AN ABSTRACT OF THE THESIS OF

Robert C. Gaynor for the degree of Master of Science in Crop Science presented on May 27, 2010.

Title: Quantitative Trait Loci Mapping of Yield, its Related Traits, and Spike Morphology Factors in Winter Wheat (*Triticum aestivum* L.).

Abstract approved:

C. James Peterson

Increasing grain yield in wheat (*Triticum aestivum* L.) is a challenging task, because yield is a complex trait controlled by many genes and highly influenced by environmental factors. The genetic control of yield components and other traits associated with yield may be less complex and thus more manageable for breeding. This study seeks to identify quantitative trait loci (QTLs) for these traits. Two new genetic linkage maps were constructed from recombinant inbred lines (RILs) derived from crosses between the Oregon soft white winter wheat variety Tubbs and a Western European hard red winter wheat variety, Einstein. A third linkage map was constructed from RILs from a cross with Tubbs and a Western European experimental hard red winter wheat line. A combination of Diversity Arrays Technology (DArT), Simple Sequence Repeat (SSR), *orw5,* and *B1* markers were used to construct genetic linkage maps. Two replications of the RIL populations were grown in yield trial sized plots at Corvallis, OR and Pendleton, OR in 2009. The RILs were evaluated for grain

yield, spikes per m^2, fertile spikelets per spike, sterile spikelets per spike, seeds per spike, seeds per fertile spikelet, average seed weight, growing degree days (GDD) to flowering, GDD to physiological maturity, GDD of grain fill, plant height, test weight, and percent grain protein. Composite interval mapping (CIM) detected 146 QTLs for these traits spread across all chromosomes except for 6D. Thirty six percent of all of the QTLs detected were in close proximity to four loci: *Rht-B1*, *Rht-D1*, *B1*, and *Xgwm372*. The use of factor analysis to aid in QTL mapping for correlated traits related to spike morphology was explored. Quantitative trait loci mapping of factor scores for these traits potentially showed an increase in statistical power to detect QTLs and a decrease in the probability of type I error over mapping the traits individually.

©Copyright by Robert C. Gaynor
May 27, 2010
All Rights Reserved

Quantitative Trait Loci Mapping of Yield, its Related Traits, and Spike Morphology Factors in Winter Wheat (*Triticum aestivum* L.)

by
Robert C. Gaynor

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 27, 2010
Commencement June 2011

Master of Science thesis of Robert C. Gaynor presented on May 27, 2010.

APPROVED:

Major Professor, representing Crop Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Robert C. Gaynor, Author

ACKNOWLEDGEMENTS

I would like to thank Adam Heesacker for his work in constructing the genetic linkage maps used in this study. Thanks to Mark Larson, Eddie Simons, Mary Verhoeven, Dolores Vasquez, and the wheat project's summer crew for helping collect data. I would also like to thank Lan Xue, Oscar Riera-Lizarazu, Mike Flowers, Jim Peterson, and Jeff Leonard for advising on technical aspects of this research. Finally, I would like to thank Jeffery Stone for serving as a graduate council representative on my committee.

TABLE OF CONTENTS

Page

1. INTRODUCTION..1

2. LITERATURE REVIEW..6

2.1. Yield Components and Crop Ideotype......................................6

2.2. Donald's Ideotype..10

2.3. Genetic Markers..12

2.4. Linkage Analysis..15

2.5. QTL Mapping...17

2.6. Factor Analysis..22

2.7. Crop Models..26

3. MATERIALS AND METHODS...28

3.1. Plant Populations..28

3.2. Data Collection and Summary Statistics...................................29

3.3. Linkage and QTL Mapping..30

3.4. Factor Analysis..32

4. RESULTS AND DISCUSSION...34

4.1. Genetic Linkage Maps...45

4.2. QTL Mapping...45

4.3. Factor Analysis..56

5. CONCLUSION..64

6. BIBLIOGRAPHY...66

TABLE OF CONTENTS (Continued)

Page

7. APPENDICES..73

LIST OF TABLES

Table Page

Table 1 Measured traits....................31

Table 2 Least square means for parents and RILs in TxN population....................36

Table 3 Least square means for parents and RILs in ExT1 population....................37

Table 4 Least square means for parents and RILs in ExT2 population....................38

Table 5 Summary of across location analyzes for the three populations of RILs........39

Table 6 Pearson product-moment correlations for the Tubbs x NSA 98-0995 RILs...41

Table 7 Pearson product-moment correlations for the ExT1 population....................42

Table 8 Pearson product-moment correlations for the ExT2 population....................43

Table 9 QTLs for yield, grain protein, test weight, and spikes per m^2....................46

Table 10 QTLs for flowering, ripening, and grain fill duration....................47

Table 11 QTLs for plant height and seeds per spike....................48

Table 12 QTLs for fertile spikelets per spike and sterile spikelets per spike...............49

Table 13 QTLs for seeds per fertile spikelet and seed weight....................50

Table 14 Pearson product-moment correlations between initial factors and yield, grain protein, and test weight....................57

Table 15 Factor loadings for rotated spike morphology traits....................59

Table 16 QTLs for spike morphology factors....................60

LIST OF APPENDIX FIGURES

Figure Page

Figure 1 Genetic linkage map for the Tubbs x NSA 98-0995 population (TxN) with QTL positions....74

Figure 2 Genetic linkage map for the larger Einstein x Tubbs population (ExT1) with QTL positions....78

Figure 3 Genetic linkage map for the smaller Einstein x Tubbs population (ExT2) with QTL positions....82

Quantitative Trait Loci Mapping of Yield, its Related Traits, and Spike Morphology Factors in Winter Wheat (*Triticum aestivum* L.)

1. INTRODUCTION

Increasing grain yield is a primary goal of wheat (*Triticum aestivum* L.) breeders. Accomplishing this goal is a challenging task due to the complexity of the trait. Grain yield is a complex trait, because it is controlled by many genes and its expression is heavily influenced by environmental factors. Wheat breeders produce large numbers of crosses every year to develop higher yielding varieties. These crosses are in turn used to generate numerous inbred lines with unknown yield potential that have to be evaluated for a litany of agronomic and quality traits before any of these lines can become eligible for release as a variety. For example, Oregon State University's (OSU) wheat breeding program typically generates over 600 crosses and evaluates around 40 thousand F_4 and F_5 rows of inbred lines each year. Accurately measuring yield of these lines is only possible in subsequent generations when there is sufficient seed to plant replicated yield trials at multiple locations. Thus, several cycles of selection have to be carried out on those lines before they are ever evaluated for yield. Once sufficient seed is generated, the high cost of running yield

trials severely limits the number of lines that can be evaluated. At OSU, less than 300 of the initial 40 thousand rows will be screened for yield in replicated trials. Many of these lines will prove to have insufficient yields to be candidates for variety release. Finding a way to screen lines earlier and at a lower cost could vastly improve a breeder's ability to generate higher yielding varieties.

Marker assisted selection (MAS) is a potential method for selecting for yield at early generations. This method uses genetic markers to select for favorable plants based on knowledge of associations between those markers and traits. Genetic markers only require a small sample of DNA, so limited quantities of seed are not a problem. Screening a line with genetic markers is also significantly cheaper than screening that line in a replicated yield trial. However, there are several limitations to MAS that have to be addressed first.

The most common way of finding markers for MAS is to identify quantitative trait loci (QTLs) using QTL mapping (Lande and Thompson, 1990). Statistical limitations of QTL mapping for complex traits, like yield, can make such an approach impractical (Doerge, 2002). However, QTL mapping for traits that influence yield may be more successful, because they often have simpler genetic control and are less influenced by environment (Yin *et al.*, 2000). MAS for traits that influence yield may then be an indirect means of applying early generation selection for yield.

The traits that influence yield include disease resistance, stress tolerance, maturity, and plant architecture. A great deal of research has gone into elucidating the

genetic control for disease resistance (e.g. Singh *et al.*, 2000), stress tolerance (e.g. Zhou *et al.*, 2007), maturity (e.g. Yan *et al.*, 2003) and plant architecture, particularly yield components (e.g. Li *et al.*, 2007). This has lead to successful, albeit limited, implementation of MAS techniques for disease and stress response in wheat (e.g. Murphy *et al.*, 2009; Baerstmayr *et al.*, 2009; Kumar *et al.*, 2010). Markers are available for vernalization genes, genes that partially influence maturity, which can be used in MAS (e.g. Sherman *et al.*, 2004). Dwarfing genes, major determinants of plant height, also have markers that can be used for MAS (Ellis *et al.*, 2002). Besides plant height, there is little literature to suggest that successful MAS strategies for yield components and other architectural traits have been implemented.

Successfully incorporating QTLs for these traits into a MAS program will likely not be accomplished by simply pyramiding genes. This is due to the lack of an obvious "more is better" philosophy for these traits that is present with disease resistance and stress tolerance. Breeders will instead have to manipulate QTLs for yield components and other architectural traits to achieve a crop ideotype. An ideotype is an idealized structure of a plant that obtains the maximum performance under production conditions. Finding optimum values for yield components and architectural traits that obtain this ideotype will require consideration of their interactions with each other, related traits, and the environment (Loomis *et al.*, 1979). Combining knowledge of these interactions with knowledge of the QTLs that control these traits can be accomplished with "QTL-based ecophysiological modeling" (Yin *et*

al., 2003). This approach uses a statistical model to estimate yield of a variety based on its composition of QTLs and the environmental and production conditions that variety is grown under.

Before such an approach can be considered, markers for QTLs that are relevant to germplasm that a breeder is using have to be available. The QTLs that have already been identified may not be relevant to OSU's germplasm. These QTLs may simply not be present in the germplasm. QTL effects are dependent on genetic background, so the QTLs already identified may perform very differently in OSU's germplasm than they did in the studies that identified them (Babu *et al.*, 2004). This could mean that the already identified QTLs for traits related to yield have no practical value to OSU's breeding project.

The goal of this study is to identify QTLs for traits related to yield in crosses that are representative of elite germplasm in OSU's wheat breeding program. The genetic mapping populations used herein contained crosses with an Oregon variety, Tubbs, and varieties from Nickerson, a private breeding program based in Western Europe. Material from Nickerson has been a source of well adapted lines for enhancing germplasm at OSU and has also been a source of lines for direct release as varieties. Knowledge of important QTLs in the genetic background of these crosses could be applicable to a wide range of crosses in OSU's wheat breeding program.

Furthermore, this study examines using QTL mapping of factor scores as a way of dealing with numerous correlated traits in QTL analyzes. The use of factor

scores would create new variables that are not correlated and can reduce the number of variables that are analyzed with QTL mapping. This should result in a reduction of type I error compared to analyzing traits individually. It may also be possible to increase the power to detect QTLs with this method. Factor scores will be calculated for traits related to spike morphology. The results of QTL mapping for these factors will be compared to mapping done on the original traits to see if any of these potential benefits are realized.

2. LITERATURE REVIEW

2.1. Yield Components and Crop Ideotype

The study of yield in cereals has often been based on exploring the relationships between yield and other traits. In particular, traits that are referred to as yield components have been the subject of considerable study. Engledow and Wadham (1923) may have been the first to partition the yield of cereal crops into its components. Their list of components included, among others: plants per unit area, number of ears per plant, number of grains per ear, and weight per grain. They proposed that if the favorable combinations of these traits were known, producing higher yielding lines is possible by generating crosses that obtain these combinations. Although they didn't phrase it as such, this favorable combination of traits can be considered an ideotype. An ideotype simply refers to the idealized appearance of a plant variety. Subsequent research has produced some support for Engledow and Wadham's proposal by showing evidence that changes in the distribution of yield components in the highest yielding varieties can partially explain increases in yield over time (Calderini, 1995). Changes in maturity and plant height have also been attributed to increased yield and were the driving forces behind the “Green

Revolution" (Borlaug, 1983). Thus, determining a wheat ideotype requires consideration of various components of plant architecture and maturity.

Varieties are ultimately judged on their performance in production settings. Therefore, studies that evaluate performance in markedly different conditions may not be meaningful for determining an ideotype (Donald, 1968). For example, many studies on yield components and plant architecture have been conducted on individual plants or thinly seeded plots (e.g. Quarrie *et al.*, 2006; Li *et al.*, 2006). The performance of these individual plants and thinly seeded plots is not likely to be representative of how the plants will perform under production settings (Nerson, 1980). For this reason, these studies may have limited to no applicability when it comes to constructing an ideotype. Only studies of yield components as they behave in a community of plants are likely to further the creation of a crop ideotype.

It is widely known that the optimum maturity for wheat in one environment can be different in another environment. Thus, the ideotype for one region can be different from the ideotype in another region. This difference may extend beyond maturity. For example, seeds per m^2 and tillering have been shown to be the two most correlated components with yield in a study conducted in Kansas (Donmez *et al.*, 2001). This is consistent with the widely accepted view that gains in yield are primarily due to increases in seeds per unit area (Slafer *et al.*, 1996). However, selection based on kernel weight has been shown to be a more effective way of breeding for higher yielding lines in a study conducted in Oklahoma (Alexander *et al.*,

1984). Calderini *et al*. (1995) examined a historical accession of varieties in Argentina and found an increase in seeds per unit area as the driving force for increasing yield in varieties released before 1987 and found seed weight to be more important in varieties released after 1987. These discrepancies could be the product of the different populations studied, but it may also represent differences in ideotypes for these regions. If that is the case, findings from one region may not be of any value to another region.

In addition to variation between environments there is also variation between years that have to be considered. The variation in wheat yields from one year to the next in one field can be large (Washmon *et al.*, 2002). Thus, data from a single year may not be representative of an average year. This year to year variation will have to be taken into account when formulating an ideotype. First, care must be taken to not give too much weight to a single year's data. The ability of an ideotype to have stable yields over multiple years will have to be considered.

Large degrees of correlation between yield components and other traits such as maturity and height have been observed in several studies (e.g. Li *et al.*, 2007; Kato *et al.*, 2000). These correlations can impact studies that attempt to determine the structure of an ideotype. For example, multiple regression of yield on its components is an approach that previously was often used for studying yield components (Walton, 1971). The results from these analyzes may be misleading, because the interpretation of the calculated coefficients is how much a change in one trait will change yield when

all other traits are held constant. Since the traits are correlated, it is not likely and perhaps not possible to change just one while keeping the others constant. Furthermore, the estimates of standard errors for those coefficients are not accurate when traits are correlated. Thus, consideration of these correlations is crucial. This has lead to research into yield components that include a statistical analysis involving the correlation matrix such as path analysis (e.g. Fonseca and Patterson, 1968) or factor analysis (e.g. Walton, 1971). These sorts of analyzes give information on how changing one yield component effects yield and the other traits simultaneously. How much an individual trait can be manipulated and how consistent the effects of those manipulations are over a large range of changes is not determined with these analyzes. This drastically limits the ability of these approaches to predict the performance of varieties not studied.

More recent studies into yield components to identify QTLs have largely ignored correlations between traits (e.g. Börner *et al.*, 2002; Marza *et al.*, 2005). Even when the correlations are considered, the extent of the examination is often limited to reporting a correlation matrix and suggesting that collocated QTLs for different traits represent close linkage or pleiotropy (e.g. Kato *et al.*, 2000; Li *et al.*, 2007; Cuthbert *et al.*, 2008). These correlations could partially be examined with multiple-trait QTL analyzes, and such an approach would increase the power of QTL detection (Mi *et al.*, 2008). The massive computational requirement of analyzing many traits at once has been a limiting factor to the adoption of this technique. An alternative is to use

principal component analysis to reduce the correlated traits into latent, uncorrelated variables. This approach has been used with tassel morphology in maize (Upadyayula *et al.*, 2006) and for morphological factors in wheat (Addisu *et al.*, 2009). The study on tassel morphology found QTLs for the principal components that were not identified with single trait analysis. This might have represented increased power from the use of principal components. The study in wheat didn't find any additional QTLs. This could be due to the inclusion of too many traits in the principal component analysis. This large number of traits were probably controlled by many genes, so there was no realized benefit of increased statistical power.

2.2. Donald's Ideotype

Donald (1968) presented a model for a wheat ideotype in an article titled "The Breeding of Crop Ideotypes". He attempted to draw on the research available at the time to predict what a wheat ideotype would look like. His proposed ideotype would have short, strong stems to withstand lodging. He pointed to the work being done by others with dwarfing genes and the general trend towards shorter wheat as evidence supporting this view. The subsequent "Green Revolution" in wheat that is in large part attributed to the introduction of dwarfing genes verifies Donald's conclusion (Borlaug, 1983). Donald's ideotype also had erect leaves to allow for maximum

illumination of the photosynthetic surfaces. The plant would have a few small leaves. Small leaves were deemed better than large leaves, because they are more likely to be erect. He also noted that theoretical work suggested that an arrangement of several uniformly dispersed small leaves was better at capturing light than a few large leaves. Donald proposed that the ear should be large to assure an adequate sink for photosynthates and to allow for high levels of grain production. The ear should also be erect and have awns to allow for maximum capture of light. A high proportion of seminal roots were deemed desirable for gathering soil nutrients. Donald also stated that the degree to which tillering is desirable is a product of production practices. He suggested that it would be best to grow an ideotype that only had a single culm, because the production of tillers is wastes a plant's resources when it results in overproduction of tillers that are subsequently aborted. Changing the production practices, namely seeding rate, would then be needed to best fit this plant type.

The main characteristic of Donald's ideotype, as proposed by him, is that the variety should be a poor competitor. Donald argued that varieties that are strong competitors use their resources to out compete their neighbors. In a well managed production setting, those neighbors are primarily plants of the same variety. Thus, there is no benefit to the yield of the crop when the plants spend resources trying to out-compete other plants that also contribute to the overall yield. Therefore, it would be better to have a variety that more efficiently use those resources for the production of yield. This trait of Donald’s ideotype is probably the most important for breeders,

because early generation selections are likely to favor strong competitors. For example, F3 generation plants are typically grown in bulk plots in breeding projects. Under these conditions each plant will have neighbors of different genetic composition next to them. The strongest competitors will out compete their weak neighbors. Thus, the stronger competitors will yield more and appear to have a more favorable phenotype than the weak competitors. This will likely bias the breeder towards selecting for the strong competitors. It may therefore be more prudent for breeders to select plants that don't necessarily look the best at these generations. These plants may actually produce higher yielding varieties according to Donald.

2.3. Genetic Markers

Understanding the genetic control of yield components and other traits related to yield requires the use of genetic markers. Genetic markers are powerful tools for exploring the genetic composition of organisms. They can be used to construct genetic maps that can subsequently be used to locate genes controlling various traits within an organism. They can also be used by breeders for identifying favorable genotypes in MAS.

For genetic markers to be useful, they have to be able to distinguish between different alleles at a locus. When this is the case, a genetic marker is said to be

polymorphic. The ability of a marker to detect polymorphisms can be classified as either dominate or co-dominate. Dominate markers have a binary response; they are either expressed or not expressed. These types of markers are only able to detect the presence of one particular allele at a locus, and they cannot definitely distinguish between homozygous and heterozygous individuals. Negative results (lack of expression) are not very informative with these markers, because there are potentially multiple genotypes that cause this result. Thus, two individuals that share a positive result can be said to share at least one allele at a locus. Two individuals that share a negative result can't be considered a match, unless it is known that there are only two alleles present at that locus. Co-dominate markers are more informative than dominate markers. These types of markers are able to distinguish between multiple alleles at a locus. This gives these markers the ability to detect heterozygotes and they are more frequently polymorphic than dominate markers (Paux and Sourdille, 2009).

The majority of genetic markers use at least one of two important biotechnological tools (Paux and Sourdille, 2009). These tools are restriction endonucleases (or restriction enzymes) and polymerase chain reaction (PCR). Restriction enzymes cleave double stranded DNA at specific recognition sites. These recognition sites are short sequences of DNA that are interspersed in plant and animal DNA (Voet and Voet, 2004). Markers that utilize restriction enzymes detect polymorphisms based on differences in the locations of these recognition sites (Paux and Sourdille, 2009). PCR is a technique that amplifies DNA using a set of genetic

probes called primers and a heat stable polymerase. The polymerase is used to exponentially increase the number of copies of DNA in regions that are flanked by primers. Correct primer selection allows polymorphic regions of DNA to be amplified and used as genetic markers (Voet and Voet, 2004). Primers can also be chosen to amplify fragments created using restriction enzymes. This can be accomplished by ligating a piece of DNA, called an adapter, to the end of these fragments that contains a sequence of DNA that the primers bind to (Voet and Voet, 2004).

The most commonly used genetic markers for constructing wheat genetic linkage maps are Amplification Fragment Length Polymorphisms (AFLPs), Simple Sequence Repeats (SSRs), and the proprietary Diversity Array Technology (DArT) markers (Paux and Sourdille, 2009). AFLPs are dominate markers that are generated by cutting DNA with two restriction enzymes, ligating DNA to the end of the fragments, and amplifying fragments with PCR. The main benefit of these markers is that they don't require any development. SSRs are codominate makers that amplify regions of the genome that contain short, repeated sequences of DNA. These markers are highly polymorphic and fairly high throughput, but they are very expensive to develop. DArTs are dominate markers that are used in large capacity microarrays to screen fragments of DNA for hybridization with probes. These markers are ultrahigh throughput. On a per marker basis they are inexpensive, but thousands of markers are screened at once making them expensive to use.

2.4. Linkage Analysis

To improve the usefulness of genetic markers, it is best to know where those markers are located in the genome. The method for determining this is linkage analysis. Linkage refers to the tendency of genes to be inherited together. This occurs when they are located on the same chromosome and in relatively close proximity to each other. Linkage can also be observed between marker loci. The examination of linkage through linkage analysis is the foundation for constructing genetic maps (Wu *et al.*, 2007). Traits that are inherited qualitatively and by only one gene can be directly incorporated into those genetic maps using linkage analysis (Kaeppler *et al.*, 1993). Wheat is an allohexaploid, so many of its qualitative traits are controlled by three genes. Thus, if more than one of those genes are segregating in a population this method cannot be used for mapping the trait. The real power in genetic maps is the ability to place quantitative traits on those maps using QTL analysis (Wu *et al.*, 2007). Genetic maps that contain the locations of important traits can then be used by breeders for identifying markers for MAS.

The first step in linkage analysis is obtaining a population of individuals that share the same parents. Either F2, backcross, doubled haploid (DH), or recombinant inbred line (RIL) populations are typically used. The DH and RIL populations are the most common, because subsequent generations can be obtained through self-crossing

that will not differ genetically. This makes it easy to repeat experiments using these populations in additional years. All of these populations allow for the measurement of linkage between sets of markers that are polymorphic in the parents (Wu *el al.*, 2007). If the loci are unlinked, equal proportions of the parental, non-recombinant, and non-parental, recombinant, pairings of marker alleles would be expected. The lower the proportion of recombinants between two loci, the higher the linkage between them is likely to be. The greater the linkage between loci the closer they are in terms of genetic distance. It is thus a matter of ordering loci according to their linkage to find where they fit on a genetic map.

In theory, finding the optimum order of loci can be determined simultaneously by maximizing the likelihood function for all possible orders. In practice there are x!/2 orders for x loci, so the number of calculations that are required with an actual set of markers would be impractical for current computing technology (Wu *et al.*, 2007). To overcome this limitation, the order of loci is determined using sets of only two or three loci at a time (Wu *et al.*, 2007). After considering many of these sets it is possible to determine an overall order for all loci. The distance between these loci is a function of the recombination rates between them. Recombination rates are not additive, so for the purpose of creating a linkage map the recombination rates are transformed to additive measures of genetic distance called Morgans. This transformation is performed using one of several available mapping functions. The most popular mapping functions are Haldane and Kosambi (Wu *et al.*, 2007). The

difference between these functions is how they handle the occurrence of crossover events between two markers. Haldane's original formula called for a parameter that controls the occurrence of crossovers called coincidence (1919). The mapping function known as Haldane's mapping function sets this parameter to one (Crow and Dove, 1990). This means the function assumes that crossovers occur randomly and independently of each other, so they can be modeled using a Poisson distribution (Wu *et al.*, 2007). Kosambi (1944) introduced a function that assumes the coefficient of coincidence is proportional to the recombination rate between the markers. This results in interference at short distances that diminishes as the distance increases (Crow and Dove, 1990).

2.5. QTL Mapping

After constructing genetic maps, the next step for locating areas of the genome that control a trait of interest is QTL mapping. QTL mapping is a statistical method that aims to find associations between traits and regions of DNA. The regions that associate with the trait are known as QTLs. Markers that are closely linked to these QTLs are commonly used in MAS to select for the desired allele of the target QTL. Many methods for conducting QTL mapping have been proposed.

The simplest technique for QTL mapping is Single Marker Analysis (SM). This technique tests if different marker alleles at a locus are associated with different values of the phenotype. This technique doesn't require a genetic map and there can be multiple marker alleles tested at once (Wu *et al.*, 2007). Analysis of variance (ANOVA) is used to test the significance of the different marker alleles. This approach has severe limitations. First, estimation of the QTL's effect is confounded with the genetic distance between the marker and the QTL (Lander and Botstein, 1989). That is, the less linkage the marker has with the QTL, the more the effect of the QTL will be underestimated. Multiple comparisons are also problematic for single marker analysis. Due to the large number of markers that are commonly tested, a large number of type I errors are expected (Jansen, 1994). Multiple comparison techniques can be used to reduce the incidence of type I error, but they will also reduce the power of the test. Performing SM in a genetic mapping population can offer some graphical tools that can help interpreting the results. In a genetic mapping population there are only two alleles possible at each locus, so a t-test is used to test the significance of different alleles. The genetic map also allows for the t-values to be plotted according to their map position. Then multiple markers that test significant for a QTL and are located next to each other on the genetic map can be assumed to be measuring a single QTL. This makes using a genetic map with SM a better way of interpreting the results when it comes to estimating how many QTLs are present and their locations.

Another way of using markers to find QTLs would be to use them as predictors in a multiple regression model. However, simultaneously estimating the partial regression coefficients is statistically challenging, because the number of markers may be greater than the number of individuals and many of the coefficients are close to zero (Wu *et al.*, 2007). Xu (2003) proposed using a Bayesian approach to estimating the coefficients when marker density is high to cope with this simultaneous estimation problem. His approach is referred to as whole-genome marker analysis. Compared to SM this approach gives fewer significant peaks, and better estimates the effects and locations of the QTLs (Xu, 2003).

Lander and Botstein (1989) formulated a QTL mapping approach that is referred to as Interval Mapping (IM). This approach utilizes a genetic map in performing the calculations for detecting QTLs. The genetic map allows IM to test for the presence of a QTL in the intervals between markers. This is accomplished by maximizing the likelihood function for QTL effect with given values of the flanking markers, considered random variables, and constant values for recombination frequencies between the QTL and each marker. A likelihood ratio (LR) is then constructed comparing this model to the model that assumes there is no QTL present. Traditionally the expectation-maximization algorithm developed by Dempster *et al.* in 1977 is used, but a regression approximation to the LR statistic that is computationally simpler can be used as an alternative (Haley and Knott, 1992). The procedure is repeated at multiple locations and the resulting LR statistics are plotted against their

location on the genetic map. With either method, significance of the LR statistics can be tested using a permutation test that calculates the experiment-wise error (Churchill and Doerge, 1994). Peaks on the map that surpass the significance threshold are used as estimates for QTL locations and effects. A permutation test for significance that sets levels differently for major and minor QTLs has been proposed, but has not been widely adopted (Doerge and Churchill, 1996). IM only tests for the presence of one QTL at a time, so the presence of multiple QTLs can impact the accuracy of the estimates for location and effect of a QTL (Zeng, 1993). Like SM, IM is prone to type I and type II error (Jansen, 1994).

Zeng (1994) along with Jansen and Stam (1994) independently proposed a QTL mapping technique that is called Composite Interval Mapping (CIM). This approach expands IM to include covariates chosen from markers outside the test interval. The covariates work to offset the impact that QTLs outside of the test interval have on the performing the likelihood test. There are several methods for choosing markers as covariates available. Zeng (1994) originally tested using all unlinked markers or all markers. Using all unlinked markers gave greater power and using all markers gave better estimates of QTL effects. An additional method using multiple regression modeling techniques of forward, reverse, or forward-reverse selection to choose markers has become popular. These techniques chose from markers not used in the test site and outside of a certain distance from the test site. They are chosen at a distance from the test site to prevent from eliminating the QTL

signal. Viewing results and determining significance in CIM is conducted in a similar fashion as with IM.

Jiang and Zeng (1994) proposed an extension of CIM to handle more than one trait that is called Multiple-Trait Composite Interval Mapping (MT-CIM). This test only tests one site at a time and uses covariates like CIM, but it looks at more than one trait at a time. This approach offers greater power to detect QTLs when traits are inversely correlated. It also provides a mechanism to test a site that has effects on more than one trait for pleiotropy or close linkage. This can be a valuable tool to breeders, because it allows them to assess if it is possible to break unfavorable combinations of traits. Permutation tests can be used to set significance thresholds as in CIM, but there is some debate as to what is an appropriate null model. A null model that assumes evenly spaced QTLs for the traits might be more favorable than a model that assumes there are no QTLs. In practice, MT-CIM is severely limited by the computational difficulty of handling many traits at one time.

Multiple Interval Mapping (MIM) provides a mechanism for testing multiple QTLs simultaneously (Kao *et al.*, 1999). This allows for interactions between QTLs, epistasis, to be examined. However, the level of computation required for this analysis is greater than those previously mentioned. Thus, putative QTLs detected using another method, such as CIM, are best used in the first steps of model building in MIM (Hackett *et al.*, 2000). The models can then be further refined by optimizing the positions and effects of the QTLs and testing for interactions (Kao *et al.*, 1999). A

permutation test for determining significance levels doesn't exist for this method. Thus, minimizing a model selection criterion is the preferred method for fitting models (Hackett *et al.*, 2000). However, this method is not able to fully protect against over fitting a model, so other factors such as not exceeding the heritability of a trait are often considered. MIM can be expanded to consider multiple traits.

2.6. Factor Analysis

There are two ways to deal with multiple correlated traits in QTL mapping: map each trait separately or use a multiple-trait mapping procedure. Both approaches have disadvantages. Mapping single traits inflates type I error, because multiple analyzes are being performed. Multiple-trait analyzes have greater power to detect QTLs than single trait analyzes, but they are computationally prohibitive and there is no universally accepted model for determining significance of a QTL. Combining factor analysis, a multivariate dimension reduction analysis, with QTL mapping may be a beneficial alternative to those two methods. Factor analysis would reduce the number of variables to be studied and those variables would be uncorrelated.

The beginnings of factor analysis date back to attempts by Karl Pearson, Charles Spearman, and others in the early 20th century to quantify intelligence (Johnson and Wichern, 2007). The technique has been used by cereal breeders and

agronomists as a way of examining correlated traits such as yield components (e.g. Walton, 1971). The intent of factor analysis is to explain inter-correlations between a set of variables as coming from a few underlying latent variables called common factors. The analysis uses the covariance matrix resulting from this set of variables or its standardized version the correlation matrix. The factors are then extracted by one of three methods: principle component, common factor, or maximum likelihood. The number of factors to be considered is determined and then their coefficients are either examined directly, or after an orthogonal or oblique factor rotation. Factor scoring coefficients can also be calculated. These can be used to determine the factor values of individuals for use in subsequent analyzes.

The principle component method for factor extraction is the most popular (Johnson and Wichern, 2007). The first factor is extracted by finding a linear combination of the variables that explains as much of the total variance as possible. The second factor is then determined by finding a linear combination of the variable that is orthogonal to the first and explains as much of the remaining variance as possible. This is process is then iterated until the desired number of factors are obtained, to a maximum of a number of factors equal to the number of variables. The researcher can use multiple methods for determining how many factors to retain. The eigenvalue greater than unity and the scree test are the most popular, but are not necessarily the most accurate (Tanguma, 2000). Parallel analysis is widely considered the most accurate method for selecting the number of factors to retain (Hayton *et al.*,

2004). This method compares the eigenvalues of the observed factors to those of randomly generated data to determine if a factor should be retained. Since the principal component method uses all of the variance and not just the shared variance to extract factors it is not considered a true method of factor analysis (Costello and Osborne, 2005). That is, it is really a method of data reduction and not a method for identifying latent variables.

A true factor analysis can be performed with the common factor method. The common factor method is performed on a reduced correlation matrix (Johnson and Wichern, 2007). This reduced correlation matrix has the variance that is specific to a variable removed from the diagonal of the matrix. Thus, the matrix only represents the variance shared between variables. There are multiple methods for determining what values should be used for the diagonals. After one of these methods is chosen, the factor solution is then obtained iteratively. However, using the reduced matrix now makes it possible to obtain negative eigenvalues for some of the factors since the reduced matrix doesn't represent all of the variance. The common factor method more accurately reproduces population patterns than the principal component method, but these differences converge when the number of variables is large and the loading patterns increase (Snook and Gorsuch, 1989).

The maximum likelihood method is another form of a true factor analysis that assumes the common and specific factors are normally distributed (Johnson and Wichern, 2007). This method can give fairly different results from the common factor

method. It is also possible to perform hypothesis tests for the adequacy of the number of factors retained when the assumptions of the maximum likelihood method are met. This gives an additional tool to researcher for deciding how many factors to retain. However, the normal distribution restriction can make the maximum likelihood method unusable for some sets of data.

Factor rotation is often applied to aid in the interpretation of factors. The mathematical justification for factor rotation is that orthogonally rotated factors will explain the covariance just as well as the original factors, so there is no unique solution for finding factors (Johnson and Wichern, 2007). The varimax rotation, the most common orthogonal rotation, seeks a rotation that makes as many coefficients close to zero as possible (Kaiser, 1957). The resulting rotated factors can then be easier to interpret than the originals, because individual factors have fewer variables making large contributions to them. In cases where the underlying factors are not assumed to be independent, oblique rotations may be more appropriate (Johnson and Wichern, 2007). These types of rotations seek to explain variables with as few factors as possible. Even after performing one of these types of rotations there can still be a good deal of ambiguity in the interpretation of the factors. The researcher has to rely on his experience to make sense of the factors.

The lack of uniqueness in the factor solution that allows for rotation is also a reason that many researchers are opposed to using factor analysis (Kaiser, 1957). The large number of subjective decisions that a researcher has to make can bias the results

of the analysis. These biases can have the greatest impact when it comes to selecting the number of factors to retain and interpreting the resulting factor coefficients (Costello and Osborne, 2005). Overall, factor analysis can appear to be more of an art than a science and great care should be taken when interpreting the results (Johnson and Wichern, 2007).

2.7. Crop Models

Simply finding QTLs for a trait is of no use unless it is known what the ideal value for that trait is. A method for determining the optimum values for yield components and other traits related to yield is using crop models. Crop models are statistical models that are typically used to predict the yield of a crop. They use predictors based on weather patterns, production practices, nutrient levels, and components known as genetic coefficients (Loomis *et al.*, 1979). Crop models can potentially be used to assist in the design of crop ideotypes by maximizing the contribution from genetic coefficients (Haverkort and Kooman, 1997). The main benefit of this approach is that it can be used to manipulate the genetic coefficients to test plant types not currently available to the breeder (Loomis *et al.*, 1979). Some estimates of genetic coefficients are actually based on observed phenotypic values, so they may really be measuring genotype by environment (GxE) interactions and not

actual genetics (Baenziger *et al.*, 2004). For models to have a useful role in structuring a crop ideotype, they have to have input parameters that are firmly based on real genetics. If the parameters are not based on genetics, breeding will not easily be able to produce plants with the desired phenotypic values of those parameters, because it is based on the manipulation of genetics (Baenziger *et al.*, 2004).

A recent approach to crop modeling called "QTL-based ecophysiological modeling" has replaced genetic coefficients with QTLs in crop models to incorporate actual genetics into them (Yin *et al.*, 2003). QTLs are regions of DNA that are associated with one or more traits. A QTL is often thought of as a gene, but that is not necessarily the case. A QTL can be a gene, a group of genes, or some other part of DNA that influences the trait. The effects of individual QTLs can be highly dependent on GxE interactions (Jansen *et al.*, 1995). This is particularly the case for complex traits such as yield (Campbell *et al.*, 2003). To overcome this, QTL-based ecophysiological modeling uses QTLs for the component traits of yield which are likely to be less impacted by GxE interactions (Yin *et al.*, 2000). The effects of those QTLs on yield and how the environment influences yield are then considered in the framework of a crop model to produce an estimate for yield. The first attempt at this approach was made in barley with poor results, because the model lacked the power to distinguish small changes in yield (Yin *et al.*, 2000). However, success using the same modeling approach for less complex traits suggests that this approach may be successful for yield in the future (Buck-Sorlin, 2002; Yin *et al.*, 2005).

3. MATERIALS AND METHODS

3.1. Plant Populations

The populations used for linkage mapping and QTL analysis were obtained by randomly selecting heads from two F_2 populations in OSU's wheat breeding program. One population consisted of a cross between Tubbs, an awned, soft white winter wheat variety widely grown in Oregon, and NSA 98-0995, an awnless experimental hard red winter wheat line from the Nickerson breeding program. 280 F_5 derived RILs were generated from this population. The other population was a cross of Einstein, an awnless, hard red winter wheat variety from Nickerson that is widely grown in Western Europe, and Tubbs. 276 F_5 derived RILs were produced from this population.

OSU's wheat breeding program produced these populations by combining two separate crosses. Each individual cross is typically only pollinated with one spike, but a second spike may be used when low levels of pollen are obtained. No record is kept of the precise individuals used in these crosses; just what lines they came from. Thus, it is possible for there to actually be up to four true populations within a nominal population and up to six separate parents. However, only one female plant and one or two males are typically used to create a population. To account for this, the segregation of multi-allelic loci was examined in these populations in an attempt to

determine the number of true populations present in each nominal population. This information was used to divide the RILs into “true” populations if necessary and possible. Lines that could not be assigned to a population with confidence were discarded.

3.2. Data Collection and Summary Statistics

The ExT and TxN RILs were grown in separate experiments with local varieties to serve as agronomic checks and their parent lines. Since actual parents weren't known, the parental varieties were used in their place. Each experiment was grown at two locations with two replications in a randomized complete block design planted in the fall of 2008. Each replication was planted in a plot that was 1.52 m wide by 4.27 m long using about 90 grams of seed. The two locations were Corvallis, OR and Pendleton, OR. The Corvallis plots contained six evenly spaced rows and the Pendleton plots had seven rows. These locations represent two of the highest yielding locations in Oregon. Each location was managed according to standard fertility and weed control practices in their respective locations. All seed was treated with a fungicidal seed treatment prior to planting. No subsequent fungicides or insecticides were used.

Fourteen traits were measured on each plot. A list of those traits and the methods used to measure them are presented in table 1. Each trait was analyzed separately in each population using the general linear model procedure in the SAS software package (The SAS Institute, Cary, NC). Environment was considered a fixed effect and genotype was considered a random effect. Type III sum of squares were used to test significance of the terms in the model. Least square means of RILs were calculated on a location basis for subsequent use in determining Pearson product-moment correlation coefficients and for QTL mapping. Heritabilities for each trait were calculated across locations on a plot mean basis using the sum of squares from the ANOVA output as outlined by Bernardo (2002).

3.3. Linkage and QTL Mapping

Each population was screened with one STS (*orw5*), 123 SSR, and 6,500 DArT markers. Markers that were polymorphic were used with the phenotypic marker for the awn repressor gene, *B1*, to create genetic linkage maps. The regression mapping algorithm in JoinMap 4.0 was used to generate linkage maps (Van Ooijen and Voorrips, 2006). Large clusters of markers were replaced with "delegate" markers to keep from exceeding resolution power and causing unreliable mapping (Korol *et al.*, 2009).

Table 1 Measured traits.

Trait	Method of Measurement	Trait Categories
Flowering	GDD from planting to 50% of spikes have extruded anthers.	Maturity Traits
Ripening	GDD from planting to 50% of flag leaves have senesced (physiological maturity).	Maturity Traits
Grain Fill Duration	Ripening minus Flowering	Maturity Traits
Height	Distance from ground to top of an average spike in the plot. Measured no earlier than two weeks before physiological maturity.	
Yield	Weight of grain harvest in a plot divided by area of a plot.	
Test Weight	Sample of grain from a plot analyzed with an automated test weight machine (model GAC 2100b, Dickey-John, St. Louis, MO)	
Grain Protein	Sample of grain from a plot analyzed using NIR-spectroscopy (model Infratec™ 1241 Gain Analyzer, Foss, Laurel, MD)	
Spikes / m^2	Measured from a 0.5 meter section of row at Corvallis and two 0.5 meter sections at Pendleton.	Yield Components
Fertile Spikelets / Spike	The average of 10 randomly chosen spikes in a plot.	Yield Components; Spike Morphology Traits
Sterile Spikelets / Spike	The average of the same 10 randomly chosen spikes in a plot.	Yield Components; Spike Morphology Traits
Seeds / Fertile Spikelet	The 10 randomly chosen spikes were threshed, the seed counted, and divided by the total number of fertile spikelets on those spikes.	Yield Components; Spike Morphology Traits; Spike Fertility
Seeds / Spike	The 10 randomly chosen spikes were threshed, the seed counted, and divided by 10.	Yield Components; Spike Morphology Traits; Spike Fertility
Seed Weight	The seed from the 10 randomly chosen spikes was weighed and divided by the number of seeds counted.	Yield Components; Spike Morphology Traits

"GDD" stands for Celsius growing degree days.

Linkage groups were assigned to their appropriate chromosome based on marker locations documented at the public source website Grain Genes 2.0 (http://wheat.pw.usda.gov/GG2). Multiple linkage groups that were attributed to the same chromosome were joined on a single linkage group by relaxing criterion for joining if possible.

QTL mapping was conducted in Windows QTL Cartographer using the composite interval mapping procedure (Wang *et al.*, 2010). Each location was analyzed separately using the least square means of the RILs. Covariates were selected using forward-reverse regression with a probability of 0.1 for adding to and removing from the model. Covariates were excluded if they were within 10 cM of the test site to prevent elimination of the QTL signal at that site. The significance threshold for declaring a QTL was set to 0.05 using 1000 permutations. QTLs were scanned for automatically with the requirement that they are at least 5 cM apart and there is at least a 1 LOD drop between peak and trough. QTLs that appeared to be detected due to incorrectly ordered loci in the genetic map were removed.

3.4. Factor Analysis

Factor analysis was performed on the spike morphology traits (Table 1) in the TxN population using the factor procedure in the SAS software package (The SAS

Institute, Cary, NC). Each location was analyzed separately. The principal component method was used to extract factors from the correlation matrix. The number of factors retained was determined by looking at correlations between factors and grain yield, test weight, and grain protein. The smallest number of factors possible was chosen such that a factor with a significant correlation with grain yield, test weight, or grain protein was not discarded. Extracted factors were rotated with the varimax function to aid in interpretation. Scoring coefficients were calculated for each factor and used to determine factor scores for the RILs. These factor scores were used for QTL mapping. QTL mapping of factor scores was performed using the same methods as QTL mapping of the original traits.

4. RESULTS AND DISCUSSION

The TxN RILs showed no marker loci with more than two alleles, so all 280 lines in the population were considered to have come from a single true population (results not shown). Three alleles were present at some marker loci in the ExT RILs. The segregation pattern of those alleles suggested there were two populations with a common Einstein parent and different Tubbs parents in this cross (results not shown). RILs from this cross were divided into two populations, ExT1 and ExT2. ExT1 included 153 RILs and ExT2 included 87 RILs. The remaining 36 lines were not analyzed due to lack of sufficient marker data to assign them to a population, or they had alleles that appeared to come from two different Tubbs parents at a locus. The latter occurrence could be an indication of out-crossing in this population. Lines lacking sufficient seed for the field trial were included in linkage mapping and not QTL mapping. This brought the total lines for QTL mapping to 271, 142, and 81 in the TxN, ExT1, and ExT2 populations respectively.

The accuracy of population assignments is critical to the resulting linkage and QTL mapping procedures. Individuals incorrectly assigned to populations, could generate outliers that bias linkage and QTL maps. For this reason, all results obtained from these populations must be interpreted with caution. This is particularly the case for those from the ExT populations, because a high number of lines couldn't be

assigned to a population. Further examination of codominate markers in these lines is needed to split them into their appropriate populations. This would increase the sizes of these populations and add statistical power to the subsequent linkage and QTL mappings.

The range of phenotypic values for RILs in each population was broad and transgressive segregates were present for all traits (Tables 2, 3, and 4). The broad range of these values should give good statistical power to detect QTLs, if the complexity of the trait is low. The presence of transgressive segregates indicates that QTLs with positive contributions to these traits are present in both parents or epistatic interactions between QTLs exist.

The majority of traits showed significant GxE interactions (Table 5). The genotype term was significant with a probability less than 0.001 for all traits and in all populations except for grain fill duration in the ExT2 population which has a p-value of 0.019 (data not shown). A large number of degrees of freedom were used to test these terms, so significant results don't necessarily indicate the presence of a meaningful effect. Correlations between locations and estimates of heritability may give better indications of meaningful GxE interactions (Table 5). For example, plant height showed significant or suggestive GxE interactions in all populations, but the correlations between locations (0.82-0.96) and heritability estimates (0.89-0.97) were high in those populations. This indicates that the GxE interactions are probably not very meaningful compared to the simple effect of genotype.

Table 2 Least square means for parents and recombinate inbred lines (RILs) in the Tubbs x NSA 98-01995 population (TxN).

	Parents		RILs		
Trait	NSA 98-0995	Tubbs	Range	Mean	Std. Err.
		Corvallis			
Flowering (GDD)	1610	1589	1541-1681	1600	8
Ripening (GDD)	2200	2217	2173-2347	2215	20
Grain Fill Duration (GDD)	590	628	554-701	615	20
Plant Height (cm)	94	113	75-140	100	3
Test Weight (kg/hL)	74.6	74.1	67.5-78.3	73.7	0.4
Grain Protein (% on dry basis)	10.4	9.9	8.5-14.7	10.5	0.6
Yield (metric tons/ha)	8.3	8.3	4.4-9.6	7.8	0.3
Spikes / m^2	289	339	217-449	321	35
Fertile Spikelets / Spike	22.4	19.1	17.1-25.2	20.9	0.5
Sterile Spikelets / Spike	1.2	2.7	0.4-3.6	1.6	0.2
Seeds / Fertile Spikelet	2.9	2.9	2.2-3.6	2.9	0.1
Seed Weight (mg)	43	49	33-57	46	2
Seeds / Spike	66	56	42-77	61	3
		Pendleton			
Flowering (GDD)	1469	1448	1403-1538	1458	12
Ripening (GDD)	2127	2099	2017-2249	2126	22
Grain Fill Duration (GDD)	659	651	556-739	669	20
Plant Height (cm)	77	94	63-108	83	3
Test Weight (kg/hL)	73.1	74.6	67.1-77.5	73.2	0.7
Grain Protein (% on dry basis)	9.4	8.9	7.6-12.7	9.6	0.6
Yield (metric tons/ha)	5.5	6.0	3.2-6.9	5.3	0.4
Spikes / m^2	369	415	276-526	397	44
Fertile Spikelets / Spike	21.4	16.8	15.1-21.8	18.9	0.5
Sterile Spikelets / Spike	1.3	3.6	0.6-4.1	2.0	0.3
Seeds / Fertile Spikelet	2.6	2.8	2.1-3.2	2.7	0.2
Seed Weight (mg)	37	43	32-52	40	1
Seeds / Spike	55	47	38-64	50	4

"Std. Err." refers to the standard error of a RIL least square mean assuming no missing data.
"GDD" refers to Celsius growing degree days.

Table 3 Least square means for parents and recombinate inbred lines (RILs) in the larger Einstein x Tubbs population (ExT1).

	Parents		ExT1 RILs		
Trait	Einstein	Tubbs	Range	Mean	Std. Err.
	Corvallis				
Flowering (GDD)	1580	1591	1541-1690	1602	8
Ripening (GDD)	2219	2227	2173-2457	2247	16
Grain Fill Duration (GDD)	639	637	575-776	645	17
Plant Height (cm)	86	112	58-160	100	4
Test Weight (kg/hL)	74.3	75.2	70.7-80.1	75.9	0.4
Grain Protein (% on dry basis)	10.3	10.1	9.3-15.1	11.1	0.5
Yield (metric tons/ha)	9.0	8.7	3.3-9.8	7.8	0.3
Spikes / m^2	322	315	228-472	329	36
Fertile Spikelets / Spike	21.2	19.6	17.3-24.5	20.8	0.6
Sterile Spikelets / Spike	1.9	2.5	0.9-4	2.4	0.3
Seeds / Fertile Spikelet	3.1	3.0	1.4-3.5	2.7	0.2
Seed Weight (mg)	47	48	35-58	46	1
Seeds / Spike	66	59	31-75	57	4
	Pendleton				
Flowering (GDD)	1484	1455	1412-1582	1475	14
Ripening (GDD)	2140	2117	2004-2317	2163	30
Grain Fill Duration (GDD)	656	662	584-795	688	29
Plant Height (cm)	68	91	48-123	81	3
Test Weight (kg/hL)	73.3	75.1	69.4-78.3	74.5	0.4
Grain Protein (% on dry basis)	8.4	8.2	7.5-13.7	9.2	0.5
Yield (metric tons/ha)	5.2	5.5	1.6-6.6	4.9	0.5
Spikes / m^2	351	356	223-462	330	42
Fertile Spikelets / Spike	20.4	16.9	15.6-22.3	18.8	0.6
Sterile Spikelets / Spike	2.5	3.6	1.3-4.8	2.8	0.4
Seeds / Fertile Spikelet	2.5	2.8	0.9-3.2	2.5	0.2
Seed Weight (mg)	39	43	32-46	39	1
Seeds / Spike	51	47	18-63	48	3

"Std. Err." refers to the standard error of a RIL least square mean assuming no missing data. "GDD" refers to Celsius growing degree days.

Table 4 Least square means for parents and recombinate inbred lines (RILs) in the smaller Einstein x Tubbs population (ExT2).

	Parents		ExT2 RILs		
Trait	Einstein	Tubbs	Range	Mean	Std. Err.
		Corvallis			
Flowering (GDD)	1580	1591	1561-1713	1605	8
Ripening (GDD)	2219	2227	2173-2362	2239	16
Grain Fill Duration (GDD)	639	637	534-711	634	17
Plant Height (cm)	86	112	65-138	103	3
Test Weight (kg/hL)	74.3	75.2	70.3-78.9	75.0	0.5
Grain Protein (% on dry basis)	10.3	10.1	8.7-11.9	10.5	0.6
Yield (metric tons/ha)	9.0	8.7	5.6-9.9	8.4	0.4
Spikes / m^2	322	315	220-492	347	36
Fertile Spikelets / Spike	21.2	19.6	16.5-23.3	20.2	0.7
Sterile Spikelets / Spike	1.9	2.5	0.9-4	2.5	0.4
Seeds / Fertile Spikelet	3.1	3.0	2.3-4	2.9	0.2
Seed Weight (mg)	47	48	38-58	47	2
Seeds / Spike	66	59	44-83	58	4
		Pendleton			
Flowering (GDD)	1484	1455	1431-1654	1490	18
Ripening (GDD)	2140	2117	2012-2317	2164	36
Grain Fill Duration (GDD)	656	662	551-725	675	37
Plant Height (cm)	68	91	58-118	85	3
Test Weight (kg/hL)	73.3	75.1	70.3-78	73.9	0.6
Grain Protein (% on dry basis)	8.4	8.2	7.3-11.1	8.6	0.5
Yield (metric tons/ha)	5.2	5.5	3.3-6.5	5.1	0.5
Spikes / m^2	351	356	225-434	331	47
Fertile Spikelets / Spike	20.4	16.9	15.8-21.4	18.8	0.5
Sterile Spikelets / Spike	2.5	3.6	1.2-5.1	3.1	0.3
Seeds / Fertile Spikelet	2.5	2.8	1.7-3.4	2.7	0.1
Seed Weight (mg)	39	43	32-47	40	1
Seeds / Spike	51	47	32-70	50	3

"Std. Err." refers to the standard error of a RIL least square mean assuming no missing data. "GDD" refers to Celsius growing degree days.

Table 5 Summary of across location analyzes for the three populations of recombinate inbred lines (RILs).

	Tubbs x NSA 98-0995 Population (TxN)				Larger Einstein x Tubbs Population (ExT1)				Smaller Einstein x Tubbs Population (ExT2)			
Trait	GxE	Correlation	R^2	h^2	GxE	Correlation	R^2	h^2	GxE	Correlation	R^2	h^2
Flowering	***	0.72***	0.98	0.84	***	0.72***	0.98	0.82	**	0.64***	0.96	0.78
Ripening	*	0.60***	0.89	0.75	***	0.61***	0.88	0.73	ns	0.58***	0.80	0.74
Grain Fill Duration	*	0.38***	0.80	0.54	**	0.31***	0.76	0.45	ns	0.12	0.65	0.28
Plant Height	ns	0.82***	0.93	0.89	***	0.95***	0.97	0.96	*	0.96***	0.97	0.97
Test Weight	***	0.74***	0.92	0.85	***	0.83***	0.96	0.90	***	0.77***	0.93	0.87
Grain Protein	**	0.29***	0.71	0.44	***	0.55***	0.88	0.71	ns	0.26*	0.83	0.52
Yield	***	0.45***	0.93	0.60	***	0.58***	0.95	0.72	**	0.54***	0.95	0.69
Spikes / m^2	ns	0.28***	0.69	0.44	ns	0.32***	0.65	0.45	ns	0.24*	0.67	0.47
Fertile Spikelets / Spike	***	0.69***	0.90	0.80	**	0.76***	0.90	0.85	ns	0.76***	0.87	0.87
Sterile Spikelets / Spike	**	0.65***	0.81	0.79	**	0.60***	0.81	0.75	**	0.65***	0.84	0.78
Seeds / Fertile Spikelet	***	0.43***	0.78	0.60	***	0.48***	0.85	0.68	ns	0.58***	0.80	0.74
Seed Weight	***	0.67***	0.91	0.78	***	0.73***	0.94	0.82	ns	0.60***	0.92	0.70
Seeds / Spike	***	0.51***	0.84	0.67	***	0.45***	0.88	0.63	***	0.53***	0.82	0.78

"GxE" is the probability of a genotype by environment interaction from the ANOVA. *, **, and *** represent 0.05, 0.01, and 0.001 significance levels, respectively. "ns" stands for not significant. "Correlation" is the Pearson product-moment correlations between RIL means at Pendleton and Corvallis. "R^2" is the coefficient of determination from the ANOVA. "h^2" is the across location narrow sense heritability estimates for the RILs on a plot mean basis.

Under these conditions, QTLs detected at one location should have a high degree of correspondence with QTLs detected at the other location. Low correlations and heritability estimates with a lack of a significant GxE interaction indicates that a trait was measured with low precision. A low coefficient of determination also indicates this (Table 5). Spikes per m^2 is an examples of such a trait. QTL mapping for these traits is unreliable due to a lack of statistical power. The relatively low heritabilties and coefficients of determination for grain protein are likely due to environmental variability within each test location.

Correlations between traits are presented in tables 6, 7, and 8 for the TxN, ExT1, and ExT2 populations respectively. The relationships between these traits may not be strictly linear, so the correlation coefficients won't accurately represent the relatedness between these traits. That is because correlation coefficients make the assumption that traits are linearly related. Even if the relationship is linear, the simple correlations between traits may not fully represent relatedness between the traits (Fonseca and Patterson, 1968).

The relationship between height and yield in the ExT populations does not appear to be linear. Biplots of the traits suggest an increase in yield as height increase until the highest values of height (results not shown). At these higher values, there appears to be a decrease in yield. Thus, the height to yield correlations in the ExT populations can be misleading.

Table 6 Pearson product-moment correlations for recombinate inbred lines in the Tubbs x NSA 98-0995 population (TxN). Corvallis correlations are above the diagonal and Pendleton correlations are below the diagonal.

	Flowering	Ripening	Grain Fill Duration	Height	Test Weight	Grain Protein	Yield	Spikes / m^2	Fertile Spikelets / Spike	Sterile Spikelets / Spike	Seeds / Fertile Spikelet	Seed Weight	Seeds / Spike
Flowering		0.67***	-0.09	-0.11	-0.35***	-0.09	-0.17**	-0.03	0.39***	0.14*	-0.24***	-0.30***	0.02
Ripening	0.58***		0.68***	0.10	-0.16**	0.16**	-0.06	0.06	0.35***	0.03	-0.24***	-0.25***	0.00
Grain Fill Duration	-0.13*	0.72***		0.24***	0.13*	0.30***	0.09	0.11	0.09	-0.10	-0.09	-0.04	-0.02
Height	-0.09	0.11	0.19**		0.29***	0.03	0.35***	-0.08	0.01	0.11	0.02	0.19**	0.01
Test Weight	-0.23***	-0.10	0.07	0.06		0.33***	0.02	0.05	-0.22***	0.14*	-0.10	0.34***	-0.22***
Grain Protein	0.12*	0.21***	0.15*	0.03	0.30***		-0.19**	0.15*	-0.12	-0.14*	-0.24***	0.18**	-0.27***
Yield	-0.24***	-0.13*	0.04	0.17**	-0.14*	-0.20**		0.05	0.09	-0.14*	0.23***	0.18**	0.24***
Spikes / m^2	-0.16*	-0.09	-0.00	0.01	-0.13*	-0.03	0.13*		-0.11	-0.07	-0.22***	-0.20***	-0.24***
Fertile Spikelets / Spike	0.19**	0.18**	0.06	-0.00	-0.15*	-0.18**	0.15*	-0.13*		-0.15*	-0.03	-0.33***	0.57***
Sterile Spikelets / Spike	-0.17**	-0.16**	-0.04	0.22***	0.25***	0.03	0.04	-0.13*	-0.40***		-0.29***	0.10	-0.33***
Seeds / Fertile Spikelet	0.01	0.11	0.12*	-0.10	0.06	-0.00	0.21***	-0.17**	0.08	-0.22***		-0.13*	0.80***
Seed Weight	-0.22***	-0.24***	-0.08	0.23***	0.07	-0.01	0.16**	-0.22***	-0.09	0.40***	-0.19**		-0.30***
Seeds / Spike	0.11	0.18**	0.13*	-0.09	-0.04	-0.10	0.25***	-0.20***	0.60***	-0.39***	0.85***	-0.20***	

*, **, and *** represent 0.05, 0.01, and 0.001 significance levels, respectively

Table 7 Pearson product-moment correlations for recombinate inbred lines in the larger Einstein x Tubbs population (ExT1). Corvallis correlations are above the diagonal and Pendleton correlations are below the diagonal.

	Flowering	Ripening	Grain Fill Duration	Height	Test Weight	Yield	Spikes / m^2	Grain Protein	Fertile Spikelets / Spike	Sterile Spikelets / Spike	Seeds / Fertile Spikelet	Seed Weight	Seeds / Spike
Flowering		0.70***	-0.11	-0.23**	-0.15	-0.41***	-0.11	0.18*	0.01	0.04	-0.40***	0.11	-0.37***
Ripening	0.64***		0.63***	-0.38***	-0.18*	-0.38***	0.01	0.41***	0.09	-0.02	-0.46***	-0.09	-0.38***
Grain Fill Duration	0.11	0.83***		-0.28***	-0.09	-0.09	0.13	0.38***	0.11	-0.08	-0.20*	-0.25**	-0.14
Height	-0.34***	-0.43***	-0.31***		0.58***	0.41**	-0.08	-0.07	-0.08	0.32***	0.14	0.37***	0.09
Test Weight	-0.11	-0.25**	-0.23**	0.41***		0.16	0.03	0.26**	-0.24**	0.36***	-0.10	0.27***	-0.22**
Yield	-0.35***	-0.37***	-0.24**	0.37***	-0.16		0.09	-0.49***	0.04	0.09	0.58***	0.01	0.56***
Spikes / m^2	-0.05	-0.02	0.01	-0.16	-0.19*	0.29***		-0.04	-0.20*	0.06	-0.08	-0.34***	-0.19*
Grain Protein	0.33***	0.42***	0.32***	-0.26**	0.24**	-0.55***	-0.19*		-0.03	-0.12	-0.48***	0.07	-0.48***
Fertile Spikelets / Spike	0.07	0.06	0.03	-0.25**	-0.20*	-0.06	-0.06	0.03		-0.25**	-0.14	-0.11	0.42***
Sterile Spikelets / Spike	-0.23**	-0.20*	-0.10	0.30***	0.15	0.22**	-0.03	-0.11	-0.28***		-0.18*	0.22**	-0.29***
Seeds / Fertile Spikelet	-0.25**	-0.24**	-0.12	0.18*	-0.21*	0.53***	0.07	-0.50***	-0.26**	-0.13		-0.14	0.84***
Seed Weight	0.01	-0.04	-0.07	0.28***	0.16	0.05	-0.31***	0.04	-0.09	0.41***	-0.25**		-0.17*
Seeds / Spike	-0.22**	-0.19*	-0.09	0.03	-0.32***	0.49***	0.04	-0.48***	0.24**	-0.27**	0.86***	-0.30***	

*, **, and *** represent 0.05, 0.01, and 0.001 significance levels, respectively

Table 8 Pearson product-moment correlations for recombinate inbred lines in the smaller Einstein x Tubbs population (ExT2). Corvallis correlations are above the diagonal and Pendleton correlations are below the diagonal.

	Flowering	Ripening	Grain Fill Duration	Height	Test Weight	Yield	Spikes / m^2	Grain Protein	Fertile Spikelets / Spike	Sterile Spikelets / Spike	Seeds / Fertile Spikelet	Seed Weight	Seeds / Spike
Flowering		0.66***	-0.16	-0.29**	-0.33**	-0.28*	-0.04	-0.05	0.06	-0.05	-0.13	-0.06	-0.09
Ripening	0.61***		0.64***	-0.26*	-0.22	-0.00	0.11	0.24*	0.18	0.02	-0.22*	-0.13	-0.08
Grain Fill Duration	-0.20	0.64***		-0.04	0.05	0.28*	0.19	0.37***	0.18	0.08	-0.16	-0.11	-0.02
Height	-0.37***	-0.25*	0.06		0.64***	0.09	-0.06	0.26*	-0.09	0.28*	-0.05	0.45***	-0.12
Test Weight	-0.15	-0.21	-0.10	0.46***		0.07	-0.10	0.37***	-0.16	0.37***	-0.23*	0.26*	-0.33**
Yield	-0.40***	-0.39***	-0.07	0.22	-0.02		0.19	-0.08	-0.17	0.06	0.35**	0.22	0.21
Spikes / m^2	-0.05	-0.20	-0.20	-0.12	-0.13	0.36***		-0.01	-0.11	0.10	-0.13	-0.34**	-0.22
Grain Protein	0.27*	0.14	-0.10	0.12	0.48***	-0.28*	-0.13		0.04	0.26*	-0.24*	-0.14	-0.20
Fertile Spikelets / Spike	0.19	0.22	0.05	-0.17	-0.12	-0.06	-0.21	0.03		-0.32**	-0.29**	-0.24*	0.41***
Sterile Spikelets / Spike	-0.15	-0.11	0.02	0.30**	0.27*	0.05	0.01	0.06	-0.45***		-0.12	0.06	-0.33**
Seeds / Fertile Spikelet	-0.07	0.00	0.08	-0.09	-0.22	0.34**	0.01	-0.34**	-0.03	-0.25*		0.08	0.74***
Seed Weight	-0.15	-0.08	0.08	0.21	0.26*	0.02	-0.04	0.07	-0.40***	0.24*	-0.07		-0.09
Seeds / Spike	0.05	0.12	0.09	-0.20	-0.25*	0.25*	-0.10	-0.26*	0.51***	-0.45***	0.84***	-0.28*	

*, **, and *** represent 0.05, 0.01, and 0.001 significance levels, respectively

All other traits appear to be at least close to linearly related over the ranges measured (not shown). However, it should not be assumed that these relationships will occur beyond the ranges measured.

The correlations between the spike fertility traits, seeds per fertile spikelet and seeds per spike, were consistently the strongest across all locations and populations. Mathematically, fertile spikelets per spike multiplied by seeds per fertile spikelet equals seeds per spike. Thus, it is expected that the spike fertility traits are under vary similar genetic control. The correlations between other traits depended to a large degree on the population and the location where they were measured. Thus, there may be different genes segregating in the different populations and/or the presence of gene by environment interactions. There is a lack of consistently strong correlations between yield and any of the other traits in all populations at every location (Tables 6, 7, and 8). Generally, yield is correlated with the spike fertility traits and inversely correlated with flowering. In the TxN population no trait has a correlation with yield stronger than 0.35. However, correlations greater than 0.5 for the spike fertility traits with yield are present in the ExT1 population. This lack of consistently strong correlations makes identifying traits to manipulate for increased yield difficult. Ultimately, selection of QTLs for these traits will require considerable knowledge of the genetic backgrounds that they are in.

4.1. Genetic Linkage Maps

The TxN population mapped 373 markers on 21 linkage groups. This map covered 1387.6 cM, 1245.5 cM and 1163 cM for the A, B, and D genomes respectively. The ExT1 population mapped 353 markers on 22 linkage groups covering 1318.6 cM, 1552.5 cM, and 1444.7 cM for the A, B, and D genomes. The ExT2 population mapped 358 markers on 22 linkage groups covering 1548.9 cM, 1423.9 cM, and 955.2 cM for the three genomes. All chromosomes were represented in each population. Relaxing criterion for merging linkage groups from the same chromosome resulted in only two gaps in all three maps. However, there were 57 marker intervals that were greater than 50 cM. Seven of those intervals were over 100 cM. Any QTLs detected in these intervals need to be interpreted with care, because the markers are nearly unlinked at these distances.

4.1. QTL Mapping

Multiple QTLs were detected for all traits for a total of 146 QTLs (Tables 9, 10, 11, 12, and 13). Grain protein (Table 9), seed weight (Table 13), and spikes per m^2 (Table 9) had at least one population that failed to identify QTLs for that trait.

Table 9 QTLs for yield, grain protein, test weight, and spikes per m^2

Trait	Pop.	Chrom.	Loc.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
Yield	ExT1	4D	Corv	wPt-0472cluster	20.8	3.296	0.31682	0.1
		7A	Corv	wPt-0031	20.5	4.84	-0.3644	0.13
	ExT2	2B	Pend	wPt-2110	184.1	3.387	0.2889	0.16
	TxN	2B	Pend	wPt-2327cluster	90.9	4.342	0.17337	0.07
Grain Protein	ExT2	3D	Pend	wPt-6066	114	5.012	0.398	0.23
		6B	Pend	381509cluster	48.7	3.179	-0.2728	0.12
	TxN	4A	Corv	wmc617	14.9	3.188	-0.2622	0.08
Test Weight	ExT1	4A	Corv	wPt-4544	93.5	3.877	1.14463	0.28
		4A	Pend	wPt-1007cluster	151.7	6.197	0.72838	0.14
		4B	Corv	barc90cluster	138.9	4.054	-0.58725	0.08
		5A-2	Corv	wPt-7061	4.2	3.64	0.57225	0.08
		5B	Corv	348314cluster	74.6	3.588	0.51	0.07
		6B	Corv	wPt-6441	216	3.623	1.13775	0.31
	ExT2	4D	Corv	wPt-3058	0	4.686	0.815	0.15
		5A	Pend	gwm291	155.3	3.669	1.24788	0.18
	TxN	5A	Corv	B1	136.4	3.356	0.41863	0.05
		5A	Pend	barc330	42.2	3.622	0.47363	0.06
		5A	Pend	B1	134.4	7.617	0.63588	0.11
		6B	Pend	barc79cluster	80.1	4.932	0.49213	0.06
		7D	Pend	cfd14	170.3	3.882	-0.631	0.1
Spikes / m^2	ExT2	3A	Corv	wPt-0714	184.8	3.743	23.7785	0.15
		3D	Pend	gdm72	95.2	4.952	25.5167	0.24
		4B	Corv	barc90cluster	150.4	3.315	-21.45	0.12
		5D-1	Corv	wPt-1197	89.5	3.749	34.8061	0.34
		7A	Pend	wPt-3425	251.1	4.144	-28.7746	0.18
	TxN	1A	Corv	wmc278	96.1	5.102	13.3076	0.09
		1A	Pend	wPt-4765cluster	116	6.092	24.3792	0.28

"Pop." refers to mapping population. "TxN" refers to Tubbs x NSA 98-0995 population. "ExT1" and "ExT2" refer to the large and small Einstein by Tubbs populations respectively. "Chom." is chromosome. "Loc." is location. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R_2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Table 10 QTLs for flowering, ripening, and grain fill duration.

Trait	Pop.	Chrom.	Loc.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
Flowering	ExT1	1B	Corv	wPt-6690	104.9	3.48	8.6361	0.07
		3A	Corv	tPt-1143	166.8	4.897	11.9123	0.12
		3A	Pend	tPt-1143	164.8	7.312	12.9231	0.18
		4D	Corv	wPt-0472cluster	20.8	5.12	-11.3638	0.13
		5B	Corv	barc4cluster	56.7	5.989	-12.2117	0.13
	ExT2	5A	Pend	wPt-4419	56.6	2.799	-11.0962	0.11
	TxN	5B	Corv	wPt-1895cluster	122.7	5.396	-7.4737	0.08
		5B	Pend	wPt-1895cluster	122.7	7.742	-11.4246	0.11
		7A	Corv	wPt-1928	99.1	7.018	-8.4414	0.1
Ripening	ExT1	1B	Corv	wPt-6690	104.3	4.751	12.7814	0.09
		3A	Corv	gwm666	168.8	3.957	13.5863	0.1
		3B	Corv	381874cluster	154.6	4.153	11.5948	0.08
	ExT2	4B	Corv	tPt-0602	154.8	3.905	-14.7623	0.12
		4B	Pend	gwm251	158.8	4.106	-19.7188	0.18
		4D	Corv	wPt-3058	0	4.628	-15.0344	0.14
		6B	Corv	barc101cluster	55.8	3.233	11.748	0.09
	TxN	4A	Corv	wPt-7924	66.9	3.473	13.4196	0.14
		4A	Pend	wPt-7924	62.9	3.853	15.0062	0.16
		5B	Corv	303931cluster	119.2	4.501	-12.0272	0.08
		5B	Corv	wPt-9467	77.8	4.581	14.0095	0.07
		6A	Pend	wPt-7063	74.7	3.857	9.3901	0.06
Grain Fill Duration	ExT1	1A	Pend	wPt-3904	31	5.129	16.0089	0.14
	ExT2	1B	Pend	barc8cluster	35.1	3.459	15.7915	0.15
		4B	Pend	gwm251	161.2	3.746	-15.1368	0.16
		5B	Corv	wPt-1881	100.1	4.277	-11.3228	0.15
	TxN	2A	Corv	gwm312	145.5	4.635	7.2015	0.07
		2B	Corv	wPt-8760	211.1	3.711	6.1693	0.05
		3B	Pend	wPt-3107	110.3	5.293	9.4677	0.1
		5B	Corv	311648cluster	248.8	3.167	5.776	0.04
		7D	Pend	cfd14	166	4.17	-9.2513	0.09

"Pop." refers to mapping population. "TxN" refers to Tubbs x NSA 98-0995 population. "ExT1" and "ExT2" refer to the large and small Einstein by Tubbs populations respectively. "Chom." is chromosome. "Loc." is location. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R_2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Table 11 QTLs for plant height and seeds per spike.

Trait	Pop.	Chrom.	Loc.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
Plant Height	ExT1	1D	Pend	barc229cluster	66	3.42	3.5724	0.05
		2A	Pend	wPt-7024	212.8	3.182	3.4297	0.04
		4B	Corv	barc20cluster	143	15.44	-11.6664	0.27
		4B	Pend	barc20cluster	143.6	17	-9.6777	0.31
		4D	Corv	wPt-0472cluster	20.8	24.08	16.2917	0.52
		4D	Pend	wPt-0472cluster	18.8	26.89	13.904	0.64
		5A-1	Corv	wPt-4419	102.4	3.483	4.4523	0.05
	ExT2	2B	Corv	wmc175	234	5.241	6.8583	0.13
		2B	Pend	wmc175	234	5.437	5.6995	0.13
		4B	Corv	barc20cluster	146.2	8.395	-9.3857	0.2
		4B	Pend	barc20cluster	146.2	7.61	-7.1513	0.18
		4D	Corv	wPt-0472cluster	14.7	9.64	11.3276	0.3
		4D	Pend	wPt-0472cluster	14.7	8.467	9.0307	0.29
		7A	Corv	wPt-0961cluster	243.8	5.826	6.9774	0.15
		7A	Pend	wPt-0961cluster	245.2	6.142	5.4881	0.14
		7B	Corv	tPt-8569cluster	26.2	4.315	5.7242	0.1
	TxN	2A	Pend	gwm515	98.2	3.759	2.0655	0.09
		4B	Corv	gwm513	36.2	3.868	-2.0761	0.06
		4B	Pend	gwm513	36.2	6.203	-2.1745	0.1
		5A	Corv	barc330	40.2	5.992	2.6729	0.1
		5A	Pend	barc330	38.2	4.531	1.9373	0.08
		5B	Pend	305092cluster	124.2	4.159	1.6379	0.06
Seeds per Spike	ExT1	2A	Corv	wPt-1142cluster	171.8	6.066	-3.2037	0.15
		3B	Corv	wPt-3041	7.7	5.428	3.4932	0.18
		5A-1	Corv	barc186	56.2	3.298	2.498	0.09
	ExT2	3A	Pend	wPt-3697cluster	73.4	3.015	2.4179	0.11
		4A	Pend	wPt-2301	199.5	3.142	-2.2382	0.1
		4B	Pend	wPt-0391	179.6	6.654	4.3063	0.39
	TxN	2A	Corv	gwm372cluster	110.7	12.37	-2.8105	0.18
		2A	Pend	wPt-1142cluster	115.8	11.01	-2.1352	0.17
		4A	Corv	rPt-2478cluster	116	4.39	1.5737	0.05
		5B	Corv	310624	71.3	3.078	-1.298	0.04

"Pop." refers to mapping population. "TxN" refers to Tubbs x NSA 98-0995 population. "ExT1" and "ExT2" refer to the large and small Einstein by Tubbs populations respectively. "Chom." is chromosome. "Loc." is location. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R_2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Table 12 QTLs for fertile spikelets per spike and sterile spikelets per spike.

Trait	Pop.	Chrom.	Loc.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
Fertile Spikelets	ExT1	2A	Corv	gwm372cluster	175.2	6.7	-0.5348	0.11
per Spike		2A	Pend	wPt-1142cluster	173.8	6.864	-0.4372	0.12
		4A	Corv	wPt-6757	55.5	5.307	-0.473	0.08
		4B	Pend	gwm149	5.8	3.54	-0.413	0.09
		5A-2	Pend	wPt-1200	0	3.591	-0.3091	0.06
		7A	Corv	wPt-5987	190.5	16.26	0.8717	0.29
		7A	Pend	wPt-5987	191.5	14.12	0.6934	0.29
		7B	Corv	gwm111	136.8	6.445	-0.5216	0.1
	ExT2	5A	Corv	gwm666	16.3	5.252	-0.7029	0.2
		7A	Corv	wPt-5987	225.8	13.3	1.01	0.4
		7B	Corv	gwm111	83.2	4.237	-0.5263	0.11
	TxN	2A	Corv	348413cluster	117.8	9.019	-0.4952	0.13
		2A	Pend	gwm95	108.4	8.312	-0.363	0.12
		3B	Corv	wPt-0280cluster	179.8	3.198	0.2744	0.04
		4B	Pend	gwm251	41.5	3.925	-0.2389	0.05
		5A	Pend	B1	136.4	4.573	-0.2656	0.06
		5B	Corv	wPt-2273cluster	93	4.018	-0.3364	0.06
		5B	Pend	wPt-2273cluster	90.1	4.184	-0.2404	0.05
Sterile Spikelets	ExT1	2A	Pend	wPt-0094cluster	205.5	4.108	0.2084	0.09
per Spike		5A-2	Corv	wPt-7061	4.2	5.097	0.2326	0.14
		5A-2	Pend	wPt-7061	4.2	7.46	0.3165	0.2
	ExT2	2D	Pend	wPt-9580	5.9	5.248	0.3206	0.14
		3A	Pend	wmc11	214.4	3.959	-0.2741	0.11
		4B	Corv	rPt-6847	167.6	5.42	-0.2985	0.21
		4B	Pend	rPt-6847	171.6	5.514	-0.4274	0.25
		5A	Corv	B1	156.7	6.631	0.2995	0.23
		5A	Pend	B1	156.7	6.206	0.3669	0.19
	TxN	2D	Corv	cfd51cluster	8.2	5.667	-0.1537	0.08
		2D	Pend	gwm296	2.2	3.797	-0.1414	0.05
		4D	Corv	wmc473cluster	11	7.313	0.1861	0.11
		4D	Pend	wmc473cluster	13	10.78	0.2503	0.17
		5A	Corv	B1	134.4	11.01	0.2221	0.17
		5A	Pend	B1	134.4	9.663	0.2097	0.12

"Pop." refers to mapping population. "TxN" refers to Tubbs x NSA 98-0995 population. "ExT1" and "ExT2" refer to the large and small Einstein by Tubbs populations respectively. "Chom." is chromosome. "Loc." is location. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R_2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Table 13 QTLs for seeds per fertile spikelet and seed weight.

Trait	Pop.	Chrom.	Loc.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
Seeds per	ExT1	3B	Corv	wPt-3041	11.7	6.021	0.1531	0.17
Fertile Spikelet		4A	Pend	wPt-1007cluster	153.7	4.725	-0.1164	0.12
		5A-1	Corv	barc186	56.2	3.212	0.1082	0.09
	ExT2	4A	Pend	wPt-6900	198.3	5.252	-0.129	0.17
		4B	Pend	wPt-0391	179.6	3.925	0.1443	0.21
		6A	Corv	tPt-6278	135.3	3.269	-0.1058	0.12
		6A	Pend	tPt-6278	135.3	3.254	-0.1005	0.1
	TxN	2A	Corv	gwm372cluster	110.7	3.268	-0.0609	0.06
		2A	Pend	wPt-1142cluster	112.7	3.823	-0.053	0.06
		3A	Pend	gwm2	161.9	3.864	-0.0604	0.07
		5A	Pend	wPt-4203cluster	138.4	4.213	0.0553	0.06
Seed Weight	ExT1	1A	Pend	barc83cluster	58.8	3.451	-0.8668	0.08
		5A-2	Pend	wPt-7061	2.2	3.416	0.822	0.07
		7A	Corv	gwm332	202.4	5.306	-1.7144	0.17
		7B	Corv	wPt-1266	132.6	4.423	1.3819	0.11
		7B	Pend	wPt-4258cluster	128.2	6.272	1.1377	0.14
	TxN	2B	Corv	wPt-4417cluster	115.7	3.356	0.8225	0.05
		4A	Corv	311358cluster	98.7	7.414	-1.2428	0.1
		4D	Corv	wmc331	15	7.699	1.5327	0.16
		4D	Pend	wmc331	15	5.61	0.8952	0.1
		5A	Pend	gwm291	130.4	3.487	0.5972	0.04
		5B	Corv	wPt-3457cluster	137.2	6.937	1.5656	0.16
		5B	Corv	barc4cluster	67.5	4.122	0.8847	0.05
		5B	Pend	wPt-3329	149.2	5.66	1.2411	0.19

"Pop." refers to mapping population. "TxN" refers to Tubbs x NSA 98-0995 population. "ExT1" and "ExT2" refer to the large and small Einstein by Tubbs populations respectively. "Chom." is chromosome. "Loc." is location. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R_2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Significant differences existed between RILs for grain protein, seed weight and spikes per m^2. Thus, the failure to detect QTLs for these traits is likely the result of a lack of statistical power. For spikes per m^2, the lack of statistical power is probably the result of a lack of precision in the measurement of that trait. The relatively few (81) RILs mapped in the ExT2 population could explain the failure to detect QTLs for seed weight in that population. A smaller number of lines used in QTL mapping results in less statistical power (Zeng, 1994). It is not immediately clear why QTLs weren't mapped for grain protein in the ExT1 population. Grain protein had higher heritability in the ExT1 population than it did in the other two populations (Table 5). Since power to detect QTLs is inversely proportional to heritability, this should have been the population that had the greatest power to detect QTLs for grain protein (Beavis, 1994). The best explanation for the failure to detect these QTLs is that genetic control of grain protein is complex or that it is more heavily influenced by environment than genetics.

QTLs for a trait that occur at nearly the same position on the genetic map for both locations likely represent a single QTL that is effective across those locations (Jiang and Zeng, 1995). These QTLs are useful to breeders, because they can be selected for with minimal consideration of GxE interactions. QTLs of this type were detected for all traits except for grain fill duration, grain protein, and yield. The lack of across location QTLs detected for grain fill duration may be attributed to the high standard error of the RIL means relative to their spread (Tables 2, 3, and 4). This high

relative standard error indicates that the measured values for this trait are too inaccurate for consistent QTL mapping. Grain protein is highly influenced by environmental factors (Gonzalez-Hernandez *et al.*, 2004). Failure to account for in field variability of those environmental factors could be limiting the ability to detect QTLs for grain protein. For yield, the lack of across environment QTLs is probably attributable to GxE interactions and complex genetic control of the trait. The genetic control of this trait may poorly fit the additive genetic model that is assumed by QTL analysis using CIM (Zeng, 1994). Techniques that allow for epistatic interactions, such as MIM, could have greater power to detect QTLs for yield (Kao *et al.*, 1999). However, an overall lack of statistical power to detect sufficient quantities of QTLs to explain a meaningful percent of variation for yield will probably still exist.

There are several instances of QTLs occurring on the same chromosome for a particular trait in more than one population. This may indicate that a single QTL occurs in more than one population. If this is the case, it would provide some measure of conformation for that QTL. Such an assumption should be made with great care, because of the large differences between linkage maps of different populations and the sparseness of those maps. However, these finding at least show that one chromosome is important to a trait in more than one population. Examples of this include 7A for fertile spikelets per spike in ExT1 and ExT2 populations and 2D for sterile spikelets per spike in the ExT2 and TxN populations (Table 12). Finer mapping is needed to determine if only one QTL is actually present in those populations.

There are four loci that show close linkage or pleiotropy with 36% of the detected QTLs. These loci are *Rht-B1*, *Rht-D1*, *B1*, and *Xgwm372* on chromosomes 4B, 4D, 5A, and 2A respectively. The large numbers of QTLs found at these loci indicates that they are strong determinants of the variability present in these populations and further examination is warranted.

Rht-B1 and *Rht-D1* are the loci for the two major dwarfing genes in these populations. Since the ExT populations are segregating for dwarfing alleles at *Rht-B1* and *Rht-D1*, the strong QTLs for height on 4B and 4D can be explained by the dwarfing genes (Table 11). The TxN population is fixed at these loci, so the QTL for height on 4B in that population represents a minor QTL for height. This QTL appears to be closed linked to *Rht-B1*, because its nearest marker, *gwm513*, maps closely to the QTLs for *Rht-B1* in both ExT populations. The height QTLs account for 10 of the 28 QTLs detected in close proximity to these loci. Pleiotropy may explain the occurrence of many of the other QTLs. If the other QTLs are not due to pleiotropy, it may be possible to break unfavorable linkages between QTLs. Finer mapping of these chromosomes combined with multiple-trait QTL mapping would be able to better position the QTLs on these chromosomes and be able to distinguish between pleiotropy and close linkage (Jiang and Zeng, 1995). This is needed to determine how useful the QTLs detected at these loci are to breeders. Pleiotropic effects are probably not very useful, unless they indicate that one dwarfing gene is more favorable than the other.

The *B1* locus is the location of the awn inhibitor gene that is segregating in all of the populations. A total of 15 QTLs showed pleiotropy or close linkage with this gene. The awned allele is associated with more sterile spikelets and fewer fertile spikelets. For sterile spikelets per spike there are QTLs at both locations for all three populations at the *B1* locus (Table 12). QTLs for fertile spikelets per spike are present at Corvallis for the TxN and ExT1 populations and Pendleton for the ExT2 population near the *B1* locus (Table 12). Based on QTLs detected for seed weight (Table 13) and test weight (Table 9) close to *B1*, it appears that awned varieties have higher seed weights and greater test weights. Laperche *et al.* (2007) also detected QTLs for seed weight at the *B1* locus. However, they found greater seed weight for awned varieties at low soil nitrogen levels and greater seed weight for awnless varieties at high soil nitrogen levels. Thus, the effects of these QTLs appear to depend on production and environmental practices.

As with the *Rht-B1* and *Rht-D1* loci, finer linkage mapping and multiple-trait QTL mapping can be used to find more precise positions of these QTLs and test for pleiotropic effects (Jiang and Zeng, 1995). If the QTLs prove to be pleiotropic, then it appears that there is a trade-off of number of fertile spikelets for greater seed weight when choosing awned varieties under the conditions studied. These findings indicate that selection of an awned or awnless variety is not trivial, but it is not obvious if one type is universally preferred over the other. A preference for awned wheat has been shown in studies in the Great Plains, attributed to awns the available photosynthetic

area of a variety (Donald, 1968). However, it has been suggested that the awnless trait makes varieties less prone to pre-harvest sprouting and thus preferable in regions prone to pre-harvest sprouting (King and Richards, 1984). Further examination is needed to identify which plant type is preferable for Oregon.

Another 8 QTLs were detected on 5A in the general vicinity of *Xbarc330*. When combined with the QTLs detected at *B1*, at least one QTL was mapped on chromosome 5A for sterile spikelets per spike, fertile spikelets per spike, flowering, height, seed weight, seeds per fertile spikelet, seeds per spike, and test weight. Other researchers have also observed QTLs for yield and yield components on chromosome 5A (Kato *et al.*, 1998). This indicates that 5A is a good candidate for further research. Important genes that have been mapped to this chromosome include *Vrn-A1*, a gene that controls vernalization requirement; and *Rht12*, a dwarfing gene that is thought to be tightly linked with *B1* (Worland *et al.*, 2006). Nothing is known about the allelic variation of these loci in the populations.

Thirteen QTLs were detected near the *Xgwm372* locus in the TxN and ExT1 populations. It is not immediately apparent what gene or genes are responsible for these QTLs. QTLs found near this locus include those for seeds per fertile spikelet (Table 13), seeds per spike (Table 11), and fertile spikelets per spike (Table 12). These spike fertility traits are the yield components that are most consistently correlated with yield (Tables 6, 7, and 8), so these QTLs might be useful for increasing yield. It appears that all of the favorable alleles are coming from the

Nickerson parent in their respective populations. Thus, it is only of minor consequence if the QTLs are pleiotropic or closely linked. Selecting for the favorable allele at the *Xgwm372* locus could provide a bump to yield. However, no QTLs for yield were detected at this locus. Thus, a better understanding of the interactions with this locus and others is probably needed to maximize any potential gain.

4.2. Factor Analysis

Four factors were retained for the Corvallis spike morphology traits and three factors were retained for the Pendleton traits based on significant correlations with yield, test weight, or grain protein (Table 14). This criterion for selecting the number of factors was used to guard against throwing out potentially informative factors. Selection of factors based on the proportion of variation explained can result in disregarding a factor that only explains a small portion of the total variation, but is important for explaining the economically important traits: yield, test weight, and protein. The significant correlation criterion is quite conservative, because the many degrees of freedom used to test correlations can identify correlations as significant that may not be very meaningful.

Table 14 Pearson product-moment correlations between initial factors and yield, grain protein, and test weight.

Factor	Yield	Test weight	Protein
Corvallis			
F1	0.19**	-0.22***	-0.28***
F2	0.20**	0.04	0.19**
F3	-0.04	-0.30***	0.01
F4	0.22***	-0.01	0.15*
F5	-0.03	0.06	0.09
Pendleton			
F1	0.16**	-0.09	-0.12
F2	0.27***	-0.00	0.20***
F3	0.12*	-0.15*	-0.12
F4	0.01	-0.08	0.12
F5	-0.05	0.03	-0.01

*, **, and *** represent 0.05, 0.01, and 0.001 significance levels, respectively

After rotation, Corvallis factors three and four are essentially factors for single traits (Table 15). These traits are seed weight for factor three and sterile spikelets for factor four. Since QTL mapping was already carried out for these traits individually, these factors were not considered further. The remaining three spike morphology traits primarily loaded on the first two factors (Table 15). Factor one can be thought of as a factor for the seeds per fertile spikelet contribution to seeds per spike and factor two can be thought of as a factor for the fertile spikelets per spike contribution to seeds per spike. Using this reasoning, the QTLs detected for factor one would be expected to be the same as those for seeds per fertile spikelet. The QTLs detected for factor two would be expected to be the same as those for fertile spikelets per spike.

One QTL was detected for factor one at Corvallis (Table 16) and that QTL was found at the same site as the only QTL detected for seeds per fertile spikelet at that location (Table 13). Three QTLs were detected for factor two (Table 16) and three QTLs were detected for fertile spikelets per spike (Table 12). Each had matching QTLs on 2A and 5B, but the third QTL differed in the two traits. The LOD threshold for declaring a QTL significant was 3.1 for both of these traits. On 3B, the LOD score for seeds per fertile spikelet exceeded this threshold by a score less than 0.1 LOD over and factor two was below the threshold by a score less than 0.1 LOD under (results not shown). Thus, QTL mapping of these traits at that locus was essentially the same. On chromosome 3A a significant QTL was found for factor two (Table 16), but none was found for fertile spikelets per spike (Table 12).

Table 15 Factor loadings for rotated spike morphology traits.

Factor	fertile spikelets / spike	sterile spikelets / spike	seeds / fertile spikelet	seed weight	seeds / spike	variance explained
			Corvallis			
1	0.06	-0.16	0.98	-0.09	0.84	34.40%
2	0.98	-0.07	-0.11	-0.16	0.50	25.05%
3	-0.16	0.04	-0.05	0.98	-0.14	20.26%
4	-0.07	0.98	-0.14	0.04	-0.15	20.23%
Communuality	1.00	1.00	1.00	1.00	1.00	100%
			Pendleton			
1	0.14	-0.11	0.99	-0.11	0.87	35.42%
2	0.95	-0.52	-0.03	0.06	0.48	28.31%
3	-0.03	0.66	-0.12	0.92	-0.12	26.37%
Communuality	0.93	0.73	0.99	0.86	0.99	90.11%

"Communuality" represents proportion of the variation for that variable explained by the factors.

Table 16 QTLs for spike morphology factors.

Factor	Chrom.	Nearest Marker	Position (cM)	LOD Score	Add. E.	R^2
		Corvallis				
1	2A	gwm372cluster	110.7	5.391	-0.296	0.086
2	2A	348413cluster	117.8	8.753	-0.365	0.132
	3A	barc346	145.5	3.482	0.3069	0.094
	5B	wPt-2273cluster	93	3.449	-0.23	0.051
		Pendleton				
1	2A	wPt-1142cluster	113.8	4.616	-0.255	0.063
	3A	gwm2	161.9	3.27	-0.248	0.061
	5A	wPt-4203cluster	138.4	6.068	0.3005	0.088
2	2A	gwm95	108.4	6.139	-0.296	0.087
	5A	B1	136.4	8.067	-0.345	0.117
	1B	rPt-9074	147.5	3.184	0.2203	0.047
	4B	gwm513	36.2	3.374	-0.221	0.047
	5B	378650cluster	88.1	3.005	-0.195	0.037
	6B	wmc494cluster	59.7	3.071	-0.197	0.038
3	5A	B1	134.4	6.618	0.2897	0.083
	5B	wPt-3329	149.2	5.099	0.4005	0.159
	4D	wmc331	15	9.329	0.3989	0.155

"Factor" corresponds to factors in table 14. "Chom." is chromosome. "Add. E." is the additive effect for replacing an allele from a Nickerson parent with an allele from Tubbs. "R^2" is the coefficient of determination for how the QTL explains the residuals after accounting for the covariates.

Overall, the two Corvallis factors detected or nearly detected the same significant QTLs that were found for seeds per fertile spikelet and fertile spikelets per spike as expected. The factors also identified an additional QTL not found mapping the individual traits. These results indicate that factor analysis may have increased power to detect QTLs. However, there was a QTL for seeds per spike on 4A that didn't show up with the factors or the other traits (Table 11). This could indicate that the factor analysis didn't account for some of the genetic variation in seeds per spike. On the other hand, it is possible that this QTL is false. Analyzing seeds per spike for QTLs after already scanning the two traits that should account for all of its variation inflates the probability of type I error (Jiang and Zeng, 1995). It is important to know which of the QTLs detected were real and which are false positives in order to adequately rate the benefits of using the factor analysis approach. The best way to do this is to collect more years of data and compare the results of multiple years.

The rotated factors for Pendleton didn't show single trait factors like Corvallis (Table 15). Factor one can be considered a factor for the seeds per fertile spikelet contribution to seeds per spike just like factor one at Corvallis. Factor two can be considered the fertile spikelets per spike contribution to seeds per spike, but it also has a meaningful negative loading of sterile spikelets per spike. Biologically this could make sense. The number of total spikelets per spike is determined at an early growth stage and when the plant experiences stress at a later growth stage it will abort some of these spikelets (Acevedo *et al.*, 2002). The abortion of spikelets results in sterile

spikelets. There is a higher proportion of sterile spikelets to fertile spikelets at Pendleton than there is at Corvallis (Table 2). This is probably explained by greater water and heat stress at Pendleton that caused the formation of sterile spikelets. It may be that lines that produced more spikelets per spike were also less likely to abort those spikelets. This reasoning is supported by the significant inverse correlation between fertile spikelets per spike and sterile spikelets per spike that was observed at Pendleton (Table 6). Factor three loaded seed weight and sterile spikelets per spike (Table 15). The biological explanation for this factor could be that lines that aborted more spikelets had relatively less seeds to act as a sink for photosynthates compared to their production of photosynthates. Thus, these lines produced larger seeds on average.

For the sake of comparing the QTL mappings of the Pendleton factors to those of the single traits, factor one QTLs are compared to QTLs for seeds per fertile spikelet, factor two is compared to fertile spikelets per spike, and factor three is compared to seed weight. Factor one mapped QTLs (Table 16) that were essentially the same as those for seeds per fertile spikelet (Table 13). Factor two mapped QTLs (Table 16) that included essentially the same QTLs that were mapped for fertile spikelets per spike (Table 12). Factor two also identified QTLs on 4B and 6B that were not identified by fertile spikelets per spike. Factor three found QTLs (Table 16) that were essentially the same as those for seed weight (Table 13).

Pendleton factors one and three behaved the same as their equivalent single traits. Factor two appeared to have greater power to detect QTLs than its equivalent

single trait. The increased power may be attributable to considering multiple traits at once with the factor scores. In particular, the inverse correlation of sterile spikelets per spike with fertile spikelets per spike may be increasing detection power in a similar way to the power increase observed in multiple-trait QTL analysis (Jiang and Zeng, 1994).

There is a lack of conclusive evidence supporting using factor analysis as a method for improved QTL mapping. However, the method did find QTLs that weren't previously detected. This provides some anecdotal evidence that the procedure has merit. Without knowing the true composition of QTLs for these traits, it cannot be determined if this is the result of increased power or simply error. Ultimately, the best way of analyzing this method is through simulation studies were the true distribution of QTLs is known. The results in this study indicate that such a simulation study should be performed. If the simulation study does show increased power and reduce error using factor analysis, the results from this method are probably more useful than those gathered using the traits individually.

5. CONCLUSION

A large number of QTLs were detected for traits related to yield in three genetic mapping populations representative of populations that produce elite germplasm in OSU's wheat breeding program. These QTLs may eventually be useful for developing MAS procedures for lines in OSU's program. For example, selection for the favorable allele at the *Xgwm372* locus can be implemented in relatively short order. This should result in wheat that has greater spike fertility, and more fertile spikelets. However, greater understanding of all of the detected QTLs and how they are interrelated is needed to maximize the effectiveness of a MAS program.

This greater understanding includes finer genetic mapping to identify genetic markers that are more closely linked to the identified QTLs and are inexpensive to use. Additional years of data are also needed to expand the scope of inference for this study. More complicated analyzes could also improve the usefulness of the data collected. Multiple-trait QTL mapping can be used to distinguish between pleiotropy and close linkage of QTLs (Zeng, 1994). This would allow a breeder to determine if it is possible to break unfavorable linkages. The impact of epistasis on these traits also needs to be examined and can be accomplished using MIM (Kao *et al.*, 1999). These two techniques would greatly enhance understanding of genetic control of the examined traits in these populations. Since these populations have similar genetic

backgrounds to other populations used to generate varieties in OSU's wheat breeding program, the effects of the detected QTLs may be very similar in those other populations.

Using factor analysis to handle correlated traits showed promising initial results. Previously undetected QTLs were identified with this approach. However, there is insufficient evidence to conclude that factor analysis increased power to detect QTLs and decreased type I error. A simulation study needs to be conducted to definitively answer these questions.

6. BIBLIOGRAPHY

Acevedo E, Silva P, Silva H (2002) Wheat growth and physiology. In Curtis BC, Rajaram S, Gomez Macpherson H (eds) Bread wheat: improvement and production. Food and Agriculture Organization of the United Nations, Rome

Addisu M, Snape JW, Simmonds JR, Gooding MJ (2009) Reduced height (Rht) and photoperiod insensitivity (Ppd) allele associations with establishment and early growth of wheat in contrasting production systems. Euphytica 166:249-267

Alexander WL, Smith EL, Dhanasobhan C (1984) A comparison of yield and yield component selection in winter wheat. Euphytica 33:953-961

Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker-assisted selection in crop breeding – Prospects and challenges. Current Science 87:607-619

Baenziger PS, McMaster GS, Wilhelm WW, Weiss A, Hays CJ (2004) Putting genes into genetic coefficients. Field Crops Research 20:133-143

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings 49th Annual Corn and Sorghum Industry Research Conference, Washington DC, ASTA

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodberry pp 121-126

Borlaug NE (1983) Contributions of conventional plant breeding to food production. Science 219:689-693

Börner A, Schumann E, Fürste A, Coster H, Leithold B, Röder MS, Weber WE (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L). Theoretical Applied Genet 105:921-936

Buck-Sorlin GH (2002) L-system model of the vegetative growth of winter barley (*Hordeum vulgare* L.). In Polani D, Kim J, Martinetz T. (eds) Fifth German workshop on artificial life, abstracting and synthesizing the principles of living

systems, March 2002, Lübeck, Germany. AKA Akademische Verlagsgesellschaft, Berlin, pp 53-64

Buerstmayr H, Ban T, Anderson JA (2009) QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breeding 128:1-26

Calderini DF, Dreccer MF, Slafer GA (1995) Genetic improvement in wheat yield and associated traits. A re-examination of previous results and the latest trends. Plant Breeding 114:108-112

Campbell BT, Baenziger PS, Gill KS, Eskridge KM, Budak H, Erayman M, Dweikat I, Yen Y (2003) Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. Crop Science 43:1493-1505

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Costello AB, Osborne JW (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation doi:10.1.1.110.9154

Crow JF, Dove WF (1990) Mapping functions. Genetics 125:669-671

Cuthbert JL, Somers DJ, Brûlé-Babel AL, Brown PD, Crow GH (2008) Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). Theoretical Applied Genetics 117:595-608

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39:1-38

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nature Genetics 3:43–52

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285-294

Donald CM (1968) The breeding of crop ideotypes. Euphytica 17:385-403

Donmez E, Sears RG, Shroyer JP, and Paulsen GM (2001) Genetic gain in yield attributes of winter wheat in the Great Plains. Crop Science 41:1412-1419

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA (2002) 'Perfect' markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theoretical and Applied Genetics 105:1038-1042

Engledow FL, Wadham SM (1923) Investigation on yield in the cereals, Part 1. Journal of Agricultural Science 13:390-439

Fonseca S, Patterson FL (1968) Yield component heritabilities and interrelationships in winter wheat (*Triticum aestivum* L.). Crop Science 8:614-617

Gonzalez-Hernandez JL, EM Elias, SF Kianian (2004) Mapping genes for grain protein concentration and grain yield on chromosome 5B of *Triticum turgidum* (L.) var. *dicoccoides*. Euphytica 139:217-225

Hackett CA, Meyer RC, Thomas WTB (2001) Multi-trait QTL mapping in barley using multivariate regression. Genetic Research 77:95-106

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. Journal of Genetics 8:299-309

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324

Haverkort AJ, Kooman PL (1997) The use of systems analysis and modeling of growth and development in potato ideotyping under conditions affecting yields. Euphytica 94:191-200

Hayton JC, Allen DG, Scarpello V (2004) Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. Organizational Research Methods 7:191-205 doi:10.1177/1094428104263675

Jansen RC (1994) Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138:871-881

Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. Theoretical and Applied Genetics 91:33-37

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455

Jiang C, Zeng Z-B (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111-1127

Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis, sixth edition. Pearson Education, Upper Saddle River

Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. Psychometrika 23:187-200

Kao C-H, Zeng Z-B, Teasdale RD (1999). Multiple interval mapping for quantitative trait loci. Genetics 152:1203-1216

Kaeppler SM, Phillips RL, Kim TS (1993) Use of near-isogenic lines derived by backcrossing or selfing to map qualitative traits. Theoretical and Applied Genetics 87:233-237

Kato K, Mirura H, Akiyama M, Kuroshima M, Sawada S (1998) RFLP mapping of the three major genes, *Vrn1*, *Q* and *B1*, on the long arm of chromosome 5A of wheat. Euphytica 101:91-95

Kato K, Miura H, Sawada S (2000) Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. Theoretical Applied Genetics 101:1114–1121

King RW, Richards RA (1984) Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. Australian Journal of Agricultural Research 35: 327-336

Korol A, Mester D, Frenkel Z, Ronin Y (2009) Methods for Genetic Analysis in the *Triticeae*. In Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the *Triticeae*. Springer, New York, pp 163-199

Kosambi DD (1944) The estimation of map distance from recombination values. Annals of Eugenics 12:172-175

Kumar J, Mir RR, Kumar N, Mohan A, Prabhu KV, Balyan HS, Gupta PK (2010) Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. Plant Breeding DOI:10.1111/j.1439-0523.2009.01758.x

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanoeq E, Devleen-Barret F, Gouis J (2007) Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theoretical Applied Genetics 115:399-415

Li JZ, Huang XQ, Heinrichs F, Ganal MW, Röder MS (2006) Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley. Genome 49: 454-466

Li S, Jia J, Wei X, Zhang X, Li L, Chen H, Fan Y, Sun H, Zhao X, Lei T, Xu Y, Jiang F, Wang H, and Li L. (2007) A intervarietal genetic map and QTL analysis for yield traits in wheat. Molecular Breeding 20:167-178

Loomis RS, Rabbinge R, Ng E (1979) Explanatory models in crop physiology. Annual Review of Plant Physiology 30:339-367

Marza F, Bai GH, Carver BF, Zhou WC (2005) Quantitative trait loci for yield and related traits in the wheat population Ning7840 · Clark. Theoretical Applied Genetics 21:1-11

Mi X, Eskridge K, Wang D, Baenziger PS, Campbell BT (2008) Multiple-trait QTL mapping using a structural equation model [abstract]. American Society of Agronomy-Crop Science Society of America-Soil Science Society of America Annual Meeting, November 5-9, Houston, Texas

Murphy LR, Santra D, Kidwell K, Yan G, Chen X, Campbell KG (2009) Linkage maps of wheat stripe rust resistance genes *Yr5* and *Yr15* for use in marker-assisted selection. Crop Science 49:1786-1790

Nerson H (1980) Effects of population density and number of ears on wheat yield and its components. Field Crops Research 3:225-234

Paux E, Sourdille P (2009) A toolbox for Triticeae genomics. In Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the Triticeae. Springer, New York, pp 255-283

Quarrie1 SA, Quarrie SP, Radosevic R, Rancic D, Kaminska A, Barnes JD, Leverington M, Ceoloni C, Dodig D (2006) Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. Journal of Experimental Botany 57:2627–2637

Sherman JD, Yan L, Talbert L, Dubcovsky J (2004) A PCR marker for growth habit in common wheat based on allelic variation at the *VRN-A1* gene. Crop Science 44:1832-1838

Slafer GA, Calderini DF, Miralles DJ (1996) Yield components and compensation in wheat: opportunities for further increasing yield potential. In Reynolds MP, Rajaram S, McNab A (eds) Increasing Yield Potential in Wheat: Breaking the Barriers. CIMMYT, Mexico, pp 101–132

Singh RP, Nelson JC, Sorrells ME (2000) Mapping *Yr28* and Other Genes for Resistance to Stripe Rust in Wheat. Crop Science 40:1148-1155

Snook SC, Gorsuch RL (1989) Component analysis versus common factor analysis: A Monte Carlo study. Psychological Bulletin 106:148-154

Tanguma J (2000) Determining the number of factors to retain. Paper presented at the annual meeting of the southwest educational research association (Dallas, TX, January 27-29, 2000)

Upadyayula N, Wassom J, Bohn MO, Rocheford T R (2006) Quantitative trait loci analysis of phenotypic traits and principal components of maize tassel inflorescence architecture. Theoretical Applied Genetics 113:1395-1407

Van Ooijen JW, Voorrips RE (2006) JoinMap 4.0. Plant Research International, Biometris, Wageningen, Netherlands, (http://www.joinmap.nl/)

Voet D, Voet JG (2004) Biochemistry, 3rd edition. Wiley, New York

Walton PD (1971) The use of factor analysis in determining characters for yield selection in wheat. Euphytica 20:416-421

Wang S, Basten CJ, Zeng ZB (2010) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC, (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Washmon CN, Solie JB, Raun WR, Itenfisu DD (2002) Within field variability in wheat grain yields over nine years in Oklahoma. Journal of Plant Nutrition 25:2655-2662

Worland AJ, Sayers EJ, Börner A (2006) The genetics and breeding potential of Rht12, a dominant dwarfing gene in wheat. Plant Breeding 113:187-196

Wu R, Ma CX, Casella G (2007) Statistical genetics of quantitative traits. Springer, New York

Xu S (2003) Estimating polygenic effects using makers of the entire genome. Genetics 163:789-801

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Sciences 100:6263-6268

Yin X, Chasalow SC, Dourleijn CJ, Stam P, Kropff MJ (2000) Coupling estimated effects of QTLs for physiological traits to a crop growth model: Predicting yield variation among recombinant inbred lines in barley. Heredity 85:539-549

Yin X, Stam P, Kropff MJ, Schapendonk HCM (2003) Crop Modeling, QTL Mapping, and Their Complementary Role in Plant Breeding. Agronomy Journal 95:90-98

Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. Journal of Experimental Botany 56:967-976

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5:505-510

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Zhou L, Bai G, Ma H, Carver BF (2007) Quantitative trait loci for aluminum resistance in wheat. Molecular Breeding 19:153-16

APPENDICES

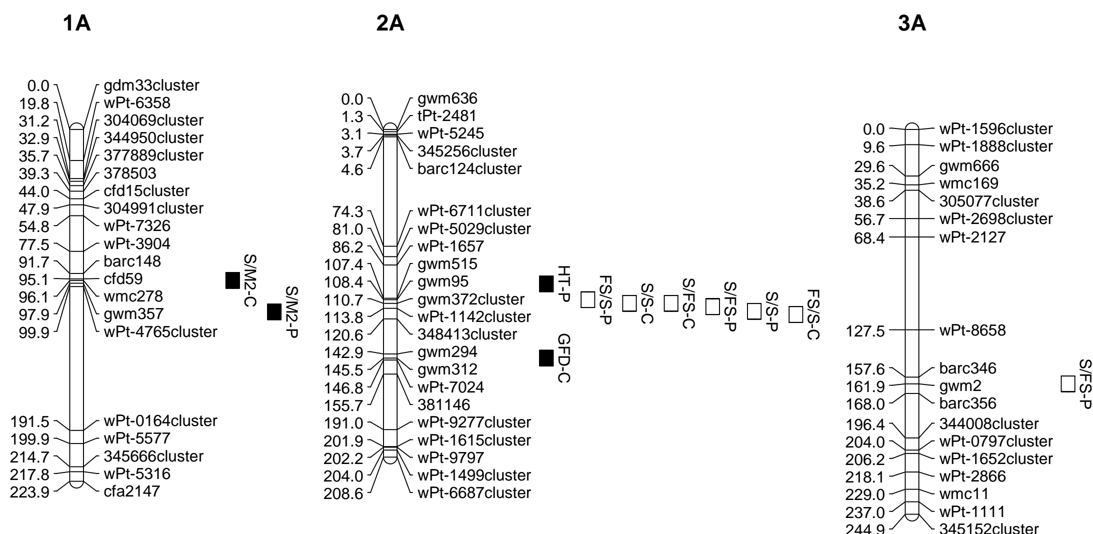

Figure 1 Genetic linkage map for the Tubbs x NSA 98-0995 population (TxN) with QTL positions. Distances are in Kosambi cM. "Cluster" at the end of a marker name denotes the use of a delegate marker. Solid boxes indicate the position of a QTL that has a positive additive effect for the Tubbs allele. Open boxes indicate a negative additive effect for the Tubbs allele. The trait abbreviations are "YLD" for grain yield, "PROT" for grain protein, "TWT" for test weight, "FLW" for growing degree days to flowering, "RIPE" for growing degree days to ripening, "GFD" for growing degree days of grain fill duration, "HT" for height, "S/M2" for spikes per square meter, "FS/S" for fertile spikelets per spike, "SS/S" for sterile spikelets per spike, "S/FS" for seeds per fertile spikelet, "S/S" for seeds per spike, and "SW" for seed weight. At the end of the trait abbreviations "-C" indicates the QTL is for Corvallis and "-P" indicates the QTL is for Pendleton.

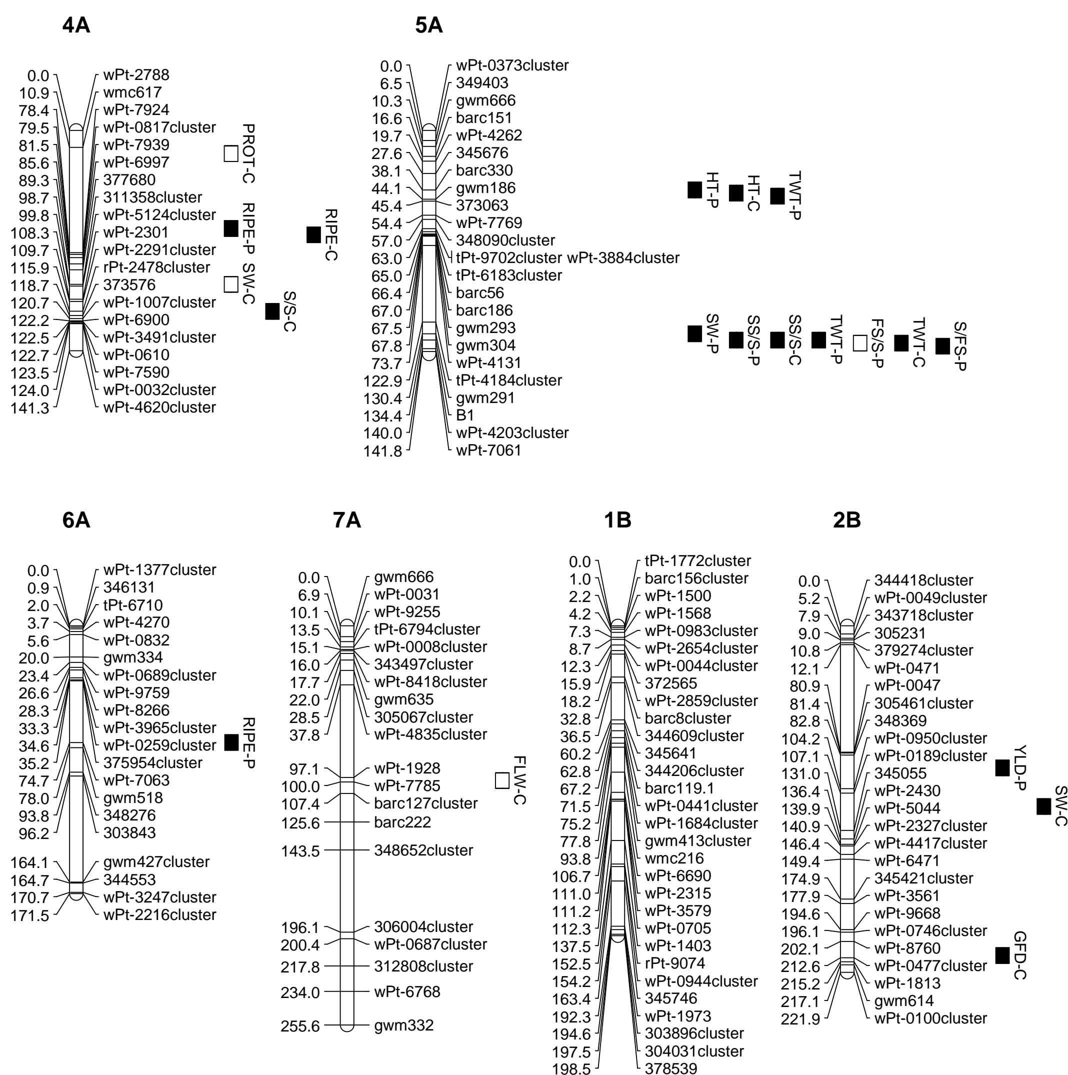

Figure 1 Continued

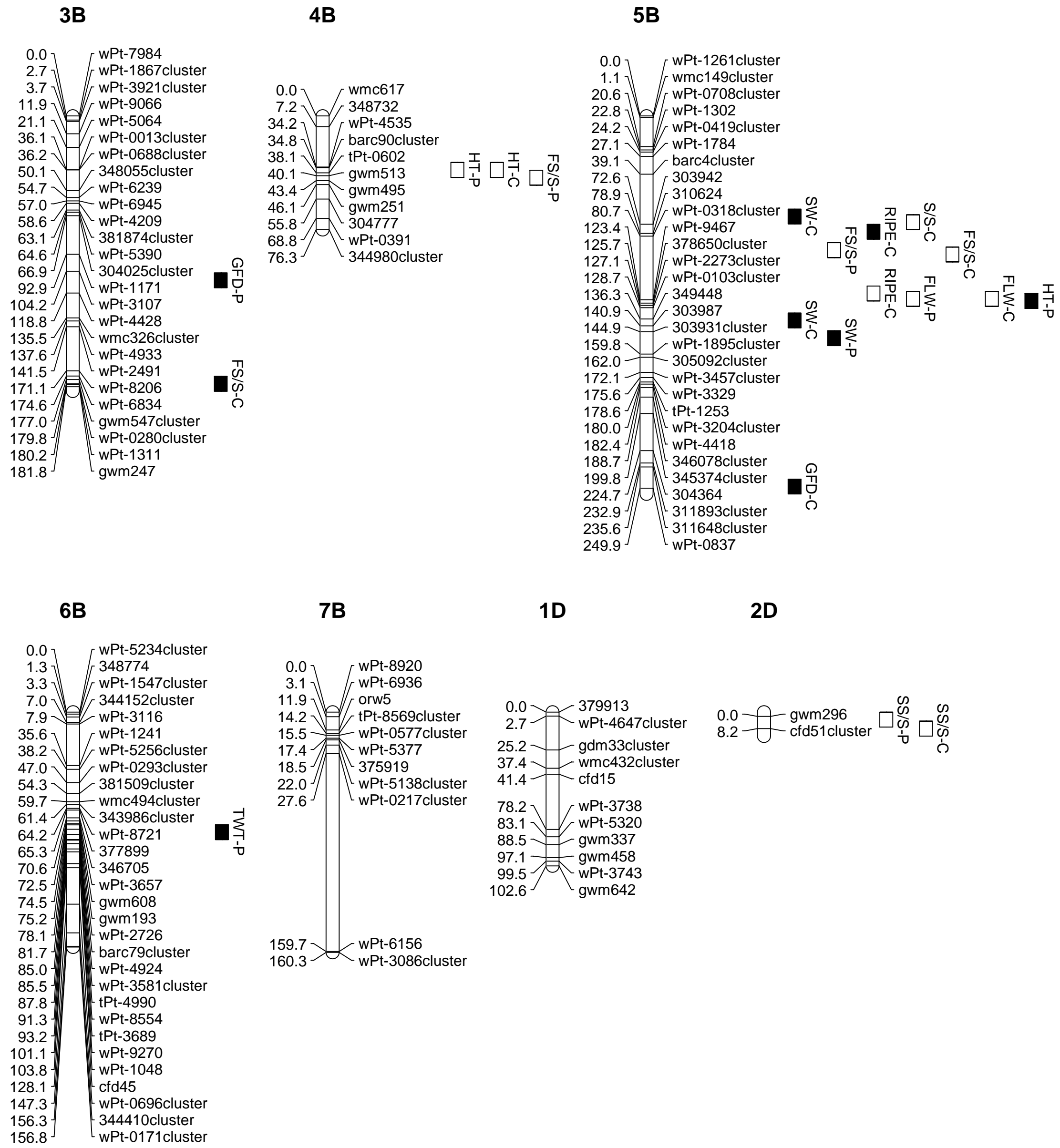

Figure 1 Continued

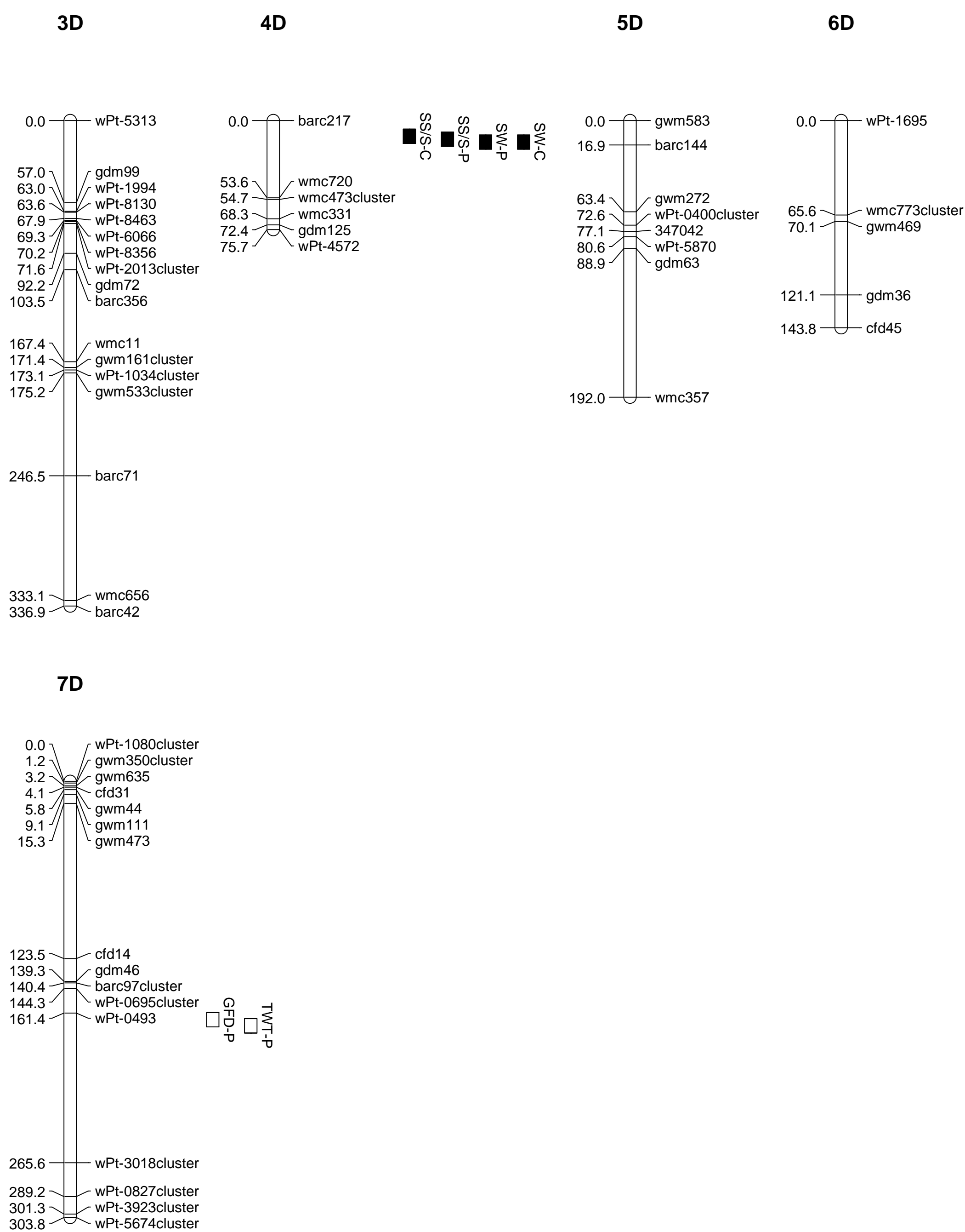

Figure 1 Continued

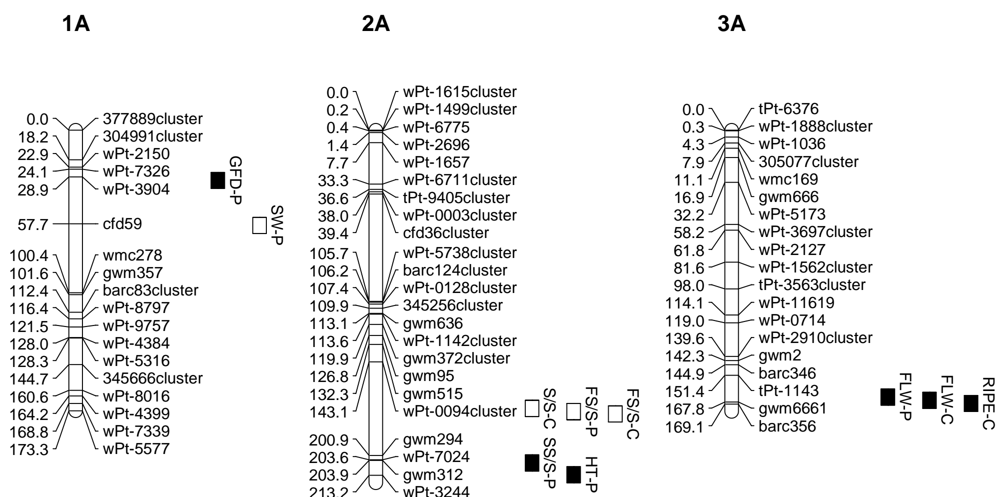

Figure 2 Genetic linkage map for the larger Einstein x Tubbs population (ExT1) with QTL positions. Distances are in Kosambi cM. "Cluster" at the end of a marker name denotes the use of a delegate marker. Solid boxes indicate the position of a QTL that has a positive additive effect for the Tubbs allele. Open boxes indicate a negative additive effect for the Tubbs allele. The trait abbreviations are "YLD" for grain yield, "PROT" for grain protein, "TWT" for test weight, "FLW" for growing degree days to flowering, "RIPE" for growing degree days to ripening, "GFD" for growing degree days of grain fill duration, "HT" for height, "S/M2" for spikes per square meter, "FS/S" for fertile spikelets per spike, "SS/S" for sterile spikelets per spike, "S/FS" for seeds per fertile spikelet, "S/S" for seeds per spike, and "SW" for seed weight. At the end of the trait abbreviations "-C" indicates the QTL is for Corvallis and "-P" indicates the QTL is for Pendleton.

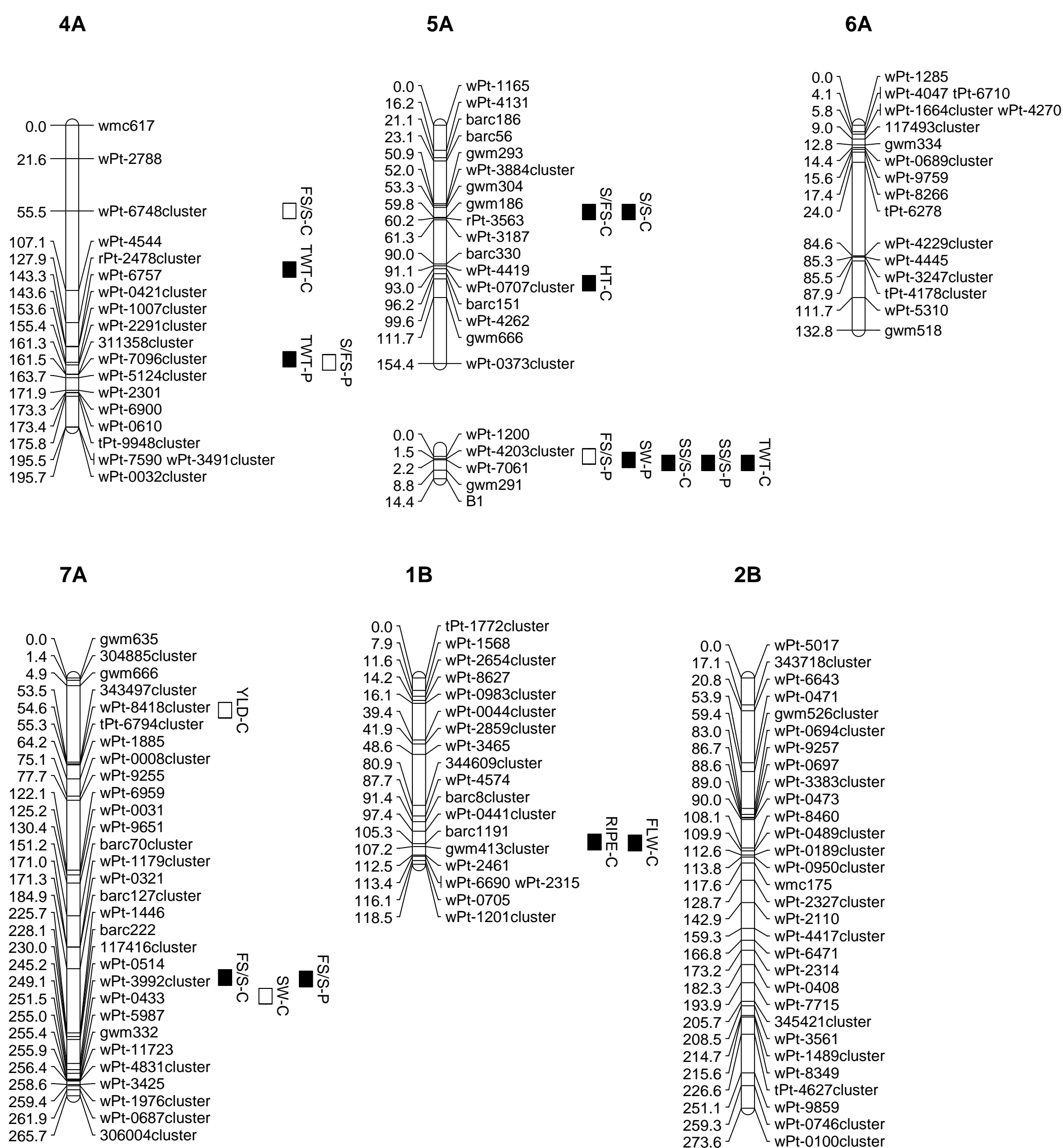

Figure 2 Continued

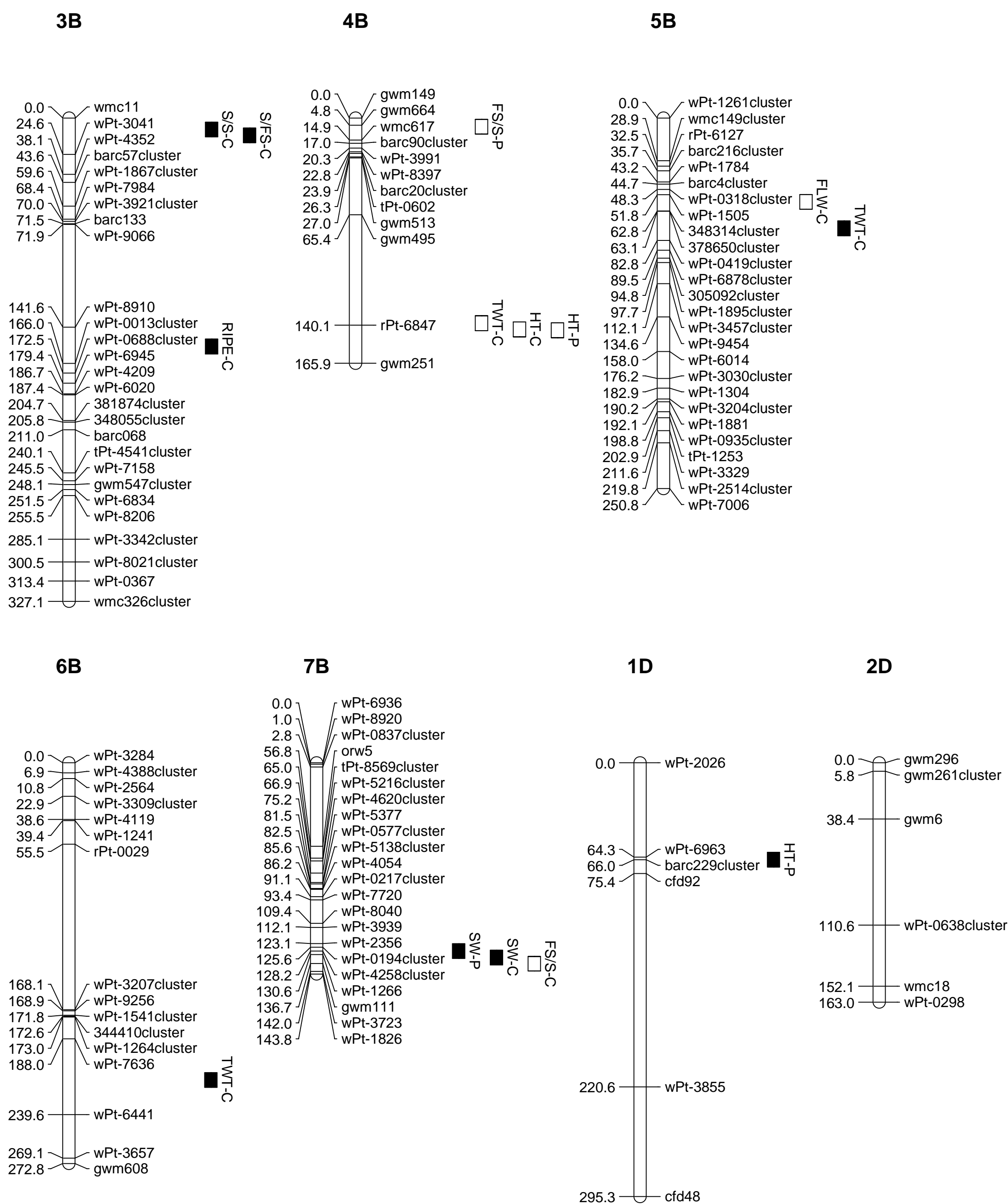

Figure 2 Continued

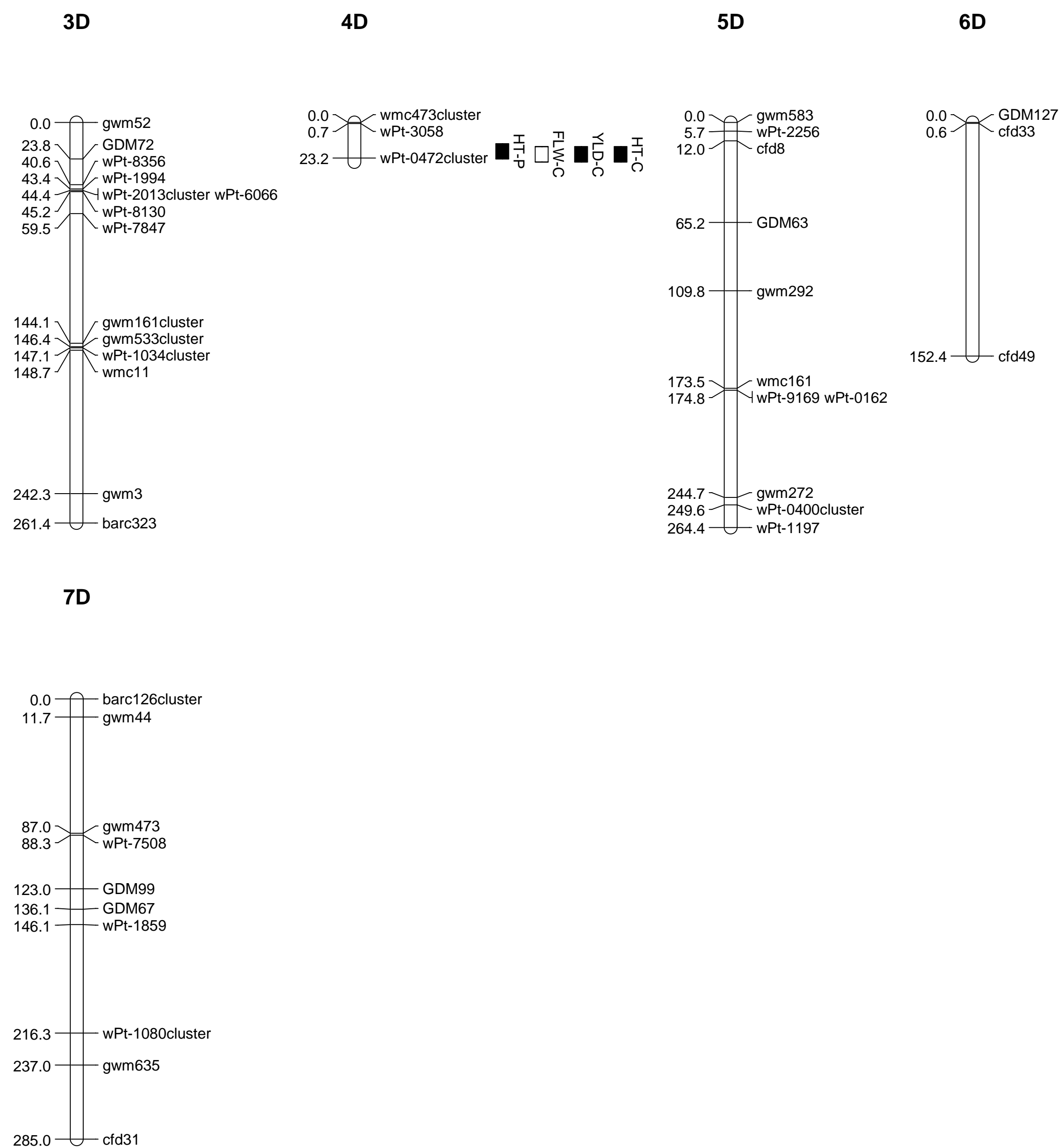

Figure 2 Continued

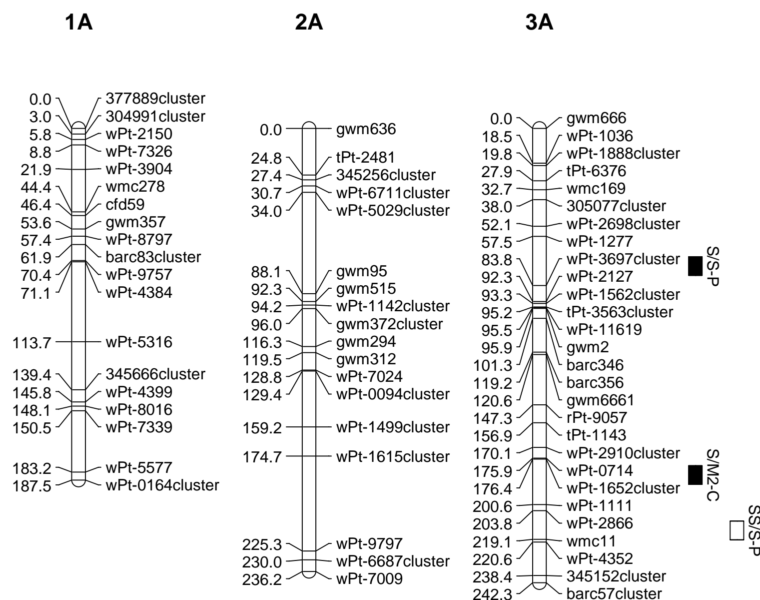

Figure 3 Genetic linkage map for the smaller Einstein x Tubbs population (ExT2) with QTL positions. Distances are in Kosambi cM. "Cluster" at the end of a marker name denotes the use of a delegate marker. Solid boxes indicate the position of a QTL that has a positive additive effect for the Tubbs allele. Open boxes indicate a negative additive effect for the Tubbs allele. The trait abbreviations are "YLD" for grain yield, "PROT" for grain protein, "TWT" for test weight, "FLW" for growing degree days to flowering, "RIPE" for growing degree days to ripening, "GFD" for growing degree days of grain fill duration, "HT" for height, "S/M2" for spikes per square meter, "FS/S" for fertile spikelets per spike, "SS/S" for sterile spikelets per spike, "S/FS" for seeds per fertile spikelet, "S/S" for seeds per spike, and "SW" for seed weight. At the end of the trait abbreviations "-C" indicates the QTL is for Corvallis and "-P" indicates the QTL is for Pendleton.

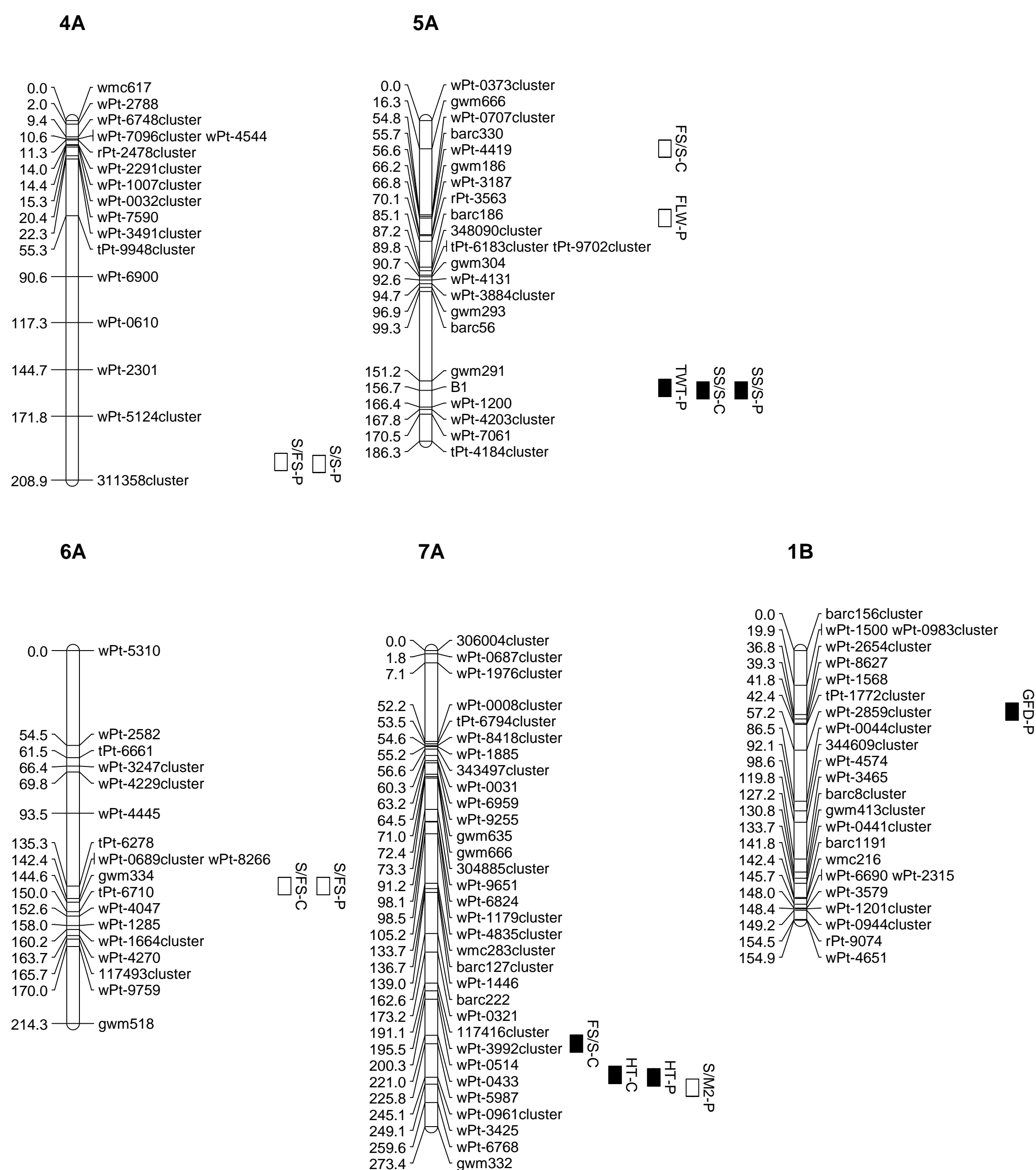

Figure 3 Continued

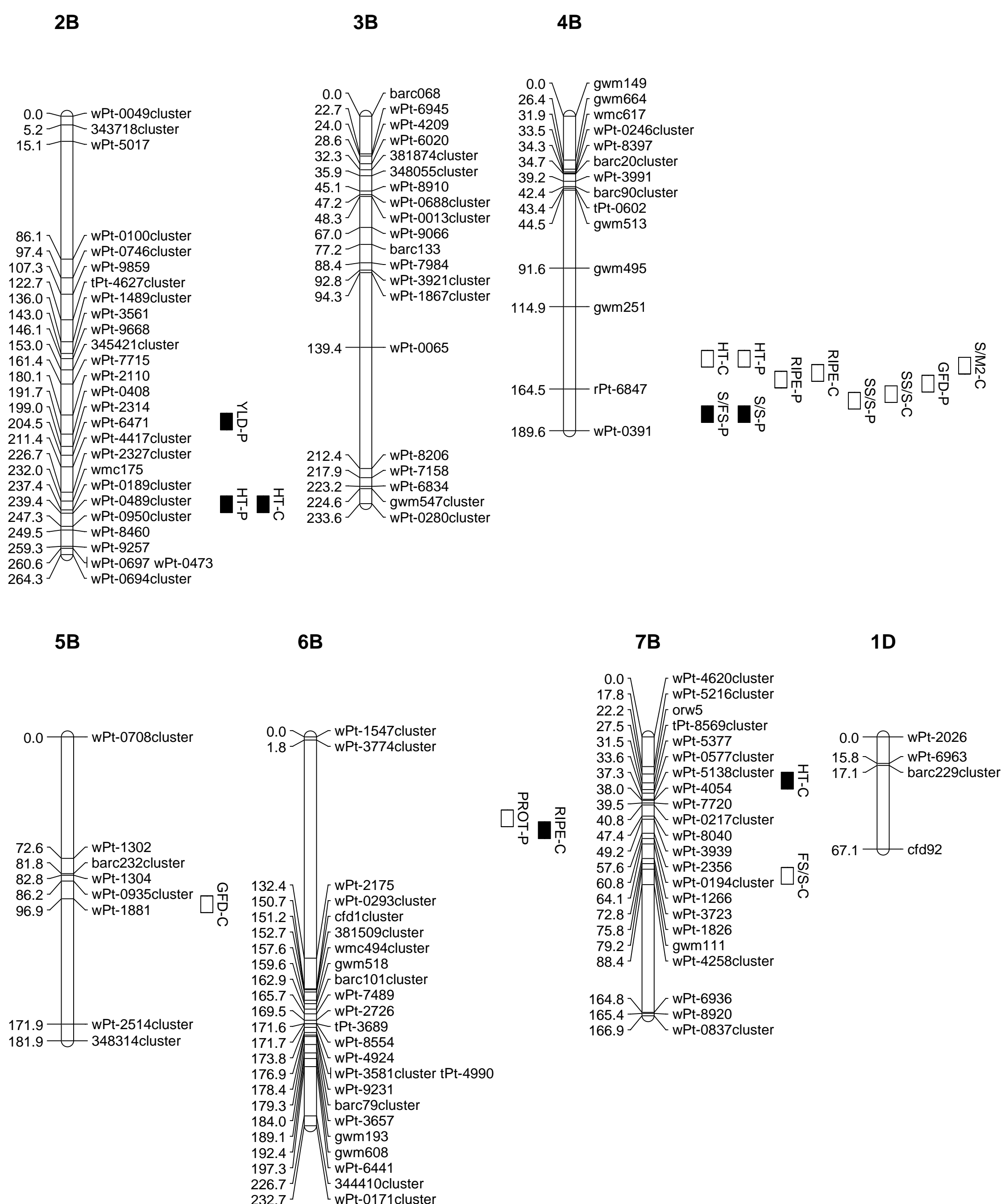

Figure 3 Continued

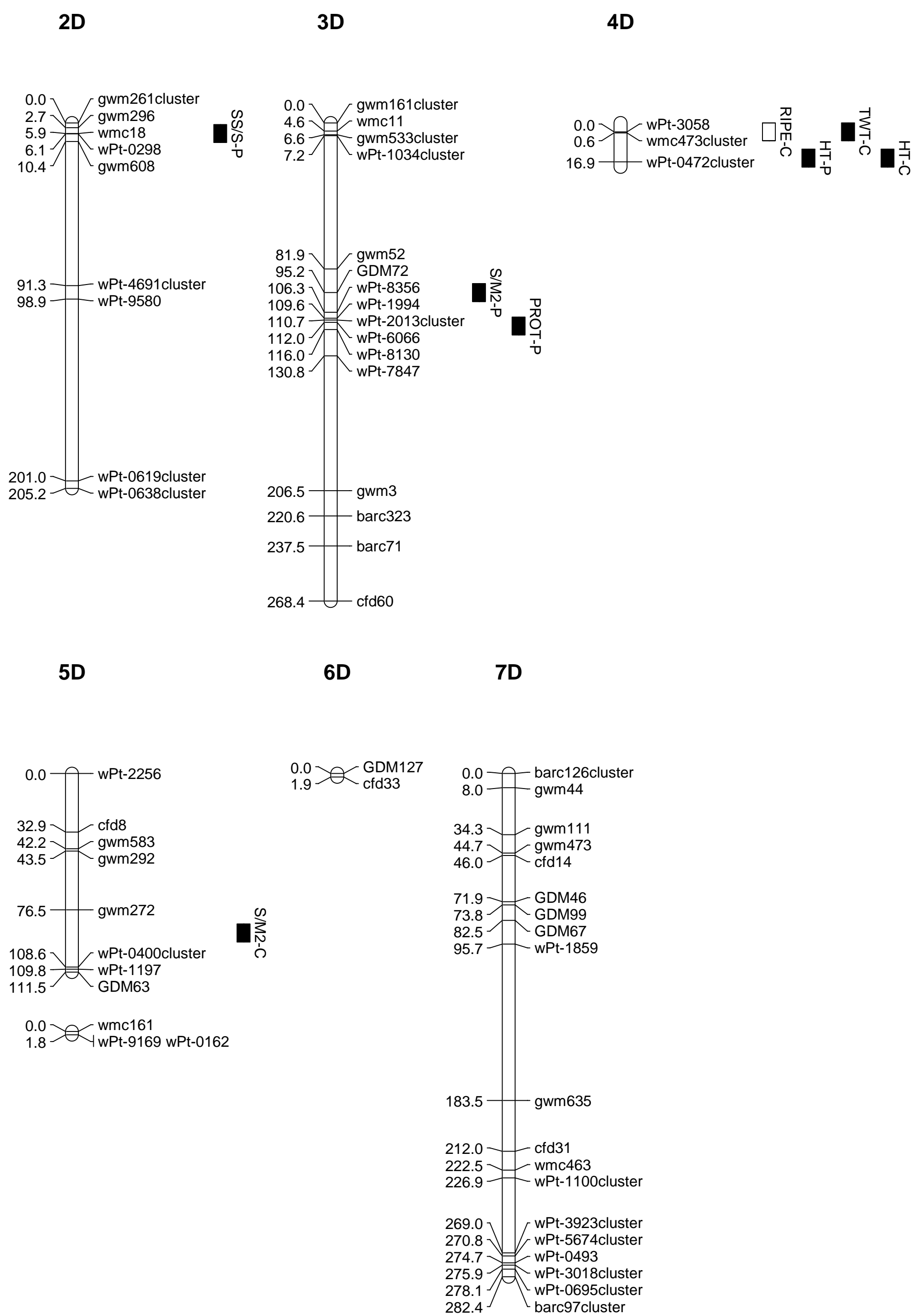

Figure 3 Continued