Regulation of Inflorescence Architecture in Maize

Summary

Sarah Hake, University of California, Berkeley, P.I.
Volker Brendel, Iowa State University, Co-P.I.
David Jackson, Cold Spring Harbor, Co-P.I.
Elizabeth Kellogg, University of Missouri, St. Louis, Co-P.I.
Rob Martienssen, Cold Spring Harbor, Co-P.I.
Torbert Rocheford, University of Illinois, Champaign, Co-P.I.
Robert Schmidt, University of California, San Diego, Co-P.I.
Erik Vollbrecht, Cold Spring Harbor, Senior Personnel
Virginia Walbot, Stanford University, subcontract

Nearly all grasses are characterized by the spikelet, a short branch that contains floral meristems. The arrangement of these spikelets in different grasses, and the branches on which they are borne, reflects differing fates of the meristems produced during inflorescence development. Identifying the genes that determine meristem fates and understanding the mechanism by which these genes integrate their activities would be of immense value for developmental biology, evolutionary biology, and applied genetics and breeding. We propose to accomplish this goal by integrating genomic-based approaches with existing genetic and molecular resources developed in maize. We will identify inflorescence genes that will serve as tools for three different disciplines: investigation of meristem development, quantitative trait analysis, and comparative biology in the grasses. The "use" of these tools in each discipline feeds information back to the other disciplines, creating a synergistic picture of inflorescence architecture. Specifically, normalized cDNA libraries from both tassel and ear will be sequenced and arrayed. We will use the microarrays to profile the changes in gene expression that occur during ear and tassel development and to compare developmental profiles of selected inflorescence mutants to wild type. A subset of genes expressed at the earliest stages of development, and also correlating with the proliferation of specific meristem types in selected mutants, will become the focus for further study. Utilizing existing resources, we will obtain insertional mutants and map locations for a subset of the inflorescence genes. We will carry out a phenotypic analysis of inflorescence architecture primarily in existing mapping populations to identify relevant quantitative trait loci (QTL). Map positions of the inflorescence genes may provide a link to QTL, and to mutations in maize and other grasses. We will make genomic libraries from selected cereals and identify and sequence orthologs of cloned genes whose function has been uncovered by a mutant phenotype in maize. Expression patterns in developing cereal inflorescences will be determined. This will provide a valuable data set comparing sequence divergence and gene expression differences across 60 million years of grass evolution. Our first set of candidate genes includes *indeterminate spikelet1*, *ramosa1*, and *fasciated ear2*. These are involved in regulating, respectively, the number of florets in a maize spikelet, the switch from long branches to spikelets, and the size of the inflorescence meristem and thus the number of branch meristems initiated. Additional genes, whose loss of function phenotype indicates a role in regulating aspects of inflorescence architecture, will be similarly analyzed as they are identified. We will then be in a position to ask how these genes have been modified over evolutionary time, whether they function in other grasses the same way as they do in maize, and whether the variation in sequence or expression helps us understand and generalize the mechanism by which they act in maize.

I. Introduction

The maize inflorescence is an appropriate system for functional genomics from a practical, agronomic, developmental and evolutionary viewpoint. On the practical side, tassel branch variation and ear row number are easy to quantify. Tassel and ear primordia are large, facilitating RNA analysis. From an economic viewpoint, QTL (quantitative trait loci) analysis suggests a few loci of major effect regulate much of the variation for numbers of tassel branches and seed rows. Hybrid seed production generally uses inbreds with different tassel architectures. Ear size has a direct effect on kernel yield. From a developmental viewpoint, a number of mutations are known that influence tassel and ear architecture by affecting meristem fate and size. The broad spectrum of inflorescence phenotypes manifested by these maize mutants resembles the diversity of inflorescence architecture across the grasses. Despite the importance of maize as a crop and its rich genetic history, only a small number of genes have been identified that function in tassel and ear architecture. We will use a combination of mutant screens and microarray expression studies to identify inflorescence genes. Function will be determined for a subset by reverse genetics, and mapping studies will link others to QTL for tassel and ear characters. Genes with known functions in maize will be characterized in other grasses to test the hypothesis that changes in the timing or expression of key regulatory genes have contributed to the diversity among the grasses. The premise of this proposal is that understanding the mechanisms controlling the ordered progression of meristem fates will suggest how changes in expression of specific genes or groups of genes underlie the diversity of architectural designs that have evolved among grass inflorescences.

Our collective group has expertise in maize genetics, developmental biology, molecular biology, genomics, quantitative trait analysis, grass phylogeny and bioinformatics. We bring our common enthusiasm for inflorescence architecture to bear on the goals of this research proposal. The information and tools produced will strengthen our separate research programs and at the same time, provide a resource for others studying maize and cereal biology.

<u>Our Specific Aims are to:</u>

1. Identify ear- and tassel-expressed genes, not yet available in the public databases, through sequencing of normalized libraries.
2. Identify genes expressed at specific stages of inflorescence development using microarray-based analyses of gene expression to profile ear and tassel development.
3. Identify genes (using microarrays) with altered expression in selected inflorescence mutants arrested at specific stages of inflorescence meristem development.
4. Use bioinformatics analysis on microarray expression profiles to group subsets of genes unique to specific developmental stages and/or inflorescence mutants, and to identify sets of genes regulated coordinately.
5. Target 100 genes, selected on the basis of their interesting sequence and expression profile, for mapping, and determine the function of 30 of these genes using existing reverse genetic resources.
6. Identify new inflorescence genes by screening *Mutator* insertion libraries and EMS populations derived from inbred lines.
7. Identify and characterize QTL that affect ear and tassel architecture.
8. Analyze gene sequence and expression differences across 8 selected grass species by comparing orthologs of those cloned maize genes for which a mutant inflorescence phenotype exists.

II. Background

IIA. Maize Inflorescence Development

Maize produces two types of inflorescences, the tassel and the ear. The tassel forms directly from the shoot apical meristem after production of a defined number of leaves, whereas an ear forms from a meristem at the tip of a compact axillary branch. Although the mature tassel and ear are very different in appearance, their underlying organization and development is remarkably similar. The inflorescence meristem in the tassel produces four types of axillary meristems: branch meristems, spikelet pair meristems, spikelet meristems, and floral meristems. The ear produces the latter three (Fig. 1A). Alteration in the growth of these different meristems leads to variation in architecture between tassel and ear, between different maize inbreds, and between different grasses. Both ear and tassel produce short branches called spikelet pairs, each of which produces two spikelets (Fig. 1B). In addition to the spikelet

pairs, the tassel produces long branches towards its base (Fig. 1C & D). In both tassel and ear, each spikelet meristem produces two sterile leaves, called glumes, followed by two lemmas, each with a floral meristem in its axil. The floral meristem forms a palea, followed by lodicules, stamens and a pistil [1]. Although the spikelets on the ear and tassel appear similar as they initiate, their differentiation is distinct. There is a pedicellate and sessile spikelet in the tassel, whereas in the ear, this distinction is not obvious. In the tassel, pistils abort to produce a pair of staminate florets, whereas in the ear, lower florets in each spikelet abort and stamens abort to produce pistillate florets [1]. Additional features such as length of glumes, distinguish the male and female flower [2].

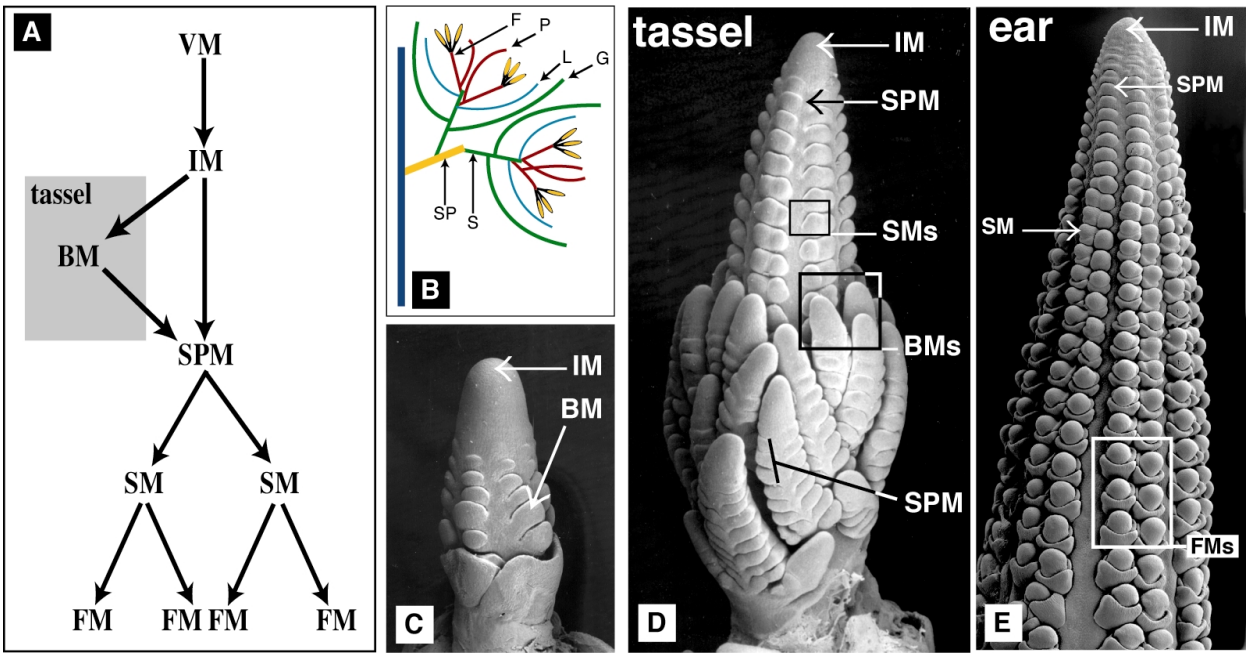

Figure 1. A) Diagram illustrating the progression of meristem fates during grass inflorescence development. Vegetative meristem (VM) becomes the inflorescence meristem (IM), which produces (in tassels) branch meristems (BM) and spikelet pair meristems (SPM). The SPM produces two spikelet meristems (SM), each of which makes two floral meristems (FM). **B)** Diagram of a tassel spikelet pair (SP) showing each spikelet (S) composed of two florets (F), each with a pair of glumes (G), lemma (L), palea (P), stamens and lodicules (not illustrated). **C and D)** SEM of, respectively, an immature tassel at 1.5 mm (developmental stage #2) and 0.4 cm (late stage #3). **E)** SEM of upper 2/3 of immature ear (late stage #8).

IIB. Maize Inflorescence Architecture Mutants as a Developmental System

A rich collection of mutants has demonstrated that distinct genetic steps can be defined in the progression from inflorescence to floral meristem in maize. These steps, the production of long branch, spikelet pair and spikelet meristems, differentiate maize from other model organisms such as Arabidopsis, and unite maize with other grasses. Our focus is primarily mutations that affect these meristems, thus producing an architectural change. Some of these mutants proliferate a specific meristem type early in their development and thus provide a valuable resource for examining the spectrum of gene expression changes that accompany these developmental steps during inflorescence maturation. In addition, the phenotypes produced by these mutants mimic some of the natural variation in the grasses. Here, in the order in which they affect the progression of inflorescence development, we will introduce those mutants that are central to this proposal. Table 1 provides a list of these and other mutants affecting inflorescence development.

<u>Inflorescence meristem produces branch meristems.</u>

bif2 mutants fail to produce branch or spikelet pair meristems (Fig. 2), providing a source of material that is only inflorescence meristem (McSteen and Hake, unpub.). *fae2* mutants have a larger ear inflorescence meristem and thus initiate more branch meristems. Branch, spikelet and floral meristems may also be fasciated, leading to spikelets that are occasionally found in threes rather than in pairs, as well as enlarged pistils. Arrangement of spikelets in threes is common in a number of grasses, including those closely

related to maize. *fae2* has recently been cloned by transposon tagging (Shiobara, Yuan, Hake and Jackson, manuscript in prep.). *fae2* encodes a protein with homology to CLAVATA2, a leucine rich repeat receptor-like protein that regulates meristem size in Arabidopsis [3]. Consistent with a proposed role in regulation of inflorescence meristem size, *fae2* maps to a QTL for row number (see IIC). The *thick tassel dwarf* (*td1*) mutant is similar to a *fae2* mutant, but also has a pronounced affect on the tassel. Spikelet and floral meristems are not affected (Running, Vollbrecht and Hake, unpub.).

Table 1. Meristem mutants define stages in maize development

Meristem	Mutant	Phenotype
inflorescence	*Fas1*, *fae1*, *fae2, td1*	fasciated, more spikelet pairs
	kn1	fewer branches and spikelet pairs
	tls1	short tassel, fewer spikelet pairs
branch	*ra1*	spikelet pair to branch
	ts4	branch meristem reiterates
	ifa1	extra spikelets
	ub1	fewer branches
	fae2	branch meristem fasciates
	ba1, *Bif1*, *bif2*	fewer branches and spikelets
spikelet	*bd1*	spikelet meristem reiterates
	ids1, *rgo1*, *ifa1*, *Ts6*	extra florets
	ra2	pedicellate spikelet to branch
	bif2	fewer florets
floral	*ifa1*	proliferative meristem
	zag1	extra carpels
	fae2	enlarged gynoecium
	bif2	fewer stamens

Branch meristems produce spikelet meristems.

Several mutants either proliferate or fail to produce branches, providing a source of material enriched or deficient in branch meristems. In the *ramosa1* (*ra1*) mutant, spikelet pairs convert to long branches [4-6]. The spikelet pairs on the long branches are normal. *ra1* encodes a zinc finger transcription factor closely related to the Arabidopsis gene *SUPERMAN* (Vollbrecht, Goh, Springer and Martienssen, unpubl). *ra1* maps to a QTL with large effects on tassel branch number (see IIC). *ramosa2* (*ra2*) has a similar phenotype to *ra1* except that in *ra2* the pedicellate spikelet is converted to a branch [5]. All spikelet pairs are affected in *ra2* mutants, including spikelet pairs on long branches. In both *ra1* and *ra2*, branches produce fewer and fewer spikelets acropetally giving the tassel a "Christmas tree-like" appearance (Fig. 2). A similar phenotype occurs in many wild grass species.

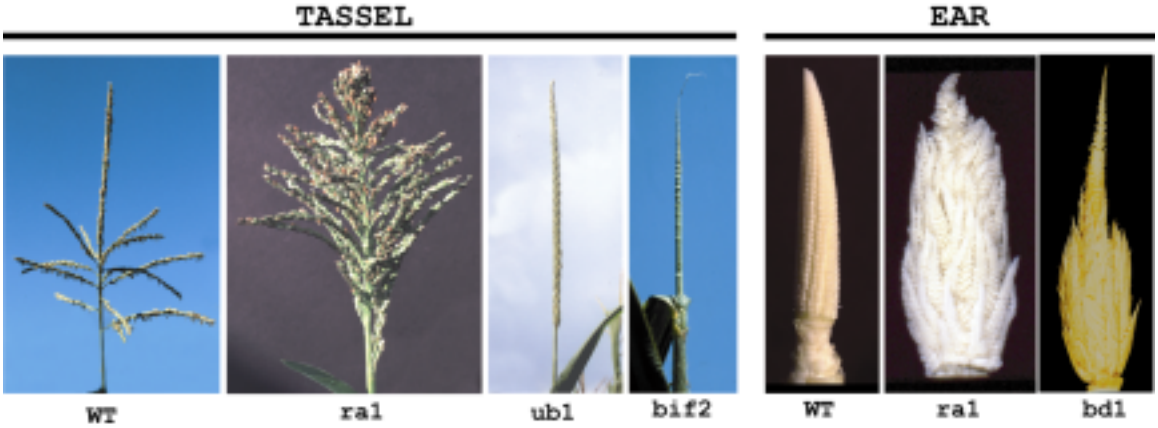

Figure 2. Maize Inflorescences. A) Mature wild type and mutant tassels. **B)** Immature wild type (silks removed) and mutant ears.

Ears are affected in both *ra1* and *ra2* mutants, creating a highly branched female inflorescence with disorganized rows and some sterility (Fig. 2). *unbranched1* (*ub1*) gives a phenotype opposite to *ramosa* mutations. Tassels of *ub1* plants (Fig. 2) have only spikelet pairs, and no long branches [7]. This phenotype occurs in such wild taxa as *Heteropogon* (a noxious weed), *Phacelurus*, and *Coelorachis*, among many others. Genes up-regulated in *ra1* might not be expressed in *ub1* mutants. A different class of mutants, represented by *tasselseed4* (*ts4*), results in branch meristems that produce more branch meristems [8]. This highly reiterative phenotype suggests that *ts4* may be required for spikelet meristem identity. Early development of foxtail millet (*Setaria italica*) looks remarkably like early development of *ts4* mutants (Doust and Kellogg, unpub.), suggesting a similar developmental program.

Each spikelet meristem produces two floral meristems.

Another class of mutants affects the transition from spikelet meristem to floral meristem, providing a source of material enriched in spikelet meristems. Still others affect meristem determinacy, a trait that may underly some of the phenotypic diversity among cereal inflorescences. In ears of *branched silkless1* (*bd1*) mutants, floral meristems are replaced by branches that proliferate spikelet meristems (Fig. 2), suggesting that *Bd1* is required for floral meristem identity and has a role in branch meristem suppression. In the tassel, extra spikelets form, showing that *bd1* is also required to promote spikelet meristem determinacy in the tassel [2, 9, 10]. *indeterminate spikelet (ids1)* affects determinacy of the spikelet meristem in male and female inflorescences. *ids1* mutants make more florets per spikelet, suggesting that *ids1* is required to suppress floral meristem formation in the spikelet or promote the conversion of the spikelet to a floret [11]. *ids1* encodes an AP2-like protein that is expressed in spikelet and branch meristems as well as certain floral organs [11]. Grasses vary in the number of florets produced per spikelet; this could correlate with differences in expression of *ids1* orthologs.

IIC. Inflorescence Architecture from an Agronomic and Quantitative Trait Viewpoint

Tassel architecture in maize inbreds and F1 hybrids.

F1 hybrid seed production in maize relies on separation of the tassel and ear and uses detasseling and male sterility systems to avoid self-pollination of seed producing plants. The ideal male parent has a large tassel, shedding copious pollen over a long time. The ideal female parent has a relatively large ear that produces large numbers of kernels. This strategy ensures good pollination and enables growing fewer rows of male pollinators. Maize breeders, however, have indirectly selected for smaller tassels in F1 hybrids since smaller tassels appear to be associated with higher grain yields [12]. These contrasting needs prompt the desire to manipulate the tassel such that male parents have large tassels but F1 hybrids have small tassels. Additionally, the new ®TopCross pollination system for high kernel oil concentration corn (HOC) involves growing fields of male sterile, high yielding commercial hybrids with a small percentage of pollinator plants of extremely high oil concentration. The pollinators need to have very large tassels that shed pollen for long periods of time. Understanding the developmental mechanisms behind tassel and ear architecture and isolating important genes controlling this process may improve strategies for hybrid seed corn production, and enhance yields.

Maize inflorescence architecture as a model and applied system for quantitative traits.

Most important agronomic traits are quantitatively inherited, the result of numerous genes having relatively small effects on the phenotype [13]. Such genes are termed quantitative trait loci (QTL). Inflorescences are ideally suited for QTL analysis, which detects chromosomal regions containing a gene(s) with minor allelic differences [14]. Maize inbreds vary considerably for tassel branch number, branching pattern, branch size, branch angle, spikelet density, ear row number and kernels per row. These traits can be measured easily, precisely, and accurately, enhancing the power and reliability of detecting QTL. When their map position correlates with a QTL, maize mutants affecting inflorescence morphology provide candidate genes for tassel and ear architecture QTL. QTL mapping analysis, coupled with new resources such as inflorescence EST libraries and microarray analysis, will provide additional and complementary clues towards identifying genes affecting inflorescence architecture, particularly those for which mutant alleles are presently not available.

In current breeding practices, QTL are selected based on linkage to nearby markers, which unfortunately can result in linked agronomically undesirable alleles also being selected and introgressed

[13]. Promising experimental inbreds associated with high F1 hybrid grain yield are often discarded because they do not have appropriate tassel architecture. Identification of ESTs that are tightly linked to, or are the actual QTL will make marker assisted selection more efficient, as smaller flanking regions could be introgressed [15]. We have three promising candidate genes. A major QTL for tassel branch number mapped to the chromosome region near *ra1* [16]. Based on subsequent allelism tests, the QTL may be allelic to *ra1*. A QTL for ear row number maps to the same chromosome region as *fasciated ear2* (D. Jackson, unpub.). A QTL for spikelet density maps near the *td1* gene (Rocheford, unpub.). Because *ra1* and *fae2* are cloned, this linkage can be tested by sequencing *ra1* and *fae2* alleles from the parental lines used to generate the QTL mapping population. These genes may provide useful agronomic tools and also provide lead examples for studying genetic relationships of mutant alleles and putatively related QTL alleles influencing tassel or ear architecture.

IID. Comparative Approaches to the Study of Inflorescence Architecture

Crops that are considered cereals, like maize, rice, wheat, sorghum, barley, millet, and oats, account for the majority of calories consumed in the world. In addition to the cereal crops, there are approximately 10,000 species of wild grasses, which together cover about 1/5 of the earth's land surface [17, 18]. Genetic information on the regulation of maize inflorescence architecture should be readily transferable between other species, both cultivated and wild. Development of maize as a model system assumes that information from maize should be applicable to other cereals [19], and indeed to any other plant. Comparisons between maize and the other cereals, and between maize and wild grasses, will test this basic assumption of model system development. Preliminary data from the Kellogg lab suggests that the sequences of such major regulatory genes as *kn1* and *lfy* are highly conserved.

Of the grasses, maize has certain advantages for the study of inflorescences, notably transposon tagging, a growing EST collection, floral mutants, and ease of genetics due to the separation of male and female flowers. Many of the other cereals also have a rich collection of mutant phenotypes including barley, sorghum, oats, and millet [20-22]. For the barley mutants, map locations exist (http://wheat.pw.usda.gov/ggpages/bgn/).

Phylogenies of the grass family have been recently updated [23-27]. Maize is closely related to sorghum and sugar cane and is somewhat more distantly related to the millets. Rice, wheat, barley, rye, and oats are more closely related to each other than any of them are to maize. The actual similarity and difference between these species at specific genetic loci remains to be tested.

Many morphological differences among the cereals are differences in inflorescence architecture. Wheat and barley do not produce any long branches in their inflorescences, unlike most other species which do produce long lateral branches. In the pooid grasses (e.g., oats), long branches are produced in a strictly distichous pattern, whereas in rice and the panicoids, long branches initiate in a spiral [28-30]. The genetics of these differences are unknown. The genes involved might be the same as those in maize; *ra1, ra2, ub1* are candidates.

Production of a spikelet pair meristem is shared among maize, sorghum, sugar cane, and about 100 other genera (e.g. *Paspalum*, *Digitaria*) [30]. Because spikelet pairs apparently do not occur in other grasses, they may be caused by a novel mutation that occurred ca. 20-30 million years ago. Candidate genes that might mediate this switch are *bd1*, *ids1*, and *ifa1*. All grasses (except the peculiar species in the subfamily Anomochlooideae) have flowers arranged in spikelets. The millets, sorghum, and sugar cane are like maize in having two and only two flowers per spikelet, but wheat, barley, rye and oats often have more than two, and are thus reminiscent of maize *ids1* mutants. Rice, like a number of turfgrasses, has only one flower per spikelet, and some species of ryegrass (*Lolium*) have as many as 20 flowers. Flower number in at least some of these species may be controlled by genes such as *ids1*, *ifa1*, or *bd1*. Foxtail and pearl millet are particularly interesting architecturally. The inflorescences are made up of contracted branches (probably "long branches"), each bearing one or a few spikelets and multiple bristles. The bristles are reminiscent of sterile, proliferative branches seen in some maize mutants, such as *ts4*. This phenotype also could reflect differences in regulation of the corresponding *ids1* genes.

Just as characters of inflorescence architecture vary quantitatively (in the case of QTL) or strikingly (in the case of mutant alleles) within maize, these characters display a broad range of diversity across the grass family. Thus, comparative analysis of inflorescence-specific gene function and regulation among grasses provides an extensive survey of potential allelic diversity. Such diversity points to characteristics that might have been acted on by past natural or, in the case of crops, human selection.

Doebley's work provides an excellent example of how developmental biology and QTL analysis together elucidate mechanisms of crop domestication and evolution. The morphological and architectural differences between maize and teosinte map to 5 major QTL [31], one of which corresponds to the maize gene, *teosinte branched1* (*tb1*) [32, 33]. Sequence comparisons of *tb1* alleles showed that promoter sequences were highly conserved amongst maize lines but not teosinte lines, suggesting that regulation of *tb1* was an important contribution to the difference between maize and teosinte [34].

III. Experimental Plan

Our primary source for uncovering inflorescence genes will be cDNA library sequencing and gene profiling. A smaller but informative set of genes will be identified by phenotypic screens of existing *Mutator* insertion populations and EMS-generated mutants. Comparison of microarray data sets from complete developmental profiles of wild type with those of specific mutants, whose arrested development provides a source of a particular meristem type, will uncover markers for specific meristem fates and identify sets of genes important to each step in inflorescence development. Expression profiling will begin with EST-based microarrays produced by U. of Arizona. We anticipate switching to an oligo-based technology in Year 3 (see letter #2 from V. Walbot). We will also determine whether QTL for maize inflorescences can be identified with genomic-based tools developed as part of this proposal. Finally, we will compare gene sequence and in situ expression profiles of functionally defined inflorescence genes across 8 species of grasses spanning 60 million years of evolution and covering an array of inflorescence architectural types. All tools generated through this proposal including ESTs, mircroarrays, libraries and mutants will be made available to the public.

<u>Specific Aim 1.</u> Identification of Ear- and Tassel-expressed Genes

The Maize EST database (ZmDB), a product of the NSF-funded Maize Gene Discovery Project, has provided a set of unique cDNA clones (www.zmdb.iastate.edu). Although more than 74,000 ESTs have been identified, this is far from saturation, and a number of genes known to be expressed in ear or tassel have yet to be uncovered (e.g. *te1*, *ra1*, *ids1*). Schmidt has recently provided an immature tassel (1-2 mm) library (inbred OH43A) to ZmDB for sequencing and 5000 ESTs have been sequenced and deposited. The overlap between this young tassel stage and the more mature (1-2 cm) tassels is around 15%. Plans for a 2-4 mm ear library are underway. Although further sequencing of these libraries will certainly yield additional unique genes, it will be more cost-effective to sequence from normalized libraries [35, 36]. Normalized cDNA libraries will be subtracted for all ear and tassel cDNAs in the existing Maize EST database. This approach should allow us to identify even cDNAs whose abundance is comparatively low and will dramatically increase our success rate at identifying new ESTs. Sequence analysis will be subcontracted through the Stanford Sequencing Center.

<u>III 1a. cDNA Normalization and Subtraction Procedures.</u>

First strand cDNAs will be generated from RNA isolated from developing tassels and ears between 0.1 and 1.5 cm in length. This covers stages of inflorescence development from branch meristems and spikelet pair primordia through establishment of floral organs. To insure good representation of cDNAs from earliest developmental time points, the ratio of 0.1 cm ears and tassels to that of 2.0 cm samples will be 50 to 1. For constructing the libraries we will use inbred OH43A, which was used for previous ear and tassel libraries, and is intermediate among inbred lines for inflorescence size, branching and time to flowering.

An excess of single stranded cDNA will be hybridized to a limiting amount of denatured, genomic DNA for normalization. The genomic DNA is prepared by partially digesting with *Sau*3A and *Mae*III and size selecting in the 1-4 kb range. A Klenow fill in reaction incorporates biotin-dUTP into the ends of fragments followed by denaturation and incubation with streptavidin-coated magnetic beads. The single stranded cDNA is mixed with the beads in hybridization buffer for 16 hrs at 65°C. After hybridization, unbound cDNA is removed by washing; hybridized cDNA is eluted and recovered by ethanol precipitation. Normally at this point, the recovered cDNA is processed using standard procedures for constructing cDNA libraries (Stratagene). However, to avoid sequencing many clones already in ZmDB, we will subtract existing ESTs after normalization. A small aliquot (1 µl, equaling about 1 ng) of PCR-amplified tassel and ear ESTs (already available from D. Gailbraith's lab, U. of Arizona) will be

pooled and amplified by bulk PCR using biotinylated primers unique to the vector sequence flanking the cDNA inserts. Amplified products will be denatured and incubated with streptavidin-coated magnetic beads. The normalized, first strand cDNA from developing ear and tassel is mixed with the beads in hybridization buffer for 16 hrs under conditions stringent enough to prevent association of related, but non-identical, cDNAs. After hybridization, beads containing the subtracted cDNAs are removed and the unbound unique cDNAs are ready for cDNA library construction. Schmidt already has considerable expertise in constructing libraries and we anticipate no significant difficulties.

Although we realize that sequences resulting from normalization and subtraction may not provide full length clones, incomplete sequences are still useful for making oligo-based probes, but very small fragments (< 300 bp) will be omitted by size selection.

III 1b. Sequencing.

The Stanford sequencing unit, already skilled in rapid and large scale cDNA sequencing, has agreed to sequence these normalized libraries (see accompanying letter and budget justification). All sequence information will immediately enter the publicly accessible databases established for the Maize Gene Discovery Project (www.zmdb.iastate.edu). Stanford will sequence a total of 100 96-well plates (9600 reactions in the 5' and 9600 in the 3' direction) beginning on the tassel library and switching to the ear library after 60 plates. Initial sequences will be compared with the existing database to determine the level of new gene discovery. If the rate of gene discovery drops quickly with the tassel library, we will switch to the ear library sooner. Additional libraries from RNA of mutant tassels could be sequenced if gene discovery rates become low. For example, we would sequence libraries of *bd1* or *ts4* mutants, which are enriched in branch meristems. Based on past sequencing effort, we anticipate about 6600 ESTs will be identified and transferred to the U. of Arizona for preparation of microarrays. Using 3' sequence information, we can confidently identify duplicate genes and reduce the redundancy of ESTs for microarraying.

Contig analysis (finding ESTs with overlap) will be performed on sequences from the normalized libraries and the existing 12,000 ESTs from ear and tassel libraries. Past experience has shown that approximately one third to one half of the ESTs sequenced are unique. We anticipate that the 18,600 ESTs will be reduced to around 6,000 total unique tassel/ear ESTs. This unique set will be used for constructing an ear and tassel chip that will provide the basis for all expression profiling studies.

Specific Aim 2. Identify Genes Expressed at Specific Stages of Inflorescence Development through Expression Profiling on Microarrays

We are interested in identifying genes that function in tassel and ear architecture. Through hybridizations to microarrays, we will determine which genes are expressed at specific times and which are expressed throughout development. The cluster analysis [37] that results will be compared to gene expression patterns of mutants that either proliferate or eliminate particular meristem types. We will include labeled targets from vegetative meristems (to identify genes shared with the vegetative meristem), *bif2* mutants (to identify inflorescence meristem markers), *ra1* and *ub1* (to identify branch meristem markers), and *bd1* (for markers of spikelet meristems). The microarray profiling will provide information to select a subset of genes to map and obtain gene knockouts. Although we will begin with EST-based microarrays, we anticipate switching to oligo-based profiling when such arrays become available (see 3c).

The Galbraith group has recently printed microarrays using ESTs from a young seed library and an ear library (see www.zmdb.iastate.edu). Members of the Schmidt and Hake labs traveled to Arizona in December to learn how to do hybridizations and scans. Martienssen and Vollbrecht also have experience with microarrays through facilities at CSHL. From these initial experiments, we realize the importance of abundant, high quality RNA, sufficient replicates and a thorough application of statistics and data management. A scanner has been purchased by the genomic facility where Hake has her lab (see attached letter) and a core facility at UCSD will provide on-site access for members of the Schmidt lab. Two scanners and support personnel for their operation are also available at the CSHL. Thus, the consortium has considerable expertise available for this emerging technology, and we are confident that we can rapidly begin generating meaningful and exciting results.

III 2a. Hybridization to Tethered Probes on Microarrays

Our normalized cDNA libraries will be administered according to the implemented protocols of the Stanford and Arizona labs funded by the NSF Maize Gene Discovery. Sequenced cDNAs will be archived at –80°C at the Stanford DNA Sequencing and Technology Center, with duplicate plates at the Plant Gene Expression Center and Dept. of Plant Science, Arizona. Only 1-2% of the DNA purified for sequencing is required. "Extra DNA" is shipped to the University of Arizona in indexed plates, along with pre-constructed tabular data on each clone. In Arizona, DNA samples are reracked to make a unique representation of the clones, and amplified by PCR for construction of microarrays. All PCR reactions are purified and analyzed on a gel to determine size and quality of product. Only cDNAs with high quality single PCR reactions are arrayed. Each PCR reaction produces enough template for a 1000 slides. These chips will provide the foundation for our profiling studies and will be made available to other scientists at minimal cost (ca. $100 per slide) through ZmDB.

III 2b. RNA Preps and Labeling

We will use the microarray nomenclature in which ESTs placed on the slide are "tethered probes", and RNA is "labeled target" [38]. We will include on our tassel slide negative controls such as human genes and positive controls such as constitutively expressed genes. Microarrays will also include several genes whose expression during maize inflorescence development has been well characterized in our laboratories by in situ hybridization, including *knotted1*, *zag1*, *zag2*, *ids1*, *zap1*, *ra1*, *fae2* and *silky1* among others. These will be important positive controls, because we know when during development these genes are first expressed and how their expression changes over developmental time. The slides can fit 20,000 DNA elements. Therefore, there is room for 3 replicates of the library and numerous controls.

To minimize background variation from environmental effects, we will obtain all target RNA from field-grown B73 plants in San Jose, CA where fields are uniform and irrigated. RNA for generating cDNA libraries and the labeled targets will be extracted from material pooled from numerous plants to minimize effects of plant to plant variation. To minimize the time at room temperature, we will immediately freeze dissected samples and sort frozen ears and tassels into distinct size classes while keeping them on dry ice. RNA will be isolated using Trizol (GIBCO/BRL), and polyA+ RNA isolated with Dynal bead kits. We typically average about 25 µg total RNA from a 0.5 cm tassel or ear, so would need 30 plants to obtain 600 µg total RNA which should yield 16 µg of poly(A)+ RNA. PolyA+ RNA is reverse transcribed into cDNA with Cy3 or Cy5 labeled dCTP or dUTP using Sigma's RT-PCR kit. 4 µg of polyA+ RNA are presently used per labeling but we expect to optimize hybridizations with 2 µg.

A potential problem is obtaining sufficient poly (A)+ RNA from earlier points in development (0.1-0.2 cm ear and tassel) and from vegetative meristems and *bif2* mutants. These may require over several hundred dissections to obtain sufficient RNA for a complete experiment. To circumvent this difficulty, we plan to immortalize the initial source of poly(A)+ in the form of a cDNA library. With this approach we would have an abundant and reproducible source of target for microarray experiments. This requires more up-front work, but in the long run greatly minimizes variation (i.e. the probe is the same in all experiments from a given developmental stage). For this project it may be necessary to make several libraries, but that is less work and more cost-effective than re-isolating RNAs from dissected tissues grown in different environments, especially if library construction is done in parallel.

The Schmidt lab is testing this procedure using a newly made, maize endosperm library (14 day after pollination, DAP) and endosperm microarrays that have been created as part of the Maize Gene Discovery Project. To minimize background hybridization to tethered probes on the microarrays, the cDNA library was constructed using 5' and 3' primers distinct from those in the cDNAs on the microarrays. In our initial experiment we compared a microarray hybridized with labeled target generated from the 14 DAP poly(A)+ message with labeled targets generated from excised plasmids of the endosperm cDNA library. Although the patterns are not identical, they are sufficiently similar that we are encouraged to continue this approach (Fig. 3). Such immortalized libraries will not only provide material for generating reproducible array targets for this project, but also a valuable resource for others.

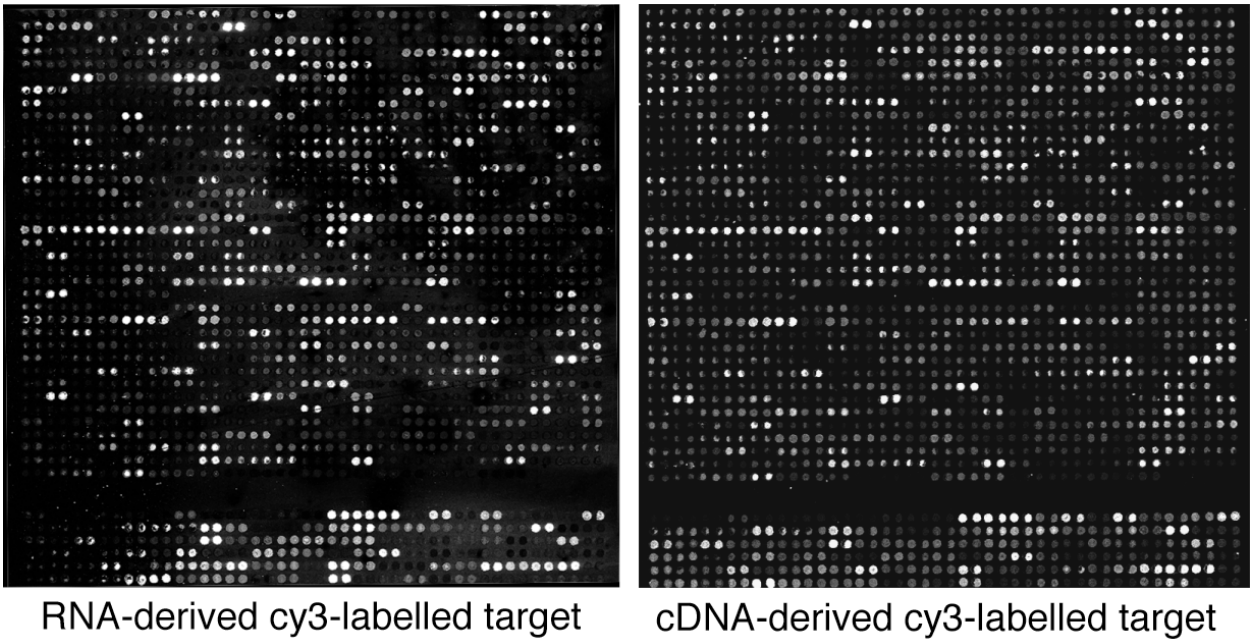

Figure 3. A comparison of microarrays hybridized with labeled target generated from 14 DAP endopserm poly (A+) RNA (left panel) and labeled target generated from a cDNA library made from an aliquot of the same poly (A+) RNA sample (right panel).

III 2c. Hybridization Format.

Each experiment will have three repeats of the following hybridizations.
A). RNA 1 is labeled with cy3 and RNA 2 is labeled with cy5 and hybridized.
B). RNA 2 is labeled with cy3 and RNA 1 is labeled with cy5 and hybridized.
We will follow up on targets that reproducibly vary in expression 3-fold or more. Three repeats of A and B, with the entire unique gene set arrayed three times, will give 18 hybridizations. The inherent variability in microarray analyses requires redundancy for meaningful results. With three replicates of each experiment we can calculate a mean and standard error, and use t-tests to determine significance.

Following denaturation of the tethered probes on the slide, labeled target is applied and spreads under a coverslip. Slides are incubated at 65°C in a humidity chamber, then washed, and immediately transported in the dark for scanning. Brendel will manage data analysis (see Specific Aim 4).

Table 2. Developmental Stages

#	Age	Enrichment for Meristem Type
#1.	3 wk	vegetative meristem + some leaf primordia.
#2.	4-5 wk tassel (0.1-0.2 cm)	long branch and some spikelet pair meristems.
#3.	5-6 wk tassel (0.2-0.4 cm)	spikelet pair and spikelet meristems predominate.
#4.	5-6 wk tassel (0.5-0.8 cm)	spikelet and floral meristems predominate.
#5.	6 wk tassel (0.9-1.5 cm),	floral organs predominate.
#6.	7 week ear (0.1-0.2 cm)	inflorescence and spikelet pair meristems predominate.
#7.	7-8 week ear (0.2-0.4 cm)	spikelet pair and spikelet meristems predominate.
#8.	7-8 week ear (0.5-0.8 cm)	spikelet and floral meristems predominate.
#9.	8 week ear (1.0-1.5 cm)	floral organs predominate.

III 2d. Developmental Profiles

We will investigate changes in gene expression over developmental time. We will use inbred B73 because the inflorescence mutants have been introgressed into this line, and B73 is one of the parents for the QTL mapping studies. Material will be collected at specific times of development and inflorescence sizes. These timepoints (Table 2) have been determined for B73 under our field conditions. Careful observation and dissection will provide the appropriate stages. Because tassel branches reiterate earlier stages of the main axis, we will remove side branches before freezing in liquid nitrogen.

We will compare the stages sequentially. RNA from stage #1 versus #2 should identify genes that may be important in the transition from vegetative to reproductive development. Stage #2 versus #3 may identify genes unique to long branches or to spikelet meristems. Stage #3 versus stage #4 may identify genes unique to spikelet pair or floral meristems, etc. Cluster analysis (see Specific Aim 4) will be used to identify groups of genes whose expression is unique to each time point and coordinately changing during development. We expect to see some genes, like *kn1* and *fae2*, expressed at all stages. Others such as *zag1*, would initiate expression as floral meristems are produced. We will be interested in any genes appearing in unique stages. Similar comparisons will be performed between selected stages of ear development: #6 vs #7, #7 vs #8, and #8 vs #9.

Specific Aim 3. Expression Profiling of Maize Mutants

We have picked four mutants (*bif2*, *bd1, ra1* and *ub1*) to help us identify genes expressed at specific time points in inflorescence development (Table3). These analyses, to be conducted during Years 2 and 3, will hopefully highlight about 100 genes with which to carry out genetic analyses (Specific Aim 5). Expression profiles from additional mutants (see III 3b) may be generated in Year 3 and 4 to contribute information about the genes on the array and help place newly identified mutants into existing pathways.

III 3a. Use of Mutants to Identify Genes Unique to Different Developmental Timepoints.

We will collect tassels from *bif2* mutants and normal siblings. *bif2* mutant tassels in a B73 background fail to make any branches and so lack spikelets and florets. Genes unique to branch, spikelet pair, spikelet and floral meristems would not appear in *bif2* mutants. Genes expressed in *bif2* mutant tassels, but not in vegetative meristems, may be unique to the inflorescence meristem. The *bif2* mutant is male and female sterile so populations that are segregating 1:3 will be used.

We will also carry out microarray hybridizations with ear RNA from *bd1* mutants and their normal siblings. *bd1* ears initiate spikelets, but not floral meristems, suggesting that *Bd1* promotes the transition from spikelet to floral meristems. This RNA will be enriched for transcripts expressed in spikelet meristems and will provide an informative comparison with the developmental timeline. Genes normally expressed in stage #4 but not #3 should be missing in *bd1* mutants. These may be direct or indirect targets of *Bd1*. *bd1* mutants, which are female sterile, will be crossed as heterozygotes by homozygous mutants to generate 1:1 segregating populations. RNA will be isolated from mutants and normal siblings at the 0.5-1 cm stage.

The *ra1* mutation converts spikelet pairs to branches, so we may enrich for expression of genes promoting long branch meristems by collecting *ra1* tassels at 4-5 weeks. expression of these same genes should be absent or near background levels in all ear inflorescence stages. As a negative control to the hybridization seen with *ra1* tassel RNA, we will also make RNA from *ub1* tassels, which only produce spikelet pairs and lack all long branches. Again, we predict that genes unique to long branches will not be expressed in *ub1* tassels. Finally, because *ra1* ears proliferate branches, we predict that comparing the profile of *ra1* mutant ears from stage #6 to that of wild-type should reveal gene expression differences unique to long branches.

Table 3. Enrichment for developmental stages using mutants

Meristem type	Expression Stages	Enriched in	Reduced in
Vegetative	#1	-----	-----
Inflorescence	#2, 3,	*bif2, fae2*	-----
Long branch	#2	*ra1*	*bif2, ub1*
Spikelet pair	#2, 3, 6	*ts4*	*bif2*
Spikelet	#3, 4, 6, 7	*bd1*	*ra1, ts4, bif2*
Floral	#4, 5, 7, 8	-----	all

III 3b. Expression Profiling of Additional Mutants

We have several other mutants ready for gene expression analysis. If time and resources permit, we will start with analysis of the *fae2* mutant, recently cloned by Jackson. It will be interesting to compare the expression profile of *fae2* mutants, which have an enlarged inflorescence meristem, to that of *bif2* mutants, which fail to make branch and spikelet meristems but make a normal inflorescence meristem.

Genes unique to both mutants may be specific to the inflorescence meristem. Given time and resources, we would then hope to compare these profiles with analysis of *ra2*, and possibly *ts4*.

III 3c.Oligo Arrays.

Detailed gene expression analysis may be better served by Affymetrix-type oligo-based methods, primarily because closely related members of a gene family are not resolved when the entire cDNA is arrayed. Genes with greater than 85% sequence similarity will cross-hybridize using these microarrays (D. Galbraith, personal communication), so related genes could give false positive signals. Alternatively, oligos can be designed to 3' ends and thus take advantage of differences in non-coding regions between related genes. Use of multiple oligos that span the 3' ends allows statistical analysis of a positive hybridization. In addition, mismatch oligos can provide an internal negative control for each gene arrayed [39]. Comparisons between separate experiments are also improved and much less RNA is required per hybridization. If the Walbot proposal to make Affymetrix corn chips is funded (see letter), we would have the option of switching to that methodology. We would use their CornChip1, which will include sequences from 2 mm tassels and ears in year 3, and CornChip2 in year 4, which will include additional genes we discover through this grant.

One potential obstacle to oligo arrays is the high polymorphism in maize. Will the labeled targets need to be from the same inbred background as the printed oligos? If so, we have already prepared by introgressing many of our mutants into OH43A, the main source of the ear and tassel sequences. Our limited data set, comparing a few genes cloned from different inbreds, suggest that 3' untranslated regions vary by 1-2%. If the oligos span the 3' end, then these 1-2 bp changes between inbreds should not be a problem. We should be able to take advantage of the Walbot Affymetrix proposal to make that determination.

Specific Aim 4. Bioinformatics and Data Management

Most computational data management and analysis tasks of the project will be handled by the group of co-PI Brendel at Iowa State University. This group is currently NSF-funded to provide the database and bioinformatics support for the Maize Gene Discovery project (PI: V. Walbot, Stanford University). The public interface for that project is the ZmDB web site (http://www.zmdb.iastate.edu/). The scope of ZmDB and available gene identification tools are described in [40-42]. Data derived in the proposed project will also be made accessible at ZmDB. In the first year, this will involve merely EST sequence data from the normalized ear and tassel cDNA libraries. These ESTs will be processed in the same way as EST collections from the current Maize Gene Discovery project. Clones will be available at cost from the Stanford Sequencing Center via web ordering at ZmDB. Beginning in Year 2, a subset of unique ESTs will be selected from the ZmDB collection for preparation of a project-specific microarray chip at the University of Arizona. There is increasing recognition of statistical issues relevant to the design (and analysis) of microarray chips (see sites of Gary Churchill http://www.jax.org/research/churchill /research/expression/index.html and Terry Speed http://www.stat.berkeley.edu/users/terry/zarray/Html). We will review current recommendations and incorporate suggestions into the design of our chip with respect to number and arrangement of duplicates and controls.

Bioinformatics work in Years 2-5 will concentrate on analysis of the microarray data. Because this field is rapidly evolving, we cannot anticipate all the details of the specific analyses. Our general strategy, based on current procedures, is as follows. Scanning of microarrays will be conducted at Berkeley, UCSD and CSHL and eventually other locations. Scanning will produce about 20MB TIF files for each dye used on each slide. These data will be stored as primary data on dedicated disks at ZmDB. With about 50MB per experiment, several hundred experiments can easily be stored on disk. Data will be analyzed by publicly available software. For example; there are several excellent web sites providing such resources, including Stanford's MicroArray Database (http://genome-www4.stanford.edu/MicroArray/ SMD), Michael Eisen Lab (http://rana.lbl.gov/), Molecular Pattern Recognition at Whitehead/MIT (http://www.genome.wi.mit.edu/MPR), and National Human Genome Research Institute MicroArray Project (http://www.nhgri.nih.gov/DIR/LCG/15K/HTML). Secondary data from this analysis will be ratios of inferred gene expression of experimental and controls for all ESTs on the microarray chip. These ratios will be displayed at ZmDB as log-log scatter plots of intensities (such that same ratios are represented by 45 degree lines, with highly under- or over-expressed genes far off the main diagonal).

Different regions of the scatter plot will be hot-linked to HTML tables displaying all genes expressed in the 3-5 or >5 ratio ranges, for example. Sequence information on these ESTs will be accessible through ZmDB entries. Java scripts implementing such displays are already in common use, including those developed by collaborators of Brendel in Bielefeld (see below).

Analysis of secondary data will involve clustering of similarly expressed genes. Again, this is an area of much current research, so we will rely on methods publicly available at that time. Currently employed methods include applications based on hierarchical clustering [43], self-organizing maps [44], and support vector machines [45]. Temporal patterns of gene expression have been successfully derived from microarray data using singular value decomposition [46, 47]. For all these methods, software implementations for microarray data analysis are freely available. We are also collaborating with the group of Prof. R. Giegerich, Praktische Informatik, University of Bielefeld, Germany, on the design of microarray data analysis tools. This group in turn collaborates with the laboratory of Terry Gaasterland at Rockefeller University. We currently have an account on the Rockefeller microarray server (http://arrays.rockefeller.edu/xenopus/). Because much if not all of our needs are met with that software, we anticipate mirroring their publicly available components or licensing their commercially developed software, executed by the ISU Bioinformatics Center that was set up to support exactly such infrastructure needs.

Specific Aim 5. Targeting Genes for Mapping and Knockouts

Sequencing and expression profiling will identify genes that are expressed in the inflorescence. We can begin to make inferences of function by following their expression pattern, i.e., genes not expressed in *bif2* but up-regulated in *bd1* may function in production of spikelets. To truly understand function, however, will require identification of allelic differences. Loss of function alleles may provide a mutant phenotype that is readily visible, but gene duplication may mask loss of function, necessitating knockouts in both genes and double mutant analyses. The gene may also contribute to a quantitative difference in inflorescence architecture, in which case, the gene may map near an established QTL (see Aim 7).

III 5a. Which Genes to Pursue.

We will pick 100 genes for further analysis. We anticipate that many genes important in influencing inflorescence architecture will be expressed early in development and encode regulatory or signaling proteins. Many of these will correlate with a specific developmental stage and a subset are expected to be affected in the mutants under study in this proposal. For example, genes important in branch meristem determination can be predicted to be present in developmental stages #2 and be highly expressed in RNA from tassels or ears of *ra1* mutants but down-regulated in RNA of *ub1* tassels. Although we will emphasize those genes whose sequence suggests a regulatory function, we will also target "unknowns" that have interesting expression profiles. To bias our samples towards mutations affecting only the inflorescence, the final 100 will include those having expression patterns preferentially restricted to inflorescences. If necessary, this could be determined by routine RNA blots, but we propose arraying several hundred candidates (selected by the criteria indicated above) and hybridizing arrays with RNAs from various organs (roots, shoots, leaves, endosperm, embryo, ear and tassel).

III 5b. Map Positions.

We will blast the sequences of our gene cluster with the rice genomic sequencing project. This analysis may provide a tentative map position if rice and maize are syntenic in that particular chromosomal region. Many ZmDB genes are being placed to map position by different groups (P. Schnable, High-throughput mapping tools for maize genomics; E. Coe, Comprehensive Genetic, Physical, and Database Resources for Maize; R. Phillips, A Radiation Hybrid and Cloning System for the Genetic and Physical Mapping of the Corn Genome). We will check these web sites monthly for the map position of our candidate genes. If the candidate genes are not mapped, we will map them ourselves with the Mo17/B73 recombinant inbred population. For most genes, this requires two Southern blots per gene.

III 5c. Obtaining Gene Knockouts.

We have two avenues to obtain knockouts (MTM and *RescueMu*), both relying on the *Mutator* transposon. *Mutator* (*Mu*) remains the most effective transposon tagging element in maize due to high

copy number, high forward mutation rate, and the fact that the elements preferentially insert into genes and move throughout the genome [48-50]. We will use a reverse genetics approach [51, 52] that relies on the fact that in a large population of *Mu* plants, there is a likelihood of finding an insertion in any given gene. PCR is used to amplify fragments between *Mu* primers and primers from the gene. DNA from a large number of plants is pooled to limit the number of PCR reactions required to find an insertion.

MTM (Maize Targeted Mutagenesis http://mtm.cshl.org/) is an NSF-funded reverse genetics project led by Martienssen and Freeling. DNA has been prepared from 44,000 F1 plants in pools, and seed has been harvested from the resulting F2 ears. So far, 23 genes have been screened for insertions and 14 have been verified in the F2, giving a 61% success rate (B. May and R. Martienssen, unpub.). This is a low estimate, since many of the genes had previously failed in the Pioneer TUSC *Mu* knockout system. Using previously untested genes, the Schmidt and Hake labs have each had success with MTM. Sequence is sent to the MTM curator at CSHL, Bruce May, who designs primers, carries out the PCR, and verifies that insertions are germinally transmitted.. Seeds are sent to the investigator and to the stock center. The cost is $1000 per gene, which covers the cost of PCR reagents and primers and results in 2-3 alleles on average.

We will also screen the *M u* plants from the Maize Gene Discovery project (www.zmdb.iastate.edu). By summer 2002, 15 grids (about 2,000 plants per grid) will be available. We will screen DNA from these fields for insertions of normal *Mu* elements or genetically engineered *RescueMu* elements (designed for plasmid rescue). Hake and Schmidt are directly involved with generating these resources and so are very familiar with them.

<u>Specific Aim 6.</u> Screen *Mutator* and EMS Populations for New Inflorescence Mutants

We will screen several mutagenized populations to look for new inflorescence phenotypes. These will include a recently generated *Mutator* insertion population in inbred W22 and an EMS-generated population in inbred B73, produced as part of this project. Defined inbreds are preferable to mixed populations for detecting subtle phenotypes and for comparing a mutant to its wild type sibling using microarray-based expression profiling. The populations will be grown in Illinois under the supervision of co-PI Rocheford. Illinois offers abundant field space and support staff for this endeavor and is centrally located for all participants.

<u>III 6a. Mutator Populations.</u>

We will screen a *Mu* population generated by Don McCarty and colleagues (see letter of collaboration). They presently have 5000 self-pollinated ears and, through an NSF-funded genome project, plan to generate 40,000 total. One of our group will travel to Florida to prepare seed packets to be shipped to Illinois. We will screen 20 kernels from each of 5000 ears the first summer and a similar number the second summer. A powerful advantage of this population over other *Mu* populations, such as that generated in the *RescueMu* project or MTM, is that the material is in a uniform, inbred background (W22). Thus, even subtle phenotypes, often masked in mixed backgrounds where alleles at different loci are segregating, can be observed.

Other sources are the Maize Targeted Mutagenesis (MTM) at Cold Spring Harbor, and *RescueMu* populations. A portion of the MTM population is screened every summer by Hake, Martienssen and Jackson lab members for new floral mutants. During the summer of 2001, 5000 F2 families from the *RescueMu* population will be grown by the Maize Stock Center in Illinois. These mutants are in mixed inbred backgrounds and so not ideal for expression profiling or visualization of subtle phenotypes but may still provide new mutants. Mutants we identify will be introgressed into inbred B73. New mutant phenotypes will be posted at the ZmDB web site and available to others through the Stock Center.

<u>III 6b. Mapping.</u>

Inflorescence mutants identified from the McCarty W22 population will be crossed to three different inbreds, Mo17, B73 and W22. Crosses to B73 and Mo17 will provide material for mapping. The F1 crosses will be selfed in Hawaii and the F2 populations grown the following summer in Illinois. Tissue will be collected from pooled mutants and pooled normal siblings. Bulked segregant mapping [53] will be carried out using SSR primers and protocols [54] modified and optimized in the Rocheford lab. We will screen bulked DNA from different F2s with a series of primers that provide basic coverage of the

genome. Once a potential association is detected, we will put 4-5 SSR markers in that region on approximately 100 individual F2 plant DNA samples using programs such as JoinMap or MapMaker [55, 56] to establish a map location. We also expect to find new alleles of already identified mutants, so will plant reference mutants listed in Table 1 in the screening field. The W22 *Mu*-induced mutants will be crossed with the appropriate reference allele, and the F1 will be scored for complementation. For new alleles, we will see the mutant phenotype segregating, though we will be careful to further test for non-allelic non-complementation, which does occur with fasciated mutants. Mutations that do complement can be selfed to generate material for double mutant analysis at a later point. These double mutants may provide insight into the function of the newly identified gene.

II 6c. Microarray Profiling.

Crosses to W22 will reduce the number of other *Mu*-induced mutations present in the family while keeping the mutant of interest in W22. These crosses will be selfed in the winter nursery and the F2 planted the following summer. We will confirm heritability of the mutation and determine if it is a new mutation. If so, we will collect material for RNA expression profiling. Mutants that show similar expression changes on the array are likely to function in related developmental programs. Depending on time and technology, we may analyze only 2-3 new mutants. The results, however, will not only inform us as to the function of that particular gene, but should also provide additional information about the clones on the microarray.

III 6d. EMS Mutagenesis and Screening.

In addition to screening for mutants caused by *Mu* insertions, we will also make and screen an ethyl methane sulfonate (EMS)-induced population [57]. EMS is about 10 times more effective at making mutations than *Mu*. We have chosen a different inbred for this mutagenesis given that some phenotypes are background dependent [58]. B73 pollen will be mutagenized and crossed onto B73 females. 3000 F1 plants will be selfed and the F2 families screened the following summer in Illinois. New mutants will be analyzed as described above for *Mu.* We will backcross into B73 to eliminate other mutations and cross to Mo17 to generate a mapping population. Because these mutations will be identified in a defined inbred, these new mutants will also be suitable for gene profiling.

Specific Aim 7. Identification and Characterization of QTL for Ear and Tassel Architecture

A major goal for plant breeder/geneticists has been identification of QTL controlling quantitatively inherited traits, and uncovering genes underlying QTL [13, 14, 59-61]. Our goal is to detect QTL for tassel and ear morphology and use new genomic resources to identify candidate genes associated with QTL. These genes may correspond to known mutants [62], or to genes that do not have a mutant phenotype due to redundancy, an essential role in viability, or simply because a mutant allele has not been found yet. The QTL approach therefore is complementary to mutagenesis. From a practical standpoint, some inflorescence mutants are difficult to work with due to infertility, whereas QTL may be weak alleles that do not have this problem.

We will use existing mapping populations to develop an extensive inflorescence architecture QTL data. We will develop a series of near-isogenic lines (NILs) [63] to confirm a small subset of QTL, focusing on those with the largest effect. The NILs will also be used for allelism tests, and selected microarray analyses, which could identify candidate genes for the QTL, assuming that differences in gene transcript level underlie QTL. Candidate genes strongly associated with QTL in mapping populations and NILs will be sequenced in the parents of the corresponding mapping population. This will also provide baseline sequence data on non-mutant alleles.

For two cloned genes, *ra1* and *fae2*, we will examine the relationship between these mutant loci and QTL that map to the same chromosome regions. Comparisons will be performed at sequence, expression, and phenotypic levels. These studies will take advantage of ongoing genetic studies on *ra1* and *fae2* funded by other sources. *ra1* and *fae2* will therefore bring specific resources into the larger genomics effort, and provide a paradigm for functional genomic analysis of mutant and QTL alleles.

III 7a. Comprehensive Identification of QTL with Mapping Populations.

Only a subset of loci influencing quantitative variation are likely to be polymorphic between any two parents. To identify a representative number of QTL that influence tassel and ear architecture, we will therefore use a set of populations involving genetically diverse parents. Included in this set is the Illinois Mo17xB73 (IBM) population, which will be central to the entire project. IBM has become the main mapping population for the maize community and the NSF funded Missouri Genome Project (www.agron.missouri.edu/top.html), which includes physical and comparative mapping. Most of our mutants show a difference in phenotype between B73 and Mo17. For example, in B73, the *bif2* phenotype is stronger, while more branched tassel phenotypes (eg. *ra1*), are dampened. Similarly, there are differences in the inbreds themselves: B73 tassels are sparse with short branches, and Mo17 tassels are more robust; B73 ears have more kernel rows and kernels per row than any other inbred with which we work, whereas Mo17 has relatively low values for these traits (www.ars-grin.gov).

The other populations we will use for inflorescence QTL studies were developed at Illinois for study of other traits and have already been marker genotyped. Therefore we can simply grow the lines, make phenotypic measurements, and perform statistical analyses to identify QTL. In some cases we may need to add SSR markers to more finely map a particular QTL. The mapping populations include those derived from Illinois Long Term Selection Experiment strains: Illinois High Oil (IHO), Low Oil (ILO), High Protein (IHP) and Low Protein (ILP). The IHOxILO, IHPxILP and Mo17xB73 mapping populations all underwent cycles of random mating at the F2 stage during their development, to increase recombination and allow more precise fine mapping [64] (Dhijkhuzien, Rocheford and Dudley, in preparation). Some mapping populations are at advanced stages of inbreeding (e.g. F6-F7 lines approx 96.8-98.4% homozygous), and are essentially recombinant inbred lines, which increases the power for detection of QTL and facilitates the rapid extraction and development of near-isogenic lines (NILs) to confirm QTL (Section 7c) [65, 66].

III 7b. Measurement of Tassel Traits and Statistical Analysis

Each QTL population will be grown on the South Farms of the Illinois campus in multi-year replicated trials according to well established procedures for field QTL studies of maize agronomic traits [16, 67, 68]. Initial results will influence which populations will be evaluated subsequently to complement findings to that point.

Given the diverse phenotypes of maize mutants (Table 1) and the broad variation in inflorescence architecture in other cereals, our measurements will target specific tassel and ear architecture parameters: length of main rachis (main axis of the tassel), distance from tip of rachis to first long branch, ratio of rachis area with branch meristems versus spikelet pair meristems, primary long branch number, length of the lowest long branch, presence of secondary long branching, spikelet density on the main rachis and the lowest long branch, frequency of unpaired spikelets (more or less than two), primary long side branch angle, number of days from initiation of anthesis to completion of pollen shed, tassel weight after pollen shed; number of kernel rows per ear, number of kernels per row, total number of kernels per ear, and shelled cob weight. Ears will be given a binary score for row organization, since disorganized rows indicate altered spikelet meristem function. Not all measurements will be taken on all populations. Rocheford has a large field pollinating crew that take measurements directly in the field, saving considerable time. Measurements will be taken on 5-7 random tassels in each family row. Ear and spikelet density measurements will be taken after harvest.

Rachis length is required for calculating ratio of long branch meristems to spikelet pair meristems. This transition is relevant to phenotypes associated with maize mutants such as *ra1, ra2, ra3,* and *ub1*. Spikelet density and kernels per row is relevant to phenotypes similar to *fae2*, *kn1, bif2, ba1, td1*, and *Fas1* mutants. Frequency of unpaired spikelets is relevant to *ifa*, *ts4*, and *bd1*. Length of the lowest side branch and total number of spikelets on it are relevant to variation in the extent of total spikelet coverage of long side branches among the cereals. Variation in primary branch angle is associated with *ra2* and seen among other grasses. Number of kernel rows is relevant to *fae2* and *td1* and row organization is relevant to *fae2, ra1, ra2, ra3, ts4* and *ifa1*.

Statistical analyses will identify molecular marker-trait associations (QTL) [14, 59]. The goal is to develop a comprehensive set of QTL locations, and to identify a small subset of unique, major QTL for

subsequent analyses. Initially, single factor analysis of variance will be performed with Statistical Analysis System (SAS) software [69]. Subsequent analyses will depend on the structure of the specific population. For populations where a molecular marker map is available or can be easily constructed, composite interval analysis will be performed with programs such as PLABQTL and QTL Cartographer [70, 71]. These programs adjust for QTL segregation in other parts of the genome and may provide more reliable results [72]. Multiple regression models will be developed for all populations, either with simple regression or with interval regression [73]. These programs identify the few loci that contribute the largest effects in a given population. Analyses for epistatic interactions will be performed for significant and non-significant markers, including mapped ESTs and cloned mutants, with the EPISTACY program [74]. For mapped ESTs associated with epistatic interactions, relevant microarray results will be examined. The few QTL selected for further analysis will first be confirmed by NIL analysis (III 7c) [66, 75].

III 7c. QTL Associations with Mutants & ESTs, and QTL Assessment with Near-isogenic Lines.

At this time, maize inflorescence architecture QTL data are limited, but we know of QTL that map close to three known inflorescence mutants. The QTL with the largest effect on branch number (35% of variation) in an IHOxILO standard $F_{2:3}$ mapping population is in the same region as *ra1* [16]. Allelism tests suggest the QTL in IHO may be an allele of *ra1*. A QTL for ear row number in an H99xMo17 mapping population was detected in the same chromosome location as *fae2* [76]. The *fae2* probe maps to the same location as the row number QTL on the H99xMo17 mapping population (M.Lee unpub), making it a candidate gene. Our initial analyses of spikelet density in the IHOxILO $F_{2:3}$ population identified a large QTL linked to marker NPI449, which maps within 2 cM of *td1* in a maize composite map. Together these three candidate genes allow lead QTL studies related to the function of specific meristem subsets.

We expect to identify many QTL in the mapping populations and will focus on some of those with large effects that do not map to known mutants. These will represent additional genes beyond the 19 mapped mutants in Table 1. QTL mapping combined with microarray analyses (Aims 2-4) may generate a significant set of ESTs that are candidate genes for some QTL. We will only pursue the most promising candidates, using our high resolution mapping populations to identify only the small subset of ESTs that are tightly linked to QTL.

Near-Isogenic Lines have been used for genetic analysis of mutants for decades and more recently for study of QTL, including recent efforts leading to cloning two QTL in tomato [60, 61]. We will use NILs to study QTL alleles influencing inflorescence architecture. Since some of our mapping populations use recombinant inbred lines, we can simply identify a few families with rare heterozygotes for markers linked to the QTL, self within this family (15-20 plants), and genotype progeny to generate NIL sublines. Since individual families used to develop NILs are already highly inbred, most background segregation is minimized. The sublines are grown in replicated comparisons to confirm QTL.

We will use the NILs for the *ra1* associated QTL (and possibly *td1* or *fae2*) as a lead microarray study. RNA from paired NIL sublines will be collected from developing tassels at stage #2 (Table 2) and used for expression profiling. Genes whose expression profiles differ between the two NIL sublines should function to orchestrate development of the phenotypic character associated with that QTL. Ideally, one of the genes whose expression differs should represent the QTL itself. The QTL subline expression profile will also be compared with that of *ra1* versus non-mutant sib at the same stage. For NIL of QTL not associated with a cloned, characterized candidate gene, mapping studies of ESTs showing differences in NIL microarrays would help narrow the candidate genes. The candidates may represent the QTL itself, and/or identify downstream targets of the QTL.

For a few of the most promising candidate genes, such as *ra1* and *fae2,* the genomic sequence of parental alleles from the relevant QTL population will be cloned and sequenced to identify allelic differences. However, ultimate proof that a QTL corresponding to a mutant locus or EST is indeed the same gene, may require transformation, and is outside the scope of this proposal. The QTL work will benefit from resources developed as a part of other aims, e.g. microarrays, and will provide biological information, such as new candidate inflorescence genes that could correspond to mapped ESTs.

Specific Aim 8. Analysis of Inflorescence Genes in Related Grass Species

Cloned genes with a known function, as defined by mutant phenotype, will provide the basic tool for cross-species comparisons. We will investigate up to 10 gene sets beginning with the isolation of putative

orthologs to *fae2*, *ra1* and *ids1*. Other genes will be investigated as they become available. *fae2* encodes an LRR-receptor protein similar to *CLAVATA2* of *Arabidopsis*, and *ra1* and *ids1* encode putative transcription factors of the C2H2 zinc-finger and *AP2*-domain classes, respectively [11] (EV and RM unpub, DJ unpub). In maize, *fae2*, *ra1* and *ids1* regulate, respectively, inflorescence meristem size and thus branch number, spikelet vs. branch meristem identity, and number of florets in a spikelet. These characters vary among grasses and may have played a role in diversification of grass inflorescence architecture. Using maize as our reference, we hypothesize that relative differences in orthologous gene function, manifested as differences in gene sequence or expression, will correlate with differences in related grass inflorescence architectures. The Kellogg lab has already isolated fragments of *kn1*, *lfy* and *tasselseed2* from a variety of grass species, and the same approach can be applied to the genes studied in this proposal.

Our "target cereals" will include oat, barley, sorghum, pearl millet, common millet, foxtail millet, rice and maize. These species were chosen for their economic importance, and for their range of inflorescence morphologies. To facilitate gene identification we will prepare a resource of genomic DNA libraries from the first 6 species, but rely on existing resources for rice and maize. Maize genomic libraries are already on hand, and rice sequences are available via arrangements with private companies, although we anticipate that public rice sequences should become available during the time frame of this grant. We will use these libraries to identify candidate gene orthologs including regulatory regions. Once full gene sequences are in hand, sequence comparisons will reveal whether coding regions and/or regulatory regions are conserved. This will provide a rich and unparalleled resource for detailed comparison of gene divergence across 60 million years of grass evolution.

We anticipate partitioning cloned genes into those whose expression is conserved among the grasses and those whose expression varies. Genes with conserved expression will be of particular interest to breeders working in the cereals. Non-conserved expression patterns reflect evolutionary variation, indicating that gene function may have changed over the evolutionary history of the grasses. These data can suggest alternative interpretations of function, and may indicate which genes have been important for the generation of morphological diversity. Taken together, our sequence and expression analyses will determine the extent of allelic diversity, and suggest regulatory mechanisms that might not be obvious from study of model systems alone.

III 8a. Genomic DNA Gel Blots and Library Preparation

DNA gel blot hybridizations will be performed under reduced stringency [77] to determine whether a single ortholog or small gene family is present. Copy number will guide library screening and ultimately be relevant to determination of orthology. We will start with relatively long probes derived from cDNAs of *fae2*, *ra1* and *ids1* from maize, including conserved regions (*e.g.* putative DNA-binding domains). Initial experiments using *ra1* and *fae2* as probes suggest the existence of only a few homologs in several species examined (Fig. 4).

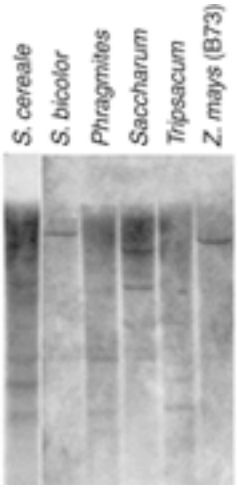

Figure 4
***ra1* homologs were detected by reduced stringency DNA gel blot hybridization.** This gene-specific maize probe detects homologs in related panicoid grasses sorghum, *Phragmites*, sugar cane and *Tripsacum*, as well as in the phylogenetically distant pooid grass, rye. BamH1-digested genomic DNA was electrophoresed, blotted, hybridized at 50˚C, and washed at 55˚C in 0.2x SSC, 0.2% SDS.

Genomic libraries will be constructed with partially digested *SauIII*A DNA cloned into a commercial phage lambda cloning vector. We will screen sufficient recombinant phage for full genome coverage by adjusting the number according to each species' C-value (http://www.rbgkew.org.uk/cval/database1. html).

Gene fragments homologous to the species being investigated can be isolated by PCR (see III 8b), and may provide better hybridization than heterologous maize sequences. Homologous sequences can be tested on Southerns by hybridization at high stringency, and similarly used in library screening, thus adding specificity to this step.

III 8b. PCR Primers and Development of Probes.

All genes identified in maize (including those from microarray experiments) will immediately be compared to rice genome sequences by a standard BLAST search. Alignment of the maize and rice sequences will identify regions of conserved sequence outside those common to all members of a gene family. These sequences can be used to develop sets of PCR primers that will amplify fragments of the gene. These can be chosen to be gene specific, or to amplify all members of a particular gene family. By designing several sets of primers they can be chosen to amplify across introns (if any) or not, depending on the goal of the experiment. The Kellogg lab has used such an approach to design primers to *kn1* that avoid the homeodomain and thus appear to be gene specific; we have also designed primers to *ts2*. These are only slightly degenerate, and have been used successfully to amplify genomic sequences from most grasses tried. We have also used the highly degenerate, nested primers designed by Frohlich for *leafy*, again on a broad range of grasses. Primers can also be designed easily using CODEHOP [78]. The fragments generated by PCR can be used to screen libraries, as probes in Southern and Northern blots, and for in situ hybridization.

III 8c. Sequencing and Analysis of Genomic Clones.

Our goal is to obtain full-length gene sequence, including transcribed and flanking regions. Genomic library screening will rapidly produce a collection of overlapping clones from each species derived from one to several different loci, depending on hybridization conditions. Clones containing common elements will be identified to eliminate redundant sequencing effort. One quick way to begin identifying individual genes is to use the DNA fragments produced in the intial PCR studies above (III 8a). This strategy has already proven successful in isolating a *ra1* homolog from sugar cane in Martienssen's lab. Sequences will be analyzed by BLAST and gene-finding programs and compared to sequences of RT-PCR products to verify intron-exon structure. In particular, we will make use of the SplicePredictor and GeneSequer programs developed by co-PI Brendel to identify potential splice sites and gene products by spliced alignment with cognate or related ESTs or putative protein homologs [41, 42]. Kellogg will conduct phylogenetic analysis, including phylogenetic trees constructed using maximum parsimony and maximum likelihood with programs in PAUP*4.0 [79]. Sequencing infrastructure resources are available at all participating institutions, and are especially well developed at Cold Spring Harbor. The Kellogg lab also has their own sequencer and has considerable expertise in optimizing sequencing reactions for non-model organisms. Thus, we expect to do most sequencing on-site.

III 8d. Expression Analysis.

To test whether changes in timing or expression pattern of a particular gene correlate with differences in inflorescence architecture across the cereals, we will use RNA gel blot analysis and in situ hybridization to developing inflorescences.

a.) RNA gel blots. Hybridization experiments will use species-specific, gene-specific probes from transcribed regions outside of highly conserved sequence motifs. We will first develop a basic expression profile using an RNA gel blot for each cereal, probed with each candidate ortholog. Plant material will be separated into roots, leaves and a few stages of inflorescence development. The Kellogg lab has extensive SEM data on development in the millets, and in sorghum, and information on oats and barley is available in the literature. Development in these species is standardized by size of inflorescence rather than by time as in maize. For each species, we know the size of the inflorescence when primary, secondary, tertiary and higher order branches are being initiated, and also the size at which spikelet parts and floral organs

form. Stages will be chosen to correspond to those investigated in maize (Table 2). RNA gel blot experiments will provide basic parameters for in situ hybridizations, and allow comparison with the known expression profiles of the maize genes.

For a given gene, expression levels between different species will be roughly compared by normalizing to the constitutively expressed *ubiquitin* or *actin* genes. Comparable expression levels in all species would be consistent with a hypothesis of conserved function. Reduced or absent expression could mean that expression is below the level of detection or that the gene is not active in the species investigated. In the former case, we should be able to detect the gene with RT-PCR. Reduced or lack of expression would be a marked departure from the easily detected expression of *fae1*, *ra1* or *ids1* in maize, and would be a significant regulatory difference. A gene may be non-functional in some species due to mutations, which would be a direct explanation for lack of expression, and should be observable in the genomic sequence. We may find duplications of genes investigated; if so one copy may accumulate mutations rapidly because it is no longer constrained. This can be seen on a gene tree, and the apparent difference in evolutionary rate tested statistically. A rigorous study to demonstrate a change in gene function will require experiments transferring genes among and between species, and is beyond the scope of this grant. Our data, however, will identify alleles and species most valuable to test by such intensive investigation.

b). Comparative in situ hybridization. To determine if changes in location or developmental timing of gene expression are associated with differences in inflorescence architecture, we will use in situ hybridization with gene specific probes to inflorescence sections. Several of our labs (DJ, SH, RS) are skilled with in situ techniques [80, 81]. Standard in situ methods have been used successfully in rice and barley apices [82, 83] so we do not expect technical difficulties. If necessary we will modify important parameters such as fixative and fixation time, and degree of protease predigestion. The highly expressed gene *kn1* will be used as a positive control; orthologues have been cloned from rice and barley and gene specific fragments have been amplified from all other species to be investigated. Most developmental genes, including *ids1* [11], are expressed in restricted domains and are generally easy to detect by in situ. Our in situ results will reveal tissue specific and temporal patterns of expression that may correlate with differences in inflorescence architecture. For example, based on our (Doust and Kellogg) unpublished results which suggest that the bristles of foxtail millet represent spikelets with indeterminate growth, we hypothesize that *ids1* is down-regulated in these structures given the known expression pattern of *ids1* in maize. Detailed expression patterns of *ra1* and *fae2* are presently being characterized by in situ hybridization (EV and RM, DJ). Like *ids1*, each gene is expressed at high levels and is relatively easy to detect by in situ hybridization. Because we will have some idea of gene function from mutant phenotypes, and will know the expression patterns of our candidate genes in maize, this will allow us to formulate specific and testable models about the role of our chosen genes in controlling cereal inflorescence architecture.

IV. Role of Participants

Although already indicated in the Project Description, we will outline the role of participants here. Schmidt will prepare the normalized libraries. Hake and Schmidt will work together on RNA preps and with Martienssen, Vollbrecht and Jackson on microarraying experiments. Brendel will have responsibility for bioinformatics and data management. He will implement microarray data visualization and analysis tools that can be accessed over the web by project participants or mirrored locally. Additionally, he will advise the participants in all technical and statistical aspects of the analysis. Rocheford will have primary responsibility for organizing the mutant screens in Illinois, but different groups will be responsible for assembling the seed packets, and all members will participate in mutant screens. Martienssen will be responsible for overseeing the MTM knockouts, while Hake will be responsible for identifying knockouts in *RescueMu.* Schmidt, Hake, Martienssen and Jackson will share the task of introgressing and selfing the EMS and *Mutator* mutants. Rocheford will have responsibility for establishing QTL and mapping genes to Mo17/B73 populations. He will also develop PCR methods to map the mutants we identify by phenotype. Kellogg will make genomic libraries with help from Schmidt. She will have primary responsibility for the comparative phylogenetic sequence and expression analysis of *ids1* and other genes as they become available. Cloning and sequencing of *ra1* homologs in other species is already underway

in Martienssen's lab and sequencing of *fae2* homologs is being pursued by Jackson, but Kellogg will collaborate with them on phylogenetic aspects of the analysis.

V. Training and Diversity

The proposed studies will provide postdoctoral fellows, graduate and undergraduate students a diverse and broad education, including classical and molecular genetics, developmental and evolutionary biology and state of the art training in genomics and bioinformatics. All locations offer excellent departmental seminar programs, and participation in journal clubs and lab-group meeting provides opportunities for students to hone speaking skills. In our effort to foster the broadest training, we will encourage our students to interact with members of the other groups. The interactions will be fostered by the yearly Maize Genetics Conference, trips to Illinois during the summer mutant screens, workshops, CSHL short courses, and visits to the other labs. A number of participants will go to Illinois South Farms for part of July to help with crosses, targeted screens, and phenotypic analysis for QTL studies. This affords opportunities to interact with the maize geneticist/breeders and members of the Maize Genetics Stock Center. When possible, visiting members of the group will be housed by the host labs graduate students, postdocs, or faculty to facilitate interaction. Part of a workshop or group meeting may be held at facilities such as Banbury Center (CSHL) or Allerton Conference Center (IL) in which participants stay in close retreat-style quarters with plenty of opportunity for interaction. In addition, we plan to have an annual workshop for research participants. The location and topic of these workshop/meetings will rotate; St. Louis - phylogeny and grass morphology; Iowa State - bioinformatics; Illinois - quantitative trait analysis (possibly including CPSC 430 short course in Molecular Marker Data Analysis); Berkeley - Microscopy; CSHL - Genome Analysis, taking advantage of a short course available, such as the NSF funded course in Plant Genetics. These meetings would have both a training focus and organizational purpose.

The recruitment and diversity plans are designed to take advantage of unique resources and opportunities associated with each institution. Some institutions have access to centers promoting outreach while others are fortunate to be in large, culturally diverse population centers. Each institution has a diversity plan in place for both undergraduate and graduate education. Each institution will seek to hire members of under-represented groups, with those near large metropolitan areas more likely to do so, and some have track records of employing minorities in various capacities (e.g. UM Saint Louis, U. of Illinois, UC Berkeley, UC San Diego). Summer lab and field work provide an excellent opportunity to train undergraduate students in genetics and basic molecular biology techniques (DNA and RNA isolation, gene cloning and molecular genotyping). As nearly all the PIs are involved in graduate and undergraduate teaching, we have direct contact with hundreds of students every year. This allows first hand knowledge of students through which we can identify and recruit minority talent. In addition, outreach efforts can involve the Preuss High School at UCSD for talented minorities, with a PI giving an annual talk and discussing opportunities to work in the summer at UCSD and/or other locations. Having minority members attend the workshop or short course at an institution that traditionally does not have strong minority representation will serve to enrich the diversity exposure at that site. The Biodiversity program at UMSL and Missouri Botanical Garden attracts a number of international students of non-Caucasian background, some of these students will come to Illinois for mutant screens, and attend workshops. Other opportunities include inviting local high school teachers to participate in field research or one of our meetings. For example, in coordination with The DNA Learning Center (an exhibition center at CSHL with community outreach activities) we will invite local high school teachers and their students to participate in short workshops in which we describe the overall goals of our project and highlight opportunities for their participation in field or laboratory research.

When appropriate, we will apply to the NSF and other funding agencies to further support minority high school and undergraduate internships and travel, and minority graduate fellowships.

References

1. Weatherwax, P., The development of the spikelets of Zea Mays. *Bull. Torrey Bot. Club*, *44*, 483-496 (1917).
2. Veit, B., R.J. Schmidt, S. Hake, and M.F. Yanofsky, Maize floral development - new genes and old mutants. *Plant Cell*, *5*, 1205-1215 (1993).
3. Jeong, S., Trotochaud, A. E., and S.E. Clark, The Arabidopsis *CLAVATA2* gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell*, *11*, 1925-1933 (1999).
4. Gernert, W.B., A new subspecies of Zea Mays L. *Am. Nat.*, *46*, 616-622 (1912).
5. Nickerson, N.H. and E.E. Dale, Tassel modifications in Zea mays. *Ann. MO Bot. Gard.*, *42*, 195-211 (1955).
6. Postlethwait, S.N. and O.E. Nelson, Characterization of development in maize through the use of mutants I. The polytypic (Pt) and ramosa-1 (ra1) mutants. *Am. J. Bot.*, *51*, 238-243 (1964).
7. Neuffer, M.G., E.H. Coe, and S.R. Wessler, *Mutants of maize*. 2nd ed. Plainview, New York: Cold Spring Harbor Laboratory Press (1997).
8. Irish, E.E. and T.M. Nelson, Development of *tassel seed 2* inflorescences in maize. *American Journal of Botany*, *80*(3), 292-299 (1993).
9. Kempton, J.H., Heritable characters in maize: XLVII Branched silkless. *J. Hered.*, *25*, 29-32 (1934).
10. Colombo, L., G. Marziani, S. Masiero, P.E. Wittich, R.J. Schmidt, M.S. Gorla, and E.M. Pe, *Branched silkless* mediates the transition from spikelet to floral meristem during *Zea mays* ear development. *Plant Journal*, *16*, 353-363 (1998).
11. Chuck, G., R. Meeley, and S. Hake, The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. *Genes and Development*, *12*, 1145-1154 (1998).
12. Geraldi, I.O., J.B. Filho, and R. R. Vencovsky, Estimates of genetic parameters for tassel characters in maize (Zea mays L.) and breeding perspectives. *Maydica*, *30*, 1-14 (1985).
13. Lee, M., DNA markers and plant breeding programs. *Adv. Agron.*, *55*, 265-344 (1995).
14. Lynch, M. and B. B. Walsh, *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates Inc. (1998).
15. Joshi, S.P. and P.K. Ranjekar, Molecular markers in plant genome analysis. *Current Science*, *77*(2), 230-240 (1999).
16. Berke, T.G. and T.R. Rocheford, Quantitative trait loci for tassel traits in maize. *Crop Science*, *39*, 1439-1443 (1999).
17. Shantz, H.L., The place of grasslands in the earth's cover of vegetation. *Ecology*, *35*, 143-145 (1954).
18. Clayton, W.D. and S.A. Renvoize, *Genera graminum*. *Kew Bulletin Additional Series XIII*. London: Her Majesty's Stationery Office (1986).
19. Moore, G., K.M. Devos, Z. Wang, and M.D. Gale, Cereal Genome Evolution - Grasses, Line Up and Form a Circle. *Current Biology*, *7*, 737-739 (1995).
20. Bossinger, G., W. Rohde, U. Lundquist, and F. Salamini, Genetics of barley development: Mutant phenotypes and molecular aspects, in *Barley: Genetics, biochemistry, molecular biology and biotechnology*, P.R. Shewry, Editor, Alden Press Ltd: Oxford, UK. 231-264 (1992).
21. Ayyangar, G.N.R., T.R. Narayanan, and T. Narayana Rao, The inheritance of characters in Setaria italica (Beauv.), The Italian millet. Part IV, Spikelet-tipped bristles. *Indian J. Agricultural Sci.*, *3*(3), 552-556 (1933).
22. Huang, H. and H. Ma, FON1, an Arabidopsis gene that terminates floral meristem activity and controls flower organ number. *Plant Cell*, *9*(2), 115-34 (1997).
23. Soreng, R.J. and J.I. Davis, Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Bot. Rev.*, *64*, 1-85 (1998).
24. Kellogg, E.A., Relationships of cereal crops and other grasses. *Proc Natl Acad Sci U S A*, *95*(5), 2005-2010 (1998).

25. Kellogg, E.A., Molecular and morphological evolution in Andropogoneae, in *Proceedings of the Second International Conference on the Comparative Biology of the Monocotyledons. Vol. 2. Symposium on Grass Systematics and Evolution.*, J.E. Everett and S.W.L. Jacobs, Editors, CSIRO: Melbourne (2000).
26. Mathews, S., R.C. Tsai, and E. Kellogg, Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *Amer. J. Bot.*, *Submitted*, (2000).
27. GPWG, G.P.W.G., A phylogeny of the grass family (Poaceae), as inferred from eight character sets, in *Proceedings of the Second International Conference on the Comparative Biology of the Monocotyledons. Vol. 2. Symposium on Grass Systematics and Evolution.*, J.E. Everett and S.W.L. Jacobs, Editors, CSIRO: Melbourne (2000).
28. Evans, M.W. and F.O. Grover, Developmental morphology of the growing point of the shoot and inflorescence in grasses. *J. Agric. Res.*, *61*, 481-520 (1940).
29. Moncur, M.W., *Floral initiation in field crops.* Melbourne: CSIRO (1981).
30. Kellogg, E.A., A model of inflorescence development, in *Proceedings of the Second International Conference on the Comparative Biology of the Monocotyledons. Vol. 1.*, K. Wilson and C. Quinn, Editors, CSIRO: Melbourne (2000).
31. Doebley, J. and A. Stec, Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics*, *134*, 559-570 (1993).
32. Doebley, J., A. Stec, and C. Gustus, *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, *141*(1), 333-346 (1995).
33. Doebley, J., A. Stec, and L. Hubbard, The evolution of apical dominance in maize. *Nature*, *386*, 485-488 (1997).
34. Wang, R.-L., A. Stec, J. Hey, L. Lukens, and J. Doebley, The limits of selection during maize domestication. *Nature*, *398*, 236-239 (1999).
35. Hillier, L.D., G. Lennon, M. Becker, M.F. Bonaldo, and B.E.A. Chiapelli, Generation and analysis of 280,000 human expressed sequence tags. *Genome Research*, *6*(9), 807-28 (1996).
36. Kopczynski, C.C., J.N. Noordermeer, T.L. Serano, W.Y. Chen, J.D. Pendleton, S. Lewis, C.S. Goodman, and G.M. Rubin, A high throughput screen to identify secreted and transmembrane proteins involved in Drosophila embryogenesis. *Proc. Natl. Acad. Sci. (USA)*, *95*(17), 9973-8 (1998).
37. Cho, R.J. and M.J. Campbell, Transcription, genomes, function. *Trends in Genetics*, *16*(9), 409-415 (2000).
38. Duggan, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent, Expression profiling using cDNA microarrays. *Nature Genetics*, *21*(1 Suppl), 10-4 (1999).
39. Lockhart, D.J. and E.A. Winzeler, Genomics, gene expression and DNA arrays. *Nature*, *405*, 827-836 (2000).
40. Gai, X., S. Lal, L. Xing, V. Brendel, and V. Walbot, Gene discovery using the maize genome database ZmDB. *Nucl. Acids. Res.*, *28*, 94-96 (2000).
41. Usuka, J., W. Zhu, and V. Brendel, Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics*, *16*, 203-211 (2000).
42. Usuka, J. and V. Brendel, Gene structure prediction by spliced alignment of genomic DNA with protein sequences: Increased accuracy by differential splice site scoring. *Jour Mol Biol*, *297*, 1075-1085 (2000).
43. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, *95*, 14863-14868 (1998).
44. Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub, Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA*, *96*, 2907-2912 (1999).
45. Brown, M.P., W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M.J. Ares, and D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA*, *96*, 262-267 (2000).
46. Holter, N.S., M. Mitra, A. Maritan, M. Cieplak, J.R. Banavar, and N.V. Fedoroff, Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc. Natl. Acad. Sci. USA*, *96*, 8409-8414 (2000).

47. Alter, O., P.O. Brown, and D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA*, *96*, 10101-10106 (2000).
48. Walbot, V., The *Mutator* transposable element family of maize. *Genetic Engineering*, *13*, 1-37 (1991).
49. Chandler, V.L. and K. Hardeman, The *Mu* elements of *Zea mays*. *Advances in Genetics*, *30*, 77-122 (1992).
50. Bennetzen, J.L., The Mutator transposable element system of maize. *Curr Top Microbiol Immunol*, *204*, 195-229 (1996).
51. Bensen, R.J., G.S. Johal, V.C. Crane, J.T. Tossberg, P.S. Schnable, R.B. Meeley, and S.P. Briggs, Cloning and characterization of the maize *An1* gene. *Plant Cell*, *7*, 75-84 (1995).
52. Das, L. and R. Martienssen, Site-selected transposon mutagenesis at the *hcf106* locus in maize. *Plant Cell*, *7*, 287-294 (1995).
53. Kesseli, R., H. Witsenboer, M. Stanghellini, G. Vandermark, and R. Michelmore, Recessive resistance to Plasmopara lactucae-radicis maps by bulked segregant analysis to a cluster of dominant disease resistance genes in lettuce. *Molecular Plant-Microbe Interactions*, *6*, 722-728 (1993).
54. Senior, M.L., E.C.L. Chin, M. Lee, J.S.C. Smith, and C.W. Stuber, Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. *36*(6), 1676-1683 (1996).
55. Stam, P. and J.W.V. Ooijen, Software for the calculation of genetic maps. CPRO-DLO: Wageningen (1995).
56. Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daley, S.E. Lincoln, and L. Newburg, MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations (1987).
57. Neuffer, M.G., Mutagenesis, in *The Maize Handbook*, M. Freeling and V. Walbot, Editors, Springer-Verlag: New York. 212-219 (1993).
58. Vollbrecht, E., L. Reiser, and S. Hake, Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, *knotted1*. *Development*, *127*(4), 3161-3172 (2000).
59. Liu, B.H., *Statistical Genomics*. Boca Raton: CRC Press. (1997).
60. Frary, A., T.C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley, fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, *289*, 85-88 (2000).
61. Fridman, E., T. Pleban, and D. Zamir, A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc Natl Acad Sci*, *97*(9), 4718-4723 (2000).
62. Robertson, D.S., A possible technique for isolating genic DNA for quantitative traits in plants. *Jour. Theor. Biol*, *117*, 1-10 (1985).
63. Koester, R.P., P.H. Sisco, and C.W. Stuber, Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines of maize. *Crop Science*, *33*(6), 1209-1216 (1993).
64. Liu, S.C., S.P. Kowalski, T.H. Lan, I.A. Feldmann, and A.H. Paterson, Genome-Wide High-Resolution Mapping By Recurrent Intermating Using Arabidopsis Thaliana As a Model. *Genetics*, *142*, 247-258 (1996).
65. Burr, B., F.A. Burr, K.H. Thompson, M.C. Albertson, and C.W. Stuber, Gene mapping with recombinant inbreds in maize. *Genetics*, *118*, 519-526 (1988).
66. Tanksley, S.D. and J.C. Nelson, Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted gerplasm into elite breeding lines. *Theoretical and Applied Genetics*, *92*, 191-203 (1996).
67. Beavis, W.D., D. Grant, M. Albertson, and R. Fincher, Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor. Appl. Genet.*, *83*, 141-147 (1991).
68. Goldman, I.L., T.R. Rocheford, and J.W. Dudley, Quantitative trait loci influencing protein and starch concentration in the Illinois long-term selection maize strains. *Theorerical and Applied Genetics*, *87*, 217-224 (1993).

69. SAS Institute Inc, SAS/STAT. SAS Institute Inc: Cary, NC (2000).
70. Utz, H.F. and A.E. Melchinger, *PLABQTL: A program for composite interval mapping of QTL.* Stuttgart, Germany: Univ. of Hohenheim, Institute of Plant Breeding, Seed Science, and Population Genetics (1996).
71. Zeng, Z.B., Precision mapping of quantitative trait loci. *Genetics*, *136*(4), 1457-68 (1994).
72. Jansen, R.C. and P. Stam, High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, *136*(4), 1447-55 (1994).
73. Haley, C.S. and S.A. Knott, A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, *69*, 315-324 (1992).
74. Holland, J.B., EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. *Jour of Heredity*, *89*, 374-375 (1998).
75. Melchinger, A.E., H.F. Utz, and C.C. Schön, Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, *149*(1), 383-403 (1998).
76. Veldboom, L.R. and M. Lee, Molecular marker facilitated studies of morphological traits in maize. II Determination of QTLs for grain yield and yield components. *Theor Appl Genet,*, *1989*, 451-458 (1994).
77. Vollbrecht, E., B. Veit, N. Sinha, and S. Hake, The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature (London)*, *350*(6315), 241-243 (1991).
78. Rose, T.M., E.R. Schultz, J.G. Henikoff, S. Pietrokovski, S.M. McCallum, and S. Henikoff, Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. *Nucleic Acids Research*, *26*(7), 1628-1635 (1998).
79. Swofford, D., *PAUP*4.0.* Sunderland, MA: Sinauer Associates (1998).
80. Jackson, D., *In situ* hybridization in plants, in *Plant Molecular Pathology: A Practical Approach*, S.J. Gurr, M.J. McPherson, and D.J. Bowles, Editors, Oxford University Press: Oxford. 163-174 (1992).
81. Jackson, D., B. Veit, and S. Hake, Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development*, *120*(2), 405-413 (1994).
82. Sentoku, N., Y. Sato, N. Kurata, Y. Ito, H. Kitano, and M. Matsuoka, Regional expression of the rice KN1-type homeobox gene family during embryo, shoot, and flower development. *Plant Cell*, *11*, 1651-1664 (1999).
83. Williams-Carrier, R.E., Y.S. Lie, S. Hake, and P.G. Lemaux, Ectopic expression of the maize *kn1* gene phenocopies the *Hooded* mutant of barley. *Development*, *124*, 3737-3745 (1997).

A-1. Intellectual Property

Sequencing will be carried out at the Stanford Sequencing Center (SSC). The SSC will immediately deposit sequences in GenBank. Annotated sequences will be maintained at ZmDB (http://www.zmdb.iastate.edu/). Public sector individuals can request clones through the ZmDB web site for a small shipping and handling fee. The Maize Gene Discovery Project will manage the storage and retrieval of these clones in return for use of the DNA and clones for maize gene analysis. Hybridization profiles from the microarrays will also be published at ZmDB. Given that the PCR reactions for the microarray produces enough DNA for approximately 1000 slides, there will be slides available for purchase by the public. These would also be purchased through ZmDB for a reasonable cost. Aliquots of cDNA libraries would be available for a modest cost from ZmDB. Map positions placed on the Mo17/B73 population would be published by the Missouri Maize Database (www.agron.missouri.edu/). Phenotypes for new mutants will be deposited at ZmDB. Once stocks have had two introgressions into an inbred background, we will provide all seed to the stock center, except for reverse genetic alleles derived from MTM which will follow the guidelines of MTM. MTM lines can be requested from the MTMDB website (http://mtm.cshl.org/) as soon as they are posted, at which time they are deposited in the stock center without waiting for backcrosses. They are then propagated at the stock center where they are available within 12 months. Other resources from the research, such as the sequence of *ids1*, *fae2* and *ra1* alleles in other cereals, and map positions of QTLs would be published in peer reviewed manuscripts. After publication, QTL information will be deposited into the Missouri Maize Database so that it is more readily accessible.

A-2. Management Plan

Coordination. Hake has ultimate responsibility for the project and will make sure that we are on track and that funds are distributed appropriately. If need be, funds or tasks will be redistributed to meet our specific aims. The PI and CoPIs are very well-acquainted with each other professionally and all share a mutual interest in the focus of this project. Many of the participants have either collaborated formally, through government-sponsored projects, or informally, through projects of common interest. Jackson was formerly a postdoc in the Hake lab; Vollbrecht was formerly a graduate student in the Hake lab and is now a postdoc in the Martienssen lab; Hake and Schmidt have collaborated in the past on tassel development; Hake, Schmidt and Brendel are present collaborators on the Maize Gene Discovery grant; Martienssen has collaborated with Schmidt and Rocheford on *ra1* for many years; Vollbrecht, Hake and Schmidt have participated in mutant screens in the summer at University of Illinois with Rocheford; Kellogg and Hake have talked about collaborating on maize inflorescences for years; Kellogg recently took the Molecular Marker Data Analysis short course at Illinois that Rocheford co-teaches. Because we have worked well together in the past and know each other's strengths and weaknesses, we are confident of our ability to collaborate effectively on this project.

Meetings and communication. In addition to emails, PI Hake will talk to each group at regular intervals. We will meet three times a year as a group, once every summer in Illinois, once at the Maize Genetics Conference in March, and once in December at a rotating location. In Illinois, each group will assist with mutant screens, introgressions and allelism tests (see Timeline) during the day and, during the evenings, discuss data, look at digital images of new mutants, and plan for the next few months. The Maize Genetics Conference in March provides an easy and inexpensive opportunity to meet as everyone comes to this conference. Rocheford co-teaches a 12-day "Molecular Marker Data Analysis" course in early June. We will send 1-2 of our personnel to attend this course. We will use the December meeting for both training and management. The training component will vary by site as indicated below:

1) Cold Spring Harbor, to coincide with a genomics meeting
2) Iowa State, workshop in bioinformatics
3) Missouri, phylogenetics workshop and tour of the herbarium
4) Cold Spring Harbor, to coincide with a genomics meeting
5) Berkeley, cytogenetics from Zac Cande and share analysis into wild grasses with Freeling

Project Timeline

Year 1

EMS mutagenesis, selfs in Hawaii	Jackson		
Screen *Mutator* populations in Illinois	All		
Crosses to possible alleles or inbreds for mapping	Rocheford		
Subtracted and normalized libraries	Schmidt		
Sequencing	Walbot subcontract		
Developmental time point RNA preps	Hake		
Manage sequence data at ZmDB	Brendel		
Prepare genomic libraries from cereals and screen	Kellogg w/ help from Schmidt, Martienssen		
Development of NILs from established data	Rocheford		
Mapping populations grown for phenotypic evaluation	Rocheford		

Year 2

Screen EMS mutagenesis populations	All		
Allelism tests and introgressions for *Mu* mutants	Hake	Schmidt	
Map positions for *Mu* mutants	Rocheford		
Tassel/ear cDNAs arrayed by Arizona	subcontract		
Microarray analyses of ear and tassel development	Hake	Schmidt	Mart./Jackson
RNA preps for microarrays of mutants	Hake	Schmidt	
Manage microarray data at ZmDB	Brendel		
Sequence *ra1*, *fae2*, *ids1* from different grasses	Kellogg	Martienssen	Jackson
Design PCR primers to amplify genic regions in grasses	Kellogg		
Generate near isogenic lines/ NIL field evaluation	Rocheford		
Mapping populations grown for phenotypic evaluation	Rocheford		

Year 3

Continue EMS and *Mutator* screens	All		
Continue microarray analyses	Hake	Schmidt	Mart./Jackson
Allelism tests and introgressions for EMS mutants	Martienssen	Jackson	
Map positions for EMS mutants	Rocheford		
Arrays hybridized with RNA from classic mutants	Hake	Schmidt	Martienssen
Manage expression profile data	Brendel		
Initiate knockouts for 30 selected genes	Martienssen	Hake	Schmidt
Continue evaluation of QTL populations and generating near isogenic lines and NIL field evaluation	Rocheford		
Sequence *ra1*, *fae2*, others, in parents of QTL populations	Rocheford	Jackson	Martienssen
Continue cloning, sequencing and expression analyses on targeted cereals	Kellogg	Jackson	

Year 4

Isolate RNA from newly defined mutants for oligo chip	Hake		
Continue knockouts for selected genes	Martienssen	Hake	Schmidt
Examine knockout phenotypes, cross to maize mutants	Hake	Schmidt	
Continue array analyses on mutants	Hake	Schmidt	Mart./Jackson
Clone orthologs of genes with mutant phenotypes from knockout expt. cont. in situ analyses	Jackson	Kellogg	Martienssen
Map genes based on sequence & expression profiling	Rocheford		
Complementation crosses of NILs with mutants	Rocheford		
Manage expression profile data	Brendel		

Year 5

Continue with phenotypic characterizations of knockouts	Hake	Schmidt	Jackson
Continue with microarray analyses on selected mutants	Hake	Schmidt	Mart./Jackson.
Place new mutants in pathway based on cluster analysis	Brendel		
Design future experiments to test cluster analysis	Martienssen	Schmidt	Hake
Cont. sequence/expression analysis of orthologs for genes with mutant phenotypes	Jackson	Kellogg	Martienssen
Sequence candidate genes for QTLs	Rocheford		
RNA from near isogenic lines for oligo chip analysis	Rocheford		
Manage expression profile data	Brendel		

Field planting timelines.

Summer 1: *Mutator* screens, EMS mutagenesis, QTL populations, NIL advancement

Summer 2: EMS screens, plus allelism tests for *Mu* crosses, QTL populations, NIL advancement and evaluation

Summer 3: More EMS and *Mutator* screens, plus allelism tests for EMS crosses, QTL populations, NIL advancement and evaluation

Summer 4: Reverse genetics grow-outs, further introgressions for new EMS and *Mu* mutants, QTL populations, NIL advancement and evaluation

Summer 5: Further introgressions of mutants and construction of double mutants, NIL evaluation.

Integration with Related Funded Projects

V. Brendel is a computational biologist who has developed ZmDB as comprehensive web-accessible database for EST data, phenotypes of maize mutants, microarray protocols and data, and genomic sequence links (funded by NSF). The group's main interest is computational identification of gene structure, with development of programs for species-specific splice site prediction and exon/intron identification by spliced alignment. These resources and his expertise are invaluable to this project. Like Hake and Schmidt, he is a Co-PI on the Maize Gene Discovery Project.

S. Hake studies meristems and their role in plant development (supported by NSF, NIH, USDA). Genetic approaches have been taken to study determinate and indeterminate fates of meristems and meristem maintenance. This approach involves study of and cloning of genes, such as *ids*, which will be used in this project. Ongoing efforts towards cloning of *ra2*, *bif2*, and *td1* supported by other projects will likely provide useful clones during the course of the project. Participation in the Maize Gene Discovery Project (NSF) provides linkage to development of EST library and microarray analyses, which provides a useful foundation for this project.

D. Jackson seeks to gain a better understanding of the control of shoot meristem function in maize, with particular emphasis on phyllotaxy and meristem size (supported by NSF, USDA). His lab is characterizing mutations that cause enlargement of and fasciation of the shoot meristems, most notably the ear infloresence meristem. These efforts have led to the isolation of the *fasciated ear2* gene which encodes a leucine rich repeat receptor like protein. This gene is associated with a QTL for ear row number and provides a useful resource for this project.

E. Kellogg pursues evolutionary research on the grasses which requires a combination of phylogeny, developmental morphology, and molecular genetics (supported by NSF). Experience in studying morphological variation, availability of the grass phylogeny, and ongoing developmental studies provides an excellent context to integrate and interpret the sequencing and expression experiments in this project.

R. Martienssen's research in developmental genetics concerns organogenesis in maize and Arabidopsis (supported by USDA). Over the last 10 years his lab has tagged and cloned *ra1*, a classical mutation in spikelet pair meristem fate, which encodes an EPF zinc finger transcription factor related to the Arabidopsis floral meristem determinancy gene SUPERMAN. The lab has a strong interest in genomics (funded by USDA and NSF) including participating in the Arabidopsis and Rice Genome sequencing projects. The Maize Targeted Mutagenesis (MTM) project (NSF) generated a population of 50,000 plants each with a spectrum of new mutations, which will be used in this project.

T. Rocheford performs genome wide mapping studies for QTL and candidate genes controlling quantitatively inherited traits. He has a series of large mapping populations already used for QTL analysis for other traits, particularly kernel composition (supported by NSF, DOE, USDA) in addition to those listed in the proposal. These populations are of various size, 150–350 progeny, of different genetic parents, W64AxA632, ILPxB73, Tex6xB73 and others, and of inbreeding level, F3 – F7, of which a subset will be used in the later years of this project. No other member of the consortium has expertise in this area.

R. Schmidt studies the regulation of storage protein gene expression in maize (USDA) and is exploring the conservation in floral organ identify gene function between maize and model dicots (supported by NSF). His lab uses a combination of genetics, transposon mutatgenesis, heterologous hybridizations and reverse genetics to identify and clone genes controlling floral morphogenesis in maize. His studies on the evolutionary comparisons between maize and Arabidopsis floral genes, and his experience with generating cDNA libraries for Maize Gene Discovery, provide valuable linkage to components of this project. Like Hake and Brendel, Schmidt is a Co-PI on the NSF-funded Maize Gene Discovery Project, providing additional linkage to the resources available through that project.

A-3. Coordination with Outside Groups

We intend to closely coordinate with other groups that are mapping maize genes, notably the group at Missouri which all of us have interacted with for years (Missouri Maize Project, E. Coe). As indicated in the Project Description, our proposal integrates closely with the NSF-funded Maize Gene Discovery project (PI. V. Walbot, Stanford) for EST sequencing and microarray development. In addition, our proposal will make use of tools developed or being developed by two other NSF-funded projects, "Maize Targeted Mutagenesis" for PCR-based screens to identify insertional mutants (PIs R. Martienssen and M. Freeling, CSHL and UC Berkeley, respectively), and "Functional Genomics of Endosperm Development in Maize" for genetic screens of *Mutator* insertional lines in inbred W22 (PI D. McCarty, University of Florida). We have established collaborations with groups that are building bioinformatics tools and databases in other cereal crops (see attached letters) and we will continually be monitoring their web sites. Rocheford or an individual from Illinois will visit collaborators at John Innes Centre in years 1, 3 and 5. We will apply for NSF and other travel funds to invite some of the scientists at the John Innes Centre, Norwich, U.K., to attend some of our annual workshops.

A-4. Conflict of interest

Included in Biographical sketches, section E, page 2 and sent separately to NSF-PGRP as requested.