GENETIC CONTROL OF FLOWERING TIME AND BIOMASS YIELD IN SORGHUM

BY

ELIZABETH MARIE HAWKINS

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, 2013

Urbana, Illinois

Adviser:

Assistant Professor Patrick J. Brown

ABSTRACT

Sorghum is a C_4 grass grown for grain, forage, sugar, and most recently biomass, though there has been very little genetic improvement for biomass yield. Most biomass sorghums are very late flowering to maximize vegetative growth. In order to understand the genetic control of flowering time and biomass yield in sorghum, we used a diverse panel of more than 400 exotic sorghum inbreds to perform association mapping for flowering time, biomass yield, and yield component traits. We also examined correlations between traits, correlations between inbred and hybrid performance, and the relationship between traits of interest and genetic diversity. Significant marker-trait associations were detected for maturity, plant height, and lodging. Forty percent of the variance in biomass yield is explained by plant height and lodging, inbred yield explained 82% of the variance in hybrid yield, and inbreds from all genetic subpopulations were represented in the top 5% of yield entries over two years. To better understand the reported epistatic interaction between two major flowering time loci, Ma5 and Ma6, we also performed linkage mapping in two biparental populations (Tx623 x Tx2909 and Tx623 x Tx2910) thought to be segregating for functional variation at both Ma5 and Ma6. Linkage mapping results suggest that the architecture of the photoperiod-sensitivity response differs between these two populations. Overall, our results suggest there is abundant genetic variation to quickly improve sorghum biomass yields by incorporating novel alleles from exotic sorghum.

TABLE OF CONTENTS

INTRODUCTION

Finding a sustainable and affordable renewable fuel to supplement petroleum fuels has been discussed for decades. Corn ethanol and soy biodiesel have worked their way into the mainstream fuel market in the United States, but production costs are high and the industry is subsidized to keep biofuels competitive with traditional petroleum fuels. Also, there is concern that using food crops grown on prime acres for bioenergy production will drive up food costs. Both of these factors have prompted the search for alternative sources of bioenergy based on dedicated biomass crops. In 2011, Congress enacted the US Renewable Fuels Standard (RFS), which mandates that by 2022, 36 billion gallons of fuel must be biofuels, and 21 billion gallons must be in the form of advanced biofuels. Advanced or second-generation biofuels are defined by the EPA as a renewable fuel other than ethanol produced from corn starch, that is derived from renewable biomass. This includes cellulosic biofuels and well as biomass-based diesel. In December 2012, it was determined that sorghum qualifies as an advanced biofuel if the starch is used for ethanol and the stover is gasified or combusted (www.epa.gov).

Sorghum is a crop traditionally grown in drier regions of the United States because of its ability to produce yields under drought and heat stress (J. R. Quinby, 1974). It is grown for grain, forage, sugar, and most recently for biomass. However, there has been very little genetic improvement for biomass production. Many farmers already produce sorghum and many others are familiar with the agronomic practices and equipment required to cultivate it, since these are very similar to the practices and equipment required for corn. This makes sorghum an attractive, practical option as an easy-to-adopt biomass crop.

Sorghum is a short-day plant that originated in tropical areas of Africa (J. R. Quinby, 1974). Most exotic accessions of sorghum are photoperiod-sensitive, meaning they flower very late or not at all at temperate latitudes. Photoperiod-sensitivity maximizes total biomass production by using the entire growing season for vegetative growth. The National Center for Genetic Resources Preservation (NCGRP) maintains over 40,000 accessions of sorghum. Most of these accessions (>75%) are photoperiod-sensitive and not suitable for grain or sugar production, but they represent a vast untapped resource for the genetic improvement of biomass sorghum. In sum, efforts to improve biomass sorghum are just beginning, and there is tremendous genetic diversity available. Together, these facts suggest that very rapid genetic improvement of biomass sorghum is feasible.

Sorghum can be classified into five races --caudatum, kafir, durra, guinea, and bicolor-- based on phenotypic characteristics and area of origin. Phenotypic classification relies chiefly on the appearance of the spikelet and panicle (Harlan and de Wet, 1972). Caudatums originated in eastern Africa and are commonly used for grain sorghum breeding. Kafirs were collected from southern Africa and have also been used in grain sorghum improvement. Durras originated in India and the Horn of Africa, and they are characterized by a compact panicle and are typically drought tolerant. Guineas originated in western Africa, perform well in moist climates and usually have a loose, open panicle structure. Bicolors are considered to represent the primitive progenitor race (Brown et al., 2011).

In photoperiod-sensitive (PS) sorghum, the flowering date remains the same regardless of planting date. The initiation of reproductive growth in PS sorghum requires a reduction in day

length rather than an accumulation of growing degree days (GDDs; Jones et al., 2013). Similar photoperiod-sensitivity responses have been observed in other species of grasses including maize, wheat, rice, and barley. Long-day delay of flowering in PS sorghum is only attenuated when the day length reaches the critical point of 12 hr 20 min or shorter, when floral initiation begins to occur (Smith and Frederiksen, 2000).

The genetic control of photoperiod sensitivity in sorghum was first attributed to a single major maturity locus called *Ma1* (J. R. Quinby, 1974). In an interspecific cross between *S. bicolor* (photoperiod insensitive; PI) and *S. propinquum* (PS), *Ma1* explained ~85% of the variation in flowering time and had an additive effect of over 40 days, meaning that the two homozygotes flowered nearly three months apart. An individual with a dominant *Ma1* allele will exhibit a photoperiod response, while an individual with homozygous for recessive *ma1* will flower early regardless of daylength. In 2011, *Ma1* was cloned and identified as pseudoresponse regulator protein 37 (PRR37; Murphy et al., 2011). Natural variation in the rice ortholog of this gene, *OsPRR37*, was subsequently shown to affect flowering time in a diverse panel of Asian and European rice accessions (Koo et al., 2013). PRR genes are distantly related to the CONSTANS (CO) gene family from Arabidopsis; both gene families contain C-terminal CCT domains and are regulated by both light and the circadian clock (Cockram et al., 2012).

Ma3 was the first maturity locus to be cloned in sorghum, and encodes a phytochrome B (phyB; Childs et al., 1997). PhyB is a red/ far-red light receptor and is necessary for the plant to regulate flowering time through light perception. Sorghum plants with the complete loss-of-function *ma3R* allele of *Ma3* are insensitive to photoperiod and flower early regardless of day length and

Ma1 allele. The epistatic interaction between *Ma1* and *Ma3* can be exploited to generate a late-flowering sorghum hybrid *(Ma1/ma1; Ma3/ma3R)* from two early flowering inbred parents *(ma1/ma1; Ma3/Ma3* and *Ma1/Ma1;ma3R/ma3R;* Childs et al., 1995; Brown lab unpublished data).

Several other maturity loci exist, including *Ma2, Ma4, Ma5,* and *Ma6*, but little is known about their function. *Ma4* is thought to be on chromosome 10 (Mace and Jordan, 2010), and the early, recessive allele is known only from the sorghum variety Hegari (Quinby, 1966). *Ma2* is unmapped and shows a complex interaction with *Ma1* (Brown and Paterson, 2013). The dominant *Ma2* allele causes lateness in a *Ma1* background and earliness in a *ma1* background. Also, overdominant late flowering of *Ma1/ma1* heterozygotes is observed in *ma2* but not *Ma2* backgrounds (J. R. Quinby, 1974).

Allelic variation at the recently reported *Ma5* and *Ma6* loci affects photoperiod sensitivity. Rooney and Aydin first reported a very late flowering hybrid produced from a cross between two early flowering inbreds (Aydin and Rooney, 1999). One of the parental inbreds (EBA-3, a grain sorghum line from Argentina) reproducibly produced late flowering hybrids when crossed with other inbred lines. This phenomenon was hypothesized to result from a complementary dominant interaction between two new loci, *Ma5* and *Ma6*, in which at least one dominant allele of each locus was necessary to confer photoperiod-sensitivity. In this case, EBA-3 is *ma5ma5/Ma6Ma6* and most other inbreds are *Ma5Ma5/ma6ma6,* and both are PI. The resulting hybrids are *Ma5ma5/Ma6ma6* and are PS. This interaction is conceptually similar to the interaction between *Ma1* and *Ma3*, but in practical terms the *Ma5-Ma6* interaction appears to be more useful as it

produces hybrids that are more strongly PS and later flowering, with higher biomass yields There are only two publically available lines that incorporate the *ma5*/*Ma6* allele combination from EBA-3: Tx2909 and Tx2910.

The *Ma5-Ma6* interaction is widely used by the biomass sorghum seed industry to create PS hybrids from two PI parents. The system saves money because it avoids the need to make crosses in tropical winter nurseries. Both Tx2909 and Tx2910 are male lines, meaning that they are restorers of fertility in A1 cytoplasm. Creating publically available female lines with the desired *ma5ma5*/*Ma6Ma6* genotype and traits favorable for biomass production would greatly advance the feedstock industry and make production of PS hybrid seed more economical. Recently, an additional maturity locus *(Ma7)* has been reported. *Ma7* is thought to play a role in the photoperiod sensitivity response by interacting with *Ma5* and/or *Ma6* (Mullet et al., 2010). However, the contribution of *Ma7* to the biomass sorghum hybrid seed industry is not clear.

Temperature also affects flowering time in plants. High temperatures promote flowering by accelerating the accumulation of GDDs. Conversely, in cool season grasses and many other plants, promotion of flowering by low temperatures is known as vernalization (Jones et al., 2013). Warm-season grasses like sorghum do not display vernalization, but an interaction between the photoperiod sensitivity response and low temperatures has been reported in sorghum. Growth chamber experiments suggest that when the minimum night temperature falls below 20°C, early flowering replaces the normal photoperiod sensitivity response in sorghum lines carrying a dominant Thermosensitive (T) allele (Tarumoto et al., 2003). Further growth chamber experiments confirmed that the T locus lengthened the critical day length for flowering induction

to13’ from 12’20” in a recessive tt background at any temperature or a T/- background when exposed to temperatures below 20C (Tarumoto et al., 2005).

Linkage mapping and association mapping can both be used to identify quantitative trait loci (QTL) by genotyping and phenotyping segregating populations (Flint-Garcia et al., 2003). Linkage mapping uses populations created from a small number of parents (usually two) that differ for the trait of interest. Linkage mapping requires few markers, due to high linkage disequilibrium (LD), but has low resolution. Only a small number of parental haplotypes are evaluated (two in a biparental cross), but these haplotypes are at high frequency, increasing the power to detect phenotypic associations. In contrast, association mapping uses a large number of diverse lines with low levels of LD, and requires a large number of markers to conduct a genome-wide scan. Major benefits of association mapping include high resolution and the ability to evaluate multiple haplotypes, but this is balanced by lower power to detect low-frequency QTL. Another major limitation of association mapping is genetic heterogeneity. This occurs when a single phenotype can result from multiple different mutations. This can occur either through multiple mutations in the same gene, or through mutations in different genes that cause the same phenotype. Genetic heterogeneity can cause synthetic associations, in which the mutation underlying the phenotype is incorrectly identified as a single large effect QTL. For example, the sorghum shattering locus *Sh1* harbors three independent loss-of-function alleles, and association mapping for shattering at the *Sh1* locus identifies a single major synthetic association that is more significant than any of the individual loss-of-function mutations (Lin et al., 2012).

In order to understand the genetic control of biomass yield in sorghum, we: i) study trait correlations between biomass yield, maturity, plant height, and other plant architectural traits in a diverse panel of more than 400 exotic sorghum inbreds; ii) create and evaluate the performance of ~150 exotic sorghum hybrids to assess the contribution of additive and non-additive genetic effects; iii) examine the relationship between our traits of interest and genetic diversity using principal component analysis, and iv) perform association mapping for flowering time and biomass yield in a diverse panel of exotic sorghum accessions. In an effort to better understand the interaction between *Ma5* and *Ma6*, we also use linkage mapping to identify maturity QTL in two biparental populations (Tx623 x Tx2909 and Tx623 x Tx2910) thought to be segregating for functional variation at both *Ma5* and *Ma6*.

MATERIALS AND METHODS

Germplasm development for association mapping

In 2011, 196 exotic sorghum accessions, including 183 exotic progenitors of the sorghum conversion program, were grown and phenotyped for maturity in three Illinois locations. Photoperiod-sensitive accessions were selected and planted in the 2011 winter nursery in Mexico, where they were bulked and used to create hybrids with ATx623, a cytoplasmic male sterile "female" version of the inbred used for sequencing the sorghum genome. In 2012, 187 inbred accessions and 172 hybrids with ATx623 were grown and phenotyped for maturity, yield, and yield component traits in Urbana with two replications. 369 new accessions acquired from GRIN were also phenotyped for maturity and photoperiod sensitive new accessions from this set were again selected to include in future trials. In 2013, 437 inbred lines selected on the basis of maturity, lodging, and yield were grown and phenotyped for maturity, yield, and yield component traits. The 2012 trial included four commercial hybrids: Pacesetter and Pacesetter-BMR (MMR Genetics), CHR-FS9 (Chromatin, Inc.), and ES5200 (Ceres), whereas the 2013 trial included only Pacesetter.

Germplasm development for linkage mapping

Two initial crosses, BTx623-ms3 x Tx2909 and BTx623-ms3 x Tx2910, were made in 2011. The nuclear-encoded *ms3* locus provides recessive genetic male sterility for breeding and genetics experiments. F_1 progeny of these crosses were grown in the 2011 winter nursery in Mexico and selfed to create F_2s. In 2012, F_1 and F_2 plants were grown. A single F_1 row from each cross was regrown in order to confirm that the hybrids were photoperiod sensitive in a temperate environment. Ten rows of F_2 individuals from each population

were also grown, and PI F_2 individuals were selfed. A single F_3 seed from each selfed F_2 panicle was sent to the 2012 winter nursery along with additional F_2 seed. These F_2s and F_3s were selfed to create F_3s and F_4s. In 2013, the resulting F_3 (segregating for photoperiod sensitivity) and F_4 (theoretically fixed for photoperiod-insensitivity) populations were grown and phenotyped.

Field design for association mapping

Field tests were conducted over three years. In 2011, two to four replications were grown at three locations in Urbana, Perry, and Dixon Springs, Illinois in single-row plots with a row length of 6 m (20'), 1.5 m (5') alleys, 76 cm (30") row spacing, and a target density of 123,550 plants/ha (50,000 plants/acre). In 2012, two replications were grown in Urbana, Illinois in single row plots 7.6 m (25') in length with 1.5 m alleys and 76 cm (30") row spacing. The target density was approximately 207,570 plants/ha (84,000 plants/acre). In 2013, a single trial was grown in Urbana, Illinois in an incomplete block design with a commercial hybrid replicated in each block. Each genotype was planted in a four-row plot, 3 m (10') long with 1.5 m alleys and 76 cm (30") row spacing. The target density was 207,570 plants/ha (84,000 plants/acre), however due to a combination of planting error and climatic factors the actual density was approximately 88,960 plants/ha (36,000 plants/acre).

Field design for linkage mapping

In 2013, 95 F_3 families were grown without replication at three locations; Urbana, (Energy Farm), Dixon Springs, and Monmouth, IL. Forty-six F_4 families derived from photoperiod-

insensitive F_2 selections the previous summer were grown in the Urbana location only. Each genotype was planted in single row plots measuring 4.6 m (15') with 1.5 m (5') alleys and 76 cm (30") row spacing. The target density was 143,320 plants/ha (58,000 plants/acre).

Phenotyping for association mapping

Only maturity phenotypes were collected in 2011. All other phenotypes were measured in both 2012 and 2013 except for stalk circumference, which was measured in 2012 only. Stand counts were taken to verify the population of each row. The 30- day growth rate of each genotype was collected by measuring the height of the plants from the ground to the top of the whorl every thirty days after planting. A total of four height measurements were collected. For the 3rd (90 day) and 4th (120 day) height measurements, plants that had initiated flowering were measured to the top of the panicle rather than the top of the whorl. Stalk circumference (cm) was measured 30 cm above the base of the plant, on 3 random plants per row. Leaf length and width (mm) were measured by taking a representative leaf from a plant in the middle of the plot. Maturity was measured prior to harvest using a 0-4 maturity scale (Table 1). A simple "days to anthesis" maturity measurement was not feasible because most entries in any given season never reached anthesis. In 2013, a three-level nitrogen status rating was assigned to each plot during the October harvest: "green", "green-yellow", or "yellow." Total plot wet weight (kg) was collected by machine harvest with a John Deere forage harvester. The sample was chopped and then weighed in a forage wagon modified with a load cell accurate to within 1 kg. A subsample from each plot was taken, then weighed before and after oven drying at 60°C for 72' to determine the moisture

content using the following calculation: Plot moisture = (subsample wet weight- subsample dry weight)/subsample wet weight. The yield in dry metric tons per hectare was calculated as follows: Dry metric tons/ ha = (total plot wet weight (kg) * (1-plot moisture (%))/1000)/(plot area (m^2)/10,000).

Phenotyping for linkage mapping

Phenotypes related to maturity were collected at all locations. Flowering was measured by scoring the percentage of plants that had reached anthesis within each row. To capture the variation throughout the season, this phenotype was collected every 1-3 weeks beginning in July and ending in early October. These phenotypes were used to calculate the "area under the flowering progress curve" (AUFPC; similar in principle to the area under the disease progress curve) for each row. A distinct short, highly-branched "tillery" phenotype was also observed, and was scored as a qualitative trait on a three-point scale: rows where all plants were short and tillery were scored as a 1, rows where all plants were phenotypically normal were scored as a 0, and rows that were segregating for this trait were scored as 0.5. Due to concerns that tillery individuals might suffer elevated early-season mortality, a "plant stand" phenotype was also scored as the percentage of each row with adult plants.

Genotyping and data analysis for principal component analysis and association mapping

DNA was extracted from dark-grown, etiolated seedling tissue in 96-well plates using a modified CTAB protocol. Illumina libraries were created using two pairs of restriction enzymes (*PstI-HF/HinP1I* and *PstI-HF/BfaI*). Restriction-ligation reactions were performed in 96-well plates, and unique barcoded adapters were ligated to each DNA sample. All 96

DNA samples per library were then pooled into a single tube for all subsequent steps including bead cleaning, PCR amplification, and a second round of bead cleaning. Single-end, 100 bp sequencing reads were obtained for all libraries on the Illumina HiSeq2000 instrument following submission to the Keck Center at the University of Illinois. The TASSEL GBS pipeline (www.maizegenetics.net) was then used to identify SNPs. Only reads that perfectly matched a barcode and restriction site overhang were retained. After barcode trimming, a set "master tags" was generated from the set of unique 64 bp sequences that were present at least ten times in the dataset and that mapped uniquely to the sorghum genome. SNPs were called in each individual by comparing the tags in that individual to the set of master tags. SNPs and individuals with more than 95% missing data were discarded. Missing data were imputed in two steps using first the imputation plugin implemented in the TASSEL GBS pipeline, and then BEAGLE (Browning and Browning, 2011). R (www.cran.org), an open source statistical software, was used for all data analyses. Best linear unbiased predictors (BLUPs) for each phenotype were calculated using the lme4 function. In unbalanced experiments where entries are not equally replicated within and between environments, BLUPs adjust the values of measurements with fewer replications towards the mean. BLUPs were created for each phenotype in the 12EF location, the 13EF location, and both 12EF and 13EF locations combined before other analyses were run. The R function prcomp was used to perform principal component analysis (PCA) on all genotypes using all SNPs with minor allele frequencies greater than 10%. The R package GAPIT (Lipka et al., 2012) was used to perform association mapping for each trait. We discarded SNPs with minor allele frequencies less than 10%, and also

trimmed the dataset for adjacent SNPs with identical genotypes, before association mapping with a false discovery rate (FDR) threshold of 0.10.

Genotyping and data analysis for linkage mapping

Genotyping-by-sequencing was conducted as described above, with several changes. 192 entries were genotyped per lane instead of 96, to reflect the reduced number of haplotypes in biparental populations. SNP imputation in the biparental populations was performed using a set of undocumented TASSEL plugins provided by Peter Bradbury (USDA-ARS/Cornell). Dr. Bradbury's method is designed to address the problem of undercalling heterozygous genotypes due to limited sampling, and uses the Viterbi algorithm and user-inputted inbreeding coefficients to estimate the most likely crossover positions in biparental populations. BLUPs were created for all phenotypes as described above, and QTL were identified using the r/qtl package in R. Significance thresholds determined by permutation using an α of 0.05.

RESULTS AND DISCUSSION

Correlations between traits, environments, and inbreds and hybrids

In 2011, trials were grown at 3 locations across Illinois (Urbana, 40.1 N; Perry, 39.8 N; Dixon Springs 37.4 N.) These trials were phenotyped using a 0-4 maturity rating at the end of the growing season. Because minimal differences in maturity were observed between the locations, trials were only grown at a single location in the following years.

Correlations between all traits were calculated for 2012 inbreds, 2012 hybrids, 2013 inbreds, and inbreds in both years combined (Figures 1-4). In general, the r^2 values for trait correlations in 2012 are lower than the r^2 values for 2013. This suggests that the 4-row plots in 2013 provided a better estimate of trait values than the single-row plots in 2012. This is especially true for traits such as biomass yield and lodging that are strongly influenced by neighbor effects.

Plant height shows the strongest correlation with biomass yield. The 90- day height measurement has the highest correlation with biomass yield in 2013 and in both years combined, whereas the 120- day height measurement has the highest correlation with biomass yield in 2012 inbreds and hybrids. This discrepancy is likely due to differences in maturity between entries and years. If all entries were completely photoperiod-sensitive (vegetative), the correlation between biomass yield and height would likely be much higher, since initiation of flowering causes plants to become much taller without accumulating a proportionate amount of biomass.

Maturity also correlates with height. The strength of this correlation depends on the number of short, early entries included in each trial. The correlation between 90-day height and maturity is higher in 2013 than in 2012. This could be caused by the lines that flowered earlier in 2013 than in 2012 as a result of environmental conditions.

The traits most highly correlated with yield were used to build models to explain yield variance. In 2012, lodging explains more of the yield variance than height. This is probably because of later harvest date: plots were harvested between October 12th and November 7th in 2012 and between October 5th and 9th in 2013. An additional possible explanation is that entries with excessive lodging in 2012 were not planted in 2013. In 2012, 120- day height and lodging explained 36% of the variance in yield, and in 2013, 90- day height explained 40% of the variance in yield and lodging was not significant. When both years are combined, 90- day height and lodging explain 39% of the variance. Approximately the same percentage of the variance in biomass yield (40%) is explained by a combined height/lodging model in each year separately and over both years combined. In the hybrids, 34% of the variance is explained by 120- day height and lodging. This is most likely because there was less variance in height within the hybrid lines, which all shared a common female parent.

Maturity was not correlated with biomass yield. This was likely the result of strong selection for photoperiod-sensitivity in 2011, so that nearly all the lines tested in 2012 and 2013 were photoperiod sensitive. If early flowering, photoperiod-insensitive lines had been included in the study, the correlation between yield and maturity would likely have been very strong.

Nitrogen status in relation to biomass yield

In 2013, nitrogen stress negatively affected yield. Of the twenty check plots, fourteen were scored as "green," three as "green-yellow," and three as "yellow." When nitrogen status of the check plots was compared to their biomass yield, rows rated as yellow or green-yellow yielded much lower on average than rows rated green, and nitrogen status explained 38% of the check plot variance in biomass yield. Nitrogen stress was not evenly distributed throughout the field, but was localized to the low-lying areas (Figure 5). This suggests that the severe early season flooding in 2013 contributed strongly to nitrogen stress, and that the effects of this early season stress persisted throughout the season and were detrimental to biomass yield.

Differences between inbreds and hybrids in 2012

The mean values for maturity, height, biomass yield, and lodging were compared between the hybrids and the inbreds tested in 2012. In general, the hybrid lines were earlier and taller than their exotic inbred parents (Figures 6 and 7). ATx623, the common parent for the hybrids, is recessive for three of the four major dwarfing mutations *(dw1, dw3,* and *dw4)* and is dominant for *Dw2.* As most of the hybrids were not dwarfed, this suggests that most exotic parents are dominant at *Dw1, Dw3,* and *Dw4.* Hybrids yielded more on average than their inbred parents, though some inbred parents out yielded their hybrids (Figure 8). Lodging was not consistently higher in hybrids or inbreds, but explained 25% of the variance in the yield difference between inbreds and hybrids (Figure 9). In other words, if an inbred lodged less than its corresponding hybrid, then it tended to yield more, and vice

versa. Because single-row plots were used in 2012, many of these differences in lodging are probably non-genetic and due to neighbor effects.

Ma1 is usually considered to be completely dominant, so that a *Ma1/ma1* hybrid will flower in synchrony with the *Ma1/Ma1* parent (J. R. Quinby, 1974). However, *Ma1/ma1* hybrids in this study usually showed earlier maturity than their corresponding *Ma1/Ma1* inbreds, suggesting that *Ma1* is not functioning as a completely dominant allele. One possible explanation is that there is an allelic series at *Ma1* with a range of additive and dominance effects. Most classical studies of *Ma1* inheritance (J. R. Quinby, 1974) compared a "wild type" *Ma1* allele *(SbPRR37)* with the *Sbprr37-1* allele (Murphy et al., 2011), which has a frameshift mutation upstream of the pseudoreceiver domain. The common female parent used for hybrids in this study, ATx623, carries the *Sbprr37-3* allele, which contains a premature stop codon downstream of the pseudoreceiver domain. Therefore, it is possible that the *Sbprr37-3* allele is not completely recessive, in contrast to the *Sbprr37-1* allele. One line (PRE0210) was also unique in this study in that the hybrid flowered significantly later than both the inbred parents. This suggests a possible *Ma5/Ma6* interaction in the PRE0210 hybrid similar to that observed for EBA-3 hybrids with elite grain sorghum.

Inbred yields combined over 2012 and 2013 explained 82% of the variance in hybrid yields from 2012 (Figure 10). This surprisingly high correlation could result in part from the large range in yield across the diverse exotic accessions that were evaluated. This would cause the correlation to be higher than if only elite biomass inbreds and hybrids were evaluated. Also, hybrids in this experiment were only generated from a single female tester. However,

our initial results suggest that genetic gains can be made in biomass sorghum hybrids based on evaluation of their inbred parents, which would make breeding for high yielding hybrids more efficient.

Difference in maturity between 2012 and 2013

Mean maturity ratings for inbreds in 2012 were compared to the their ratings in 2013. A subset of inbreds flowered considerably earlier in 2013 than in 2012 (Figure 11). We hypothesize that unseasonably cool nights during the 2013 growing season lengthened the critical daylength required for induction of flowering in inbred lines with a dominant T (Thermosensitive) allele (Tarumoto et al., 2005). However, this hypothesis requires validation in a controlled environment.

Relationships between traits and genetic structure

Principal component analysis (PCA) was performed to examine the relationship between traits of interest and population structure. The first two principal components separate the major sorghum races included in this experiment: PC1 separates the durras from the guineas, and PC2 separates the caudatums from the durras and guineas. The inset plot shows that PC3 splits the durras into two subgroups (Figure 12).

Superimposition of biomass yield values onto the PCA plot shows that the highest-yielding lines are distributed across all major races, with durras, guineas, and caudatums all represented in the top 5% of biomass yield values (Figure 13). This suggests that there is vast potential to stack beneficial alleles across races of sorghum. Superimposition of maturity values shows that caudatums tend to be early flowering, and durras are only

moderately photoperiod-sensitive whereas guineas show the strongest photoperiod response (Figure 14). For the lodging severity trait, caudatums represent a large proportion of the most lodging-resistant lines. PC3 shows that one subpopulation of durras lodges severely while the other does not (Figure 15). Durras are the tallest of the subspecies, representing the majority of the tallest 5% of the lines tested (Figure 16). Superimpositions of additional phenotypes onto the PCA plot (30-day, 60-day, and 120-day height, leaf length and width, and moisture) are shown in the Appendix in Figures 27-32. These results can be used to create a representative ideotype for each class. In general, durras are tall, susceptible to lodging, and moderately photoperiod sensitive with short leaves. Caudatums are moderately tall, lodging-resistant, with a relatively weak photoperiod response (earlier maturity) and wide leaves. The Guineas are short and relatively lodging resistant with the strongest photoperiod sensitivity response (latest maturity) and long, narrow leaves.

Association mapping for maturity in grain and biomass sorghum

In order to better understand the QTL contributing to maturity in exotic sorghum, we compared association results for maturity in three different panels of sorghum; the biomass sorghum lines evaluated in this study, a set of grain sorghum lines evaluated in a sister study, and both biomass and grain types combined. GAPIT, a statistical package available in R, was used to test for phenotype-genotype associations using a false positive rate (**α)** of 0.05 and all the default parameters. A single significant maturity QTL located at the end of chromosome 9 is detected in grain sorghum lines (Figure 17). This maturity QTL maps close to the uncloned *Dw1* locus, but appears to be a separate QTL based on multiple

independent lines of evidence including low linkage disequilibrium between SNPs associated with height and maturity in the Dw1 region (Thurber et al., 2013) and separate incidence of height and maturity QTL in this region in different biparental populations (R. Higgins, personal communication).

When just biomass lines are analyzed, a single QTL at the beginning chromosome 6 is observed (Figure 18). The closest gene to the most significant SNP is this region is the sorghum ortholog of *Ghd7*, which is responsible for natural variation in photoperiodic flowering in both rice (Liu et al., 2013) and maize (Hung et al., 2012). A search of the *SbGhd7* gene in PubMed (www.ncbi.nlm.nih.gov) revealed that SbGhd7 has been identified as *Ma6* (J. Mullet, Texas A&M University, unpublished). It is likely that the accompanying manuscript was still under peer review when this thesis was completed. When both grain and biomass sorghum lines are analyzed together, three significant QTL for maturity are present (Figure 19). Both QTL regions detected in the individual populations are present again. The same SNP near *SbGhd7/Ma6* is again detected in the combined analysis, but the chromosome 9 QTL has moved 2 Mb towards the centromere and may represent a different QTL. The third QTL that is mapped is near the known location of *Ma1*. The lack of a *Ma1* QTL in grain sorghum and biomass sorghum panels individually suggests that both panels lack functional variation at *Ma1*, with recessive *ma1* fixed in grain sorghum, and functional *Ma1* fixed or nearly fixed in biomass sorghum. When these populations are combined the diversity between these two types of sorghum allow the *Ma1* QTL to be detected.

Association mapping of yield and yield component traits

Several traits, including biomass yield, leaf length, and leaf width had no significant associations. Possible reasons for this include the highly complex, quantitative nature of the trait itself, the small number of environments tested, and the relatively small number of accessions evaluated (187 and 437 in 2012 and 2013 respectively). 90- day and 120- day height analyses show a barely significant association on the end on chromosome 5 (Figure 20; the dashed line indicates a Bonferonni correction with an **α** of 0.01). No plant height QTL have previously been reported in this region of the sorghum genome, so this QTL may only have significant effects in a photoperiod-sensitive background. This same genomic location showed a smaller peak in 60- day height that did not exceed the significance threshold. A significant association with lodging was detected on the short arm of chromosome 9 in the 13EF environment, which used four-row plots, but not in the 12EF environment, which used single-row plots, or in the 12EF-13EF combined analysis (Figure 21). This suggests that multi-row plots are necessary for accurate phenotyping of lodging. Since lodging is an important yield component trait with few reported QTL, the lodging QTL reported here warrants further study. No significant QTL were discovered for biomass yield (Figure 22).

Linkage mapping with bi- parental populations

In the BTx623 x Tx2909 population, QTL peaks are present on chromosomes 1 and 6 (Figure 23A and Figure 33). The peak on chromosome 1 is close to the reported location of *Ma7*, while the peak on chromosome 6 is close to the presumed location of *Ma6*. In the BTx623 x Tx2910 population, one significant peak is present on chromosome 6 close to the

cloned *Ma1* locus at 40.3 Mb, and other significant peaks are found on chromosomes 3 and 9 (Figure 23B). This result is compelling since both BTx623 x Tx2909 and BTx623 x Tx2910 hybrids were thought to be photoperiod sensitive because of the *Ma5/Ma6* interaction. However, the mechanism of photoperiod-sensitivity clearly differs between the two populations. The BTx623 x Tx2910 population appears to confer photoperiod sensitivity to its hybrids through action of *Ma1* and QTL on chromosomes 3 and 9, and the BTx623 x Tx2909 population appears to confer photoperiod-sensitivity through the combined action of *Ma6* and *Ma7*. The dominant *Ma6/SbGhd7* allele causes plants to flower later, but in a homozygous *ma7/ma7* background, the effects of *Ma6* are suppressed such that all plants flower relatively early (Figure 24). The absence of QTL for *Ma5* in this study is surprising, and may indicate that the effects of *Ma5* are reduced in our Illinois environments compared to the previously- evaluated effects of *Ma5* in Texas.

A significant QTL for plant stand was detected on chromosome 3 in the BTx623 x Tx2910 population and a nearly significant QTL for the tillery trait was found on chromosome 5 in this population (Figure 25). This suggests that variation in these traits is under separate genetic control, and that the reduced plant stands observed in this experiment were not due to high early-season mortality of tillery individuals.

CONCLUSIONS

Biomass sorghum is poised to become a viable renewable fuel source. Results from this study suggest several possible breeding strategies for improving biomass sorghum. Forty percent of biomass yield variance was explained by plant height and lodging over the two years of yield trials in this study. Much of the remaining sixty percent unexplained variance might be explained by inconsistent plant stands across the field. Great variation in seed size and germination rate made it difficult to achieve consistent stands across genotypes, locations, and years. A recent study showed that maximum yields of biomass sorghum were obtained using very narrow (7.5") row spacing and relatively low seeding rates (116,000 seeds ha^{-1}; Snider et al., 2012). Nearly equidistant seed placement may help achieve more uniform stands by allowing surviving plants to more successfully compensate for neighbors that did not germinate or survive. Identifying female parents that increase seed size and improve hybrid germination in colder soils would also be helpful for improving plant stands. The *ma1* allele present in ATx623 *(Sbprr37-3)* is not ideal for making biomass sorghum hybrids since it appears not to be fully recessive. High-yielding inbreds are distributed across all genetic sub-populations, suggesting that breeders have sufficient genetic diversity to make rapid genetic gains in biomass yield. Biomass yields in inbreds and their corresponding hybrids show a remarkably high correlation ($r > 0.8$) when a common, dwarf female parent is used, validating the unusual strategy of selecting hybrid parents based on inbred performance. Since only one tester (ATx623) was used for creating hybrids, we cannot yet estimate combining ability within and between genetic sub-populations. Creation of new female testers from each of the three major genetic sub-populations would be a logical first step for the eventual exploitation of heterosis in

biomass sorghum hybrids. The proposed ideotype of high- yielding biomass sorghum is a plant that is tall, lodging resistant, and at least moderately photoperiod-sensitive in order to maximize vegetative growth. Leaf architecture was not shown to contribute to yield in this experiment, but the shorter leaves characteristic of durras might prove beneficial for light capture when grown under narrower row spacing.

TABLES

Rating	Phenotype
0	Completely vegetative, showing no signs of flowering
1	Booting stage
2	Anthesis
3	Grain fill
4	Mature grain stage

Table 1: Maturity rating scale taken to record developmental stage of plants following a killing frost.

FIGURES

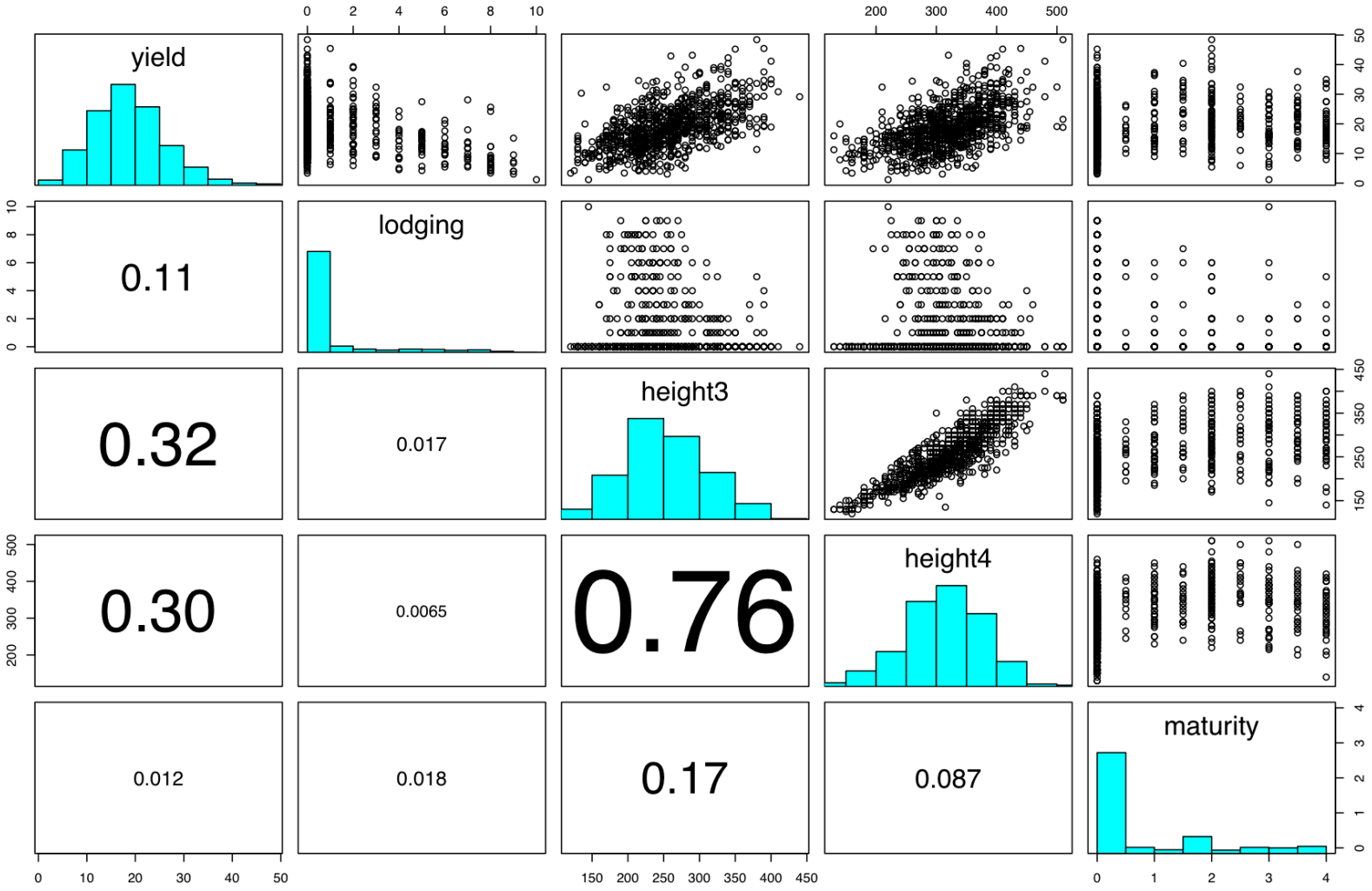

Figure 1: Inbred trait correlations over two seasons (2012 Energy Farm and 2013 Energy Farm). The diagonal shows a histogram for each trait. For each pair of traits, the upper triangle shows scatterplots and the lower triangle shows r^2 values. Traits units are as follows: maturity and lodging (categorical); 90-day height (height3) and 120-day height (height4) (cm); and yield (dry metric tons per hectare).

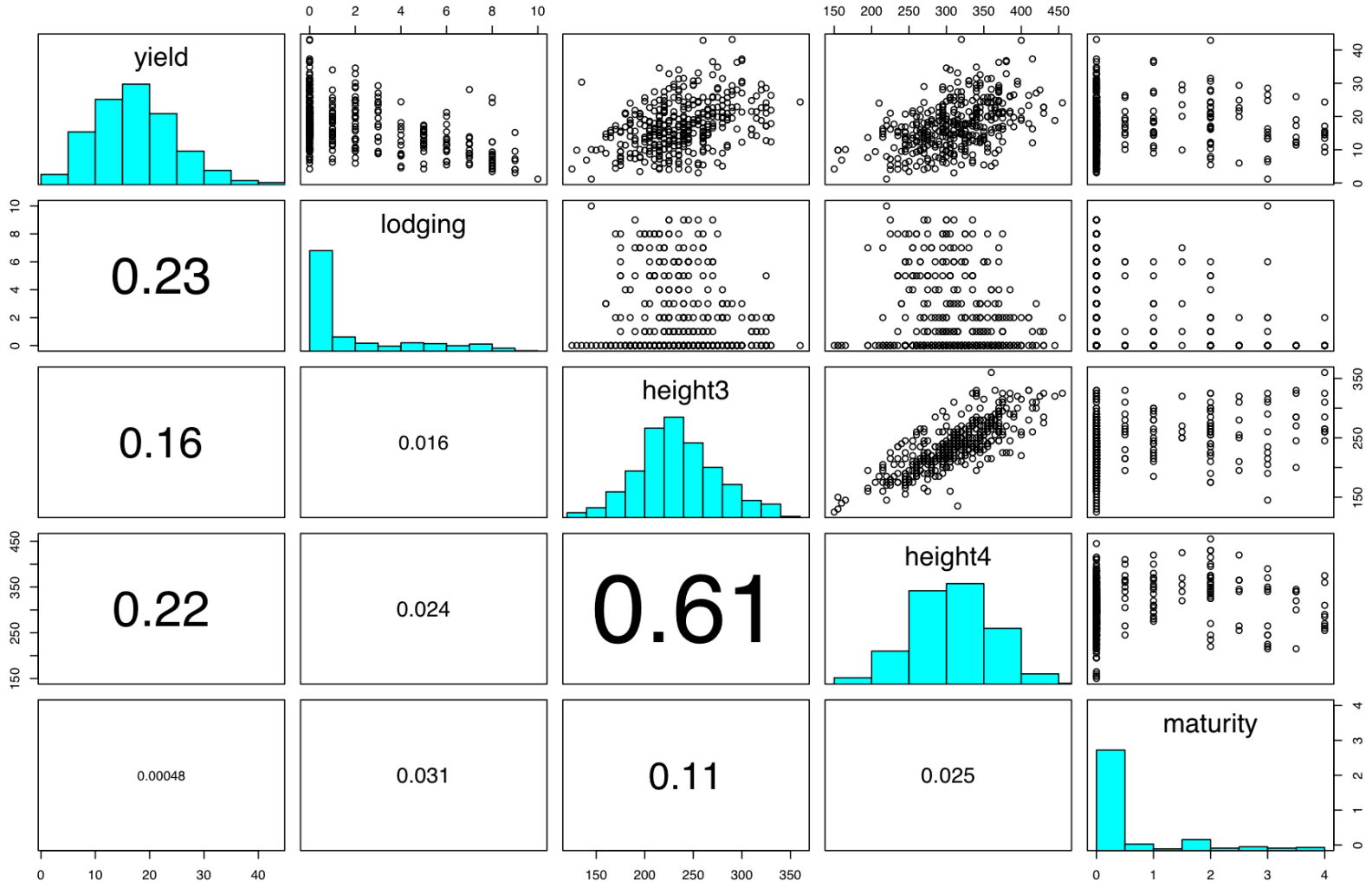

Figure 2: Inbred trait correlations for 2012 Energy Farm. The diagonal shows a histogram for each trait. For each pair of traits, the upper triangle shows scatterplots and the lower triangle shows r^2 values. Traits units are as follows: maturity and lodging (categorical); 90-day height (height3) and 120-day height (height4) (cm); and yield (dry metric tons per hectare).

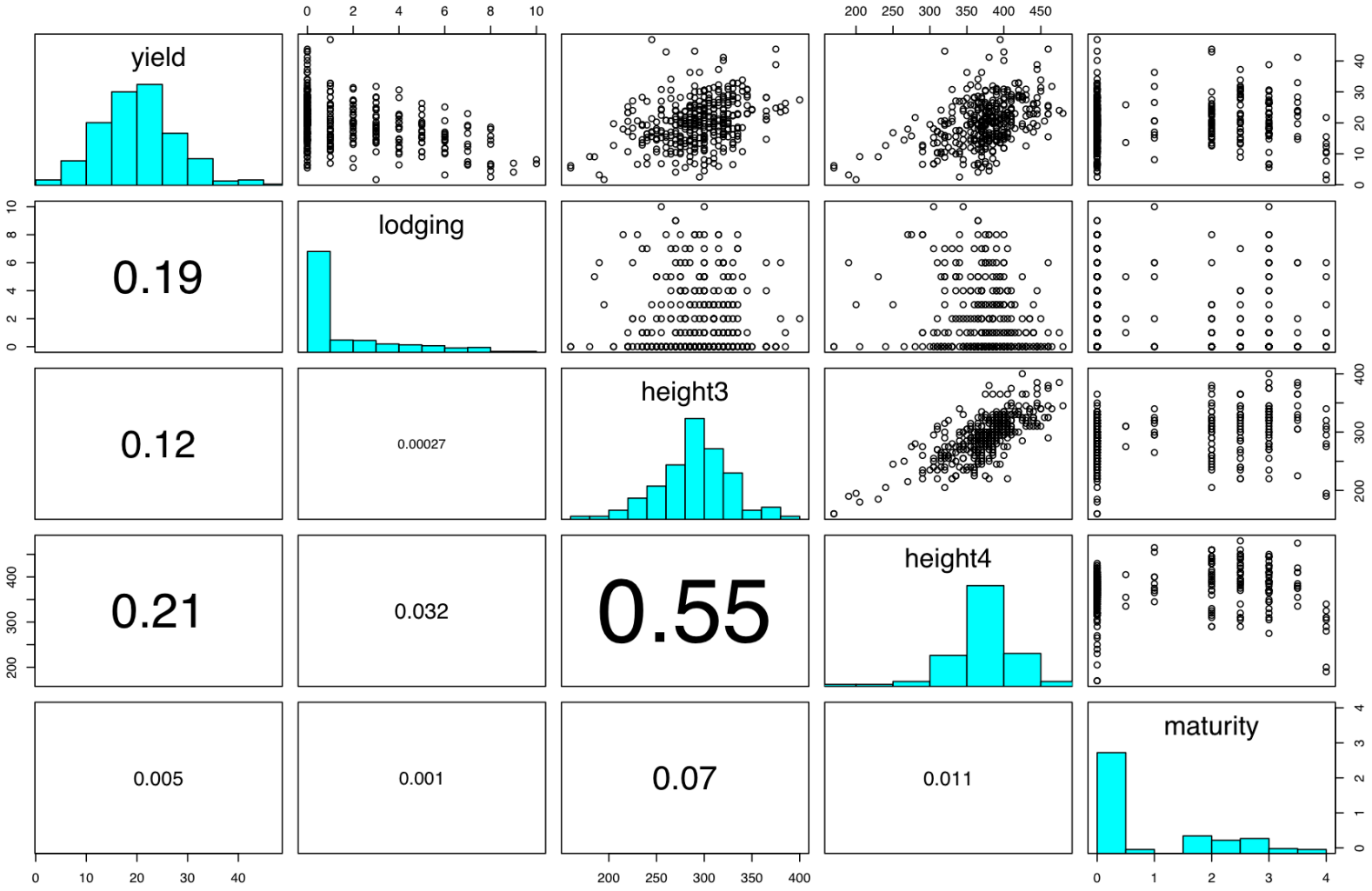

Figure 3: Hybrid trait correlations for 2012 Energy Farm. The diagonal shows a histogram for each trait. For each pair of traits, the upper triangle shows scatterplots and the lower triangle shows r^2 values. Traits units are as follows: maturity and lodging (categorical); 90-day height (height3) and 120-day height (height4) (cm); and yield (dry metric tons per hectare).

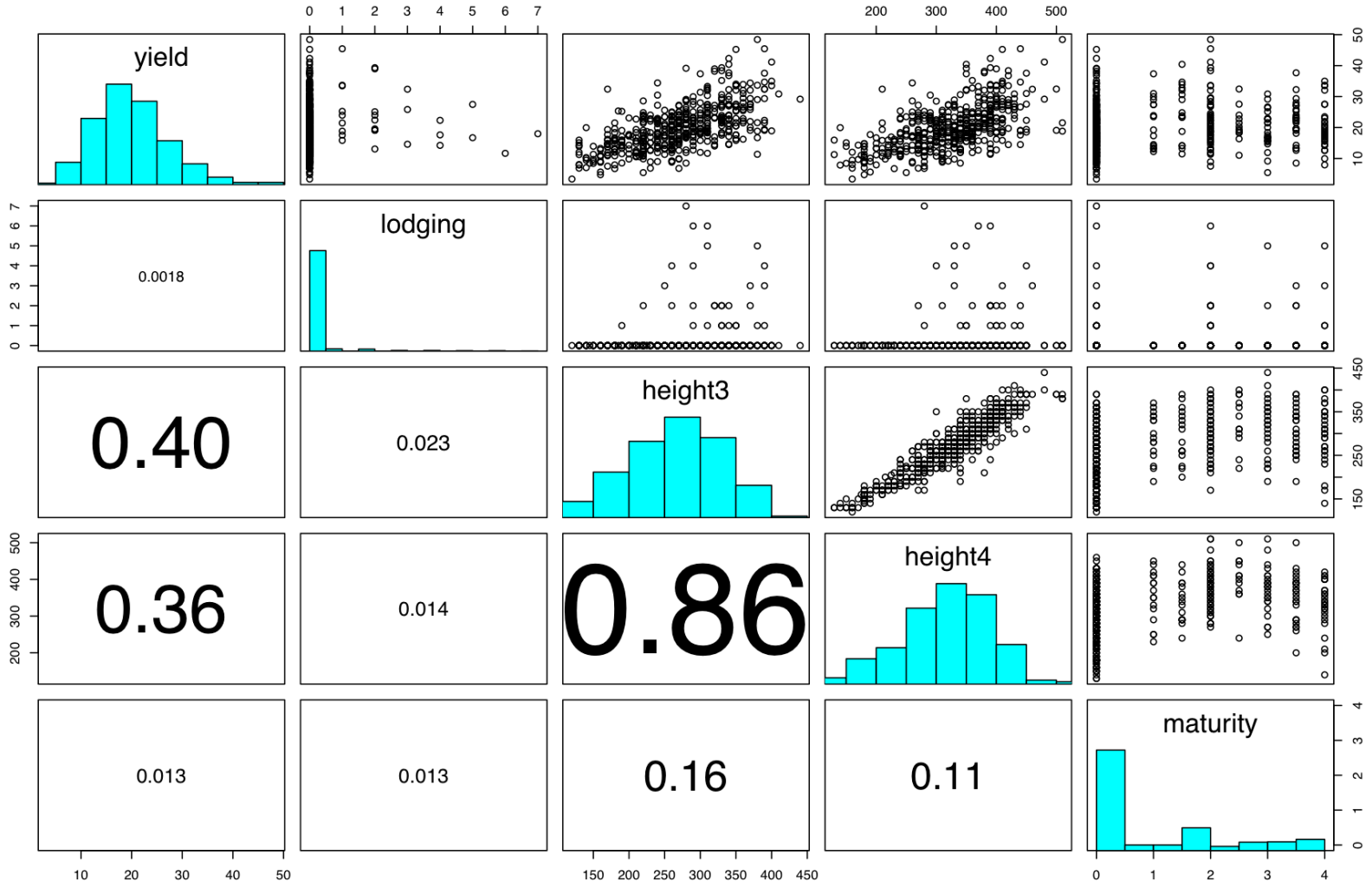

Figure 4: Inbred trait correlations for 2013 Energy Farm. The diagonal shows a histogram for each trait. For each pair of traits, the upper triangle shows scatterplots and the lower triangle shows r^2 values. Traits units are as follows: maturity and lodging (categorical); 90-day height (height3) and 120-day height (height4) (cm); and yield (dry metric tons per hectare).

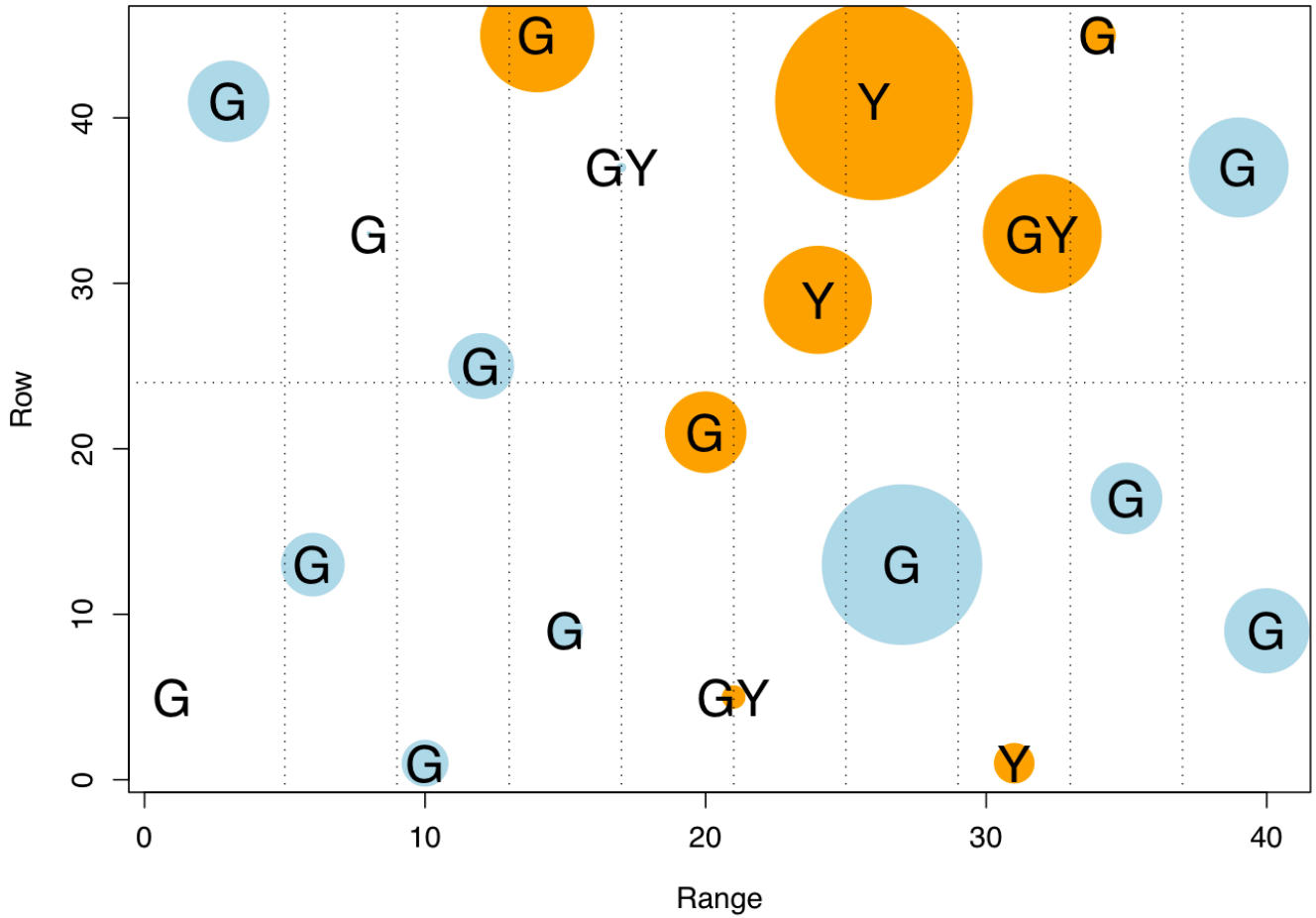

Figure 5: Relative yield and nitrogen status of replicated commercial hybrids in the 2013 Urbana biomass yield trial. An augmented design was used with the same commercial hybrid included in each incomplete block. Each circle represents one 4-row hybrid plot. Green and red circles represent biomass yields above and below the mean, respectively, and the size of each circle is proportional to the standard deviation from the mean. Letters in each circle represent a qualitative nitrogen status score assessed at harvest: G= green; GY= green-yellow; Y= yellow.

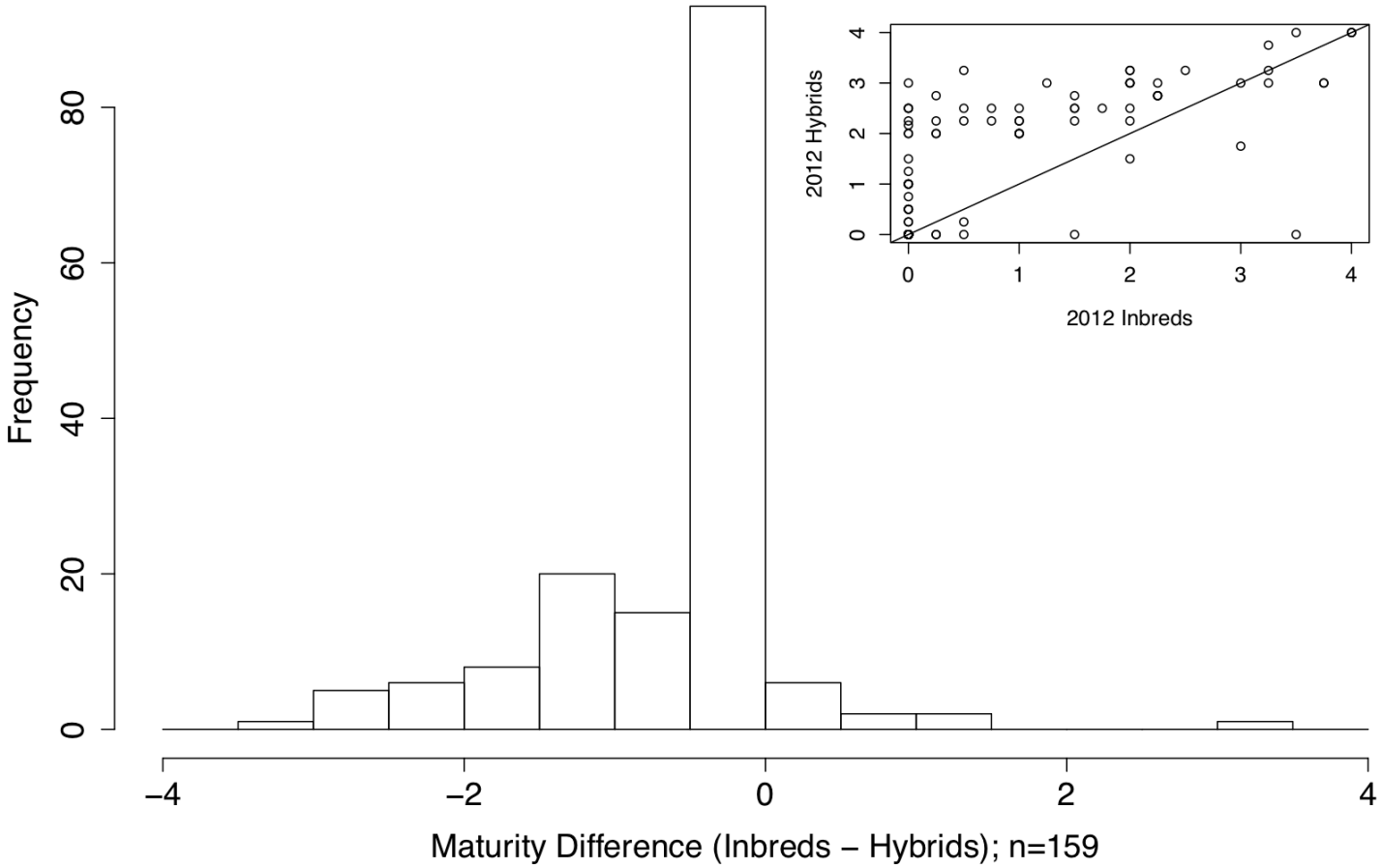

Figure 6: Maturity differences between inbreds and hybrids in 2012. The maturity rating for a hybrid was subtracted from the maturity rating for its exotic inbred parent. Inset is a scatter plot with a solid line indicating the 1:1 relationship.

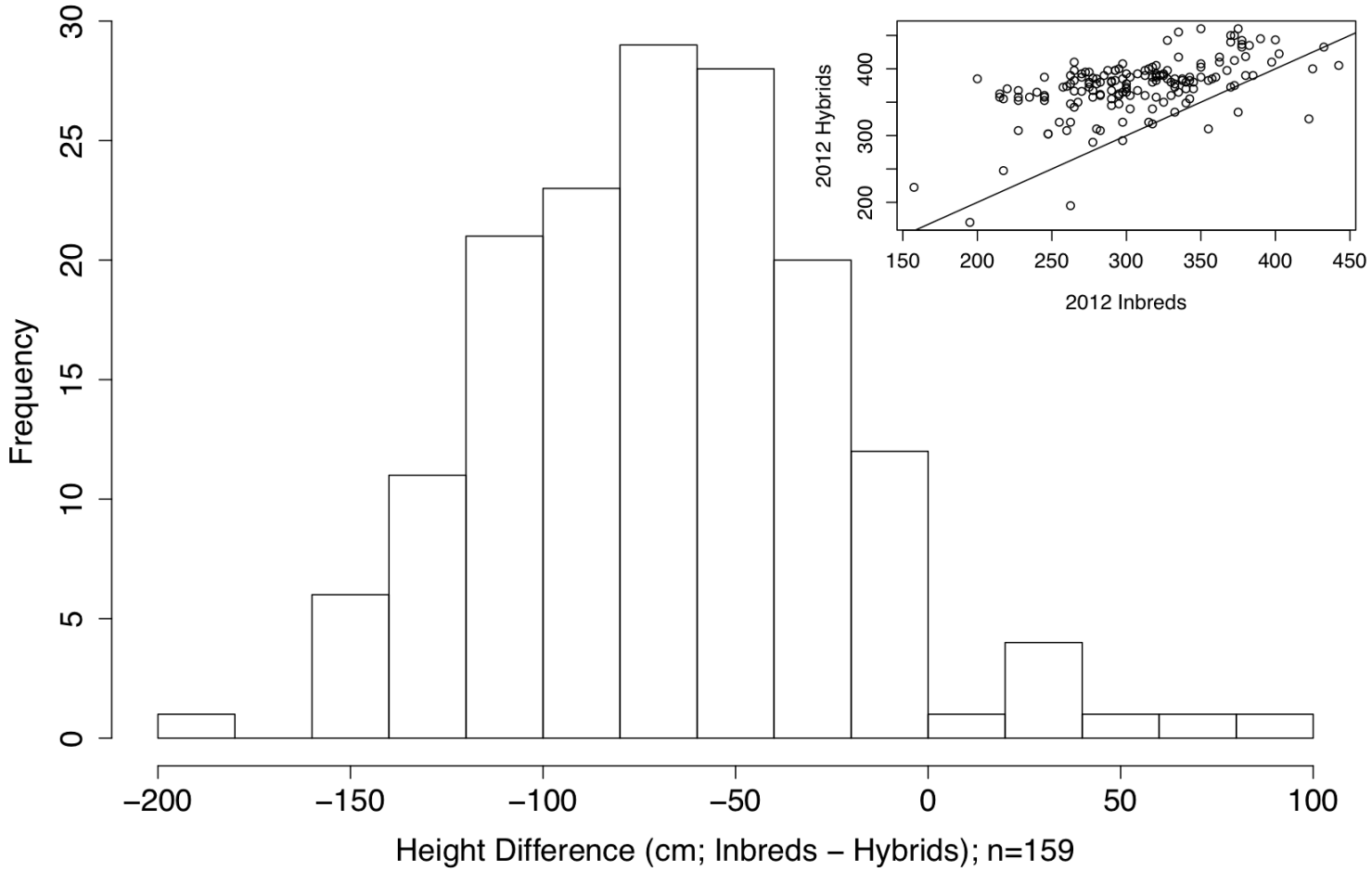

Figure 7: Plant height differences between inbreds and hybrids in 2012. The plant height of a hybrid line was subtracted from the height of its inbred parent. Inset is a scatter plot with a solid line indicating the 1:1 relationship.

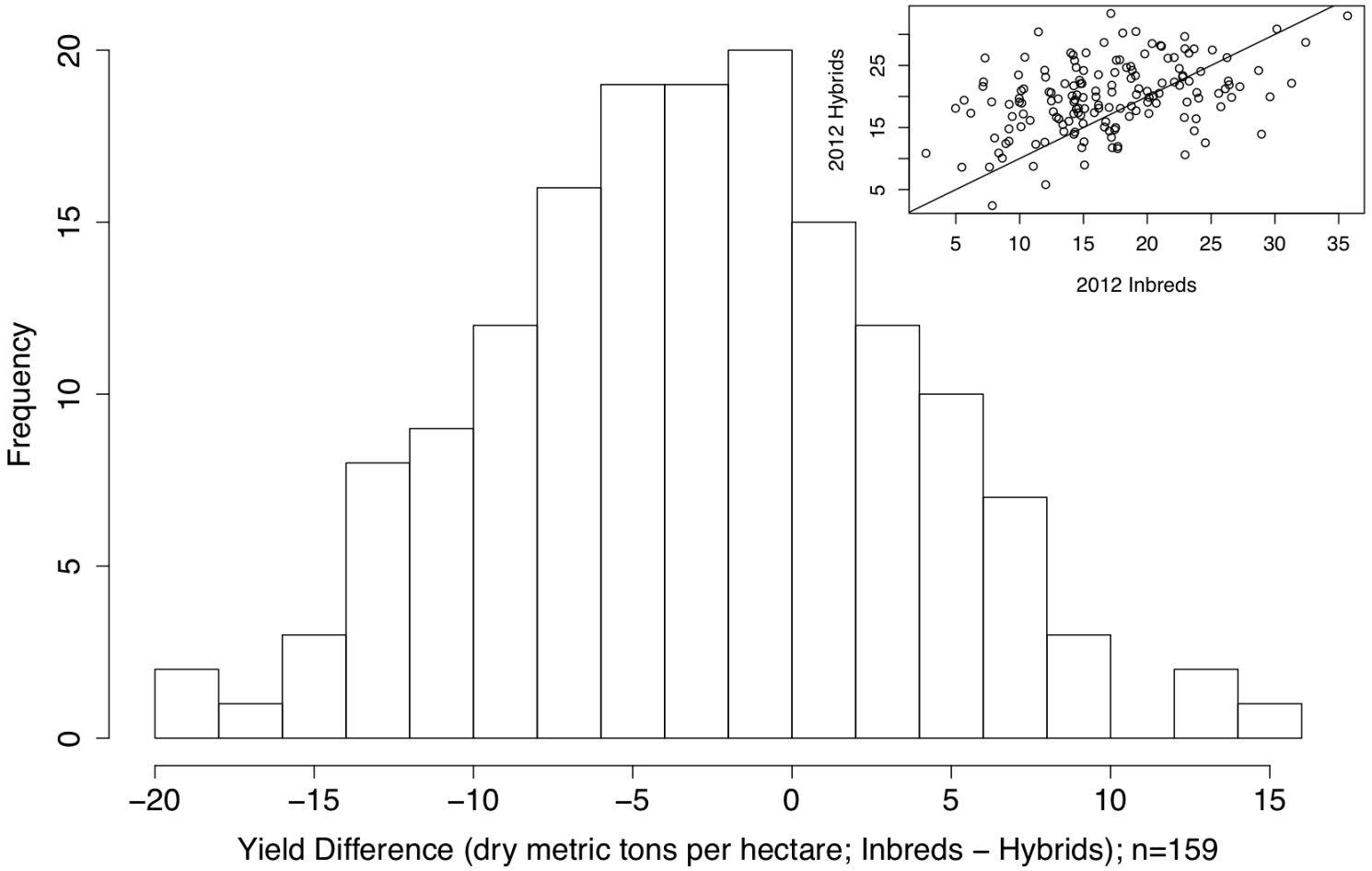

Figure 8: Biomass yield differences between inbreds and hybrids in 2012. The biomass yield of a hybrid line was subtracted from the biomass yield of its inbred parent. Inset is a scatter plot with a solid line indicating the 1:1 relationship.

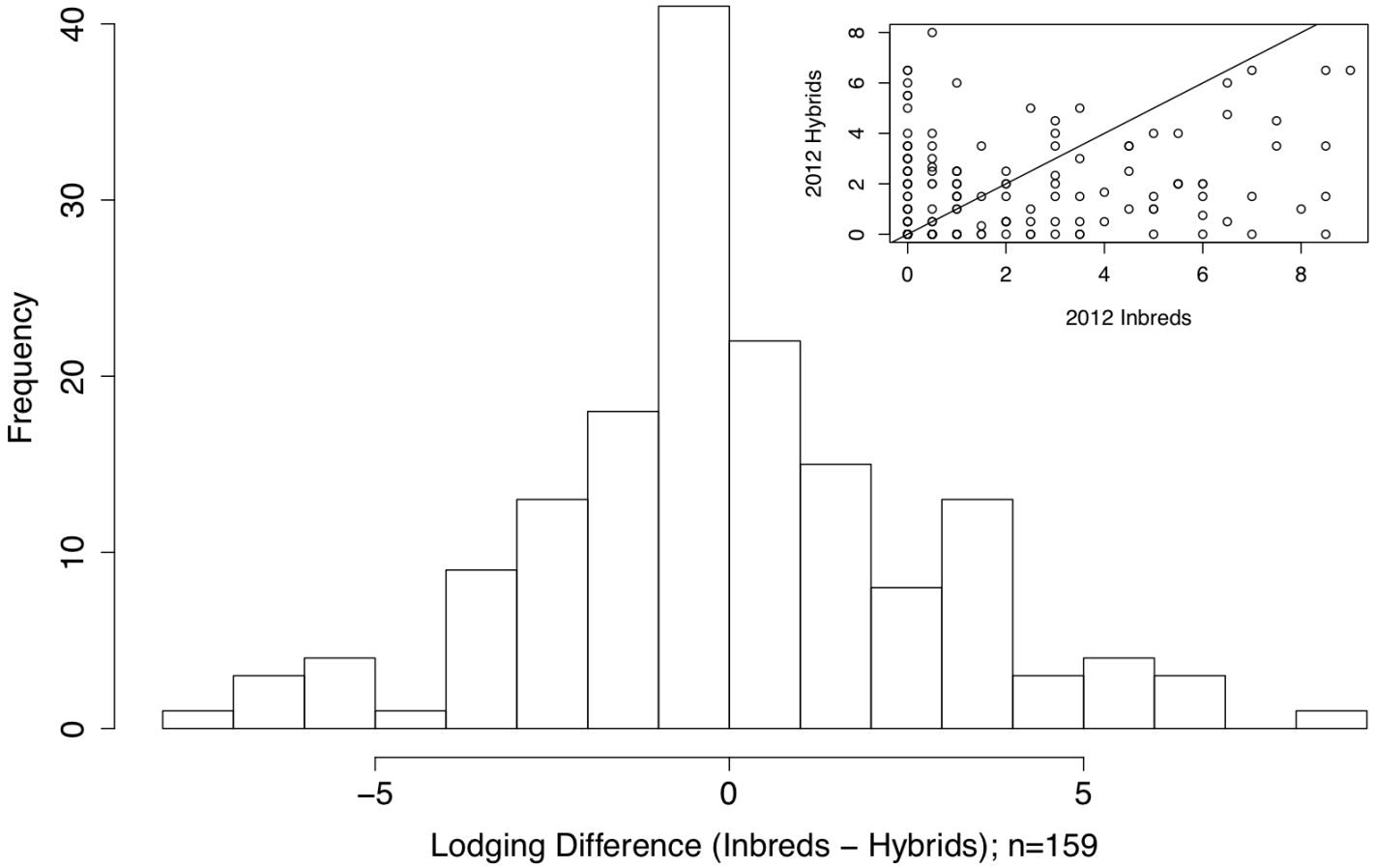

Figure 9: Lodging severity differences between inbreds and hybrids in 2012. The lodging score of a hybrid line was subtracted from the lodging score of its inbred parent. Inset is the scatter plot. The solid line is the 1:1 ratio.

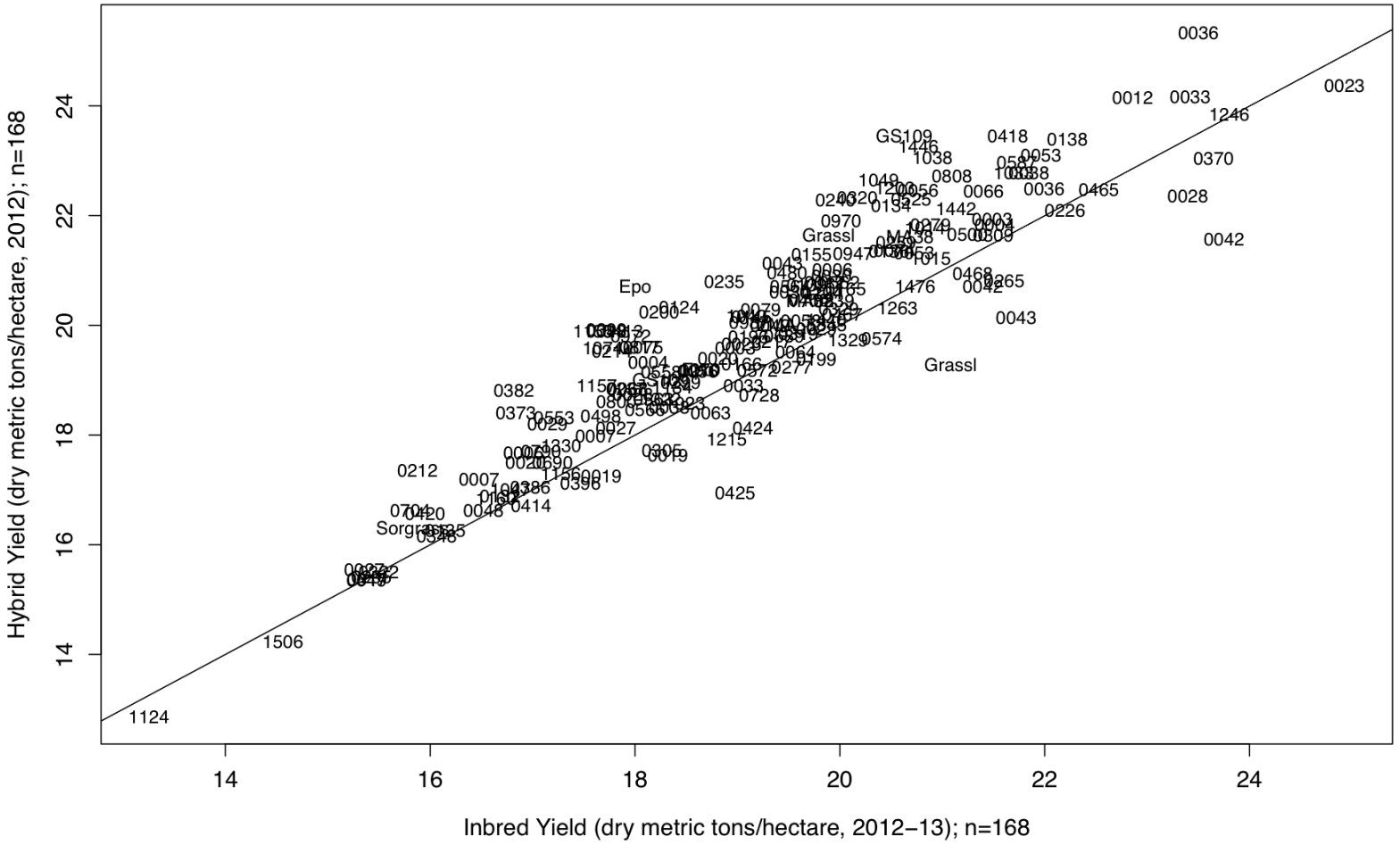

Figure 10: Regression of hybrid biomass yield (2012) on inbred biomass yield (2012, 2013). The r^2 is 0.82. The points are labeled as the line number for identification.

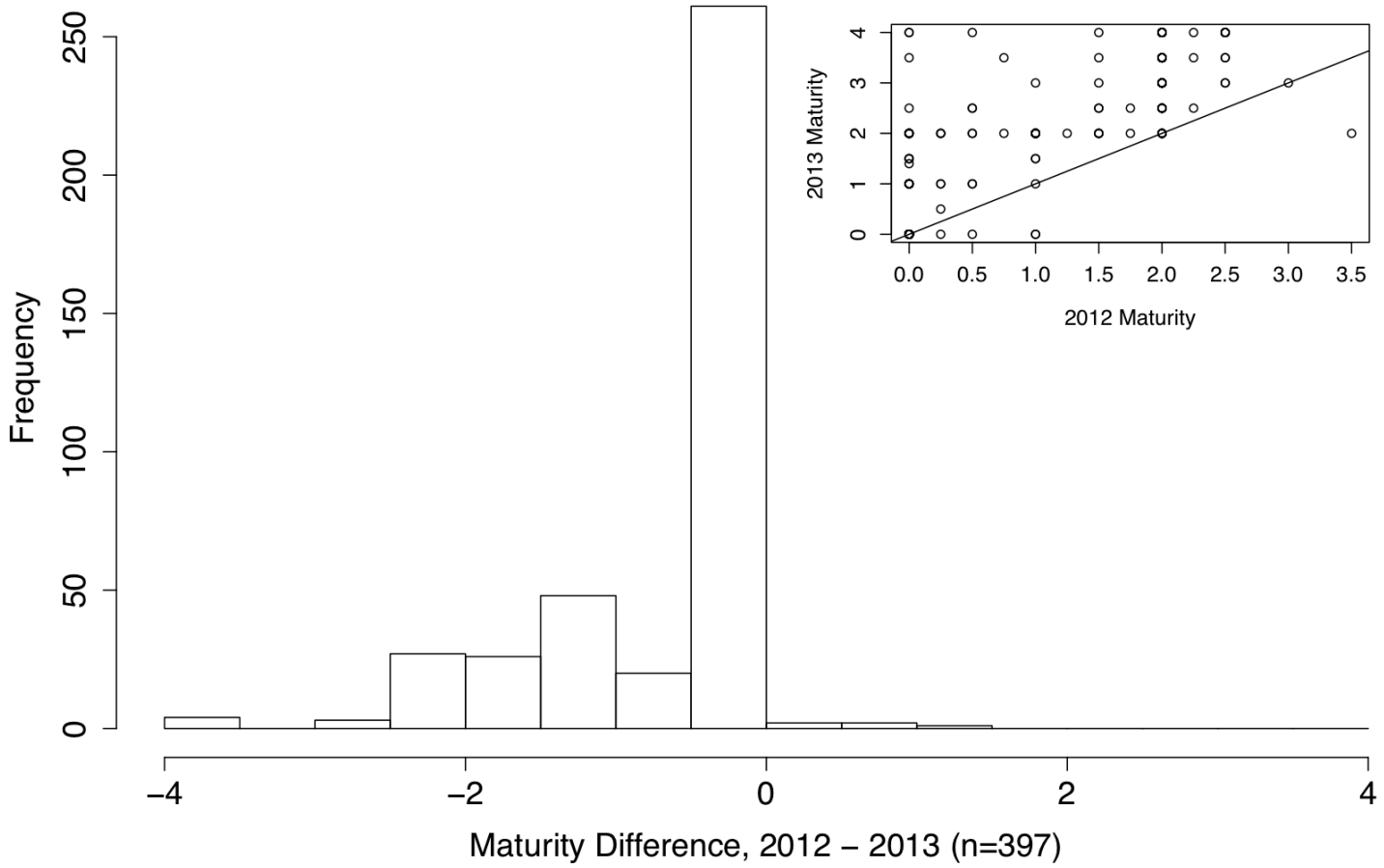

Figure 11: Maturity differences between inbreds in 2012 and 2013. The maturity rating for an inbred line in 2013 was subtracted its maturity rating in 2012. Predominantly negative values indicate that most lines flowered earlier in 2013 than in 2012. Inset is a scatter plot with a solid line indicating the 1:1 relationship.

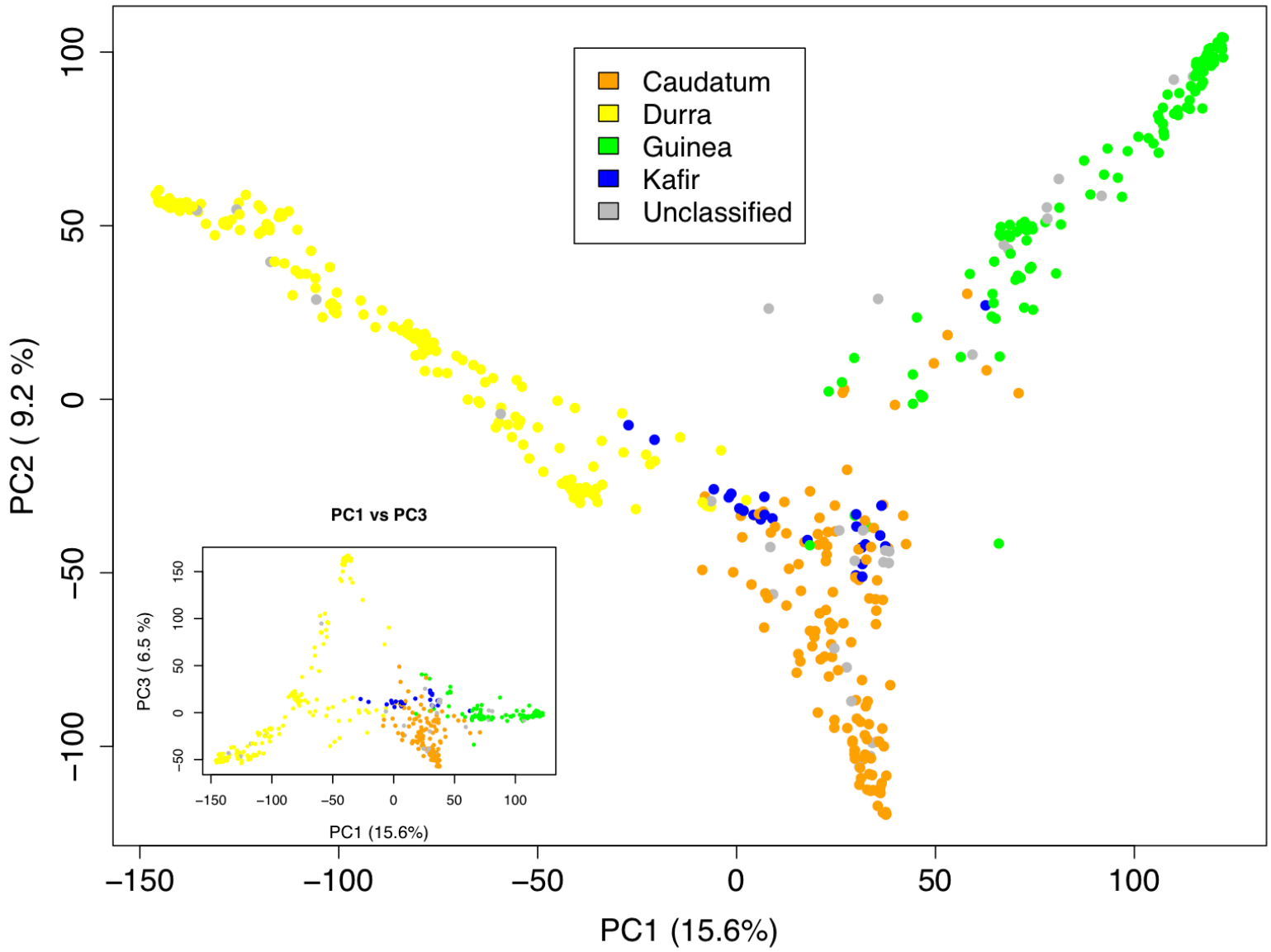

Figure 12: Principal component analysis of sorghum genetic structure and its relationship with sorghum morphological race. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

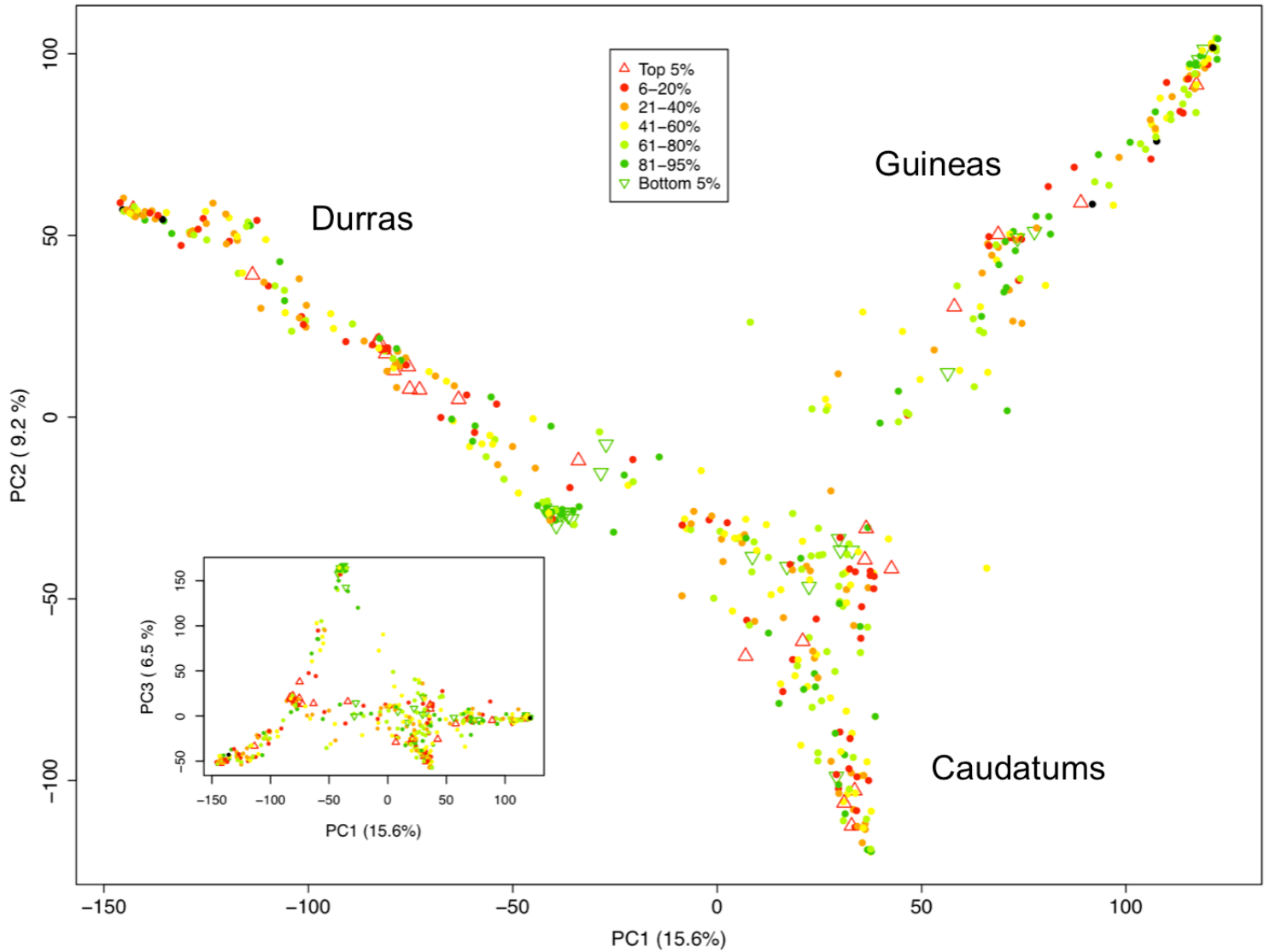

Figure 13: Principal component analysis of sorghum genetic structure and its relationship with the biomass yield phenotype. Each dot represents a sorghum inbred (n=451) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

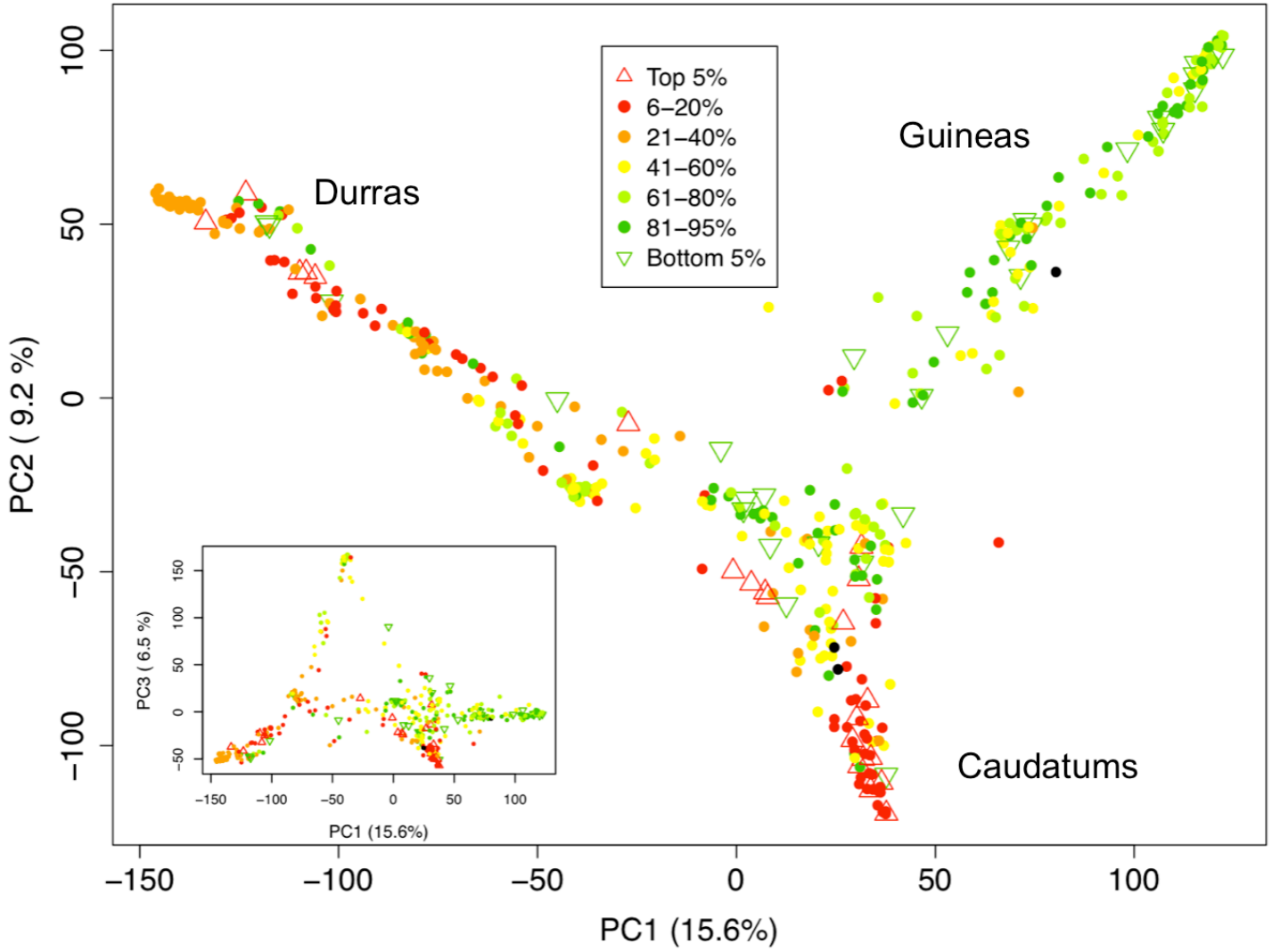

Figure 14: Principal component analysis of sorghum genetic structure and its relationship with the maturity phenotype. Each dot represents a sorghum inbred (n=453) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

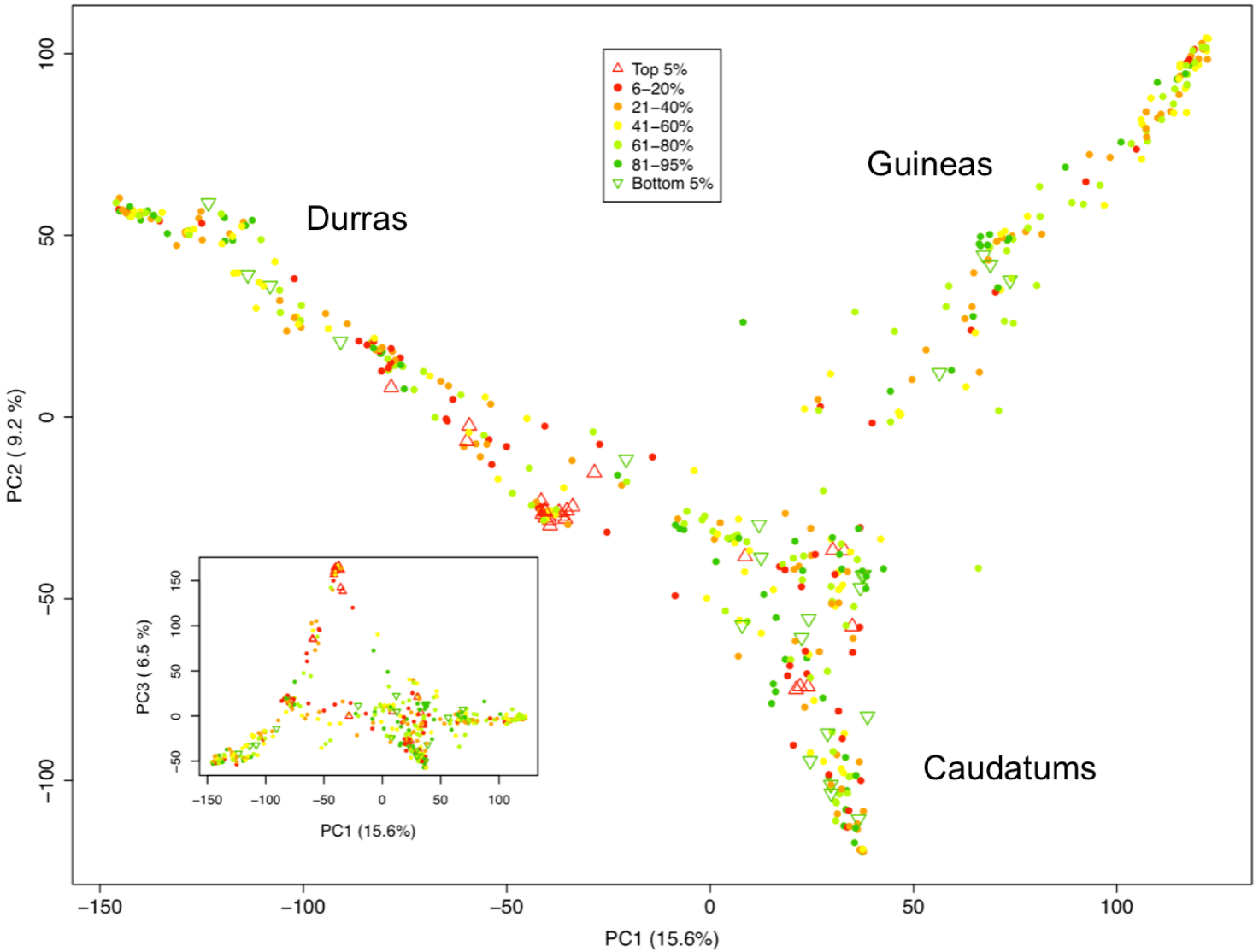

Figure 15: Principal component analysis of sorghum genetic structure and its relationship with the lodging phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

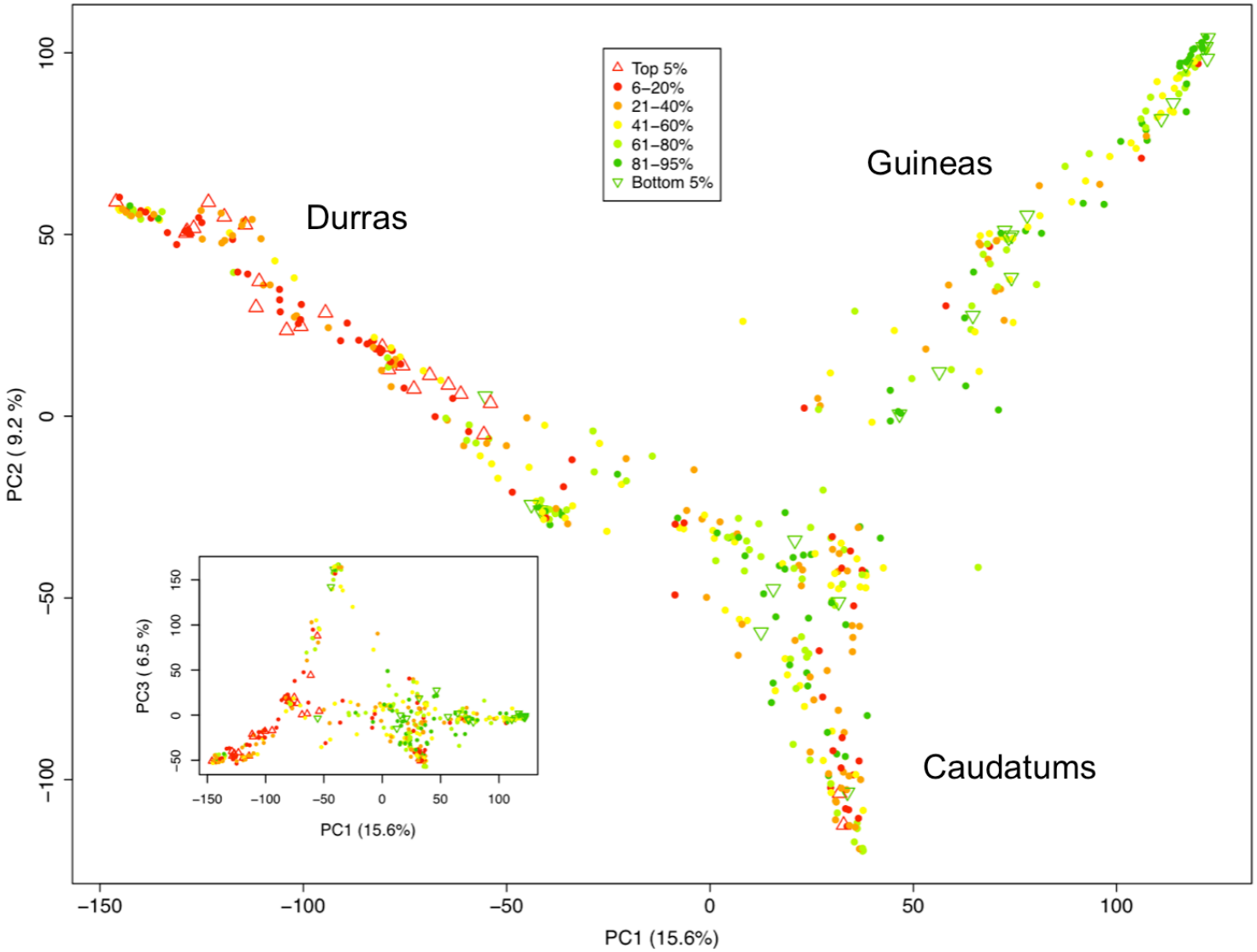

Figure 16: Principal component analysis of sorghum genetic structure and its relationship with the 90- day height phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

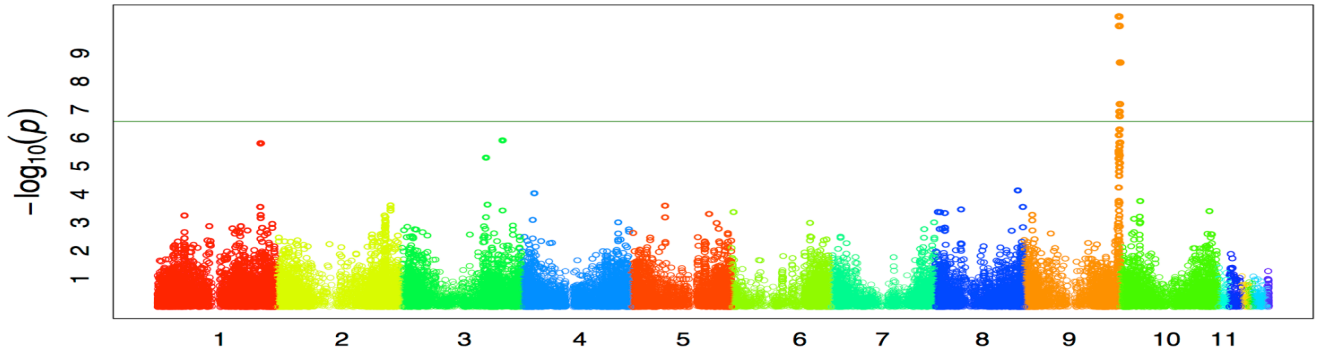

Figure 17: Genome wide association analysis of maturity in grain sorghum. Maturity measured as days to anthesis. The horizontal green line is the Bonferroni- corrected significance threshold at α= 0.01. (# of observations = 769, # of SNPs = 37,025)

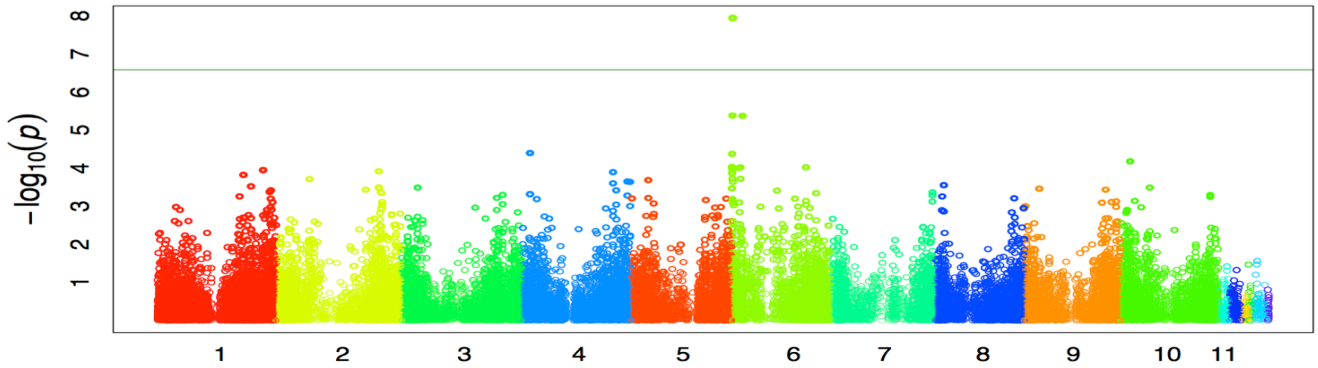

Figure 18: Genome wide association analysis of maturity in biomass sorghum. Maturity measured on 0-4 scale at end of season. The horizontal green line is the Bonferroni- corrected significance threshold at α= 0.01. (# of observations = 619, # of SNPs = 38,135)

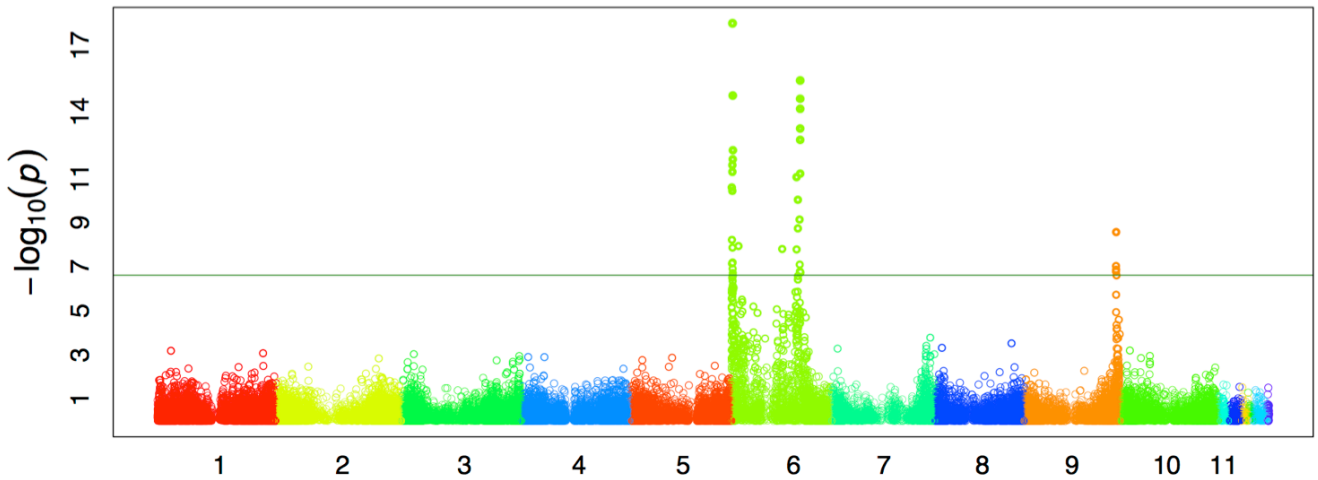

Figure 19: Genome wide association analysis of maturity in grain and biomass sorghum combined. The horizontal green line is the Bonferroni- corrected significance threshold at α= 0.01. (# of observations = 1,391, # of SNPs = 38,995)

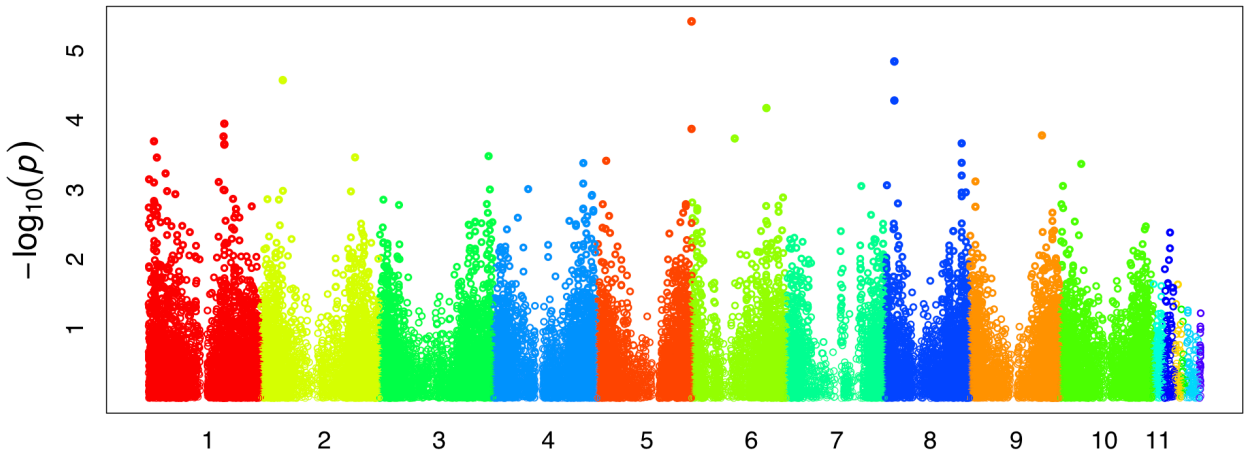

Figure 20: Genome wide association analysis of 90- day height in biomass sorghum in 2013. Height is measured from the ground to the top of the whorl 90 days after planting. (# of observations = 456, # of SNPs = 37,307)

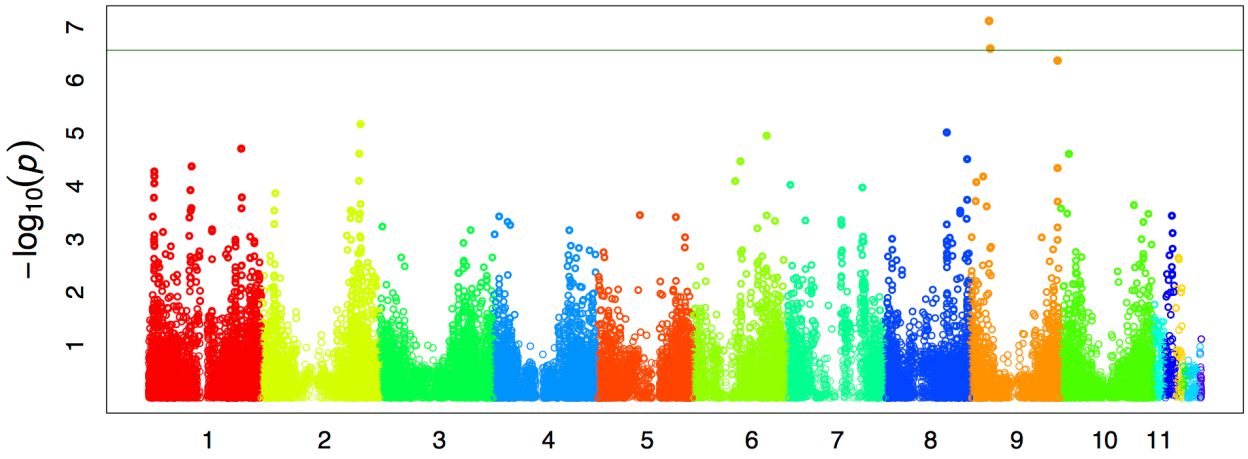

Figure 21: Genome wide association analysis of lodging in biomass sorghum in 2013. Lodging was measured on a 0-9 scale. The horizontal green line is the Bonferroni- corrected significance threshold at α= 0.01. (# of observations = 374, # of SNPs = 36,778)

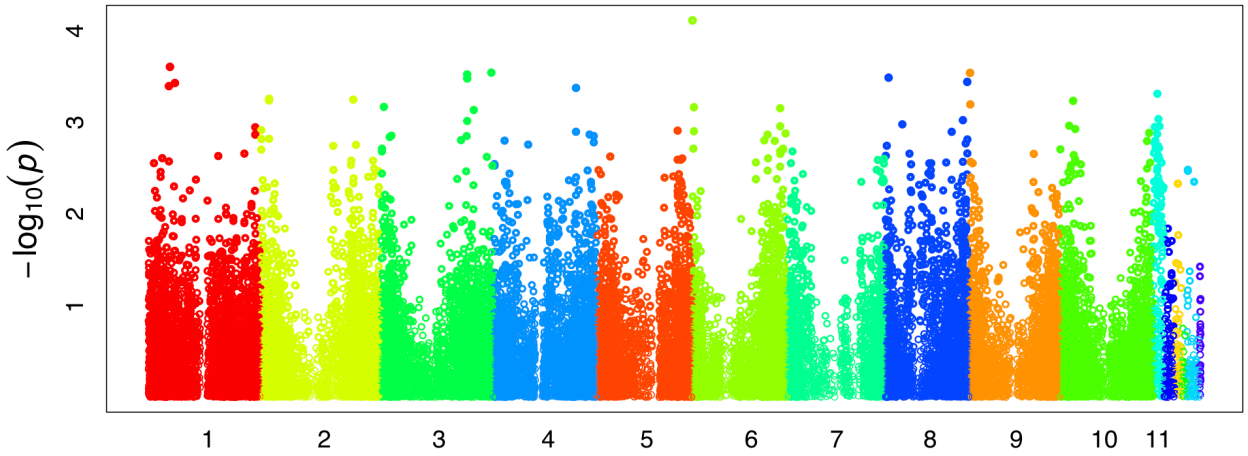

Figure 22: Genome wide association analysis of biomass yield. (# of observations = 451, # of SNPs = 37,307)

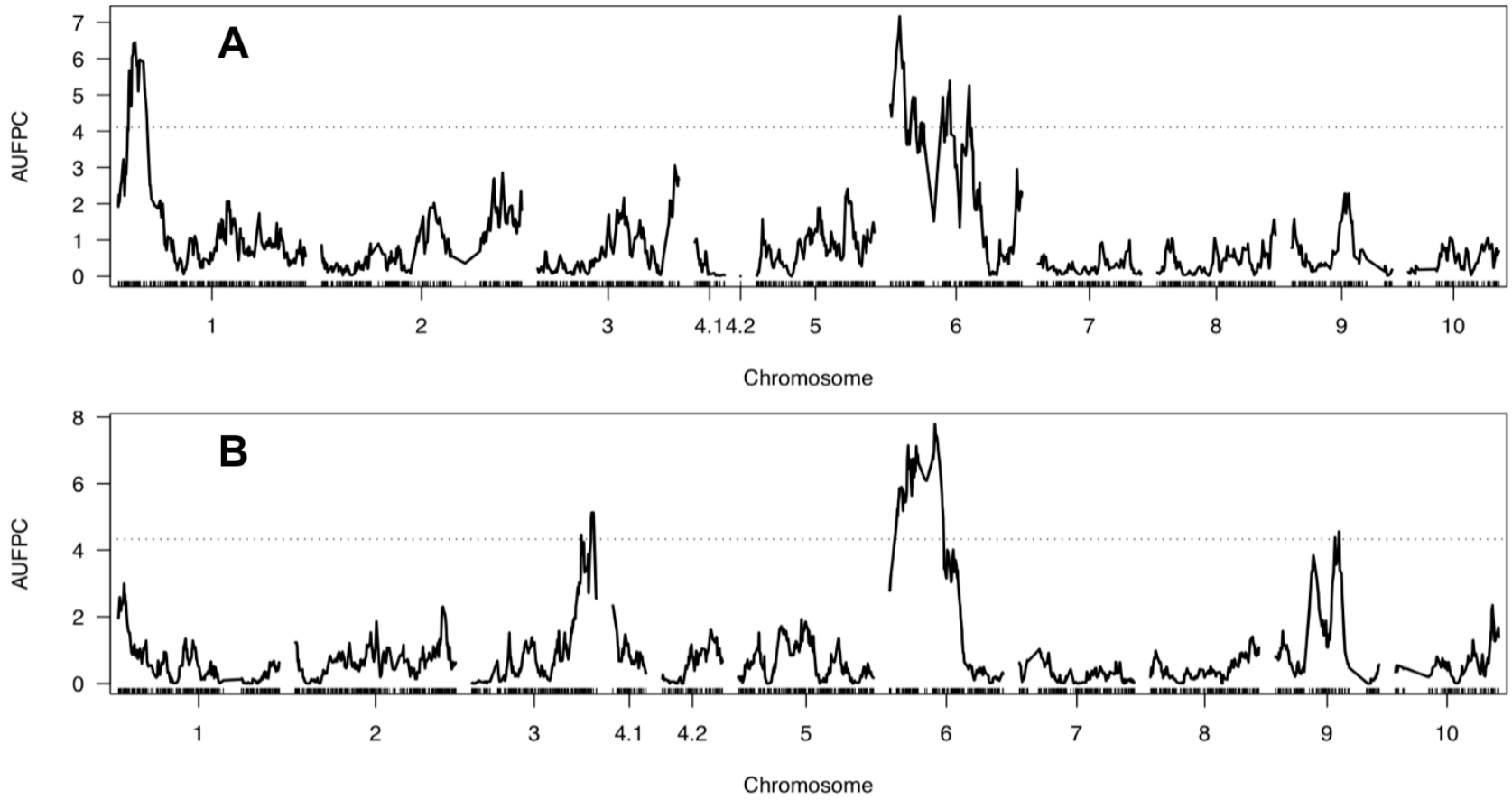

Figure 23: QTL analysis of the AUFPC trait for (A) BTx623 x Tx2909 and (B) BTx623 x Tx2910 populations. The dotted line is the significance threshold at α=0.05.

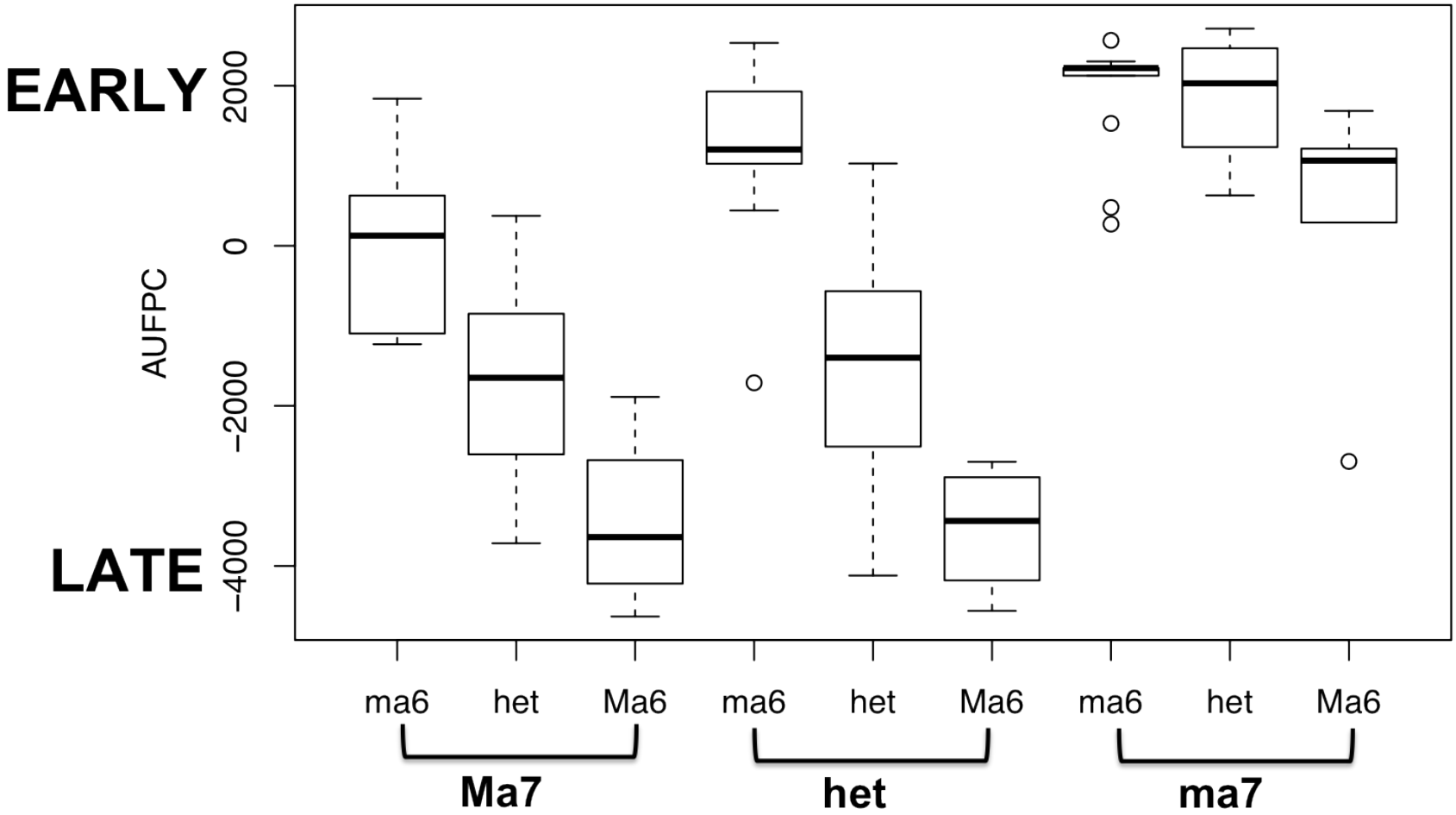

Figure 24: Boxplots showing the interaction between Ma6 and Ma7 in the Tx2909 population. *ma7/ma7* individuals flower early regardless of *Ma6* genotype.

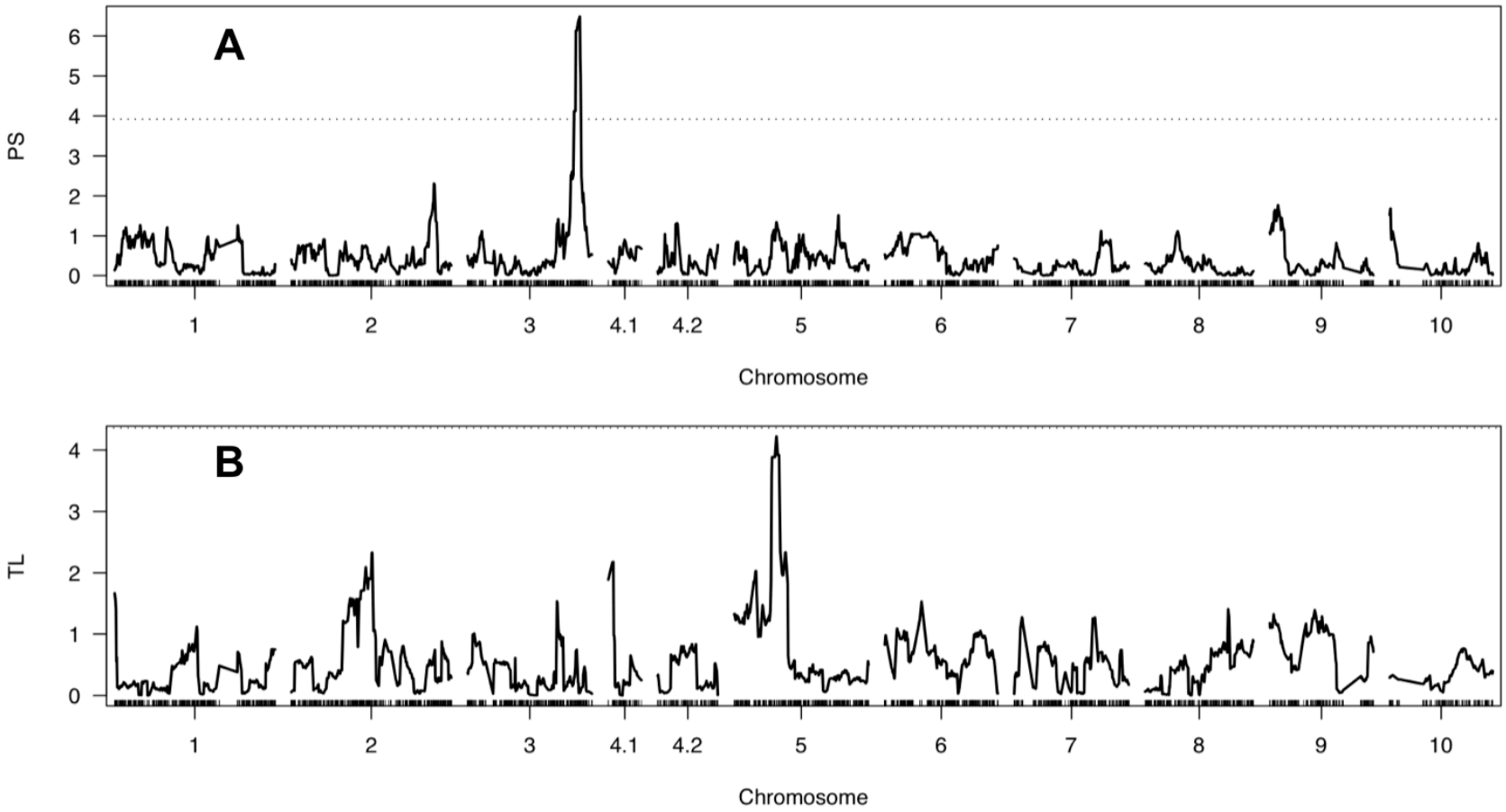

Figure 25: QTL analysis in the BTx623 X Tx2910 population for (A) plant stand and (B) tillery trait. The dotted line is the significance threshold at α=0.05.

LITERATURE CITED

.

Aydin, S., and W.L. Rooney. 1999. Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. Crop Sci. 39(2): 397.

Brown, P.J., S. Myles, and S. Kresovich. 2011. Genetic Support for Phenotype-based Racial Classification in Sorghum. Crop Sci. 51(1): 224.

Brown, P.J., and A.H. Paterson. 2013. Bridging Classical and Molecular Genetics of Sorghum Plant Stature and Maturity. p. 333–346. *In* Genomics of the Saccharinae. Springer, New York.

Browning, B.L., and S.R. Browning. 2011. A Fast, Powerful Method for Detecting Identity by Descent. Am. J. Hum. Genet. 88(2): 173–182.

Childs, K.L., J.L. Lu, J.E. Mullet, and P.W. Morgan. 1995. Genetic Regulation of Development in Sorghum bicolor (X. Greatly Attenuated Photoperiod Sensitivity in a Phytochrome-Deficient Sorghum Possessing a Biological Clock but Lacking a Red Light-High Irradiance Response). Plant Physiol. 108(1): 345–351.

Childs, K.L., F.R. Miller, M.M. Cordonnier-Pratt, L.H. Pratt, P.W. Morgan, and J.E. Mullet. 1997. The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiol. 113(2): 611–619.

Cockram, J., T. Thiel, B. Steuernagel, N. Stein, S. Taudien, P.C. Bailey, and D.M. O'Sullivan. 2012. Genome Dynamics Explain the Evolution of Flowering Time CCT Domain Gene Families in the Poaceae. Plos One 7(9)Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3454399/ (verified 8 November 2013).

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler 4th. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54: 357–374.

Harlan, J.R., and J.M.J. de Wet. 1972. A Simplified Classification of Cultivated Sorghum1. Crop Sci. 12(2): 172.

Hung, H.-Y., L.M. Shannon, F. Tian, P.J. Bradbury, C. Chen, S.A. Flint-Garcia, M.D. McMullen, D. Ware, E.S. Buckler, J.F. Doebley, and J.B. Holland. 2012. ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl. Acad. Sci. 109(28): E1913–E1921.

J. R. Quinby. 1974. Sorghum Improvement and the Genetics of Growth. The Texas Agricultural Experiment Station.

Jones, R.L., H. Ougham, and H. Thomas. 2013. The molecular life of plants. Wiley-Blackwell, Oxford [etc.].

Koo, B.-H., S.-C. Yoo, J.-W. Park, C.-T. Kwon, B.-D. Lee, G. An, Z. Zhang, J. Li, Z. Li, and N.-C. Paek. 2013. Natural Variation in OsPRR37 Regulates Heading Date and Contributes to Rice Cultivation at a Wide Range of Latitudes. Mol. Plant.

Lin, Z., X. Li, L.M. Shannon, C.-T. Yeh, M.L. Wang, G. Bai, Z. Peng, J. Li, H.N. Trick, T.E. Clemente, J. Doebley, P.S. Schnable, M.R. Tuinstra, T.T. Tesso, F. White, and J. Yu. 2012. Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44(6): 720–724.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinforma. Oxf. Engl. 28(18): 2397–2399.

Liu, T., H. Liu, H. Zhang, and Y. Xing. 2013. Validation and Characterization of Ghd7.1, a Major Quantitative Trait Locus with Pleiotropic Effects on Spikelets per Panicle, Plant Height, and Heading Date in Rice (Oryza sativa L.). J. Integr. Plant Biol. 55(10): 917–927.

Mace, E.S., and D.R. Jordan. 2010. Location of major effect genes in sorghum (Sorghum bicolor (L.) Moench). Tag Theor. Appl. Genet. Theor. Angew. Genet. 121(7): 1339–1356.

Mullet, J.E., W.L. Rooney, P.E. Klein, D. Morishige, R. Murphy, and J.A. Brady. 2010. Discovery and utilization of sorghum genes (ma5/ma6).

Murphy, R.L., R.R. Klein, D.T. Morishige, J.A. Brady, W.L. Rooney, F.R. Miller, D.V. Dugas, P.E. Klein, and J.E. Mullet. 2011. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc. Natl. Acad. Sci. U. S. A. 108(39): 16469–16474.

Quinby, J.R. 1966. Fourth Maturity Gene Locus in Sorghum1. Crop Sci. 6(6): 516.

Smith, C.W., and R. Frederiksen. 2000. Sorghum: origin, history, technology and production. Wiley, New York.

Snider, J.L., R.L. Raper, and E.B. Schwab. 2012. The effect of row spacing and seeding rate on biomass production and plant stand characteristics of non-irrigated photoperiod-sensitive sorghum (Sorghum bicolor (L.) Moench). Ind. Crops Prod. 37(1): 527–535.

Tarumoto, I., M. Yanase, Y. Iwahara, Y. Kuzumi, T. Morikawa, and S. Kasuga. 2003. Inheritance of a Thermo-sensitivity Gene Controlling Flower Initiation in Sorghum. Breed. Sci. 53(4): 353–357.

Tarumoto, I., M. Yanase, H. Kadowaki, T. Yamada, and S. Kasuga. 2005. Inheritance of photoperiod-sensitivity genes controlling flower initiation in sorghum, Sorghum bicolor Moench. Grassl. Sci. 51(1): 55–61.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14(6): R68.

Xue, W., Y. Xing, X. Weng, Y. Zhao, W. Tang, L. Wang, H. Zhou, S. Yu, C. Xu, X. Li, and Q. Zhang. 2008. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat. Genet. 40(6): 761–767.

APPENDIX

Chromosome	Position	P-value	Minor Allele Freq	# of Obs	FDR Adj P-values	Trait	Year
1	4084607	2.05E-05	0.466	453	0.070	Maturity	2012, 2013
1	48129300	1.52E-05	0.108	372	0.086	Maturity	2013
1	67405870	2.00E-06	0.208	372	0.074	Maturity	2013
2	19892831	3.82E-05	0.103	453	0.084	Maturity	2012, 2013
2	20017588	5.75E-06	0.108	453	0.043	Maturity	2012, 2013
2	60932808	1.18E-05	0.390	453	0.056	Maturity	2012, 2013
2	65208767	6.79E-06	0.143	374	0.062	Lodging	2013
4	58856631	1.36E-05	0.411	453	0.056	Maturity	2012, 2013
5	61286161	3.30E-05	0.105	453	0.084	Maturity	2012, 2013
5	62094784	6.12E-07	0.135	374	0.023	120- day height	2013
6	197807	2.94E-05	0.332	453	0.084	Maturity	2012, 2013
6	198475	7.59E-07	0.343	453	0.028	Maturity	2012, 2013
6	198860	1.29E-05	0.340	453	0.056	Maturity	2012, 2013
6	263833	1.28E-05	0.338	453	0.056	Maturity	2012, 2013
6	340855	1.75E-06	0.106	453	0.031	Maturity	2012, 2013
6	595560	4.18E-06	0.391	619	0.054	Maturity	2011, 2012, 2013
6	619807	3.54E-05	0.485	453	0.084	Maturity	2012, 2013
6	700785	1.14E-08	0.289	619	0.000	Maturity	2011, 2012, 2013
6	782644	2.24E-05	0.192	453	0.070	Maturity	2012, 2013
6	886632	3.32E-06	0.121	453	0.031	Maturity	2012, 2013
6	6484472	2.15E-05	0.204	453	0.070	Maturity	2012, 2013
6	6484472	6.82E-06	0.203	372	0.075	Maturity	2013
6	6857792	4.26E-06	0.192	619	0.054	Maturity	2011, 2012, 2013
6	32221721	3.72E-05	0.429	453	0.084	Maturity	2012, 2013
6	46066888	8.17E-06	0.120	372	0.075	Maturity	2013
6	48778518	1.12E-05	0.175	374	0.068	Lodging	2013
6	54898134	1.64E-05	0.483	372	0.086	Maturity	2013
8	40335883	9.70E-06	0.238	374	0.068	Lodging	2013
8	49431394	6.13E-06	0.140	372	0.075	Maturity	2013
9	9125531	1.08E-05	0.145	372	0.079	Maturity	2013
9	12634006	7.61E-08	0.154	374	0.003	Lodging	2013
9	13433421	2.53E-07	0.130	374	0.005	Lodging	2013
9	47615903	2.52E-06	0.107	453	0.031	Maturity	2012, 2013
9	57516609	4.27E-07	0.317	374	0.005	Lodging	2013

Table 2: Table of all significant SNPs at an FDR of 0.10 from the association analyses. The duplicate SNP (6-6484472) was identified when the data was analyzed for combined years 2012 and 2013 and also in 2013 only.

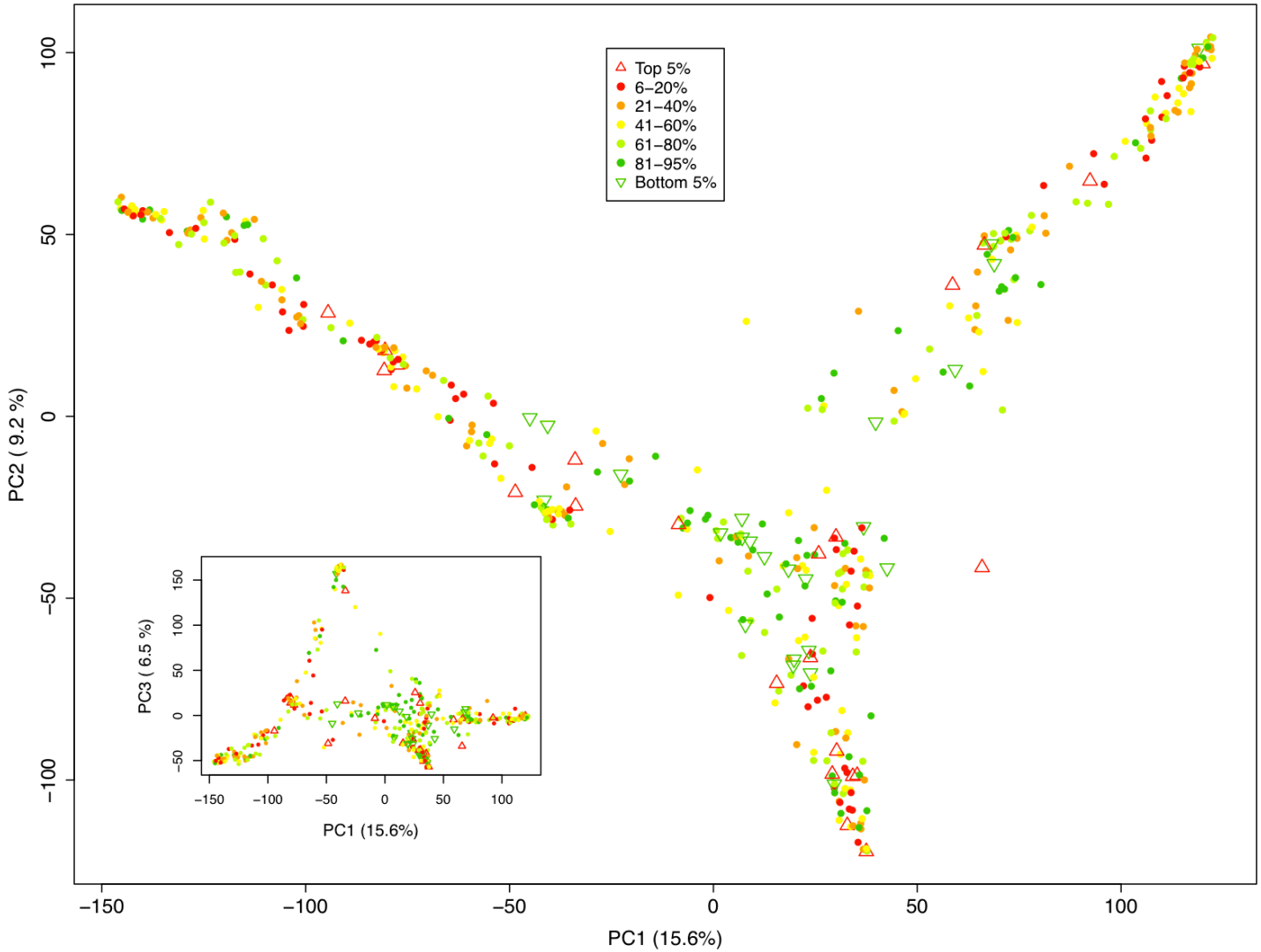

Figure 26: Principal component analysis of sorghum genetic structure and its relationship with the 30- day height phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

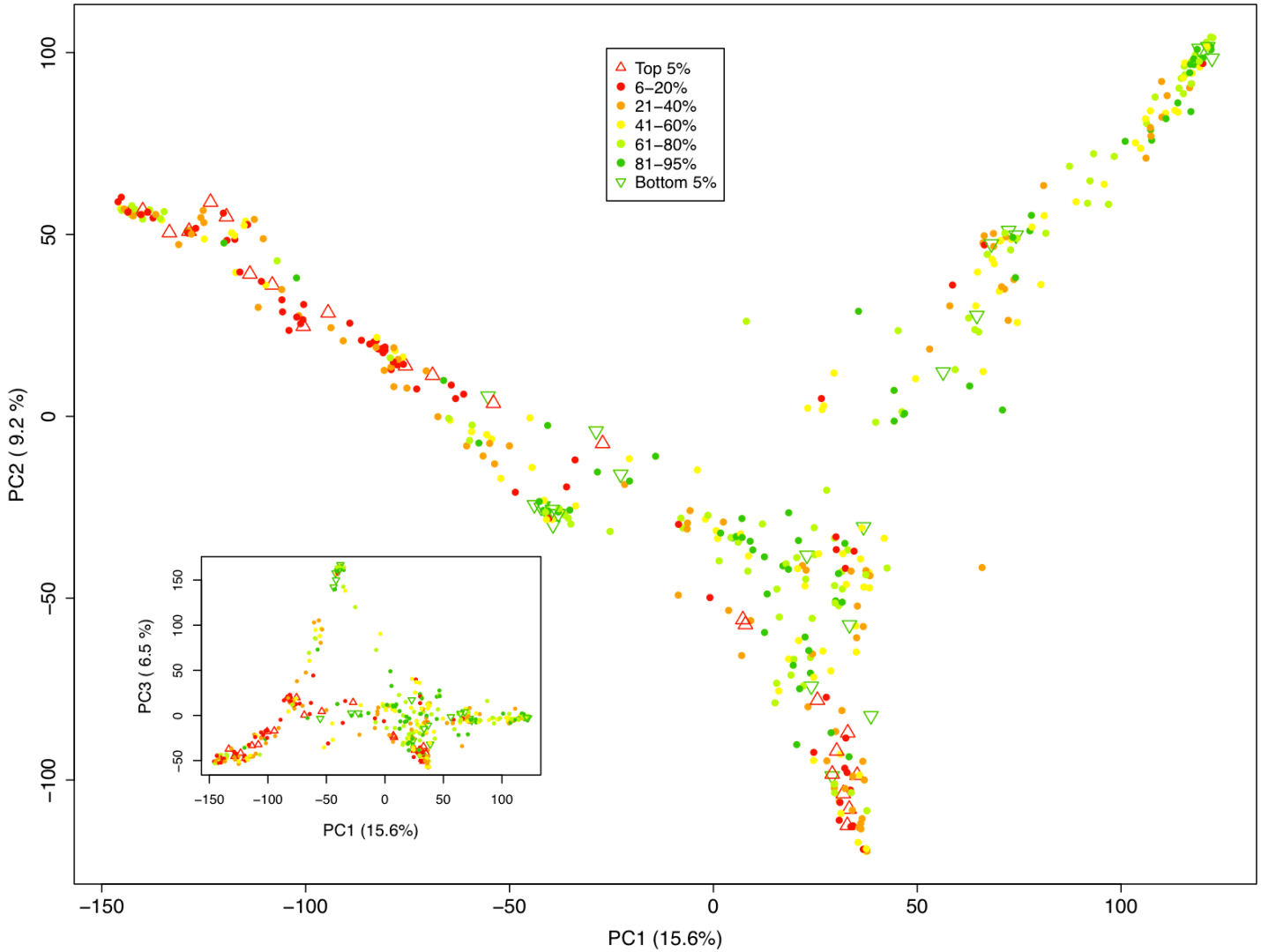

Figure 27: Principal component analysis of sorghum genetic structure and its relationship with the 60- day height phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

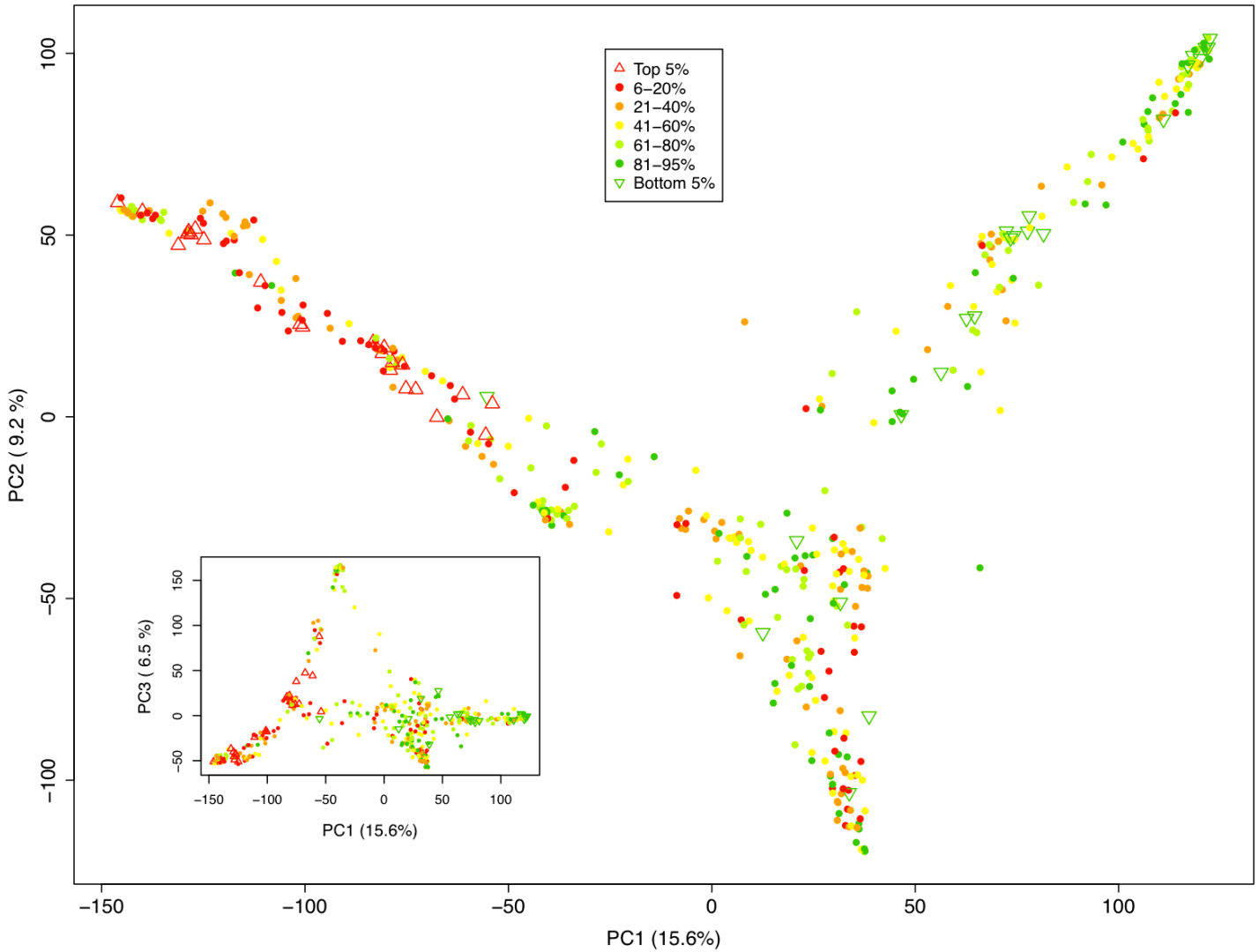

Figure 28: Principal component analysis of sorghum genetic structure and its relationship with the 120- day height phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

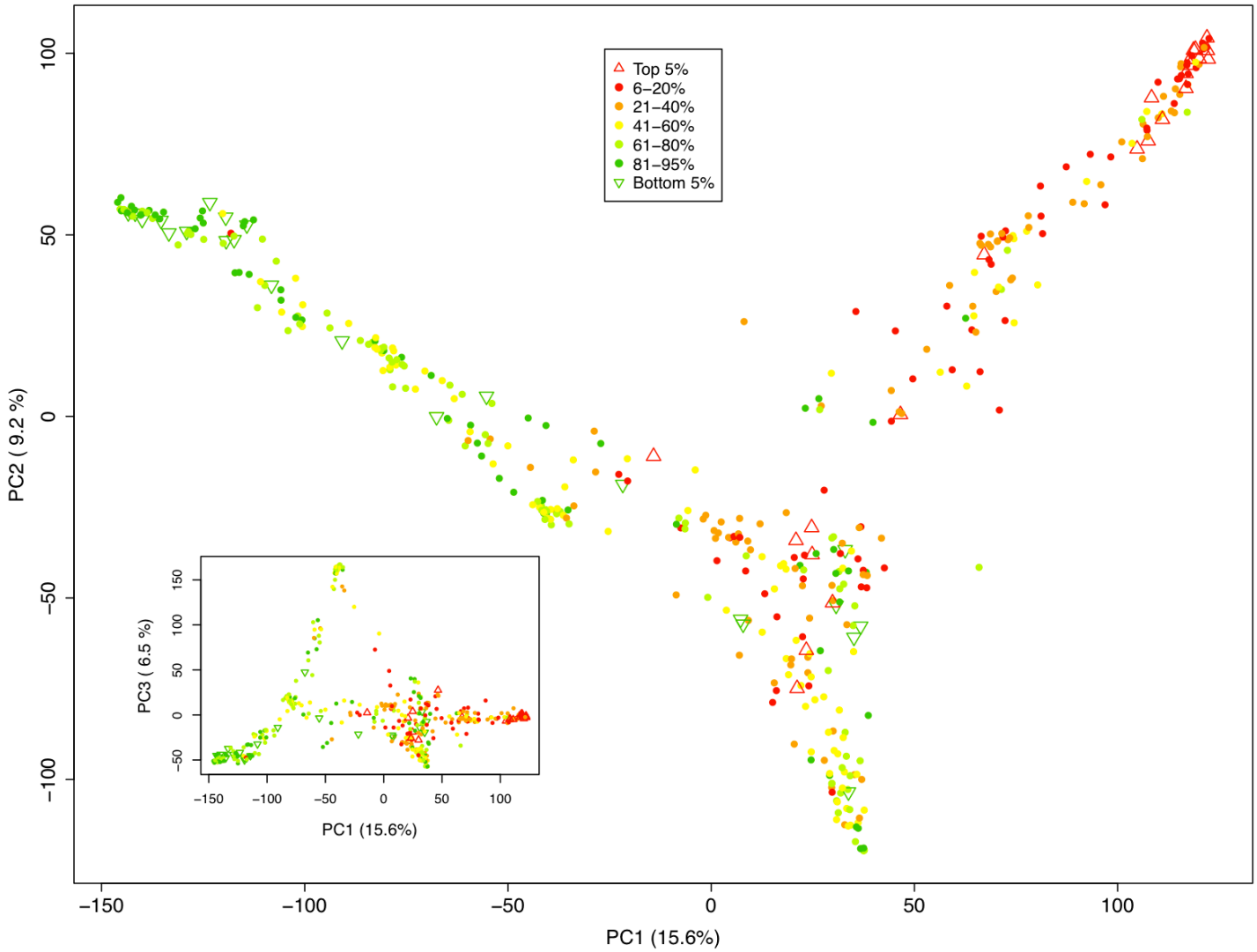

Figure 29: Principal component analysis of sorghum genetic structure and its relationship with the leaf length phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

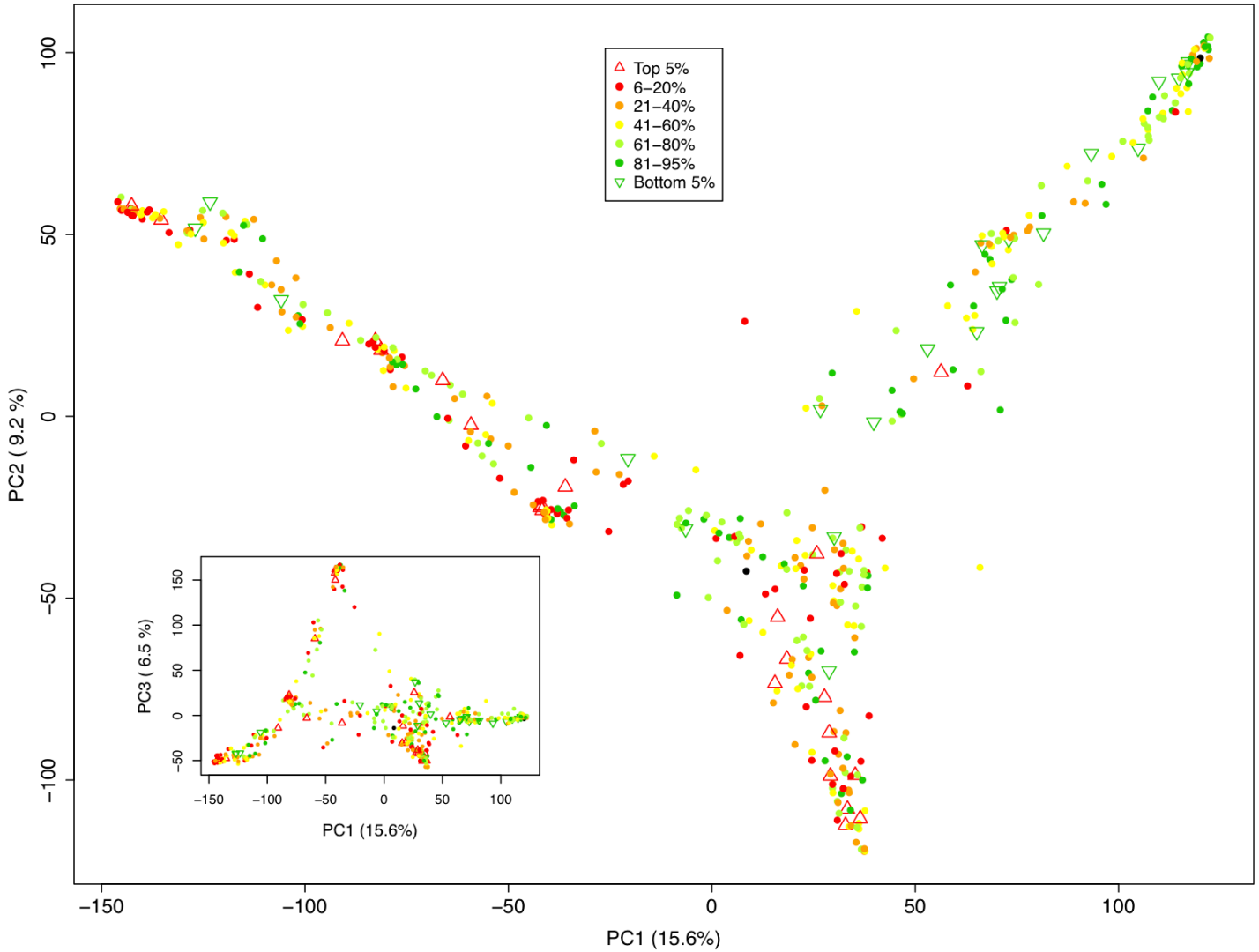

Figure 30: Principal component analysis of sorghum genetic structure and its relationship with the leaf width phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

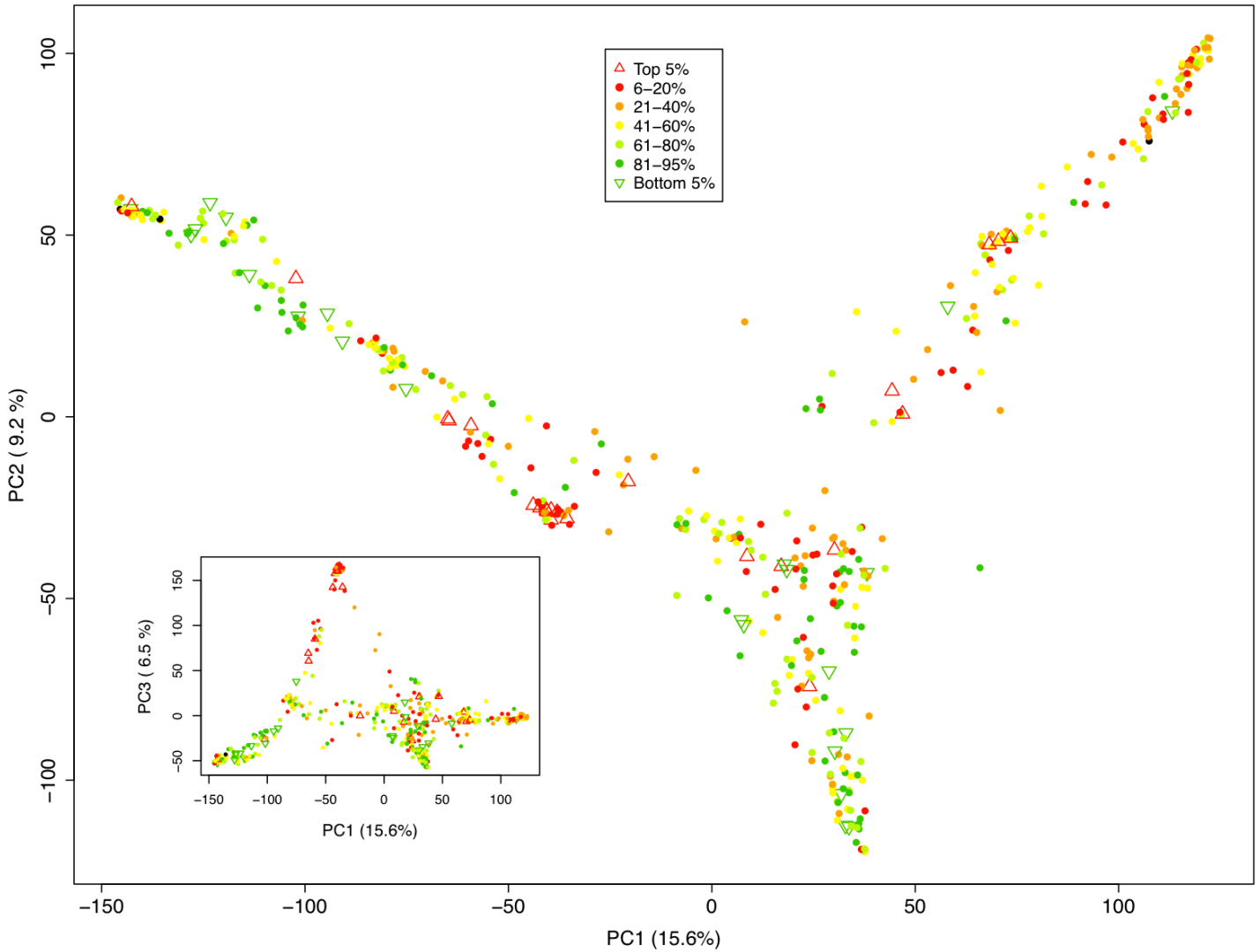

Figure 31: Principal component analysis of sorghum genetic structure and its relationship with the moisture phenotype. Each dot represents a sorghum inbred (n=456) colored according to its racial classification. 37,307 SNPs with minor allele frequencies greater than 10% were used in the analysis.

Family	Trait	Effect	Chr	Mb	LOD	% var. exp.
Tx2909	AUFPC	qtl1	1	7.769	6.45	30.2
Tx2909	AUFPC	qtl2	6	1.023	7.16	33.3
Tx2909	AUFPC	qtl1*qtl2	NA	NA	3.23	7.6
Tx2910	AUFPC	qtl1	3	72.933	5.13	10.4
Tx2910	AUFPC	qtl2	6	40.065	7.79	23.4
Tx2910	AUFPC	qtl3	9	48.703	4.56	8.6
Tx2910	Plant Stand	qtl1	3	69.936	6.48	18.1

Table 3: Table of all significant QTL detected through linkage mapping.