PRELIMINARY QUANTITATIVE TRAIT LOCI ANALYSIS FOR BIOMASS TRAITS IN *MISCANTHUS SINENSIS*

BY

JUSTIN MICHAEL GIFFORD

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Adviser:

Professor John A. Juvik

Abstract

In light of rising energy costs, lignocellulosic ethanol has been identified as a renewable alternative to the petroleum based transportation fuels. In an attempt to reach government mandated ethanol production levels, potential biofeedstock candidates have been investigated and accessions within the genus *Miscanthus* have been identified as leading contenders in the Midwestern United States due to their high yield and efficiency. The sterile nature of the widely studied *M. x giganteus* clone precludes crop improvement via traditional breeding methods. An alternative strategy to introduce genetic variability and advance performance is through the breeding of *M. sinensis*, one of the progenitor species of *M. x giganteus*. The obligatory outcrossing nature of *Miscanthus* prohibits the formation of inbred lines making it necessary to evaluate large segregating populations from crosses between heterozygous parents. In addition, a plant establishment period of three years is needed to accurately determine the true phenotypic performance of progeny. Establishing correlations between traits exhibited during the establishment phase and plot yields and developing a marker assisted selection program would allow earlier selection. With this in mind, a QTL study was conducted in a pseudo testcross mapping population segregating for flowering time, height, leaf width, and yield. An initial look at the genetic architecture underlying traits important to biomass production in a population of 221 progeny from the cross of *M. sinensis* accession 'Grosse Fountaine' with *M. sinensis* 'Undine' identified 42 QTLs across 15 traits. The use of spring emergence as a covariate to account for variation resulting from the establishment effect increased the power to detect QTLs. Synteny between sorghum and *Miscanthus* has been utilized to identify a plausible gene candidate underlying a flowering time QTL. This analysis sets

the foundation for fine mapping, positional cloning, and the development of a marker assisted selection program in the ongoing effort to improve *Miscanthus* as a biofuel crop.

Acknowledgments

I would like to express my appreciation to my advisor Dr. Jack Juvik for his guidance throughout my master's project. Additionally, I would like to extend my thanks to my graduate committee members Dr. Erik Sacks and Dr. Stephen Moose for their consideration and helpful feedback. I also owe a debt of gratitude to Dr. Ray Ming for the use of his QTL software that made the analysis of my data possible. Lastly, I am grateful to my colleagues Won Byoung Chae, Kang Mo Ku, Talon Becker, and Chris Kaiser as well as my undergraduate assistants Chris Posta, Jane Christensen, and Lukas Looby for their intellectual expertise and assistance with fieldwork and data collection.

Table of Contents

Introduction

Rationale

According to the U.S. Energy Information Administration (US EIA), annual global energy consumption is predicted to rise from 2008 to 2030 by 43% to 721 quadrillion Btu (International Energy Outlook 2011). While a majority of this increase in demand will emanate from developing countries, developed countries such as the United States could see increases in energy demands in excess of 18% over the same time period (Ethanol Myths & Facts 2011). The main driving forces behind these trends are an increase in the world population and improved standards of living.

Breaking down world energy consumption by source highlights the world's dependence on fossil fuels. The US EIA estimates that in 2006 fossil fuels provided slightly over 80% of the energy consumed worldwide with the largest constituent of this being petroleum based fuels followed by coal and natural gas at 42.5%, 32.2%, and 22%, respectively (World Energy Outlook 2008). A very similar story can be told for the U.S.; the latest data approximates that fossil fuels constitute 83.1% of our energy consumption (2010 Annual Energy Review 2011).

A number of concerns arise from relying so heavily on fossil fuels. First and foremost, they are a finite resource. Although proven reserves exist that can likely meet the world's increasing energy demands over the next 40 years, fossil fuels won't last indefinitely. A look at the U.S. energy balance will reveal yet another reason to be concerned about our dependence on fossil fuels. The U.S. produces approximately 75 quadrillion Btu of energy per year yet consumes 98 quadrillion Btu annually (2010 Annual Energy Review 2011). This disparity in the energy balance is reconciled by imports, mostly in the form of petroleum. This dependence on

foreign oil makes the U.S. vulnerable as political instability in oil producing regions, among others things, can cause interruptions in the supply and lead to highly volatile oil prices. Lastly, fossil fuel usage releases millions of tonnes (Mt) of carbon dioxide into the atmosphere every year resulting in the greenhouse effect and global warming; in 2008, the amount of carbon dioxide released from combustion of fossil fuels was estimated to be 29,381 Mt (Key World Energy Statistics 2010). In summary, fossil fuels are a finite resource and our continued dependence on them leads both to a national security risk and global warming. Consequently, it would be prudent to develop alternative energy sources.

The most promising prospective sources of renewable, alternative energy include solar energy, wind power, hydroelectric power, tidal power, geothermal energy, and energy derived from biomass. While it will be necessary to develop each one of these renewable energy sources to help meet our increasing energy demands, energy from biomass may prove to be the most important as it is currently the most suitable alternative for the high octane, liquid fuels that dominate the transportation sector. Ethanol blended fuels are nothing new; they have been in the market for a number of years. Moving forward, the major differences will be the scale at which ethanol is used and the source from which it originates. In accordance with the Energy Independence and Security Act of 2007 (EISA), the U.S. must integrate 36 billion gallons of renewable transportation fuels into the transportation sector by 2022. In order to prevent competition between food and fuel, the EISA prohibits more than 15 of these 36 billion gallons to come from corn grain and further stipulates that 16 billion gallons must be derived from renewable sources of cellulose, hemicellulose, or lignin.

Prior to any further discussion regarding energy derived from biomass, it is valuable to explicitly define potentially ambiguous terminology. Biomass, for the purpose of this paper, refers to all plant and plant-derived materials including animal manure (DOE/USDA, 2005). In saying that, the focus of this investigation isn't to cover the entire realm of biomass derived energy sources; instead, the focus is on cellulosic biofuels and more precisely lignocellulosic ethanol derived from dedicated energy crops. A cellulosic biofuel is defined by the U.S. government as a renewable fuel derived from any cellulose, hemicellulose, or lignin coming from renewable biomass (EISA of 2007). The overarching purpose behind the subsequent study is to develop a high yielding, dedicated energy crop in the Midwestern United States to be used as an alternative, renewable, carbon neutral energy source.

Considering the many hurdles still standing in the way of large scale cellulosic ethanol production, it should come as no surprise that the goals set forth by the U.S. government aren't going to be reached overnight. Fortunately, the government has provided substantial financial aid for research and development, producers of dedicated energy crops, and cellulosic ethanol producers through the Food, Conservation, and Energy Act of 2008 (FCEA). One imperative issue that must be addressed is the efficiency of the conversion technology utilized to transform lignocellulosic material into more readily usable ethanol; it must become more efficient if lignocellulosic ethanol is to become practical. Current technology yields only a fraction of the practical maximum yield of 120 gallons/dry ton of biomass (West et al. 2009). Another prerequisite to stimulate the cellulosic ethanol industry noted by West and colleagues (2009) is the need for protection against volatile oil prices for cellulosic ethanol producers. The risk associated with the large capital investment is too great when profitability hinges on high

oil prices. Particularly in the beginning stages of cellulosic ethanol production, government subsidies or tax credits will be essential to establishing this industry and maintaining its profitability.

Second generation biofuels derived from lignocellulosic crop residue will play an integral role in providing sufficient feedstock to biorefineries but without the development of dedicated energy crops, won't be enough to meet the ethanol production goals set forth in the EISA (West et al. 2009). Any crop, annual or perennial, that is grown expressly for the purpose of producing feedstock for a renewable biofuel that is not typically used for food, feed, or fiber qualifies as a dedicated energy crop according to the FCEA. Since this definition includes any crop that is not regularly grown, it encompasses an extensive number of species. It goes without saying that not every species has an equal probability of becoming a dedicated energy crop. An ideal dedicated energy crop should be high yielding under low input systems. In addition, it should be native or noninvasive, relatively inexpensive to establish and maintain, have high genetic diversity, and be beneficial to the environment.

The most promising feedstock candidates in the United States include herbaceous, perennial grasses such as *Miscanthus*, switchgrass, reed canarygrass and woody species such as hybrid poplar, willows, and black locust. Like all crops, particular species or genotypes of dedicated energy crops will be adapted to certain environments. The sheer size of the U.S. almost guarantees that no single energy crop will be universally adopted because many environments marked by different temperature extremes, precipitation levels, soils, pest pressures, etc. transect the land. In fact, microclimatic differences within a single parcel of land will likely result in distinct niches for multiple energy crops. This study will henceforth focus

specifically on the potential of *Miscanthus* as a feedstock in the Midwestern United States although many inferences will be made from research conducted throughout Europe.

Miscanthus

Miscanthus is a genus of perennial, C_4 grasses that have shown excellent potential to produce high yielding genotypes in temperate environments (Clifton-Brown et al. 2004; Heaton et al. 2008). Despite extensive inquisition, minimal agreement has been reached over the years regarding the number of species within the *Miscanthus* genus since its creation by Andersson in 1856 (Bentham 1882; Honda 1930; Keng 1959; Lee 1964; Liu 1994; Renvoise 2003; Hodkinson 2002a; Sun 2010). Honda (1930) claimed the genus includes 20 species, but more recent evaluations using both molecular markers and morphological data have suggested that six to eight species is more appropriate (Hodkinson et al. 2002a; Sun et al. 2010). Of these, *Miscanthus sacchariflorus*, *Miscanthus sinensis*, and their hybrid *Miscanthus x giganteus* have been identified as particularly strong candidates to become dedicated energy crops (Clifton-Brown et al. 2008).

One attribute that makes these grasses particularly attractive as feedstock crops is the C_4 photosynthetic pathway. This pathway has been proven to convert solar energy into biomass more efficiently than the C_3 pathway (Beadle and Long, 1985). Ribulose-1,5-biphosphate carboxylase/oxygenase (RubisCO), a key enzyme in the initial assimilation of carbon dioxide, has affinity for both carbon dioxide and oxygen. When RubisCO catalyzes the oxygenation of Ribulose-1,5-biphosphate (RubP), energy is consumed. Although it takes more energy for C_4 plants to reduce carbon dioxide, the loss is compensated by the elimination of the oxygenation of RubP. C_4 plants accomplish this by concentrating carbon dioxide around RubisCO in bundle

sheath cells of leaf tissue (Beadle and Long 1985). Concurrently, C_4 plants have greater water use efficiency due to decreased transpiration resulting from the ability to concentrate carbon dioxide around sites of concentrated RubisCO (Brown 1978).

Additionally, nitrogen use efficiency is significantly higher in C_4 species (Ehrlinger and Monoson 1993). Once again, this is attributed to the plants ability to concentrate carbon dioxide in areas where RubisCO is isolated. A significant amount of nitrogen must be devoted to constructing RubisCO as it is one of the most abundant plant proteins (Evans 1989). Since each molecule of RubisCO is more efficient in C_4 plants, they require less RubisCO. Ku et al. (1979) estimated that C_3 plants have 3-6 times more RubisCO in their leaves compared to C_4 plants. In the case of *Miscanthus,* nitrogen use efficiency is further increased due to its perennial nature. In temperate adapted *Miscanthus* genotypes, nitrogen is translocated to underground storage organs in the fall for growth the following year. Empirical evidence has confirmed the high nitrogen use efficiency of *Miscanthus x giganteus* (Lewandowski and Schmidt 2006; Clifton-Brown et al. 2007).

The previous few points have all dealt with the efficiency of *Miscanthus* so now is an excellent opportunity to revisit why efficiency matters. From a business perspective it is intuitive. In order to maximize profits, you want to maximize output per unit investment. This same thought process can be applied to biomass crops. In order to maximally reduce our dependency on foreign oil, we need to produce the most biomass while using the least amount of energy to do so. Consideration of this energy balance is another reason why perennials such as *Miscanthus* have an advantage over high-input annual crops (Heaton et al. 2004a). The majority of agronomic costs go into cultivation and chemical inputs such as nitrogen fertilizers.

Perennials have a huge advantage because they do not require annual cultivation and recycle nitrogen from year to year in belowground storage organs. Other beneficial traits of perennial species such as M*iscanthus* include carbon sequestration and erosion control. Clifton-Brown and colleagues (2007) estimated carbon sequestration for *M. x giganteus* to be 8.9 t C ha^{-1} over a 15 year period while carbon mitigation could be as high as 7.2 t C $ha^{-1}yr^{-1}$. At this juncture, the many upsides of using *Miscanthus* as a feedstock for ethanol production should be apparent.

As with any crop, genetic diversity is essential to sustained crop improvement. In combination with its obligate outcrossing nature, the large geographic distribution of *Miscanthus,* spanning more than 70° of latitude in Eastern Asia from Siberia to the Pacific islands, has led to high genetic diversity (Hodkinson et al. 2002a). Despite collection efforts by a few public and private organizations in the U.S., germplasm collections represent only a small fraction of the natural genetic variation within *Miscanthus* species. The main obstacle impeding germplasm transfer of *Miscanthus* outside of its natural range is import and export restrictions (Jakob et al. 2009). Although these delays may initially restrict *Miscanthus* improvement, they should not prove to be detrimental in the long term. While the proponents of *Miscanthus* view these restrictions as a nuisance, it is wise to remember why these rules are in place. History has shown that moving species out of their natural range can have major environmental and economic implications (National Research Council 2002). This issue may be particularly pressing for potential bioenergy crops as many of the traits that make crops ideal candidates overlap with traits common to invasive species (Raghu et al. 2006). This issue must be addressed if a nonindigenous species such as *Miscanthus* will be widely accepted by the public.

Miscanthus Genetics

A variety of ploidy levels exist within the *Miscanthus* genus. *M. sinensis* is considered to be diploid based on the high prevalence of bivalent chromosome pairings during meiosis, *M. giganteus* a triploid, and *M. sacchariflorus* either diploid or tetraploid (Burner 1991; Lafferty and Lelley 1994). The base chromosome number of the genus *Miscanthus* of x=19 is nearly twice that of most genera in *Saccharinae* where x=10. Recently, three groups independently established the fact that *Miscanthus* has undergone a whole genome duplication event accompanied by a chromosome fusion sometime between its divergence from closely related genera but before its own diversification (Swaminathan et al. 2012; Kim et al. 2012; Ma et al. 2012). The *Miscanthus* genome has been further enlarged by an abundance of highly repetitive DNA sequences (Swaminathan et al. 2010). The ancestral tetraploidy event along with increased number of repetitive sequences accounts for the monoploid DNA content of *Miscanthus* being three times larger than that of sorghum (Swaminathan et al. 2012).

Despite these differences, the genomes of sorghum and *Miscanthus* show a high level of synteny (Swaminathan et al. 2012; Kim et al. 2012; Ma et al. 2012). Swaminathan et al. (2012) were able to place nearly 90% of the SSR and SNP markers used to construct their *M. sinensis* genetic linkage map on the sorghum genome; they found only one small inversion in respect to sorghum chromosome four and one fusion event where a single copy of sorghum chromosome seven was inserted into the centromeric region of sorghum chromosome four. This conservation of synteny may prove useful for validating QTLs and identifying the underlying genes in *Miscanthus*.

Miscanthus x giganteus

A singular sterile triploid *M. x giganteus* clone has been the focus of the majority of the efforts to establish a *Miscanthus* biomass industry due to its high yields, low input requirements, and broad adaptability (Clifton-Brown et al. 2001; Heaton et al. 2004b; Sacks et al. 2012; Guader et al. 2012). *M. x giganteus* is a tall, late flowering genotype with an intermediate spreading nature. Although it had long been suspected of being a hybrid between a tetraploid *M. sacchariflorus* and a diploid *M. sinensis*, only in the last ten years has genetic data been available to prove this hypothesis (Hodkinson et al. 2002b; Swaminathan et al. 2010). Beale et al. (1996) discovered that unlike most C_4 species, *M. x giganteus* doesn't appear to suffer declines in photosynthetic rate at low temperatures early in the growing season. This unique ability leads to rapid canopy formation and an extended growing season.

Clonal field establishment via rhizomes takes between three and four years to mature but are estimated to remain productive for 15-20 years (Lewandowski et al. 2000). Confirmation of the previous claim cannot be drawn from four recent studies of *M. x giganteus* yields conducted independently across Europe over a period ranging from 11-16 years (Clifton-Brown et al. 2007; Christian et al. 2008; Angelini et al. 2009; Guader et al. 2012). Angelini (2009), Clifton-Brown (2007), and their respective coworkers found a decline in yield following the ninth and tenth years of production respectively, while the other two studies found no conclusive evidence of declining yields. The small scale of the plots studied by these researchers in comparison to production fields may limit the scope of their inference.

Miscanthus sacchariflorus

M. sacchariflorus is the northern most adapted species in the genus with a native range from Siberia to approximately 30^{o} N in East Asia. This adaptation to cooler environments has led to the prevalence of early flowering along with a strong dormancy response (Sacks et al. 2012). The potential of this species can be seen in a recently published article from Germany. A *M. sacchariflorus* genotype produced 10.7 t DM ha^{-1} year^{-1} on average over a 14 year period (Gauder et al. 2012). The strong rhizomatous nature along with its affinity for wetlands has made this species appear on several invasive species lists in the United States. The intraspecific improvement of *M. sacchariflorus* in the United States is severely limited by the number of commercially available cultivars and their sparse pollen production (Data not shown). For these reasons, improvement efforts should be directed towards the importation of new varieties and producing interspecific hybridizations from currently available varieties.

Miscanthus sinensis

M. sinensis has the broadest natural range of any species in the genus. It can be found from Borneo to northern China and parts of Russia (Sacks et al. 2012). Within this range it can be found in a number of different environments and elevations (Chiang 1993). This adaptation to diverse environmental conditions should provide ample variation for breeders to increase biomass production and to integrate stress tolerance into elite germplasm. *M. sinensis* is a tufted grass with heights ranging nearly three meters. The findings of Guader et al. (2012) indicate that *M. sinensis* can produce comparable yields to *M. sacchariflorus* but current varieties cannot match the long-term yields of *M. x giganteus*. These results should be applied cautiously as only eight genotypes were evaluated. Presumably, breeding efforts over the

period since these selections occurred has advanced yields in *M. sinensis* while *M. x giganteus* has remained stagnant.

M. sinensis was introduced to the U.S. from Japan in the late 1800s and has since been used extensively as an ornamental plant (Quinn et al. 2010). As such, traits counterproductive to biomass productivity have been favored, namely, early flowering and short plant stature. Still, this species has the most diversity within the genus in U.S. germplasm collections. Over 85 cultivars are sold commercially varying greatly in height, flowering time, and a variety of other traits of agronomic importance (Sacks et al. 2012).

Miscanthus Improvement

Understanding the inherent risks of planting large monocultures of a single or a few very similar genotypes, it has been suggested by many to produce alternative varieties to *M. x giganteus*. The sterile nature of the *M. x giganteus* clone is a double-edged sword. It does cut down on the risk of invasiveness, but it also eliminates any possibility of improvement through traditional breeding techniques. A number of plausible solutions to circumvent this problem exist. One such approach is to collect naturally occurring triploid varieties in areas where sympatric populations of *M. sinensis* and tetraploid *M. sacchariflorus* exist (Stewart et al. 2009). Another recourse is to restore fertility through polyploidization. Protocols have been developed to generate polypoid plants in *Miscanthus* species (Peterson et al. 2002; Yu et al. 2009). Yu et al. (2009) recovered a number of hexaploid *M. x giganteus* lines, yet fertility hasn't been established beyond pollen viability (W.B. Chae, personal communication). Ongoing research will determine the viability of introducing new genetic variability via fertility recovery.

Another strategy presently being exploited is intra- and inter-specific improvement of the two *M. x giganteus* progenitor species. Due to the wide latitudinal and altitudinal distribution of these species throughout the Far East, it is reasonable to conclude that breeders will be able to adapt cultivars to a wide range of environments. Traditional breeding methods should be able to take advantage of heterosis and transgressive segregation. Additionally, synthetic cultivars could be bred as an alternative to clonal propagation in these fertile species. Establishing *Miscanthus* by seed would greatly reduce the estimated establishment cost of 128 dollars $acre^{-1}$ associated with vegetative propagation, one of the biggest limitations of *Miscanthus* (Lewandowski et al. 2000). Initial efforts to create doubled haploid lines via anther culture have shown some promise (K. Glowacka, personal communication); if a successful doubled haploid procedure could be established, *Miscanthus* breeding would be revolutionized. Regardless of the means, these improved lines could be used in their own right or be used to recreate the *M. x giganteus* cross.

When considering the narrow *Miscanthus* germplasm available to U.S. breeders, few viable improvement options exist. To reiterate, ploidy manipulation, intraspecific improvement of *M. sinensis*, interspecific hybridizations, and re-creation of the *M. x giganteus* cross should all be exploited to create alternative cultivars to *M. x giganteus*. Because interspecific improvement of *M. sinensis* can be beneficial to all aforementioned strategies, it will be the main focus of the subsequent discourse.

Currently, the major limitation to *M. sinensis* breeding is the need to grow large segregating populations to maturity before making selections. The findings of Clifton-Brown et al. (2001) indicate that the true potential of a clone cannot be accurately predicted based off

the yield of the first two years following planting. This requirement limits the number of individuals and populations that can be evaluated. Some studies have been conducted to correlate measurable traits with yield in the same year (Jezowski 2008; Gauder et al. 2012), but less work has been done in the public sector to find appropriate characteristics in establishment years to predict third year yields. Clifton-Brown and Lewandowski (2002) showed that of the traits measured the number of shoots and the combination of shoot x height in the first year where the best predictors of third year yield (r^2=0.61 and 0.64 respectively). It is prudent to keep in mind that these correlations are based off of 20 to 25 m^2 plots using preselected genotypes.

Marker Assisted Selection

Any strategy that could be used to preselect promising individuals and discard inferior plants would save substantial resources and greatly accelerate the improvement of *M. sinensis.* One option suitable for this purpose would be marker-assisted selection (MAS). MAS exploits linkage disequilibrium (LD) between a quantitative trait locus (QTL) of interest and markers (Hospital 2009). As long as the LD between the QTL and markers is high, selection can be made based on the markers as opposed to the phenotype. MAS is purported to be most effective when phenotyping is difficult, costly, or cannot be evaluated at an early stage of development as in the case of *Miscanthus*. The most routinely and successfully employed form of MAS in both private and public breeding programs is marker assisted backcross breeding (MAB) (Bouchez et al. 2002; Helguera et al. 2003; Kuchel et al. 2007). Other successful examples of MAS include gene pyramiding schemes and population screening (Ashikare et al. 2005; Zhang et al. 2006; Barloy et al. 2007; Nocente et al. 2007). Success can be defined differently depending

on the goals of the individuals performing the MAS. Speaking from a breeder's point of view, MAS selection is only practical when it provides a greater return per dollar invested as compared to the current breeding scheme. This has yet to be evaluated in a *Miscanthus* breeding program.

On the other hand, there are instances where MAS failed to produce the expected results. Shen et al. (2001) conducted a MAB program in rice to create near isogenic lines containing one of four previously identified quantitative trait loci (QTLs) controlling root characteristics. The authors discovered that one of the four QTLs had vanished upon transfer to the recurrent parent. This can happen if the original QTL was simply a false positive or if it was a true positive but depended on a certain effect absent in the recurrent parent or environment. Generally, these effects are caused by a QTL x QTL, QTL x genetic background, or a QTL x environment interaction. Another factor that greatly influences the effectiveness of MAS is the genetic architecture of traits being selected. MAS tends to work best with simple traits controlled by one or a few large-effect QTLs with limited interactions. Many disease resistant traits are a good example of such. Complex traits controlled by many small-effect QTLs such as yield tend to be less suited for MAS.

The question to ask then is how appropriate of a strategy is MAS to the improvement *Miscanthus*? This depends on many factors. It has been well established that phenotyping *Miscanthus* is difficult, costly, and cannot be done at an early stage of development. The major unknown is the underlying genetic architecture of traits important to biomass production in *Miscanthus*. Additionally, how closely can we associate genetic markers with the traits of interest in a breeding population? Little has been done thus far to address these questions.

The overall objective of this study is to perform a QTL analysis on traits important to above ground biomass yield in a pseudo testcross mapping population of *Miscanthus sinensis*. This will help elucidate the genetic architecture and the mode of inheritance underlying selectable traits influencing yield. QTL analysis will be performed on the following traits: heading date, 50% anthesis, plant height, tiller diameter, number of tillers, compressed circumference, basal circumference, compressed circumference to basal circumference ratio, average leaf width, maximum leaf width, spring emergence date, percent moisture, growth rate, leaf length, and above ground biomass yield. Descriptions of these measurements can be found in Table 2. These traits are of interest because they are either important to the potential biomass yield or more broadly to developing a successful dedicated energy crop.

Literature Review

Quantitative Trait Locus Analysis

Quantitative trait locus (QTL) analysis is a statistical method that links phenotypic data and genotypic data in an attempt to explain the genetic basis of variation in complex traits (Falconer and Mackay 1996). QTL analyses allow researchers to link complex phenotypes to specific regions of chromosomes and understand the action, interaction, number, and precise location of these regions (Miles and Wayne 2008). Like marker assisted selection, QTL analysis takes advantage of linkage disequilibrium between a genetic marker and a QTL influencing the trait under investigation. Markers that are genetically linked to a QTL will segregate more frequently with high or low trait values than unlinked markers.

The concept of using simply inherited genotypic markers linked to factors effecting complex traits to isolate and study the size of each individual effect contributing to the observed phenotype was first proposed by Karl Sax in 1923. Using qualitative seed pigmentation characteristics following Mendelian inheritance, Sax (1923) performed the first QTL analysis on seed weight in the common bean. A quick search of scientific literature will reveal the extent of the popularity of QTL studies. Historically, the availability of genetic markers was the limiting factor in QTL analyses (Miles and Wayne 2008). With the advent of new DNA molecular marker platforms and advancements in next generation sequencing technology, a seemingly endless supply of molecular markers are becoming available at increasingly reduced costs (Pop and Salzberg 2008). Simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), sequence-based genetic markers, have made phenotyping the limiting factor in QTL analyses (Appleby et al. 2009; Miles and Wayne 2008).

In addition to the advancements in marker technology, statistical tools have become increasingly sophisticated over the years. Single marker analysis using simple t-tests, ANOVA, or regression to detect significant associations between genetic markers and phenotypic classes was the dominant method of analysis up until the late 1980s. One of the largest disadvantages of single marker tests is the inability to separate the effect of a QTL from its proximity to the marker being tested. Lander and Botstein (1989) proposed the idea of interval mapping (IM) to estimate the location of a QTL on a genetic map. IM uses a likelihood function or linear regression to scan chromosomes at set increments in between ordered pairs of markers to test whether a QTL is likely to be present at the location within the interval or not. This method assumes that a maximum of one QTL is located on each chromosome although in reality this assumption is often violated. When this assumption is not fulfilled 'ghost QTL' can appear and the predicted location of a QTL can be skewed (Martinez et al. 1992).

Shortly after IM was proposed, Zeng (1993) and Jansen (1993) proposed the idea of composite interval mapping (CIM) still popular today. Theoretically, this technique considers every position in the genome simultaneously while searching for a potential QTL. This technique is advantageous in that it can detect quantitative trait loci (QTLs) acting independently, linked, or interacting epistatically with other loci (Doerge 2002). Unfortunately, this method is too computationally intense so shortcuts have been developed to make this method feasible. A typical shortcut is accomplished by first scanning the genome for the most significant markers, including these in the model as a cofactors, and then rescanning the genome using the new model. This process is repeated until no significant QTLs are detected

and the estimated QTL positions are fixed. After the model is complete, interactions between either the detected QTLs or QTLs and every marker can be investigated.

Regardless of the statistical method used to search for QTLs, the issue of declaring a significance threshold must be addressed. The best way to do this is through non-parametric randomization permutations (Churchill and Doerge 1994). This is accomplished by randomizing the phenotypic data with respect to the genotypic data. After randomization, the full model is used to test for significant marker associations and the lowest p-value for each run is recorded. This process is repeated N times. In order to determine the experimentwise significance threshold, the lowest p-values are ordered sequentially and their $100(1 - \alpha)$ percentile is the appropriate critical value (Churchill and Doerge 1994).

Once QTLs are located, the next objective it to define a confidence interval (CI) around its location. Such a CI can be a useful guide in further experiments whether it is to fine map the region, isolate the underlying gene, develop a marker assisted breeding program, or move the QTL into another background for the purpose of crop improvement. Unfortunately, the various methods of QTL detection don't lend themselves to a straightforward calculation of CI.

Lander and Botstein (1989) recommended the use of 1- or 2-LOD support intervals to form approximate CIs while Dupuis and Siegmund (1999) found that 1.5-LOD support intervals provided 95% coverage when high marker density was used to detect QTLs. LOD support intervals are very simple to calculate. If one wishes to use a 2-LOD support interval to define a CI around a QTL, then they need only find the location on either side of the maximum LOD score where the LOD score has dropped by two. These two points on the linkage group represent the ends of the CI. This method has been found to be both adequate by computer

simulations (Van Ooijen 1992; Dupuis and Siegmund 1999; Manichaikul et al. 2006) and inadequate by the same means (Mangin et al. 1994; Visscher et al. 1996).

Visscher and his colleagues (1996) proposed using nonparametric bootstrapping to develop appropriate CIs. This is accomplished by making new data sets by randomly drawing with replacement N number of individuals from a population of N individuals. QTL analysis is then conducted on the new data set and the location of the maximum LOD score recorded. This process is repeated at least 200 times. In order to construct the CI, first order the locations of the peak LOD scores along the pertinent chromosome. The positions of the peak LOD scores marking the 2.5 and 97.5 percentiles specify the two end points of the 95% CI. This method has been accepted by some as an improvement to the LOD support method yet has also been shown to be unreliable (Bennewitz et al. 2002; Manichaikul et al. 2006). The search to find the most accurate way to produce CIs continues today as increasingly complex strategies are developed (Sen and Churchhill 2001; Bennewitz et al. 2002; Manichaikul et al. 2006; Hengde 2011). Many of these methods are developed for traditional population structures and are not easily adapted to cross pollinated populations.

The final subject matter pertinent to QTL detection is power, i.e., the percentage of real QTLs detected. Beavis (1994; 1998) showed through simulation studies that the power to detect QTLs increases with increasing number of individuals and that the amount of phenotypic variation explained was grossly overestimated with small populations (100 individuals) and slightly overestimated in large populations (1000 individuals). In addition to population size, the size and variance of the QTL effect, stringency of the significance threshold, and population type all affect power. QTL with larger effects are naturally easier to detect because the

genotypic class trait means will be further apart. The variation of a QTL effect is most easily explained visually (Figure 1). Population type affects power by changing the number of potential QTL genotypes. Backcross type populations are more powerful than F2 populations because only two QTL genotypes have to be estimated and compared as opposed to three. A trade off exists between the power to detect QTL and the number of false positives detected. This is controlled by the stringency level. A decrease in the desired alpha, i.e., the lower tolerance for false QTL discovery, results in less power to detect QTLs.

Previous QTL studies in *Miscanthus*

To date, four published QTL studies have been conducted on *Miscanthus*–all stemming from one segregating population (Atienza et al. 2003abcd). All four were published in 2003 by S.G. Atienza, a graduate student at the Institute for Sustainable Agriculture in Cordova, Spain. The mapping population consisted of 89 individuals from a full-sib cross originating from an intraspecific *M. sinensis* hybridization. Data was collected for stem diameter, height, yield, and numerous combustion quality traits over a two year period which included the establishment year. The QTL analyses were performed using the cross pollinated population type within MapQTL 4.0 (Van Ooijen et al. 2000). The genetic map consisted of 257 random amplified polymorphic DNA (RAPD) markers covering 28 linkage groups (Atienza et al. 2002); the expected number of linkage groups for *M. sinensis* is 19.

While this group did report a number of QTLs explaining much of the phenotypic variation over the traits investigated (Table 1), there were many weaknesses in these experiments. First, the genetic map they utilized was incomplete and consisted of a relatively

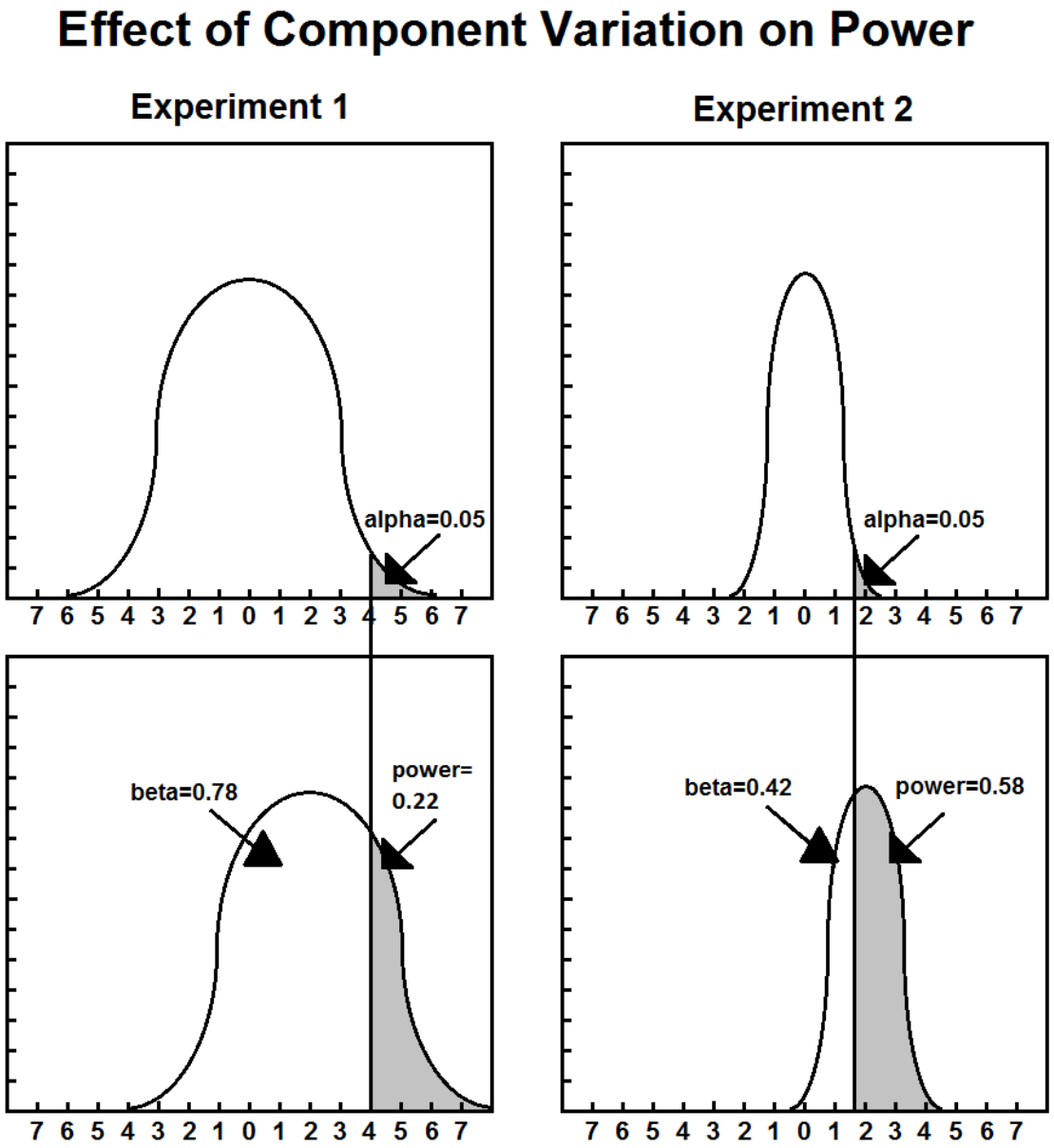


Figure 1. Component variation increases both due to QTL acting elsewhere in the genome and variation caused by non-genetic factors. The top graphs in each experiment represent the phenotypic distributions of one genotypic class or allele while the bottom graphs represent the phenotypic distributions of individuals having the alternative allele. Both experiments have a QTL with identical effects yet the power to detect the QTL in experiment two is much larger compared to experiment one.

Table 1. Summary of Atienza et al. (2003abcd) QTL analyses showing the percent of phenotypic variation explained for the QTL deemed significant.

Trait	QTL	PVE Year 1	PVE Year 2
Height	H1	24.4	NA
	H2	10.4	18
	H3	12.4	NA
	H4	NA	24.1
	H5	NA	34.9
Stem Diameter	D1	29	14
	D2	24.5	NA
	D3	14.5	NA
Yield	Y1	19.5	26.9
	Y2	21.3	NA
	Y3	20.1	29.6
	Y4	42.7	NA
	Y5	2.5	NA
	Y6	NA	14.2
Calcium	C1	17.5	13.4
	C2	NA	44.6
Phosphorus	P1	44.5	NA
	P2	39.1	NA
	P3	NA	36.6
	P4	NA	38.6
	P5	NA	34.8
Chlorine	C1	35	NA
	C2	NA	28.6
	C3	18.5	NA
	C4	NA	31.7
	C5	NA	30.2
Potassium	K1	11.8	NA
	K2	34.6	NA
Sulfur	S1	7	NA
	S2	NA	38.2

NA-QTL not detected

small number of RAPD markers. The inconsistent nature of RAPD markers makes the comparison between studies problematic. Second, the population size was extremely small. This makes the analyses prone to the Beavis effect and severely limits the power to detect QTLs. Lastly, half of the phenotypic data was collected in the establishment period of the population. The lack of consistency between years may be due to QTL x year interactions or to differences in the genetic control of traits between the establishment year and subsequent years. As mentioned, the large percentage of phenotypic variation explained by each QTL is likely an artifact of small population size. Due to these weaknesses and the nature of the genetic map employed in previous *Miscanthus* QTL studies, limited comparisons between previous work and the ensuing results can be made.

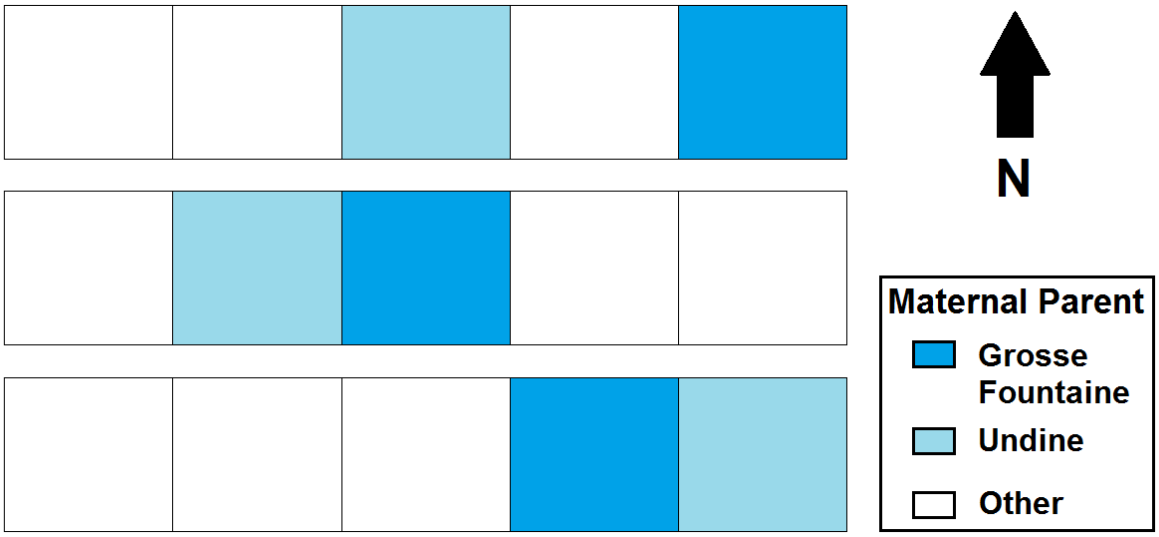
Materials and Methods

Population Development

The mapping population consists of 221 F1 individuals derived from an intraspecific cross between two ornamental varieties of *M. sinensis*. 'Grosse Fountaine (GF)' and 'Undine (UN)' were reciprocally cross-pollinated in isolation during the 2009 fall semester in the Plant Science Laboratory at the University of Illinois Urbana-Champaign. 500 seeds collected from each parent were planted in the aforementioned greenhouse in the winter of '09-'10 in 96 well flats using LC1 Sunshine Mix. Greenhouse temperatures were kept between 22.2-29.4^{o}C, and supplemental light was provided at a threshold of 600 W/m^{2} between 6 am and 8 pm. When the root balls were adequately strong to remain intact during transplantation, seedlings were transferred to four inch diameter pots to continue growth. After the seedlings were large enough to clonally propagate, they were divided into four equal parts and planted into 36 well flats with Metromix 510 soil. Three weeks following propagation, flats were moved to outdoor greenhouse bays for a one week hardening off period prior to field planting.

In May of 2010, 200 progeny from each parent were randomly selected from a pool of progenies with at least three surviving clones and were planted at the Energy Biosciences Institute farm just south of Urbana, Illinois. The remaining clone, when applicable, was maintained in the greenhouse as a backup. Plugs were planted on five foot centers in blocks of 13 x 16 individuals with a border row on all sides. Blocks were randomized and complete within each direction of the cross (Figure 2). In other words, three blocks containing all 200 progeny originating from seed harvested from GF were organized within each block in a random fashion. The other three blocks were designed similarly with the 200 individuals harvested from UN.

a



b

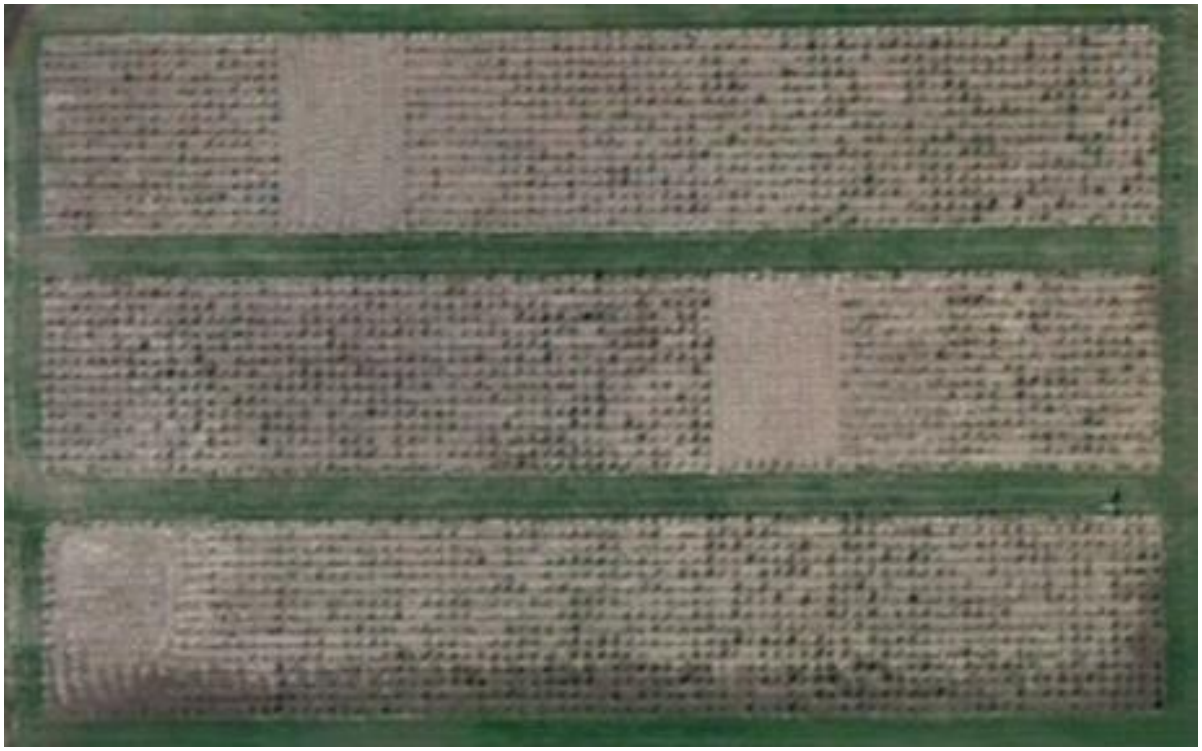


Figure 2. a) The field design showing all six blocks. The progeny were separated into two groups based on which cultivar the seed was collected from and planted in two sets of three randomized complete blocks resulting in each individual being replicated three times. b) A Google Earth image of the plot from October 4, 2010.

Each block also contained eight controls: four GF and four UN plants. In summation, 400 progeny were planted in the spring of 2010 each with three replications. The number of individuals was reduced to 221 based on progeny exhibiting alleles absent in both parents. Presumably this resulted from pollen contamination. Further details on the DNA extraction can be found in Swaminathan et al. (2012).

Genetic Map

The genetic map was constructed from the same population described above using the multipoint maximum likelihood model and the Haldane mapping function in JoinMap 4.1 (Van Ooijen 2011). This approach takes advantage of markers that are heterozygous in both parents to integrate crossover events into one map as opposed to two as the traditional pseudo-testcross strategy yields. Because of the nature of the population, the map is based on meiosis that occurred in the grandparents of the mapping population. As such, each locus could potentially have four alleles segregating.

The resulting map contains the expected number of linkage groups (19) spanning a map length of ~1890 cM with an average intermarker spacing of 2.7 cM (Figure 3). The map consists of 653 SNPs and 193 SSRs. Deep transcriptome sequencing from the two parental accessions was used to develop a GoldenGate^{TM} genotyping array of 1536 single nucleotide variants. Sugarcane SSR primers derived from expressed sequence tags and intergenic sequences designed and characterized by James et al. (2011) were used to screen genomic DNA extracted from GF and UN for polymorphisms. The segregation ratio of markers in the mapping population was used for validation.

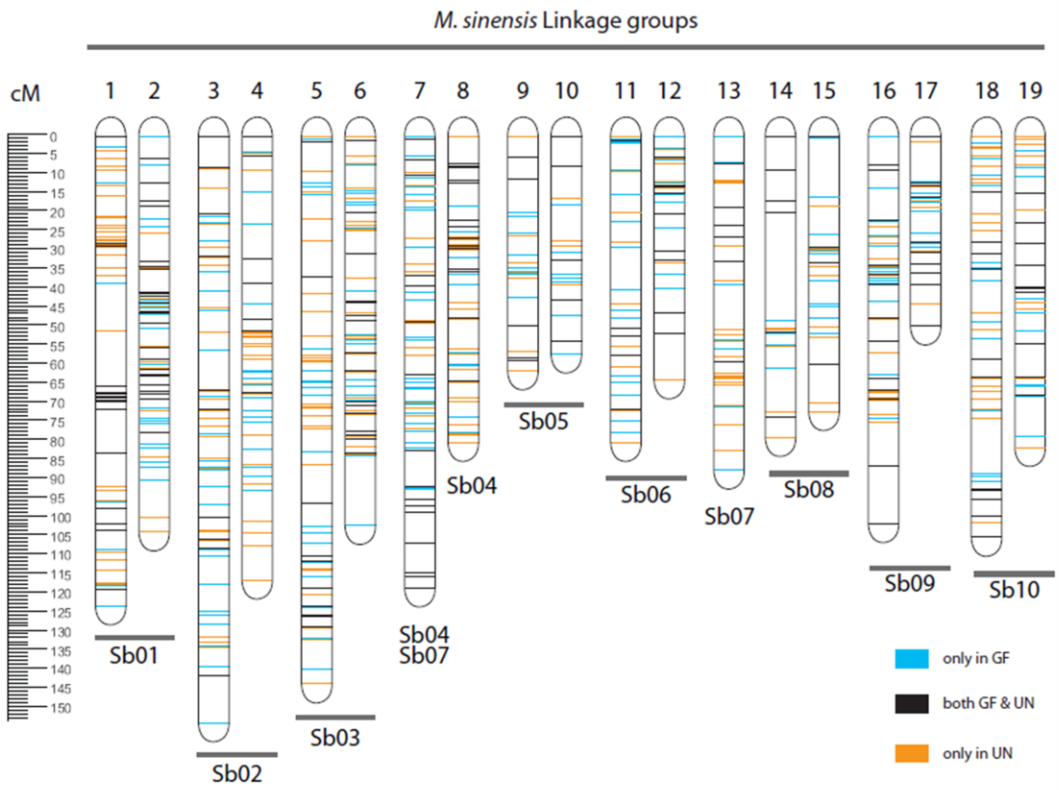


Figure 3. The genetic linkage map used to analyze QTL indicating the cultivar from which the marker is derived as well as the homologous sorghum chromosome (Swaminathan et al. 2012).

Phenotypic Data Collection

Fewer measurements were taken in the establishment year, due to the fact that much of the observed variation was caused by differences in plant health at the time of transplanting. Monthly counts of the number of tillers (any tiller visible above the soil surface) and plant vigor (rated 1 to 5 with one being weak and 5 being vigorous growth) were taken to assess plant establishment. In addition, flowering data was recorded on a weekly basis after the first signs of reproduction were observed.Two dates were recorded per plant: the first date at which the plant was observed to have at least one tiller heading and the date at which approximately half of the tillers were or had undergone anthesis. Beyond assessing plant establishment, these measurements were taken to see if any correlations could be established from year one to year three yield data and to determine if any loci important to establishment could be identified. Table 2 contains the traits measured in the second year along with a brief description of each.

Data Analysis

Genotype means needed for the QTL analyses were determined using the MIXED procedure of SAS (SAS Institute Inc. 2012). Means were analyzed separately for each direction of the cross each as a randomized complete block design with blocks being considered random. The data was analyzed in this manner due to the field design, although admittedly the averages are remarkably similar to the simple arithmetic means. In addition to finding the genotypic means, correlations between traits were determined using SAS software (SAS Institute Inc. 2012).

Although second year growth is less affected than the first year data by the condition of the plant plug at the time of planting and the plants ability to overcome the transplant shock,

Table 2. A description of the phenotypic data collected on the mapping population in 2011.

Spring Emergence	Measurements were initiated in the spring at the first sign of regrowth and taken twice weekly thereafter until all plants were either considered emerged or dead. A plant was recorded as emerged once photosynthetic tissue reached a height of 25 cm.
Plant vigor	Qualitative rating was taken monthly at the beginning of June, July, and August. The largest plants received a rank of five while the smallest a score of 1.
Plant height	Plant height was taken both at the time the vigor ratings were recorded and at harvest with the QTL analysis of plant height conducted on the later. Plant height was taken from the base of the plant to the tallest point of the plant.
Flowering time	Data collection was initiated at the first sign of flowering. Observations were recorded twice weekly. The date was recorded both when the first tiller of a plant began heading and when the plant reached 50% anthesis.
Leaf traits	Three leaves per replicate were measured using a LI-3000C Portable Leaf Area Meter (LI-COR, Inc.). Measurements were taken on the fourth leaf from the top of the plant in mid-September. The device simultaneously records leaf length, average leaf width, and maximum leaf width.
Tiller diameter	Prior to harvest, three tillers per replicate were measured at half the total tiller height using a Mitutoyo Series 500 Absolute Coolant Proof digital caliper accurate to 0.02 mm.
Compressed circumference	Prior to harvest, the compressed circumference was measured to the nearest cm at half the plant's height. This was done by measuring the circumference of the plant after compressing the tillers together with a zip tie.
Tiller number	Tiller number was estimated from the following equation which is simply the area of the compressed circumference divided by the area of a cross section of an average tiller $T = {C^2}/{4\pi^2 r_s}$ where T=tiller number, C=compressed circumference, and r_s= average stem radius
Aboveground biomass yield	In late November, after senescence replicates were zip-tied and cut with a sickle bar mower at approximately 10 cm above the ground. Plants were weighed on site using a temperature adjusting hanging scale accurate to 0.02 kg.
Basal circumference	After harvest, plant circumference was measured to the nearest cm at ground level.
Percent Moisture	A subsample approximately 0.2 kg in weight was taken at harvest from each replicate. This was weighed, dried, and reweighed. Percent moisture was calculated by subtracting the weight of the dried subsample from the weight of the subsample taken from each plant at the time of harvest (wet weight) then dividing by the wet weight.
Growth Rate	An estimate of growth rate was taken by subtracting the height measurements taken in June from those taken in July (growth in m/month).

considerable variation was still observed due to this establishment phenomenon. In hopes of accounting for some of this variation, an analysis of covariance was ran using spring emergence as a covariate. One of the key assumptions inherent in an analysis of covariance is that the covariant not be affected by the treatment. In this case, the treatment is the genotype. Therefore, in order for this analytical method to be valid, the timing of spring regrowth following the first year of growth must be due to the establishment effect as opposed to genetic predisposition towards early or late spring regrowth. The biological explanation behind this covariate analysis is that plants failing to establish well will have impoverished nutrient reserves resulting from a decreased ability to produce and store nutrients in the first year. Thus, poorly established plants would meet the regrowth criteria later in the spring. The SAS code follows:

```
Proc mixed data=means method=type3;
class block genotype;
model trait=genotype / ddfm=kr;
random block;
lsmeans genotype;
run;

Proc corr data=correlation;
var traits
run;

Proc mixed data=covariate method=type3;
class genotype;
model trait = genotype covariate / solution ddfm=kr;
lsmeans genotype;
run;
```

The QTL analysis was performed with MapQTL 5.0 using the cross pollinated (CP) population type (Van Ooijen 2004). This population type allows the fitting of four different QTL genotypes (Table 3). This is imperative because there are potentially four different alleles

Table 3. The model for the genotypic QTL affects in a cross pollinating species as presented by Van Ooijen (2004).

ab x cd	**c**	**d**
a	ac $\mu - \alpha - \gamma - \tau$	ad $\mu - \alpha + \gamma - \tau$
b	bc $\mu + \alpha - \gamma + \tau$	bd $\mu + \alpha + \gamma - \tau$

μ-overall mean
α-difference between a and b allele
γ-difference between c and d allele
τ-intralocus interaction

segregating at each loci, one allele from each grandparent. Linkage phases determined in JoinMap 4.1 are used to identify each allele's most likely origin. MapQTL 5.0 employs the maximum likelihood approach to the segregation of a mixture of probability distributions (Titterington et al. 1985; McLachlan and Basford 1988). Under the CP population type four distributions are fitted, one for each QTL genotype. The expectation-maximization (EM) algorithm is used to predict the most likely genotype of each individual and estimate the mean and variance of each component distribution at each locus (Dempster et al. 1977). Because many of the markers in the CP population type are not fully informative, surrounding markers are utilized to determine the probabilities of each QTL genotype. The algorithm stops when the likelihood of successive iterations increases less than a designated amount coined the functional tolerance. The calculated maximum likelihood is then compared to the likelihood under the null hypothesis that no QTL is segregating at that locus (Figure 4). The maximum number of neighboring markers used to calculate the probability of each QTL genotype at a locus, the maximum number of iterations used in the EM algorithm, and the functional tolerance were all kept at their default values of 5, 200, and $1x10^{-8}$, respectively.

The significant likelihood of odds (LOD) threshold for each trait was set by permutation testing as described by Churchill and Doerge (1994). The test was implemented in the MapQTL software using 1,000 permutations. The significance value was set to 0.05 on a genome wide basis. This means that the chance of detecting a false positive by random chance somewhere along the 19 chromosomes is approximately 5%. This is equivalent to an experimentwise error rate. The alternative is to use a separate significance threshold for each chromosome. While this would lead to the declaration of more QTLs, the number of false positives would

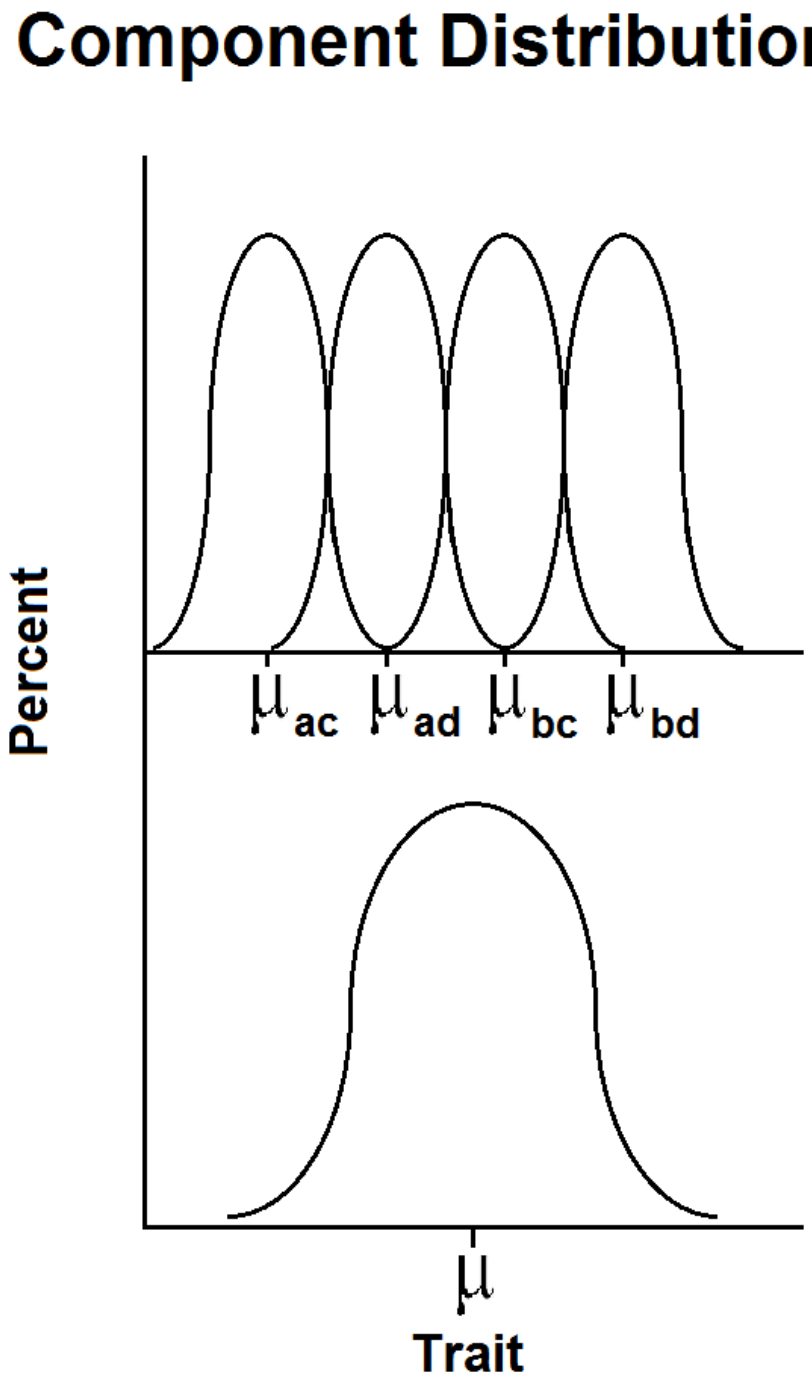


Figure 4. The expectation-maximization algorithm calculates the likelihood that a QTL is segregating at a particular locus by dividing the population into groups based on the genotype of each individual at a particular locus and estimating the trait mean and variance of each group. The likelihood that each group has its own distribution represented by the top portion of the figure is compared to the null hypothesis that no QTL is segregating, i.e., the bottom portion where trait values all come from a single distribution.

dramatically increase. Since only one year of data is available for most of the traits, a more conservative approach to declaring QTLs was taken.

This approach was initially taken to conduct interval mapping (IM) with a map step size of 1 cM. The map step size indicates the positions along the genome which are searched for QTL. Any chromosome having a LOD score above the permuted threshold was considered to have a QTL segregating. The marker nearest to the highest LOD score was recorded as the most likely position of the QTL. As mentioned previously IM operates under the assumption that only a single QTL is segregating. This assumption is often violated when analyzing complex quantitative traits as they tend to be controlled by many genes; the result is a decrease in power to detect QTL.

A more powerful search for QTL can be accomplished using composite interval mapping which doesn't make any assumption on the number of QTL segregating. As mentioned previously, it would be most appropriate to consider every locus simultaneously when searching for QTL, but this approach is computationally prohibitive. The shortcut employed by the MapQTL software is to use an approximate multiple-QTL model (MQM). This is an extension of the previously described IM approach with the addition of previously identified QTLs to the model as cofactors. The cofactors chosen are those markers found to be most closely linked to the QTL in IM. By removing variation accounted for by other QTLs, the power of MQM mapping is greatly enhanced. The overall means of the component densities as shown in Table 3 have additional terms representing the additive effects of each cofactor. For this reason, each individual will have its own set of component densities.

The number of calculations needed for MQM mapping can quickly bog down the MapQTL software. The program has to determine the probability of each cofactor's genotype and the probability of the QTL genotype then fit the different component distributions at every locus for every. For example, assume three cofactors are being used in MQM mapping with incomplete genotypes and the locus being tested for a QTL is in between markers. This means that the program will have to estimate 256 component distributions and therefore likelihoods every iteration for every individual at every locus. Additionally, MapQTL 5.0 (in fact all versions of MapQTL) is a 32-bit MS-Windows software which runs as a single thread meaning that the program can utilize a maximum of 4 GB of RAM and only one processing core. These limitations along with the complexity of running MQM analyses in the CP population type made it impossible to run the QTL analyses with the full set of markers. Fortunately, by removing some less informative markers, the program was able to run the analysis in a practical amount of time. Even though the marker density was reduced, it likely came at little penalty due to the ability of IM and CIM to utilize linked markers to accurately predict genotypes between markers. Besides using fewer markers and the addition of cofactors in the MQM analysis, all other analysis settings remained the same as in IM.

Confidence intervals were developed using the 2-LOD dropoff method. This strategy was adopted due to both its simplicity and the lack of congruence in the literature as well as the time prohibitive nature of performing bootstrapping using MapQTL software.

Results and Discussion

Environment

Environmental conditions at the Energy Biosciences Institute farm outside of Urbana, Illinois were slightly different between the growing seasons of 2010 and 2011. The most noticeable difference between the two years was the prolonged dry period in 2011 that spanned from mid June to August. Although the difference between total rainfall over the two years between March and October was only 10.2 cm, the majority of the rainfall in 2011 occurred at the very beginning of the growing season while in 2010 the rainfall was spread more evenly throughout the growing season (Figure 5). The temperatures between the two growing seasons do not show a stark contrast (Figure 5).

Environmental conditions have long been known to play a major role in gene regulation and likewise in QTL studies. In plant breeding it is important to identify cultivars that do well over the entirety of the target environment. For this reason, many studies have investigated the consistency of QTLs over a number of environments. While this investigation was not designed specifically for this purpose, year to year weather variation naturally creates unique growing environments resulting in some expected inconsistency in QTLs between years.

As a perennial crop, *Miscanthus* goes through different developmental phases. Different loci may be controlling the same trait in different phases. Since results from this preliminary investigation can only be compared between the establishment year and one year of established growth, it is impossible at this point to distinguish whether the inconsistency between years is due to the developmental phase, the environment, or a combination of the

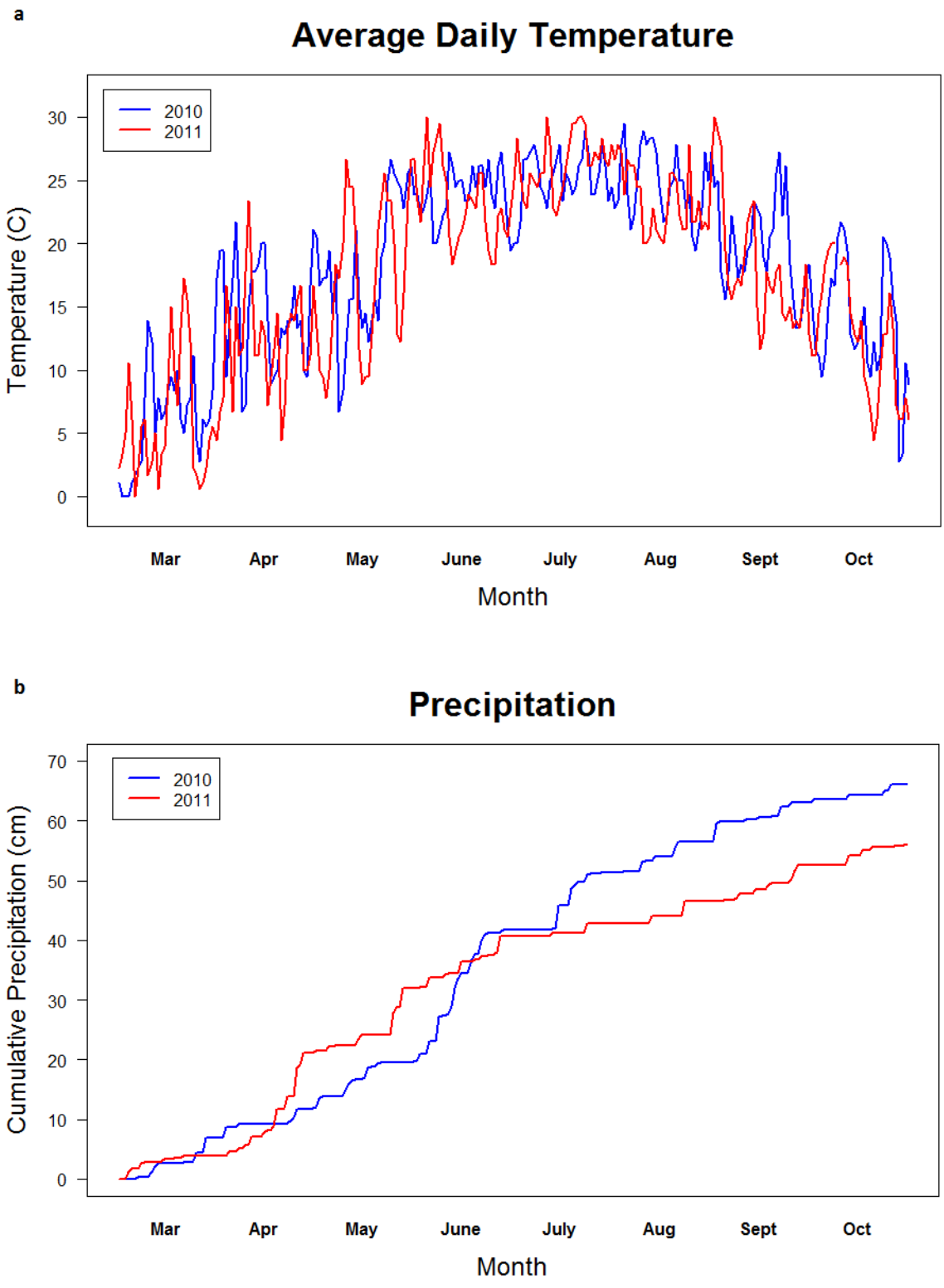


Figure 5. Environmental conditions. a) average daily temperatures over the growing seasons of 2010 and 2011 b) cumulative precipitation in 2010 and 2011.

two. Moreover, sporadic irrigation was practiced in 2010 to avoid desiccation of the newly transplanted, vulnerable plants. A more informative comparison of environmental data will be between 2011 and 2012.

Phenotypic Values

Phenotypic values for traits measured in 2010 and 2011 are given in Table 4. Due to the limited number of traits observed in 2010, it is difficult to compare between the two years. Plants headed nearly a month earlier and reached 50% anthesis nearly two weeks earlier in 2011. While the number of tillers did drastically increase from 2010 to 2011, the difference is magnified by the change in measurement technique.

Trait Correlations

A major impediment to *Miscanthus* improvement is the presence of an extended establishment period preceding the time when a genotype's true potential can be reliably evaluated. This results in a lengthened generation time leading to a reduction in breeding gain per unit time. Establishing strong correlations between traits in the first or second year of growth with that of mature yield or traits correlated with mature yield may allow selection to occur prior to maturity. Due to the complex nature of yield, it may be more realistic to cull the losers from the population as opposed to picking the winners. By removing the need to phenotype a percentage of individuals in a large breeding population, breeders can screen many more individuals in the search for superior genotypes. Additionally, understanding the correlation between traits may help optimize resource allotment. If a trait measured in the establishment phase has no correlation to mature yield, the question of whether measuring

Table 4. Phenotypic values for traits measured in 2010 (above the dashed line) and 2011 (below the dashed line). Phenotypic means are given for each parent in the last two columns.

Trait (unit of measure)	Average	Maximum	Minimum	Standard Deviation	'Grosse Fountaine'	'Undine'
Number of Tillers	13.40	31.00	3.33	5.63	7.61	4.50
Heading Date	Sept. 20th	Oct. 4th	Sept. 7th	5.6 days	Sept. 25th	Sept. 19th
50% Anthesis	Sept. 22nd	Oct. 8th	Sept. 10th	6.7 days	Sept. 30th	Sept. 25th
Height (m)	1.49	2.00	0.99	0.16	1.94	1.46
Tiller Diameter (cm)	6.24	8.85	4.45	0.75	7.33	5.70
Wet Yield (kg)	1.54	2.94	0.08	0.49	2.17	0.93
Dry Yield (kg)	1.17	2.11	0.07	0.36	1.69	0.75
Compressed Circumference (cm)	35.75	53.00	7.33	7.62	38.96	28.52
Number of Tillers	375.23	1081.81	20.48	156.96	293.71	291.65
Basal Circumference (cm)	80.22	116.00	33.00	13.20	86.75	64.21
Percent Moisture	0.20	0.33	0.09	0.04	0.18	0.15
Heading Date	Aug. 24th	Oct. 7th	Jul. 22nd	16.03 days	Aug. 20th	Aug. 4th
50% Anthesis	Sept. 8th	Oct. 14th	Jul. 31st	14.88 days	Aug. 31st	Aug. 23rd
Spring Emerge	Apr. 5th	Apr. 15th	Apr. 1st	2.1 days	Apr. 5th	Apr. 6th
Leaf Length (cm)	61.48	82.01	41.98	6.55	74.15	55.79
Average Leaf Width (cm)	0.77	1.62	0.24	0.34	0.85	0.50
Maximum Leaf Width (cm)	1.30	2.40	0.58	0.44	1.37	0.94
Growth Rate (m/month)	0.38	0.74	0.08	0.10	0.43	0.28
Compressed to Basal Circumference Ratio	0.45	0.61	0.24	0.06	0.43	0.38

said trait is a productive use of resources or not must be considered. Likewise, if two measurements show a high correlation, should resources be spent to measure both?

The correlation of 0.98 between wet and dry yield exemplifies the previous point (Table 5). The same measurement is essentially being taken twice, yet drying down subsamples of each plant to calculate the dry yield nearly doubles the amount of work required to harvest individual plants. On the other hand, without both wet and dry yield, percent moisture (PM) cannot be determined without developing an alternative method. This trait is important because it heavily influences the cost of transportation, a major factor in assessing the feasibility of cellulosic ethanol production. An alternative means to quickly calculate PM will be needed if breeding programs wish to develop cultivars exhibiting increased dry down. The development of a marker assisted selection (MAS) program for the percent moisture at harvest would sidestep phenotyping altogether. Currently, three QTLs in our population for PM have been identified but all are of relatively small effect.

Another high correlation exists between dry weight (yield) and compressed circumference (CC) (Table 5). The correlation of 0.81 makes CC a potential surrogate for the much more time intense quantification of yield. At the very least, it is a great starting point to developing a regression model for yield. It is important to keep in mind that the relationship between CC and yield needs to be tested over multiple populations and multiple years to validate its stability and utility.

Although neither heading date nor 50% anthesis show very strong correlations to traits outside of flowering time, it is interesting to note that all of the correlations are stronger with heading date as opposed to 50% anthesis (Table 5). A number of potential explanations exist as

Table 5. Correlation between height (H), tiller diameter (TD), wet yield (WY), dry yield (DY), compressed circumference (CC), number of tillers (NT), basal circumference (BC), heading date (HD), 50% anthesis (A), spring emergence (E), leaf length (LL), average leaf width (LW), and compressed circumference to basal circumference ratio (CCBC). Data is from 2011 unless marked by an asterisk. All correlations are significant at an α=0.05.

	H	TD	WY	DY	CC	NT	BC	HD	A	E	LL	LW	CCBC	NT	HD	A
Height (H)	1.00															
Tiller Diam (TD)	0.23	1.00														
Wet Yield (WY)	0.53	0.30	1.00													
Dry Yield (DY)	0.55	0.28	0.98	1.00												
Comp Circ (CC)	0.38	0.25	0.83	0.81	1.00											
Num of Tiller (NT)	0.19	-0.31	0.56	0.56	0.79	1.00										
Basal Circ (BC)	0.38	0.21	0.68	0.66	0.68	0.50	1.00									
Head Date (HD)	-0.21	0.19	NS	NS	-0.08	-0.16	NS	1.00								
50% Anthesis (A)	-0.10	0.11	NS	NS	NS	-0.12	NS	0.59	1.00							
Spr Emerge (E)	-0.35	NS	-0.45	-0.45	-0.45	-0.33	-0.52	0.17	0.11	1.00						
Leaf Length (LL)	0.37	0.27	0.40	0.41	0.35	0.17	0.26	-0.11	NS	-0.18	1.00					
Leaf Width (LW)	0.24	0.44	0.33	0.30	0.29	NS	0.29	NS	NS	-0.18	0.14	1.00				
Comp/Bas (CCBC)	0.16	0.15	0.53	0.52	0.75	0.61	NS	NS	NS	-0.18	0.25	0.14	1.00			
*Num Tiller (*NT)	0.14	NS	0.42	0.42	0.41	0.35	0.47	NS	NS	-0.19	0.09	NS	0.13	1.00		
*Head Date (*HD)	-0.09	0.08	-0.14	-0.16	-0.20	-0.19	-0.20	0.53	0.32	0.16	NS	NS	-0.11	-0.32	1.00	
*50% Anth (*A)	NS	0.10	NS	-0.11	-0.14	-0.15	-0.14	0.51	0.31	0.11	NS	NS	NS	-0.25	0.87	1.00

* 2010 data

NS-Non-significant

to why heading date appears to be more accurate measure of flowering time. One explanation deals with the subjectivity of declaring a plant to have reached 50% anthesis. Human error is much less of a factor in recording heading data as the observation is highly distinguishable. An additional source of error in the 50% anthesis measure came from a number of plants displaying impaired inflorescence emergence. The inflorescence began to emanate from the flag leaf but then appeared to become immobilized by the tightness of the leaf sheath. In these instances, anthesis occurred within the sheath. This made it difficult to accurately determine the date at which the plant eclipsed the 50% anthesis mark. Lastly, heading date is closer to the reproductive phase change and therefore less likely to be influenced to the same extent as 50% anthesis is by the environment. Regardless of the explanation, it appears that heading date was a better measure of flowering time in 2011.

Interestingly, flowering time, although generally considered a key factor in high yielding varieties, shows no significant correlation with yield. This very well may be due to the fact that yield here is on a per plant basis as opposed to a per plot basis. Presumably, correlations with yield on a per area basis will be different than on a per plant basis. The expectation would be that the ability for a clone to spread would be favored on a per plant basis as well as in the establishment years of a plot while height, tiller density, and tiller diameter would increase in importance in mature stands. Of the few examples found in the literature, this holds true (Clifton-Brown and Lewandowski 2002; Jezowski 2008; Gauder et al. 2012).

Lastly, the strong negative correlation between spring emergence and a number of other traits including yield, basal circumference, and compressed circumference is logical in that a shorter growing period will lead to less growth. This explanation is weakened due to the

positive correlation between spring emergence and heading date. Individuals that emerged later in the spring had a tendency to flower later; this relationship equilibrates the length of the growing season between individuals emerging at different times. An alternative explanation is that the late emergence wasn't caused by the genotype but rather the establishment effect. Plants that did not establish well in the first year would have less opportunity to assimilate carbon and inorganic nutrients and translocate them to their rhizomes for overwintering. In turn, this leads to a delayed emergence as fewer nutrients are available to support strong spring regrowth. This relationship between carbon reserves in rhizomes and spring regrowth has long been established in the literature (White 1973; Dhont et al. 2002; Dhont et al. 2004). The hypothesis that spring emergence can be used as an indicator of establishment would be strengthened if it were found to not have a genetic basis, if in subsequent years the correlations are weakened, and if when used as a covariate, it increases the power to detect QTLs. While these correlations did yield telling information, it is yet unknown how the correlations will hold over subsequent years, over different populations, and into larger scale trials.

Permuted Significance Thresholds and Interval Mapping

Genome wide significance thresholds were determined empirically through permutation testing (Table 6). The significant thresholds are higher than the frequently used, arbitrary logarithm of odds (LOD) threshold of 3.0 because of the incorporation of an experimentwise error rate. While this cuts down on the number of false positives, it does come at the cost of failing to detect some true positives. The significance threshold will likely be relaxed when another year's data is available for cross validation of QTLs.

Table 6. The initial set of cofactors used in the MQM analysis based on the IM results for yield (Y), basal circumference (BC), compressed circumference (CC), compressed circumference to basal circumference ratio (CCBC), tiller diameter (TD), height (H), number of tillers (NT), 50% anthesis (A), heading date (HD), leaf length (LL), leaf width (LW), maximum leaf width (MLW), spring emergence (E), growth rate (GR), and percent moisture (PM).

Trait	GW[a]	LG[b]	Original Data[c]		Covariate Adjusted Data[d]	
			Position[e]	Marker[f]	Position	Marker
Yield	4.4	5	X	X	21.664	EBI 262
		6	78.275	EBI 372	78.653	EBI 373
		12	30.285	EBI 598	X	X
Basal Circumference	4.3	1	83.041	EBI 043	83.041	EBI 043
		3	13.699	EBI 135	13.699	EBI 135
		6	69.649	EBI 365	81.511	EBI 377
		10	X	X	38.358	EBI 532
Compressed Circumference	4.4	6	78.275	EBI 372	78.275	EBI 372
		7	X	X	15.243	EBI 392
		12	12.026	EBI 580	X	X
Compressed to Basal Circumference	4.2	6	78.844	EBI 374	X	X
		7	X	X	18.676	EBI 395
Tiller Diameter	4.5	3	13.699	EBI 135	13.699	EBI 135
		5	9.166	EBI 257	9.166	EBI 257
		7	81.699	EBI 432	81.699	EBI 432
Height	4.4	11	22.546	EBI 551	22.546	EBI 551
***Number of Tillers**	4.5	12	0	EBI 572	0	EBI 572
Number of Tillers	4.4	6	78.275	EBI 372	83.876	EBI 380
***50% Anthesis**	4.4	5	9.166	EBI 257	21.664	EBI 262
		16	47.816	EBI 722	47.816	EBI 722

Table 6. (Cont.)

Anthesis	3.8	5	13.448	EBI 259	9.166	EBI 257
		13	65.062	EBI 639	62.331	EBI 632
		16	66.7	EBI 729	66.7	EBI 729
***Heading Date**	4.3	5	9.166	EBI 257	9.166	EBI 257
		16	47.816	EBI 722	47.816	EBI 722
		19	54.425	EBI 838	54.425	EBI 838
Heading Date	4.4	3	108.375	EBI 184	110.041	EBI 185
		5	9.166	EBI 257	9.166	EBI 257
		13	X	X	57.671	EBI 630
		16	26.256	EBI 697	28.944	EBI 704
Leaf Length	4.4	5	102.456	EBI 292	102.456	EBI 292
		6	78.275	EBI 372	78.653	EBI 373
Leaf Width	4.5	3	13.699	EBI 135	13.699	EBI 135
Maximum Leaf Width	4.3	3	13.699	EBI 135	13.699	EBI 135
		7	81.699	EBI 432	92.028	EBI 434
Emergence	4.4	X	X	X	X	X
Growth Rate	4.4	1	X	X	28.421	EBI 022
		5	X	X	12.15	EBI 258
Percent Moisture	4.4	6	X	X	81.256	EBI 376

[a] Genome wide significant threshold
[b] Linkage group
[c] Interval mapping ran on the original data
[d] Interval mapping ran on the data adjusted for spring emergence
[e] Nearest marker position in cM to the maximum LOD score
[f] Marker name
X-LOD score failed to eclipse the permuted threshold
* 2010 trait data

Interval mapping (IM) was performed to get an initial look at the genetic architecture underlying the traits segregating in the population and establish linkage disequilibrium between markers and QTLs for the multiple-quantitative trait locus model (MQM). Since IM tests whether a single QTL is segregating on each chromosome, it is expected that both environmental variation and variation caused by other QTLs will lead to a less powerful test. Even so, interval mapping is a powerful QTL detection tool. It effectively identified a number of loci with segregating QTLs and highlighted closely linked markers to use as cofactors in the more powerful MQM mapping (Table 6).

Multiple-QTL Model

A total of 31 QTLs were identified over the 15 traits analyzed using the original data with the MQM model (Table 7). The largest number of QTLs was identified for heading date: three in 2010 and an additional four in 2011. No significant QTLs were identified for spring emergence, percent moisture, growth rate, and compressed circumference to basal circumference ratio. Of the limited number of traits with multiple years of data, A1, A2, and HD1 were found to be stable across years. Significant LOD scores ranged from 3.8-46.0 with an average of 9.8 while the percent variation explained (PVE) by individual QTLs ranged from 3.3-67.5 with an average of 14.7. Out of the 11 traits with significant QTLs, genetics accounted for the least amount of variation in height at only 11.2% and the most with maximum leaf width at 78.5%.

Covariate Analysis

In general, more QTLs were found when genotypic means were adjusted for spring emergence. An additional eleven QTLs were discovered when using spring emergence as a covariate, but six previously discovered QTLs failed to meet the significance threshold (Table 7).

Table 7. Results from the MQM analyses based on the original phenotypic data and the data adjusted for differences in establishment using spring emergence as a covariate for yield (Y), basal circumference (BC), compressed circumference (CC), compressed circumference to basal circumference ratio (CCBC), tiller diameter (TD), height (H), number of tillers (NT), 50% anthesis (A), heading date (HD), leaf length (LL), leaf width (LW), maximum leaf width (MLW), percent moisture (PM), growth rate (GR), and emergence (E).

		Original Data			Covariate Adjusted Data		
QTL	LG[a]	CM (LOD)[b]	PVE[c]	2-LOD CI[d]	CM (LOD)	PVE	2-LOD CI
Y1	5	27.6 (6.8)	10.8	20.7↔41.1	26.7 (8.1)	13.8	14.2↔27.3
Y2	6	74.9 (12.6)	20.1	71.6↔81.5	77.4 (11.8)	19.1	72.6↔81.5
Y3	12	32.7 (4.7)	6.7	30.3↔37.5	X	X	X
BC1	1	81.9 (4.7)	9.5	73.9↔92.1	X	X	X
BC2	3	X	X	X	9.4 (4.9)	9.2	end[e]↔13.7
BC3	6	69.6 (6.3)	10.8	63.7↔72.9	80.4 (5.8)	9.9	76.9↔83.9
BC4	7	X	X	X	27.7 (4.6)	6.2	18.3↔29.1
BC5	10	X	X	X	37.0 (4.4)	7.6	29.6↔42.8
BC6	19	X	X	X	59.4 (6.7)	10.8	54.8↔61.9
CC1	5	X	X	X	25.7 (4.6)	7.2	20.7↔41.1
CC2	6	74.9 (10.2)	18.1	71.6↔81.5	78.4 (10.6)	16.5	72.6↔81.5
CC3	7	X	X	X	15.3 (4.8)	8.5	10.2↔19.2
CC4	12	8.0 (5.1)	8.9	7.3↔15.2	X	X	X
CCBC1	7	X	X	X	19.2 (4.7)	11.3	14.2↔26.7
TD1	3	15.7 (7.3)	15.3	8↔21.0	16.7 (6.3)	10.3	8.0↔21.0
TD2	5	5.6 (4.8)	9	0↔14.7	4.6 (6.0)	10	0↔14.7
TD3	7	82.5 (4.9)	7.6	75.3↔92.0	82.5 (4.8)	7.4	76.3↔92.0
H1	11	23.4 (5.0)	11.2	22.4↔27.9	25.0 (5.9)	13.4	19.1↔37.9
***NT1**	12	0 (5.7)	12.6	end↔5.3	0 (6.2)	13.8	end ↔4.3
NT2	6	79.4 (6.9)	13.8	71.6↔81.5	83.9 (7.4)	15.2	81.4↔97.9
***A1**	5	9.2 (6.6)	11.4	0↔14.7	19.7 (7.2)	12.4	14.2↔27.3
***A2**	16	47.8 (12.5)	19.6	45.2↔56.7	47.8 (13.4)	21.3	45.2↔56.7

Table 7. (Cont.)

A1	5	9.2 (6.8)	12.3	0.6↔14.7	0.6 (30.0)	34.9	0↔4.6
A3	13	59.5 (3.9)	6	57.8↔63.3	63.3 (26.8)	37.9	62.5↔65.3
A4	15	56.5 (3.8)	6.2	51.6↔69.7	X	X	X
A2	16	55.8 (24.9)	32.5	47.8↔56.7	56.7 (27.2)	38.3	47.8↔57.7
***HD1**	5	6.6 (7.1)	12	0↔14.7	5.6 (8.2)	13.4	0↔14.7
***HD2**	16	47.8 (9.5)	14.5	43.2↔56.7	47.8 (10.6)	15.5	43.2↔56.7
***HD3**	19	55.8 (4.4)	6.8	49.3↔58.4	56.8 (5.4)	8.1	50.3↔58.4
HD4	3	100.0 (10.7)	10.1	96.0↔103.2	98.6 (9.6)	10.5	96.0↔103.2
HD5	4	X	X	X	46.7 (5.7)	6.5	41.1↔59.3
HD1	5	9.2 (15.7)	21.4	2.6↔14.7	8.6 (14.3)	19.3	0.6↔14.2
HD6	13	57.8 (5.5)	5.8	50.9↔60.5	60.5 (8.8)	9.7	57.8↔63.3
HD7	15	29.3 (5.3)	5.7	23.9↔30.6	X	X	X
HD8	16	32.1 (10.3)	11	26.1↔34.1	32.1 (8.4)	8.3	26.1↔33.1
LL1	5	97.1 (4.9)	9.3	84.9↔106.4	97.1 (4.7)	9.2	86.0↔106.3
LL2	6	77.4 (5.9)	10.4	71.6↔81.5	78.4 (5.0)	9.4	72.6↔81.5
LW1	3	12.4 (46.0)	67.5	10.4↔13.7	12.4 (54.7)	78.6	9.4↔13.7
LW2	6	69.6 (4.9)	3.3	63.7↔72.9	X	X	X
MLW1	2	20.5 (8.5)	6.6	16.5↔21.4	20.5 (8.9)	6.7	16.5↔21.4
MLW2	3	12.4 (46.0)	67.5	8.4↔13.7	12.4 (47.4)	67.8	9.4↔13.7
MLW3	7	82.5 (5.7)	4.4	75.3↔92.0	82.5 (6.4)	4.8	76.6↔92.0
PM1	3	X	X	X	8 (5.3)	9.9	end↔13.7
PM2	4	X	X	X	49.3 (4.6)	8.9	41.1↔59.3
PM3	6	X	X	X	80.4 (4.9)	9.2	76.9↔83.9
GR	X	X	X	X	X	X	X
E	X	X	X	X	X	X	X

* 2010 data, [a] Linkage group, [b] Peak position in centiMorgans (Maximum LOD score), [c] Percentage of phenotypic variance explained by QTL, [d] 2-LOD support confidence interval, [e] CI goes beyond the last marker on the LG, X LOD score failed to eclipse the permuted threshold

The largest difference between adjusted and unadjusted means was seen for basal circumference; four additional QTLs were discovered. Other traits where spring emergence increased the number of QTLs found were compressed circumference, compressed to basal circumference ratio, heading date, and percent moisture. Previously discovered QTLs were not detected using adjusted means for yield, basal circumference, compressed circumference, 50% anthesis, heading date, and leaf width although only yield, 50% anthesis and leaf width resulted in a net loss of QTLs.

It is expected that using a covariate to account for some of the phenotypic variation within the mapping population would lead to mixed results because some traits are less influenced by the establishment effect than others. This is seen in which traits benefited most from the analysis of covariance. If a plant has trouble establishing, it won't aggressively spread or send up a large number of tillers; this results in decreased circumferences and yield. Even though Y3 wasn't detected using a genome wide threshold when means were adjusted for the emergence date, the peak LOD score of 3.3 on chromosome 12 easily surpasses the chromosome wide threshold of 2.8. More often than not, entirely new LOD peaks aren't discovered; rather, the height of the peaks is altered. The MQM results for basal circumference show this phenomenon well (Figure 6). As observed, traits that are less affected by the number of tillers like leaf width, tiller diameter, and leaf length wouldn't be expected to benefit from the covariate analysis. In summary, covariate analysis can be a useful way to remove some residual variation caused by the establishment effect for a number of traits. A more direct measure of plant establishment may further increase the utility of covariate analysis. One potential measurement that may be well suited for use is a measure of root ball health at the

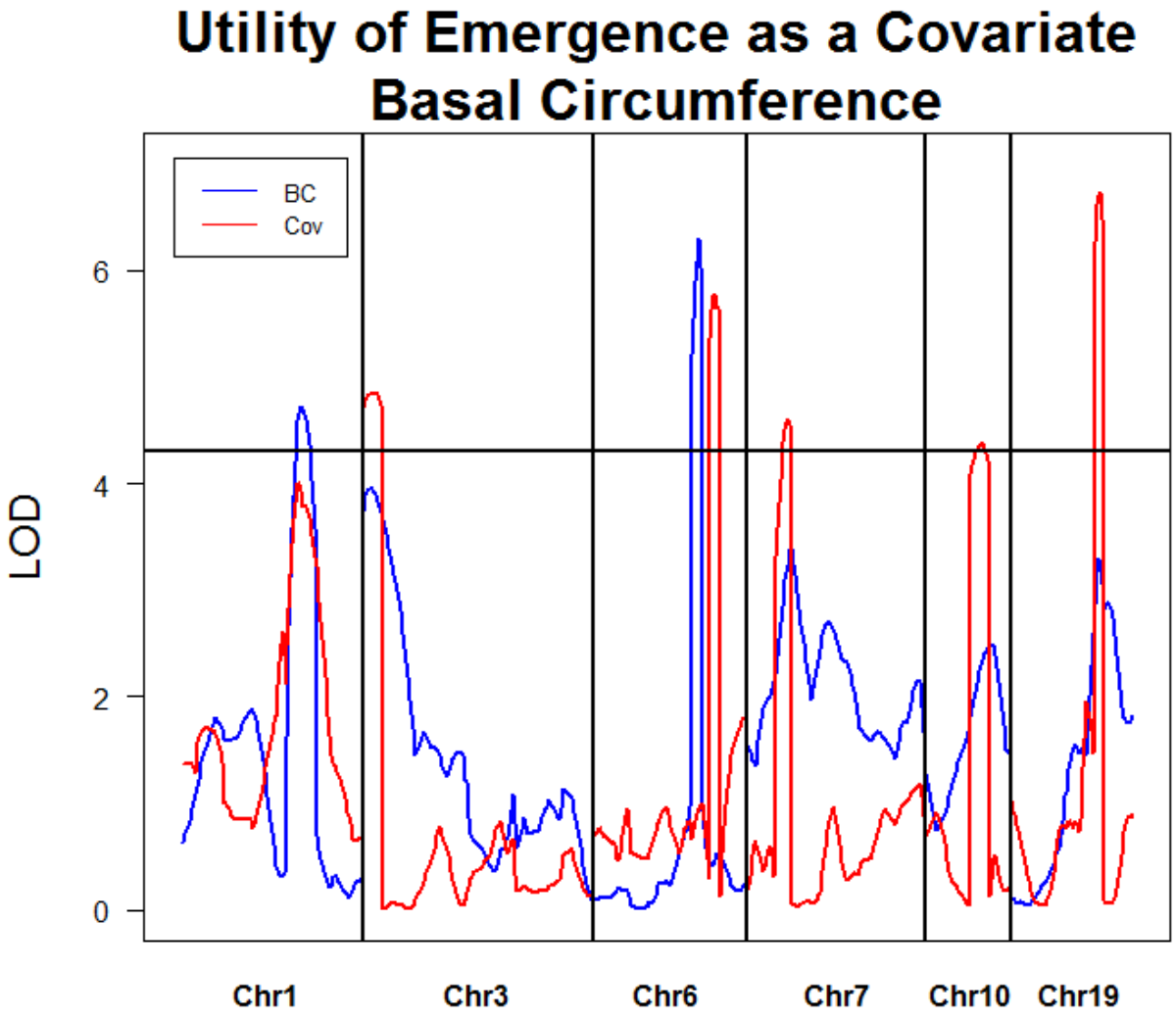


Figure 6. P-value plots for basal circumference showing the effect that using a covariate has on the QTL analysis. Only the chromosomes on which QTL are located are depicted. The horizontal line represents the significance threshold. New peaks aren't generally discovered using the adjusted means; the LOD scores are either accentuated or unaccentuated.

time of transplanting. Henceforth, QTLs from both analyses will be combined. When identical loci were found in both analyses, the one reported will be the analysis that found the larger number of QTLs for that particular trait. If equal numbers of QTLs were found, results are reported from the original analysis.

Yield

Three QTL were identified for yield (Y) which cumulatively explain 37.6% of the variation (Table 7). Of the three, Y2 is the largest accounting for 20.1% of the variation. Y2 colocalizes with QTLs for basal circumference, compressed circumference, leaf length, average leaf width, percent moisture, and the number of tillers on chromosome six giving us more confidence that these QTLs are true positives (Figure 7). At this stage, it is impossible to know for certain whether this region represents a number of different genes each affecting different traits in tight linkage or one or two genes that have a pleiotropic effect. It is possible that a gene involved early in the regulatory process of axillary bud formation and elongation may be underlying this locus as this would affect a number of the traits. Monoculm1 (Moc1) was isolated and characterized in rice to control the initiation and outgrowth of tillering buds (Li et al. 2003). This gene belongs to a family of GRAS transcription factors and is homologous to the Lateral Suppressor gene in tomato and *Arabidopsis* (Schumacher et al. 1999; Greb et al. 2003). Due to the expression patterns of other key tillering genes in relation to Moc1, it is believed that this gene participates early in tiller initiation (Li et al. 2003). A gene of this nature is one possible explanation for the colocalization of so many traits.

The only previous QTL study on yield in *Miscanthus* found a total of six QTLs over a two year period (Atienza et al. 2003b). As noted, a number of problems exist with this study. One of

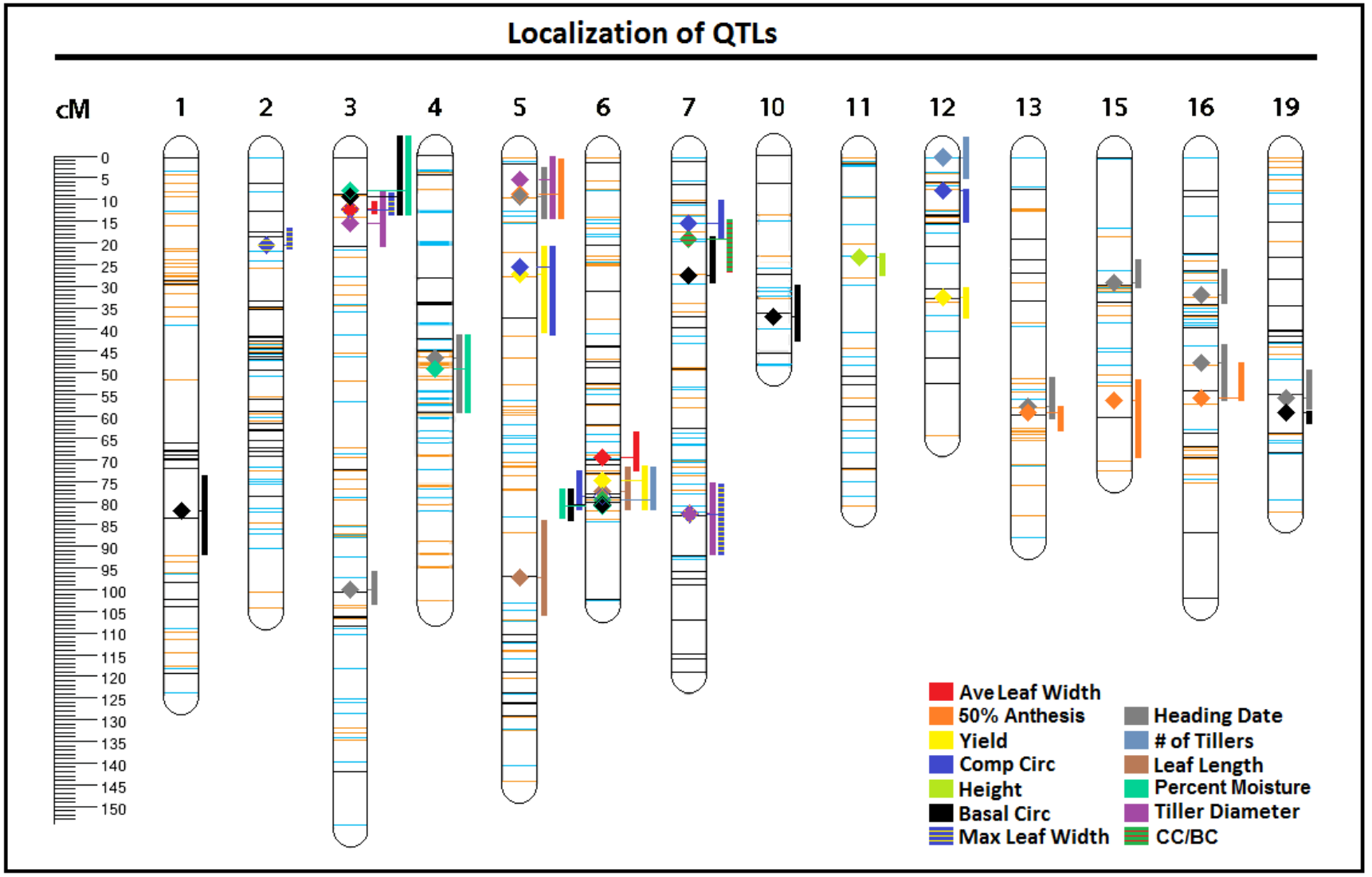


Figure 7. Localization of QTLs found in the MQM analysis for indicated traits. Chromosome numbers are indicated across the top and map units in cM along the left hand side.

the main reasons Atienza and his colleagues found a larger number QTLs was their use of a relaxed significance threshold. In fact, none of the six QTLs found by this research group would have been considered significant using the same significance criteria used herein. Unfortunately, cross validation of QTL positions is impossible due to the design of their genetic map.

In most agronomically important crops, yield refers to a particular sink tissue. Therefore, yield gains can be realized by manipulating the allocation of resources within the plant. This is a bit different in *Miscanthus* since yield refers to the entirety of the above ground portion. For this reason biomass production must be increased by lengthening the duration of growth, increasing the rate of growth, or a combination of the two. Attempts to link biomass QTLs with underlying genetic causes have been made. Lisec et al. (2008) found a relation between primary metabolism QTLs and biomass QTLs in *Arabidopsis*. Cross and colleagues (2006) proposed that variation in the growth rates of 24 *Arabidopsis* accessions was caused by variation in enzymatic capacity to use carbon and nitrogen and not variation in assimilation rates. These studies suggest genes influencing metabolomics can underlie the yield QTLs.

Complex quantitative traits such as biomass are expected to be influenced by many genes with most having a small effect (Kroymann and Mitchell-Olds 2005; Lisec et al. 2008). The reason only three QTLs were discovered for yield most likely stems from the inability to detect small effect QTLs. One factor limiting the power of this analysis is the high stringency level in place to guard against false positives. Power is also limited in the second year by the larger QTL effect variance caused by the establishment effect.

Compressed Circumference, Basal Circumference, and their Ratio

Four, six, and one QTLs were identified for compressed circumference (CC), basal circumference (BC), and the ratio between the two traits (CCBC), respectively (Table 7). Seven of these QTLs were identified exclusively when analyzing covariate adjusted genotypic means (CC1, CC3, BC2, BC4, BC5, BC6 and CCBC1). PVE by individual QTLs ranged from 6.2 for BC4 to 18.1 for CC2. CC2 and BC3 colocalize on chromosome six with five other traits as mentioned in preceding comments (Figure 7). CC3, BC4, and CCBC1 colocalize on chromosome seven providing further evidence of their validity. Another interesting aggregation of QTLs occurs at the beginning of chromosome three. BC2 colocalizes with the two leaf width traits, tiller diameter, and percent moisture. This region is of importance because it controls the thickness of tillers and the spreading habit, both very important traits for a biomass crop. Once again, this may be the result of a pleiotropic gene affecting both rhizome growth and stem primordia.

As this is the first time these traits have been analyzed for QTLs in *Miscanthus*, no comparisons can be made to previous work within the genus. Recent work regarding rhizomatous nature of rice has focused on understanding the genetic control in hopes of creating a perennial crop (Hu et al. 2003; Hu et al. 2011). Hybridizations between perennial wild rice and domesticated annual rice have revealed the complex nature of this trait. Gene expression experiments have revealed the tissue specific nature of regulatory pathways and the involvement of the plant hormones auxin and gibberellin (Hu et al. 2011). It is difficult to pinpoint a particular gene accounting for variation in BC and CC, but any gene regulating the levels of plant hormones within the rhizomes could account for variation in the spreading habit of individuals in the mapping population. Regardless of the underlying mechanism, the high

correlation between Y, CC, and BC in addition to the large number of QTLs found for these traits may make these QTLs suitable targets for marker assisted selection. Of course this depends if correlations are maintained through subsequent years and to plot testing as well if variation at this locus exists in the breeding population.

Tiller Diameter, Height, and Number of Tillers

A total of six QTLs were found for tiller diameter (TD), height (H), and number of tillers (NT) collectively (Table 7). Three QTLs were identified for TD at mid-plant height, a single QTL for H, and two QTLs for the NT. Cumulative PVE ranged from 11.2 to 31.9 for H and TD, respectively. While H1 was the lone QTL found on chromosome 11, two of the three TD QTLs coincided with a leaf width QTL (Figure 7). Another colocalization occurred between the two flowering time traits and TD on chromosome five. In general, larger tiller diameter was associated with wider leaves and later flowering. Evolutionarily speaking this is justifiable as later flowering genotypes tend to be taller and need thicker stems to remain erect.

The fact that stem diameter and leaf width seem to be so closely linked in this population yet there is such a disparity in the cumulative PVE between these two traits (31.9 vs. 78.5) raises concern regarding the effectiveness of the data collection. Data could be affected by many variables: inconsistency in the height where the diameter is measured, deviation of the measurement from the internode, differences in leaf sheath thickness, error caused by variation in size between tillers on the same plant, or equipment error. A more comprehensive data set has been collected on a subsample of the population and will be used to refine the measuring strategy if necessary.

The lack of QTLs found for plant height is a bit of a surprise. Previous QTL analysis on height and stem diameter in *Miscanthus* revealed five and three QTL, respectively (Atienza et al. 2003a). Perhaps a better measurement of height would be an average tiller height as opposed to the tallest tiller. While this will introduce subjectivity, it may provide a better representation of the genetic potential of a genotype. As for the number of tillers, no consistency was observed between years. This may be partially explained by the difference in measuring technique. In 2010, actual count data was taken while 2011 data was calculated based on tiller diameter and compressed circumference due to the time prohibitive nature of counting tillers beyond the first year. Surprisingly, the QTL analysis on the number of tillers did not benefit from the use of the covariate analysis.

Flowering Traits

The two flowering traits, 50% anthesis (A) and heading date (HD), yielded a total of ten unique loci over the two years (Table 7). A1 and HD1 colocalize on the fifth chromosome and were found in both 2010 and 2011 (Figure 7). A2 and HD2 colocalize on chromosome 16 but HD2 was only found in 2010. HD8, found in 2011, is also on chromosome 16 but the two QTL's confidence intervals are separated by approximately 10 cM. 50% anthesis had a cumulative PVE of 31 in 2010 and 52.1 in 2011 although the later is highly suspect for reasons explained shortly. HD has a cumulative PVE of 33.3 in 2010 and 54.0 in 2011. This increase in ability to explain flowering time over subsequent years following establishment is expected because the establishment effect diminishes over time.

The reason why the 2011 PVE for A should be cautiously interpreted is because of the abnormally high LOD score and associated PVE of A2 on chromosome 16. The LOD score and

PVE under IM for A2 was 5.55 and 11, respectively, while the MQM analysis yielded a LOD score and PVE of 24.9 and 32.5. The reason for skepticism is because the maximization-likelihood (ML) algorithm assumes the component distributions behave according to a normal distribution. When this is not the case, the iterative procedure of the ML algorithm fails to converge quickly resulting in artificially high LOD scores. While this can be a problem for both IM and MQM, it isn't surprising to see the effect solely in the MQM analysis. In MQM analysis, component distributions are adjusted for other QTLs in the model; this is not the case in IM. If linkage between a large effect QTL and the marker used as a cofactor in the analyses is not tight, then the distributions could be skewed enough to break the normality assumption. Since the program does not report statistics beyond the mean of the component distributions, the only way to check if this is indeed the problem is to look at the number of iterations prior to convergence at the locus in question. In this case, the number of iterations was 15 at the peak LOD position under the MQM and only four in IM.

Flowering time is one of the most extensively studied plant traits due to its importance to local adaptation, stress avoidance, and yield. A considerable amount of knowledge of the flowering pathway has come from the model species *Arabidopsis*, but this has been applied to most important agronomic crops. Descriptions of the flowering pathway have been described for grasses such as rice, maize, wheat, and sorghum (Tsuji et al. 2011; Xu et al. 2012; Shimada et al. 2009; Murphy et al. 2011). Sorghum being the closest in relation will likely be the most useful in comparative studies to better understand flowering time in *Miscanthus*. A quick comparison of the QTLs identified for flowering time in this study to known sorghum flowering

genes mapped onto the sorghum consensus map by Mace et al. (2009) revealed the coincidence of the Ma5 gene and HD4.

Leaf Traits

The Licor 3000 simultaneously measures leaf length (LL), average leaf width (LW), and maximum leaf width (MLW). Two QTLs were identified for LL and LW while three QTLs where identified for MLW (Table 7). LL1 and LL2 explain 9.3 and 10.4 percent of the variation respectively. MLW and LW share a common large effect QTL at the front end of chromosome three. The PVE should once again be viewed somewhat cautiously due to the large number of iterations taken by the ML algorithm. Unlike the previous case, other QTLs were not affecting the normality of the component distributions as IM had a comparable number of iterations to MQM (11 vs. 14). In fact, the LOD score was actually higher for IM (55 vs. 46). Coupled with the fact that estimated means for the component distributions were 1.07, 0.96, 1.83, and 1.21 for QTL genotypes ac, ad, bc, and bd, respectively, these equally high LOD values provide support that a large effect QTL is segregating, and that it explains much of the variation observed in leaf width. Further evidence for a large segregating QTL can be seen in the trait distribution for LW and MLW (Figure 8). The roughly bimodal shape suggests a single gene segregating in the progeny.

The colocalization of LL2 and LW2 on chromosome six is not surprising as LW takes into account both leaf length and width. Due to the large effect of a single segregating gene, leaf width acts similar to a qualitative trait. If leaf width was difficult to phenotype or was highly affected by establishment, which it is not, it would be a good candidate for marker assisted

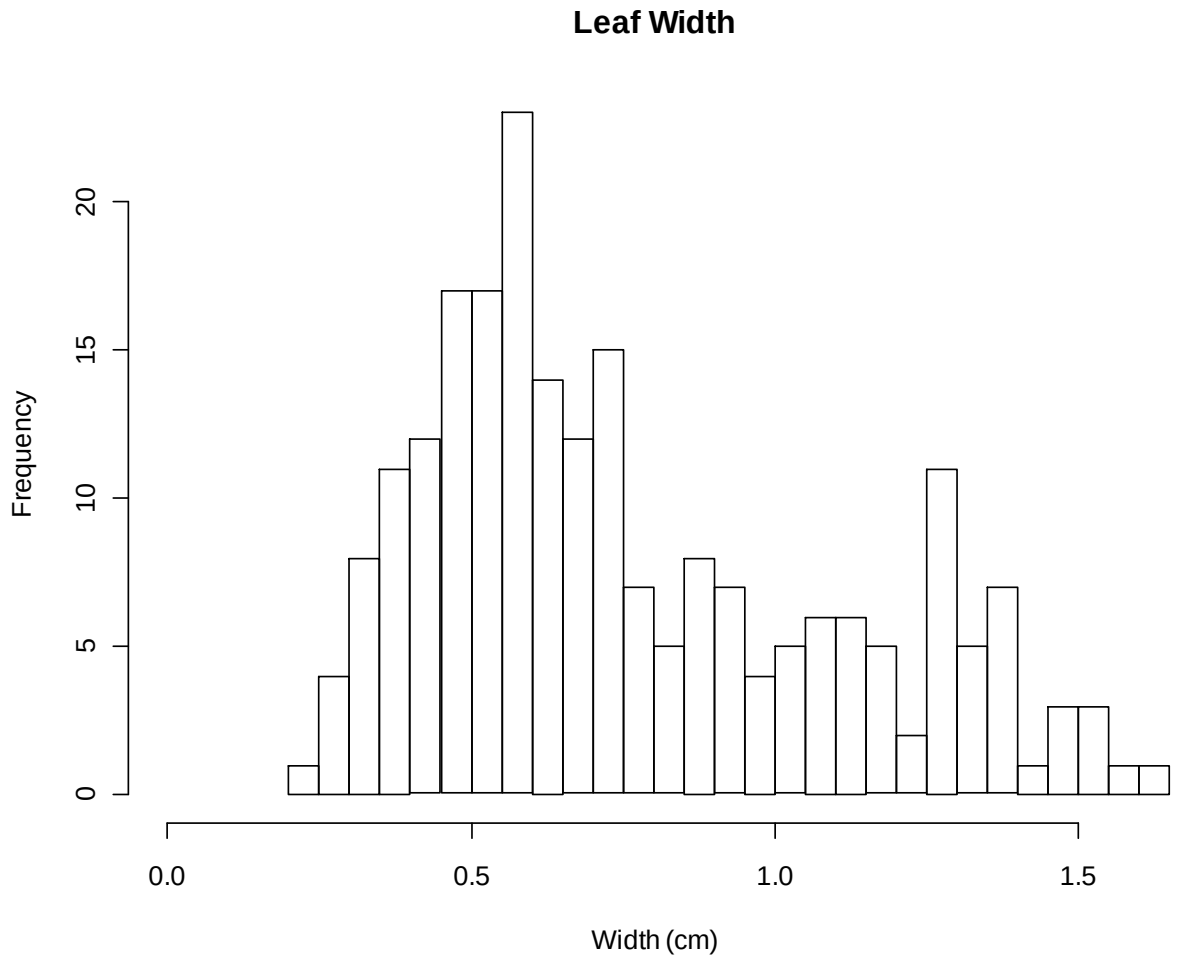


Figure 8. Histogram showing the roughly bimodal distribution expected of a trait controlled by a single large effect gene.

selection. A more useful application of leaf width may be to estimate tiller diameter due to their correlation and colocalization.

Percent Moisture, Growth Rate, and Spring Emergence

Interval mapping of percent moisture (PM), growth rate (GR), and spring emergence (E) using the original data uncovered no QTLs. The fact that E has no QTLs provides support for its use as a covariate. Since there is not enough power to detect any QTLs, it is assumed that environmental variation outweighs any genetic effects controlling the trait. Three QTLs were found for PM using the adjusted data all with PVE below ten. Power to detect GR QTLs, calculated from changes in height over time, did not benefit from the adjustment. GR was not one of the initial traits of interest in this mapping population so the phenotypic data collection was not optimized for such a characteristic. Li et al. (2006) report finding different QTLs in rice for growth rate depending on the time of year the measurement was taken. Half of the alleles found to enhance growth rate early in the season were found in one parent while the QTLs found to promote late season growth came from the opposite parent. Further evidence of this phenomenon is provided by Xu et al. (2004) who found different height QTLs in rice depending on the timing of the data collection. The possibility of stacking positive growth rate alleles for different parts of the growing season may be a useful way to increase yield in *Miscanthus*. In order to do so, more appropriate growth rate indicators will need to be found and the germplasm surveyed for variation.

Future Work

Identifying QTLs is the first step towards understanding the genetic complexity underlying quantitative traits. A deeper understanding of the genetic pathways affecting key

agronomic traits can in turn influence plant improvement programs, albeit unnecessary to identify genes underlying QTLs to use the information to benefit breeders. Establishing linkage between QTLs and markers is the first step towards fine mapping, positional cloning, and developing a marker assisted breeding program. In a perennial crop like *Miscanthus* with a three year establishment period, MAS has the potential to drastically accelerate improvement.

Conclusion

Many of the traits presumed to be important to biomass production show strong correlations to dry biomass yield and amongst themselves. The correlation of wet and dry yield of 0.98 suggests the dry weight calculations are not necessary to identify superior yielding genotypes in this particular population. The flaw with this time and money saving endeavor is the need to develop an alternative method to measure percent moisture. Another noteworthy correlation is the 0.81 correlation between dry biomass yield and compressed circumference. A correlation of this magnitude suggests a regression model for second year yield may be able to sufficiently explain yield. Lastly, the correlation between flowering time traits and the other traits measured suggest heading time is a more accurate measure of flowering time than a measure of 50% anthesis. In saying this, it is important to remember the real interest lies in how these correlations change over time and if any correlations can be established from individual plant traits to plot yield.

Beyond correlations, this analysis has revealed the first look at the genetic architecture underlying important quantitative traits using a complete genetic map in *Miscanthus sinensis*. Interval mapping was successfully employed along with the use of permutated genome wide thresholds to determine an initial set of cofactors to be used in the more powerful composite interval mapping strategy. Since strong correlations between spring emergence and a number of traits existed yet no genetic basis seemed to be underlying the phenotypic variation observed for spring emergence, it was tested as a covariate to capture some of the variation seen in other traits due to the establishment effect. This strategy resulted in heterogeneous success across the measured traits. In general, traits most likely affected by uneven plant

establishment such as compressed circumference and basal circumference benefit from this covariate analysis while other less affected traits remain neutral to this statistical adjustment to phenotypic means.

In total, 42 QTLs were identified on 14 different chromosomes across the 15 traits. A wide range of PVE and cumulative trait PVE was observed. Two spots in particular appear to be important to a plethora of traits and might be interesting candidates to investigate further. Due to the nature of quantitative traits and second year data, many small effect QTLs were likely left undiscovered. Both the analysis of third year data and the comparison between the third and second year data will yield more power and presumably discover a larger number of these small effect QTLs. We can infer from these results which traits seem to have a larger amount of heritability in the establishment phase. Flowering and leaf width appear to be selectable in early stages while plant height and number of tillers are not.

The limited number of QTLs found for certain traits suggests either our phenotypic data collection method is flawed, the establishment effect causing variation in some traits is not captured by the covariate, environmental variation drowns out the genetic effect, QTLs are not segregating in this population, or all QTLs have effects beyond the power of detection. The modification of certain phenotypic data collection protocols as well as another year's data will help alleviate this problem.

This work sets the foundation for fine mapping of QTLs, positional cloning of underlying genes, and marker assisted selection. The highly conserved synteny with sorghum can be used to not only find gene candidates for QTLs as demonstrated herein, but also to compare QTL between the two species. MAS has huge potential in *Miscanthus* due to the long establishment

period, difficulty in phenotyping, and outbreeding nature. A successful MAS breeding program would decrease the generation time and allow a larger number of individuals to be screened.

Even without the aid of MAS, these QTL analyses have given a more complete understanding of the traits important to biomass and how they are affected by the establishment period. This allows for a more efficient allocation of resources and improved phenotypic data collection. Ultimately, this work will help to increase *Miscanthus sinensis* yields. Potentially these improved lines could become cultivars in their own right or be used to recreate the *M. x giganteus* cross to provide alternative genotypes and increase the genetic variability of clones available for use as a lignocellulosic biomass feedstock crop. Improved cultivars will be needed for the cellulosic ethanol industry to meet the requirements set forth by the government, to break our dependence on foreign oil, and provide an environmental friendly, renewable energy source.

Literature Cited

2010 Annual Energy Review (2011) Department of Energy/U.S. Energy Information Administration-0384

Angelini LG, Ceccarini L, Nassi o Di Nasso N, Bonari E (2009) Comparison of *Arundo donax* L. and *Miscanthus × giganteus* in a long-term field experiment in Central Italy: analysis of productive characteristics and energy balance. Biomass and Bioenergy 33:635–643

Appleby N, Edwards D, Batley J (2009) New technologies for ultrahigh throughput genotyping in plants. Methods Molecular Biology 513:19-39

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2002) Preliminary genetic linkage map of *Miscanthus sinensis* with RAPD markers. Theoretical and Applied Genetics 105:946–952

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martın A (2003a) Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anderss. I. total height, flag-leaf height and stem diameter. Theoretical and Applied Genetics 107:123–129

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003b) Identification of QTLs associated with yield and its components in *Miscanthus sinensis* Anderss. Euphytica 132:353–361

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003c) Identification of QTLs influencing combustion quality in *Miscanthus sinensis* Anderss. II. Chlorine and potassium content. Theoretical and Applied Genetics 107:857–863

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003d) Influencing combustion quality in *Miscanthus sinensis* Anderss. Identification of QTLs for calcium, phosphorus and sulphur content. Plant Breeding 122: 141–145

Barloy D, Lemoine J, Abelard P, Tanguy AM, Rivoal R, Jahier J (2007) Marker-assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat. Molecular Breeding 20:31–40

Beadle CL, Long SP (1985) Photosynthesis-is it limiting to biomass production? Biomass 8:119–168

Beale CV, Bint DA, Long SP (1996) Leaf photosynthesis in the C4-grass *Miscanthus x giganteus*, growing in the cool temperate climate of southern England. Journal of Experimental Botany 47:267-273

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in Proceedings of the Forty-Ninth Annual Corn & Sorghum Industry Research Conference. American Seed Trade Association, Washington, DC

Beavis WD (1998) QTL analyses: power, precision, and accuracy, pp. 145–162 in Molecular Dissection of Complex Traits, edited by A. H. PATERSON. CRC Press, New York

Bennewitz J, Reinsch N, Kalm E (2002) Improved confidence intervals in quantitative trait loci mapping by permutation bootstrapping. Genetics 160:1673-1686

Bentham G (1882) Notes on Gramineae. Journal of the Linnean Society, Botany 19:14-134

Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A (2002) Marker assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics 162:1945–1959

Brown RH (1978) A difference in N use efficiency in C3 and C4 plants and its implication in adaptation and evolution. Crop Science 18:93-98

Burner D (1991) Cytogenetic analyses of sugarcane relatives (*Andropogoneae: Saccharinae*). Euphytica 54:125-133

Chiang YC (1993) The biochemical tolerance of *Miscanthus floridulus* populations to lead. Master's thesis, Faculty of Sciences, Department of Botany, National Taiwan University, Taipei

Clifton-Brown JC, Lewandowski I, Andersson B, Basch G, Christian D, Bonderup-Kjeldsen J, Jørgensen U, Mortensen J, Riche AB, Schwarz KU, Tayebi K, Teixeira F (2001) Performance of 15 *Miscanthus* genotypes at five sites in Europe. Agronomy Journal 93:1013–1019

Clifton-Brown JC, Lewandowski I (2002) Screening *Miscanthus* genotypes in field trials to optimize biomass yield and quality in Southern Germany. European Journal of Agronomy 16:97-110

Clifton-Brown JC, Stampfl PF, Jones MB (2004) *Miscanthus* biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. Global Change Biology 10:509–518

Clifton-Brown JC, Jones MB (2007) Carbon mitigation by the energy crop, *Miscanthus*. Global Change Biology 13:2296–2307

Clifton-Brown JC, Chiang YC, Hodkinson TR (2008) *Miscanthus*: genetic resources and breeding potential to enhance bioenergy production. In: Genetic Improvement of Bioenergy Crops (ed. Vermerris W), pp. 273–294. Springer Science, New York

Christian DG, Riche AB, Yates NE (2008) Growth, yield and mineral content of *Miscanthus × giganteus* grown as a biofuel for 14 successive harvests. Industrial Crops and Products 28:320–327

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

Cross JM, von Korff M, Altmann T, Bartzetko L, Sulpice R, Gibon Y, Palacios N, Stitt M (2006) Variation of enzyme activities and metabolite levels in 24 *Arabidopsis* accessions growing in carbon-limited conditions. Plant Physiology 142:1574–1588

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J.R. Statistical Society Ser. B 39:1-38

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental population. Nature Reviews 3:43-52

DOE/USDA (2005) Biomass as feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply

Dhont C, Castonguay Y, Nadeau P, Bélanger G, Chalifour FP (2002) Alfalfa root carbohydrates and regrowth potential in response to fall harvests. Crop Science 42:754–765

Dhont C, Castonguay Y, Nadeau P, Bélanger G, Drapeau R, Chalifour FP (2004) Untimely fall harvest affects dry matter yield and root organic reserves in field-grown alfalfa. Crop Science 44:144–157

Dupuis J, and Siegmund D (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151:373–386

Ehleringer JR, Monson RK (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. Annual Review of Ecology, Evolution, and Systematics 24:411–439

Ethanol Myths & Facts (2011) U.S. DOE http://www1.eere.energy.gov/biomass/ethanol_myths_facts.html

Evans JR (1989) Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia 78:9-19

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics 4th ed. London, Prentice Hall

Gauder M, Graeff-Honninger S, Lewandowski I, Claupein W (2012) Long-term yield and performance of 15 different *Miscanthus* genotypes in southwest Germany. Annals of Applied Biology 160:126-136

Greb T, Clarenz O, Schafer E, Müller D, Herrero R, Schmitz G, Theres K (2003) Molecular analysis of the LATERAL SUPPRESSOR gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. Genes Dev. 17:1175-1187

Heaton EA, Clifton-Brown J, Voigt TB, Jones MB, Long SP (2004a) Miscanthus for renewable energy generation: European union experience and projections for Illinois. Mitigation and Adaptation Strategies for Global Change 9:433–451

Heaton E, Voigt T, Long SP (2004b) A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature and water. Biomass and Bioenergy 27:21–30

Heaton E, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. Global Change Biology 14:1-15

Helguera M, Khan IA, Kolmer J, Lijavetzky D, Zhong-qi L, Dubcovsky J (2003) PCR assays for the Lr37-Yr17-Sr38 cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Science 43:1839–1847

Hodkinson TR, Chase MW, Lledo MD, Salamin N, Renvoize SA (2002a) Phylogenetics of *Miscanthus Saccharum* and related genera (*Saccharinae Andropogoneae Poaceae*) based on DNA sequences for ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. Journal of Plant Research 115:381-392

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA (2002b) The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (*Poaceae*). American Journal of Botany 89:279–286

Honda M (1930) *Monographia, Poacearum, Japonicarum, Bambusoideis exclusis*. J Fac Sci, Imperial University Tokyo Section 3 Bot 3:1–484

Hospital F (2009) Challenges for effective marker-assisted selection in plants. Genetica 136:303–310

Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li ZK (2003) Convergent evolution of perenniality in rice and sorghum. Proceedings of the National Academy of Sciences 100:4050-4054

Hu F, Wang D, Zhao X, Zhang T, Sun H, Zhu L, Zhang F, Li L, Li Q, Tao D, Fu B, Li Z (2011) Identification of rhizome-specific genes by genome-wide differential expression analysis in *Oryza longistaminata*. BMC Plant Biology 11:18

International Energy Outlook 2011 (2011) Department of Energy/U.S. Energy Information Administration-0484

Jakob K, Zhou FS, Paterson A (2009) Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. In Vitro Cellular and Developmental Biology-Plant 45:291–305

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Jezowski S (2008) Yield traits of six clones of *Miscanthus* in the first 3 years following planting in Poland. Industrial Crops and Products 27:65–68

Keng YL (1959) Flora illustralis plantarum primarum sinicarum–Gramineae. Beijing: Science Press, 748–755

Key World Energy Statistics (2010) International Energy Administration

Kim C, Zhang D, Auckland SA, Rainville LK, Jakob K, Kronmiller B, Sacks EJ, Deuter M, Paterson AH (2012) SSR-based genetic maps of *Miscanthus sinensis* and *M. sacchariflorus*, and their comparison to sorghum. Theoretical and Applied Genetics 124:1325–1338

Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature 435:95–98

Kuchel H, Fox R, Reinheimer J, Mosionek L, Willey N, Bariana H, Jefferies, S (2007) The successful application of a marker-assisted wheat breeding strategy. Molecular Breeding 20:295–308

Ku MSD, Schmitt MR, Edwards GE (1979) Quantitative determination of RuBP carboxylase-oxygenase protein in leaves of several C_3 and C_4 plants. Journal of Experimental Botany 114:89-98

Lafferty J, Lelley T (1994) Cytogenetic Studies of Different *Miscanthus* Species with Potential for Agricultural Use. Plant Breeding 113:246-249

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

Lee YN (1964) Taxonomic studies on the genus *Miscanthus* (1), new species and varieties. The Journal of Japanese Botany 39: 115–123

Lewandowski I, Clifton-Brown JC, Scurlock JMO, Huisman (2000) *Miscanthus*: European experience with a novel energy crop. Biomass Bioenergy 19:209–277

Lewandowski I, Schmidt U (2006) Nitrogen, energy and land use efficiencies of *Miscanthus*, reed canary grass and triticale as determined by the boundary line approach. Agriculture, Ecosystems and Environment 112:335–346

Li SB, Zhang ZH, Hu Y, Li CY, Jiang X, Mao T, Li YS, Zhu YG (2006) Genetic dissection of developmental behavior of crop growth rate and its relationships with yield and yield related traits in rice. Plant Science 170:911-917

Li H (2011) A quick method to calculate QTL confidence interval. Journal of Genetics 90:355-360

Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjék O, Selbig J, Altmann T, Willmitzer L (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. Plant Science 53:960–972

Liu L (1994) Diandranthus. In: The Comprehensive Scientific Expedition to the Qinghai-Xizang Plateau, Chinese Academy of Sciences, eds. Vascular plants of the Henduan Mountains, Vol. II. Beijing: Science Press, 2296–2298

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J (2003) Control of tillering in rice. Nature 10:618-621

Mangin B, Goffinet B, Rebai A (1994) Constructing confidence intervals for QTL location. Genetics 138:1301–1308

Manichaikul A, Dupuis J, Sen S, Broman KW (2006) Poor Performance of Bootstrap Confidence Intervals for the Location of a Quantitative Trait Locus. Genetics 174:481–489

Martinez O, Curnow RN (1992) Estimating the locations and the size of the effects of quantitative trait loci using flanking markers. Theoretical and Applied Genetics 85:480–488

Ma XF, Jensen E, Alexandrov N, Troukhan M, Zhang M, Thomas-Jones S, Farrar K, Clifton-Brown J, Donnison L, Swaller T, Flavell R (2012) High Resolution Genetic Mapping by Genome Sequencing Reveals Genome Duplication and Tetraploid Genetic Structure of the Diploid *Miscanthus sinensis.* PLoS ONE 7: e33821

McLachlan GJ, Basford KE (1988) Mixture Model: inference and applications to clustering. Marcel Dekker, New York

Miles C, Wayne M (2008) Quantitative trait locus (QTL) analysis. Nature Education 1:1

Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proceeding of the National Academy of Sciences 108:16469–16474

National Research Council (2002) Predicting Invasions of Nonindigenous Plants and Plant Pests. National Academies Press, Washington DC, pp 194

Nocente F, Gazza L, Pasquini M (2007) Evaluation of leaf rust resistance genes Lr1, Lr9, Lr24, Lr47 and their introgression into common wheat cultivars by marker-assisted selection. Euphytica 155:329-336

Petersen KK, Hagberg P, Kristiansen K (2002) In vitro chromosome doubling of *Miscanthus sinensis.* Plant Breeding 121:445-450

Pop M, Salzberg (2008) Bioinformatics challenges of new sequencing technology. Trends in Genetics 24:142-149

Quinn LD, Allen DJ, Stewart JR (2010) Invasiveness potential of *Miscanthus sinensis*: implications for bioenergy production in the United States. Global Change Biology-Bioenergy 2:310-320

Raghu S, Anderson RC, Daehler CC, Davis AS, Wiedenmann RN, Simberloff D, Mack RN (2006) Adding biofuels to the invasive species fire? Science 313:1742

Renvoize SA (2003) The genus *Miscanthus*. The Plantsman New Series 2:207–211

Sacks EJ, Juvik JA, Lin Q, Stewart JR, and Yamada T (2012). The gene pool of Miscanthus species and its improvement. In: A.H. Paterson (ed.). Genomics of the Saccharinae

SAS Institute. 2012. SAS v. 9.2. SAS Inst., Cary, NC

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K (1999) The Lateral Suppressor (Ls) gene of tomato encodes a new member of the VHIID protein family. Proceeding of the National Academy of Sciences 96:290–295

Sen S and Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159:371–387

Shen L, Courtois B, McNally KL, Robin S, Li Z (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theoretical and Applied Genetics 103:75–83

Shimada S, Ogawa T, Kitagawa S, Suzuki T, Ikari C, Shitsukawa N, Abe T, Kawahigashi H, Kikuchi R, Handa H, Murai K (2009) A genetic network of flowering-time genes in wheat leaves, in which an *APETALA1*/*FRUITFULL*-like gene, *VRN1*, is upstream of *FLOWERING LOCUS T*. The Plant Journal 58:668–681

Stewart RJ, Toma Y, Fernández FG, Nishiwakis A, Yamada T, Bollero G (2009) The ecology and agronomy of *Miscanthus sinensis*, a species important to bioenergy crop development, in its native range in Japan: a review. Global Change Biology – Bioenergy 1:126–153

Sun Q , Lin Q, Yi ZL, Yang ZR, Zhou FS (2010) A taxonomic revision of *Miscanthus s*. l. (*Poaceae*) from China. Botanical Journal of the Linnean Society 164:178–220

Swaminathan K, Alabady MS, Varala K, De Paoli E, Ho I, Rokhsar DS, Arumuganathan AK, Ming R, Green PJ, Meyers BC, Moose SP, Hudson ME (2010) Genomic and small RNA sequencing of *Miscanthus x giganteus* shows the utility of sorghum as a reference genome sequence for *Andropogoneae* grasses. Genome Biology 11:R12

Swaminathan K, Chae WB, Mitros T, Varala K, Xie L, Barling A, Glowacka K, Hall M, Jezowski S, Ming R, Hudson M, Juvik JA, Rokhsar DS, Moose SP (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. BMC Genomics 13:142

Titterington DM, Smith AFM, Makov VE (1985) Statistical analysis of finite mixture distributions. Wiley, New York

Tsuji H, Taoka KI, Shimamoto K (2010) Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. Current Opinion Plant Biology 14:1-8

Van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theoretical and Applied Genetics 84:803–811

Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2000) MapQTL™ version 4.0: software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, the Netherlands

Van Ooijen JW (2011) JoinMap ® 4.1 Software for the calculation of genetic linkage maps in experimental population. Kyazma BV, Wageningen, Netherlands

Van Ooijen JW (2004) MapQTL 5™, Software for the Mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, the Netherlands

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020

West T, Dunphy-Guzman K, Sun A, Malczunski L, Reichmuth D, Larson R, Ellison J, Taylor R, Tidwell V, Klebanoff L, Hough P, Lutz A, Shaddix C, Brinkman N, Wheeler C, O'Toole D (2009) Feasibility, economics, and environmental impact of producing 90 billion gallons of ethanol per year by 2030. Sandia National Laboratories, Livermore, CA

White LM (1973) Carbohydrate reserves of grasses: a review. Journal of Range Management 26:13-18

World Energy Outlook (2008) International Energy Agency

Xu CG, Li XQ, Xue Y, Huang YW, Gao J, Xing YZ (2004) Comparison of quantitative trait loci controlling seedling characteristics at two seedling stages using rice recombinant inbred lines. Theoretical and Applied Genetics 109:640–647

Xu J, Liu Y, Liu J, Cao M, Wang J, Lan H, Xu Y, Lu Y, Pan G, Rong T (2012) The Genetic Architecture of Flowering Time and Photoperiod Sensitivity in Maize as Revealed by QTL Review and Meta Analysis. Journal of Integrative Plant Biology 54:358–373

Yu CY, Kim HS, Rayburn AL, Widholm JM, Juvik JA (2010) Chromosome doubling of the bioenergy crop *Miscanthus x giganteus*. Global Change Biology-Bioenergy 1:404-412

Zeng ZB (1993) Theoretical basis of precision mapping of quantitative trait loci. Proceedings of the National Academy of Science USA 90:10972–10976

Zhang J, Li X, Jiang G, Xu Y, He Y (2006) Pyramiding of Xa7 and Xa21 for the improvement of disease resistance to bacterial blight in hybrid rice. Plant Breeding 125:600–605

Appendix

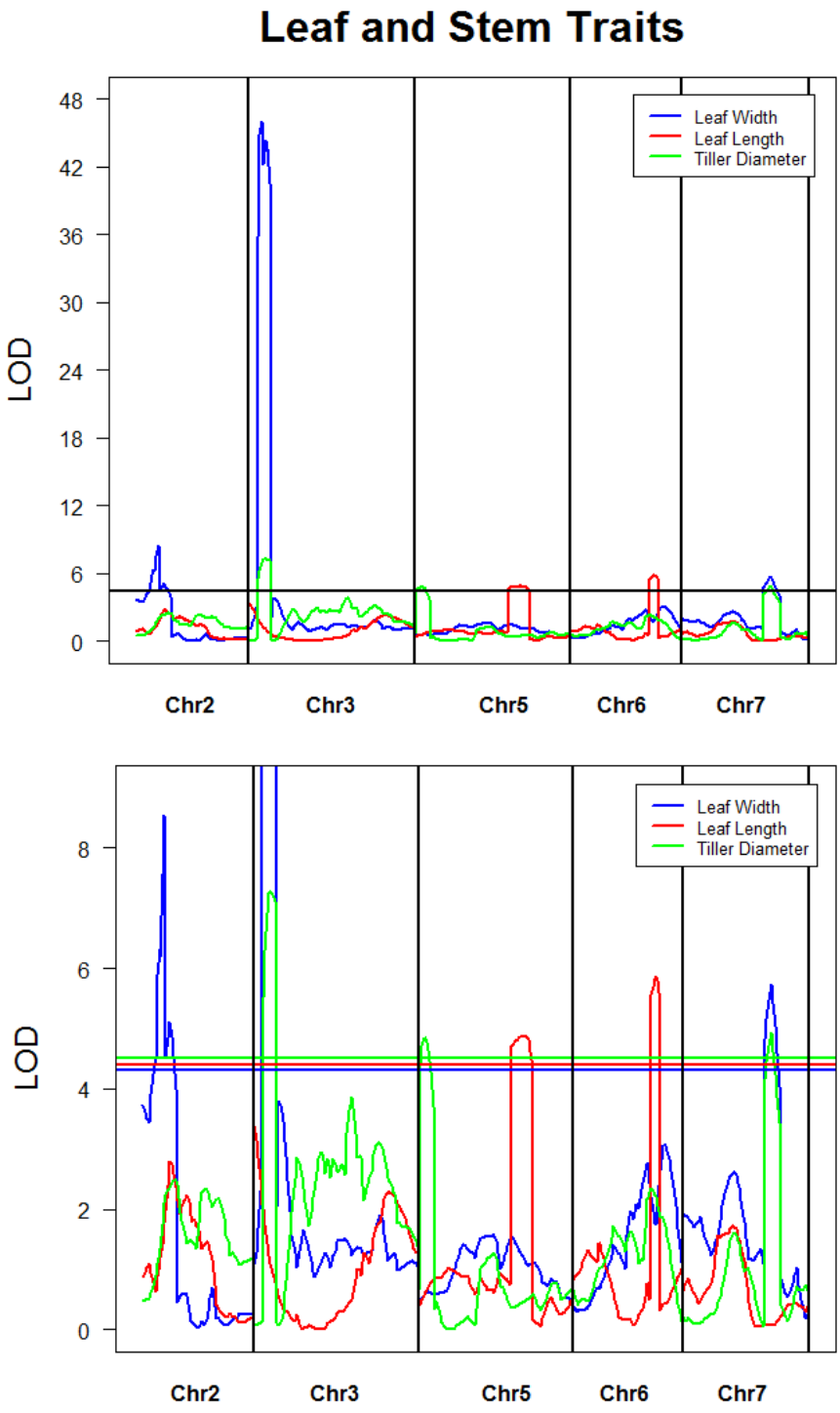


Figure A1. P-value plots for leaf width, leaf length, and tiller diameter showing only those chromosomes on which QTL are located. The bottom figure is a close up view that shows a more detailed picture. Horizontal lines represent the significance level for each trait.

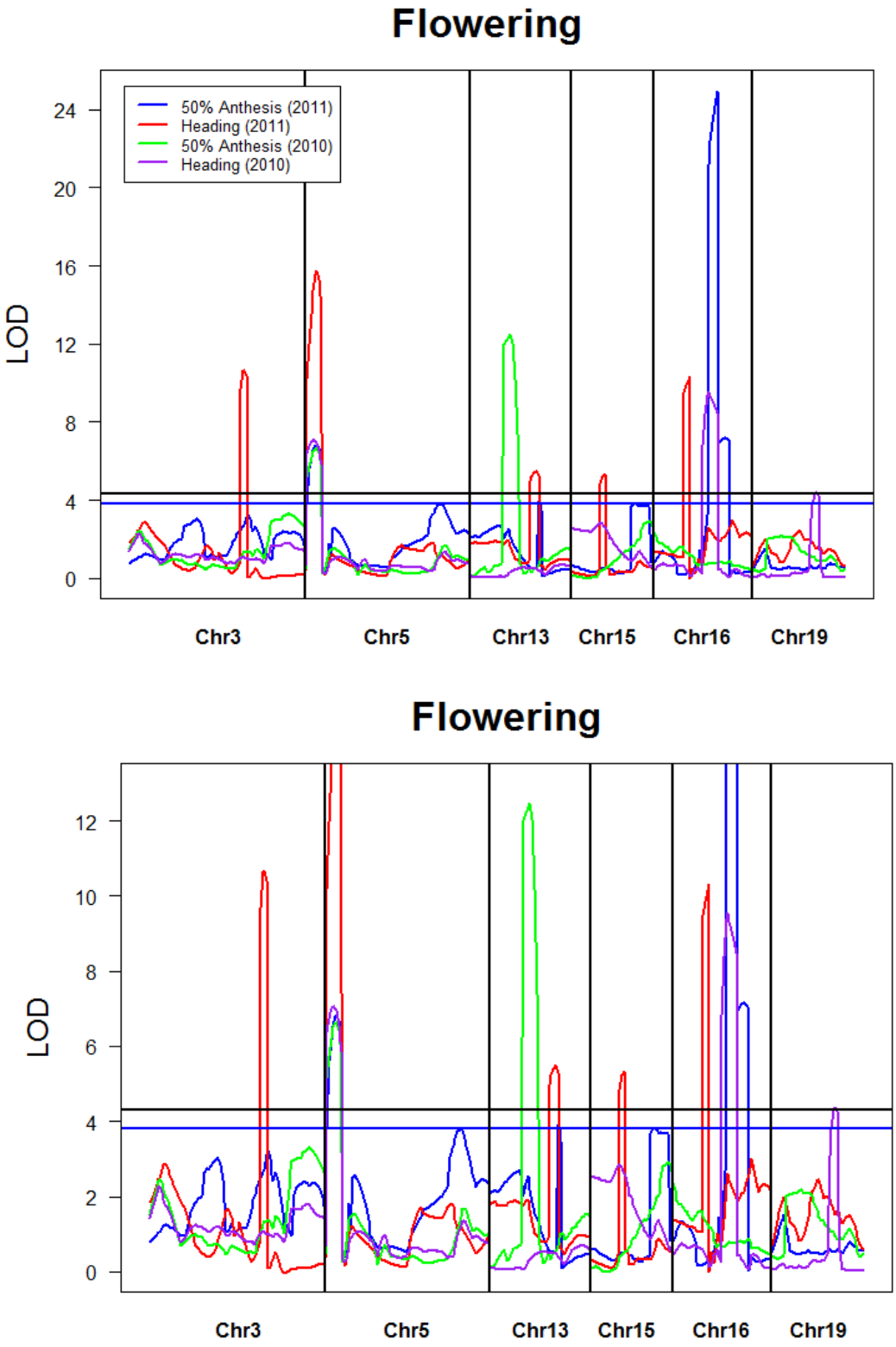


Figure A2. P-value plots for flowering traits showing only those chromosomes on which QTL are located. The bottom figure is a close up view that shows a more detailed picture. Horizontal lines represent the significance level for each trait.

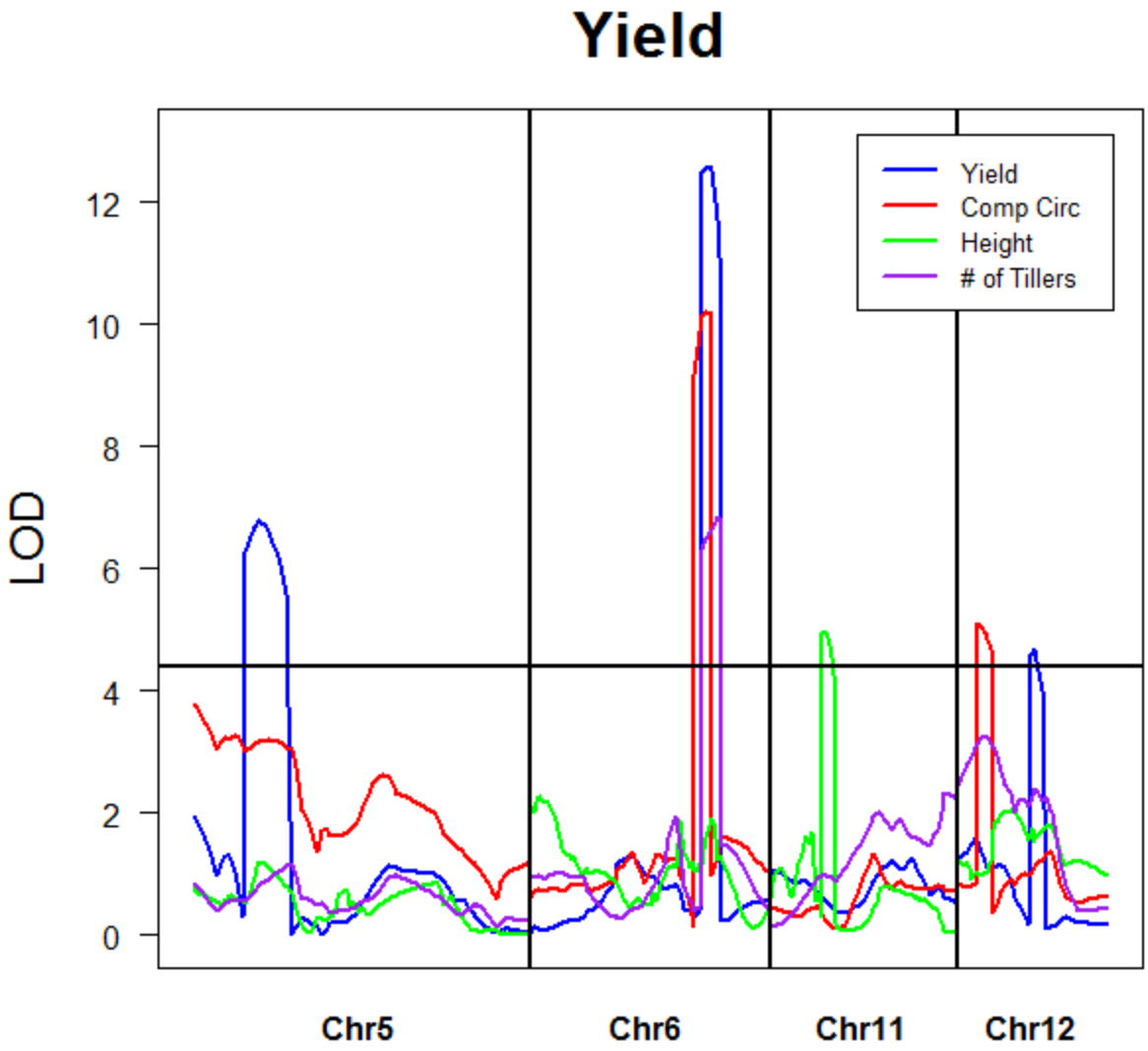


Figure A3. P-value plots for yield and its components showing only those chromosomes on which QTL are located. The horizontal line represents the significance level which happened to be the same for each trait.