INTRODUCTION

The research and educational activities proposed here weave together and address three emerging trends: the distancing of the general public from the practice and economic importance of plant breeding (Knight, 2003), increasing attention to the genetics not only of yield and agronomic traits but of quality traits that enhance a crop’s nutritional value (e.g. entire Jan.-Feb. 2006 issue of Cereal Foods World devoted to barley quality), and the advent of marker systems and statistical analysis methods that enable population-wide association between markers and causal alleles that improve valuable quantitative traits (Yu et al., 2006; URL 1 [see Website References]). Public support for applied plant breeding, as measured by science-years devoted to this activity by state agricultural experiment stations, is declining: from 542 in 1990 to 529 in 1994 (Frey, 1996) to only 420 in 2001 (Traxler et al., 2005; URL 2). This trend has no doubt continued unabated in the lean budget years since then. Ironically, these cuts in effort have most deeply affected the crops that are least-well served by the private sector: between 1994 and 2001, state science-years devoted to genetic improvement of oat (Avena sativa L.) and barley (Hordeum vulgare L.) changed by –40% and –36%, respectively, while for corn and soybean the respective changes were +8% and –4% (Traxler et al., 2005). These declines are problematic because small grains play an important and increasing role in U.S. food supply even as their domestic production weakens, challenging our food security. A potential cause for the declines has been decreasing public awareness of plant breeding, as our nation becomes more urban (Thro and Zankowski, 2003). At the same time, interest in plant breeding as a source of foods with new quality and health attributes has been on the rise. Processors see these attributes as a marketing tool (URL 3), while the general public discovers that quality may have a genetic basis (e.g. Grey, 2006). Though processors are well-aware of the need for public support of plant breeding (see Bair letter of support), we believe that little attempt has been made to bring this connection to the general public. Greater effort in that regard would benefit small grain crops directly by giving them a constituency arguing for scientific attention. Equally importantly, the connection between plant breeding and food quality might speak more directly to the interests of our urban population, creating a new and interested public from which to draw future plant breeders.

Finally, these dynamics are taking place at a time when the tools of plant breeding are rapidly changing: the throughput, cost, and density of DNA markers have changed dramatically, as have the statistical methods for analyzing resulting data. Routine use of high numbers of markers has become feasible, but there is currently no consensus on even the basic framework to use these markers optimally within recurrent selection breeding programs. New statistical techniques have been developed and applied in simulation settings (Meuwissen et al., 2001; Wang et al., 2005; Whittaker et al., 2000; Xu, 2003b), but not confronted to rigorous selection experiments to identify their strengths and weaknesses under true genetic models. The lack of empirical evidence on these techniques represents a serious impediment to their adoption, or to their rejection and further development as demanded by the attributes of real genetics.

At the broadest level, our long-term goals are to enhance the security and nutritional value of grain consumed in the United States by 1) contributing to the genomic and methodological knowledge base needed to efficiently improve grain for functional food attributes, 2) devising, refining, and disseminating efficient breeding methods that optimally use data conferred by cutting-edge biotechnologies, and 3) developing educational materials and programs to stimulate the curiosity and invite the next generation of plant breeders to become involved in this profession. The objectives of this proposal are to 1) identify loci affecting β-glucan quantity and structure through association genetic methods in elite North American oat and use these loci to
compare two competing marker-assisted selection (MAS) methods with phenotypic selection, 2) identify loci affecting β-glucan quantity and structure on oat accessions from the National Plant Germplasm System (NPGS) divergent for β-glucan content and determine whether NPGS accessions carry alleles that complement those of elite germplasm, and 3) to develop and teach curricula to professionals on association-based MAS and to draw intelligent young students toward plant breeding as a stimulating and rewarding career.

Our central research hypothesis is that the number of loci affecting β-glucan content and structure, and the size of their effects, are such that important loci will be detected through association genetic approaches. The importance of β-glucan to the health impact of oat has long been appreciated, and substantial research experience, including unpublished preliminary data of ours, shows that β-glucan content is highly heritable, behaves additively, and is controlled by a limited number of loci, some with substantial effects. These attributes make it an ideal trait for association analysis. The purposeful use of data and germplasm from an elite cooperative nursery, the Uniform Oat Performance Nursery (URL 4), tests a second hypothesis that is critical to the applied value of this research. Namely, that association genetics provides an innovative paradigm for the cooperation among many small breeding programs to leverage each other’s efforts thereby increasing their overall efficiency. Thus, association genetics may contribute in unique ways to the effectiveness of public breeding programs. The eight PDs teamed up in this proposal cover the set of expertise needed to successfully complete this research. We bring together expertise in β-glucan breeding and structure function relationships, statistical genetics and particularly association approaches, high-throughput biochemical analysis of grain, whole-genome expression profiling, bioinformatics, and resident and distance education. We are prepared to address the objectives proposed here, thereby creating knowledge to allow the efficient improvement of oat’s health attributes, a rational and optimal application of association analysis in MAS programs, and the realization of exciting educational media to stimulate quality students to become interested in plant breeding.

In brief, our research objectives will be pursued as follows:

Objective 1. **Association mapping of β-glucan content and structure in elite oat germplasm and replicated comparison of phenotypic with marker-assisted selection.** We will apply association analysis to a population of elite oat lines tested in the USDA UOPN since 1996, represented by about 410 lines of known pedigree. The β-glucan content of lines is known; we will also measure the molecular weight of the polymer. We will genotype lines using DArT markers (Jacquod et al., 2001) developed through a consortium of 15 oat research groups (see Kilian letter of support; URL 5). We will apply both mixed-model association analysis controlling for kinship using pedigree information and whole genome analysis (Meuwissen et al., 2001). Results will feed into two cycles of MAS using both analyses. Conversion of UOPN data files on GrainGenes to a relational database that will also hold genotypic data will provide informatic support and a future resource for the oat community.

Objective 2. **Association mapping of β-glucan content from the National Plant Germplasm System: complementation of elite oat.** The GRIN system contains β-glucan content observations on 5382 oat accessions. We will pick 200 lines from each tail of the distribution to genotype with the marker system described above. Their kinship matrix will be constructed using genotypic data. The tails of the distribution differ by 3% in β-glucan content and this selective genotyping will confer substantial power to the analysis (Darvasi and Soller, 1992). Objective 2 will either confirm loci identified in Objective 1, or to identify new loci. For loci mapped in Objectives 2
but not 1, we will determine whether the elite population carries the allele conferring higher β-glucan content as shown by analysis of the NPGS population. **This Objective will identify not only accessions but alleles from the NPGS oat collection useful for oat improvement.**

Objective 3. *Educational initiatives to pipeline students into plant breeding and to educate professionals.* We will develop an interactive distance short course entitled “Association analysis and its application to plant breeding.” The course syllabus and content will emerge from the research proposed here and from the barley CAP (see Muchlbauer letter of support). The target audience will be plant breeding professionals (see Hable letter of support) and upper-level graduate students in plant breeding nationwide. Two educational activities will further help draw new students to plant breeding. First, a one-hour presentation oriented toward high-school students will be used at recruitment venues organized by ISU. Second, a week-long teaching module on biotechnology and computational biology in plant breeding will be introduced into two lower-level Agronomy courses and one upper-level Food Science course to show plant breeding to be an exciting and rewarding area for graduate study and professionally. **These aims will serve our stakeholders by attracting quality students to plant breeding and by providing continuing education for their employees.**

**REVIEW OF LITERATURE RELEVANT TO THIS APPLICATION**

**β-glucan is a healthy food ingredient.** β-glucans are classified as a water-soluble fiber and occur in certain cereals, such as barley and oat. β-glucan from oat and barley can avert or mitigate a number of diseases associated with a highly refined diet. The U.S. Food and Drug Administration (FDA) approved a health claim that “a diet high in soluble fiber from whole oat (oat bran, oatmeal and oat flour) and low in saturated fat and cholesterol may reduce the risk of heart disease” (21CFR101.81). After review of 37 studies in which oat were consumed the FDA concluded that ≥ 3 g of β-glucan from oat should be consumed daily to achieve a clinically relevant decrease in total serum cholesterol concentration. Barley also contains β-glucan, and the FDA just approved a health claim for it’s consumption (URL 6). These health claims can have enormous economic impact: the retail volume of a nearly one-hundred year-old traditional food product, Cheerios®, jumped 11% in 1999 after marketing based on this claim!

In addition to lowering cholesterol, β-glucan consumption has other benefits. 1) It can improve glycemic response when consumed before high-glycemic foods (Bourdon 1999, Pick et al. 1996), making it useful for treating diabetes (Jenkins et al 2002). 2) β-glucan can assist with weight loss by increasing satiety after food consumption, thus reducing the temptation to overeat (Anonymous 2006). 3) β-glucan has anti-tumor and immune stimulating properties (Estrada et al. 1997, DiRenzo et al. 1991). It binds to receptors on macrophages that may be pre-adapted to fight yeast whose cell wall contains glucans (Lavigne et al., 2006). Most clinical studies have been conducted on β-glucan from oat; however, the β-glucan from barley has been shown to have many similar effects (Bengtsson et al. 1990).

**β-glucan structure is important to its function.** Although β-glucan is believed to be the key oat factor lowering cholesterol, its mechanism is unclear. It may act by increasing viscosity of intestinal contents which decreases bile acid and cholesterol reabsorption (Lia et al. 1997; Andersson et al. 2002; Uusitupa et al. 1997). Viscosity induction may depend on β-glucan solubility and molecular weight (MW) (Kerckhoffs et al. 2003). Our previous research showed that β-glucan from two oat lines had both higher MW and greater variance in MW than a commercial variety (Colleoni-Sirghie et al., 2003b). Solution viscosity of the lines was higher
than that of the variety at the same β-glucan concentration. Nevertheless, the role of MW in lowering serum cholesterol is not well established. Other studies, including ours, suggest that while β-glucan MW may play a role, it alone cannot predict β-glucan cholesterol lowering potency (Yao et al. 2006; Braaten et al. 1994; Kerckhoff et al. 2003). In chicken the cholesterol lowering was linked to the degradation rate of β-glucan in the intestine (Bengtsson et al. 1990, Bach et al. 1990). We showed that the impact of β-glucan on bile acid binding was affected by synergistic interactions with other components (Sayar et al. 2006). Studies to date have not been able to evaluate near-isogenic lines (NIL) for which the β-glucan differed in MW because the tools to create such NIL have not existed. Thus, rigorous study of the role of MW in β-glucan effects is hampered by a lack of breeding tools to create the necessary study materials.

Studies on the fine structure of mixed-linked β-glucan have revealed that about 90% of water soluble fractions were comprised of randomly arranged cellobiosyl and cellotetraosyl units, separated by single β-(1-3) linkages with the remaining 10% consisting of up to 10 or more adjacent β-(1-4) linkages (Staudte et al. 1983, Woodward et al. 1983, Vårum and Smidsrød 1988). We showed that widely divergent oat varieties and experimental lines had the same molar ratio of cellobiosyl /cellotriosyl units after complete hydrolysis with lichenase. Nevertheless, they had different lichenase specificities, as evidenced by different end product appearance time-courses (Colleoni-Sirghie et al., 2003a). These specificities may influence β-glucan degradation rates and, therefore, functionality (Bengtsson et al., 1990, Bach et al., 1990). The firmest conclusion available from prior research on β-glucan structure is that genetic variability exists for key parameters that influence its health impact. Tools to dissect that genetic variability will enable understanding of β-glucan function to improve the crop’s health benefits.

**Candidate loci for 1-3, 1-4 β-glucan content have been identified.** Cellulose synthase-like (Csl) gene families have been reported to play a vital role in the synthesis of cellulose in higher plants and have been divided into subgroups with designation of CslA to CslH (Burton et al. 2006; Hazen et al. 2002; Richmond and Somerville, 2000). At least 37 Csl genes have been identified in rice (Hazen et al. 2002). Some Csl subfamilies such as CslA, CslC, and CslD are found in all land plants, but CslF and CslH are found only in grasses (Hazen et al. 2002; Keegstra and Walton, 2006). Comparative mapping showed that rice CslF genes coincided with a major QTL explaining variation for β-glucan content in barley (Han et al. 1995; Burton et al. 2006). *Arabidopsis* lacks the CslF genes and does not produce β-glucan. Small amounts of β-glucan were detected, however, in *Arabidopsis* transformed with CslF isolated from rice (Burton et al. 2006).

**Association genetics is an effective approach for MAS.** Identification of QTL in segregating populations developed from crosses between two inbred lines is limited by 1) the expense of generating such populations, 2) their limited diversity, 3) their frequent separation from the process of breeding itself, and 4) the low number of informative meioses occurring in the population (Jannink et al., 2001). A more powerful approach for mapping traits in a collection of germplasm is association mapping (Thornsberry et al., 2001). This method of analysis is based on the non-independence of alleles in a population called linkage disequilibrium (LD). Association mapping circumvents the need for constructing genetic mapping populations and instead utilizes existing populations of genotypes. In such populations, association between marker alleles and causal alleles arises not from experimental crossing but from historical drift and mutation events. Decay of LD between causal loci and markers far from them result from all historical recombinations that have occurred since the associating event. This possibly large
number of recombinations make it possible to map causal loci more accurately through association than traditional linkage analyses (Flint-Garcia et al., 2003).

The genetic distance over which LD is maintained in a population determines the resolution of mapping that is possible and the marker density required for association analyses. In elite barley LD remains significant between markers that are 5-10 cM apart (Kraakman et al., 2004; Rostoks et al., 2006), such that relatively few markers are needed to supply adequate genome coverage. The recently funded barley coordinated agricultural project (CAP) proposes a marker density of about 3000 markers (URL 1) as compared to densities from 500,000 to 1,000,000 projected to be necessary in human (de Bakker et al., 2005; Pe'er et al., 2006). No published analysis of LD exists in oat, though we have preliminary data.

To date, MAS in small grains has been employed to either select for desirable alleles at target regions in early generation populations or introgress alleles using marker-assisted backcrossing. MAS using these approaches has been successful for large effect QTL or traits controlled by single genes (e.g. in barley, Coventry et al. 2003; Collins et al. 2003) and to pyramid several QTL to enhance disease resistance (again in barley, Castro et al. 2003a; 2003b). No reports of MAS in oat exist, principally due to the lack of low-cost markers and the difficulty of mapping in oat due to its allohexaploid nature and chromosomal translocations (Wight et al. 2003).

Besides backcrossing approaches, Lande and Thompson (1990) showed in their seminal paper how DNA marker information could improve estimates of breeding values for the purpose of selection. In this case, a marker score is calculated based on the effects of alleles carried by each line under selection. The marker score per se can be used or combined with the phenotype of the line into an index that accounts for the trait heritability. Simulation studies using a limited number of lines for QTL effect estimation have shown that this approach to MAS is more efficient than phenotypic selection under conditions (low trait heritability and low number of loci affecting the trait) that are not likely to hold for most quantitative traits. The reason for the poor gain in efficiency has been dubbed the “catch-22 of MAS” (Holland, 2004): effective MAS requires accurate QTL estimates, which in turn require accurate phenotypic data. But in the presence of accurate phenotypic data, phenotypic selection itself is effective (Bernardo, 2001; Moreau et al., 2000). The key advantage of association mapping over traditional linkage mapping is that the former can pool vast amounts of phenotypic data across many lines and many breeding programs. This advantage allows it to overcome the catch-22 of MAS.

Summarizing, association analyses have four advantages making them particularly useful in plant-breeding programs. First, they exploit large breeding populations for mapping, increasing the power and accuracy of the process (Schon et al., 2004). Second, breeding programs invest great effort into careful phenotyping by evaluating in multiple, diverse environments. That effort is leveraged through these analyses. Third, the inbreds tested are those of the breeding programs themselves, making the QTL inferences immediately applicable. And finally, to assemble a wide sample of the germplasm and genetic backgrounds relevant for breeding, the association analysis can bring into cooperation many small breeding programs. Thus, it offers a new paradigm for small programs to surmount the difficulties of applying the latest in marker and computational technologies, by pooling resources for variety development, gene discovery, and MAS.

**Two competing association analysis methods exist: mixed-model and whole genome analysis.**

In mixed-model analysis, marker alleles are considered fixed effects and population structure is accounted for by a random effect (Kennedy et al., 1992). Accounting for structure ensures that identified loci are truly associated with the trait rather than with subpopulation differentiation.
Project Narrative

(Pritchard et al. 2000). Fitting the apposite random effect requires determining the kinship of individuals observed (Lynch and Walsh, 1998; Yu et al., 2006), which can be obtained either from molecular markers or pedigree records (Ritland, 2000). Simulation studies have confirmed mixed-model analysis to be useful in both cross- and self-pollinated crops (Arbelbide et al., 2006; Yu et al., 2005). The analysis has also been successful with real data in maize (Parisseaux and Bernardo, 2004) and wheat (Arbelbide and Bernardo, 2006; Breseghello and Sorrells, 2006).

A weakness of these analyses is that they evaluate QTL effects one marker at a time in a process called model selection. This process necessarily entails performing many statistical tests and requires high significance thresholds to identify significant effects, lowering QTL detection power. A corollary is that detected QTL have inflated effect estimates (Beavis, 1994; Schon et al., 2004; Xu, 2003a). Recent developments in model selection seek to avoid it altogether by including all markers as predictors in the model, rather than choosing a “best” set among them. Such a high-dimensional model is complex, however, which may lead to unstable QTL estimates (Jansen, 2001). It is therefore necessary to constrain the allowed effect estimates. In ridge regression, the least squares effect estimators \( \hat{\beta} = (X'X)^{-1}X'y \) are replaced by \( \hat{\beta} = (X'X + \lambda I)^{-1}X'y \) (Whittaker et al., 2000). A high value for the parameter \( \lambda \) causes a penalty for large \( \beta \) thereby avoiding inflated estimates. This approach has strong affinities with the estimation of \( \beta \) using best linear unbiased prediction or using random Bayesian models. Ridge regression is akin to the solution of the random component of Henderson’s mixed-model equations for estimating BLUP (Gianola, 2001), with \( \lambda = \sigma^2_e / \sigma^2_\beta \), where \( \sigma^2_e \) is the residual, and \( \sigma^2_\beta \) is the estimator variance. The Bayesian interpretation amounts to assuming a prior distribution for \( \beta \): \( \beta \sim N(0, \sigma^2_\beta) \). Boer et al. (2002) provide a clear exposition of these relationships.

A drawback of the ridge regression solution for including all markers is that all marker effects are equally penalized. To remove this constraint, Xu (2003b) proposed a hierarchical model that allowed for a different variance for each \( \beta_i (\sigma^2_\beta_i) \). He showed that the posterior distributions of all parameters could be readily estimated using Markov chain Monte Carlo. His method performed well for both real and simulated datasets, though important improvements to the model were proposed by ter Braak et al. (2005). An interesting feature of Xu’s (2003) model is that it severely shrinks marker effects toward zero, more so as the effect becomes small. A consequence of this severe shrinkage is that the model reverses the usual bias of QTL effect estimates: under Xu’s model, small effects are underestimated while little bias is present in the estimates of large effects (Wang et al., 2005). To date, Xu’s model has been applied to linkage data (see for example our preliminary data) but not to association analysis. A very similar model, however, has been applied to association analysis simulations by Meuwissen et al. (2001) in a complex pedigree context. The method worked admirably. Furthermore, there is nothing about the model that requires knowledge of the genetic positions or recombination frequencies of the markers involved. Marker data are simply used as predictors. A surprising aspect of Meuwissen et al.’s (2001) application was that it worked well in the absence of a polygenic effect taking into account variable kinships among individuals in the pedigrees. The explanation for this success is that the polygenic effect accounts for genetic effects in regions of the genome that are not otherwise included in the statistical model. But with whole genome analysis there are no such regions, and correlations in the residuals that otherwise invalidate the analysis do not arise.
PRELIMINARY FINDINGS RELEVANT TO THE PROPOSED RESEARCH.

Levels of linkage disequilibrium in oat.

No published analysis of LD exists in oat. Two lines of evidence suggest that LD in elite oat populations is quite high and extends for long distances. First, we have reanalyzed historic marker data in oat used by O’Donoghue et al. (1994) to determine relationships among North American oat cultivars. Her dataset is relatively small for association purposes, but nevertheless useful. A total of 83 varieties were genotyped using RFLP with 56 probes, of which 40 were mapped in the Kanota x Ogle mapping population (Wight et al. 2003). A total of 239 polymorphic bands were identified, though the allelism of these bands could often not be determined given the allohexaploid oat genome (that is, the same RFLP probe may anneal to more than one and often three different loci). Still, 13 probe pairs were on the same linkage group. The LD between probe pairs was determined while controlling for population structure as determined by the markers themselves using the program Structure (Pritchard et al. 2000). Figure 1 shows the relationship between a measure of LD, the coefficient of determination ($r^2$) between pairs, as a function of their recombination frequency. It also shows the average $r^2$ between pairs of probes determined to be on different linkage groups (for which the recombination frequency is 0.5). Though represented by very few points, the figure is similar to ones observed for barley (Kraakman et al., 2004; Rostoks et al., 2006). If anything, it appears that LD extends further in elite oat than elite barley. A possible mechanism that might maintain LD in oat is the fact that a number of distinct chromosomal translocation patterns exist in elite oat, and in fact, oat may generate translocations on a regular basis (Jellen and Beard, 2000; Jellen et al., 1993; Singh and Kolb, 1991). These translocations hinder recombination and could therefore slow the decay of LD. The translocations have been a great hindrance to genetic map development in oat. Indeed, to date, there is still no consensus map of oat (Wight et al. 2003), nor can there be one. While the absence of a consensus map is problematic for traditional linkage mapping in oat (typical interval mapping requires an established map), it presents no particular problem for association mapping in which the map position of markers is not explicitly used. Further, given that the highest density marker system in oat (the DArT markers to be used here) have moderate rather than high density, the extent of LD will be beneficial to association mapping of causal loci.

![Figure 1: Reanalysis of data from O'Donoghue et al. (1994). 13 probe pairs could be ascertained to be on the same linkage group. Recombination frequencies calculated based on the Kanota x Ogle map (Wight et al. 2003).](image)
**Preparation to perform whole-genome analysis.** Whole-genome QTL analyses are relatively new and no user-friendly software exists to implement them. We demonstrate here simply that we have developed in-house the necessary software and that, on real oat data sets, it performs remarkably well, true to simulation results presented in the literature (Xu, 2003). Figure 2 shows the original least-squares analysis of β-glucan content in the Kanota x Ogle mapping population, and our reanalysis of the data using whole genome Bayesian shrinkage analysis. The reanalysis produces remarkably sharp QTL peaks that correspond to those identified in the original analysis. Estimated effects of genomic regions with no putative QTL are shrunken close to zero. The analysis also appears to identify three new QTL with quite significant effects (all three happen to be on the right side of the graph). These results would need to be validated, but they appear promising.

**Summary of linkage mapping studies of β-glucan content in oat.** β-glucan has been measured in three oat RIL populations, Kanota x Ogle and Kanota x Marion (Kianian et al., 2000), and Ogle x TAM-O-301 (Unpublished data). Since the previous analysis, the KM population has been scored with an additional 100 AFLP markers. We have analyzed populations separately and combined results using the Kanota x Ogle map as reference (Fig. 3). The outcome is a composite showing up to 13 genomic regions in oat, four of which have been confirmed in more than one population. Because of uncertainty of synteny between the KM and KO maps, two QTL detected in KM cannot be matched up with the consensus, and therefore may or may not coincide with other detected QTL. None of the effects detected are as large as those estimated by Kianian et al. (2000) who found effects up to 0.35% and no less than 0.15%. Shrinkage of effects is expected in the Bayesian analysis. Given typical β-glucan phenotypic standard deviations around 0.5% (Cervantes Martinez et al., 2001; Kibite and Edney, 1998 Chernyshova et al. submitted), the effects we find are nevertheless substantial. The preliminary data indicate that power to detect β-glucan in the association analysis we propose will be reasonable.
Available germplasm and β-glucan records from the UOPN and the NPGS.
A total of 655 entry-year combinations have been evaluated in the Uniform Early and Uniform Midseason Oat Performance Nurseries since 1996 (URL 4). Given the presence of five checks per nursery and the repetition of some experimental entries across years, a total of 378 unique entries have been evaluated in that time. With inclusion of the 2007 UPON, the number of entries will increase to over 410. To date, we have identified seed sources for 85% of these entries. Pedigrees for all these entries have been entered in the Pedigree of Oat Lines (POOL) database (URL 7), which can automate the calculation of kinship matrices. As part of this project, we will ensure that all pedigrees are accessible from POOL. Entries are evaluated in 10-12 and 14-20 locations for the Early and Midseason nurseries, respectively. Hardcopy evaluation data is available for all years. In a given year, the phenotypic range of β-glucan content is typically 2.5% to 3% (URL 4).

The Germplasm Resource Information Network (GRIN) system contains β-glucan records for 5,382 oat accessions, of which 42% were evaluated in 1991 and 58% in 1995. The range of β-glucan content is 2.3% to 8.5%. A stratified selection across the two evaluations gives β-glucan means for the bottom and top 200 accessions of 3.3% and 6.3%, respectively. The countries of origin of the 400 selected accessions do not suggest obvious biases in selected accessions. Of the selected accessions, 35% originated from the United States while 39% of all accessions have that origin. Of countries whose accessions represent >1% of the collection, Brazilian accessions are over-represented in the high β-glucan sample, and Greek accessions are under-represented in the low β-glucan sample. Such disparities can easily be offset in the final selected set to avoid gross structure in the NPGS population at the outset. Passport data from the GRIN system is only a crude indicator of possible relatedness. Measures of relatedness based on multi-locus genotypic data (Pritchard et al., 2000; Ritland, 1996) will therefore be used in the analyses.

RATIONALE AND SIGNIFICANCE
Rationale. The genomic tools and validated methods to use them that will emerge from this research will enable oat breeders to efficiently improve oat’s unique health attributes and manipulate its composition to meet processor’s needs. These activities will increase oat’s economic value for producers and processors, but more importantly, they will enhance oat’s role in combating heart disease, diabetes, and obesity that are on the rise in our society.

The outreach proposed will familiarize and interest young people in plant breeding thereby expanding the pool of quality students who enter this profession. The curriculum development proposed will educate professionals on association-based MAS. Both activities will ultimately boost the productivity of U.S. plant breeders that, in turn, is a key to our food security.
**Significance.** The outcomes will favorably impact 1) plant breeders in general and oat breeders in particular through the tools and knowledge developed; 2) oat producers and processors through the increased value and marketing opportunities for the crop; and 3) the American consumer who will benefit from enhanced oat nutritional function. In particular, plant breeders continue to need published examples of MAS applied to quantitative traits. Over 15 years after the publication of Lande and Thompson’s (1990) foundational paper, we know of only a single empirical application of it (Moreau et al., 2004). For oat breeders and geneticists, the situation is more dire yet. The complexity of oat’s allohexaploid genome with its translocations has stymied linkage analysis and marker development efforts. The concurrent emergence of DArT markers and association analysis provide an opportunity to bring oat into the genomic era. Stakeholders from the processing industry have played an important role in steering us toward β-glucan research. Small increases in the value added to oat during processing leverage enormous revenues for processors and provide new marketing opportunities. Importantly, the mechanisms of oat improvement we propose are accepted by the organic industry of which our stakeholder group includes a representative (Kathryn Begeal). Furthermore, increased demand for oat grain provides incentives to diversify of U.S. agricultural production with this sustainable, versatile, low-input crop in ways that can reduce pest problems and chemical use. Increased demand for oat therefore improves farmers profitability and sustainability. For the American consumer, advancements in breeding may reasonably double β-glucan content. Consider then the attractiveness of obtaining recommended intakes of soluble fiber through a single serving of oat cereal rather than two such servings (Davy et al. 2002). We believe that the greatest economic impact of our proposed research will be through increasing consumer intake of soluble fibers to the benefit of their health and vitality as well as the health and vitality of our society as a whole. An ounce of prevention is worth a pound of cure.

**APPROACHES**

**OBJECTIVE 1. Association mapping of β-glucan content and structure in elite oat germplasm and replicated comparison of phenotypic with marker-assisted selection.** (Jannink, Rai, Anderson, Scott, White)

The research proposed here will identify loci affecting β-glucan in the North American elite oat population and use the associations to test two MAS methods as compared to phenotypic selection. The rationale is two-fold. First, while loci that affect β-glucan content have been identified in wide experimental crosses of oat, their application to oat improvement has not occurred. Here we develop a paradigm to use cooperation between many small oat breeding programs through the Uniform Nurseries to obtain genomic information that can be useful to all. Second, no rigorous comparison of mixed-model and whole-genome MAS has been published or even proposed to our knowledge to date. The empirical evidence we generate will serve as a critical first reference point for the adoption and / or modification of these methods.

Aim 1.1. Association analysis in North American elite oat. (Jannink, Rai) We will request seed from entries submitted to the UOPN over the past 12 years from cooperators. DNA will be extracted from seedlings in Rai’s lab and submitted to DArT P/L according to given protocols. We will work with Nick Tinker at Agriculture and Agri-Food Canada (a member of our stakeholder group) to insure that all UOPN entries have pedigrees in the POOL database (URL 7) and use these pedigrees to calculate kinships. We will also calculate kinships on the basis of their genotypes using established methods (Ritland, 1996) and publicly available software (Hardy and Vekemans, 2002). Pedigree- and marker-based kinships will be compared using
Mantel tests (Fortin and Gurevitch, 1993). Kinships between oat varieties based on pedigree have been published (Souza and Sorrells, 1991) and the oat breeding community should be interested to know of their accuracy. In general, we will resolve conflicts between the two methods of calculating kinship in favor of the marker-based kinship matrices for two reasons. First, breeders may err in recording their pedigrees. Second, the pedigree-based kinship assumes that each parent of a cross contributes equally to an inbred progeny. But such equal contribution will not occur for sampling reasons during the process of inbreeding, and because each progeny is subject to intense selection, which may further bias sampling of genomic regions from each parent.

The genotypic data obtained will allow the first large-scale linkage disequilibrium study in elite oat. About 250 of the DArT markers will have been mapped on the Kanota x Ogle population (see Kilian letter of support). For many marker pairs, both will be on the same chromosome and we will obtain a denser version of Fig. 1 of the preliminary data above. Comparisons with LD data emerging on barley through the barley CAP should illuminate the impacts of their differing genomic structures (i.e. allohexaploid with translocations versus well-behaved diploid).

The genotype and kinship data will be subjected to two analyses. First, we will perform mixed-model association analysis as described in (Arbelbide and Bernardo, 2006; Parisseaux and Bernardo, 2004; Yu et al., 2006) and implemented in The Hordeum Toolbox (URL 1). We have significant experience with such mixed models (Jannink, 2006). Second, we will apply the whole-genome analysis proposed by Meuwissen et al. (2001) and Xu (2003b) and incorporating improvements (ter Braak et al., 2005). We have implemented these methods in our own software.

**Aim 1.2. Management of oat phenotypic and DArT genotype data.** (Anderson) Currently the data from the UOPN is available on the GrainGenes website as spreadsheets in HTML format (URL 4). This information will be migrated to a relational database holding both it and the DArT marker data from Aim 1.1. Public software interfaces to the database will be developed to allow the combined data to be queried and exported to external analysis tools such as TASSEL. This project database will be maintained on a server at the GrainGenes site, administered by co-PI Anderson. In addition the data will be integrated into the GrainGenes Database itself. Placing data curation responsibilities under co-PI Anderson will facilitate its coordination with other GrainGenes work on phenotype and genotype data thereby insuring its compatibility across other genomic efforts and enabling its long-term maintenance as a growing data resource.

We intend the oat phenotype/genotype database to be a resource that will be extended by new data from other projects beyond the duration of this grant. To this end, procedures will be developed to facilitate and automate as much as possible the submission of new data. These procedures will include defining rules for the input data, documenting the rules as instructions for submitters, and developing a software pipeline for processing and validating the data.

**Aim 1.3. Develop and apply a high-throughput method to measure β-glucan molecular weight.** (Scott, White) Both amount and structure of β-glucan have an impact on its dietary benefit. We will develop high-throughput methods to measure β-glucan structure, to be used for association analyses and for breeding purposes. We will use the methods as screening tools to identify likely extreme samples. We will then re-analyze these samples with a slower but accepted methods. We have already increased the throughput of the enzymatic method to measure β-glucan content in this manner (Chernyshova et al. submitted).

Our measures of β-glucan structure will be the average β-glucan molecular weight (MW) and the distribution of β(1-3) linkages in the polymer. We will determine the concentration of reducing ends in a ground grain sample before and after enzymatic hydrolysis of β-glucan with lichenase,
which specifically hydrolyzes β-glucan at β(1-4) linkages adjacent to β(1-3) linkages. The change in reducing end concentration will indicate the frequency of β(1-3) linkages. The β-glucooligosaccharides remaining in the sample will be hydrolyzed to glucose with β-glucosidase. The ratio of resulting glucose concentration to the initial concentration of reducing ends will indicate average β-glucan MW. Rapid methods to determine reducing end concentrations exist (Fox and Robyt, 1991). The average mass of β-glucans will be confirmed by purification of β-glucan followed by size exclusion chromatography (Colleoni-Sirghie et al., 2003a; Colleoni-Sirghie et al., 2003b). If the large amount of α-glucan or other components of grain extracts interferes with the assay, we will use a rapid 96-well plate β-glucan purification method involving water extraction, treatment with amylases and proteases to hydrolyze the α-glucan and proteins in the extract followed by ethanol precipitation of the β-glucan.

Aim 1.4. **Compare three selection methods.** (Jannink, Scott, Rai) The ultimate purpose and validation of these analyses is marker-assisted selection. We will use three methods to perform selection over two cycles: phenotypic selection, mixed-model analysis and whole-genome selection using DNA marker data. Our *rationale* is that neither of the MAS methods have been applied using association analysis, nor have they been compared to each other. Though a single selection experiment is insufficient to come to definitive conclusions (Hill and Caballero, 1992), the comparisons generated here will provide an important reference. Each method will be performed in duplicate to improve our ability to make inferences. Oat is a convenient crop to perform selection experiments because it is easy to grow three generations per year including a field evaluation and two greenhouse seasons (Frey et al., 1988). Selection will begin in grant Year 2 according to the following schemes:

**Phenotypic selection.** Given the availability of pedigrees, we will reanalyze all UOPN data jointly and take advantage of relationships between entries using mixed-model analysis (Lynch and Walsh, 1998). Observed phenotypes are modeled as

\[
y = X\beta + Zg + e
\]

where \(\beta\) are environmental effects and \(X\) relates the phenotypes to the environments where they were observed, \(g\) are experimental line polygenetic effects and \(Z\) relates phenotypes to the experimental lines on which they were observed, and \(e\) are residual errors. The kinship matrix is required to model the variance-covariances of the vector \(g\). The predicted genotypic values that result from this analysis are

\[
\hat{y} = Zg.
\]

We will select the twelve UOPN entries with the highest predicted β-glucan contents and employ the rapid cycle recurrent selection scheme proposed by Frey et al. (1988) to recombine them. Because that scheme requires segregating \(S_0\) lines rather than inbreds, we will cross selected inbred parents using a partial diallel in the greenhouse in Jan. 2009 and generate 48 crosses. In the greenhouse Sept. 2009, each cross will be selfed to create two \(F_2\) progeny. In the greenhouse Jan. 2010 each \(F_2\) will be increased. \(F_{2:3}\) families thus derived will be evaluated in the field in 2010 and measured in the lab for β-glucan content in August 2010. Phenotypic selection will identify the twelve highest β-glucan lines. Greenhouse crossing, increase and then field evaluation will be repeated in Sept. 2010 to Summer 2011.

**Mixed-model MAS.** After association analysis described in Aim 1.1, all loci identified as contributing to β-glucan content will be entered into a single statistical model as follows. Phenotypes are now represented as

\[
y = X\beta + M\alpha + Zg + e
\]

where model terms are as in phenotypic selection and \(\alpha\) are genetic effects associated with all DArT markers identified as significant and \(M\) is a matrix containing the marker allele identities for all loci and all lines in the analysis. The \(g\) effects now need to be interpreted as experimental line polygenetic effects not
associated with markers. The predicted genotypic values that result from this analysis are 
\[ y' = M\alpha' + Zg \]  
As in the phenotypic analysis, we will select UOPN entries with the highest predicted \( \beta \)-glucan contents and begin by crossing them in the greenhouse in Jan. 2009. In the greenhouse in Sept. 2009, each cross will be selfed to create two \( F_2 \) progeny. In the greenhouse Jan. 2010 each \( F_2 \) will be increased. At the same time, tissue will be sampled for DNA extraction and genotypes will be scored for the relevant DArT markers. At the same time, \( F_{2:3} \) families will be evaluated in the field in 2010 and measured in the lab for \( \beta \)-glucan content in August 2010. For selection after this first cycle, we wish to maintain some flexibility. On the one hand, we will rerun an association analysis using now all both the UOPN population and the selection population: 
\[ y = X\beta + M'\alpha + Zg + e \]  
where the ' indicate revised estimates of genetic effects accounting for additional data. We will then compare \( y' = M'\alpha' + Zg \) to \( y = M\alpha \), where, in the latter case, the marker effects were estimated from the original analysis of the UOPN population. If these predictions are divergent, further analysis to determine whether epistatic interactions are occurring will be warranted (Jannink, 2006), and we will use the \( y' \) predictions for selection. Otherwise, we will use the \( y \) predictions, which will provide evidence of the ability to select for \( \beta \)-glucan solely on the basis of marker information and therefore to accelerate the process and avoid phenotyping costs. In any event, the twelve lines with highest predicted \( \beta \)-glucan contents will be selected and crossing will take place in the greenhouse in Sept. 2010 and new families will be available for analysis in Summer 2011.

**Whole genome selection.** In whole genome selection, all markers are retained in the model, and consequently we change the model notation to 
\[ y = X\beta + N\gamma + e \]  
where \( \gamma \) are the shrunken genetic effects of all the markers and \( N \) contains the marker allele identities for all loci and all lines. Note that in this case there is no polygenic effect, consistent with the practice found successful by Meuwissen et al. (2001). The predicted genotypic values are now \( y' = N\gamma \). Selection and crossing according to the same schedule as for the mixed-model MAS.

**Comparison of selection methods.** By summer 2011, all three selection schemes will have gone through two cycles of selection and intermating using different procedures and sources of information, but the same selection intensities. We have purposefully maintained each selection program small (each protocol entails evaluating 96 lines, which can be accomplished quickly) so that each program can be replicated twice. The 2011 evaluation, therefore, will involve not only the lines from each program but a comparison of programs. Lines within programs will be considered random effects, as will the replication of each program. Selection methods will be considered fixed effects. This evaluation will provide a clean comparison of the methods.

**Outcomes from Objective 1**
- A set of markers associated with \( \beta \)-glucan content and structure in elite North American oat
- Phenotypic and genotypic data for oat will become available to modern bioinformatics tools
- A validated high-throughput method to assess \( \beta \)-glucan structure
- Elite oat germplasm improved for \( \beta \)-glucan content using three selection methods
- Rigorous comparison of two competing association analyses and their application to MAS

**Anticipated problems and their resolution.** We provide evidence here that LD in elite oat extends quite far, likely farther than in barley. Nevertheless, the marker system we will use, while incredibly economical and of higher density than anything else available in oat, is only of moderate density and may not provide complete genome coverage. This potential problem will not hinder any of the Obj. 1 outcomes. The consequence of this problem would be that the association analyses proposed would not identify markers associated with *all* loci causing
variation in β-glucan. The preliminary data we show on the number and size of β-glucan loci detected to date, on LD in oat, and the marker density we will have all suggest that we will identify an important fraction of all causal loci. The high-throughput method we propose to evaluate β-glucan structure is unproven. The method, however, rests on previously well-established analyses and with which we have significant expertise. We anticipate being able to overcome any hitches in that specific aim.

**OBJECTIVE 2. Association mapping of β-glucan content from the National Plant Germplasm System: complementation of elite oat. (Jannink, Scott, Rai)**

The research proposed here will identify loci associated with β-glucan content in world-wide oat collection and compare these loci identified in elite North American oat. The rationale for the research is two-fold. First, the greater diversity of the world-wide collection may allow the identification of loci that are not segregating in the elite population. In that case, we will be able to determine whether the elite population carries the allele conferring high β-glucan content. If not, this locus would be a prime target for marker-assisted introgression, and this objective provides a focused, rational exploration of the genetic diversity present in the NPGS. Second, more rapid decay of LD is expected in the world-wide collection than the elite population. Thus, loci identified in both populations may be mapped more accurately in Objective 2 than 1.

Aim 2.1. **Association analysis in NPGS oat.** (Jannink, Rai) The GRIN system has made it easy to identify desirable germplasm for this objective (see preliminary data). We will request accessions from the National Small Grains Collection in Aberdeen, Idaho. For each accession, we will plant an increase in the greenhouse. In case of genetic heterogeneity within accessions we will pick a single plant as representative of that accession. DNA will be extracted from tissue sampled from the plant and seed collected from the plant for field evaluation. All accessions will be grown in the field in hill plots in Summer 2008. These plots will serve both to increase seed of the accession and to re-evaluate β-glucan content on all accessions in a common environment.

DNA will be sent to DArT P/L for genotyping. We will calculate marker-based kinships using SPAGeDi (Hardy and Vekemans, 2002). Association analyses will then proceed as in Aim 1.1.

Aim 2.2. **Compare association in the NPGS population with those in elite oat.** (Jannink) Association analysis in two separate populations as proposed here will provide a wealth of information about the identified loci not possible from a single analysis. For a given marker four outcomes are possible from the comparison of the two results:

<table>
<thead>
<tr>
<th>Marker Outcomes</th>
<th>Associated in Elite</th>
<th>Not associated in Elite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated in NPGS</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Not associated in NPGS</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

For each outcome we outline the different questions and interpretations entailed:

1. Are the allelic effects in both populations consistent? If yes, the marker is confirmed and can be reliably associated with the trait and accurately mapped. If no, the marker may be weakly linked to the locus causing phenotypic effects and/or there may be heterogeneity of LD across the populations due to drift.

2. Is the marker highly polymorphic in the elite population? If no, does the elite population have the favorable allele for β-glucan content? This outcome will provide a list of candidate loci for introgression from NPGS accessions to elite oat lines. If yes, the marker may again be only weakly linked to a causal locus.
3. Is the marker highly polymorphic in the NPGS population? If yes, the lack of association in NPGS may result from more rapid decay of LD in that population, hence a warning that the marker may not be tightly linked to the causal locus.
4. Uninteresting marker.

**Outcomes from Objective 2.**
- Relative to Obj. 1, additional markers associated with β-glucan content and structure.
- Analyses providing validation and interpretation of Obj. 1 results.
- A rational entry point for the exploitation of oat genetic diversity present in the NPGS for the improvement of North American oat. Will allow request of accessions from the NPGS based on specific alleles to complement existing germplasm rather than based solely on phenotype.

**Anticipated problems and resolutions.** We anticipate that LD will decay more rapidly in the NPGS than in the elite oat population because of the more distant common origins of the NPGS accessions. Thus, the currently moderate density of our marker system may not give us adequate genome coverage. Countering this fact is the selective genotyping we have performed in the NPGS population. The great variance in the trait that selective genotyping causes increases QTL detection power (Darvasi and Soller, 1992). Furthermore, given an associated marker from Obj. 1, non-detection in Obj. 2 due to lack of LD in the NPGS population is incorporated into the interpretive scheme we provide above and gives valuable information on the linkage status of the marker relative to the causal locus.

**OBJECTIVE 3. Educational initiatives to pipeline students into plant breeding and to educate professionals. (Gibson, Moore, Drew, Schultz, Jannink, Minear)**

National efforts are underway to increase the numbers of students studying and embarking on careers in science and technology. The first activities proposed will introduce students interested in science to careers in plant breeding. The rationale for these activities is that the traditional pool of students drawn to plant breeding is shrinking. Unless efforts are made to reach out to different populations and excite new students about the opportunities available in plant breeding and biotechnology, the number of scientists engaging in plant breeding research may become critically low. Ben Hable, a member of our stakeholder committee (see Hable letter of support), writes, “One of the growing concerns of industry breeding managers is the shortage of plant breeders. We have more openings and there are fewer candidates to fill positions.” The second activity will develop a distance short course for continuing education of professionals. Hable writes, “Another need of the seed industry is continuing education for current research scientists. Plant breeding has changed considerably in the past 10 years and the future will see accelerating change. … Industry plant breeders are scattered about the Agricultural areas of the country, making distance education the best way to reach the most scientists.”

**Aim 4.1. Recruiting potential students to the study of plant breeding. (Gibson, Jannink, Minear)**

**High school level.** We plan to link the science of plant breeding to improvements in people’s health, a connection that we believe will speak more directly to the urban/suburban audience from which we need to recruit than a discussion of crop yields. We will develop a series of presentations that inform potential students about careers in plant breeding and the impact plant breeding research has on the larger community, using as our leading example the story of breeding oats and barley with enhanced β-glucan. Our program will specifically target outreach toward women and minority students, including Native Americans, African Americans, and Hispanics, populations under-represented in the plant breeding profession in the U.S. In our project, we will partner with two established campus organizations, Women in Science and
Engineering (see Zunkel letter of support) and Science Bound (see Hargrave letter of support), groups that educate middle and high school students about careers in science. We will develop presentations and hands-on activities that describe plant breeding research and its impact on the health of the community. Activities will be tailored to meet the needs of specific targeted groups. We have budgeted funds to give graduate students willing to serve as presenters a small increase in their annual stipend. An interactive, online version of the presentation will also be available on Agronomy related Web sites (URL 10). In addition, we will develop presentations about plant breeding career opportunities for on-campus student groups, including MANRRS (Minorities in Agriculture, Natural Resources, and Related Sciences; Nina Grant, ISU College of Agriculture Minorities Liaison, is a stakeholder for our proposal. The ISU MANRRS chapter has won national recognition), the Agronomy Club, the Food Science Club, Collegiate 4-H (Gibson has extensive organizational experience with FFA and Collegiate 4-H), and others.

Undergraduate level. In addition to the class designed for advanced students and plant breeding professionals, we propose to create curriculum focusing on the science of plant breeding, basic breeding techniques and processes, and the impact of plant breeding programs on the larger community. This curriculum will be inserted into introductory undergraduate programs currently taught in the College of Agriculture at Iowa State University and will introduce undergraduates to opportunities in pursuing advanced work in plant breeding. Courses that will benefit from the improved curriculum are Agronomy 114 Principles of Agronomy, Agronomy 212 Grain and Forage Crops, and Food Science 411 Experimental Foods. Pairing this early course work with the active mentoring program already established by the Agronomy faculty at Iowa State University for undergraduates interested in plant breeding and biotechnology will further support, guide, and pipeline these students into graduate training in plant breeding.

Aim 4.2. Education of professionals, advanced plant breeding students, and undergraduate students. (Moore, Drew, Schultz, Jannink) We will develop a graduate credit course that will teach both theoretical and practical aspects of applying DNA marker information to breeding. The course will be titled “Association Genetics and its Application to Plant Breeding.” The course and course materials will be accessed and completed from the Internet or from a disc. The class will be designed to deliver course content in an interactive multimedia format combining text, graphics, video, audio and other animation tools. Synchronous and asynchronous communication, course calendar, assignments, and other computer resources will be included in a student notebook system via WebCT, a Learning Management System enabling the user to determine the sequence of content. An html and JavaScript web page template will be developed and implemented to support materials used for the course. The system used will be based on the model used to deliver materials for the Iowa State University of Master’s Agronomy Distance Education Program (see Thompson letter of support; URL 11). Evaluation of the effectiveness of the program will be based on research done in distance education. Formative and Summative evaluations will be included in all stages of content development and instructional delivery (Moore & Kearsley, 2005). Components used for evaluation will include 1) lesson reflections written by students on course content and delivery, 2) overall course evaluations assessed at close of course, 3) student focus group meeting to discuss effectiveness of course, and 4) Distance Education Impact Assessment Survey (DEIAS), a survey measuring the impact of distance classes on the careers of participating professionals (Drew, 2006).

Outcomes from Objective 3

• Plant breeding professionals and advanced graduate students educated in application of association genetics to plant breeding
• Updated curriculum for undergraduate courses on biotechnology and computational biology in plant breeding, showing it to be a rewarding field of study and work
• Presentations and hands-on activities tailored to specific audiences that inform students and potential students about the study of plant breeding
REFERENCES


Drew, J. (2006, November). Impact of the Master of Science in Agronomy Distance Education Program on Student Careers. Paper presented at the meeting of the American Society of Agronomy, Indianapolis, IN.


Frey, K.J. 1996. National plant breeding study. Special Report 98, Iowa Agriculture and Home Economics Experiment Station. Iowa State University, Ames, IA.


Grey, S. 2006. Amid trans-fat fears, the time appears right for Dr. Fehr's beans The Wall Street Journal, New York, NY.


WEBSITE REFERENCES
URL 1. http://www.barleycap.org
URL 5. http://ca.groups.yahoo.com/group/OatDArTs/