding domain, fused to eEF1A or actin and tested to determine whether they interact with themselves. The same constructs in yeast was transformed with a library of plasmids containing maize endosperm cDNAs fused to the GAL4 activation domain. By screening for transcription of GAL4- and HIS-dependent reporter genes, many clones were identified which interact with eEF1A and actin maize cDNAs. The amino acid sequences where the clones interacted with eEF1A were mapped. Finally, the hypothesis that clones found with eEF1A screen could interact with actin, and vice-versa was tested.

MATERIALS AND METHODS

Yeast and Bacteria strains and plasmids

The DNA-binding domain vector, pAS2, and the GAL4 activation domain vector, pACT2, are illustrated in figure 4.1. eEF1Aa and eEF1Ab cDNAs were cloned into the *NcoI-BamHI* sites of the plasmids pAS2 (pAS2-eEF1Aa and pAS2-eEF1Ab) and pACT2 (pACT2-eEF1Aa and pACT2-eEF1Ab). The cDNAs corresponding to clones eEF1Aa and eEF1Ab were described in the Chapter 3. Actin was cloned into the *BamHI* site of pAS2 (pAS2-actin) and pACT2 (pACT2-actin). The PCR actin fragment was generated by the primers *BamHI* (GGGATCCAAATGGCTGATGGTGAG) and T3 (AATTAAC CCTCACTAAAGGG). The template for the PCR reaction was the cDNA clone 5C02F11 (Chapter 2). The product of PCR reaction was digested with *BamHI* and the appropriate fragment isolated from the gel using Gene Clean according to the manufacture's instructions (BIO101). The purified DNA was cloned into the *BamHI* site of pAS2 and pACT2. The appropriate 5' to 3' direction of the actin sequence was checked by *EcoRI* digestion. All clones were sequenced to confirm that the coding sequences were in the correct reading frame.

The positive control plasmids were pSE1111 (SNF4 fused to the activation domain of GAL4 in pACT) and pSE1112 (SNF1 fused to the DNA-binding domain of GAL4 in pAS1) (Fields and Song, 1989). The pAS1 plasmids used to test the pACT2-cDNA plasmids for false positives were pAS1-CDK2 (Fields and Song, 1989), pAS1-SNF1 (Fields and Song, 1989), pAS1-p53 (Fields and Song, 1989) and pAS1-lamin (Fields and Song, 1989).

The yeast strains used for the two hybrid system were Y190 (MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112, +URA3::GAL -->LacZ, LYS2::GAL(UAS) -->HIS3 cyhr) and Y187 (MAT gal4 gal80 his3trp1-901 ade2-101 ura3-52 leu2-3, -112 met- URA3:: DAL -->lacZ).
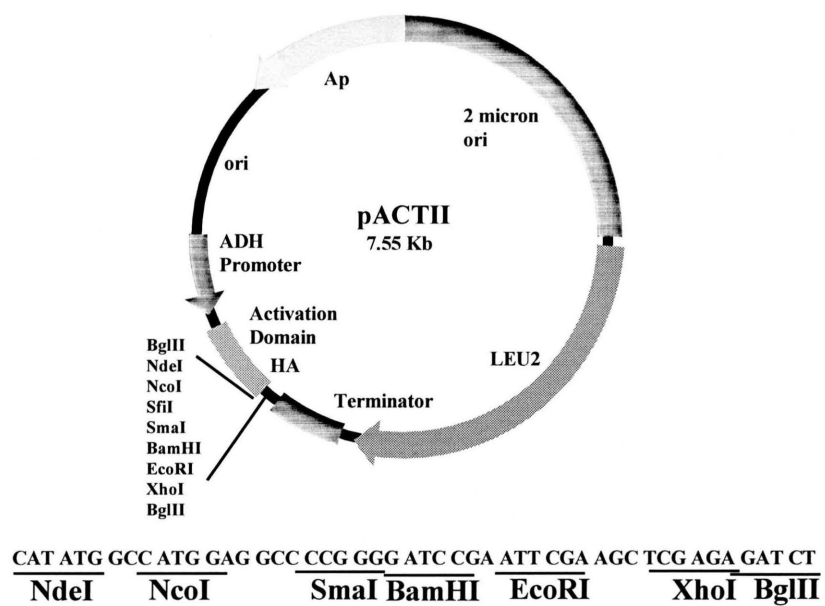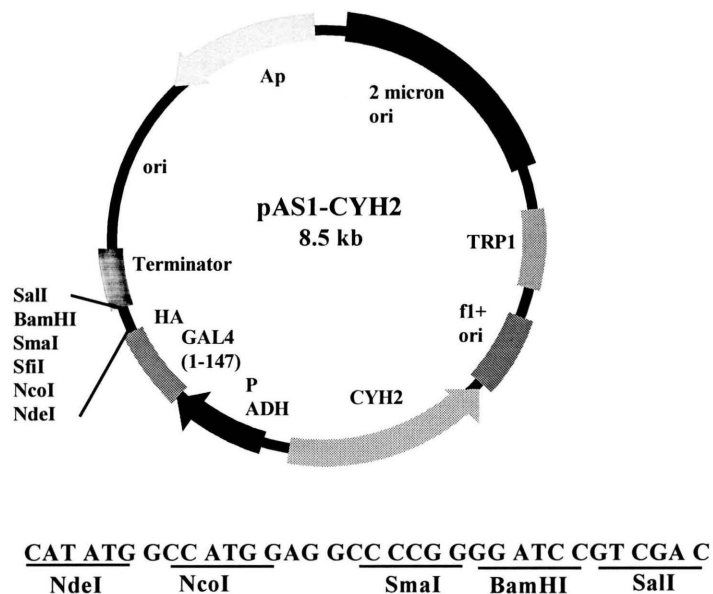
Figure 4.1 - Map of the DNA-binding domain vector pAS2 and the activation domain vector pACT2. The reading frames of the polylinkers are shown in coding triplets (Bai and Elledge, 1996).

The *Escherichia coli* strain BNN:132 was used to convert the λpACT2 cDNA library to the pACT plasmid cDNA library *in vivo*. BNN132 (JM107 λKC lysogen) *end*A1 *gyr*96 *hsd*R17 *rel*A1 *sup*E44 *thi* Δ(*lac-pro*AB)[F' *tra*D36 *pro*AB⁺ *lacI*ᑫZ M15] λKC (*Kan-cre*). The *E. coli* strain used to amplify the plasmids was DH5α *sup*E44 Δ*lac*U169 (ø80*lac*ZM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96*thi*-1 *rel*A1.

Construction of the cDNA library in the activation domain plasmid

An endosperm cDNA library was constructed using poly(A) RNA from free polysomes of maize endosperm at 14 to 26 DAP. The poly(A) RNA was isolated using oligo(dT) cellulose, according to manufacture's instructions (Promega - Madison, WI).

The cDNA libraries were constructed with the vector λpACT2 by directional cloning (5'-*EcoRI* - *XhoI* - 3'), according to the manufacture's instructions (Strategene - La Jolla, CA).

The cDNA library was amplified and transferred to plasmids using the bacteria strain BNN:132. BNN:132 colonies containing pACT2 plasmids were selected in 50 μg/ml ampicillin LB plates. The plasmid DNA from these colonies was CsCl purified (Sambrook *et al.*, 1989) and used for yeast transformation.

The conversion from phage to plasmid did reduce the number of recombinant plasmids. The colonies were grown in LB plates containing ampicillin at 50 μg/ml and bulked for plasmid preparation. The pACT2 plasmids containing the cDNA library was used to transform yeast Y190 cells containing pAS2-eEF1A and pAS2-actin.

Yeast transformation

A colony of the yeast strain Y190 was grown in YEDP media (1% yeast extract, 2% peptone and 2% dextrose) to A600 = 0.5 to 0.8. Cells were harvested by centrifugation at 2,500 x g for 5 min., washed with sterile distilled water, repelleted at the

same speed, and then resuspended in 50 ml of Lisorb buffer (100 mM LiOAC, 10 mM Tris-HCl [pH 8], 1 mM EDTA, 1 M Sorbitol). After incubation at 30°C for 30 min., the cells were harvested by centrifugation at 2,500 x g for 5 min., resuspended in 650 µl of Lisorb and then kept on ice.

For the preparation of the carrier DNA, 200 µl of 20 mg/ml sheared salmon sperm DNA was boiled for 7-10 min., cooled at room temperature and then 800 µl Lisorb (RT) was added with 40 µg the pAS2-eEF1A or pAS2-actin plasmid. One hundred µl of the mixture containing the pAS2-eEF1A or pAS2-actin plasmid was added to the same volume of yeast cells. Nine hundred µl of 40% PEG 3350 in 100 mM LiAc/TE (100 mM LIOAC, 10 mM Tris, pH 8, 1 mM EDTA) was added to 100 µl of cells containing DNA. The cells were incubated at 30°C for 30 min., heat shocked at 42°C for 7 min and then plated on SC-Trp media. SC: Synthetic medium was used for selection of yeast cells containing plasmids with specific nutrient markers. SC lacking a particular amino acid is referred to as dropout medium (Table 4.1). A dropout mixture is made by combining and grinding into a fine powder with a mortar and pestle the components mentioned in Table 4.1, leaving out those that will be selected in the media. For example SC -Trp media contains 0.13% dropout mixture (minus tryptophan), 2% dextrose, 0.5% $(NH_4)_2 SO_4$ and 2% agar. The media was sterilized by autoclaving. One hundred ml of 10X YNB was added to the medium just prior to pouring plates. 10X YNB contains 6.7 % yeast nitrogen base without amino acids. This stock solution is filter sterilized and stored in the dark.

Cells containing the pAS2-eEF1A or pAS2-actin plasmids were transformed with the maize endosperm cDNA library in pACT2 using the protocol described above, and then selected in SC -His, -Trp, -Leu +25 mM AT (3-aminotriazole) media.

The molecular weight of the fusion proteins synthesized in yeast was verified by Western blots, using the commercially available monoclonal antibody, mAb 12CA5 (Babco, Richmond, CA) against hemagglutinin (HA) epitope. The protein from yeast was extracted by boiling two ml of yeast cells (A600 = 0.8) for 10 min in 1X Laemmli Buffer (0.18 M Tris-HCl pH 6.8, 7% SDS, 33% glycerol, 15% β- mercaptoethanol, 0.3%

Table 4.1

Proportion of amino acids (mg) in the dropout media

Component	Weight in mg	Component	Amount mg
Adenine	800	Arginine	800
Aspartic acid	4000	Histidine	800
Leucine	2400	Lysine	1200
Methionine	800	Phenylalanine	2000
Threonine	8000	Tryptophan	800
Tyrosine	1200	Uracil	800

bromophenol blue).

X-Gal assay

The yeast colonies growing in SC -His, -Trp, -Leu +25 mM AT were transferred to new media, and after overnight growth, a nitrocellulose membrane was placed on plates containing the colonies and allowed to wet completely. The membrane was lifted carefully to avoid smearing the colonies and placed in liquid nitrogen for 10 sec to permeabilize the cells. The membrane was removed from nitrogen liquid, allowed to thaw, and placed colony side up in a petri dish with Whatman (Clifton, NJ) 3MM chromatography paper soaked with autoclaved Z-buffer (1.61 % Na_2HPO_4 $7H_2O$, 0.55 % NaH_2PO4 H_2O, 0.075% KCl, 0.025% $MgSO_4$ $7H_2O$, [pH 7.0]; 0.27% 2-mercaptoethanol and 1% X-Gal were added just before use). X-GAL: 5-Bromo-4-chloro-3-indolyl-β-D-galactoside was dissolved in DMF (N,N-dimethylformamide) at 100 mg/ml as a stock. Yeast colonies that were positive in the X-GAL assay were grown in SC -Leu liquid media to eliminate the pAS2-eEF1A plasmid, and the pACT2 plasmid containing the cDNA was isolated.

Small scale DNA isolation from yeast

Two ml of Y190 cells containing the pACT2 plasmids grown in SC-Leu (A600 = 0.8) were pelleted by centrifugation. The cells were suspended in 200 μl of yeast lysis solution (300 mM NaCl, 10 mM Tris, pH 8, 1 mM EDTA, pH 8 and 0.1% SDS w/v) with some glass beads, 200 μl of PCI (49% Phenol [v/v]; 49% CHCl3 [v/v], 2% Isoamyl alcohol [v/v]) and vortexed for 1 min. The phenol was saturated with 10 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM EDTA (pH 8). After centrifugation, the aqueous phase was extracted once with PCI and once with $CHCl_3$. The DNA was precipitated with 2 vol. of ethanol and incubated at -20°C for 1 hr. The tubes were then centrifuged at 14,000 x g at

room temperature, and the DNA pellet was washed with 1 ml of 70% ethanol. The DNA was dissolved in 10 μl of TE. The pACT2 plasmids containing the cDNAs were transformed into *E. coli* DH5α using the CaCl$_2$ protocol (Sambrook *et al.*, 1989). After DNA sequencing, they were retransformed individually into yeast Y187 containing the p53, lamin, CDK2 and SNF1 genes in pAS1 to test for false positives.

Nucleotide sequence

The plasmids containing putative eEF1A-interacting proteins were amplified and then isolated from *E. coli* DH5α using an alkaline extraction procedure (Sambrook *et al.*, 1989). The cDNA inserts were characterized by dideoxy manual sequencing of the double-stranded DNA using ^{35}S-dATP, as per the manufacturer's instructions (USB). Sequencing was from the presumed 5' end of the mRNA. The primer used to initiate sequencing was 5' GGCTTACCCATACGATGTTCC 3'. After the DNA sequence was obtained, it was compared to GenBank entries to identify the encoded protein.

Deletion on the eEF1Aa sequence

Deletion mutants were made to map the regions where the eEF1A interacting protein binds to eEF1A. Two deletions of eEF1Aa were obtained by PCR amplification with eEF1A pAS2 as template, using the primers NcoI 930 (5'-GGCCCATGGACCACG AGGCAC-3') and 262 (5'-CAGTTGAAGTGAACTTGCGGG-3') and the primers 530 (5'-GGCCCATGGATGCAACCACTCCC-3') and 262 (5'-CAGTTGAAGTGAACTTG CGGG-3'). The 1.0 and 0.6 kb PCR fragments were isolated from 1.2% agarose gel using Gene Clean according to the manufacture's instructions (BIO101). The DNA product was cloned into the *NcoI- BamHI* site of the pAS2 and pACT2 plasmids. The pAS2 plasmids containing the eEF1A 1.0 and 0.6 kb deletions were transformed into Y190 as described above; they were selected on SC -Trp media and transformed with the

plasmids containing cDNAs encoding the putative eEF1A-interacting proteins. The full length pAS2-eEF1A was used as positive control, and pAS2-SNF1 was used in as a negative control. Protein interactions were tested by the X-gal assay and by monitoring growth in SC -Trp, -Leu plates and in SC -Trp, -Leu, -His+3-AT plates.

RESULTS

The principle of using the yeast two-hybrid system to detect protein-protein interactions is shown in figure 4.2. A gene fusion is generated that encodes the protein under study (eEF1A or actin) as a hybrid with the DNA-binding domain of GAL4. To verify that a complete eEF1A protein was produced, the size of the fusion protein was examined using the eEF1A and HA antibodies (Fig. 4.3). The fusion protein had an apparent molecular weight that was larger than the eEF1A protein itself (50 kD), since it contains the GAL4 DNA binding domain (4 kD). The actin fusion protein was tested with only HA antibodies (result not shown).

A variety of activities (binding, bundling and severing) have been reported for the association of eEF1A with components of the cytoskeleton (Condeelis, 1995). To test if eEF1Aa or eEF1Ab interact with themselves and actin in the yeast two-hybrid system, pAS2-eEF1Aa, pAS2-eEF1Ab and pAS2-actin were co-transformed in yeast with pACT2-eEF1Aa, pACT2-eEF1Ab and pACT2-actin (Table 4.2). Only the combination pAS2-actin and pACT2-actin gave a positive result. The interactions of eEF1Aa and eEF1Ab with themselves and eEF1A with actin (in both directions) did not result in β-galactosidase activity.

To screen for eEF1A- and actin-interacting proteins, Y190 cells expressing eEF1A and actin cDNAs fused to the Gal 4 DNA-binding domain were transformed with an activation domain-tagged endosperm cDNA library. The initial titer of the endosperm cDNA library in λpACT2 was 6 X 10^7 pfu/μg of cDNA. PCR analysis of 50 random phage revealed that around 85% contained recombinant cDNAs, and that the inserts ranged from 0.5 to 1 kb in length. The λpACT2 cDNA library was converted to pACT2 plasmids using the bacteria BNN:132 and used to transform yeast cells containing pAS2-eEF1Aa plasmid. Around 200 colonies were obtained after 10 days under selection of -His, -Leu, -Trp plus 25 mM 3-AT selection. Lower levels of 3-AT (25 mM) was used to detect the weak interaction.

DNA binding domain hybrid

X

GAL4 (1-147)

USAG   **GAL1-lacZ**

Activation domain hybrid encoded by a library

Y   GAL4 (768-881)

GAL4 (1-147)

USAG   **GAL1-lacZ**

Interation between DNA-binding domain hybrid and a hybrid from the library

X **Y**

GAL4 (1-147)

USAG   **GAL1-lacZ**
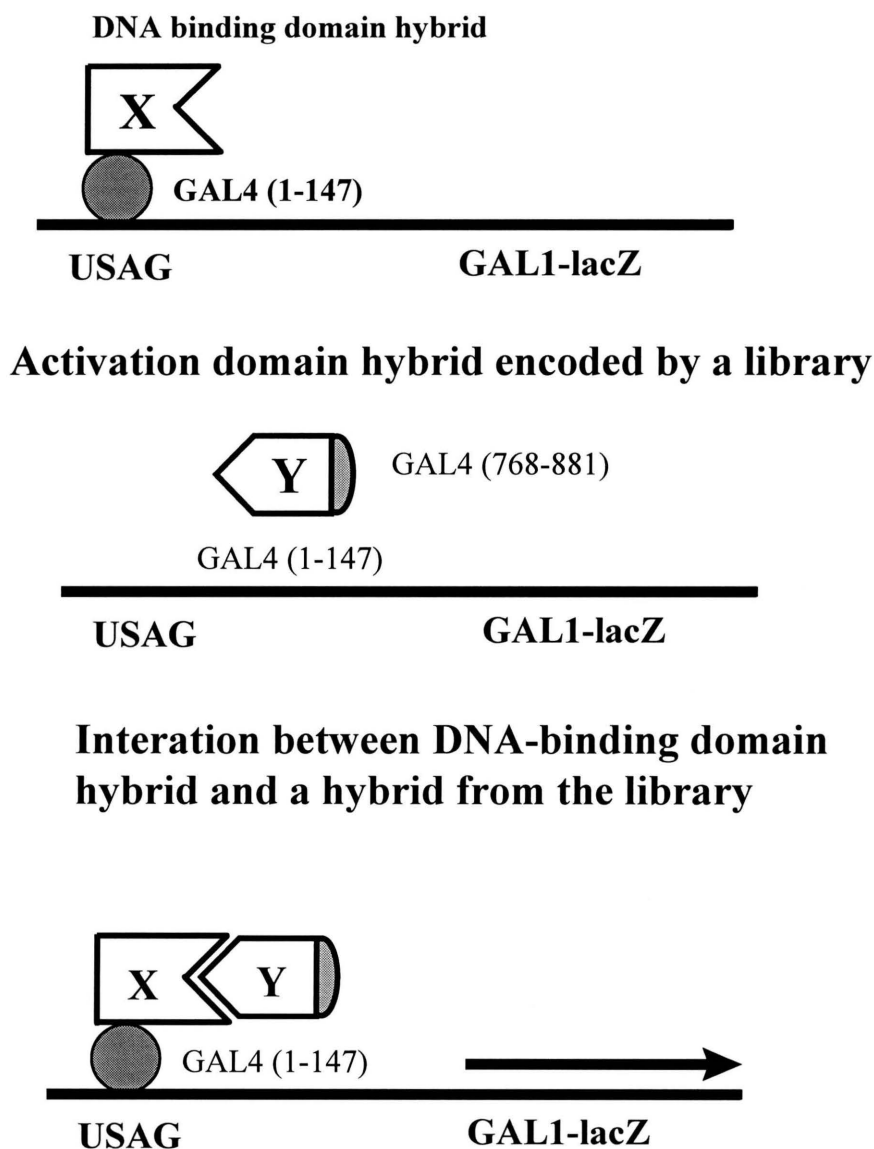
Figure 4.2 - Strategy to detect interacting proteins using the two hybrid system. USAG is the upstream activation sequence for the yeast GAL gene, which binds the GAL4 protein. The library of activation domain hybrid are constructed in the pGAD vectors (Chien *et al.*, 1991).
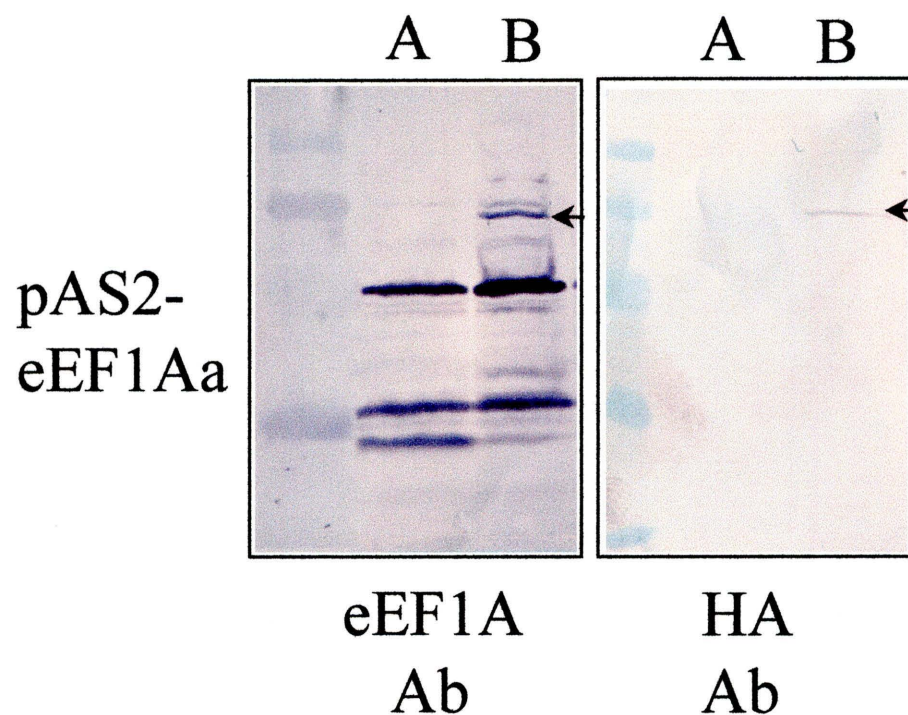
Figure 4.3 - Immunodetection of maize eEF1A in protein from Y190 (A) and Y190 transformed with pAS2-eEF1Aa (B). Proteins were extracted from yeast cells as described in Materials and Methods. The antibodies for eEF1A detection were made against maize protein (Habben *et al.*, 1995). A commercially available monoclonal antibody, mAb 12CA5 (Babco, Richmond, CA), was used for hemagglutinin (HA) epitope. The arrows show the position of the maize eEF1A polypeptide.

Table 4.2

Interaction of eEF1Aa, eEF1Ab and actin in the yeast two-hybrid system. Positive interaction (+); negative interaction (-).

	pAS2- eEF1Aa	pAS2-eEF1Ab	pAS2-actin
pACT2-eEF1Aa	-	-	-
pACT2-eEF1Ab	-	-	-
pACT2-actin	-	-	+

Cells containing plasmids encoding proteins that potentially interact with eEF1A and actin proteins were subsequently screened for β-galactosidase (β-gal) activity. This secondary screen eliminated His+ revertants and plasmids bearing the His gene of the organism from which the cDNA library was derived. Colonies that are His+ and blue were considered positives and were isolated for further analysis.

Thirteen cDNA clones were identified that encode eEF1A-interacting proteins. Based on difference in growth on SC -His, -Leu, -Trp plus 25 mM 3-AT plates (B) and the intensity of the signal in the X-gal assay (A), the nature of interaction with eEF1A might be stronger for some of clones than others (Fig. 4.4). The cDNAs encoding these proteins ranged in length between 0.2 to 1.5 kb. The 5' end sequences of some eEF1A-interacting cDNA clones is shown in the Table 4.3.

At present, there is only one clone (number 5) that matches with protein data bases. This clone, which corresponds to RNA pol II, was isolated twice (the two isolates are not separately represented in Figs. 4.4 and 4.5). Considering that one of the cDNAs encoding RNA pol II is shorter by about 300 bp, it appears that the N-terminus of RNA pol II is not important for the interaction with eEF1Aa. Using this screen, no pACT2 plasmid was recovered without a cDNA insert.

To investigate how the eEF1Aa-interacting proteins react with eEF1Ab, the clones identified with the pAS2-eEF1Aa construct were tested for interaction with pAS2-eEF1Ab (Fig. 4.4). Even though eEF1Aa and eEF1Ab differ by only one amino acid (Chapter 3), some of the clones that interacted with eEF1Aa reacted differently with eEF1Ab, suggesting that the two proteins may have different physiological roles.

To demonstrate that the eEF1Aa-interacting clones were specific for eEF1Aa, they were also tested with CDC2, p53, lamin and SNF1. The lack of interaction by CDK2, p53, lamin and SNF1 demonstrated that the eEF1A interactions are specific to the pAS2-eEF1Aa clone (Fig. 4.4).

Table 4.3

Sequence of the 5' end of some eEF1A interacting proteins

Clone	Sequence
2	AATTCGGCACGAGATCGGAATCGCCATCGTCCGCACAAACACC TTTGGGTTTGTTTATGGACCGAGAGATGATGATCGAGCAACTG ATGTGGTGCAGGACAAGGTGCAGTCGGTGGAACGACACAAGG GAAGGGATGGAACTTAACCAGTGGATGCTTCGCCAAGCAGTAC CCACCCATCAACACTATCTATTGGCATTAAATTAAA
3	GAACTAGTCTCGAGTTTTACATTGTTGAAGGGATGTTATTGATG TCGTGGAGTCAAACCTACTACTAAGCGTCATGGCGATTAGTTTT TGAGACAAATTGAGTCGAGGAATGATCGTGGACTGGATAAGTC AAGCTGCTGACTGGTGCAATATGAGAGA
4	TACAAGCAGTGCAGTACAACCGCCGCTTCGTCACCGCCGTCGT CGGATTCGGCAACAAACGCGCCCAACTCCTCGGAGAAGTAGAT ATAGCATGTCTTGTGTTTGGTCCCACTAGTTCCTTGTACTGATG GCAGTGCTGAGACTTATATTGCTATATATGGTTGATTCAGTGTG TTACTGCACACTATGAATT
5	GCAAGCAAACCACCGGAGCAGCGGCGCTCGCCGCGGGCGGAG CATGGCGGACGACGATTACAACGAAATCGACATGGGCTATGAG GATGAGCCTCCAGAACCTGAGATTGAGGAAGGGGCCGAAGAA GAGCTTGAGAACAA
14	ACACGCCTCCCCGCGGCGGCCTGCAGCTCCAGCGCGGGAGGAG GACGAGGGCTGCGGGACGTGTTGGCGAGGGAAGGACGAGTTTC GGAGA

Figure 4.4 - Comparison of the intensity of eEF1Aa and eEF1Ab interaction with endosperm cDNAs identified with the yeast two-hybrid screen. The numbers on the top indicate the pACT2-Y clone that encodes a putative eEF1A interacting protein. The names on the left identify the full length clones eEF1Aa, eEF1Ab, CDK2, p53, lamin and SNF1 in pAS2-X. Yeast cells containing the combination of pAS2-X and pACT2-Y plasmids were plated on SC -Leu, -Trp (A) SC -His, -Leu, -Trp plus 25 mM 3-AT (B). The X-gal assay (C) is a replica from plate A.
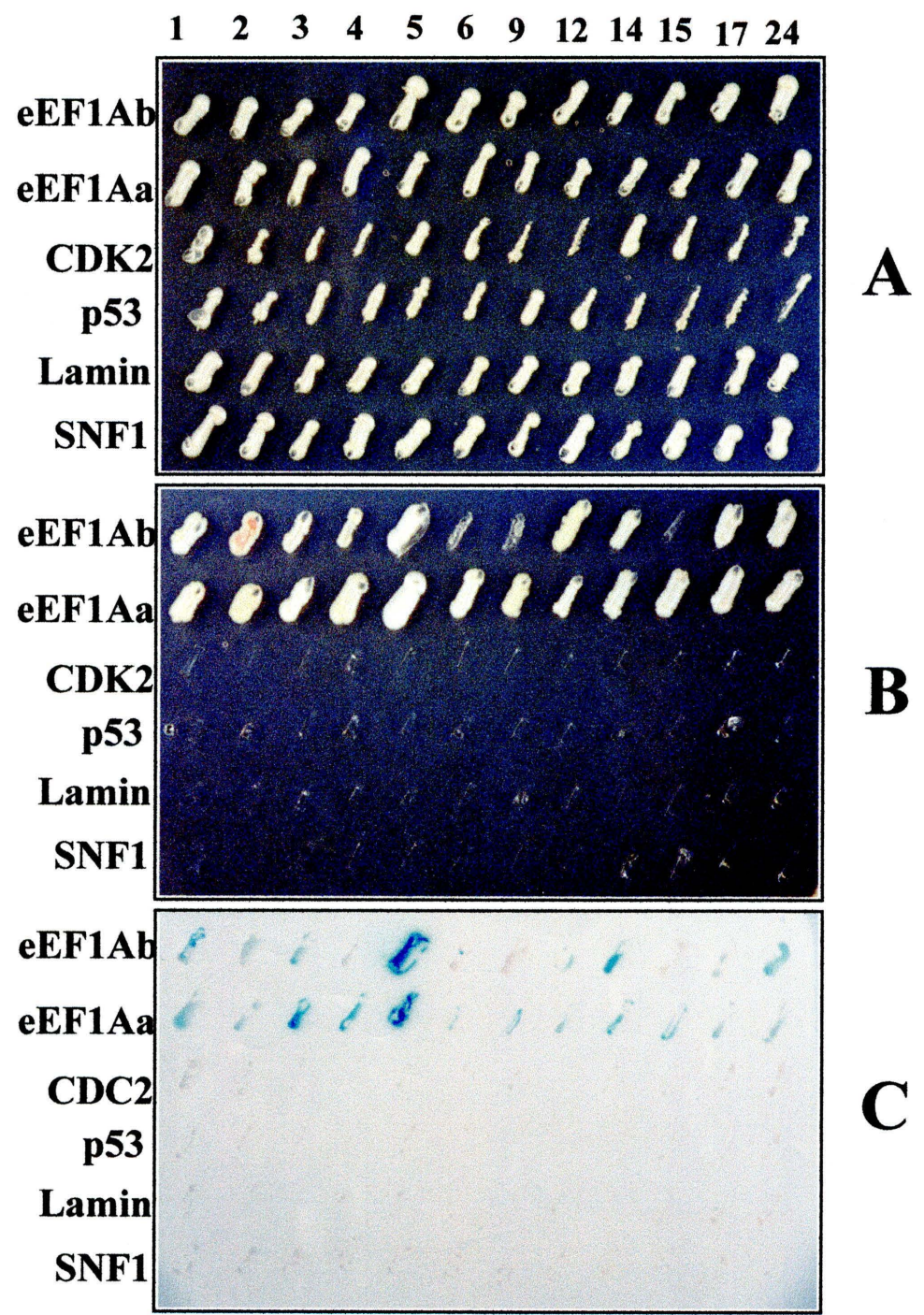
Figure 4.4

To determine if the eEF1Aa-interacting proteins depend on the N- or C-terminus of eEF1Aa, the clones that encode the eEF1A-interacting proteins were tested with two deletions of eEF1Aa mutants. One eEF1A mutant represented a 500 bp deletion from the 5' end and the other represented a 900 bp deletion from the 5' end.

The eEF1A-interacting proteins were found to bind with different efficiency to different regions of eEF1Aa. The clones numbered 1, 3, 5, 6, 12 and 17 appeared to interact with the N-terminus, clones numbered 2, 4 and 14 appeared to interact with the middle region, while clones numbered 9, 15 and 24 seemed to interact with the C-terminus of eEF1Aa (Fig. 4.5). For these experiments, pAS2-SNF1 was used as negative control. pAS2 alone can weakly activate lacZ due to residual transcriptional activation activity of amino acids 1-147 of the Gal4 protein; it is not therefore a good negative control.

Fifteen clones were identified that encode actin-interacting proteins. The 5' sequences of some of the actin-interacting cDNA clones are shown in the Table 4.4. Only four of these sequences matched with the data base: 1 = sucrose synthase; 2 = protein kinase; 3 = acidic ribosomal protein and 4 = ubiquitin conjugating enzyme. A pACT2 plasmid without a cDNA insert was found twice in the actin screen.

The clones identified in the actin yeast two-hybrid screen were also tested with eEF1Aa and eEF1Ab. None of the clones identified with eEF1Aa interacted with actin, but one of the clones that interacted with actin showed a positive interaction with eEF1A, RNA pol II (clone number 5). The interaction of this clone with p54, lamin, CDK2 and SNF1 was negative, suggesting that this interaction is specific for eEF1Aa and actin and that the encoded protein might function as a bridge, thus explaining their simultaneous interaction.
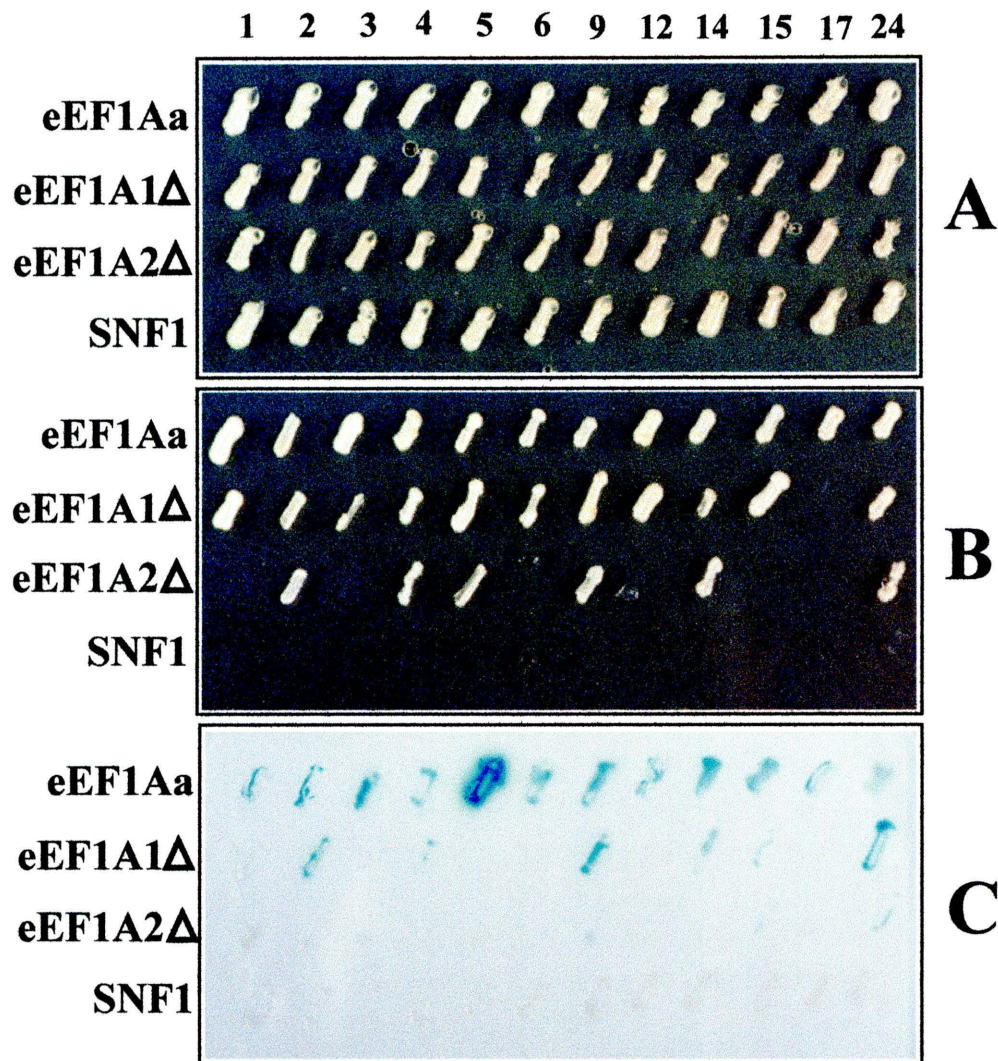
Figure 4.5 - Characterization of the eEF1A binding sites of cDNAs identified with the yeast two-hybrid system. The numbers on the top correspond to pACT2-Y clones that encode the eEF1A-interacting proteins. The names on the left indicate the full length eEF1A sequences, the deletions of eEF1A1 and eEF1A2, and SNF1, all in pAS2-X. The yeast containing the combination of pAS2-X and pACT2-Y plasmids was plated on SC -Leu, -Trp (A); SC -His, -Leu, -Trp plus 25 mM 3-AT (B). The X-gal assay (C) is a replica from plate A.

Table 4.4

Sequence of the 5' end of some actin interacting proteins

Clone	Sequence
1	CCCTGCTGAGATCATCGTGGACCGGGTATGTGGCCAGCACATTG ACCCTTTAACCCACAGCGACAAGGCCGCGGATATCCTGGTCCAA CTTGTTGACAATGCAAGGCAGATCCGAGCTACTGGACAGACTCA CAGGCGGCCTGCAGAGAATTATGAGAAGTACAGTGGAAC
2	ATAGAGATTTGAAACTAGAAAACACTCTCCTGGATGATAGTATA GCACCTCGGCTAAAGATATGTGATTTTGGTTACTCCAATTCCTCT GTGCTGCACTCTCAACCAAATCAACTGTGGACACCAGCTTACAT TGCTCCTGAGGTCCTTCTCACAAGGAATATGATGGAAGGCGGCG AT
3	AGAGCAGCTCAAGGATGTGCGAAGAACCCCGTGGATGGGTTCT CCGCCGGGCTTGTAGACGACAGCAACATCTTCGAGTGGCAGGTC ACCATCATCGGACCGCCTCACACCCTATATGATGGAGGCTACTT CAATGCAATAATGACCTTCCCCACAACTATTCCCAACAGCCCGC ATCAGTAAGATTTACTTCTGAGATGTGGGCATCGAATGTTATCC TGATGGACGTGTTGCATTCTATTCTTCATCCACCTGTCAAGATCC CAATGGTATGAGCTTGCA
4	GGTTTAGCATGTCGACCGGTTCCTCGGAGCACGATGAAGTTTGT TGCTGCCTACCTGCTTCCTGTCCTCGCTGGGAACGCCAGCCCCC CTCCGCTGACGATTTGACTGCCATTCTGGAGTCAGTTGGCTGTG AAGTTGACAATGAAAAGATGGAACTCTGCTGTCCCAACTGAGC GTAAGGACATCACCGAGCTCATTGCCGCGGCAGGAGAGTTGCTT CAGTCCT

Table 4.4 (cont.)

Clone	Sequence
5	CACGAGAGAACACCGAGCAAGACCTCCAGCTGGCCGTCGAGGA GGCAGCGACATGCCTGGACGCGGAGAAGCTGGCGATCGTCGAT TTCACCAGCTTCGTGGAGAAGTCAAGCGACCCGGCCGCTACGTC ATCTTCTGGGAGCTGAGCTCGACGCGAGCGAGGACGTCCTTGCG AAGGCTTGCGCGAACTGCC
6	ACGAGCACGAACGCCCACACGTACCTGTCCAGCGAGCCCAAGT CCCTCCGCACCTGAAACACGGAGAACGAACAACGCCATCAGAT GATTTTTTATATATATCTTCTTCTCCATTCATTGACCGCCTCCTTC GTCTTCGTTGCAGGATGGTACTGGAACTACTACTACTGCACTAT GTATGTTCAACGACTGGGTGAAGCTCACGTACGGGGTACCTCGA TATCGCAGGCGTGTGACGTCCGGGACGTGCCAATGAGCCGTAGT GG

DISCUSSION

The yeast strain, Y190, was constructed to provide a dual selection system to more efficiently screen cDNA expression libraries for clones encoding proteins that interact with a protein of interest. The Y190 strain carries two chromosomally located genes which expression regulated by Gal4. First, the *Escherichia coli lacZ* gene, is under control of the GAL1 promoter, and its usefulness in this system has been described (Fields and Song, 1989). A second reporter, the selectable HIS3 gene, was chosen because very low levels of the enzyme (imidazole glycerol phosphate [IGP] dehydratase) are required for prototrophy. To provide Gal4 control, His 3 regulatory sequences were replaced by an upstream activating sequence from GAL1 (USAG). Because Gal4 is deleted in Y190 (and its negative regulator gal80), there should be no expression of either reporter in the absence of exogenous galactose. However, the GAL1-HIS3 fusion has residual HIS3 expression sufficient to allow growth without exogenous histidine, even in the absence of Gal4. Growth inhibition can be overcome by growing cells in the presence of 25 mM 3-aminotriazole (3-AT), a chemical inhibitor of IGP dehydratase, which restores histidine auxotrophy (Kishore and Shah, 1988). The low requirement for His3 protein makes this selection very sensitive, such that proteins with only weak interactions can be detected.

The second advantage of the combination HIS3 selection/lacZ screen involves the elimination of false positives. Durfee *et al.* (1993) observed that a class of false positives appears in these library screens that depends on a nonspecific interaction with the DNA-binding domain. These positives are often transcription factors which when overproduced are thought to access the promoter DNA adjacent to the target protein. Because the HIS3 and GAL1 reporter promoters, share only a small region of common DNA sequence (150 bp that should mostly be protected by the binding of the target protein fusion), this class of false positives should be significantly diminished.

The pAS2 plasmid was constructed to facilitate the creation of a target protein fusion with the Gal4 DNA-binding domain. This plasmid contains TRP1, the 2 μ origin of replication, and the ADC1 promoter driving expression of the Gal4 DNA-binding domain (amino acid 1-147, Keegan *et al.*, 1986) fused to a polylinker. The Gal4 derivative is tagged with the hemagglutinin (HA) epitope, which is recognized by a commercially available monoclonal antibody, mAb 12CA5 (Babco, Richmond, CA). The polylinker contains several useful cloning sites, including *NcoI* and *NdeI*.

Some proteins that are not transcription factors can activate transcription when fused to the GAL4 DNA-binding domain. This activity severely interferes with the results of the yeast two-hybrid system. Prior to beginning a screen, eEF1Aa, eEF1Ab and actin were tested for lacZ activation and growth properties on SC-His plates containing varying concentrations of 3-AT (Sigma).

The use of a genetic selection in yeast to detect interacting protein allows much larger cDNA libraries to be analyzed. To facilitate construction of the cDNA libraries, a phage vector, λpACT2 (activation domain), was created that fuses sequences for the Gal4 activation domain to the cDNAs. λpACT2 is similar to YES yeast, the yeast-*E. coli* shuttle, and it has several advantages over conventional methods for the production of large plasmid libraries. This phage makes it possible to generate a large cDNA library with a high percentage of long inserts (capacity of 8.5 kb). It can replicate and express the fusion protein in yeast, and it can be converted from a phage to a plasmid using *cre-lox*-mediated site-specific recombination. The ability to convert phage to plasmids with high efficiency facilitates recovery of the inserts and the introduction of libraries into yeast. pACT2, the plasmid excised from λpACT2, contains the ColE1 origin of replication and *bla* gene for replication and selection in *E. coli*, and LEU2, 2 μ origin, and the ADC1 (*Adh1*) promoter sequences for selection, replication and expression in *Saccharomyces cerevisiae*. The ADC promoter drives expression of a hybrid protein consisting of the SV40 large T antigen nuclear localization signal and sequences

encoding the activation domain of Gal4. Fused to GAL4 at amino acid 881 is a polylinker containing a *XhoI* site into which the cDNAs are inserted.

As positive controls, pSE1111 (SNF4 fused to the activation domain in pACT) and pSE1112 (SNF1 fused to the DNA-binding domain of GAL4 in pAS1) were used for the X-Gal and 3-AT resistance assays. It was noted from previous work that pAS2 alone can activate lacZ weakly, and it is therefore not a good negative control, due to residual transcriptional activation activity of amino acids 1-147 of the Gal4 protein.

From the thirteen clones found to interact with eEF1A, only one had a match with the protein data base (RNA pol II). eEF1Bα and eEF1Aβ subunits bound to the eEF1Bγ-subunit, creating two binding sites for eEF1A (Browning, 1997). eEF1B bound to eEF1A as a complex which might explain why eEF1B was not found in the two-hybrid system screen with eEF1A.

RNA pol II is a nucleoprotein that participates in transcription of mRNAs. Its specificity to bind promoters is determined by the combination of different transcription factors. It is not obvious why RNA pol II would interact with eEF1A. One possibility is that eEF1A shares biochemical properties with some transcription factors that bind to RNA pol II. Even though these proteins are known to participate in independent processes, there are suggestions that eEF1A might participate in the translocation of some proteins to specific locations in the cell.

Plant eEF1A has been described as a phosphatidyl-4-kinase activator that binds actin and enhances actin bundling (Yang *et al.*, 1993), a calcium/calmodulin-sensitive microtubule bundling factor (Durso and Cyr, 1994) having a role in the organization of perinuclear microtubules during the transition from M to G1 phase of the cell cycle (Hasezawa and Nagata, 1993, Kumagai *et al.*, 1995), a microtubule-severing protein at low molar ratios that aacts upon polymerized tubulin and a microtubule stiffening and bundling protein when at higher ratios (Littlepage *et al.*, 1995). However, the precise molecular mechanisms of these interactions are unknown.

The reason eEF1A did not interact with actin in the two-hybrid system might be due to a requirement for a filamentous conformation of actin (F-actin). The two-hybrid system detects interactions of two monomeric proteins in the nucleus. Also, eEF1A and actin are encoded by multigene families and this interaction might occur between specific members of the gene family.

One of the clones identified in the yeast two-hybrid screen with actin was found to also interact with eEF1Aa and eEF1Ab, but not with p53, lamin, CDK2 and SNF1. This result might indicate that the interaction between actin and eEF1A occurs through a second protein.

Plant polysomes are reported to interact with components of the cytoskeleton and membranes (Davies *et al.*, 1991), and recently an acidic ribosomal protein (p40) from Arabidopsis has been shown to associate with polysomes and the cytoskeleton (Garcia-Hernandez *et al.*, 1994). An acidic ribosomal protein identified in the two-hybrid screen with actin reinforces the association of actin with components of the protein synthetic machinery. There is no obvious explanation for why no actin clones were found in the yeast two-hybrid screen with actin, since the reaction was positive when the protein was paired in the control reactions.

CONCLUSIONS

The discovery that the concentration of eEF1A serves as an index for protein-bound lysine content in the endosperm could have a profound implication for creating maize genotypes with increased nutritional quality. It is therefore very important to understand the reason why the level of eEF1A is correlated with the content of lysine in the endosperm. One approach that might give some clues about which eEF1A processes is involved and why eEF1A correlates with lysine content is by the identification of proteins that bind to eEF1A. To that end, I decided to use the yeast two-hybrid system to assay interactions between eEF1A and other endosperm proteins. A number of clones were identified that gave positive interactions in the yeast two-hybrid system. However, their encoding sequences did not lead to the identification of any known proteins. Additional experiments are required to further characterize the functional relevance and structure of the protein/nucleotide these sequences.

CHAPTER 5

SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

Endosperm development in maize seed involves multiplication, enlargement, and differentiation of cells with subsequent accumulation of starch and storage products. This tissue has been studied from a variety points of view: evolution, genetics, physiological and biochemical perspectives, and more mutants have been analyzed and mapped in endosperm than any other maize tissue. However, we still have a superficial understanding of endosperm development. Advances in molecular biology have created the possibility for a detailed study of many genetic and molecular mechanisms involved in maize endosperm development. The results of this research may help to answer basic questions about the physiology of maize endosperm and enhance the practical uses of the seed.

The goal of many maize research programs involves the isolation of genes with subsequent determination of their involvement in various aspects of plant development and crop productivity. Gene isolation and identification have often proven to be rate-limiting steps in this process. One of the ways to identify genes and determine their putative function is by random DNA sequencing analysis. This dissertation reports the analysis of a number of randomly selected cDNAs from two maize libraries. The strategy for identifing putative gene function was through a combination of gene mapping, 'single-pass' sequencing (SPS) and measurement of transcript levels. Many other questions concerning genome structure and organization, such as gene clustering and duplication were also addressed through this work.

The development of genetic markers and maps for many species has provided plant geneticists and breeders with powerful new tools for genetic analysis and plant breeding. Map location might provide some indication of gene function for a random cDNA. It seems likely that SPS could provide at least as much insight regarding

function, if not more. It also appears likely that the combination of the two approaches might be synergistic in facilitating identification of gene function. For instance, sequencing a cDNA and finding that it possesses homology with protein kinases is certainly instructive, but determining that it maps to the same location as a dominant disease resistance locus is suggestive of the role of the gene. Once genes have been identified, they can then be expressed or used in *in vitro* assays to determine if they are similar in function. Transcript levels of candidate clones can be compared in Northern blots containing normal and mutant RNA, and further analyses can be performed to determine whether changes in expression of these genes are directly or indirectly related to the mutation.

A maize endosperm cDNA library was extensively examined in this project. Clones from an endosperm library were first probed with cDNAs prepared from two different tissues (endosperm and seedling) and categorized as abundant/rare and general/tissue-specific. Individual members of each gene family were also sequenced from the putative 5' end of the mRNA to develop 400 bp of DNA information, which was compared to the Genbank database. Clones were categorized as either having no significant homology, partial homology (indicative of a putative function), or high homology (indicative of a homologous function). Clones were hybridized to genomic blots of segregating progeny from the Brookhaven recombinant inbred lines in order to determine both gene copy number and location(s) within the maize genome. The latter was compared with mapped mutants for initial correlation of genes with defective functions.

In the maize genome project described in this dissertation, a total of 576 cDNA clones from endosperm and 490 clones from seedlings were evaluated for their transcript levels. A total of 313 clones from the two libraries was submitted to define SPS and more than 300 clones were mapped using a standard maize genetic population. The results demonstrated that a combination of analyses can provide important information that might eventually relate to the function of these genes. This project has been

extended at many other institutions and has generated tremendous amounts of data used to help other researches to identify and characterize gene functions. For instance, a clone with homology to the maize *waxy* gene, which maps very close to the *su1* mutation on chromosome 4, was identified. This is an interesting mutant which not only affects the level of sugar in the kernel, but also the amount of unbranched starch (Coe *et al.*, 1988). This clone was sent to L. C. Hannah (Gainsville, FL) to test in a starch mutant population. Similarly, many protein kinase clones were checked to determine if their map location was near disease resistance genes. A homolog for an acyl carrier protein was identified and mapped to a location near 10c, a chromosomal region identified as a major QTL for oil quality in maize. A maize homolog to an *Arabidopsis* male sterility gene was identified and sent to M. Alberton at Pioneer (Hibred Inc.), who is studying this phenomenon in maize. An *Agamous* homolog from the endosperm library was mapped and sent to R. Schmidt at UC San Diego, who has been studying related maize sequences with homology to *Arabidopsis* clones (Helentjaris, personal communication).

Aspects of research concerning genome organization were also studied in this project. The 19 kD and 22 kD zeins are located in a large inverted array on chromosome 4s. Two widely separated groups of 19 kD genes are intercalated by two groups of 22 kD genes. These results support the hypothesis that zein genes might have come from a early duplication event.

In summary, this project was not simply the collection and cataloging of data. Many opportunities were created from this research to improve the knowledge of gene expression, organization and function in higher plants.

Many gene families were identified in the genome project, including heat shock proteins, zeins, actin, tubulin and eEF1A, etc. The eEF1A gene family was chosen for further characterization, because of its multifunctional nature and a lack of understanding of its biological role. Beside binding aminoacyl-tRNAs to the acceptor (A) site of ribosomes during peptide chain elongation, eEF1A appears to be involved in many other cellular processes. eEF1A has been found to interact with microtubules (Durso and Cyr,

1994), actin (Yang *et al.*, 1993), the valyl-tRNA synthase complex (Motorin *et al.*, 1988) and calmodulin (Kaur and Ruben, 1994). It has also been shown to be involved in the activation of the phosphatidylinositol 4-kinase (Yang *et al.*, 1993) and the ubiquitin-dependent proteolysis of N-acetylated proteins (Gonen *et al.*, 1994). Another aspect that made the eEF1A gene family very interesting for further characterization is its relationship with the protein-bound lysine content of maize endosperm (Habben *et al.*, 1995) and its colocalization to the cytoskeletal network surrounding protein bodies (Clore *et al.*, 1996). Plant breeders have spent an enormous amount of time and effort trying to improve the nutritional quality of protein in maize endosperm. Most of the proteins present in the endosperm are zeins, which contain low amounts of essential amino acids, such as lysine. Selection for higher levels of eEF1A will help breeding programs generate maize genotypes with better protein quality, but little is known about the physiological basis involved in the correlation between eEF1A and lysine content.

eEF1A genes have been isolated from many organisms including *Saccharomyces cerevisiae* (Schirmaier and Philippsen, 1984, Cottrelle *et al.*, 1985), *Mucor racemosus* (Hiatt *et al.*, 1982, Fonzi *et al.*, 1985. Linz and Sypherd, 1987), *Artemia salina* (van Hemert *et al.*, 1984; Lenstra *et al.*, 1986) and *Xenopus laevis* (Viel *et al.*, 1991). In plants, eEF1A has been isolated from *Arabidopsis* (Axelos *et al.*, 1989), barley (Dunn *et al.*, 1993), potato (Vayda *et al.*, 1995), soybean (Aguilar *et al.*, 1991) and maize (Berberich *et al.*, 1995). The complete sequence of four eEF1A genes, A1, A2, A3 and A4, in *Arabidopsis* has been described (Liboz *et al.*, 1990). Based upon transient expression and S1 nuclease mapping experiments, it was found that all eEF1A genes in *Arabidopsis* are actively transcribed. In maize, very little was known about the number of eEF1A genes and their expression in different tissues. The results presented here show that eEF1A in maize is encoded by a multigene family composed of 10 to 15 genes.

Around 144 eEF1A clones were isolated from endosperm and seedling cDNA libraries and were divided into 10 groups according to their 3' non-translated sequences. The expectation was that most of the genes would be represented in the two cDNA

libraries. There is the possibility that some genes are expressed in other tissues, such as tassel and ear, that were not represented in either library. Expression of some eEF1A gene family members might be restricted to specific tissue types, thus making them seem to be rare.

These maize eEF1A genes were tested for attributes which relate to their functions. One attribute is the transcript level of different eEF1A gene family members in various tissues and developing endosperm. Differentially expressed eEF1A genes have been found in *Mucor racemosus* (Linz and Sypherd, 1987), *Drosophila melanogaster* (Hovemann *et al.*, 1988), rats (Lee *et al.*, 1993) and humans (Knudsen *et al.*, 1993). The eEF1A transcript levels were measured in the present work by Northern blots and quantitative RT-PCR analysis. The results demonstrated that the eEF1A gene family members showed variation in expression at different tissues. eEF1Aa and eEF1Ab transcripts account for approximately 80% of the total eEF1A mRNA in all tissues examined. Transcript levels of the other eEF1A genes could not be detected by Northern blot analyses in the same tissues. Under some circumstances, lower levels of some eEF1A mRNAs might reflect different functions. In support of this hypothesis, the expression of eEF1Af was higher in young roots and silk and low in other tissues. eEF1Af transcript levels did not increase in rapidly dividing tissues, such as 9 DAP endosperm. Thus, eEF1Af is potentially differentially regulated in elongating tissues, and could be related to a cytoskeletal function.

The eEF1A mRNA levels in W64A+ and W64Ao2 endosperm were different from those previously described (Habben *et al.*, 1994). At least at early stages of development, there were no significant differences in eEF1A mRNAs in the normal and mutant endosperm. Sun *et al.* (1997) also found the levels of eEF1A protein were not significantly different in early stages of developing endosperm. One possible explanation for the difference between these results and those of Habben *et al.* (1994) is that eEF1A transcript levels become more divergent late in development. Another is that the

differences between W64A+ and W64Ao2 in eEF1A transcript levels described here for early developing endosperm continue until later developmental stages.

Sequence analysis was performed for seven members of the eEF1A gene family with the objective of determining if functional domains for GTP, aminoacyl tRNA, eEF1B binding are conserved. Even though the abundance of maize eEF1A mRNAs varied in different tissues, the amino acid sequences of the different genes were highly conserved. All of the genes have conserved GTP, aminoacyl tRNA, eEF1B and actin-binding domains. Studies done with aminoacyl tRNA and eEF1B binding regions did not specify the amino acids important in these regions for eEF1A binding. Collings *et al.* (1994) demonstrated that the aspartic acid in the actin-binding domain in *Dictyostelium* eEF1A is conserved. Some of the members of the maize eEF1A gene family have a substitution from aspartic acid to glutamic acid in the same region. These results imply that there are two eEF1A groups that might interact differently with actin.

Physiological differences between eEF1Aa and eEF1Ab genes were also suggested using the yeast two-hybrid system. The clones that code for eEF1Aa-interacting proteins were tested with the eEF1Ab protein. The results demonstrated that eEF1Aa-interacting proteins interacted at different levels with eEF1Ab. The differences in the protein interaction could be due to cloning artifacts to the single amino acid change. Future experiments to test the single amino acid change hypothesis could be done by repeating the PI4-Kinase and yeast two-hybrid system experiments, substituting the third base of the codon that codes for glutamic acid from A to T in eEF1Aa and vice-versa for eEF1Ab.

The interaction between eEF1Aa and eEF1Ab were also tested with actin in the yeast-two hybrid system. eEF1Aa, eEF1Ab and actin were cloned into pAS2 and pACT2, cotransformed into yeast and their interactions verified by growth of the transformed yeast in selective media containing 25 mM AT and by the X-GAL assay. This experiment demonstrated that these proteins cannot interact with themselves (except actin with actin) in either direction. The negative result of this *in vivo* interaction raises

the possibility that the co-localization of eEF1A and actin might be due to an intermediate protein. This hypothesis was tested by transforming yeast with the clones that code for the actin-interacting proteins and eEF1A and vice-versa. None of the clones that code for the eEF1A-interacting proteins were able to interact with actin, but one of the clones that interacted with actin was positive for both eEF1Aa and eEF1Ab. Even though it is premature to draw conclusions from this result, it reinforces the idea that the interaction between actin and eEF1A occurs by an intermediary protein, and it warrants further investigation.

Another interesting experiment that could be done is to verify which eEF1A gene codes for the protein that is correlated with the lysine content. Since eEF1A-specific probes can be created from the 3' end non-coding sequence, we can potentially identify which members of this gene family are responsible for the correlation with lysine content. eEF1A-specific probes created from the 3' end non-coding sequence can be used to analyze endosperm of crosses between plants with high and low endosperm lysine content. If specific eEF1A gene family members map close to loci related to lysine content, their selection would help to develop genotypes with better nutritional value by a map-based approach. Map location would also help to identify other genes involved in this trait.

Physiological differences were demonstrated with eEF1Aa and eEF1Ab. The same experiment can be performed with the other maize eEF1A genes that are transcribed at lower levels. Even though eEF1A sequences are very conserved, the less abundant maize eEF1A genes are less closely related to the two most abundant ones, and physiological differences may eventually be identified.

My goal was to understand more about the roles of eEF1A in maize endosperm and its correlation with lysine in maize endosperm. Transformation of maize endosperm with high lysine proteins has not been successful, in part due to the lack of knowledge of the biology of the endosperm. A more complete understanding of endosperm biology will help us improve its important agronomic characteristics.

REFERENCES

Aarts, M.G.M., Dirkse, W.G., Stiekema, W.J., and Pereira, A. (1993). Transposon tagging of a male sterility gene in *Arabidopsis*. *Science* 363:715-717

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Ploymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. (1991). Complementary DNA sequencing: Expressed sequence Tags and human genome project. *Science* 252:1651-1656

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C., and Venter, J.C. (1992). Sequence identification of 2375 human brain genes. *Nature* 355:632-634

Aguilar, F., Montandon, P., and Stutz, C. (1991). Two genes encoding the soybean translation elongation factor EF-1α are transcribed in seedling leaves. *Plant Mol. Biol.* 17:351-360

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipmann, D. (1990). Basic local alignment search tool. *J. Mol. Biol* 215:403-410

An, G., and Friesen, J.D. (1980). The nucleotide sequence of *tufB* and four nearby tRNA structural genes of *Escherichia coli*. *Gene* 12:33-39

Ann, D.K., Moutsatsos, I.K., Nakamura, T., Lin, H.H., Mao, P.-L., Lee, M.-J., Chin, S., Liem, R.K.H., and Wang, E. (1991). Isolation and characterization of the rat chromosomal gene for a polypeptide (pS1) antigenically related to statin. *J. Biol. Chem.* 266:10437-10437.

Ann, D.K., Lin, H.H., Lee, S., Tu, Z.-J., and Wang, E. (1992). Characterization of the statin-like S1 rat elongation factor 1α as two distinctly expressed messages in rat. *J. Biol. Chem.* 267:699-702.

Annapurna, S., and Reddy, G.M. (1971). Modified *opaque* maize for possible applied use. *Curr. Sci.* 40:581-582.

Arai, K., Clark, B.F., Duffy, L., Jones M.D., Kaziro, Y., Laursen R.A., L'Italien, J., Miller, D.L., Nagarkatti, S., Nakamura, S., Nielsen, K.M., Petersen, T.E., Takahashi, K., and Wade, M. (1980). Primary structure of elongation factor Tu from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 77:1326-1330

Azpiroz-Leehan, R., and Feldmann, KA. (1997). T-DNA insertion mutagenesis in *Arabidopsis* goind back and forth. *TIG* 13:152-156.

Axelos, M., Bardet, C., Liboz, T., LeVan Thai, A., Curie, C., and Lescure, B. (1989). The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1α: Molecular cloning, characterization and expression. *Mol. Gen. Genet.* 219:106-112.

Babiychuk, E., Fuanfthong, M., Van Montagu, M., Inze, D., and Kushnir, S. (1997). Efficient gene trapping in *Arabidopsis thaliana* using a gene trap approach. *Proc. Natl. Acad. Sci. USA* 94:12722-12727.

Bai, C., and Elledge, S.J. (1996). Gene identification using the yeast two-hybrid system. *Methods Enzymol.* 273:331-347

Baker, B., Coupland, G., Federoff, N.V., Starlinger, P., and Schell, J. (1987). Phenotypic assay for excision of the maize controlling element Ac in tobacco. *EMBO J.* 6:1547-1554

Baker, B., Schell, J., Lorz, H., and Federoff, N.V. (1986). Transposition of the maize controlling element Activator in tobacco. *Proc. Natl. Acad. Sci. USA* 83:4844-4848

Ballinger, D.G., and Benzer, S. (1989). Target gene mutation in *Drosophila*. *Proc. Natl. Acad. Sci USA* 86:9402-9406

Bancroff, I., Jones, J.D.G., and Dean, C. (1993). Heterologous transposon tagging of the DRL locus in *Arabidopsis*. *Plant Cell* 5:631-638

Bankier, A.T., Beck, S., Bohni, R., Brown, C.M., Cerny, R., Chee, M.S., Hutchison, C.A. 3d, Kouzarides, T., Martignetti, J.A., Preddie, E., *et al* (1991). The DNA sequence of the human cytomegalovirus genome. *DNA Seq.* 2:1-12

Benner, M.S., Phillips, R.L., Kirihara, J.A., and Messing, J.W. (1989). Genetic analysis of methionine-rich storage protein accumulation in maize. *Theor. Appl. Genet.* 78: 761-767.

Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S. Meeley, R.B., and Briggs, S.P. (1995). Cloning and characterization of the maize *An1* gene. *Plant Cell* 7:75-84.

Berberich, T., Sugawara, K., Harada, M., and Kusano, T. (1995). Molecular cloning, characterization and expression of an elongation factor 1 alpha gene in maize. *Plant Mol. Biol.* 29:611-615

Berg, P. (1993). "Co-chairman's remarks" reverse genetics: directed modification of DNA for functional analysis. *Gene* 15:261-264

Berridge, M.J. (1987). Inositol lipids and cell proliferation. *Biochim. Biophys. Acta* 907:33-45

Boguski, M.S., Lowe. T.M.J., and Tolstoshev. C.M. (1993). dbEST - database for "expressed sequence tags". *Nature Genet. 4:332-333,.*

Bourne, H.R., Sanders, D.A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349:117-127

Boyer, C.D., and Hannah, L.C. (1994). Kernel mutants of corn. In Specialty corns. Hallauer, A.R. CRC Press, Ames, Iowa. pp.2-18.

Brady, R.C., Cabral, F.. and Dedman, J.R. (1986). Identification of a 52 kD calmodulin-binding protein associated with the mitotic spindle apparatus in mammalian cells. *J. Cell. Biol.* 103:1855-1861

Brands, J.H., Maassen, J.A., van Hemert. F.J., Amons. R.. and Moller. W. (1986). The primary structure of the alpha subunit of human elongation factor 1α. Structural aspects of guanine-nucleotide-binding sites. *Eur. J. Biochem.* 155, 167-171

Browning, K.S.. Humphreys. J.. Hobbs. W.. Smith, G.B.. and Ravel. J.M. (1990). Determination of the amounts of the protein synthesis initiation and elongation factors in wheat germ *J. Biol. Chem.* 265:17967-17973.

Browning, K. (1996). The plant translational apparatus. *Plant Mol. Biol.* 32:107-144

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D.. *et al.,* (1996). Complete genome sequence of the Methanogenic Archaeon, *Methnococcus jannaschii. Science.* 273:1058-1073.

Burr, B.. Burr, F.A.. Thompson, K.H.. Albertson, M.C.. and Stuber. C.W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* 118:519-526

Cavallius, J., Popkie, A.P. and Merrick, W.C. (1997). Site-direct mutants of pos-translationally modified sites of yeast eEF1A using a shuttle vector containing a chromogenic switch. *Biochimica et Biophysica Acta* 1350:345-358

Cavallius, J., Rattan, S.I.S., and Clark, B.B.F.C. (1986). Changes in activity and amont of active elongation factor 1α in ageing and immortal human fibroblast cultures. *Exp. Georntl.* 21:149-157

Chen, Q., Brglez, I., and Boss, W.F. (1991). Inositol phospholipids as plant second messengers. *Symp. Soc. Exp. Biol.* 45:159-175

Chien, C.T., Bartel, P.L., Sternglanz, R.. and Field, S. (1991). The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci.USA* 88:9578-9582.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochem.* 162:156-159

Chourey, P.S., and Nelson, O.E., (1976). The enzymatic deficiency conditioned by the *shrunken-1* mutations in maize. *Biochem. Genetic.* 14:1041

Clark, L., and Carbon, J. (1980). Isolation of a yeast centromere and construction of small functional circular chromosomes. *Nature* 287:504-509

Clark, B.F., Kjeldgaard. M., la Cour, T.F., Thirup, S., and Nyborg, J. (1990). Structural determination of the functional sites of *E. coli* elongation factor Tu. *Biochim. Biophys. Acta* 1050:203-208

Clore, A.M., Dannenhoffer, J.M., and Larkins, B.A. (1996). EF-1α is associated with a cytoskeletal network surrounding protein bodies in maize endosperm cells. *Plant Cell* 8:2003-2014

Coe, E.H.. Neuffer, M.G., and Hoisington, D.A. (1988). The Genetics of Corn. In Sprague, G.F., and Dudley, J.W. (eds), Corn and Corn Improvement, Third Edition pp.81-258. ASA, CSSA, SSSA Publishers, Madison WI

Coleman, C.E., Lopes, M.A., Gillikin, J.W.. Boston, R.S., and Larkins, B.A. (1995). A defective signal peptide in the maize high-lysine mutant *floury2*. *Proc. Natl. Acad. Sci. USA.* 92:6828-6831.

Coleman, C.E., Dannenhoffer, J. M., and Larkins, B.A. (1996). The prolamin proteins of maize, sorghum and Coix. In press.

Collings, D.A, Wasteneys, G.O., Miyazaki, M., and Williamson, M. (1994). Elongation factor 1α is a component of the subcortical actin budles of characcean algae. *Cell Biol. Intern.* 18:1019-1024

Condeelis, J. (1995). Elongation factor 1α, translation and the cytoskeleton. *TIBS* 20:169-170.

Cottrelle, P., Thiele, D., Price, V., Memet, S., Micouin, J., Marck, C., Buhler, J., Sentenac, A., and Fromageot, P. (1985). Cloning, nucleotide sequence and expression of one of the two genes coding for yeast elongation factor 1 alpha. *J. Biol. Chem.* 260:3090-3096.

Datta, S.K., Peterhans, A., Datta, K., and Potrykus, I. (1990). Genetically engineered fertile Indica-rice recovered from proteoplasts. *BioTechnology* 8:736-740

Davies, E., Fillingham, B.D., Oto, Y., and Abe, S. (1991). Evidence for the existence of cytoskeleton-bound polysomes in plants. *Cell Biol Int Rep* 15:973-981

Dje, M.K., Mazabraud, A.,Viel, A., Maire, M.L., Denis, H., Crawford, E., and Brown, D.D. (1990). Three genes under different development control encode elongation factor 1-α in *Xenopus laevis*. *Nucleic Acids Res.* 18:3489-3493.

Dunn, M.A., Morris, A., Jack, P.L., and Hughes, M.A. (1993). A low temperature-responsive translation elongation factor 1α from barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* 23:221-225

Durfee, T., Becherer, K., Chen, P., Yeh, S., Yang, Y., Kilburn, A.E., Lee, W., and Elledge, S.J. (1993). The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes and Development* 7:55-569

Durso, N.A., and Cyr, R.J. (1994). Beyond translation: elongation factor-1α and the cytoskeleton. *Protoplasma* 180:99-105.

Edmonds, B.T. (1993). ABP50: an actin-binding elongation factor 1 alpha from *Dictyostelium discoideum. J. Cell Biochem.* 52, 134-139.

Edmonds, B.T., Murry, J., and Condeelis, J. (1995) pH regulation of the F-actin binding properties of *Dictyostelium* elongation factor 1-α. *J. Biol. Chem.* 270:15222-15230

Ejiri, S., Kawamura, R., and Katsumata, T. (1994). Interactions among four subunits of elongation factor 1 from rice embryo. *Biochim. Biophys. Acta* 1217:266-272

Emmons, S., and Yesner, L. (1984). High frequency excision of transposon element *Tc1* in the nematode *Caenorhabditis elegans* is limited to somatic cells. *Cell* 36:599-605

Evan, M.J., Carlton, M.B.L., and Russ, A.P. (1997). Gene trapping and functional genomics. *TIBS* 13:370-374

Federoff, N.V., Furtek, D.B., and Nelson, O.E. (1984). Cloning of the *bronze* locus in maize by a simple and generalized procedure using the transposable element Activator (*Ac*). *Proc. Natl. Acad. Sci. USA* 81:3825-3829

Feix, G., and Quayle, T. (1993). Structure and expression of zein genes of maize. *Cr. Rev. Plant Sci.* 12:111-127.

Feldmann, K.A. (1991). T-DNA insertion mutagenesis in *Arabidopsis* mutational spectrum. *Plant J.* 1:71-82

Feldmann, K.A., and Marks, M.D. (1987). *Agrobacterium* mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. *Mol. Gen. Genet.* 208, 1-9.

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

Flavell, R., O'Dell, M., and Smith, D. (1986). Repeated sequence DNA comparisons between *Tritium* and *Aegilops* species. *Heredity* 42:409-322

Fonzi, W.A., Katayama, C., Leathers, T., and Sypherd, P.S. (1985). Regulation of protein synthesis factor EF-1α in *Mucor racemosus*. *Mol. Cell Biol.* 5:1100-1103.

Fraley, R. T., Rogers, S.G., and Horsch, R.B. (1986). Genetic transformation in higher plants. *CRC Crit. Rev. Plant Sci.* 4:1-46

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D,, Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., *et al* (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 20:397-403

Fromm, M., and Walbot, V., (1987) Transient expression of DNA in plant cells. In Plant Gene research; Plant DNA Infection Agents, ed. T. Hohn, J, Schell pp. 304-310, Wien/New York: Springer.

Frydenberg, J., Poulsen, K., Petersen, A.K., Lund, A., and Olesen, O.F. (1991). Isolation and characterization of the gene encoding EF-1 alpha O, an elongation factor 1-alpha expressed during early development of *Xenopus laevis. Gene* 109:185-192

Garcia-Hernandez. M., Davies, E., and Staswick, P.E. (1994). *Arabidopsis* p40 homologue. A novel acidic protein associated with the 40 S subunit of ribosomes. *J. Biol. Chem.* 269:20744-20749

Gevers, H.O., and Lake J.K. (1992). Development of modified opaque2 maize in South Africa. In E. T. Mertz (ed). Quality Protein Maize. (The American Society of Cereal Chemists, St. Paul MN), pp. 49-78.

Gibson, S., and Somerville, C. (1993) Isolating plant genes. *Trends Biotechnol.* 11:306-313.

Gilman, A.G. (1987). G proteins: transducers of receptor-generated signals. *Annu Rev Biochem* 56:615-649

Glover, D.V., and Mertz, E.T. (1987). Corn. In R.A Olson, and K.J. Frey (eds). Nutritional Quality of Cereal Grains, Genetic and Agronomic Improvement (ASA-CSSA-SSSA, Madison, WI.), pp. 184-336.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology* 179:247-266

Goffeau, A., Barrell. B.G., Bussey, H., Davis, R.W., Dujon, B., Felmann, H., Galibert, F., Hohensel, J.D., Jacq, C., Johnston. M., Louis, E.J., Mewes, H.W., Murakami, Y., Phillippsen. P., Tettelin. H., and Oliver. S.G. (1996). Life with 6000 genes. *Science* 274:546-567.

Goldberg, R.B., Hoschek, G., Kamalay, J.C., and Timberlake, W.E. (1978). Sequence compelxity of nuclear and polysomal RNA in leaves of the tobacco plant. *Cell* 14:123-131

Gonen, H., Smith, C. E., Siegel, N.R., Kahana, C., Merrick, W.C., Chakraburtty, K., Schwartz, A.L., and Ciehanover, A. (1994). Protein synthesis elongation factor EF-1α is essential for ubiquitin-dependent degradation of certain N^{α}-acetylated proteins and may be substituted for by the bacterial elongation factor EF-Tu. *Proc. Natl. Acad. Sci. USA* 91:7648-7652.

Goodman, H.M., Ecker, J.R., Dean, C. (1995). The genome of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 92(24):10831-10835

Gordon-Kamm, W.J., Spencer, T.M. Mangano, M.L., Adams, T.R., Daines, R.J. *et al.*, (1990). Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* 2:603-618

Habben, J.E., Kirleis, A.W., and Larkins, B.A.. (1993). The origin of the lysine-containing proteins in *opaque2* maize endosperm. *Plant Mol. Biol.* 23:825-838.

Habben, J.E., Moro, G.L., Hunter, B.G., Hamaker, B.R., and Larkins, B.A. (1995). Elongation factor 1α concentration is highly correlated with the lysine content of maize endosperm. *Proc. Natl. Acad. Sci. USA* 92:8640-8644.

Hannah, L.C., Giroux, M., and Boyer, C. (1993). Biotechnological modifications of carbohydrates for sweet maize and maize improvement. In Biotechnological modification of nutritional crop, R. Jorgensen, ed. Scientia Horticulturae 55:177-197

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F., and Thompson, R. (1989). The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J.* 8:2795-2801.

Hasezawa, S., and Nagata, T. (1993). Microtubule organization centers in plant cells: localization of a 49 kDa protein that is immunologically cross-reactive to a 51 kDa protein from sea urchin centrosomes in synchronized tobacco BY-2 cells. *Protoplasma* 176:64-74.

Hayashi, Y., Urade, R., Utsumi, S., and Kito, M. (1989). Anchoring of peptide elongation factor EF-1α by phosphatidylinositol at the endoplasmic reticulum membrane. *J. Biochem.* 106:560-563.

Heidecker, G., Chaudhuri, S., and Messing, J.W. (1991). Highly clustered zein gene sequences reveal evolutionary history of the multigene family. *Genomics* 10:719-726.

Helentjaris, T. (1987). A genetic linkage map for maize based on RFLPs. *Trends in Genet.* 3:217-221

Hershey, J.W.B. (1991). Translation control in mammalian cells. *Ann. Rev. Biochem.* 60:717-755.

Hershey, J.W., Asano, K., Naranda, T., Vornlocher, H.P., Hanachi, P., and Merrick, W.C. (1996). Conservation and diversity in the structure of translation initiation factor EIF3 from humans and yeast. *Biochimie* 78:903-907

Hiatt, W.R., Garcia, R., Merrick, W.C., and Sypherd, P.S. (1982). Methylation of elongation factor 1 alpha from the fungus *Mucor*. *Proc. Natl. Acad. Sci. USA* 79:3433-3437

Hodge, R., Paul, W., Draper, J., and Scott, R. (1992). Cold-plaque screening: a simple technique for the isolation of low abundance, differentially expressed transcritps from conventional libraries. *Plant J.* 2: 257-260

Hofte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M., Charpenteau, J.L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D.Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G.D., Gigot, C., Fleck, J., Philipps, G., Axelos, M., Bardet, C., Tremousaygue, and Lescure, G. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* 4: 1051-1061.

Hoog, C. (1991). Isolation of large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acid Res.*19:6123-6127

Hooykaas, P.J.J. (1989). Transformation of plant cells via *Agrobacterium*. *Plant Mol. Biol* 13:327-336.

Hooykaas, J.J. and Schilperoort, R.A. (1992). *Agrobacterium* and plant genetic engineering. *Plant Mol. Biol.* 19:15-38

Hovemann, B., Richter, S., Waldorf, U., and Cziepluch, C. (1988). Two genes encode related cytoplasmic factor 1α (EF-1α) in *Drosophila melanogaster* with continuos and stage specfic expression. *Nucleic Acids Res.* 16:3175-3194.

Izawa, T., Miyazaki, C., Yamamoto, M., Terada. R., Iida, S., and Shimamoto, K (1991) Introduction and transposition of the maize transposable element Ac in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* 227:391-396

Janmey, P.A., and Stossel, T.P. (1989) Gelsolin-polyphosphoinositide interaction. Full expression of gelsolin-inhibiting function by polyphosphoinositides in vesicular form and inactivation by dilution, aggregation, or masking of the inositol head group. *J. Biol. Chem.* 264:4825-4831

Janssen, G.M.C., van Damme, H.T.F., Kriek, J., Amons, R. and Moller, W. (1994). The subunit structure of elongation factor 1α from *Artemia. J. Biol. Chem.* 269:31410-31417.

Jaskunas, S.R., Lindahl, L., and Nomura, M. (1975). Identification of two copies of the gene for the elongation factor EF-Tu in *E. coli. Nature* 257:458-462

Jonak, J., Petersen, T.E., Meloun, B., and Rychlik, I. (1984). Histidine residues in elongation factor EF-tu from *Escherichia coli* protected by aminoacyl-tRNA against photo-oxidation. *Eur J Biochem* 144:295-303

Joshi, C.P. (1987). An inspection of the domain between putative TATA box and translation start site in 79 plants. *Nucleic Acid Res.* 15:6643-6653

Jurss, K., Junghahn, I., and Bastrop, R. (1992). The role of elongation factors in protein synthesis rate variation in white teleost muscle. *J Comp. Physiol.* [B] 162:345-350

Kaur, K.J., and Ruben, L. (1994). Protein translation elongation factor-1α from *Trypanosoma brucei* binds calmodulin. *J. Biol. Chem.* 269:23045-23050.

Kawahara, R., Sunaborl, S., Fukuda, H., and Komamine, A. (1992). A gene expressed preferentially in the globular stage of somatic embryogenesis encodes elongation factor 1α in carrot. *Eur. J. Biochem.* 209:157-162.

Keegan, L., Gill, G., Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* 231: 699-704.

Keith, C.S., Hoang, D.O., Barrett, B.M., Feigelman, B., Nelson, M.C., Thai, H., and Baysdorfer, C. (1993). Partial sequence-analysis of 130 randomly selected maize cDNA clones. *Plant Physiol.* 101:329-332

Kempim, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D. and Yanofsky, M.F. (1997). Target disruption in *Arabidopsis. Nature* 389:802-803.

Kishore, G.M., and Shah, D.M. (1988). Amino acid biosynthesis inhibitor as herbicides. *Annu. Rev. Biochem.* 57:627-663

Knowles, R.V., and Phillips, R.L. (1988). Endosperm development in maize. *Int. Rev. Cytol.* 112, 97-136

Knudsen, S.M., Frydenberg, J., Clark, B.F., and Leffers, H. (1993). Tissue-dependent variation in the expression of elongation factor-1 alpha isoforms: isolation and characterisation of a cDNA encoding a novel variant of human elongation-factor 1 alpha. *Eur. J. Biochem.* 215:549-554

Kouchi, H., and Hata, S. (1993). Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* 238:106-109

Krieg, P.A., Varnum, S.M., Wormington, W.M. and Melton, D.A. (1989). *Dev. Biol.* 133:93-100 .

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1996). Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. *Proc Natl Acad Sci USA* 93(15):8145-8150

Kumagai, F., Hasezawa, S., Takahashi, Y., and Nagata, T. (1995). The involvement of protein synthesis elongation factor 1α in the organization of microtubules on the perinuclear region during the cell cycle transcition from M phase to G1 phase in tobacco BY-2 cells. *Bot. Acta*, in press (1995).

Kuriyama, R., Savereide, P., Lefebvre, P., and Dasgupta, S. (1990). The predicated amino acid sequence of a centrosphere protein in dividing sea urchin eggs is similar to elongation factor (EF-1α). *J. Cell Sci.* 95:231-236

Larkins, B.A., and Hurkman, W.J. (1978). Synthesis and deposition of zein in protein bodies of maize endosperm. *Plant. Physiol.* 62:256-263.

Larkins, B.A., Lending, C.R., Wallace, J.C., Galili, G., Kawata, E.E., Geetha, K.B., Kriz, A.L., Martin, D.N., and Bracker, C.E. (1989). Zein gene expression during maize endosperm development. In R.B. Goldberg (ed). The molecular basis of plant development, (Alan R. Liss, Inc. New York, NY), pp. 109-120.

Lassing I, Lindberg U (1985). Specific interaction between phosphatidylinositol 4,5-bisphosphate and profilactin. *Nature* 314:472-474

Lee, S. Francoeur, A-M., Liu, S., and Wang, E. (1992). Tissue-specific expression in mammalian brain, heart and muscle of S1, a member of the elongation factor-1α gene family. *J. Biol. Chem.* 267:24064-24068

Lee, S., Wolfraim, L.A., and Wang, E (1993). Differential expression of S1 and elongation factor-1 alpha during rat development. *J. Biol. Chem.* 268:24453-24459

Lenstra, J.A., van Vliet, A., Amberg, A.C., van Hemert F.J., and Moller, W. (1986). Genes coding for the elongation factor EF-1α in *Artemia. Eur. J. Biochem.* 155:475-483

Liboz, T., Bardet, C., Le Van Thai, A., Axelos, M., and Lescure, B. (1990). The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha are actively transcribed. *Plant Mol. Biol.* 14:107-110

Linz, J.E., Katayama, C., and Sypherd, P.S. (1986). Three genes for elongation factor EF-1α in *Mucor racemosus. Mol. Cell Biol* 6:593-600.

Linz, J.E., and Sypherd, P. (1987) Expression of three genes for elongation factor 1 during morphogenesis of Mucor racemosus. *Mol. Cell Biol.* 7:1925-1932

Littlepage, L.E., Garman, M.E., Mendenhall, J.M., and Morejohn, L.C. (1995). Plant microtubule (MT) severing, stiffening and bundling by wheat germ protein synthesis elongation factor 1α (EF-1α). *Mol. Cell Biol.* 6:258a

Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P., and Couplant, G., (1993). The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: Identification of an albino mutation induced by *Ds* insertion. *Proc. Natl. Acad. Sci. USA* 90:10370-10373

Lopes, M.A., Coleman, C.E., Kodrzycki, R., Lending, C.R., and Larkins, B.A. (1994). Synthesis of an unusual α-zein protein is correlated with the phenotypic effects of the *floury2* mutation in maize. *Mol. Gen. Genet.* 245, 537-547.

Lopes, M.A., and Larkins, B.A. (1995). Genetic analysis of opaque2 modifier gene activity in maize endosperm. *Theor. Appl. Genet.* 19, 271-281.

Lopes, M.A., Takasaki, K., Bostwick, D.E., Heletjaris, T., and Larkins, B.A. (1995). Identification of two opaque2 modifiers loci in Quality Protein Maize. *Mol. Gen. Gen.* 247:603-613

Lund, A., Knudsen, S.M., Vissing, H., Clark, B., and Tommerups, N. (1996). Assignment of human elongation factor 1 genes: EEF1A maps to chromosome 6q14 and EF!A2 to 20q13.3. *Genomics* 36, 359-361

Ma, H., Yanofsky, M.F., and Meyerowitz, E.M. (1991). AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 5:484-495

Ma, J., and Ptashne, M. (1987). Detection analysis of GAL4 defines two transcriptional activating segments. *Cell* 48, 847-853.

Ma, J., and Ptashne, M. (1988). Converting a eukaryotic transcriptional inhibitor into an activator. *Cell* 55, 433-446.

Mahe. A., Grisvard, J., and Dron, M. (1992). Fungal- and Plant-specific gene markers to follow the bean anthracnose infection process and nromalise a bean chitinase mRNA induction. *Mol. Plant Microbe Inter.* 5:242-248

Majerus, P.W., Connolly, T.M., Deckmyn, H., Ross, T.S., Bross, T.E., Ishii, H., Bansal, V.S., Wilson, D.B. (1986). The metabolism of phosphoinositide-derived messenger molecules. *Science* 234:1519-1526

Massung, R.F., Esposito, J.J., Liu, L.I., Qi, J., Utterback, T.R., Knight, J.C., Aubin, L., Yuran, T.E., Parsons, J.M., Loparev, V.N., *et al* (1993). Potential virulence determinants in terminal regions of variola smallpox virus genome. *Nature* 366:748-751

Mccabe, D.E., Swain, W.E., Martinell, B.J., and Christou, P. (1988). Stable transformation of soybean (*Glycine max*) by particle acceleration. *BioTechnology* 6:923-926.

McCormick, F., Clark, B.F.C., la Cour, T.F.M., Kjeldgaard, M., Norskov-Lauritsen, L., and Nyborg, J. (1985). A model for the tertiary structure of p21, the product of the ras oncogene. *Science* 230:78-82.

McCreery, T., and Helentjaris, T. (1994a). Production of hybridization probes by the PCR utilizing digoxigenin-modified nucleotides. In P. Isaac (ed). Protocols for nucleic acid analysis by nonradiocative techniques (Humana Press, Totowa, NJ), pp. 107-112.

McCreery, T., and Helentjaris, T. (1994b). Hybridization of digoxigenin labeled probes to Southern blots and detection by chemiluminescence. In P. Isaac (ed). Protocols for nucleic acid analysis by nonradiocative techniques (Humana Press, Totowa, NJ), pp. 112-117

McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowell, J.M., and Meagher, R.B. (1995). Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*. *Plant J.* 8:613-622

Melton, D.W. (1994) Gene targeting in the Mouse. *BioEssay.* 16:633:638.

Mena, M., Ambrose, B.A. Meely, R.B., Briggs, S.P., Yanofsky, M.F., and Schmidt, R.J. (1996). Diversification of C-function activity in maize flower development. *Science* 274:1537-1540

Mertz, E.T., Bates, L.S., and Nelson, O.E. (1964). Mutant gene that changes the protein composition and increases the lysine content of maize endosperm. *Science* 145:279-280.

Metz, A.M., Timmer, R.T., Allen, M.L., and Browning, K.S. (1992) Sequence of a cDNA encoding the alpha-subunit of wheat translation elongation factor 1. *Gene* 120:315-316

Miyajima, A., Shibuya, M., and Kaziro, Y. (1979). Construction and characterization of the two hybrid ColE1 plasmids carrying *Escherichia coli* *tufB* gene. *FEBS Lett* 102:207-210

Moldave, K. (1985). Eukaryotic protein synthesis. *Ann. Rev. Bioche* 54:1109-1149.

Morelli, J.K., Shewmaker, K., and Vayda, M.E. (1994). Biphasic stimulation of translational activity correlates with induction of translation elongation factor 1 subunit α upon wounding in potato tubers. *Plant Physiol.* 106, 897-903.

Motorin, Yu.A., Wolfson, A.D., Orlovsky, A.F., and Gladilin, K.L. (1988). Mammalian valyl-tRNA synthetase forms a complex with the first elongation factor. *FEBS Lett* 238:262-264

Murray, M., Cramer, J., Ma, Y., West, D., Romero-Sverson, J., Pitas, J., DeMars, S., Vilbrandt, L., Kirshman, J., Macleester, R., Shilz, J., and Lotzer, J. (1988). Agrigenetics maize RFLP linkage map. Maize Genet. Coop. Newsletter 62:89-90.

Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M., and Kaziro, Y. (1984). Polypeptide chain elongation factor 1α (EF-1α) from yeast: nucleotide sequence of one of the two genes for EF-1α from *Saccharomyces cerevisiae. EMBO J.* 3:1825-1830

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* 2:279-289

Negrutiu, I., Shillito, R.D., Potrykus, I., Biasini, G., and Sala. F (1987). Hybrid genes in the analysis of transformation conditions. I. Setting up a simple method for direct gene transfer to protoplast. *Plant Mol. Biol.* 8:363-373.

Nelson, O.E. (1969). Genetic modification of protein quality in plants. *Adv. Agron.* 21, 171-194.

Nelson, O.E., Mertz, E.T., and Bates, L.S. (1965). Second mutant gene affecting the amino acid pattern of maize endosperm proteins. *Science* 150, 1469-1470.

Neuffer, M.G., Jones, L., and Zuber, M.S. (1968). The mutants of maize. Crop Science Society of America, Wisconsin

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., *et al* (1992). Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. *Mol. Cell Biol.* 223:1-7

O'Hare. K. (1990). Searching for a needle in haystacks via the polymerase chain reaction. *TIG* 6:202-203

Ohta, K., Toriyama, M., Miyazaki, M., Murofushi, H., Hosoda, S., Endo, S., and Sakai, H. (1990). The mitotic apparatus-associated 51-kDa protein from sea urchin eggs is a GTP-binding protein and is immunologically related to yeast polypeptide elongation factor 1 alpha. *J. Biol. Chem.* 265:3240-3247

Opdenakker. G., Cabeza-Arvelaiz, Y., Fiten, P., Dijkmans, R., Van Damme, J., Volckaert, G., Billiau. A., Van Elsen, A., Van der Schueren, B., Van den Berghe, H., *et al* (1987). Human elongation factor 1 alpha: a polymorphic and conserved multigene family with multiple chromosomal localizations. *Hum Genet* 75:339-344

Osborne, T.B., and Mendel, L.B. (1914). Nutritive properties of protein of the maize kernel. *J. Biol. Chem.* 18:1-6.

Park, Y.S., Kwak, J.M., Kwon, O.Y., Kim, Y.S., Lee, D.S., Cho, M.J., Lee, H.H., and Nam, H.G. (1993). Generation of expressed sequences tags of random root cDNA clones of Brassica napus by single-run partial sequencing. *Plant Physiol.* 103:359-370

Payrastre, B,, van Bergen en Henegouwen, P.M., Breton, M., den Hartigh, J.C., Plantavid, M., Verkleij, A.J., and Boonstra, J. (1991). Phosphoinositide kinase, diacylglycerol kinase, and phospholipase C activities associated to the cytoskeleton: effect of epidermal growth factor. *J. Cell Biol.* 115:121-128

Pearson, W.R., Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85:2444-2448

Peter, M.E., Reiser, C.O., Schirmer, N.K., Kiefhaber, T., Ott, G., Grillenbeck, N.W., and Sprinzl, M. (1990). Interaction of the isolated domain II/III of *Thermus thermophilus* elongation factor Tu with the nucleotide exchange factor EF-Ts. *Nucleic Acids Res.* 18:6889-6893

Plasterk, R.H.A. (1993). Reverse genetics of *Caenorhabditis elegans*. *BioEssay* 14:629-633

Pokalsky, A.R., Hiatt, W.R., Ridge, N., Rasmussen, R., Houck, C.M., Shewmaker, C.K. (1989). Structure and expression of an elongation factor 1α in tomato. *Nucleic Acids Res* 17:4661

Potrykus, I., and Shillito, R.D. (1986). Protoplast: isolation, culture and plant regeneration. *Methods Enzymol.* 118:549-578

Potrykus, I., Paszkowski, J., Saul, M.W., Petruska, J., and Shillito, R.D. (1985). Molecular and general genetics of a hybrid foreign gene introduced into tobacco by direct gene transfer. *Mole. Gen. Genet.* 199:167-177.

Rao, T.R., and Slobin, L.I. (1986). Structure of the amino-terminal end of mammalian elongation factor Tu. *Nucleic Acids Res* 14:2409

Riis, B, Rattan, S.I., Clark, B.F., and Merrick, W.C. (1990). Eukaryotic protein elongation factors. *Trends Biochem Sci* 15:420-424

Rodes, C.A., Pierce, D.A., Mettler, I.J., Mascarenhas, D., and Detmer, J.J. (1989). Genetically transformed maize plants from protoplasts. *Sciences* 240:204-207

Roest, S., and Gilissen, L.J.W. (1989). Plant regeneration from protoplast: a literature review. *Acta Bot. Neerl.* 38:1-23, 1989

Rosenzweig, B., Liao, L.W., and Hirsh, D. (1983). Target sequences for the *C. elegans* transposable element *Tc1*. *Nucleic Acids Res* 11:7137-7140

Rothstein, S.J., Lahners, K.N., Lotstein, R.J., Carozzi, N.B., Jayne, S.M., and Rice, D.A. (1987). Promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation. *Gene* 53:153-161

Rushforth, A.M., Saarri, B., and Anderson, P. (1993). Site-selected insertion of the transposon *Tc1* into *Caenorhabditis elegans* myosin light chain gene. *Mol. Cell Biol.* 13:902-910.

Sacchi, G.A., Zocchi, G., and Cocucci, S. (1984). Changes in form of elongation factor 1α during germination of wheat seeds. *Eur. J. Bioch.* 139:1-4

Sambrook, J., Fritsch, E.F., and Maniatis T. (1989). Molecular Cloning 2nd ed.. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467

Schirmaier, F., and Philippsen, P. (1984). Identification of two genes coding for the translation elongation factor EF-1 alpha of *S. cerevisiae*. *EMBO J.* 3:3311-3315

Schmidt, R.J., Burr, F.A., Aukerman, M.J., and Burr, B. (1990). Maize regulatory gene *opaque2* encodes a protein with a 'leucine-zipper' motif that binds to zein DNA. *Proc. Natl. Acad. Sci. USA* 87:46-50.

Schultes, N.P., Brutnell, T.P., Allen, A., Dellaporta, S.L., Nelson, T., and Chen, J. (1996). Leaf permease I gene in maize is requred for chroloplast development. *Plant Cell* 8: 463-475

Seidler, L., Peter, M., Meissner, F., and Sprinzl, M. (1987). Sequence and identification of the nucleotide binding site for the elongation factor Tu from *Thermus thermophilus* HB8. *Nucleic Acids Res.* 15:9263-9277

Shattuck-Eidens, D.M., Bell, R.N., Neuhausen, S.L., and Helentjaris, T. (1990). DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. *Genetics* 126: 207-217

Sheehy, E.R., Kramer, M., and Hiatt, W.R. (1988). Reduction of polygalacturonase activity in tomato fruit by antisense RNA. *Proc. Natl. Acad. Sci. USA* 85:8805-8809.

Shepherd, J.C., Walldorf, U., Hug, P., and Gehring, W.J. (1989). Fruit flies with additional expression of the elongation factor EF-1 alpha live longer. *Proc. Natl. Acad. Sci. USA* 86:7520-7521

Sherman, M., and Sypherd, P.S. (1989). Role of lysine methylation in the activities of elongation factor 1A. *Archives of Biochem. Biophys.* 275:371-378.

Shibuya, M., Nashimoto, H., and Kaziro, Y. (1979). Cloning of an *EcoRI* fragment carrying *E. coli tufA* gene. *Mol. Gen. Genet.* 170:231-234

Shillito, R.D., Saul, M.W., Paszkowski, J., Muller, M., and Potrykus, I. (1985). High frequency direct gene transfer. *BioTechnology* 4:1099-1103.

Shiina, N., Gotoh, Y., Kubomura, N., Iwamatsu, A., and Nishida, E. (1994). Microtubule severing by elongation factor 1α. *Science* 266, 282-285.

Shimamoto, KR., Terada, R, Izawa, T., and Fujimoto, H. (1989). Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* 338:274-276

Shotwell, M.A., and Larkins, B.A. (1989). The biochemistry and molecular biology of seed storage proteins. In A. Marcus (ed). The Biochemistry of Plants, Vol. 15, A Comprehensive Treatise (Academic Press, San Diego, CA), pp. 297-345.

Singleton, C.K., Delude, R.L., Ken, R., Manning, S.S., and McPherson, C.E. (1991). Structure, expression, and regulation of members of the developmentally controlled V and H gene classes from Dictyostelium. *Dev. Genet.* 12:88-97

Slobin, L.I. (1980). The role of eukaryotic elongation factor Tu in protein synthesis. *Eur. J. Biochem.* 110:555-563

Slobin, L.I., Clark, R.V., and Olson, M.O.J. (1981). Functional and structural studies on a tryptic fragment of eucaryotic elongation factor Tu from rabbit reticulocytes. *Biochemistre* 20:5761-5767

Soave, C., and Salamini, F. (1984). Organization and regulation of the zein genes in maize endosperm. Phil. Trans. R. Soc. Lond. B. 304, 341-347.

Sofia, H.J., Burland, V., Daniels, D.L., Plunkett, G. 3rd, and Blattner, F.R. (1994). Analysis of the Escherichia coli genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. *Nucleic Acids. Res* 22:2576

Staiger, C.J., and Schliwa, M. (1987). Actin localization and function in higher plants. *Protoplasma* 141, 1-12.

Stankovic, B., Abe, S., and Davies, E. (1993). Co-localization of polysomes, cytoskeleton, and membranes with protein bodies from corn endosperm: Evidence from fluorescence microscopy. *Protoplasma* 177, 66-72.

Sun, Y., Carneiro, N., Clore, A.M., Moro, G.L., Habben, J.E., and Larkins, B.A. (1997). Charaterization of maize elongation factor 1A and its relationship to protein quality in the endosperm. *Plant Physiol.* 115:1101-1107

Sutoh, K., and Mabuchi, I. (1989). End-label fingerprintings show that an N-terminal segment of depactin participates in interaction with actin. *Biochemistry* 28:102-106

Taylor, B., and Powell, A. (1982). Isolation of plant DNA and RNA. *Focus* 4:4-6.

Thiele, D., Cottrelle, P., Iborra, F., Buhler, J., Sentenac, A., and Fromageot, P. (1985). Elongation factor 1α from Sacharomyces cerevisiae, rapid large-scale purification and molecular characterization. *J. Biol. Chem.* 260:3084-3089

Toriyama, K., Arimoto, Y., Uchimiya, H., and Hinata, K., (1988). Transgenic rice plants after direct gene transfer into protoplasts. *BioTechnology* 6:1072-1074

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* 76:4350-4354

Uchimiya, H., Kidou, S.I., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nishi, R., Hashimoto, H., Matsubayashi, Y., Kidou, N., Umeda, M., and Kato, A. (1992). Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice *(Oryza sativa* L). *Plant J.* 2:1005-1009

Ursin, V.M., Irvine, J.M., Hiatt, W.R., and Shewmaker, C.K. (1991). Developmental analysis of elongation factor-1α expression in transgenic tobacco. *Plant Cell* 3:583-591

van Damme, H.T.F., Amos, R., Karssies, R., Timmers, C.J., Janssen, G.M.C., and Moller, W. (1990). Elongation factor 1β of Artemia: localization of functional sites and homology to elongation factor 1δ. *Biochim. Biophys. Acta* 1050:241-247

van Damme, H.T.F., Amons, R., and Moller, W. (1992). Identification of the sites in the eukaryotic elongation factor 1α involved in the binding of elongation factor 1β and aminoacyl-tRNA. Eur. J. Biochem. 207:1025-1034.

van der Krol, A.R., Mol, J.N., and Stuitje, A.R. (1988). Antisense genes in plants: an overview. *Gene* 72:45-50

van Hemert, F.J., van Ormondt, H., and Moller, W. (1983). A bacterial clone carrying sequences coding for elongation factor EF-1 alpha from *Artemia*. *FEBS Lett.* 157:289-293

van Hemert, F.J., Amons, R., Pluijms, W.J., van Ormondt, H., and Moller, W. (1984). The primary structure of elongation factor EF-1 alpha from the brine shrimp *Artemia*. *EMBO J.* 3:1109-1113

Vayda, M.E., Shewmaker, C.K., and Morelli, J.K (1995). Translation arrest in hypoxic potato tubers is correlated with the aberrant association of elongation factor EF-1α with polysomes. *Plant Mol. Biol.* 28:751-757

Venema, R.C., Peters, H.I., and Traugh, J.A. (1991). Phosphorylation of elongation factor 1 (EF-1) and valyl-tRNA synthetase by protein kinase C and stimulation of EF-1 activity. *J. Biol. Chem.* 266:12574-12580

Viel, A., le Maire, M., Philippe, H., Morales, J., Mazabraud, A., and Denis, H. (1991). Structural and functional properties of thesaurin a (42Sp50), the major protein of the 42 S particles present in Xenopus laevis previtellogenic oocytes. *J. Biol. Chem.* 266:10392-10399

Villegas, E., Vasal, S.K., and Bjarnarson, M., (1992). Quality Protein Maize - What is it and how was it developed. In E. T. Mertz (ed). Quality Protein Maize (The American Society of Cereal Chemists, St. Paul MN), pp. 27-48.

Villemur, R., Joyce, C.M., Haas, N.A., Goddard, R.H., Kopczak, S.D., Hussey, P.J., Snustad, D.P., and Silflow, C.D. (1992). α-Tubulin gene family of maize (*Zea mays* L.) Evidence for two ancient α-tubulin genes in plants. *J. Mol. Biol.* 227:81-96

Vojtek, A.B., Hollenberg, S.M., and Cooper, J.A. (1993). Mammalian *Ras* interacts directly with the serine/threonine kinase *Raf. Cell* 74:205-214

Walbot, V. (1992). Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 43:49-82

Wang, G., Zhang, H., Ding, Q., Xie, Y., and Dai, J., (1996). Studies on the transfer techniques of three maize genes. *Chin. J. Biotechnol.* 12:53-60

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierrymieg, J., and Sulston, J. (1992). A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet.* 1: 114-123

Webster, G., Monakembashi, S., Greulich, K.O., and Wolfrum, J. (1988). Injection of DNA into plant cells with a UV laser microbeam. *Naturwissenschaften* 75:35-36

Webster, G.C., and Webster, S.L. (1984). Specific disappearance of translatable messenger RNA for elongation factor one in aging *Drosophila melanogaster*. *Mech. Ageing Dev.* 24:335-342

Wolf, M.J., Buzan, C.L., and Rist, C.E. (1952). Structure of the mature corn kernel. I. Gross analtomy and structure relationship. *Cereal Chem.* 29:321-333.

Yamamoto, K., and Sasaki, T. (1997). Large-scale EST sequencing in rice. *Plant Mol. Biol.* 35:135-144

Yang, F., Demma, M., Warren, V., Dharmawardhane, S., and Condeelis, J. (1990). Identification of an actin-binding protein from *Dictyostelium* as elongation factor 1α. *Nature* 347: 494-496

Yang, W., Burkhart, W., Cavallius, J., Merrick, W.C., and Boss, W.F. (1993). Purification and characterization of a phosphatidylinositol 4-kinase activator in carrot cells. *J. Biol. Chem.* 265, 392-398.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346(6279):35-39

Yokota, T., Sugisaki, H., Takanami, M., and Kaziro, Y. (1980). The nucleotide sequence of the cloned *tufA* gene of *Escherichia coli*. *Gene* 12:25-31

Yonezawa, N., Nishida, E., Iida, K. K., Yahara, I., and Sakai, H. (1990). Inhibition of the interactions of cofilin, destrin, and deoxyribonuclease I with actin by phosphoinositides. *J. Biol. Chem.* 265:8382-8386

Zang, W., and Wu, R. (1988). Efficient regeneration of transgenic rice plants from rice protoplasts and correctly regulated expression of foreign genes in the plants. *Theor. Appl. Henet.* 76:835-840.

Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T., and Plasterk, R.H. (1993). Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci. USA* 90(16):7431-7435

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"
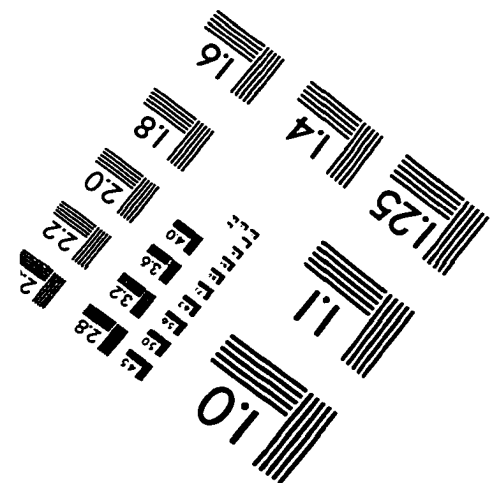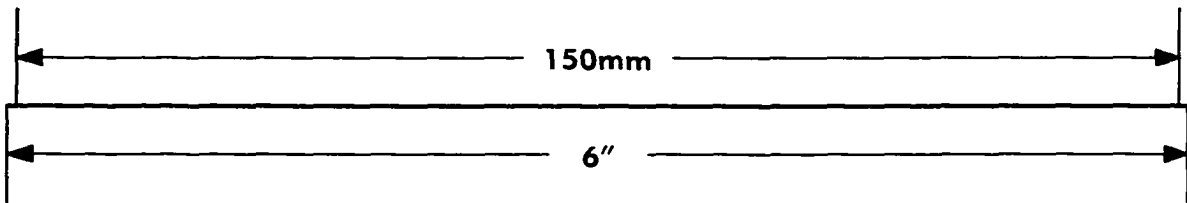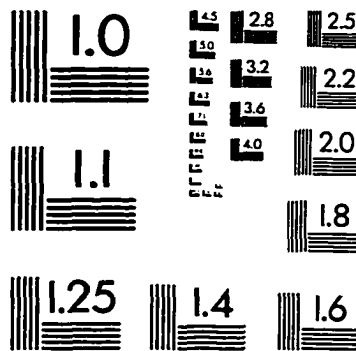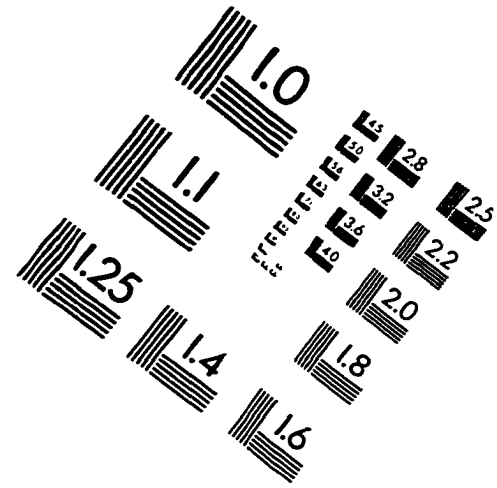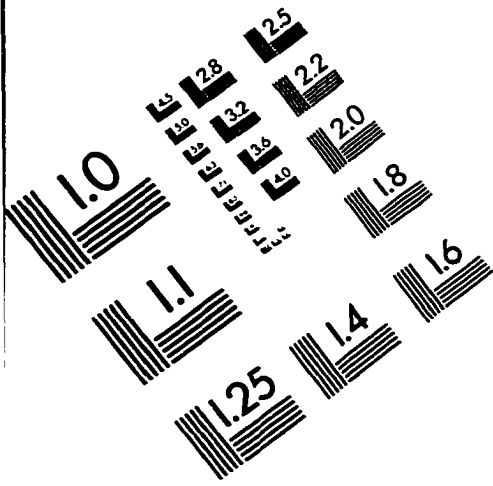
APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved