

## **A collaborative program for functional genomics of the maize endosperm.**

Our long-term goal is comprehensive genetic dissection of the molecular mechanisms underlying endosperm development and metabolism. We reason that in the near term the most informative genes will be those that have single mutant phenotypes. Molecular analysis of these mutants will lead to other genes that function in the same or interacting pathways. These will be functionally analyzed in their turn. We expect to clone 50 to 100 new endosperm mutants in this project. Our approach to identifying this critical initial set of genes can be viewed as passing all maize genes through a series of screens, each filtering genes by a different criterion. In the end we will have a large set of cloned genes that confer endosperm phenotypes as single mutants. Each of these mutants then serves as a point of entry into the network of all genes expressed in the endosperm.

**1. Screen one is phenotype.** In principle, all genes that have visible knock out phenotypes in endosperm can be tagged by forward transposon mutagenesis. We estimate that at least 2000 independent mutations will be needed to approach saturation for qualitative endosperm factors (~300 loci). We have constructed a Mu population large enough to produce the 2000 visible mutants. The end result will be a set of maize lines that are enriched for informative Mu insertions.

**2. Screen two is endosperm expression.** The goal is to identify the subset of tagged genes whose wildtype transcripts are present in the existing endosperm cDNA microarray. This is feasible because we have perfected a method for amplifying all germinal Mu insertion sites in pools of plants by TAIL PCR. By probing the microarray with Mu-TAIL products from suitably arranged pooled DNA samples, insertions that are in genes represented in the array can be traced to individual mutant lines. cDNA hits will include candidates for the mutant gene, plus incidental gene knockouts not associated with phenotypes. Independent mutant lines that hybridize to the same cDNA's define a "hybridization group". The depth of the array determines the total number of hybridization groups we will identify.

**3. Screen three is complementation testing.** Mutant lines within each hybridization group will be systematically tested for allelism. Finding independent mutant alleles of the same gene will confirm the assignment of clones to phenotypes.

**4. Screen four is bioinformatics and functional analysis.** The mutant gene sequence and phenotype will be used to infer function and guide experimental analysis of selected genes. The participating labs have complementary interests covering the essential processes underlying endosperm development. We expect to efficiently identify many mutants in relevant pathways.

### **Participants:**

Donald R. McCarty, P.I., U. of Florida

Joachim Messing, Rutgers

Brian Larkins, U. of Arizona

Philip Becraft, Iowa State University

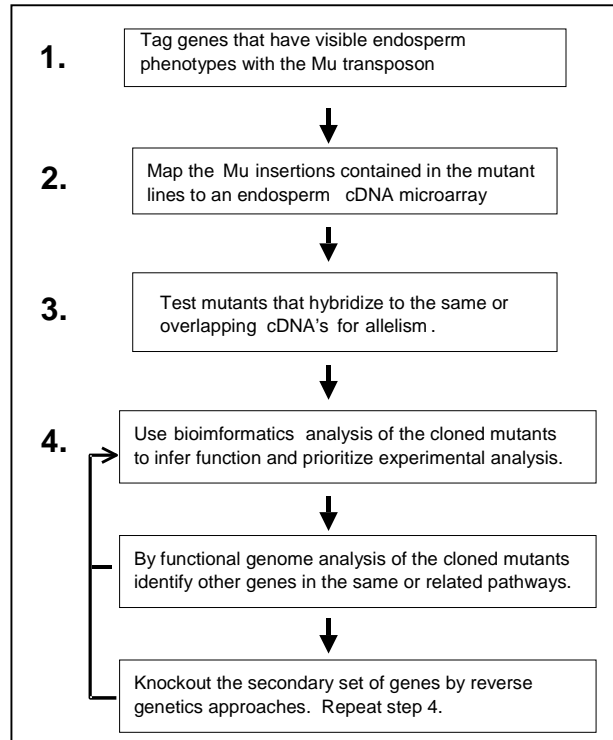
Karen E. Koch, U. of Florida

L. Curtis Hannah, U. of Florida

**Project Description: A collaborative program for functional genomics of the maize endosperm**

**1. Project Overview.** Our long-term goal is comprehensive genetic dissection of the molecular mechanisms underlying endosperm development and metabolism. We propose a functional genomics approach outlined on the right. Specifically, we address the first steps (1-4) in this program. We reason that in the near term the most informative genes will be those that have single mutant phenotypes. Molecular analysis of these mutants will lead to other genes that function in the same or interacting pathways. These will be functionally analyzed in their turn (step 4). Only a small fraction of the ~300 endosperm mutants have been molecularly analyzed to date, we expect to clone 50 to 100 new mutants in this project.

Our approach to identifying this critical initial set of genes can be viewed as passing all maize genes through a series of screens, each filtering genes by a different criterion. In the end we will have a large set of cloned genes that confer endosperm phenotypes as single mutants. Each of these mutants then serves as a point of entry into the network of all genes expressed in the endosperm.



**1. Screen one is phenotype.** In principle, all genes that have visible knock out phenotypes in endosperm can be tagged by forward transposon mutagenesis. We estimate that at least 2000 independent mutations will be needed to approach saturation for qualitative endosperm factors (~300 loci). We have already constructed a Mu population large enough to produce the 2000 visible mutants. Furthermore, this population is inbred and designed for future screens of quantitative phenotypes. The end result will be a set of maize lines that are enriched for informative Mu insertions.

**2. Screen two is endosperm expression.** The goal is to identify the subset of tagged genes whose wildtype transcripts are present in the existing endosperm cDNA microarray. This is feasible because we have perfected a method for amplifying all germinal Mu insertion sites in pools of plants by TAIL PCR. By probing the microarray with Mu-TAIL products from suitably arranged pooled DNA samples, insertions that are in genes represented in the array can be traced to individual mutant lines. cDNA hits will include candidates for the mutant gene, plus incidental gene knockouts not associated with phenotypes. Independent mutant lines that hybridize to the same cDNA's define a "hybridization group". In this way, each mutant line in the collection that detects a clone in the array will be assigned to at least one hybridization group. The depth of the array determines the total number of hybridization groups we will identify.

**3. Screen three is complementation testing.** Mutant lines within each hybridization group will be systematically tested for allelism. Finding independent mutant alleles of the same gene will confirm the assignment of clones to phenotypes. The current cDNA array (6500 clones) should yield 50 to 100 different endosperm mutants.

**4. Screen four is bioinformatics and functional analysis.** The mutant gene sequence and phenotype will be used to infer function and guide experimental analysis of selected genes. The participating labs have complementary interests covering the essential processes underlying endosperm development. We expect to efficiently identify many mutants in relevant pathways.

<b>Participating Investigators:</b>	<b>Project role:</b>	<b>Research interests:</b>
Donald R. McCarty, P.I., Florida	Project coordinator, Uniform Mu population development and screening, Mu-TAIL PCR implementation and microarray analysis, field genetics	Mutants affecting hormone signaling, aleurone organization
L. Curtis Hannah, Florida	Field genetics coordinator, allelism testing, small kernel screen	Mutants affecting regulation of starch biosynthesis
Karen E. Koch, Florida	Greenhouse genetics and DNA sampling coordinator, small kernel screen, digital image database of endosperm mutants, field genetics	Mutants affecting sugar signaling, regulation of carbon metabolism
Joachim Messing, Rutgers	Bioinformatics coordinator, early endosperm cDNA library, parental imprinting screen, EST sequencing	Mutants affecting regulation of zein genes, parental imprinting
Brian Larkins, Arizona	Microarray coordinator, endosperm microarray development and analysis, Rescue Mu screen	Mutants in cell cycle regulation, zein biosynthesis
Phill Becraft, Iowa State	<i>dek1</i> subtracted cDNA library, aleurone microarray development, aleurone mutant screen	Mutants affecting aleurone differentiation

## 2. Relevance of this project to plant genome research

We propose a new strategy for focused functional analysis of complex gene systems in plants. By coupling transposon mutagenesis and microarray technologies we can efficiently target the genes relevant to a particular organ or biological process. We will focus on the maize endosperm for the following reasons.

**The endosperm has unique economic and biological significance.** As our largest single primary food source, the cereal endosperm is among the most economically important structures in biology. The product of double fertilization, it is also among the most novel. The unique value of the endosperm of cereal grains as a food source can be traced directly to the subsequent evolution of the major grain-filling pathways, namely starch and storage protein biosynthesis, within this novel structure.

**The endosperm is a model for organ development.** The endosperm displays the basic themes of organ formation in plants, but with a simpler structure (2 tissues made up of 3 basic cell types) than other major organs. Hence, the central issues in endosperm development are directly relevant to other plant organs. Key issues include regulation of the cell cycle, partitioning of growth between cell division and cell expansion, regulation of cell expansion and terminal differentiation, cell-to-cell signaling, and determination of cell fate.

**The endosperm is a model for integrated regulation of metabolic and developmental pathways.** The high growth rate and grain filling processes place extraordinary demands on cellular metabolism and metabolic regulation making the endosperm an excellent model for biochemical genetics. Molecular genetic studies indicate that the regulation of the major grain filling pathways is highly integrated, and key metabolic intermediates have been implicated in this process.

**The maize endosperm is a powerful genetic model.** As a highly visible component of the seed, the maize endosperm is very conducive to genetic analysis. Tools such as the B-A translocation series are uniquely suited to manipulation of endosperm mutants. Hence the large number and phenotypic complexity of the endosperm mutants of maize is at once a challenge and an opportunity. If quantitative phenotypes are included, the number of genes identifiable by phenotype grows even larger.

In principle any gene that has a discernable phenotype will have a discernable biological function. For this reason we will give priority to identifying and analyzing those genes that have phenotypes. Among these, we expect to find key genes in pathways of interest. Others will be so-called “housekeeping genes”, though we suspect this conveniently vague term has outlived its utility in the era of functional genomics. The more relevant question is in what pathway or process does each phenotypic gene function? Robust bioinformatics and functional genomics tools such as expression profiling will yield clues for many of the genes we identify. Therefore, each mutant can serve as a separate point of entry into a local network of functionally related genes. The participating investigators have complementary interests that span many of the essential processes of endosperm development.

**3. Relationship of this project to other maize genomics initiatives.** The plant genomics initiative has yielded a key set of tools for high throughput functional genomics of maize. Resources especially relevant to this project are a significant maize EST collection including approximately 6,500 endosperm cDNA's (ZmDB, <http://www.zmdb.iastate.edu/>), the Stanford Rescue Mu gene discovery project ([http://www.zmdb.iastate.edu/zmdb/nsf\\_grant.html](http://www.zmdb.iastate.edu/zmdb/nsf_grant.html)), and the MTM insertional mutagenesis facility developed at Cold Spring Harbor Lab. Our strategy is centered on microarrays constructed from the large endosperm cDNA collection. While the extant cDNA's will well serve the goals of the present project, we have detected significant gaps in the coverage of endosperm expressed genes. For this reason, we propose to enhance EST coverage of specific early endosperm and aleurone stages.

We can take advantage of three *Mutator* based transposon tagging resources; NSF funded Rescue Mu and MTM populations and the large Uniform Mu population developed by the McCarty lab with support from the Florida Agricultural Experiment Station. Uniform Mu was designed for forward seed mutagenesis and addresses specific limitations of other populations (see Preliminary Results). The MTM population is designed primarily for creating directed knockouts in cloned genes, but it can be a supplemental source of new endosperm mutants for this project. More importantly, MTM will be important for eventually creating knockouts of endosperm-expressed genes revealed through subsequent analysis of our initial set of mutants.

As it was conceived, the Rescue Mu population would be an ideal resource for this project; however, the Rescue Mu system is not yet functional and indeed faces significant long-term challenges. One difficulty is achieving and maintaining adequate mutation frequency. Most of the mutations in the existing population are caused by native Mu transposons and are not due insertions of the engineered Rescue Mu transposon. In order to increase the proportion of Rescue Mu induced mutations it will be necessary to reduce the copy number of other Mu elements through breeding. Because the high mutation frequency of active Mu lines is at least in part related to high copy number, reducing the overall Mu copy number can be expected to substantially lower the overall forward mutation frequency. Breeding for higher Rescue Mu copy number may eventually compensate for this loss, however, this will also eliminate much of the benefit of tagging mutants with a marked element. In any event, Rescue Mu populations with adequate mutation frequency will not be available in the near-term.

A potentially more serious limitation to effective implementation of Rescue Mu is the efficiency of plasmid rescue from the maize genome. One haploid genome of maize contains about 3.25 pg of DNA. Thus, 100 ng of maize genomic DNA will contain about 30,000 copies of a homozygous Rescue Mu locus, or 15,000 copies of a hemizygous Rescue Mu insertion. Highly optimized plasmid transformation protocols yield between  $10^8$  and  $10^9$  colonies per 100 ng DNA for pure, super-coiled plasmid vector containing no insert (Stratagene, Inc). Even at these extreme efficiencies, only 0.3%

to 3% of the plasmid molecules in solution are recovered as colonies (e.g.  $1 \times 10^9 / 3 \times 10^{10} = 0.03$ ). If we accept these numbers as the practical upper limit for plasmid recovery, the best we can hope for recovery of a hemizygous integrated plasmid in maize DNA would be in the range of 45-450 colonies per 100 ng DNA. Importantly, this limit does not take ligation efficiency and other loss factors into account. In practice, only a small fraction of the plasmid copies released by a restriction digest of maize genomic DNA will be recovered as closed circles of appropriate size. If the realizable efficiencies fall 10-fold below the ideal limit (~5-50 colonies per 100 ng DNA ligation), then efficient sampling from pooled restriction digests and pooled plant samples becomes problematic. The current generation Rescue Mu plasmid has an additional disadvantage ([http://www.zmdb.iastate.edu/zmdb/nsf\\_grant.html](http://www.zmdb.iastate.edu/zmdb/nsf_grant.html)). Because six-base restriction sites occur with roughly equal frequencies in maize DNA, the Bam HI double digest proposed for selection against non-transposed Rescue Mu will also eliminate about 50% of the transposed copies. In other words, the probability of hitting a BamHI site first going in either direction from Rescue Mu is 50%, fragments that have different sites on each end will not circularize.

For these reasons, we will not rely on near-term output of Rescue Mu per se. Fortunately, our Mu-TAIL approach works with native Mu elements and we will still be able to take advantage of any endosperm mutants present in the existing Rescue Mu population.

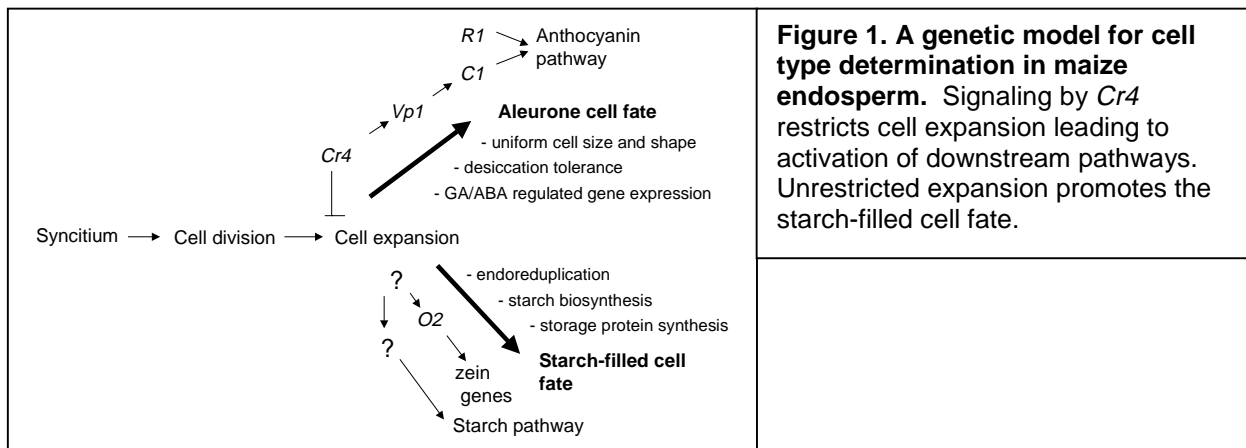
#### 4. Background.

**Genetics of endosperm formation.** Based on a statistical analysis of EMS induced mutants, Neuffer and Sheridan (1980) estimated that at least 300 maize genes condition visible endosperm phenotypes. The *Robertson's Mutator* transposon system of maize generates a similar spectrum of seed phenotypes (Scanlon et al, 1994). The Scanlon et al. study illustrates the enormous potential for using *Mutator* to analyze endosperm mutants. Of 63 *Mutator* induced mutations surveyed, at least 14 identified new loci. Of all loci represented, only 9 were previously described by Neuffer and Sheridan (1980). Moreover, only a small fraction of the known endosperm mutants have been molecularly analyzed (Scanlon and Meyer, 1998).

**Endosperm growth and development.** The mature endosperm consists of two tissues, the interior starch-filled endosperm and an epidermal layer called the aleurone. The aleurone consists of uniform single layer of small cuboidal cells that have large vacuoles and a capacity to accumulate anthocyanins (Walbot, 1993), except at base of the endosperm where specialized transfer cells form (Hueros et al., 1995). The starch-filled cells in the interior become highly enlarged and densely packed with starch grains and zein storage protein bodies (Lending & Larkins, 1989), the latter more prevalent in peripheral cells. As the kernel reaches maturity starch-filled cells senesce, apparently undergoing a form of programmed cell death (Young and Gallie, 1999), whereas, aleurone cells acquire desiccation tolerance and remain viable in the dry seed.

**Differentiation of starch-filled cells.** Key pathways relevant to differentiation of starch-filled cells include cell cycle regulation (Schweizer et al., 1995) starch biosynthesis (Hannah, 1997) and zein storage protein biosynthesis (Lending and Larkin, 1989). The transitions through syncitial, cellular and endopolyploid phases make the starch-filled endosperm an attractive model for cell-cycle studies. Multiple cell cycle factors have been isolated and their expression studied in endosperm (Sun et al., 1997; Grafi et al., 1996). The biochemical steps of the starch biosynthetic pathway are genetically defined by the *ae1*, *sh1*, *sh2*, *bt1*, *bt2*, *wx*, *du*, *su1*, and *su2* mutants (Hannah, 1997); whereas, the *fl2*, *o2*, *o6* and *o15* mutants identify genes involved in storage protein synthesis and assembly (Soave et al., 1981; Kodrzycki et al., 1989; Geetha et al., 1991; Lohmer et al., 1991; Coleman et al., 1997; Dannenhoffer et al., 1995). The *o2* gene encodes a transcriptional activator of the 22 kd zein gene family (Schmidt et al., 1987; 1990). Additional transcriptional regulators of the zein genes have been defined biochemically (Mena et al., 1998; Pysh et al., 1993).

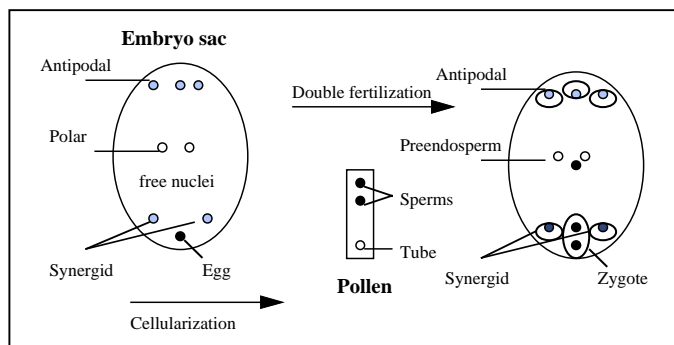
**Where are the regulatory genes?** *o2* is the only genetically defined regulatory gene in a pathway specific to the starch-filled cell fate. No regulatory mutants specific to the starch biosynthetic pathway are known. There is evidence that regulation of major grain-filling pathways is highly integrated in endosperm (Giroux et al., 1994; Maddaloni et al., 1996; Muller et al., 1997). Gene responses to sugars and C/N balance have been implicated (Koch, 1997; Motto et al., 1996). Moreover, many pleiotropic *defective kernel (dek)* mutations that fail to initiate or complete grain-filling have been identified, but not studied in detail (Neuffer and Sheridan, 1981; Scanlon et al., 1994). These are likely to include mutations in “housekeeping genes” as well as important developmental mutants. A key challenge is to devise molecular and genetics strategies that can be used to effectively analyze



this large, complex phenotypic class.

**Aleurone differentiation.** Anthocyanin biosynthesis is the best-understood pathway specific to the aleurone cell fate. Key regulators of the pathway include the *vp1*, *c1* and *r1* genes (Dooner et al., 1991). *Vp1* encodes a transcription factor required for regulation of the seed maturation pathway in both the embryo and aleurone (McCarty et al., 1991). *C1* and *R1* encode myb and H-L-H transcription factors, respectively, that interact and specifically activate structural genes in the anthocyanin biosynthetic pathway (Goff et al., 1992; Sainz et al., 1997). *Vp1* activates transcription of the *C1* gene (Hattori et al., 1992), and binds specifically to a cis-element in the *C1* promoter (Kao et al., 1996; Suzuki et al., 1997). In addition, *Vp1* represses germination specific  $\alpha$ -amylase genes (Hoecker et al., 1995; 1999). Other pleiotropic aspects of the *vp1* phenotype suggest a still broader role in aleurone gene expression (Dooner, 1985).

The *dek1(gay1)* (Neuffer and Sheridan, 1980; Neuffer, 1993), *cr4* (Becraft et al., 1996) and *Dap* (Gavazzi et al., 1997) mutants block aleurone differentiation over all or part of the endosperm. The pleiotropic *dek1* mutant has an early embryo lethal phenotype and floury textured starch-filled endosperm (Neuffer, 1993). The *cr4* mutation blocks aleurone differentiation over portions of the abgerminal surface (Becraft et al., 1996). In regions that lack aleurone, surface cells differentiate as large starch-filled endosperm cells. *Cr4* encodes a receptor kinase with a novel extracellular domain suggesting that it functions in signaling required for aleurone differentiation. In leaves, *cr4* causes abnormal expansion of epidermal cells. This suggests that *Cr4* is a negative regulator of cell expansion and that aleurone differentiation depends on the restriction of endosperm cell expansion.



**Figure 2. Double fertilization and the basis of gene imprinting in endosperm.** In the model of Messing and Graosniklaus (1999) imprinted nuclei are filled. Activated nuclei are open.

Hence restriction of cell expansion may be a critical step in determination of starch-filled v.s. aleurone cell fates (Figure 1).

**Gene imprinting and dosage interactions.** Double fertilization, fusion of a second male gamete with a differentiated somatic cell of the female gametophyte (fig.2), has a profound influence on endosperm development. Studies in maize have illuminated the parental imprinting (Lin, 1982, 1984; Kemicle and Alleman, 1991; Chaudhuri and Messing, 1994; Lund et al., 1995ab) and gene dosage interactions between maternal and paternal genomes (Birchler and Hart, 1987; Birchler, 1993; Gou and Birchler, 1994). The *medea* (Grossniklaus et al., 1998; Kiyouse et al., 1999), *fis* (Chaudhuri et al., 1999; Luo et al., 1999) and *fie* (Ohad et al., 1999) mutants of arabidopsis identify a class of polycomb and set domain-related factors that regulate reprogramming of the central cell from gametophyte formation to endosperm development. Messing and Grossniklaus (1999) hypothesize that the silenced state of a parentally imprinted gene is the ground state, which is equivalent to the premieiotic or sporophytic state (Figure 2). During the postmeiotic mitotic divisions in the embryo sac the two polar nuclei differentiate causing imprinted genes to be activated (nuclei are open circles); whereas, the egg nucleus remains in the ground state (nucleus is filled circle). Once in activated state the affected promoter remains in an inducible state throughout endosperm development. Gene expression can occur at any point later in endosperm development. Therefore, the activation process does not require a resetting for the next generation. In case of allele-specific imprinting, the effect is independent of grandparentage and only dependent on the immediate parents.

### 5. Preliminary results from the McCarty lab.

We have focused on developing two essential tools for the analysis of seed mutants:

- 1) a genetically uniform transposon tagging population (Uniform-Mu) suitable for forward and reverse genetics and quantitative phenotypic analyses;
- 2) a simple PCR method (Mu-TAIL) that allows efficient extraction of the germinal Mu insertion sites from a plant or pool of plants.

**The Uniform-Mu population.** The goals of Uniform Mu address specific limitations of existing tagging populations for seed mutagenesis. 1) *Mutator* lines are typically heterogeneous. Background variation limits quantitative phenotypic analysis and confounds interpretation of qualitative phenotypes. Consequently several generations of back crossing are typically needed to evaluate knock out phenotypes. 2) Because recessive mutations may be carried undetected in outcrossed populations, the independence of allelic mutations is not assured. 3) Unless Mu activity is under careful genetic control, somatic insertion events may interfere with reverse genetics and cloning strategies. 4) Finally, the progenitors of new mutants are frequently unknown limiting the choice of cloning strategies (see below). In their current forms, the TUSC, Rescue Mu and MTM populations are affected by one or more of these limitations. This is in part because they were not designed to target a particular class of phenotypes.

To create a genetically uniform, pedigreed tagging population, we introgressed *Mutator* into a color converted W22 inbred background by alternating generations of backcrossing with self-pollination. The *bz1-mum9* allele is carried along as a marker for presence of *MuDR* (Chomet, 1993). Each cycle,



**Figure 3.** Self-pollinated BC3 ear showing segregation of a new miniature seed mutation and *bz1-mum9*. Note segregation of stable bronze (Mu inactive) seed within the *bz1* class.

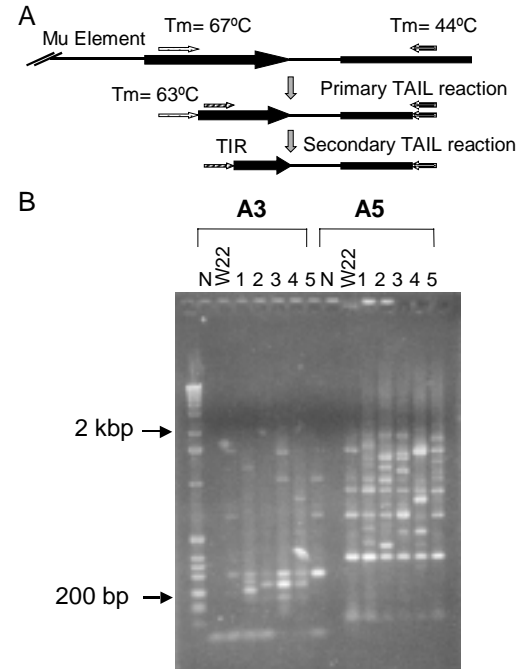
500 to 1500 backcross progeny were grown and self-pollinated to monitor the frequency of new seed mutations. For parents of the next cycle, we selected densely spotted kernels from self's that showed no visible seed phenotypes (i.e. "clean" ears). Purging of visible seed phenotypes reduces the genetic load carried forward and ensures that mutations isolated in the next generation are independent events. Thus, our Uniform-Mu population has four essential features: 1) Active *Mutator* is in an inbred background while maintaining a high mutation frequency. 2) Seed phenotypes are purged from the population each generation to ensure that new mutations are independent events in a genetically uniform background. 3) The pedigree of each line is known so that the immediate progenitor of every new mutation is unambiguous. 4) Because back cross lines are marked with bz1-mum9 and heterozygous for MuDR, Mu inactive derivatives of each new mutant can be selected in the same generation by picking stable bronze kernels (see Figure 3).

The BC3 population (527 self's scored) had an overall seed mutation frequency of 6.5%. The 45 BC3 plants selected as parents for BC4 seed produced "clean" self-ears and derive from 18 highly mutagenic BC3 families. The selected pedigrees averaged 13% new seed mutants or about twice the population average. For BC4 and BC5 we have combined the self and back cross in the same generation. In Fall 1999, we selfed and back crossed 900 plants from 45 different BC4 families to create BC5 seed. The BC4 plants are highly uniform and very W22-like in plant phenotype and flowering time. Scoring of the BC4 self's for seed mutations is not yet completed, however, we conservatively estimate that these crosses will yield between 30,000 and 50,000 BC5 seed suitable for this project.

Interestingly, color converted W22 contains a significant number of endogenous Mu elements (see below). This likely explains why we see consistently high forward mutation rates. With the W22 inbred, we may only need to introgress one or two MuDr elements to obtain reasonable forward mutagenesis.

**Efficient extraction of Mu insertion sites by Mu-TAIL PCR.** The second key to our strategy is a highly efficient method for amplifying sequences that flank Mu family elements. TAIL-PCR has been used very effectively to amplify sequences flanking insertions in *Arabidopsis* (Liu et al., 1995). This technique was adapted to amplify Mu insertion sites by Dr. Mark Settles, a post-doc in the McCarty lab (figure 4A). In brief, PCR is performed using a high  $T_m$  primer specific for the TIR sequence of Mu-family transposons, and a low  $T_m$  arbitrary primer that is designed to be promiscuous at low annealing temperatures. The TIR primer is degenerate enough to recognize all Mu family TIR sequences. By interlacing PCR cycles with high and low annealing temperatures, amplification of Mu-specific fragments is favored. A second round of TAIL is performed using a nested TIR primer to enhance the specificity of the PCR products.

**Characterization of Mu-TAIL PCR products.** We have tested six different arbitrary primers. The products from two are shown in figure 4B. The size range of the Mu-TAIL products is 150 bp to >2000 bp. Thus Mu-TAIL amplicons contain a significantly greater amount of flanking sequence information than amplicons generated by other methods such as AIMS (Frey et al., 1998). Pooled DNA samples (12 plants each) from five different BC3 Mu families and the W22 recurrent parent were analyzed. Note that Mu-TAIL detects Mu family elements in W22 that are shared across all lines, plus numerous products that are specific to different BC3 families. Cross-hybridization experiments indicated that



**Figure 4.** (A) Schematic of TAIL-PCR (B) Mu-TAIL PCR with arbitrary primers A3 and A5.

each arbitrary primer produced different but overlapping sets of products (not shown). By combining products from different arbitrary primers, we get deep representation of the *Mu* elements present.

The specificity of Mu-TAIL PCR is remarkably high, roughly 95% of the PCR products were from authentic Mu insertion sites. Mu-TAIL products from the 99-060 family were cloned into a TOPO-TA plasmid (Invitrogen, Inc) and analyzed by sequencing. The Mu-TAIL products contain a sufficiently long TIR sequence to confirm if they derive from bona fide transposon ends. Out of 38 different clones sequenced 36 were verified to include the Mu-TIR sequence located downstream of the nested primer. Two products lacked full TIR sequences, one was derived from the chloroplast genome and the second a low copy genomic fragment. Consistent with Cresse et al (1995), Mu-TAIL products have very little repetitive DNA. Mu-TAIL blots probed with total maize DNA and genomic Southern blots (not shown) probed with 99-060 TAIL products identified only one high copy number band which we traced to the PCR reaction containing the chloroplast DNA fragment.

**Functional genomics applications of Mu-TAIL: High-throughput sequencing.** To explore the potential of Mu-TAIL PCR for functional genomics applications, we undertook a pilot sequencing project supported by a grant of 960 sequencing reactions from the high-throughput sequencing lab at the University of Florida (UF-ICBR Genomics Core Lab). 96 random clones (192 in one case) were sequenced from each of 9 different microlibraries constructed in the TOPO-TA vector.

In total, we identified at least 230 unique Mu insertion sites. Excluding failed or poor quality reads (~20%), empty vector (<3%) and artifactual clones (e.g. chloroplast DNA) the sampling from the libraries was about 3X redundant overall. 25% of the unique Mu flanking sequences produced convincing BLASTX hits in genbank, and/or TBLASTX hits on plant EST's. A partial list of hits is shown in Table 1 (below).

Many of the insertions that we have sequenced would be predicted to disrupt gene function (e.g. Figure 5).

```
>dbj|BAA75813.1| (AB019240) RPR1h [ Oryza sativa]
      Length = 901

Query:  1 PESVQRLAYLDTLDVRQTRVQRLPQGVGKARKLRHILAD 117
        PES+ +L+ L TL+++QT++Q+LPQ + K +KLRH+LAD
Sbjct:  603 PESIGKLSSLLTLNIKQTKIQKLPQSIVKIKKLRHLLAD 641

Query:  130 DGSQQSEFRSFTALEPPKALTSFGELETLETVQANKDMAMKLARMMLXSVSIDNISSAL 309
        + +QS FR F ++ PK L++ ELQTLLETV+A+K+A +L ++MQL SV IDNI +
Sbjct:  644 EDEKQSAFRYFIGMQAPKELSNLEELQTLLETVEASKELAEQLMKMLQLRSVWIDNIRTTDD 703

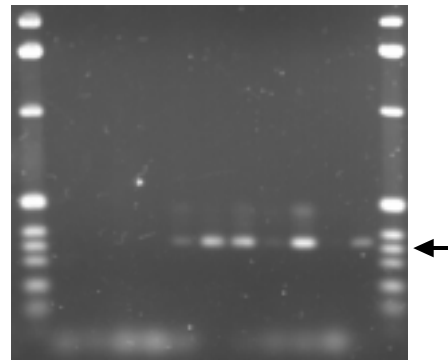
Query:  310 CAELFASVSKLQFLTSLLLSATDEHEPLSFQNLVPKSSYLSXLTVRXSWLGR 465
        CA LFA++SK+ L+SLLLSA+ E+E L + L P+S L L VR W R
Sbjct:  704 CANLFATLSKMPLLSLLLSASHENETLCLEALKPESEELHRLIVRGCWAAR 755
```

**Figure 5.** BLASTX hit in a NBS-LRR gene (RPR1-like) from the family 2 (99-060) Mu-amplicon microlibrary.

As a test case, we analyzed the insertion in the NBS-LRR resistance gene analog (#23, table 1; figure 5). To determine whether the insertion was heritable, individuals from the 99-060 family were analyzed by PCR using a Mu-TIR primer and a second gene specific primer derived from the Mu-TAIL sequence. Figure 6 shows that the insertion is absent in the W22 parent and segregating (5:4) in the 99-060 family indicating that the backcross Mutator parent was heterozygous for the insertion. We confirmed heritability of several other knockouts (e.g. #4, #16), and in at least one case detected linkage to a seed phenotype (#16).

A key goal of the pilot study was to determine how well Mu-TAIL clones represent the population of Mu elements contained in a DNA sample. Hence we sequenced 96 Mu-TAIL products cloned from our recurrent parent, color-converted W22. By Southern blotting we estimate that W22 contains

Family 2 (99-060)  
N W22 1 2 3 4 5 6 7 8 9



**Figure 6.** RPR-like insertion segregates in family 2. PCR products with a TIR primer and gene specific primer.

**Table 1. Partial list of hits (62 total) in maize genes identified by Mu-TAIL sequencing.**

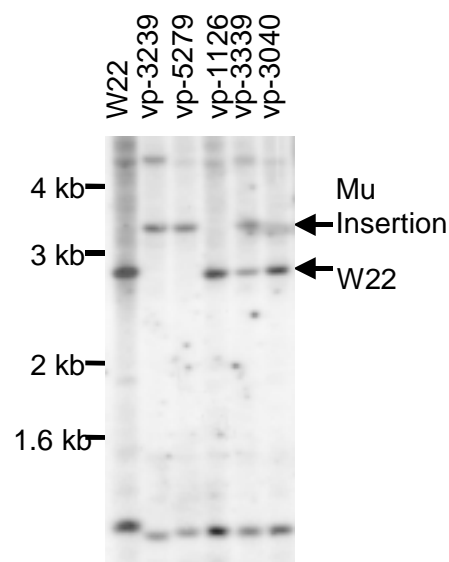
Mu-TAIL Clone:	Sequence homology detected by BLAST:
1. 99F-263B-1-1A9	Maize ear tissue cDNA
2. 99F-263B-1-1C9	Maize endosperm cDNA
3. 99F-263B-1-1C7	Proline-5-carboxylate reductase (proline biosynthesis)
4. a1-2-1-1C9	Auxin transport protein (EIR/PIN-like), presumptive efflux carrier
5. a1-2-1-1G5	Nucleotide sugar epimerase
6. 3286-1-1D8	Protein phosphatase 2c
7. 99F-235-1-1H4	Hypothetical protein Arabidopsis, maize endosperm cDNA
8. 99F-235-1-1H10	E1-E2 type ATPase
9. D1-ADP-1-1G9	Ring3 nuclear kinase like protein
10. MS407-1-1F6	SCARECROW/RGA like protein
11. MS407-1-1B10	Topoisomerase I
12. MSBE-1-1D1	KORRIGAN-like 1,2 beta-glucanase, cellulase
13. MSBE-1-1G9	Maize leaf cDNA, acyl carrier protein desaturase?
14. MSBE-1-1A10	Identical to maize catalase-1
15. MSBE-1-1A11	MADS box-like protein [ <i>Oryza sativa</i> ]
16. MSBE-1-1D11	Aminoalcoholphosphotransferase
17. MSBE-1-1B5	GDP-mannose pyrophosphorylase
18. MSBE-1-1D7	Dim1p homolog
19. MS114	Myosin-like coiled coil domain protein
20. MSW22-1-1C3	EF-hand protein
21. MSW22-1-1E11	LRR receptor protein kinase
22. 113-5F	Protein Disulfide Isomerase Er-60
23. 113-15F	NBS-LRR protein, RPR1h –rice
24. 113-28F	Maize leaf cDNA
25. 113-79F	DHHC domain protein

approximately 30 different Mu-family transposable elements. We identified at least 27 unique Mu insertions among the Mu-TAIL clones. Even though we did not sequence exhaustively from the W22 library, the results indicate that Mu-TAIL represents a high percentage of the Mu insertions present in the genome. We used the W22 results to assess the contribution of recurrent parental elements to insertions represented in a typical BC4 microlibrary. 87 good reads obtained from a microlibrary constructed from a BC4 family (12 plant pool; 99F-235), included 25 unique sequences. Of those 18 were novel and 7 were shared with W22.

Finally we constructed microlibraries from several tagged seed mutants. In each case, 15 or more unique elements were identified. Based on preliminary Southern analysis, for example, we have identified one clone that is linked to a lethal kernel mutation. Interestingly, that insertion resides in a gene for aminoalcohol phosphotransferase, an enzyme involved in biosynthesis of phosphatidyl ethanolamine lipids. No other mutants in this enzyme are known in plants.

In summary, we have with minimum resources already identified a significant number of random insertional knockouts. These results compare favorably with the output of more elaborate publicly funded approaches. The 60+ gene knockouts identified by Mu-TAIL sequencing are currently being mapped by the Missouri maize genome project.

**Probing cDNAs with Mu-TAIL products.** While sequencing is very efficient in identifying random knockouts, the goal of this proposal is to specifically target endosperm expressed genes. One way to target genes expressed in a specific organ is to use



**Figure 7.** Genomic DNA from W22 and 5 alleles of *vp13* digested with *Sst*I. The blot was probed with a cDNA isolated with the vp-5279 Mu-amplicon

Mu-TAIL products to probe a cDNA library. Effectively the cDNA library acts as a filter to select the subset of Mu insertions that are within PCR range of exons expressed in the appropriate tissue. As shown above, Mu-TAIL products are highly specific and contain little repetitive DNA. To test this idea, we amplified Mu-TAIL products from pooled genomic DNA (12 plants/pool) from each of four *viviparous13* (*vp13*) alleles (McCarty, 1995). We used these four probes to screen a wild type seedling leaf cDNA library (250,000 plaques/probe). Each of the probes identified a small number of unique cDNA's (15 total). Analysis of these candidate cDNA's is in still progress. Several cDNA's detect polymorphisms on genomic Southern blots consistent with transposon disruptions in the lines used to construct the Mu-TAIL probes (e.g. figure 7). Single pass sequencing of the cDNA clones identified two cDNAs that have sequence homology with sequenced Mu-TAIL products from the W22 parent indicating that these cDNAs also correspond to genes that carry authentic Mu insertions. These results indicate that probing cDNA libraries with Mu-TAIL products identifies genes with heritable Mu insertions.

## **6. Microarray directed transposon mutagenesis: Theory and practice.**

Microarrays of cDNA's can be probed much more efficiently than conventional libraries. Microarray hybridization is highly sensitive and can be used with very complex probes. Hence it will be feasible to use Mu-TAIL probes from pooled lines. A grid pooling scheme is particularly efficient because individual cDNA hits can be traced back to an individual line.

**Probe complexity.** Our sequencing results indicate that Mu-TAIL represents a high percentage of the Mu insertions in a genome. If we assume an average of 50 germinal copies per line, the Mu-TAIL products will include up to 100 different fragments. Because we inactivate Mu in lines we select for analysis, only germinal insertions are recovered and average copy number will not change in lines derived from an individual. About 80% of the germinal Mu insertions in an individual are shared with one or both parents. In our population roughly 30 copies came from the W22 parent and are shared by all lines. Of the remainder, about half will be unique insertions inherited from the previous generation and half (~10) will be new germinal transpositions. Hence, a probe from a pool of 24 lines from unrelated pedigrees would contain about 1000 different sequences  $((30 + (24 \times 20)) \times 2 = 1020)$ . This is a reasonable probe complexity for a microarray experiment. Moreover, we expect each probe to produce a small number of qualitative positive hybridization signals against a low background. In this way, a total of 48 probes prepared from rows and columns of a 24 X 24 grid can be used to analyze 576 mutants.

**Subtraction of parental insertions.** Microarrays are especially powerful for our application because of the built-in facility for subtraction. As noted above, most of the Mu insertions in a line are shared with the parents. Importantly, only the new germinal transpositions contribute to the forward mutation frequency. Because Uniform Mu is pedigreed we can subtract the non-informative parental insertions by using pooled Mu-TAIL products from the Mu parents as the reference probe for each row and column. Any cDNA's that hybridize to parental insertions will be labeled with both dyes and will fluoresce yellow. New insertions that are potentially linked to the selected mutation will fluoresce with a single dye (red or green).

**Hybridization groups v.s. complementation groups.** Our goal is to map Mu insertions on to the microarray from a collection of mutant lines large enough to include multiple alleles of many informative mutants. If we accept the Sheriden and Neuffer estimate of 300 genes conditioning a discernible seed phenotype as a single mutant, then a collection of 2000 mutations would contain, on average, 6 to 7 alleles of each gene. In general, allelism can be inferred from hybridization of respective TAIL-PCR products to a common cDNA clone; however, this is not definitive. For example, insertions in a pair of non-allelic but paralogous genes that share sequence homology may well detect the same cDNA. Conversely, it is possible to detect allelic mutations via hybridization to a cDNA that are not from the mutant gene. Hence allelic mutations can be identified for genes that are not actually represented in the array. In all likelihood, however, the detected cDNA could be used to clone the mutant gene.

To highlight this distinction, we define a “hybridization group” as a group of mutant lines that hybridize to the same or overlapping set of cDNA’s. Note that because each line contains multiple insertions, it is possible for a line to be assigned to more than one hybridization group. If we screen enough mutants, most lines that detect a cDNA in the microarray will be assigned to a hybridization group that contains more than one member. The mutants within each hybridization group will in turn be genetically tested for allelism to confirm gene assignments. Importantly, the depth of the cDNA array will determine the total number of hybridization groups we identify, but not the size of those groups.

**Expected results.** The total number of endosperm mutants we can expect to clone can be estimated in at least two ways. First from a forward genetics perspective, if the number of mutations screened approaches saturation, then the efficiency can be estimated as follows:

Number of mutants cloned =

$$[\text{Mu-TAIL Efficiency}] \times [\text{Microarray representation}] \times [\text{Total number of genes that have phenotypes}]$$

Where Mu-TAIL Efficiency is the fraction of all Mu insertions represented in the Mu-TAIL products; Microarray representation is the fraction of all endosperm-expressed genes represented in the array; and total number of genes that have phenotypes is the number of potential targets. Based on sequence analysis of Mu-TAIL products cloned from the W22 inbred we estimate that Mu-TAIL represents at least 90% of the Mu elements present from at least one border ( $27/30 = 0.9$ ). For this discussion we will use a more conservative estimate of 0.5. A complete statistical analysis of the number of genes represented in the 6,500 endosperm EST’s is not yet available. A preliminary analysis using ZmDB ([www.zmdb.iastate.edu](http://www.zmdb.iastate.edu)) indicates that endosperm EST’s include at least 1,900 different cDNA contigs (D.R. McCarty, unpublished). For the sake of argument, let us suppose this includes 30% of all endosperm expressed genes. Finally, an estimate for the total number of genes that have qualitative endosperm phenotypes (~300 genes) is derived from the near-saturation EMS mutagenesis studies of Neuffer and Sheridan (1980). Using these conservative estimates the expected number of endosperm mutants cloned will be about:  $0.5 \times 0.3 \times 300 = 45$  genes. If the more optimistic estimate for Mu-TAIL efficiency (0.9) is used, then roughly 80 genes will be cloned.

Alternatively, we can approach the problem from an empirical reverse genetics perspective:

Number of phenotypic insertions identified =

$$[\text{No. of lines}] \times [\text{cDNA hit rate}] \times [\text{Effective library size}] \times [\text{Frequency of knockout phenotypes}]$$

In our preliminary experiments, Mu-TAIL probes from four different *Mutator* lines detected an average of 4 different cDNA’s each in a 250,000 clone library. Hence the empirical hit rate for unique cDNA’s is  $\sim 1/60,000$ . The 6,500 endosperm EST’s represent a set of  $\sim 17,000$  random clones. While there are no firm estimates for the percentage of gene knockouts that have visible phenotypes in maize, anecdotal results from the Pioneer TUSC facility suggest that it may be as low as 10%. If we screen 2000 Mu lines, the expected number of mutants cloned is  $2000 \times (1/60000) \times 17,000 \times 0.1 = 57$ . A more optimistic estimate for discernable phenotypes (20%) raises the expectation to 114 mutants.

This theoretical treatment illustrates several key points:

1. With reasonable assumptions, the two models yield similar estimates in the range of 45 to 115 phenotypic gene knockouts detected using the existing cDNA array. However, there is much uncertainty. This project will yield much a firmer understanding of the efficiencies underlying forward and reverse genetics approaches to functional genomics of maize.
2. Our ability to detect subtle phenotypes limits both approaches. The TUSC population is highly heterogeneous and not conducive to quantitative phenotypic analysis, so perhaps we can do better than 10% with Uniform Mu.
3. The depth of coverage of the cDNA array is a key limiting factor in both models. Preliminary analysis of the existing endosperm EST’s indicates significant gaps, especially for genes expressed very early and in aleurone. We propose to enhance the EST coverage at these stages.

## 7. Experimental plan.

Our long-term goal is to saturate the genes involved in maize endosperm formation with transposon insertions. A comprehensive mutant hunt will be used to derive a set of plants that carry at least one informative Mu insertion in an endosperm-expressed gene. The germinal Mu insertion sites contained in these plants will be systematically extracted by Mu-TAIL PCR. By using pooled Mu-TAIL products to probe a large endosperm cDNA microarray we will identify candidate clones for each mutant. By using a 24X24 grid pooling scheme >1,700 mutations may be screened with less than 150 hybridizations (3 X 48). By screening a large number of mutants we identify independent alleles which facilitate confirmation of gene assignments and organize genetic complementation testing. While we will emphasize a forward genetics approach, our strategy will also identify random knockouts of endosperm expressed genes that are not associated with phenotypes. A time table for the project is outlined in Table 2 at the end of this section.

**7.1 High throughput mapping of Mu insertions on to an endosperm cDNA array.** Microarrays may be used to identify candidates for tagged mutants or to target any set of genes by insertional mutagenesis. In this case, we will target a set of 6,500 endosperm cDNA's that are currently being microarrayed by David Galbraith at U. of Arizona. In preliminary experiments, using a conventional library we have shown that Mu-TAIL PCR products can be used effectively as complex probes to identify cDNA's of genes that carry *Mu* insertions. On the basis of those results, we expect about one hit per line in the endosperm array. By preparing Mu-TAIL probes from lines that segregate visible endosperm phenotypes we ensure that each line represented in the probe carries at least one informative insertion in an endosperm expressed gene. By organizing the mutants in grid arrays and pooling probes along rows and columns, we are able to trace each cDNA hit back to an individual mutant line.

We will begin with a pilot experiment to establish conditions for probing arrays with Mu-TAIL products. Our sequencing project identified two insertions from the population that give strong BLASTN hits in the endosperm EST's. Because the location of those cDNA's in the array will be known, the Mu-TAIL products from these lines will serve as positive controls. Mu-TAIL products from W22 will provide the reference probe. Should any parental W22 insertion sites hybridize to the array, they will be shared by all Uniform Mu plants and will show up as yellow spots. Once conditions for the control probes are established, we will prepare probes from 24 different *dek* and *small kernel* mutants isolated from BC3 Uniform Mu. The progenitors of these lines will provide the reference probes. The Mu-TAIL probes will be prepared from individual and from pooled DNA samples. Our target pool size for efficient screening is 24 lines, because this would allow us to assemble mutants in 24 X 24 grids each containing 576 mutants. Hence the pilot experiment represents a typical row or column in a grid. We routinely make Mu-TAIL products from 12 plant pools with no apparent loss of diversity relative to individual plant samples. If necessary Mu-TAIL will be performed on 12 plant subpools then combined. We will prepare DNA from individual lines (see below) to allow flexible pooling strategies. Detected cDNA's will be used to probe Southern blots of corresponding mutant and progenitor lines to confirm the presence of an insertion.

Once satisfactory conditions are established, Mu-TAIL probes will be prepared from pooled DNA samples that represent grid arrays of mutant lines. A 24X24 array containing 576 mutants can be represented in rows and columns by 48 pooled DNA samples. Mu-TAIL PCR will be performed on each sample using six different arbitrary primers. The six reactions will be pooled for each sample. The PCR reactions will be done in 96 well format, i.e. 16 samples per plate, 3 plates per grid. A balanced reference probe for each of the row and column pools will be constructed by pooling DNA from the progenitor Uniform Mu parent of each mutant, and performing Mu-TAIL PCR as above. In this way, 48 hybridizations (96 with replication) containing paired mutant and progenitor probes labeled with Cy dyes, respectively, will be sufficient to map the insertions represented in the mutant collection on to a cDNA microarray. Insertions that are shared between a mutant line and its progenitor (parental Mu elements) will fluoresce yellow, whereas, transpositions events that are unique to the mutant will fluoresce red. We plan to do three 24X24 grids (1728 mutants; 144 probes)

from Uniform Mu lines and one smaller grid of mutants derived from other populations (see below). In the latter case, we will prepare the reference probes from sister lines that do not segregate the mutant.

**7.2 Resolving hybridization groups into complementation groups.** Independent mutant lines that detect the same cDNA(s) define a hybridization group. Mutant lines within each hybridization group will be tested for allelism. If ~300 genes give visible endosperm phenotypes (Neuffer and Sheridan, 1980), we estimate that an analysis of three 24X24 grids (1728 mutants; 144 probes) will identify multiple alleles of most phenotypic genes represented in the array. In other words, most assignments will be confirmed by multiple alleles.

To accommodate the expected numbers, complementation tests will be performed in two stages. In the first stage, pollen from a single mutant is used to pollinate ears of each of the other candidates. Six crosses, for example, would be made with a collection of seven candidate alleles. In practice, we will make triplicate crosses with three separate males. Ears are then scored at maturity for the presence of mutant (positive allelism test) seed. If all tests are positive, we are done. If one or more mutants give negative tests, the process will be repeated using those lines as pollen parents. If the initial tests are all negative, we will make all possible crosses among the seven alleles in the following growing season. For example, in our group of 7 mutants, 42 different crosses (including reciprocals) would be made. The data in the second growing season should sort out any complexity found in the first growing season (two or more duplicate but complementary genes, intracistronic {interallelic} complementation, etc.) Lethal mutants will be propagated as heterozygotes and require a larger number of crosses. In these cases we will make pollen mixtures from three plants grown from wildtype seed, 2/3 of which will be heterozygotes. Pooled pollen is used to pollinate the plants producing the pollen as well as ears from three plants of each mutant in question. The sib pollinations will identify heterozygotes within the group of three. The probability that all three plants are homozygous and lack the mutant is 1 in 27 or ~4%. Likewise, the probability that all three plants used as female in the crosses are homozygotes is ~4%. Assuming that mutant seed are seen in at least one of the sib-pollinations, the probability that the cross will generate mutant seed if the mutants are allelic is ~96%. In rare cases where only one of the three plants used as male is a heterozygote, the frequency of mutant seed in the cross is 1 in 12 on an allelic, heterozygous female. This is sufficient since in a sample of only 64 test cross kernels, the probability of at least one mutant seed is > 99.5%.

The use of the hybridization data as a guide makes genetic allelism testing feasible. For example, without any insight into gene identity, the number of single, non-reciprocal crosses needed for allelism in a group of 2000 mutants is  $(2000)(1999)/2$  or ~ 2 million. This does not take into account the complexities of working with lethals. With the molecular data, the number of crosses is equal to the number of mutants placed in hybridization groups minus the total number of groups identified. The largest number of crosses required would occur if, unrealistically, all 2000 mutants belonged to the same hybridization group. In that case, placement of pollen from one family onto each of the other 1999 would suffice (1999 crosses) for the first round.

**7.3 Assembly of augmented endosperm cDNA microarrays.** The overall efficiency of our approach is chiefly constrained by the size and coverage of the microarrayed cDNA libraries. The NSF funded public maize EST collection includes about 6,500 different endosperm cDNA's derived from 10 DAP developing kernels. The annotation of the endosperm collection is not completed, however, the overall level of redundancy (~3-fold) suggests as many as 2,000 unique genes may be represented. Hence the existing EST's are an excellent starting point for microarray construction and this work is already well under way in the Larkins laboratory in collaboration with David Galbraith. However, a cursory analysis of the 6,500 EST's suggests the coverage of developing endosperm is incomplete. For example, the relatively abundant *sh2* and *sh1* transcripts are represented >10 times each; whereas, other genes such as *vp1* (aleurone specific) and *bt2* (surprising because the mRNA is abundant) are not represented at all. Other important genes expected to be active very early in development such as the SET domain and polycomb-like homologs of the arabidopsis *fis* and *medea* genes are also absent (Messing, unpublished).

To obtain deeper coverage we will construct two additional cDNA microarrays that target very early stage endosperm and aleurone development, respectively. The Messing lab will construct the early endosperm library (4-6 DAP). The Becraft lab will construct a *dek1* subtracted library that will be enriched in aleurone specific cDNA's. We will determine 10,000 ESTs each from the Rutgers and Iowa State libraries, adding 20,000 ESTs to the current GenBank database. We will coordinate merging these data into maize EST databases with curators of ZmDB (Iowa State) and MaizeDB (Missouri). An aleurone cDNA library will be generated by subtracting transcripts from wild type endosperm against *dek1* mutant. The *dek1* mutant completely eliminates aleurone indicating that it acts at an early step in aleurone development. The subtracted library will therefore be enriched for aleurone specific transcripts. A range of developmental stages from 7 to 22 days after pollination will be included in the RNA source tissue. This approach has several advantages over making a standard cDNA library from aleurone tissue. It would be more likely to include developmentally important genes because it covers a range from fully differentiated aleurone to early stages when decisions governing aleurone differentiation are likely to occur. It would eliminate much redundancy because housekeeping genes would be subtracted out. It would increase the probability of including rare transcripts that would have a low chance of being sampled by standard techniques. The differential library will be generated using suppression subtractive hybridization (Diatchenko et al., 1996). This is a PCR based technique that normalizes for rare messages and then preferentially amplifies differentially expressed cDNAs. The technique is currently being employed in the Becraft lab for other research. One drawback to this method is that only short fragments of cDNA are generated. To generate a full length library, a standard wild type cDNA library will be constructed from the same developmental stages. A biotinylated probe will be generated from the differentially expressed library and used for a mass selection of homologous plasmid clones from a wild type cDNA library by binding to streptavidin beads. A similar strategy will be used to enrich endosperm cDNA's expressed in a 4 to 6 day post-pollination library by subtracting cDNA from unfertilized ovules of same age.

**7.4 Comprehensive transposon mutagenesis of endosperm genes.** While we may take advantage of Mu tagged mutants from any source, we will emphasize screening of the Uniform-Mu population because it is uniquely designed for seed mutagenesis (see Preliminary Results). Importantly, pre-existing endosperm mutations are purged from the population each generation in order to ensure that new events represent independent mutations. The frequency of new seed mutations is monitored through each backcross cycle to maintain a forward mutation frequency of ~10% in the pedigrees carried forward

We expect to isolate up to 2,000 independent visible endosperm mutations from a collection of 20,000 self-pollinated Uniform Mu (BC5) ears generated in first two years of the project. The Uniform Mu self's will be grown under intensive management at the UF Horticulture Research Unit in Gainesville over four growing seasons (2 crops per year). About 50,000 mutagenic BC5 seed were generated during the Fall 1999 season in Gainesville. Half of the seed carry the *bz1-mum9* marker for somatic activity. We already have sufficient seed to meet the goals of this project. If necessary more BC5 and BC6 families will be created and grown in subsequent years. The initial screen will be conducted for qualitative endosperm phenotypes. These will include all classical mutants (viviparous, shrunken, small kernel), a wide variety of defective kernel types (collapsed, etched, empty pericarp, etc), texture phenotypes (opaque, floury) and aleurone differentiation and pattern mutants. The selfed ears will be stored unshelled to facilitate future screens.

To obtain Mu-inactive derivatives of each mutant line, up to 10 stable *bz1* mutant seed will be selected from each segregating ear (e.g. Figure 3). Advanced generation Uniform-Mu self's typically segregate ~3:1 or rarely 15:1, spotted : stable bronze within the *bz1* mutant class indicating presence of one or occasionally two *MuDR* elements. In general, 2/3 of the *bz1* seed will be heterozygous for a new mutation allowing recovery of homozygous lethals. In rare cases where the new mutant is linked to *bz1* this may not be possible. These seedlings will be grown in a green house and pooled tissue samples will be taken from each family for DNA preparation using a Quiagen 96 well format plant kit, (Quiagen, Inc.).

Our near term goal is build a mutant collection large enough to include independent alleles of most tagged genes. Conceivably, 2,000 independent mutations will approach statistical saturation for qualitative phenotypes, i.e. an average of 5 or more alleles of an estimated ~300 genes (Neuffer and Sheridan, 1980). In practice, however, the number of genes identifiable by phenotype will depend on factors such as the sensitivity of the screen, the mutagen used and inbred background. We will estimate the coverage of Uniform Mu by tracking the mutation frequencies of classical qualitative mutants (*vp1*, *sh2*, *bt2*, *bt1* and anthocyanin mutants). If saturation is not reached (a probable outcome), we will continue screening Uniform-Mu self's (5,000 per year) in the remaining years of the project. The Koch lab will supervise digital photography and cataloging of the mutant ears. The entire 20,000 ear collection will be preserved to allow future quantitative screens of both seed and plant phenotypes.

Once the primary screen is complete, the Koch and Hannah labs will undertake a pilot screen of the remaining ears for quantitative *small kernel* phenotypes. The prominent *small kernel* class is especially interesting because this phenotype integrates many processes of central importance to cellular growth and metabolism as well as regulation of the grain filling pathways. Our preliminary studies indicate that the threshold differential for visual detection of small phenotypes segregating on a uniform ear is about 35%. Smaller differentials can be detected by plotting the kernel weight profile of 96 kernel weights from an ear. A quantitative screen will be conducted on 1,000 ears selected at random and shelled. Individual kernel weights will be determined by placing a 96 well micro titer plate on an electronic recording balance and adding single kernels to each well, the successive weight changes will be recorded to a PC. Data collection will be performed by undergraduate trainees.

Because our cloning strategy is in principle applicable to any Mu tagged gene, we are also able to supplement our collection with endosperm mutants gleaned from the Rescue Mu (Stanford maize consortium) and MTM (Cold Spring Harbor Lab) populations. The Larkins lab will screen 2000 ears from the Rescue Mu population for endosperm mutations. For our purposes it will not matter whether Rescue Mu or another Mu element (most likely) caused the mutant. Phil Becraft has already screened the MTM population for aleurone phenotypes. While the design of these populations is not ideal for selection of independent endosperm mutants, any alleles will be independent of Uniform Mu and hence useful for confirmation of candidate clones. Finally, we will include 63 Mu induced endosperm mutants from Don Robertson's collection (Scanlon et al., 1994) available from the Maize Stock Center. These partially characterized mutants provide an important set of reference mutations that will nucleate hybridization groups.

**7.5 Customized Uniform-Mu populations.** A strength of the Uniform-Mu system is that derivative populations can be tailored for specific mutant screens. For example, a long-term goal of the Messing lab will be to develop a screen for mutants that quantitatively enhance or reduce expression of a 27 kd zein promoter-GFP reporter gene in endosperm. This will be set up by first backcrossing the GFP reporter gene into our recurrent W22 inbred parent. Once the W22 GFP line is established, that line will be outcrossed to our advanced generation Uniform-Mu parental lines. The resulting plants will be self-pollinated in order to derive plants that are homozygous for the GFP transgene and *bz1-mum9* markers. Because both parents in the cross share a W22 background the progeny will be homogenous. The resulting GFP marked Uniform-Mu lines will be back crossed to the W22 GFP parent to create a marked Uniform-Mu population that can be screened for variation in GFP expression in endosperm. In a similar fashion, the Hannah lab will construct populations for two additional screens: 1) suppressors and enhancers of the *sh2-l* mutation a leaky RNA splicing defect in the *sh2* gene and 2) mutants that suppress transgene induced co-suppression the *sh2* gene. The latter phenotype was induced by sense expression of a *Sh2* transgene. The *sh2-i* allele (Lal et al 1999.) contains a G to A mutation at the 3' end of intron 2. While destroying the invariant AG dinucleotide terminus of conventional nuclear introns, this mutant acceptor site is still used in ~10% of the splicing events. This gives rise to an intermediate phenotype. Mutations increasing or decreasing this unusual form of splicing would be seen as plump or severe shrunken variants, respectively, of the original leaky phenotype. Because at least five generations are needed to create these populations, significant screens will not be conducted until the fifth year of the project.

**Future prospects.** By the end of this project our Uniform Mu population will be in at least BC9 and essentially pure W22 inbred. Any residual linkage drag from the *MuDR*'s and *bz1-mum9* will have been largely eliminated. Linkage drag by *MuDR*'s will be self-correcting because they will transpose into W22 chromosome regions with some frequency. If necessary recombination around *bz1-mum9* can be accelerated by marker assisted selected selection against linked RFLP markers. A mutagenic inbred (a slight oxymoron) will be valuable for quantitative genetics studies. For example, one could derive several sublines, propagate them separately for several generations by selfing. Each line may be crossed to a series of standard inbreds and their combining ability compared with that of the W22 parental line. The loci underlying quantitative variation in these derivatives will be tagged with Mutator. The results of our pilot sequencing experiment suggest that the ~30 resident Mu elements in W22 contribute significantly to the genetic load of this inbred.

## 7.6 Bioinformatics and functional analysis of endosperm mutants

We expect to clone 50 to 100 endosperm mutants using the current cDNA array. The participating labs have diverse skill sets and complementary interests covering the essential processes governing endosperm formation. We expect to identify many genes of definite interest to individual labs, many other genes will have uncertain relevance pending more data (see below), still other genes will fit everyone's definition of boring. cDNA's identified by our screen will be fully sequenced at the Rutgers facility. Labs that pursue a particular gene will complete the molecular characterization.

While specific outcomes are hard to predict, we can outline some general expectations:

- 1) Because the *small kernel* phenotype integrates basic processes of growth and grain-filling, this abundant class is very likely to partition into genes of interest to Larkins (endoreduplication, regulation of zein/grain filling pathways), Hannah (regulation of starch/grain filling pathways) and Koch (integration of sugar transport, metabolism and signaling).
- 2) The defective kernel class will be further classified according to whether the principle cell types are present (e.g. aleurone, starch-filled cells). For example, many *collapsed endosperm* mutants (Neuffer and Sheridan, 1980) differentiate normal cell types and enlarge to near normal size before collapsing during desiccation. These are candidates for a defect in integrated regulation of grain-filling pathways. A homology to a transcription factor or other regulatory molecule would strongly reinforce that hypothesis. Expression profiling experiments based on the microarray will be a powerful approach to analyzing such mutants. Genes of this type will be pursued jointly by the Larkins and Hannah labs. Many other *dek*'s (e.g. *etched*, *dappled*, *rough*, *dek1*) disrupt differentiation and organization of aleurone, and will likely include signaling genes of interest to McCarty and Becraft. These will be profitably analyzed using the *dek1* subtracted cDNA array. We will perform two additional genetic screens for transmission and gene dosage effects to gain still more insight into the complex *dek* class (see below).
- 3) We expect to detect many knockouts in cDNA's that have interesting homologies but no discernable phenotype. In many cases, bioinformatic analysis will suggest a basis for functional redundancy (i.e. a gene family). If expression of related genes is detected in endosperm, then targeted knockouts of the other gene family members using the MTM facility may be warranted.

The key first step in this process will be a thorough bioinformatics analysis of the identified genes. This will be coordinated by the Rutgers group.

**Bioinformatics.** The bioinformatics portion of this project will focus on the integration of phenotypic analysis with predicted functions of the tagged genes. A major advantage of this project is that hybridization groups add new information beyond sequence based homologies that can be extracted from ESTs alone. We will develop a relational database that will link hybridization groups with: 1) their defining cDNA sequence, 2) known sequence homologies, 3) the digital images of the mutant phenotypes of the members, and 4) genetic data defining complementation groups within a hybridization group. Linking this information will allow researchers throughout the community to quickly identify mutations that have multiple alleles and the variations seen in their phenotypes as well

as identifying incidental knockouts that potentially have no visible seed phenotype (based on differing mutant phenotypes and full complementation).

We will use an integrational approach to linking databases to facilitate design of cDNA arrays and analysis of tagged DNA sequences. For instance, GenBank has 43,137 maize ESTs, of which 43,023 are represented in the MaizeDB. While GenBank has some information regarding the inbred line used and the tissue that was used to isolate the mRNA, the information is not easily accessible from a computational point of view because all maize ESTs are not explicitly segregated into separate clusters. The MaizeDB on the other hand has grouped ESTs into twenty different cDNA libraries. Since these libraries have been generated from a single mRNA preparation, this information can be readily accessed in sequence searches. Furthermore, MaizeDB has the map information of 714 cDNAs and a file structure that permits access to additional layers of information including publications. Although MaizeDB is linked to GenBank through accession numbers, most DNA comparisons would be made with GenBank therefore we will also calculate the reverse links.

Linking to other online databases allows us to leverage still more information for each gene. For example, if a newly sequenced gene belongs to a gene family, the literature (e.g. MEDLINE abstracts and online full-text journals) associated with the other genes can be integrated into our sequenced tagged DNA database. Summarization techniques include the identification of keywords and key phrases shared by a document or set of documents (Frank et al., 1999). In addition, the motif analysis will provide descriptive labels for specific regions of the new gene. Annotation of this depth will provide clues to specific functional roles, and will suggest concrete experiments for verification. Sequence similarity can be swiftly calculated using BLAST and sequence alignment. Structural similarity requires more time, but is a well-established technique. Information retrieval models can be used to define similarity for literature, based on word co-occurrence and shared key phrases. Relationships between data types are explicitly represented in databases such as Entrez. We propose to use these relationships within a computational framework, building on their current use in an interactive environment. By following links from genes to the literature, to structures, perhaps back to other genes implicated in the same cellular processes, we can greatly increase the usefulness of the genomic annotation. Significantly for this project, we can use these annotations to select the genes to put on the expression array.

Another important aspect will be to analyze the hybridization groups in relation to the 20 existing and the two new EST libraries. We will be able to categorize groups that are present in more than one of the maize EST libraries or represent functionally related genes (e.g. transcription factors, transporters, biosynthetic enzymes). These classifications will be helpful in extending our functional and genetic analysis of the cloned mutants. As outlined above, hybridization group assignments can be used to infer relationships between non-allelic insertions within our population, such as may occur in members of a gene family. As more family members are identified, lines containing these insertions become useful tools for potential double and triple mutant experiments.

A key bioinformatics challenge is suggested by the fact that many putative gene sequences (75-80%) identified by the Florida (see Preliminary Results) and NSF/Rutgers (Dooner, H. pers. comm.) projects have no homology to known genes. In part, this may reflect limitations of gene prediction based on homology. In order to infer function for a greater number of maize genes, we will use motif-based tools. Techniques such as Emotif, Identify, and Ematrix (Nevill-Manning et al., 1997; Nevill-Manning et al., 1998; Wu et al., 1999a and 1999b) are capable of assigning a protein to a functional family based on a stretch of ten to thirty amino acids, even in the absence of overall homology. These techniques compare the unidentified sequence with pre-identified families of proteins. Such intensive analysis will initially be focused on unknown cDNAs defining hybridization groups and will be expanded to novel cDNAs identified from the ESTs sequenced within this project.

**Functional analysis.** The bioinformatics analysis will guide the experimental analysis of individual gene functions. As noted above other genome resources will play a major role. If the gene belongs to a gene family, it may be desirable to create targeted knockouts of other partially redundant family members using the NSF funded MTM facility. If the mutated gene encodes a transcription factor or

other regulator molecule, then using the cDNA microarray to determine an expression profile of the mutant endosperm at an appropriate stage is a powerful approach to identifying genes in the downstream pathway.

Importantly, a common feature of these approaches is identification of other endosperm genes that function in the same or related processes. Hence each mutant serves as a separate entry point in to the complex network of endosperm genes. As that knowledge base grows the incidental endosperm knockouts identified in our population will become increasingly important as a source new genetic tools for analyzing these gene networks. We expect to identify several hundred such knockouts, representing a significant fraction of the endosperm transcriptome. This resource will complement directed knockouts created by the NSF funded MTM facility. By sharing that knowledge base within the endosperm community we will begin to understand the network relationships and integration of these pathways. As various groups converge on these relationships, science driven collaborations will arise naturally within the community.

**Genetic analysis.** We expect that the combined sequence, expression and bioinformatic analyses will not solve the function of every cloned mutant. To gain more information, we will put the *dek* and *small kernel* mutants through a series of secondary genetic screens. The Messing lab will screen this group for mutants that have segregation ratios skewed toward 1:1 rather than the expected 3:1. These would include mutants that act in the female gametophyte including genes involved in gene imprinting. The *gametophytic small kernel* mutant isolated from our BC2 Uniform Mu material (D. R. McCarty, K. E. Koch, unpublished results) is a prototype for this phenotype. Candidate mutants identified in the initial screen will be tested for differential transmission through the female and male gametophytes by making reciprocal crosses with W22.

A second way to sort *dek* mutations is to ask whether they have significant dosage effects in the triploid endosperm. Mutants that affect limiting processes may often show significant differences in seed size of kernels that have 3, 2 or 1 doses of the wild type allele. For example, the *sh2* mutant which controls the rate limiting step in starch biosynthesis has a discernible dosage effect (Hannah, 1997). This may be detected as significant skewing or broadening of the profile of individual seed weights within the normal seed class. We will conduct a pilot of this approach by screening kernel weight profiles of selected *dek* mutants (up to 50 lines). If evidence of skewing is detected, then seed will be grown from the high and low weight extremes and the plants selfed. If the skewing is due to a dosage effect then the low weight kernels will be enriched in heterozygotes and the high weight seed depleted in hets. As above the kernel weight data will be taken by undergraduates

**Table 2. Project Time Table.**

Year 1.	Grow and screen 10,000 Uniform Mu self's Complete 24 line pilot microarray experiment Complete construction of subtractive EST libraries Assembly and DNA sampling of first mutant grid Bioinformatics analysis of existing cDNA's Begin backcrossing of zein-GFP, sh2-I, Sh2 transgenic lines into Uniform Mu	McCarty, Hannah, Koch McCarty, Larkins Messing, Becraft McCarty, Koch Messing Messing, Hannah
Year 2.	Grow and screen 10,000 Uniform Mu self's Complete microarray screening of first mutant grid Assemble, sample 2 <sup>nd</sup> and 3 <sup>rd</sup> mutant grids Complete submission of 20,000 EST's Begin full length sequencing of hit cDNA's Digital imaging of mutant collection Begin quantitative small kernel screen	McCarty, Hannah, Koch McCarty, Larkins McCarty, Koch Messing, Becraft Messing Koch Koch, Hannah

Year 3.	Complete microarray analysis of grids 2 and 3 Begin complementation testing Bioinformatics analysis of full-cDNA sequences Screen for dosage sensitive deks Begin screen for transmission effect mutants Construction of aleurone and early stage microarrays	McCarty, Koch, Larkins Hannah Messing McCarty Messing Becraft, Messing, Larkins
Year 4.	Complementation testing Screen aleurone and early stage microarrays with selected mutants Custom Uniform Mu population construction Molecular and functional analysis of selected mutants Integrate phenotype, sequence, expression database information	Hannah Becraft, Messing, McCarty, Larkins Hannah, Messing Everybody  Messing, Koch
Year 5.	Completion of complementation testing Molecular and functional analysis of selected mutants Expression profiling of selected mutants Screening of custom Uniform Mu populations	Hannah Everybody  Everybody Hannah, Messing

**8. Information Management - Bioinformatics Infrastructure** The Rutgers Research Collaboratory for Structural Bioinformatics ([www.rcsb.org](http://www.rcsb.org)) has just won NSF support for running the Protein Data Bank (PDB), previously at Brookhaven National Laboratory. The Principal Investigator for this project, Dr. Helen Berman, is a member of the Waksman Institute and the Rutgers Department of Chemistry. Dr. Casimir Kulikowski, a member of the National Institute of Medicine and an expert in artificial intelligence, also is a member of the Waksman Institute as well as the Rutgers Department of Computer Sciences. In addition, he is the head of the new bioinformatics initiative at Rutgers and has started a faculty recruitment effort in bioinformatics. The first appointment in this area was made last year when Dr. Nevill-Manning from Doug Brutlag's group at Stanford University joined the Rutgers faculty. Dr. Nevill-Manning will be responsible for the supervision of graduate students who participate in the data interpretation of this project. Currently, the Molecular Biosciences Graduate Program provides a track in computational molecular biology and students can get their doctoral degree by a committee consisting of biologists and computer scientists.

**9. Integration of research and education, diversity and out-reach programs.**

**Undergraduate intern program.** The field genetics, mutant screens and lab research activities of this project offer clear opportunities for undergraduate assistants. The Plant Molecular and Cellular Biology (PMCB) Program at the University of Florida has a Research Internship Program for juniors who have a keen interest in pursuing a PhD or MS degree in the plant molecular sciences. Student interns work closely with PMCB faculty members on projects that teach experimental rationale and design as well as give hands-on experience with a wide array of molecular, genetic, and biochemical techniques in plant biology. These full-time internships carry a stipend of ~\$3,600 (\$360 per week) for a Fall, Spring Semester or for the two Summer terms. Students pursuing B.S. degrees in plant physiology, biochemistry, molecular biology/genetics, microbiology, cell biology, or related disciplines are competitive in the selection process.

**Graduate training.** At Florida the graduate students supported by this project will receive degrees in the PMCB program. The interdepartmental PMCB involves 20 faculty from 8 departments. Key PMCB activities include journal clubs, a seminar series and an annual workshop in which students present their research. The UF Genetics Institute and the Interdisciplinary Center for Biotechnology Research (ICBR) provide specialized technical training and campus wide research interactions. The entire endosperm group will form a workshop at the annual Maize Genetics Conference.

The bioinformatics effort at Rutgers will involve two types of activities, the development of new computational tools and the analysis of data. Both problems provide ideal graduate training opportunities. Data mining results and new software will be reviewed by Dr. Nevill-Manning and transferred to other sites. Students from the computer and biological sciences will focus on problems related to plant gene prediction. In addition, the strength of the structural biology program at Rutgers will provide these students with opportunities, through local retreats, courses, and lectures, to learn about related fields/topics.

**High school out-reach.** It is increasingly apparent that we who practice genetics and genome science have an obligation to address a wider audience of diverse people and deal with the popular misunderstandings of genomics. Our high school out-reach program is an informal mechanism to engage non-traditional students and young people at an early age. As a case in point, a student from a local rural high school has been involved in DNA sequencing projects in the Hannah lab for the past three years, and he has incorporated his results into a series of science fair projects. Because of religious views prevalent in his community, and in fact held by himself and his family, work with DNA was not viewed favorably. Consequently, his science projects initially received harsh criticism from the local judges and were not advanced to the next level of competition. With perseverance the student was able to enlighten the local judges on the value of transgenic organisms to agriculture, and his projects were eventually allowed to go forward. Subsequently, he went on to win the state competition twice and has gone to the national competition both years. He has since taken it upon himself to convince many of the leaders of his community of the value of genetics to agriculture without offending local religious norms. He is now a state officer in Future Farmers of America, and uses recombinant DNA and transgenic plant examples in his presentation to FFA chapters around the state. It has simply amazing how helping and educating one high school student could have so much impact!

**Walking Scholars Program.** The resources developed through this NSF Plant Genome Project will be of immense benefit to other investigators studying maize and other plants. Accordingly, we have established the Walking Scholars Program. Through this program, investigators worldwide are invited to peruse the *Mu* lines and examine any maize tissue or organ of their interest. Availability of this program will be posted on the maize Web site, [http://teosinte.agron.missouri.edu/more\\_info.html](http://teosinte.agron.missouri.edu/more_info.html), the UF PMCB Web site <http://www.ifas.ufl.edu/~pmcb/index.htm> as well as through other conventional informational outlets. Scholars are free to walk the fields and examine and tag any plants of interest to them. Scholars can also visit the new UF/Fifield Hall seed room to examine mature ears and seeds. To the extent possible, ears will be left unshelled to provide easier scoring of mutants. During visits, walking scholars are highly encouraged to present a seminar on their research endeavors. Furthermore, visitors are asked to interact with UF/PMCB graduate students and post-doctoral associates. Ideally, this interaction occurs in the field or seed room with the maize materials in hand. (This program was inspired by a visit to the University of Illinois Agronomy Farm during the 1975 Maize Genetics and Maize Breeding Conference. During this session, Barbara McClintock and Charles Burnham led an impromptu tour through the maize field, giving the early history of each mutant. This provided marvelous, unique insight to those students lucky enough to have been in the group.) Mutants of interest will be furnished to each Walking Scholar. In the case of tagged plants, selected plants will be self-pollinated or crossed to wild type plants by UF/PMCB staff. Remnant seed of all mutants will be maintained within the program. In return for furnished seed, Walking Scholars will be asked to acknowledge support from NSF, through listing of this particular grant number, in all subsequent publications. Commercial usage of any mutants will follow standard policies as addressed in a separate section. Our long-term goal is to furnish the resulting mutants to the Maize Genetics Stock Center maintained at the University of Illinois.

## References.

- Becraft, P., P. S. Stinard, D. R. McCarty. 1996. CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal development. *Science* 273:1406-1409.
- Birchler J., and J. R. Hart. 1987. Interaction of endosperm size factors in maize. *Genetics* 117:309-317.
- Birchler JA. 1993. Dosage analysis of maize endosperm development. *Annu Rev Genet* 27:181-204
- Chaudhuri S., J. Messing, 1994. Allele-specific parental imprinting of *dzr1*, a posttranscriptional regulator of zein accumulation. *Proc Natl Acad Sci U S A* 91: 4867-4871
- Chaudhury A. M., L. Ming, C. Miller, S. Craig, E. S. Dennis & W. J. Peacock. 1997. Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 94: 4223-4228
- Chomet, P. S. 1994. Transposon tagging with *Mutator*. In. *The Maize Handbook*. Ed. M. Freeling and V. Walbot. Springer, New York 243-249.
- Clark, J., W. F. Sheridan. 1991. Isolation and characterization of 51 embryo specific mutations of maize. *Plant Cell*. 3:935-951.
- Coleman C. E., A. M. Clore, J. P. Ranch, R. Higgins, M. A. Lopes & B. A. Larkins. 1997. Expression of a mutant alpha-zein creates the *floury2* phenotype in transgenic maize. *Proc Natl Acad Sci U S A* 94: 7094-7097
- Coleman C. E., M. A. Lopes, J. W. Gillikin, R. S. Boston & B. A. Larkins. 1995. A defective signal peptide in the maize high-lysine mutant *floury 2*. *Proc Natl Acad Sci U S A* 92: 6828-6831
- Cresse AD, Hulbert SH, Brown WE, Lucas JR, Bennetzen JL . 1995. Mu1-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* 140(1):315-324
- Dannenhofer J. M., D. E. Bostwick, E. Or & B. A. Larkins. 1995. *opaque-15*, a maize mutation with properties of a defective *opaque-2* modifier. *Proc Natl Acad Sci U S A* 92: 1931-1935
- Dellaporta, S. 1994. Plant DNA miniprep and microprep, Versions 2.1-2.3. In *The Maize Handbook*. Ed, M. Freeling, V. Walbot. Springer Verlag, New York pp. 522-25.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., and Siebert, P. D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* 93, 6025-6030.
- Dooner H. K., T. P. Robbins & R. A. Jorgensen. 1991. Genetic and developmental control of anthocyanin biosynthesis. *Ann. Rev. Genetic* 25: 173-199
- Dooner H.K. (1985) *Viviparous-1* mutation in maize conditions pleiotropic enzyme deficiencies in the aleurone. *Plant Physiol.* 77: 486-488.
- Frey, M., C. Stettner and A. Gierl. 1998. A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant J.* 13:717-722.
- Gavazzi G., S. Dolfini, D. Allegra, P. Castiglioni, G. Todesco & M. Hoxha. 1997. *Dap* (Defective aleurone pigmentation) mutations affect maize aleurone development. *Mol Gen Genet* 256: 223-230
- Geetha K. B., C. R. Lending, M. A. Lopes, J. C. Wallace & B. A. Larkins. 1991. *opaque-2* modifiers increase gamma-zein synthesis and alter its spatial distribution in maize endosperm. *Plant Cell* 3:1207-1219
- Giroux M. J., J. Shaw, G. Barry, B. G. Cobb, T. Greene, T. Okita & L. C. Hannah. 1996. A single mutation that increases maize seed weight. *Proc Natl Acad Sci U S A* 93: 5824-5829
- Giroux, M.J., C. Boyer, G. Feix, and L. C. Hannah. 1994. Coordinated transcriptional regulation of storage product genes in the maize endosperm. *Plant Physiol.* 106:713-722.
- Goff, S. A., K.C. Cone, V. L. Chandler. 1992. Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional interaction between two classes of regulatory proteins. *Genes Dev* 6: 864-875

- Grafi G., R. J. Burnett, T. Helentjaris, B. A. Larkins, J. A. DeCaprio, W. R. Sellers & W. G. Kaelin. 1996. A maize cDNA encoding a member of the retinoblastoma protein family: involvement in endoreduplication. *Proc Natl Acad Sci U S A* 93: 8962-8967
- Grossniklaus, U., Vielle-Calzada, J.-P., Hoepfner, M. A. & Gagliano, W. B. 1998. *Science* 280, 446-450.
- Guo M, Davis D, Birchler JA. 1996. Dosage effects on gene expression in a maize ploidy series. *Genetics* 142:1349-55
- Guo, M., Birchler, J.A. 1994. Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science* 266:1999-2002
- Hannah, L. Curtis. 1997. Starch synthesis in the maize endosperm. In: "Advances in Cellular and Molecular Biology of Plants", Vol. 4. B. A. Larkins and I. K. Vasil (eds.). Cellular and Molecular Biology of Plant Seed Development. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 375-405.
- Hattori, T., V. Vasil, L. C. Hannah, D. R. McCarty and I. K. Vasil. 1992. The Viviparous-1 gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes and Development* 6:609-618.
- Hoecker, U., I. K. Vasil, D. R. McCarty. 1999. Signaling from the embryo conditions Vp1-mediated repression of  $\alpha$ -amylase genes in the aleurone of developing maize seeds. **Plant Journal** (in press)
- Hoecker, U., Vasil, I. K. and D. R. McCarty. 1995. Integrated control of seed maturation and germination programs by activator and repressor functions of *Viviparous1* of maize. *Genes and Development* 9:2459-69.
- Hueros G, Varotto S, Salamini F, Thompson RD. 1995. Molecular characterization of BET1, a gene expressed in the endosperm transfer cells of maize. *Plant Cell* 7:747-57.
- Kao, C-Y, S. M. Cocciolone, I. K. Vasil and D. R. McCarty. 1996. Localization and interaction of the cis-acting elements for abscisic acid, VIVIPAROUS1 and light activation of the C1 gene of maize. *The Plant Cell* 8:1171-1179.
- Kermicle JL, Alleman M. 1990. Gametic imprinting in maize in relation to the angiosperm life cycle. *Dev Suppl* 1990:9-14
- Kiyosue, T., N. Ohad, R. Yadegari, M. Hannon, J. Dinneny, D. Wells, A. Katz, L. Margossian, J. J. Harada, R. B. Goldberg & R. L. Fischer. 1999. Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proc Natl Acad Sci U S A* 96: 4186-4191
- Kodrzycki R, Boston RS, Larkins BA. 1989. The opaque-2 mutation of maize differentially reduces zein gene transcription. *Plant Cell* 1:105-14
- Koch, K. 1997. Molecular crosstalk and the regulation of C- and N-responsive genes. In: C. Foyer and P. Quick (ed.s) A molecular approach to primary metabolism in plants. Taylor and Francis, Inc., London, pp. 105-124.
- Lending C.R. & B. A. Larkins. Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell* 1: 1011-1023
- Lin, B.-Y. 1982. Association of endosperm reduction with parental imprinting in maize. *Genetics*. 100: 474-486.
- Lin, B.-Y. 1984. Ploidy barrier to endosperm development in maize. *Genetics*. 107: 103-115.
- Lisitsyn N., N. Lisitsyn and M. Wigler. 1993. Cloning the differences between two complex genomes. *Science*. 259:946-950.
- Liu C. M. & D. W. Meinke. 1998. The titan mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. *Plant J* 16: 21-31
- Liu YG, Mitsukawa N, Oosumi T, Whittier RF. 1995 Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* 8(3): 457-463.

- Lohmer S., M. Maddaloni, M. Motto, N. Di Fonzo, H. Hartings, F. Salamini & R. D. Thompson. 1991. The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. *EMBO J* 10: 617-624
- Lotan T., M. Ohto, K. M. Yee, M. A. West, R. Lo, R. W. Kwong, K. Yamagishi, R. L. Fischer, R. B. Goldberg & J. J. Harada. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93: 1195-1205
- Luerssen H., V. Kirik, P. Herrmann & S. Misera. 1998. FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J* 15: 755-764
- Lund G, Ciceri P, Viotti A. 1995. Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of Zea mays L. *Plant J* 8(4):571-81
- Lund G, Messing J, Viotti A. 1995. Endosperm-specific demethylation and activation of specific alleles of alpha-tubulin genes of Zea mays L. *Mol Gen Genet* 246(6):716-22
- Maddaloni M, Donini G, Balconi C, Rizzi E, Gallusci P, Forlani F, Lohmer S, Thompson R, Salamini F, Motto M. 1996. The transcriptional activator Opaque-2 controls the expression of a cytosolic form of pyruvate orthophosphate dikinase-1 in maize endosperms. *Mol Gen Genet* 250(5):647-54
- Mayer U., U. Herzog, F. Berger, D. Inze & G. Jurgens. 1999. Mutations in the PILZ group genes disrupt the microtubule cytoskeleton and uncouple cell cycle progression from cell division in Arabidopsis embryo and endosperm. *Eur J Cell Biol* 78: 100-108
- McCarty, D. R. 1995. Genetic control and integration of maturation and germination pathways in seed development. *Annual Reviews of Plant Physiology and Molecular Biology* 46: 71-93.
- McCarty, D. R., T. Hattori, C. B. Carson, Vimla Vasil, M. Lazar and I. K. Vasil. 1991. The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66:895-905.
- Mena M., J. Vicente-Carbajosa, R. J. Schmidt & P. Carbonero. 1998. An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolaminox of a native B-hordein promoter in barley endosperm. *Plant J* 16: 53-62
- Messing J, Grossniklaus U. Genome imprinting in plants. *Results Probl Cell Differ* 1999;25:23-40
- Miller ME, Chourey PS (1992) The maize invertase-deficient *miniature1* seed mutation is associated with aberrant pedicel and endosperm development. *Plant Cell* 4:297-305.
- Motto, M., H. Hartings, M. Maddaloni, S. Lohmer, F. Salamini, R. Thompson. 1996. Genetic manipulations of protein quality in maize grain. *Field Crops Research* 45:37-45.
- Motto, M., R. Thompson, F. Salamini. 1997. Genetic Regulation of carbohydrate and protein accumulation in seeds. in *Cellular and Molecular Biology of Plant seed development*. Ed. B. Larkins and I. K. Vasil. Kluwer, Boston. pp. 479-524.
- Muller M., G. Dues, C. Balconi, F. Salamini & R. D. Thompson. 1997. Nitrogen and hormonal responsiveness of the 22 kDa alpha-zein and b-32 genes in maize endosperm is displayed in the absence of the transcriptional regulator Opaque-2. *Plant J* 12: 281-291
- Neuffer G., W. F. Sheridan. 1980. Defective kernel mutants of maize. I. Genetic and lethality studies. *Genetics* 95:929-44
- Neuffer, J. 1993. Chimeras for genetic analysis. *The Maize Handbook*. Ed. M. Freeling and V. Walbot. Springer-Verlag. New York. pp. 259-61.
- Neuffer, MG and W. F. Sheridan 1980. Defective kernel mutants of maize. I. Genetic and lethality studies *Genetics* 95:929-944
- Olsen OA. 1998. Endosperm developments. *Plant Cell* 10:485-8
- Opsahl-Ferstad HG, Le Deunff E, Dumas C, Rogowsky PM. 1997. ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant J* 12(1):235-46

- Parcy F., C. Valon, A. Kohara, S. Misera & J. Giraudat. 1997. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* 9: 1265-1277
- Pysh L. D., M. J. Aukerman & R. J. Schmidt. 1993. OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. *Plant Cell* 5: 227-236
- Randolph, L.F. 1936. *J. Agric. Res.* 53: 881.
- Robertson, D.S. 1980. The timing of *Mu* activity in maize. *Genetics* 94:969-978.
- Rowland LJ, Robertson DS, Strommer J. 1989. Chromosome breakage undetectable in active *Mu* lines of maize. *Genetics* 122(1):205-9
- Sainz M.B., E. Grotewold & V. L. Chandler 1997. Evidence for direct activation of an anthocyanin promoter by the maize C1. *Plant Cell* 9: 611-625
- Sambrook J., E. F. Fritsch, T. Maniatis. 1989. *Molecular Cloning: A laboratory manual.*
- SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274(5288):765-768
- SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274(5288):765-8
- Scanlon MJ, Myers AM. 1998. Phenotypic analysis and molecular cloning of discolored-1 (*dsc1*), a maize gene required for early kernel development. *Plant Mol Biol* 37:483-93
- Scanlon MJ, Stinard PS, James MG, Myers AM, Robertson DS. 1994. Genetic analysis of 63 mutations affecting maize kernel development isolated from Mutator stocks. *Genetics*. 136(1):281-94
- Schmidt R. J., F. A. Burr & B. Burr. 1987. Transposon tagging and molecular analysis of the maize regulatory locus opaque-2. *Science* 238: 960-963
- Schmidt R. J., F. A. Burr, M. J. Aukerman & B. Burr. 1990. Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc Natl Acad Sci U S A* 87: 46-50
- Schweizer L, Yerk-Davis GL, Phillips RL, Srienc F, Jones RJ. 1995. Dynamics of maize endosperm development and DNA endoreduplication. *Proc Natl Acad Sci U S A.* 92(15):7070-4
- Scott R. J., M. Spielman, J. Bailey & H. G. Dickinson Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329-3341
- Sheridan, WF and Neuffer, MG. 1980. Defective kernel mutants of maize. II. Morphological and embryo culture studies *Genetics* 95:945-960
- Soave C, Tardani L, Di Fonzo N, Salamini F. 1981. Zein level in maize endosperm depends on a protein under control of the opaque-2 and opaque-6 loci. *Cell* 27:403-10
- Sun Y, Carneiro N, Clore AM, Moro GL, Habben JE, Larkins BA. 1997. Characterization of maize elongation factor 1A and its relationship to protein quality in the endosperm. *Plant Physiol* 115:1101-7
- Sun Y., B. P. Dilkes, C. Zhang, R. A. Dante, N. P. Carneiro, K. S. Lowe, R. Jung, W. J. Gordon-Kamm & B. A. Larkins. 1999. Characterization of maize (*Zea mays* L.) *Wee1* and its activity in developing endosperm. *Proc Natl Acad Sci U S A* 96: 4180-4185
- Suzuki, M., Y.-C. Kao and D. R. McCarty. 1997. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *The Plant Cell* 9:701-10.
- Tatusov RL, Koonin EV, Lipman DJ (1997) "A genomic perspective on protein families," *Science* 278(5338):631-7
- Tsai CY, Larkins BA, Glover DV. 1978. Interaction of the opaque-2 gene with starch-forming mutant genes on the synthesis of zein in maize endosperm. *Biochem Genet* 16:883-96

Walbot, V. 1993. Overview of key steps in aleurone development. The Maize Handbook. Ed. M. Freeling and V. Walbot. Springer-Verlag. New York. pp. 78-80.

Wang Z., T. Ueda & J. Messing. 1998. Characterization of the maize prolamin box-binding factor-1 (PBF-1) and its role in the developmental regulation of the zein multigene family. *Gene* 223: 321-332

Wobus U, Weber H. 1999. Seed maturation: genetic programmes and control signals. *Curr Opin Plant Biol* 2:33-8

Wu, T.D., Nevill-Manning, C.G. & Brutlag, D.L. (1999A) "Minimum-Risk Profiles of Protein Families Based on Statistical Decision Theory," *Journal of Computational Biology*, 6(2), 219-235.

Wu, T.D., Nevill-Manning, C.G. & Brutlag, D.L. (1999B) "Fast and Accurate Sequence Analysis Using Scoring Matrices," *Bioinformatics*, in press.

Young TE, Gallie DR. 1999. Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. *Plant Mol Biol* 39(5):915-26