RODRIGO OLIVEIRA DE LIMA

LINKAGE ANALYSIS AND ASSOCIATION MAPPING FOR PLANT HEIGHT IN MAIZE

Thesis submitted to Federal University of Viçosa, in partial fulfillment of the requirements of the Genetics and Breeding Graduate Program, for the Degree of *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRAZIL
2013

Ficha catalográfica preparada pela Seção de Catalogação e Classificação da Biblioteca Central da UFV

T
L732L
2013

Lima, Rodrigo Oliveira de, 1982-
Linkage analysis and association mapping for plant height in maize / Rodrigo Oliveira de Lima. – Viçosa, MG, 2013.
x, 52 f. : il. (algumas color.) ; 29 cm.

Texto em inglês..
Orientador: José Marcelo Soriano Viana.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 26-33.

1. Milho - Genética. 2. Locos de caracteres quantitativos. 3. Milho - Pesos e medidas. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 633.152

RODRIGO OLIVEIRA DE LIMA

LINKAGE ANALYSIS AND ASSOCIATION MAPPING FOR PLANT HEIGHT IN MAIZE

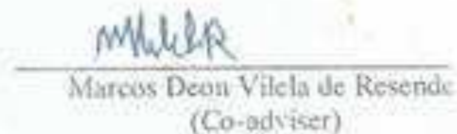
Thesis submitted to Federal University of Viçosa, in partial fulfillment of the requirements of the Genetics and Breeding Graduate Program, for the Degree of *Doctor Scientiae*.

Approved: 16 December 2013.

Lauro José Moreira Guimarães

Leandro Vagno de Souza

Fabyano Fonseca e Silva
(Co-adviser)


Marcos Deon Vilela de Resende
(Co-adviser)

José Marcelo Soriano Viana
(Adviser)

"I learned that courage is not the absence of fear, but the triumph over it. The brave man is not he who does not feel afraid, but he who conquers that fear."
(Nelson Mandela)

"The world is so competitive, aggressive, consumive, selfish, and during the time we spend here we must be all but that."
(José Mourinho)

To my parents, Geraldo and Norma,
my brother Ronaldo,
and my fiancée, Marli,
I dedicate it.

ACKNOWLEDGMENTS

I would like to thank:

God, for blessing my life and for allowing me to seek my goals and to realize my dream, to be Doctor. Thank You for having given me a fantastic family, my friends and for allowing me to live everything that I have lived until here.

My parents, Geraldo and Norma, for their love, care, help, worry, prayers and for sacrificing their dreams to help me realize mine. I cannot thank you enough for everything you have done for me. I love you so much. You are my best part.

My brother, Ronaldo for being my best friend. We lived our best time at EAF-UDI. Our first days there were very hard, but his fellowship made it possible. We had each other. I am never going to forget our moments together since our childhood in the farm in Santa Vitória, MG, until high school. I miss those moments. Thank you so much for everything. I am very proud of you.

My fiancée, Marli do Carmo Cupertino, for your love, care, friendship, understanding, advice and patience with me during all this time. In addition, I want to thank you for being my almost unique fellowship, through Skype, at American cold nights during my time in Madison, WI, USA. Even away you always encouraged me to work and study hard and realize my dream that was to study abroad. Thank you for everything. You are my soul mate. I will love you forever.

The Federal University of Viçosa, particularly to the Genetics and Breeding Graduate Program for the opportunity that it was provided to me these almost six years.

The Brazilian Federal Agency for Support and Evaluation of Graduate Education (Capes), and the National Council for Scientific and Technological Development (CNPq) for financial support and fellowships that allowed me to study 3 years and 10 months of my PhD at Federal University of Viçosa and 11 months at University of Wisconsin.

My advisor and friend Prof. José Marcelo Soriano Viana for your supervision, friendship, help, confidence, advice and every support that you have given me during my PhD course. Your supervision and friendship have been very important to my professional and personal growth. I have learnt very much with you. Thank you so much. I could not have done it without you.

My advisor at University of Wisconsin-Madison, Prof. Shawn Kaeppler for having accepted me in his research group, provided me with all the data for me to develop

and write my thesis, and supported me with everything. You always believed in my work and were an excellent advisor. To make part of my PhD in the USA was my dream and you helped me to make it possible. I am very grateful to you for helping me, Shawn. I would also like to thank my friend and "co-adviser" Prof. Natalia De Leon for helping me a lot during my trainee time at UW-Madison. We shared good moments in our research meetings and in the corn field during maize pollination time. Thank you ever so much.

All the professors who taught me during my PhD: Prof. Adair José Regazzi (Linear Models), Prof. Marcos Ribeiro Furtado (Population Genetics), Prof. Carlos Sigueyuki Sediyama (Molecular-Quantitative Genetics in Plant Breeding), Prof. Leonardo Lopes Bhering (Quantitative Genetics), Prof. Pedro Crescêncio Souza Carneiro (Biometric Models Applied to the Breeding II), Prof. José Marcelo Soriano Viana (Best Linear Unbiased Prediction in Plant Breeding) and Prof. Paulo Sávio Lopes (Mixed Models Applied to the Breeding). Thanks so much for having taught me so well. You will always be my teachers.

The members of the committee of my qualification examination: Prof. Marcos Ribeiro Furtado, Prof. Paulo Sávio Lopes, Dr. Lauro José Moreira Guimarães, Dr. Marcos Deon Vilela de Resende and Prof. José Marcelo Soriano Viana. Those five hours that we spent together discussing about BLUP, dominance, pedigree and unbalanced data were very important to increase my knowledge about the use of mixed models in annuals crops. Thank you for all the suggestions and advices.

The members of the committee of my thesis defense: Dr. Leandro Vagno de Souza, Dr. Lauro José Moreira Guimarães, Dr. Marcos Deon Vilela de Resende, Prof. Fabyano Fonseca e Silva, and Prof. José Marcelo Soriano Viana. Thank you for taking your time to read my thesis and for your valuable suggestions. Your contributions will be very important for me to improve it.

The secretaries from the Genetics and Breeding Graduate Program: Edna Maria de Oliveira, Rita Rosado Cruz and Marco Túlio Cardoso for being always available to help me. Thank you a lot for writing declarations, scheduling committee meeting, answering my questions and for all support.

The people from the Popcorn Program: José Márcio, Antônio Leonardo, Vinícius Faria, Magno Sávio, Ciro Maia, Geísa Pinheiro, Gabriel Mundim, Vinícius Almeida,

Hikmat Jan, Sardar Rahim and José Marcelo (the Boss). Thanks for helping me and sharing your time with me. We lived good moments together.

Prof. Glauco Vieira Miranda for having accepted me as trainee in the Maize Program on 20 October 2003. I could not have imagined that in that afternoon I was starting my maize breeder career.

The breeder and friend, Dr. Leandro Vagno de Souza, for having taught me the first concepts and methods about maize breeding, besides uncountable other teachings and advices. You were my first teacher in maize breeding. Thank you so much for sharing your knowledge with me and for uncountable hours working and playing together. I would also like to thank you so much for your help with my curriculum. Because of your professional influence my curriculum was delivered to Prof. Shawn Kaeppler and he invited me to go to her research group. I cannot thank you enough.

The friend and Prof. Roberto Fritsche-Neto for having encouraged me to be a professor of maize breeding.

The members and friends from the Maize Program from UW-Madison: Scott Stelpflug, German Muttoni, James Johnson, Marlies, Abadalla Zanouny, Joe Gage, Natalia de Leon and Shawn Kaeppler. Thank you so much for your support, friendship, patience and excellent moments together. You made my stay in Madison more comfortable.

Prof. Zilda Côrrea Lacerda, EAF-UDI, for having encouraged me to study at Federal University of Viçosa. You were my first advisor. Your advice were very important to me come to Viçosa and start my professional life.

The friends of life: Ronaldo Lima, Leandro Vagno, EugênioTeixeira, Flávio Maia, Vinícius Faria, Odirley Campos, Lauro Guimarães, Adérico Pimentel, Geraldo Ferreira, Izabel Figueiredo, Gabriel Mundim, Vinícius Almeida, Luís Marcos and Renato Ferreira. We do not need to have common parents to be brothers. Thank you for being part of my life.

Thanks a lot, people!

BIOGRAPHY

RODRIGO OLIVEIRA DE LIMA, son of Geraldo Francisco de Oliveira and Norma Oliveira de Lima, was born in Santa Vitória, MG, on August 4th, 1982.

In May 2001, he got the Agriculture Technician's certificate by Agro Technical Federal School of Uberlândia, Uberlândia, MG.

In January 2008, he graduated in Agronomy at Federal University of Viçosa, Viçosa, MG.

In March 2008, he began his Magister in Genetics and Breeding at Federal University of Viçosa under the supervision of Prof. Glauco Vireira Miranda. He had his Master's Degree in February 2010.

In March 2010, he began his PhD in Genetics and Breeding at the Federal University of Viçosa under the supervision of Prof. José Marcelo Soriano Viana.

SUMMARY

RESUMO ix

ABSTRACT x

1. INTRODUCTION...1

2. MATERIAL AND METHODS...6

2. 1. Plant Material 6

2.2. Field Evaluation and Phenotypic Analysis 6

2.3. Genotypic Data..8

2.4. Quantitative Trait Loci Linkage Mapping 10

2.5. Genome Wide Association Mapping...12

2.6. Candidate Gene Association Analysis...14

3. RESULTS..15

3.1. Phenotypic Assessment of the Populations for Maize Plant Height 15

3.2. Quantitative Trait Loci Mapping..15

4. DISCUSSION..21

5. CONCLUSIONS...25

7. TABLES AND FIGURES...34

8. APPENDICES 40

8.1. Supplemental Figures 40

8.2. Supplemental Tables:...42

RESUMO

LIMA, Rodrigo Oliveira de, D.Sc., Universidade Federal de Viçosa, dezembro de 2013. **Análise de ligação e mapeamento associativo para altura de planta em milho.** Orientador: José Marcelo Soriano Viana. Coorientadores: Marcos Deon Vilela de Resende e Fabyano Fonseca e Silva.

Altura de planta é um dos caracteres mais estudados em milho e está relacionado com produtividade de grãos, densidade de plantas e acamamento. Vários *quantitative trait loci* (QTL) para alttura de planta foram encontrados em milho. Entretanto, sua arquitetura genética ainda permanece desconhecida. Deste modo, o objetivo desse estudo foi explorar a arquitetura genética e fenotípica de altura de planta em milho por meio de análise de ligação e mapeamento associativo. Foram avaliados dez caracteres relacionados à altura de planta em 962 linhagens endogâmicas recombinantes (RILs) de quatro populações e 400 linhagens endogâmicas do painel de mapeamento associativo WiDiv. As linhagens endogâmicas do painel de associação WiDiv foram genotipadas com 458 151 SNP (*single nucleotide polymorphism*) através do sequenciamento de RNA (RNAseq). As populações RILs IBM, OW e NyH foram genotipadas com 8 224, 6 696 e 5 320 SNP por meio da metodologia de "genotyping-by-sequencing". A população NAM22 foi genotipada com 1 200 SNP usando a plataforma de 1 536 SNP da Illumina. Altura de planta de milho pode ser fenotipicamente decomposta em altura de espiga, número de internódios e caracteres de florescimento. Os caracteres número de internódios, número de internódios abaixo da espiga e caracteres de florescimento apresentaram elevada correlação entre eles. A altura de planta de milho pode ser geneticamente dissecada por meio de análise de ligação e mapeamento associativo. Vários QTL foram associados com caracteres de altura de planta, e alguns QTL foram identificados na mesma região genômica para altura de planta, altura de espiga e número de internódios. Alguns SNP localizados no gene candidato GRMZM2G171622 foram associados com seis caracteres de altura de planta. Os genes candidates GRMZM2G108798 e GRMZM2G012766, que foram identificados no painel de associação WiDiv, foram associados com caracteres de altura de planta em milho. Os resultados sugerem que altura de espiga, número de internódios, número de internódios abaixo da espiga, comprimento médio de internódio e caracteres de florescimento são os principais componentes de altura de planta de milho.

ABSTRACT

LIMA, Rodrigo Oliveira de, D.Sc., Universidade Federal de Viçosa, December, 2013. **Linkage analysis and association mapping for plant height in maize.** Adviser: José Marcelo Soriano Viana. Co-advisers: Marcos Deon Vilela de Resende and Fabyano Fonseca e Silva.

Plant height is one of the most studied traits in maize. It is related to grain yield, planting density and lodging resistance. Several quantitative trait loci associated with plant height have been found through linkage mapping and association mapping in maize. Nevertheless, plant height has not been yet phenotypically and genetically dissected to its components (i.e. internode length and node number) and its genetic architecture remains unknown. Thus, the objective of this study was to explore the phenotypic and genetic architecture of plant height in maize by linkage analysis and association mapping in four RILs populations and in a diverse association panel of maize inbred lines. We evaluated ten maize plant-related traits in 962 recombinant inbred lines derived from four populations and 400 inbred lines from Wisconsin Diverse population (WiDiv). The WiDiv population was genotyped with 458,151 single nucleotide polymorphism (SNP) discovered through RNA-sequencing of maize seedlings. The IBM, OW and NyH RILs populations were genotyped with 8,224, 5,696 and 5,320 SNP using the genotyping-by-sequencing methodology. The NAM22 was genotyped with 1200 SNPs using the panel of 1,536 SNPs by Illumina. Maize plant height-related traits can be phenotypically into ear height, node number, average internode length, and flowering time traits. Plant height showed strong correlation with almost all traits. Node number, below ear node number and flowering time traits were strongly correlated to each other. Maize plant height can be genetically dissected by linkage analysis and association mapping. Several QTL were associated with plant height-related traits, and some QTL hot spots were identified for plant height, ear height and node number. Single nucleotide polymorphism, located in the gene model GRMZM2G171622, was associated with six traits. The candidate genes GRMZM2G108798 and GRMZM2G012766 which were identified in WiDiv population panel were associated with plant height-related traits in our study. These results suggest that ear height, node number, below ear node number, below ear average internode length, and flowering time are the main contributors to maize plant height.

1. INTRODUCTION

Maize (*Zea mays* ssp. *mays* L.) is the third world's most important production crop and will most likely become the most important crop by 2020 (faostat.fao.org). Plant height is one of the most studied traits in maize. It is important from a genetically and from a plant breeding viewpoint due to its association with adaptation and agronomic importance. It is significantly correlated to grain yield (Beavis et al., 1991; Lima et al., 2006). Flowering time, another key trait is also related to plant height, in which early genotypes tend to be shorter. An extreme example of this relation is the Gaspé Flint variety which is virtually the earliest maize line and is about 1 m tall (120 cm shorter than B73) (Salvi et al., 2011). Additionally, from a plant breeding standpoint, it is desirable to breed for shorter plants (and shorter ear height) to reduce stalk lodging and incidence of root lodging (Duvick et al., 2005). In another way, recently, the development of taller plants has been proposed as a mechanism to increase biomass production in bioenergy crops such as maize and sorghum (Salas Fernández et al., 2009) due to the high correlation observed between these two traits (Lübberstedt et al., 1997, Yuan et al., 2008).

Maize plant height can be biologically comprised of two main components: internode number and average internode length. Lima et al. (2006) evaluated a tropical maize population to plant height and found that plant height presented significant correlations with ear height and node number. Ji-hua et al. (2007) evaluated the maize plant height components and showed that plant height had a strongest correlation to average internode length and was weakly correlated to node number. Recently, Salvi et al. (2011) employed three populations derived from cross between Gaspé Flint and B73 to dissect maize phenology in 6 morphological traits. Across three populations, plant height was positively correlated with all traits: days to pollen shed, average internode length, node number, above ear node number, and below ear node number. In another study, Teng et al. (2013) did cytological observation to investigate the internode cells size of maize and concluded that longer internodes in a genotype relative to another are caused by increased of cell lengths, not increased cells numbers. Thus, it seems that the difference among maize genotypes to plant height is further because average internode length than node number. However, there are still a few reports that have analyzed the components contributing to maize plant height.

Plant height is a complex and complicated quantitative trait controlled by a large number of genes of small effect. Quantitative trait loci (QTL) mapping techniques are commonly used to characterize the genetic architecture of complex traits and represent a powerful tool to identify genes that underlines them (Price, 2006; Mackay et al., 2009). Additionally, QTL studies may provide new impetus and opportunities for more targeted selection programs based on marker-assisted selection of the relevant QTL (Bernardo, 2008; Yu and Crouch, 2008). Beavis et al. (1991) carried out the first study with identification of QTL for plant height using four maize populations. They showed that most of the QTL detected for plant height were in close proximity to know qualitative trait loci for the trait. After the first investigation, several other studies have been done to identify QTL for maize plant height. To date, using different mapping populations, more than 219 QTL for maize plant height have been reported in numbers of studies (2013 December update to Gramene database). Although there are many QTL identified to plant height, a few reports have identified QTL for its components. For example, only 26 QTL have been reported for ear height in maize (2013 December update to Gramene database). Ji-hua et al. (2007) used a maize population of 294 recombinant inbred lines (RILs) to genetically dissect the plant height in its components. They reported six QTL for plant height, seven for node number and six for average internode length. Four of six QTL identified for average internode length were located in the same region of QTL affecting plant height suggesting that average internode length was the main contributor to maize plant height. More recently, Salvi et al. (2011) dissected the maize phenology using an intraspecific introgression library. Five QTL for flowering time were mapped, all corresponding to major QTL for node number. Additionally, the QTL for node number drove phenotypic variation for plant height node number above and below ear, but not for internode length.

Linkage mapping conducted with populations that are derived from a bi-parental cross has been a key tool for studying the genetic basis of quantitative traits in plants. However, even though hundreds of QTL have been identified for maize plant height, few of identified QTL were validated and cloned or tagged at gene level (Teng et al., 2006). This happens because QTL mapping typically localizes QTL to 10 to 20 cM intervals because limited number of recombination events that occur during the construction of mapping populations (Holland, 2007). Thus, the approaches called association mapping,

also known as linkage disequilibrium (LD) mapping offers an alternative method for mapping QTL (Soto-Cerda and Cloutier, 2012). Association mapping is a population-based survey used to identify marker-trait association based on LD between markers and functional polymorphism across a set of diverse individuals (Flint-Garcia et al., 2003). As it exploits the ancestral recombination and natural genetic diversity within a population, it offers higher resolution and finer mapping than linkage mapping (Zhu et al., 2008; Oraguzie and Wilcox, 2009). Thus, genetic dissection of complex traits with association mapping is a promising strategy (Yu and Buckler, 2006; Yu et al., 2008; Lu et al., 2010). The application of association mapping includes candidate-gene and genome scan or genome wide association testing. Candidate-gene tests polymorphism within specific genes, while genome wide association mapping or genome scan, which markers are placed across the genome, surveys genetic variation in the whole genome to identify specific functional variants linked to phenotypic differences for complex traits (Mackay, 2001; Risch and Merinkngas, 1996; Kruglyak, 1999).

Whole genome scan and candidate-gene approaches have been applied successfully to dissect and understand the genetic mechanism of complex traits in maize over last decade. In first investigation in maize, Thornsberry et al. (2001) applied candidate-gene approaches to evaluate DNA polymorphism within the *Dwarf8* locus across 94 inbreed lines and found nine polymorphism associated with flowering time. Recently, Weng et al. (2011) investigated the association of genetic variants and plant height of 284 Chinese maize diverse inbred lines genotyped with 55,000 SNPs. A total of 204 SNPs across 10 chromosomes were significantly associated with plant height. Riedelsheimer et al. (2012) applied genome scan approaches in a set of 289 diverse maize inbred lines genotyped with 56,110 SNPs to dissect leaf metabolic profiles. Significant SNP-metabolite associations were found on all chromosomes with SNPs explaining up to 32% of the observed genetic variation. In another investigation, Li et al. (2013) examined the genetic architecture of oil biosynthesis in maize kernels by genome wide association mapping using 1.03 million of SNPs characterized in 368 diverse inbred lines. They identified 74 loci significantly associated with oil composition and concentration, and the 26 loci associated with oil concentration explained up to 83% of phenotypic variation. Xue et al. (2013) employed association mapping to dissect drought tolerance in 350 maize inbred lines genotyped with 1,536-SNP developed from drought-related genes and 56,110

random SNP. Forty-two associated SNPs were identified located in 33 genes. Of these genes, three were co-localized to drought-related QTL regions.

Although association mapping have been extensively used to examine the genetic architecture of complex trait in several plant and animal species, it have presented some limitations due the effect of population structure and genetic relatedness among individuals in the population. The presence of these effects in the population often cause the identification of spurious associations reducing of power of association mapping (Yu and Buckler, 2006; Yu et al., 2006). These effects can be isolated from association mapping testing by correction of the population structure using the STRUCTURE program (Pritchard et al., 2000a, 2000b) and principal components analysis (Price et al., 2006), and by estimation of pairwise relatedness individuals using random molecular markers (Yu et al., 2006). In addition to the effects previously cited, most alleles are rare in a large proportion of individuals from diverse association panel limiting the power of mapping (Myles et al., 2009), and, consequently, part of heritability remains undetected or missed. Linkage analysis can break the correlation between population structure and phenotypes and rare allele frequencies can be inflated to enhance the QTL detection (Stich and Melchinger, 2010). Thus, joint linkage and association analysis has been proposed as alternative strategy to overcoming some limitations factors of both association and linkage approaches. Lu et al. (2010) used joint linkage-LD mapping to detecting QTL underlying maize drought tolerance in three RIL population and 305 diverse inbred lines that were achieved using 2,052 SNPs markers. Integrated mapping detected 18 additional QTL not identified by parallel mapping (independent linkage and LD analysis).

To integrate to advantages of linkage and association mapping, some special association population mapping have been designed to study and understand the genetic mechanism of complex traits. In maize, nested association mapping population (NAM) was created through controlled crosses between 25 diverse inbred lines and the reference inbred line B37, and 200 RILs were derived from each crosses, then producing ~ 5,000 RILs (Yu et al., 2008). This kind of design reduces spurious associations caused by population structure and inflates the rare allele frequency providing more power and resolution for QTL detection (Buckler et al., 2009; McMullen et al., 2009). Joint linkage mapping and genome wide association study (GWAS) in the NAM have been used as a

powerful strategy for resolving complex traits to their causal loci in maize over last three years. In maize NAM, Tian et al. (2011) identified associations yield by GWAS overlapped significantly with joint-linkage QTL intervals for leaf traits. They determined the genetic basis of leaf traits and identified some key genes. Kump et al. (2011) examined the quantitative resistance to southern leaf blight (SLB resistance) in the NAM. Thirty-two QTL with predominantly small and additive effects were identified by joint linkage analysis and 245 SNPs yield by GWAS both within and outside of QTL intervals were associated with variation for SLB resistance. In another study, Cook et al. (2012) dissected the genetic architecture of kernel composition by joint-linkage and GWAS in the NAM, and identified 21-26 QTLs explaining around 60% of the total variation, and 118-135 SNPs associated with kernel composition. Of these SNPs, between 47% and 100% overlapped with joint-linkage QTL intervals, and numerous associations genes that regulate oil composition and quality were identified. More recently, Peiffer et al. (2013) inferred the genetic architecture of maize stalk strength – measured in rind penetrometer resistance (RPR) - in the NAM. Twenty-three QTL mapped to RPR explained 81% of the phenotypic variation, and three QTL intervals overlapped with three local annotations genes that are related to cellulose synthase. In addition to QTL, 141 significant SNPs associated to RPR were identified by GWAS and the most robust associations co-localized with estimated joint-linkage mapped QTL effects.

Considering the foregoing, the objective of this study was to explore the phenotypic and genetic architecture of plant height in maize by linkage analysis four RILs populations and association mapping in a diverse association panel of maize inbred lines.

2. MATERIAL AND METHODS

2. 1. Plant Material

Four recombinant inbred line (RIL) populations and an association panel were used in this study. The OW, NyH, and NAM22 RILs populations, consisting of 255, 233 and 192 RILs, respectively, were derived from the crosses between the inbred lines, Oh43 and W64a (OW), and Ny821 and H99 (NyH), and Oh43 and B73 (NAM22), using the single-seed descendent method. The intermated population, IBM, was derived from the single-cross hybrid of inbred lines B73 (female) and Mo17 (Lee et al., 2002). One plant F_1 was self-pollinated to produce the F_2 generation. In the F_2 generation, plants were used once as male or female, in a cross with another plant so that 250 pairs of plants were mated. A single seed was taken of the other ears to form the Syn1 generation. Then, the plants were randomly intermating for four additional generations to produce the Syn5 generation. So, 282 RILs were produced by self-pollinating 300 plants in Syn5 generation using the single seed descendent method. The random intermated for some generations previously to derive the inbred lines to mapping improved genetic resolution significantly by providing additional opportunities for recombination prior to development of the mapping progeny in maize (Beavis et al., 1994).

The WiDiv association panel was recently developed to dissection of complex traits in maize by association mapping (Hansey et al., 2011). Briefly, a total of 1,411 diverse inbred lines from North Central Regional Plant Introduction were evaluated in the upper Midwestern United States, and a set of 627 lines were chosen based on flowering time within the desired interval, production of viable seed, agronomic suitability, uniformity, and pedigree information. Thus, the WiDiv association panel is a set of lines with restricted phenology that maintains genetic and phenotypic diversity allowing assessment of phenotypes in short-season environment, and exploits this diversity in association mapping studies (Hansey et al., 2011).

2.2. Field Evaluation and Phenotypic Analysis

The five maize populations were grown at Arlington Agriculture Station in Arlington, Wisconsin, USA. The 282 IBM RILs were evaluated over three years, in 2009 and 2010 with two replications in each year , and in 2011 in an unreplicated trial; the 255 OW RILs were evaluated over two years, in 2008 with 2 replications and in 2011 in an

unreplicated trail; the 233 NyH RILs were evaluated in 2011 in an unreplicated trial; and 192 NAM22 RILs were evaluated over two years, in 2009 and 2011 in unreplicated trials; and a set of 400 inbred lines from the WiDiV population were evaluated over four years, in 2008, 2009, 2010 and 2011with two replications in each year. For those years which populations were evaluated in trials with replications, the field experiment followed a randomized complete block design. Each plot included one row that was 3.6 m long and 0.76 m spacing with a density of 58,000 plants per hectare.

At the flowering stage, days to pollen shedding (DP) and days to silking (DS) were measured as the number of days from planting to the initiation of pollen shed or silk emergence for half plants within a plot. At the 20th day after flowering, five representative plants were selected from each plot and evaluated for plant height (PH, cm), ear height (EH, cm), average internode length (IL, cm), node number (NN, numbers), below ear average internode length (BEIL, cm), above ear average internode length (AEIL, cm), below ear node number (BENN, numbers) and above ear node number (AENN, numbers). PH and EH were measured from the soil surface to the collar of the flag leaf and to the uppermost ear node, respectively. The NN was defined as the number of elongated above-ground internodes. The IL was characterized as average length of the above-ground internodes, and was calculated as plant height divided by node numbers. BEIL and BENN were defined as IL and NN measured below the uppermost ear. AEIL and AENN were defined as IL and NN measured above the uppermost ear.

The plot means were used to calculate the best linear unbiased predictors (BLUPs). The BLUPs were predicted by fitting the mixed linear model in R package “lme4” for all inbred lines (Team, 2001). The mixed model included the overall mean, inbred line, year, replication, inbred line by year interaction, and random error effect. All effects of the model, except the overall mean, were considered random. Variance components were estimated by restricted maximum likelihood method (REML). Model assumptions (normality of residuals, homogenous variance of residuals and normality of random effects) were assessed in each model. Narrow-sense heritability (on an entry-mean basis) coefficients of measured traits were estimated as described by Falconer and Mackay (1996): $\hat{h}^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + (\hat{\sigma}_{GE}^2/y) + \hat{\sigma}_e^2/ry}$, where $\hat{\sigma}_A^2$ is the estimate of additive genetic variance, $\hat{\sigma}_{GE}^2$ is the estimate of the genotypic-by-environment (year) interaction variance, $\hat{\sigma}_e^2$ is the estimate of the error variance, y is the number of years, and r is the number of

replications. For the 192 NAM22 RILs were evaluated in an unreplicated trial over two years, the narrow-sense heritability coefficients for PH and EH among two years was estimated as: $\hat{h}^2 = \frac{\hat{\sigma}^2_{GA}}{\hat{\sigma}^2_A + \hat{\sigma}^2_y/y}$ where $\hat{\sigma}^2_A$ is the estimate of additive genetic variance, $\hat{\sigma}^2_y$ is the estimate of the year variance, y is the number of years. As NyH RILs were evaluated in an unreplicate trail and only one year, heritability coefficients were not estimated for PH and EH. In addition to Nyh, the heritability was not estimated for IL, NN, BEIL, AEIL, BENN, and AENN in the OW population because these traits were evaluated only in 2011 in an unreplicated trial. The BLUPs for all inbred lines were used to estimate the genotypic correlation matrix among traits using the Pearson's correlation coefficient by R package 'agricolae' and function *correl* (Team, 2001).

2.3. Genotypic Data

The inbred lines from five maize populations used in this study were genotyped using different methodology to identify SNPs. The IBM, OW and NyH RILs populations were genotyped using the genotyping-by-sequencing (GBS) methodology. GBS is a simple highly-multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform (Elshire et al., 2011). It generates large numbers of SNPs for use in genetic analyses and its advantages are: reduced sample handling, fewer polymerase chain reaction and purification steps, no size fractionation and inexpensive barcoding. Moreover, it uses restriction enzymes to reduce genome complexity and avoid the repetitive fraction of the genome. The result is tens to hundreds of thousands of genotyped SNP markers, ready to analyze.

For each of the RIL and the parents out of each population (IBM, OW and NyH), above ground seedling tissue was harvested from five to 10 plants and pooled. DNA was isolated using a modified CTAB method (Saghaimaroof et al., 1984), and subsequently barcoded and pooled according to GBS protocol previously developed (Elshire et al., 2011). Additionally, size selection for fragments of approximately 300 base pairs in length was performed. Parental and recombinant inbred lines were pooled at either 16 or 48 DNA samples per library. Sequencing was done using the Illumina Genome Analyzer II (San Diego, CA) and the Illumina HiSeq 200 (San Diego, CA) at the University of Wisconsin-Madison Biotechnology Center (Madison, WI) and the Department of Energy Joint Institute (Walnut Creek, CA). Single end reads between 74 and 100 base pairs were

generated. Read quality in the multiplexed libraries was evaluated based on phred-like quality scores. To produce the genetic map, pooled reads were cleaned using the fastx_clipper program within the FASTX toolkit (available at: http://hannonlab.cshl.edu/fastx_tookit?index.html [accessed May 2013]). The minimum sequence length was set to 15bp after clipping using both Illumina single end adapter sequences. A custom Perl script was used to parse sequence reads into individual genotype files requiring a perfect match to the barcode and *Apek*KI restriction enzyme cut site (GC[A/T]) and the barcode sequence were removed. Reads from each genotype were then mapped to the maize AGPv2 5b pseudomolecules (Schnable et al., 2009) using Bowtie version 0.12.7 (Langmead et al., 2009) requiring a unique alignment and following for up to two mismatches. SAMtools version 0.1.7 (Li et al., 2009) was used to generate unfiltered pileup files. Custom Perl scripts were also used to determine genotype calls and identify SNPs within each population. At each locus, at least one to the two parents had to have read coverage, and the reads had to support a single consensus base in both parents the locus needed to be polymorphic between the parents. Within the population, at least five of the RILs were required to have information. Additionally, only two alleles could be present at greater than 10% frequency across the population and the two alleles were required to be congruent with the parental consensus calls. When information was presented in only one of the parents the alternate parent genotype score was inferred from the population. Using GBS technology the 282 IBM, 255 OW and 233 NyH RILs were genotyped and the resulting recombination bin map contained 8,224, 5,696 and 5,320 highly informative bin markers.

In relation to NAM22 RIL population, the genotyping have been previously described (Yu et al., 2008, McMullen et al., 2009). Briefly, tissue from up to four etiolated seedlings were harvested per line, lyophilized, ground in a Gen/Grinder 2000 (BT&C/OPS Diagnostic, Bridgewater, NJ) and then extracted with a standard (Doyle and Doyle, 1987) cetyltrimethyl ammonium bromide extraction procedure. DNA was quantified using PicoGreen reagent Quant-iT™ Pico Green® dsDNA reagent (Invitrogen, Carlsbad, CA). The genotypes of all RILs were determined using the panel of 1536 SNPs by Illumina GoldenGate Assay Sistem (Fan et al., 2003; Illumina, San Diego, CA). Of out 1,536 SNPs, 974 were chosen from random genes, 329 were chosen from candidate genes

and 233 were chosen from alignments of sequence of inbred lines provided by Pionner Hi-bred International. The markers were chosen to maximize information relative to B73.

The 400 inbred lines from WiDiv population were genotyped by RNA-sequencing (RNAseq) of the two weeks old maize seedlings. Plants were grown under controlled greenhouse conditions (27°C/24°C day/night and 16 h light/8 h dark) and six seeds per inbred line were planted in each pot using Metro-Mix 300 (Sun Gro Horticulture). The whole seedling tissue collected included the roots and were collected at the vegetative stage V1 (Ritchie et al., 2008). Three plants per inbred line were pooled into a single tissue sample. RNA was isolated using TRIZOL (Invitrogen, Life Technologies) and purified with the RNeasy MinElute Cleanup kit (Qiagen). RNAseq libraries were prepared for each inbred line and sequenced on the Illumina HiSeq (San Diego, CA). To identify SNPs, sequence reads for each library were mapped to version 2 of the maize B73 reference sequence AGPv2 5b (Schnable et al., 2009) using Bowtie version 0.12.7 (Langmead et al., 2009) and TopHat version 1.4.1 (Trapnell et al., 2009). To determine the genotype of an individual at a given position, a minimum of five reads were required. Additionally the reads had to support a single allele, where at least 5% of the reads and at least two reads were required for support. A locus was considered polymorphic if at least two alleles had greater than 5% allele frequency. Thus, a total of 458,181 SNPs were discovered through RNAseq in the WiDiv inbred lines, and, then, they were filtered to keep only di-allelic SNPs, which resulted in 438,222 SNPs. These SNPs were imputed with the population-based haplotype clustering algorithm fastPHASE version 1.4.0 (Scheet and Stephens, 2006) with default settings, except for having fixed the number of clusters, which was set to 20 (i.e. K = 20). Before the imputation the percentage of missing data was 29.4. So, each one of the ten chromosomes was divided into four subset of equal number of markers to make the files manageable within the R program. After the imputation the percentage of missing data was 0.56 (Supplemental Table 1).

2.4. Quantitative Trait Loci Linkage Mapping

Quantitative trait loci (QTL) for maize plant height and its components were mapped in the IBM, OW, NyH and NAM22 RILs populations. Before QTL map, we did several quality control measures in the genotypic data to ensure accurate QTL mapping. We checked for segregation distortion, markers with too many failed individuals, markers

with identical genotypes and duplicated markers, and we removed the markers problems. After that, we use 8,175, 5,288 and 5,683 highly informative bin markers to construct the genetic map in the IBM, OW and NyH RIL populations, respectively, and 1,106 SNPs were used to constructed the genetic map in the NAM22 population. The genetic maps were constructed using the arguments *estimate.map* function in R/qtl assuming the genotyping error rate of 1 x 10^{-4} (Broman and Sen, 2009). The *estimate.map* uses the Lander-Green algorithm (i.e., the hidden Markov model technology) to estimate the genetic map for an experimental cross. The Kosambi map function (Kosambi, 1943) was used to convert recombination fractions into genetic distance. The *jittermap* function in R/qtl was used to move apart slightly two markers that were placed at precisely the same position, and the *ripple* function was used to check the order of markers on the genetic map.

The majority of QTL approach like simple interval mapping and composite interval mapping test the existence of a QTL at several points of genome, one at time, at several points of genetic map based on single QTL methods (Lander and Botstein, 1989; Jansen, 1994; Zeng, 1993). Thereby, the single QTL methods indicate individual pieces of the complex genetic architecture that underlies the phenotype. To evidence for the QTL these pieces should be bring together in a single coherent model called multiple QTL method (MQM) (Jansen, 1994; 2007). It combines the strengths of generalized linear model regression with those of interval mapping and, tests simultaneously the presence of several QTLs in the genome (Jansen and Stam, 1994) allowing more powerful and precision detection of QTL than many other methods (Arends et al., 2010). Thus, MQM provides a sensitive approach for mapping QTL in experimental populations, especially for complex traits that are controlled by many genes like maize plant height.

Considering the foregoing, models incorporating multiple QTL were fit using the *stepwiseqtl* function in R/qtl (Broman et al., 2003) which uses a forward/backward search stepwise algorithm to select the best-fitting model (Manichaikul et al., 2009; Arends et al., 2010). Multiple QTL mapping was done with an upper limit of 15 QTL contemplating only additive models. Mapping was performed using the Haley-Knott regression method (Haley and Knott, 1992). Genotypic probabilities were calculated only at marker locations with a genotyping error rate of 1 x 10^{-4}, and Kosambi map function (Kosambi, 1943) was used when converting genetic distance into recombination fractions. The

models were compared using the penalized LOD score (Manichaikul et al., 2009) with thresholds calculated from the 10,000 permutations using the *scanone* function in R/qlt and controlling the type-I error rate (α=0.05). The final chosen model was that the maximal penalized LOD score among the models visited (Broman and Sen, 2009). After the model with the QTLs was chosen, the *fitqtl* function in R/qtl was used to obtain the estimates of QTL effects. QTL support intervals were estimated with the *lodint* function in R/qtl by setting a drop of 1.5 in LOD scores from QTL peaks. As we have the relation between genetic mapping distance and physical mapping based on the B73 Reference Genome Sequence Version 2 (B73 RefGen_v2), the physical distance were used to compare the QTL interval overlapping with QTL identified in another population and to compared the QTL interval overlapping between plant-related traits.

2.5. Genome Wide Association Mapping

Genome wide association studies were performed in WiDiv population using the linear mixed model proposed by Yu et al. (2006). In this model each marker (a fixed effect) is evaluated individually one at a time (i.e. single marker regression). The equation of this model can be expressed as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Wm} + \mathbf{Qv} + \mathbf{Zu} + \mathbf{e}$$

where: $\mathbf{y}$ is a vector of phenotypic observations; $\boldsymbol{\beta}$ is a vector of fixed effects other than SNP under testing or population group effects; $\mathbf{m}$ is a vector of SNP effect under evaluation (fixed effect); $\mathbf{v}$ is a vector of subpopulation effects (fixed effect); $\mathbf{u}$ is a vector of polygene background effects (random effects - proportion of the breeding values not accounted by the SNP marker); $\mathbf{e}$ is a vector of residual effects; $\mathbf{Q}$ is an incidence matrix of principal component scores (eigenvectors) that relates the subpopulation levels to the subpopulation effects; and $\mathbf{X}$, $\mathbf{W}$ and $\mathbf{Z}$ are incidence matrices of ones and zeros relating $\mathbf{y}$ to $\boldsymbol{\beta}$, $\mathbf{m}$ and $\mathbf{u}$, respectively. The variance of random effects are assumed to be $\mathrm{Var}(\mathbf{u}) = 2\mathrm{k}\sigma_A^2$, and $\mathrm{Var}(\mathbf{e}) = \mathrm{R}\sigma_e^2$, where $\mathbf{K}$ is a matrix of kinship coefficients that define the degree of genetic covariance between a pair of individuals; $\mathbf{R}$ is a matrix in which the off-diagonal are 0 and the diagonal elements are the reciprocal of the number of observations for which each phenotypic data point was obtained; σ_A^2 and σ_A^2 are the additive genetic variance and residual, respectively, estimated by REML. The $\mathbf{K}$ matrix were estimated with a random set of 10,000 RNA-seq SNPs according to method used to estimate them,

and the population structure effect was estimated with a random set of 10,000 RNA-seq SNPs. The mixed linear model and the kinship matrix estimation were performed using the R package 'GAPIT' – Genomic Association and Prediction Integral Tool developed by Lipka et al. (2012).

To account the cryptic relationship between individuals we evaluated three different approaches to estimate the kinship matrix as variance-covariance matrix for random genotypes effects: VanRaden (VanRaden, 2008), Loiselle (Loiselle et al., 1995) and the efficient mixed-model association (EMMA) kinship approaches (Kang et al., 2008). We also used the R package GAPIT to perform principal components analysis (Price et al., 2006) of the genotypic data to control for population structure. Thus, several GWAS models - including the "Q + K", "K", "Q" and naive (without population structure and kinship) model - were evaluated to determine the most appropriate one based on the expectation that the *p*-values should follow a uniform distribution between 0 and 1 (i.e. *p*-values ~ U[0,1]) under the null hypothesis of no-associations. Expected *versus* observed quantile-quantile (Q-Q) plots were drawn to assess model fitness.

Several approaches have been suggested to derive an appropriated significance threshold for SNP-trait associations. To determine an appropriate genome-wise threshold that controls the experiment-wise type error rate without being overly we implemented the simple*M* method (Gao et al., 2008). Simple*M* method applies a Bonferroni correction to the actual number of independent test (i.e. the effective number of test) by estimating linkage disequilibrium between each pair of markers and applying principal component analysis to obtain the eigenvalues. Simple*M* method has been shown to be an effective way to control the experiment-wise error rate in genome-wide association studies (Gao et al., 2010; Johnson et al., 2010; Gao, 2011). The linkage disequilibrium was estimated as the square of the correlation between each pairs of SNPs by R package 'genetics', and the principal component analysis were performed by the R package 'adegenet'. In this study, the cumulative effective number of tests across all the chromosomes was 172,470, therefore, the genome-wise threshold is equal to 0.05/172,470 i.e. 2.9×10^{-6} ($\alpha_e = 0.05$). We also used the effective number of test per chromosome to identify chromosome-wise significant associations (Supplemental Table 1). The effective number of test for our SNPs was equal to the number of eigenvalues necessary to explain 99.0% of the variance. simple*M* is a high-speed approach that has been shown to perform as well as permutation

tests (permutations tests are very computationally demanding), which is considered the gold standard (Gao et al., 2008; Gao et al., 2010). Additionally, the permutation tests are not appropriate in the presence of the population substructure, because shuffling the genotype data with respect of the phenotypic data breaks not only SNP-trait association, but also it breaks the trait-pedigree (e.g. kinship) relationship (Aulchenko et al., 2007). In addition to simple*M* method, to account for multiple testing, we controlled the false discovery rate or FDR (Benjamini and Hochberg, et al., 1995).

2.6. Candidate Gene Association Analysis

In addition to genome wide association mapping, where all markers were tested for association with the ten traits evaluated in WiDiv population, we also performed candidate gene association mapping for the ten traits (Thornsberry et al., 2001). We used a set of 4,379 RNA-seq SNPs located inside of candidate-genes regions to maize plant height-related traits to perform genome wide association analysis on the 400 inbred lines from WiDiv population using mixed linear model approaches (Yu et al., 2006) in the R package 'GAPIT' (Lipka et al., 2012) as described above for genome wide association mapping. As we used only a set of the markers, the Bonferroni threshold across whole genome considering independent test was 1.16×10^{-4} ($\alpha_e = 0.05$).

3. RESULTS

3.1. Phenotypic Assessment of the Populations for Maize Plant Height

The estimates of heritability of the ten traits and their mean phenotypic performance based on BLUP values are showed in Table 1. We observed abundant variation in the plant height-related traits in all populations. The range of variation in the WiDiV was greater than that observed in the other populations. In the WiDiv, the traits PH, EH, IL and NN ranged from 74.0 to 274.0 cm, 18.8 to 183.0 cm, 4.6 to 17.4 cm, and 6.3 to 25.3 node numbers, respectively. The range of variation for BEIL and AEIL was greater than IL ranging from 2.2 to 15.4 cm and 7.9 to 31.4 cm, respectively. Among RIL populations, the range of variation in the IBM was smaller than that observed in other populations. The difference within range in the IBM was 92.7 cm, 71.1 cm, and 1.7 numbers for PH, EH, and AENN, respectively, and around 5.0 for other traits. OW, NyH and NAM22 displayed similar range of variation for PH. The range of variation to flowering time ranged to 49.0 to 94.0, and 49.0 to 96.0 days for DP and DS, respectively. We observed high estimates of narrow-heritability coefficients for all traits measured in all populations, ranging from 0.83 to 0.95 for AENN and DS, respectively. The interaction between genotype and year for the measured traits was not significantly (data not shown). The estimates of phenotypic correlations between traits ranged from -0.5 to almost 1.0 (Fig. 1). Weak correlations were observed between BEIL and the traits AENN, BENN, NN, AEIL and flowering time, and strong negative correlations were observed between IL and the traits NN, BENN and flowering time. Positive correlations were showed between NN and DTS, EH and BENN, BEIL and EH, and EH and flowering time. PH correlated strongly and positively with all other traits.

3.2. Quantitative Trait Loci Mapping

The genetic linkage maps were constructed using a total length of 8,175, 5,683 and 5,288 highly information bin markers, and 1,106 SNPs covered the whole genome of maize, spanning 6,144.1, 3,166.7, 3,006.4, and 1,399.3 cM with an average interval of 0.75, 0.56, 0.57 and 1.27 cM between adjacent markers in the IBM, OW, NyH and NAM22 populations, respectively (Supplemental Tables 2-5).

As showed in Table 2 and supplemental Table 6, a total of 84 QTL were identified for the eight traits measured in four populations using the multiple interval mapping

method. Of all QTL identified for plant height-related traits only three QTL were located on chromosome 8 and one was located on chromosome 7. For PH, 21 QTL were detected, located on almost all chromosomes, except on 7 and 8. Seven QTL were identified on the chromosome 1. Of the 21 QTL detected for PH, 42% (9) were identified in the OW population. Only one QTL with negative additive effect was identified in the NyH population. Half of QTL identified for PH had positive additive effect, ranging from 4.7 (*qPH5*, OW) to 6.4 cm (*qPH9*, IBM). The QTL, *qPH9* (IBM) and *qPH10* (OW), had the highest additive effects with values of 6.4 and 6.2 cm, respectively. The QTL, *qPH9* and *qPH10*, accounted for 12.28 and 9.07% of the phenotypic variance, respectively. Another QTL, *qPH4* (NAM22), had negative additive effect of -4.7 cm and explained 13.24% of phenotypic variance. The total of QTL effects detected for PH account most of phenotypic variance in populations IBM, OW and NAM22, with the phenotypic variance explained by them ranging from 41.2 to 51.3%. In relation to consensus QTL among populations, QTL for PH identified in the OW overlapped its support interval with a QTL interval detected in the IBM, and other identified in OW with QTL interval in the NAM22.

A total of 18 QTL were identified for EH, and were located on chromosomes 1, 2, 3, 4, 5, 9 and 10 spanning the four populations. Although the number of QTL identified for EH had been distributed the four populations, the OW population comprised one-third of the QTL detected for EH. As for PH, almost half of QTL identified for EH had positive additive effect ranging from 2.8 to 6.1 cm. The QTL mapped for EH explained more than 5.8% of phenotypic variance, except to the QTL, *qEH2* and *qEH1a*, identified in OW population. The QTL, *qEH9* and qEH1b, identified in the IBM and OW populations, had the highest additive effect and accounted for 14.21 and 18.86% of phenotypic variance, respectively. Among the QTL mapped for EH, the QTL, *qEH10*, identified on chromosome 10 in the NAM22 population, had negative additive effect of -3.2 cm, and it accounted for 10.51% of phenotypic variance. In relation to consensus QTL for EH among populations, there were eight QTL that overlapped their supported interval. The QTL interval of *qEH1* (NAM22) overlapped with *qEH1* interval (IBM) on chromosome 1, the *qEH2* interval (NyH) overlapped with *qEH2* interval (OW) on chromosome 2, the *qEH3* interval (NAM22) with *qEH3* interval (IBM) on chromosome 3, and the *qEH5* interval (NyH) overlapped with *qEH5* interval (OW) on chromosome 3.

For the IL, five QTL were mapped, two of which on chromosome 6 and 10 in the OW population, and the others on chromosome 1 and 9 in the IBM population (Supplemental Table 6). Four QTL identified for IL had positive additive effects, ranging from 0.2 to 0.5 cm. The QTL, *qIL10b* and *qIL9b,* identified on chromosome 10 and 9 from the OW and IBM population, had relatively large contribution rates, accounting for 12.76 and 20.78% of phenotypic variance of IL, respectively. The QTL model comprising the two QTL identified for IL in the OW population accounted 50% more than phenotypic variance than IBM QTL model. There was no overlapped between QTL interval identified for IL among OW and IBM populations.

Eleven QTL were mapped for NN and were located on five chromosomes from IBM and OW populations, three of which were located on chromosome 1 (Supplemental Table 6). Among QTL mapped for NN, most of them had negative additive effect, ranging -0.2 to -0.4 cm. The QTL, *qNN1a,* had highest positive additive effects and accounted for 15.6% of phenotypic variance. A total QTL effects detected for NN explained around 42% of phenotypic variation in both OW and IBM populations. There was no overlapped between QTL interval identified for NN among OW and IBM populations.

For the traits, BEIL, AEIL, BENN and AENN which were measurements of node number and average internode length below and above uppermost ear node, 29 QTL were mapped on the IBM and OW populations, except for AEIL to which were identified two QTL only in the OW population. Eight QTL were identified for BEIL, and four in each population. Four of which were located on chromosome 1. The QTL, *qBEIL9* and *qBEIL10,* identified for BEIL in the IBM and OW populations, had the highest positive additive effect, 0.4 and 0.5 cm, and they accounted for 15.39 and 10.40% of phenotypic variance, respectively. The QTL for BEIL, *qBEIL1b* (OW) and *qBEIL1a* (IBM), which were located on chromosome 1, overlapped their support interval among OW and IBM populations. The QTL for AEIL, *qAEIL10*, had a positive effect of 0.6 cm and accounted for 7.85% of phenotypic variance. For BENN, nine QTL were mapped, five of which were detected in the IBM population. The QTL, *qBENN5* and *qBENN1,* mapped for BENN in the OW population, had the highest positive additive effect of 0.3 cm and accounted for 9.16 and 11.29 of phenotypic variance. For AENN, of ten QTL identified, eight were mapped in the OW population, and, in all, they explained 44.02% of

phenotypic variance. Among the QTL mapped for AENN, six had additive effect of 0.1cm and other had effect of 0.2 cm and half of QTL had negative additive effect. The QTL, *qAENN3*, identified in the OW population, had relatively large contribution rate, accounting for 10.53% of phenotypic variation. There was no overlapped between QTL interval identified for BENN and AENN among both OW and IBM populations.

In relation to overlapping QTL interval mapped for plant height-related traits, there were some hot spots of overlapping QTL interval among different traits studied (Supplemental Table 7). In four chromosomal regions on chromosome 1 were located QTL identified for more than one plant height-related trait. The QTL mapped for BEIL, EH, PH and NN were nearly located in the physical interval 13.95 – 16.90 Mb, and they overlapped their support interval. In the physical interval 85.15 – 93.10 Mb were located the QTL mapped for PH, BEIL, NN, BENN and EH, all of which had positive additive effect and accounted for more than 5.0% of phenotypic variance. Besides those, in the two additional region on chromosome 1, QTL identified for PH and EH overlapped in physical interval 203.95 – 206.25 Mb, and for BEIL and EH in the physical interval 212.85 – 213.75 Mb. In addition to chromosome 1, on chromosome 3, in the physical interval 6.74 – 9.45 Mb QTL for EH, BENN, EH and BEIL were closely located and overlapped their interval, in the physical interval 182.75 - 183.30 Mb the QTL for EH and AENN were located very close, and the QTL for PH and NN overlapped in the physical interval 145.55 – 158.7 Mb. Besides hot spot of overlapping of QTL interval on chromosome 1 and 3, QTL for EH, PH and NN overlapped on chromosome 5, and QTL identified for IL and PH, and for PH and AENN overlapped in two regions on chromosome 6. Another interesting hot spot of QTL identified were located in the physical interval 99.05 – 99.50 on chromosome 9, where QTL were closely mapped for BEIL, EH, IL and PH. Most of hot spots of overlapping of QTL interval mapped for plant height-related traits involved QTL identified in the IBM population.

3.3. Association Mapping and Candidate Gene Approaches

As shown by the quantile-quantile (Q-Q) plots the mixed model that corrected for cryptic relatedness (K model) fit the data well for ear height and node number (Fig. 2). The other traits showed the same performance (Q-Q plots not showed). The K model showed a significant improvement in model fit over the simple model in which

population structure is ignored. Overall, differences in model fitness between the three different kinship approaches - VanRaden, Loiselle and EMMA – were very small, and they displayed good control of type I error rate. Thus, for plant architecture traits – PH, EH, IL, BEIL, AEIL, BENN and AENN – no population structure were detected, and Q (population structure effect) was not included in the model. For those, as K-Loiselle model adjusted slightly better than other K models, all further results refer to it. Although we argued that the K model resulted in sufficient control of false positive associations for all traits, for NN and flowering time, we consistently observed lower spurious association identified when we additionally corrected to population structure by using the first third principal components estimated with a set of 10,000 RNA-seq SNP (Q_3). Among the models assessed for those traits, the K-EMMA + Q_3 model showed little improvement in goodness fit over the other models and all further results for them refers to the K-EMMA + Q_3 model.

Based on simple*M*, genome-wise and chromosome-wise thresholds, and FDR methods to controls the experiment-wise type error rate, we found a total of 30 SNP significantly associated to seven plant height-related traits on chromosomes 1, 3, 8 and 10 (Fig. 3, Fig. S1, and Table S8). We did not find SNP associated with the traits IL, BEIL and AENN. Although the 30 associations were located in nine gene models, two third (21) of these were located in the same gene model, GRMZM2G171622, which associated was associated with PH, EH, NN, BENN and flowering time traits. It is annotated as a CBS-domain contain protein and the SNP associated to it showed substantial variation with minor allele frequency ranging from 0.22 to 0.50. In addition to gene model GRMZM2G171622, other SNP within different gene model were associated with plant height-related traits. Three SNP, located in 3 different genes models, were associated to BENN, of which two were associated to NN too. We also found an association, with limited variation (minor allele frequency=0.15), for flowering time traits located in a different gene model, the GRMZM2G169089. Moreover, we found three SNP associated with AEIL located in two gene model that were not associated to other traits. When we performed the candidate-gene association analysis, we found three SNP, located in the candidate gene GRMZM2G108798, significantly associated with NN, and three SNP, located in the candidate gene GRMZM2G01276, significantly associated with AEIL. We

did not found associations among candidate genes to plant height-related traits and other traits (Fig. S2).

We compared genome wide mapping results with linkage analysis QTL intervals. The SNP associated with BENN on chromosome 1 was located inside the QTL interval of QTL, *qBENN1a*, *qEH1b* and *qPH1c*, mapped for BENN, EH and PH in OW population, and the association were located very close to physical position of QTL, *qEH1b*. The 21 SNP associated with six plant-height traits, located in the gene model GRMZM2G171622 on chromosome 3, comprised of five different SNP. These SNP were located very close to physical position of QTL mapped for NN, *qNN3b*, in IBM, and for NN, BENN and AENN, *qNN3*, *qBENN3c* and *qANN3c*, in OW population. All these SNP were located inside QTL interval for cited QTL. The SNP associated with NN on chromosome 6 by candidate-gene analysis was located inside QTL interval for QTL mapped for IL and AEIL in the OW population. However, the three SNP associated with AEIL by candidate-gene analysis were not located inside any QTL interval mapped in the RILs populations.

4. DISCUSSION

Maize plant height has been extensively studied over last three decades because it is related to plant density and lodging resistance. Breeding activities during past decades have been slightly decreased EH, but not PH (Duvick et al., 2005), and the EH reduction relates to the selection for increasing the lodging resistance. Recently, another breeding target has been the development of taller plants to increase biomass production in bioenergy crops such as maize (Salas Fernández et al., 2009). There is an extensive morphological and genetic diversity for plant architecture traits in maize which should be used to study complex traits in this species (Li et al., 2011; Yang et al., 2011; Flint-Garcia et al., 2005). Moreover, high heritabilities are a prerequisite for reliable QTL and association results and to reduce the bias in the estimation of proportion of genetic variance accounted by the QTL (Bradbury et al., 2011). In our study, all traits displayed high heritability estimates, and, thus, large proportional of abundant variation of plant height-related traits was from genetic nature, and experimental design was efficient to control the experimental error. Moreover, plant height correlated with all traits, but it showed positive and stronger correlation with EH, NN, BEIL, BENN and flowering time traits. Ji-hua et al. (2007) found that plant height correlated positively and strongly with node number and average internode length, and increased plant height was affected by average internode length. The heritability estimates were very high, reaching the values of 0.94 for all traits. In the study of Salvi et al. (2011), plant height was positively correlated with all plant architecture traits, except proportion of node number below the uppermost ear. Additionally, node number was strongly positively correlated with days to pollen shed, but it was negatively correlated with average internode length.

Considering the four RIL populations, 84 QTL were mapped for maize plant height-related traits. Of these, QTL mapped for PH, EH and BEIL showed some overlap of their support interval among RIL population mapped. Wang et al. (2006) carried out a Meta-analysis on published QTL, and a total of 127 QTL for maize plant height, were refined to 40 consensus QTL. Compared with other reports, the QTL mapped for PH that overlapped chromosome 3 and 5 among two populations seemed to have similar chromosomal locations with QTL identified by Ji-hua et al (2007). They mapped a QTL for PH very close to hot spot QTL detected for EH and PH, located on chromosome 3 and 5, in our study. Of 14 QTL mapped on chromosome 3, 7 QTL were mapped for plant and

ear height. On chromosome 3S, Teng et al (2013) validated the QTL, *qPH3.1,* for the plant height candidate gene *ZmGA3ox2*. The *qPH3.1* accounted for 32.3% of the phenotypic variance, and the *ZmGA3ox2* is involved in the gibberellin biosynthesis pathway.

We found a lot of hot spots of overlapping QTL for different plant height-related traits across all chromosomes except to chromosome 2 and 7. Most of these hot spots were located in four different regions of chromosome 1 and 3, and they involved the traits PH, EH and some traits of NN and IL (above and below of uppermost ear). Ji-hua et al. (2007) reported that four of six QTL detected for average internode length were located on the same region of chromosomal 1 as the QTL effecting PH. They proposed that IL was the major contributor to maize plant height. In another study, Salvi et al. (2011) found that three QTL mapped for PH overlapped their interval QTL with QTL identified for NN, BENN and AENN, but not with IL, and they concluded that besides flowering time, NN drove phenotypic variation for PH. According to Price (2006) QTL mapping is a powerful way to characterize genes (in terms of numbers and contributions) underlies complex traits. In QTL mapping study, it seems that QTL for EH, NN (above and below), IL and BEIL were major contributors to genetic variation in PH.

According to Holland et al. (2007), investigations of genetic architecture of complex traits, combined with improved knowledge of the structure of plant populations, would impact our understanding of design of crop improvement strategies. Population structure and familial relatedness create genome-wide linkage disequilibrium between unliked loci resulting in large number of spurious associations (Ersoz et al., 2009). The effects of population structure and cryptic relationship between individuals can be corrected including the K and Q effects in the unified mixed model (Q + K model) proposed by Yu et al. (2006). According to them, the used of K, Q or Q + K model depend on trait nature. In our study, K model showed a good control of type I error for all traits except for flowering time and NN which were strongly associated each other. For these traits, the inclusion of three principal components as fixed effect in the model reduced the spurious associations and improved the model fitness. According to Yu et al. (2006), for flowering time, the Q + K model yielded better statistical power than other models. Maize flowering time is an adaptive trait, and for this kind of trait, population structure effect is more apparent because phenotypic variation between populations is

highly correlated with genotypic differences between them (Zhao et al., 2007). Thornsberry et al. (2001) using general linear model identified several associations for maize flowering time in an association panel consisted of 102 diverse inbred lines. Larsson et al. (2013) reanalyzed the flowering time data using Q + K model and concluded that results reported by Thornsberry et al. (2001) were false positives because did not correct sufficiently for population structure.

Genome wide association using association panel of diverse inbred line have been employed to dissect genetic bases of complex traits in maize. Weng et al. (2011) identified 204 SNP associated with plant height in a panel of 284 elite maize inbred lines. They controlled to population structure, but not use any multiple testing correcting approaches; $P \leq 0.0001$ was used as thresould which is very liberal compared to the correction approaches proposed in the literature. Riedelsheimer et al. (2012) dissected maize leaf metabolites profiles by genome wide associations using the Q + K model and FDR methods to correct for false positives, and identified 15 distinct SNP-metabolite associations. In another association study, Li et al. (2013) examined the genetic architecture of maize oil biosynthesis, and detected 74 loci associated with kernel oil composition and fatty acid composition. They used a threshold (1/total numbers of SNP) that has a little power to control false-positive associations. In the study of Xue et al. (2013), association mapping was performed using Q + K model and the threshold of 0.1/number of markers, and they identified 42 SNP associated with drought tolerance in maize. In our study, we found 31 SNP significantly associated to plant height-related traits, but 21 of them were located in the same gene model, GRMZM2G171622. Despite we used a large number of SNP in WiDiv panel with abundant variation, we did not find a lot of associations. We believe that was because the FDR and simple*M* approaches were very conservative to control type I error rates. Cook et al. (2012) conducted association mapping in a panel of panel of 282 diverse inbred lines employing similar that was used in our study and find none SNP of the 51,741 associated with kernel composition.

The gene model GRMZM2G171622 is annotated as CBS-domain containing protein. It was associated with six plant height-related traits and, thus, it seems these traits share part of their genetic base. The gene model GRMZM2G11698 which associated with DPS is annotated as *Diacylglycerol acyltransferase.* It belongs to the family of transferases, especially those acyltransferases transferring groups other than aminoacyl

groups (Styme et al., 2009). There is not report associating this enzyme of maize flowering time. The candidate gene GRMZM2G108798, identified by candidate-gene approach, is related to the pathway of gibberellin response and belongs to gene class short internode transcription factor (Strigens et al., 2013; You et al, 2012). Another candidate gene, GRMZM2G091276 belongs to *dwarf* gene class, and it was descripted as 'indole-3-acetic.acid-amido.synthetase.GH3.5'. It encodes a jasmonate-amido synthetase (JAR1) that localizes to the cytoplasm and is also a phytochrome A signaling component. JAR1 is an auxin-induced gene. Weng et al. (2013) also found four loci containing genes associated with gibberellin and auxin in the same chromosomal region of two candidate gene described above. Thus, we found two important candidate genes associated with node number and above ear internode length which are related growth hormone in maize.

We identified SNP associated with AEIL by gene-candidate association that were located inside the interval QTL for the AEIL and IL displaying the chromosomal regions is very important for average internode length trait. Additionally, the five SNPs located in the gene mode GRMZM2G171622 were located inside four QTL intervals for node number traits on chromosome 3 in different populations. This gene model seems to be important to genetic base of node number and related traits (above and below ear). Another region on chromosome 1 had overlapped among QTL interval for BENN, PH and EH and a SNP associated with BENN. Thus, it seems that node number, plant and ear height have the same genetic mechanism.

5. CONCLUSIONS

We concluded that: maize plant height can be phenotypically dissected in its components; maize plant height-related traits can be genetically dissected by linkage analysis and association mapping; there are some promises regions on chromosomes 1 and 3 that seems to be involved in control of maize plant height-related traits; the traits ear height, node number, below ear node number, below ear average internode length and flowering time were the main contributors to maize plant height; the candidate genes GRMZM2G108798 and GRMZM2G012766 which were identified in WiDiv population panel were associated with plant height-related traits in our study; and, in addition, this study constitutes an important step towards the characterization of the genetic architecture of plant height, a key trait in maize breeding and genetics.

6. REFERENCES

Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping. Bioinformatics 26: 2990-2992.

Aulchenko YS, Koning DJ, Haley C (2007) Genome wide rapid association using mixed model and regression: A fast and simple method for genome wide pedigree-based quantitative trait loci association analysis. Genetics 177: 577-585.

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theoretical Applied Genetics 83: 141-145.

Beavis WD, Lee M, Grant D, Hallauer AR, Owens T, Katt M, Blair D (1992) The influence of random mating on recombination among RFLP loci. Maize Genetics Conference 66: 52-53.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57: 289-300.

Bernardo R (2008) Molecular markers and selection for complex traits in plants: leaning from the last 20 years. Crop Science 48: 1649-1664.

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

Broman KW, Ś Sen (2009) A Guide to QTL Mapping with R/qtl. Springer, New York, NY.

Buckler ES, Holland JB, Bradbury PJ, Acharya C, Brown P, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Oropeza Rosas M, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325: 714-718.

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury PJ, Ross-Ibarra J, Buckler ES, Flint-Garcia S (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiology 158: 824-834.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.

Duvick DN (2005) The contribution of breeding to yield advances in maize (*Zea mays* L.). In: DL Sparks (Ed.). Advances in Agronomy 83-145.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6: e19379.

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th Ed. Longman Hallow, London

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54: 357-374.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchel SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. The plant Journal 44: 1054-1064.

Fan JB (2003) Highly Parallel SNP Genotyping. Cold Spring Harbor Symposia Quantitative Biology 68: 69-788.

Gao XY, Stamier J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetics Epidemiology 32: 361-369.

Gao XY, Becker LC, Becker DM, Starmer JD, Province MA (2010) Avoiding the high Bonferroni penalty in genome-wide association studies. Genetics Epidemiology 34: 100-105.

Gao XY (2011) Multiple testing corrections for imputed SNPs. Genetics Epidemiology 35: 154-158.

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

Hansey CN, Johnson JM, Sekhon RS, Kaeppler SM, de Leon N (2011) Genetic diversity of a maize association population with restricted phenology. Crop Science 51: 704-715.

Holland JB (2007) Genetic architecture of complex traits in plants. Current Opinion in Plant Biology 10: 156-161.

Jansen RC (1994) Interval mapping of multiple quantitative trait loci. Genetics 135: 205-211.

Jansen RC (2007) Quantitative trait loci in inbred lines. In: Handbook of Statistical Genetics, JD Balding, M Bishop, and C Cannings (Ed.). 599-622.

Jansen RC, Stam P (1994) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315-324.

Ji-hua T, Wen-tao T, Jian-bing Y, Xi-qing M, Yi-jiang M, Jin-Rui, Jian-Sheng L (2007) Genetic dissection of plant height by molecular markers using a population of recombinant inbred lines in maize. Euphytica 155: 117-124.

Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, **et al.** (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11:724.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, **et al.** (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709-1723.

Kosambi DD (1943) The estimation of map distances from recombination values. Annual Eugene 12: 172-175.

Kump KL, Bradbury PJ, Buckler ES, Belcher AR, Oropeza-Rosas M, Wisser RJ, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics **43**: 163–168.

Kruglyak L 1999 Prospects for whole genome linkage disequilibrium mapping of common diseases genes. Nature Genetics 22: 139-144.

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

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10: R25

Larson SJ, Lipka AE, Buckler ES (2013) Lessons from *Dwarf* on the strengths and weaknesss of structure association mapping. PLoS ONE 9: e1003246.

Lipka AE, Tian F, Wang QS, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28: 2397-2399.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J (2013) Genome-wide association study dissect the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics 45: 43-50.

Li X, Zhu C, Ye C, Wu W, Takacs EM, Petsch KA, Tian F, Bai G, Buckler ES, Muehlbauer GJ, Timmermans MCP, Scanlon MJ, Schnable PS, Yu J (2012) Genic and nongenic contributions to natural variation of quantitative traits in maize 22: 2436-2444.

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic maize with the intermated B73 x Mo17 (*IBM*) population. Plant Molecular Biology 48: 453-461.

Lima MLA, Souja Jr CL, Bento DAV, Souza AP, Carlini-Garcia LA (2006) Mapping QTL for grain yield and plant traits in a tropical maize population. Molecular Breeding 17: 227-239.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial Genetic-Structure of a Tropical Understory Shrub, Psychotria Officinalis (Rubiaceae). American Journal Botany 82: 1420-1425.

Lübberstedt T, Melchinger AE, Klein D, Degenhardt H, Paul C (1997) QTL mapping in testcrosses of European flint lines of maize .II. Comparison of different testers for forage quality traits. Crop Science 37: 1913-1922.

Lu Y, Zhang S, Shan T, Xie C, Hao Z, Li X, Farkahari M, Ribaut J, Cao M, Rong T, Xu Y (2010) Joint linkage- disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proceeding of the National Academy of Science of the United States of America, 107: 19585-19590.

Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Reviews of Genetics 35: 303-339.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10: 565-577.

Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW (2009) A Model Selection Approach for the Identification of Quantitative Trait Loci in Experimental Crosses, Allowing Epistasis. Genetics 181: 1077-1086.

McMullen MD, Kresovich S, Villeda HS, Bradbury PJ, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325: 737-740.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experiment design. Plant Cell 21: 2194-2202.

Oraguzie NC, Wilcox PL (2009) An overview of associaition mapping, In: Association mapping in plants, NC Oraguzie, EHA Rikkerink, SE Gardiner, HN de Silva, (Ed.), 1-10, Springer, Berlin, German.

Peiffer JA, Flint-Garcia S, DeLeon N, McMullen MD, Kaeppler SM, Buckler ES (2013) The genetic architecture of maize stalk strength. PLoS ONE 8:e67066.

Price AH (2006) Believe it or not, QTL are accurate! Trends in Plant Science 5: 213-216.

Price A., Patterson NJ, Plenge RM, Weinblatt, Shadick NA, Reich (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38: 904-909.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. American Society of Human Genetics 67: 170-181.

R Core Team. 2012. R: a language and environment for statistical computing, R Foundation for Statistical Computing Vienna, Austria http://www.R-project.org/

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proceedings of the National Academy of Sciences of the United States of America 109: 8872-8877.

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516-1517.

Ritchie SW, Hanway JJ, Benson GO (2008) How a corn plant developsIowa State University of Science and Technology, Cooperative Extension Service, Ames, Iowa.

Saghaimaroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA Spacer-Length Polymorphisms in Barley - Mendelian Inheritance, Chromosomal Location, and Population-Dynamics. Proceeding of National Academic Science 81: 8014-8018.

Salas Fernandez, MG, Hamblin MT, Li L, Rooney WL, Tuinstra MR, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Science 48: 1732-1743.

Salvi S, Corneti S, Belloti M, Carraro N, Sanguineti MC, Castelli S, Tuberosa R (2011b) Genetic dissection of maize phenology using an intraspecific introgression library. BMC Plant Biology 11: 1-15.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. The American Journal of Human Genetics 7: 629-644.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, et al. (2009) The B73 maize genome: complexity, diversity and dynamics. Science 326: 1112-1115.

Soto-Ceda BJ, Cloutier S (2012) Association mapping in plant genomes. . Association mapping in plant genomes In: Genetic Diversity, Mahmut Çalişkan, (Ed.), 1-26, InTech, Canadá.

Stich B, Melchinger AE (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics 10:94.

Stymne S, Stahl U, Lenman M, Banas A, Lee M, Sandager L, Ronne H(2000). Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Procceding of National Academy of Sciences 97: 6487–92.

Teng F, Zhai L, Liu R, Bai W, Huo D, Tao Y, Zheng Y, Zhang Z (2013) *ZmGA3ox2*, a candidate gene for a major QTL, *qPH3.1*, for plant height in maize. The plant journal 73: 405-416.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

Thronsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphism associate with variation in flowering time. Nature Genetics 28: 286-289.

Tian F, Bradbury PJ, Brown PJ, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler EB (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics ***43**: 159-162.*

Wang Y, Yao J, Zhang ZF, Zheng YL (2006) The comparative analysis based on maize integrated QTL map and meta-analysis of plant height QTLs.Chinese Science Bulletin 51: 2219-2230.

Weng J, Xie C, Hao Z, Wang J, Liu M, Zhang D, Bai L, Zhang S, Li X (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (Zea Mays L.) inbred lines. PLoS ONE 6: e29229.

VanRaden P (2008) Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414-4423.

Xue Y, Warburton ML, Sawkins M, Zhang X, Setter T, Xu Y, Grubloyma P, Gethi J, Ribaut JM, Li W, Zhang X, Zheng Y, Yan J (2013) Genome wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions.Theoretical Applied Genetics 126: 2587-2596.

Yang X, Gao S, Xu S, Zhang Z, Prasana BM, Li L, Li J, Yan J (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex and quantitative traits in maize. Molecular Breeding 28: 511-526.

Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN (2008) Plants to power: bioenergy to fuel the future. Trends Plant Science 13: 421-429.

Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Current Opinion Biotechnology 17: 155-160.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539-551.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model [method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203-208.

Yu X, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice 48: 391-407.

Zhao KM, Aranzana J, Kim S, Lister C, Shindo C (2007) An *Arabidopsis* example of association mapping in structured populations. PLoS ONE 3:e4

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. The Plant Genome 1: 5-20.

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping multiple quantitative trait loci. Proceedings of the National Academy of Science 90: 10972-10976.

7. TABLES AND FIGURES

Table 1. Means, ranges, difference within range, and narrow-sense heritability (h^2) estimates for plant height (PH, cm), ear height (EH, cm), average internode length (IL), node number (NN), below ear average internode length (BEIL), above ear average internode length (AEIL), below ear node number (BENN) and above ear node number (AENN), days to pollen shedding (DP) and days to silking (DS) best linear unbiased predictors in the WiDiv, IBM, OW, NyH and NAM 22 population

Trait	Population	Mean	Range	Difference	h^2
PH	WiDiv	174.6	74.0 - 274.0	200.0	0.91
	IBM	189.7	145.6 - 238.2	92.7	0.94
	OW	153.7	98.0 - 217.0	118.0	0.89
	NyH	144.7	90.0 - 201.0	111.0	-
	NAM22	190.4	88.0 - 276.0	188.0	0.87
EH	WiDiv	86.7	18.8 - 163.0	144.5	0.91
	IBM	95.4	65.0 - 136.1	71.1	0.92
	OW	70.3	35.0 - 107.0	72.0	-
	NyH	63.9	34.0 - 96.0	62.0	-
	NAM22	104.6	38.3 - 183.0	144.7	-
IL	WiDiv	10.2	4.6 - 17.4	12.8	0.84
	IBM	11.1	13.5 - 8.9	4.9	0.88
	OW	12.2	16.1 - 8.0	8.1	-
NN	WiDiv	17.4	6.3 - 25.3	18.9	0.91
	IBM	17.6	9.5 - 14.8	5.3	0.93
	OW	12.7	15.6 - 9.8	5.8	-
BEIL	WiDiv	6.8	2.2 - 15.4	13.2	0.85
	IBM	7.4	10.0 - 5.3	4.7	0.94
	OW	9.7	13.6 - 4.9	8.7	-
AEIL	WiDiv	15.9	7.9 - 31.4	23.5	0.75
	IBM	18.4	22.0 -14.9	7.1	0.82
	OW	15.4	21.1 - 9.2	11.9	-
BENN	WiDiv	13.4	7.5 - 21.0	13.5	0.92
	IBM	13.13	15.6 - 10.4	4.9	0.94
	OW	7.2	9.8 - 4.4	5.4	-
AENN	WiDiv	5.7	2.2 - 11.5	9.3	0.83
	IBM	5.4	6.3 - 4.6	1.7	0.86
	OW	5.4	8.6 - 4.0	4.6	-
DP	WiDiv	70.8	49.0 - 94.0	45.0	0.95
DS	WiDiv	72.9	49.0 - 96.0	47.0	0.94

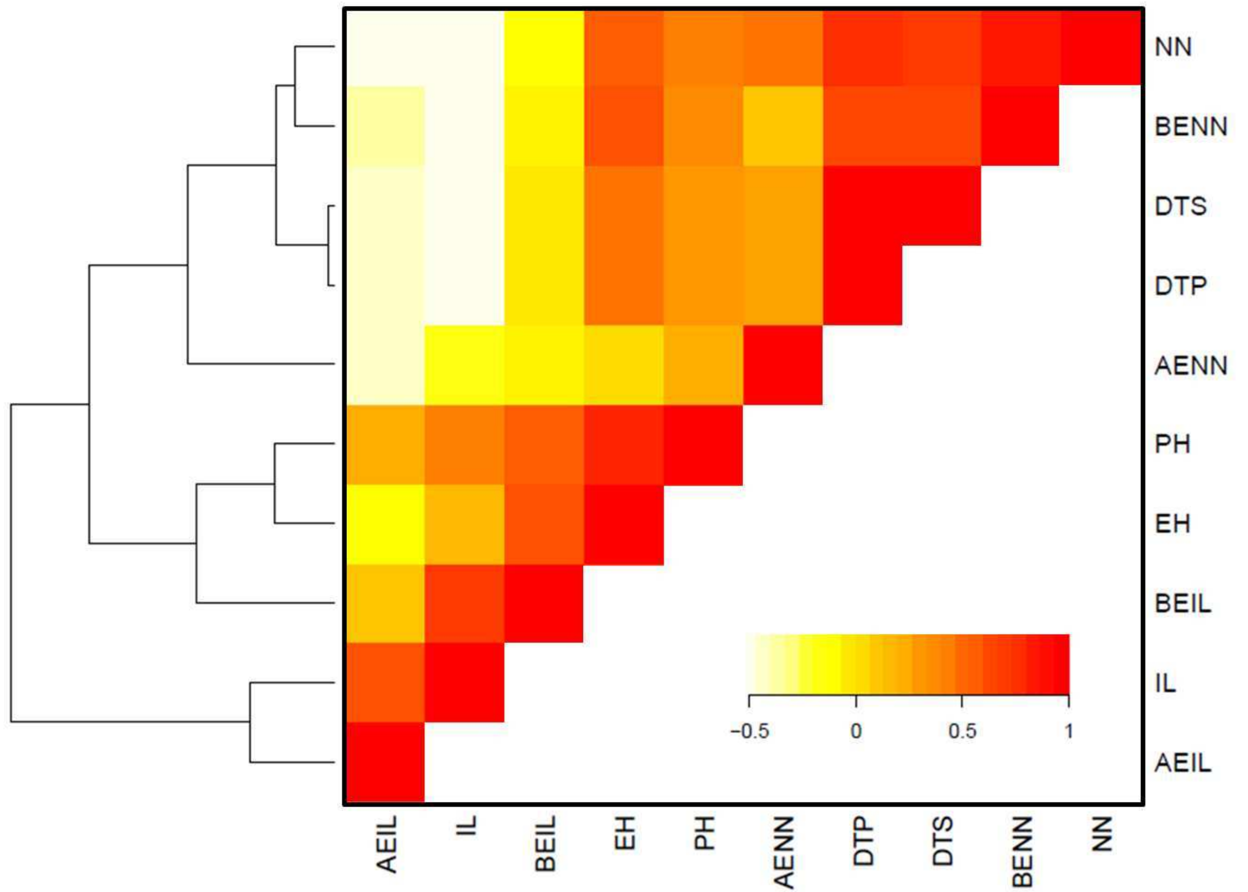


Figure 1. Spearman's rank correlations ($p < 0.05$) and clustering analysis (dendrogram) among maize plant height and its components measured in four RIL populations (IBM and OW) and the WiDiV panel. PH, plant height; EH, ear height; NN, node number, IL, average internode length; BEIL, below ear average internode length; AEIL, above ear average internode length; BENN, below ear node number, above ear node number; DP, days to pollen shedding, and DS, days to silking.

Table 2. Summary of QTL mapping results for plant height (PH), ear height EH), node number (NN), average internode length (IL), below ear average internode length (BEIL), above ear average internode length (AEIL), below ear node number (BENN) and above ear node number (AEN) evaluated in the OW, IBM, NyH and NAM22 populations

Trait	Number of QTL identified				Percent of variation explained by QTL Model (R^2)			
	OW	IBM	NyH[1/]	NAM22	OW	IBM	NyH	NAM22
PH	9	6	1	5	51.73	45.41	6.59	41.2
EH	6	4	3	5	45.96	32.52	22.17	37.2
IL	2	3	-	-	19.82	32.58	-	-
NN	4	7	-	-	42.59	41.56	-	-
BEIL	4	4	-	-	27.64	37.50	-	-
AEIL	2	0	-	-	12.93	0	-	-
BENN	4	5	-	-	36.03	32.37	-	-
AENN	8	2	-	-	44.02	16.16	-	-

[1/] - means that traits were not evaluated in this populations.

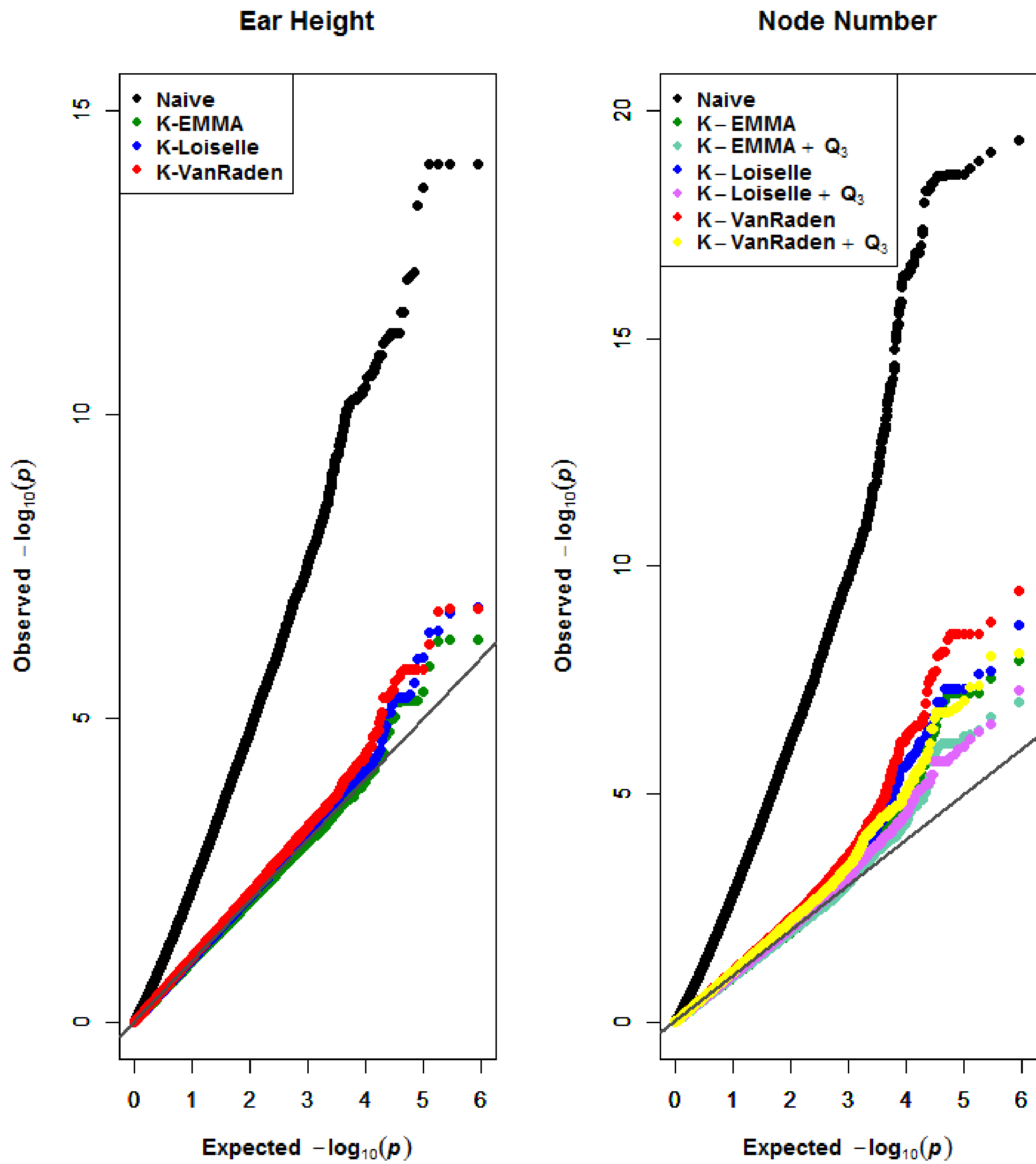


Figure 2. Comparison of QQ plots for different GWAS models for two plant height-rlated traits with contrasting extend of confounding due relatedness and population structure. Observed *vs* expected *p*-values are shown for (A) ear height using four models and (B) node number using seven models with different corrections of confounding effects (*see details in Material and Methods*).

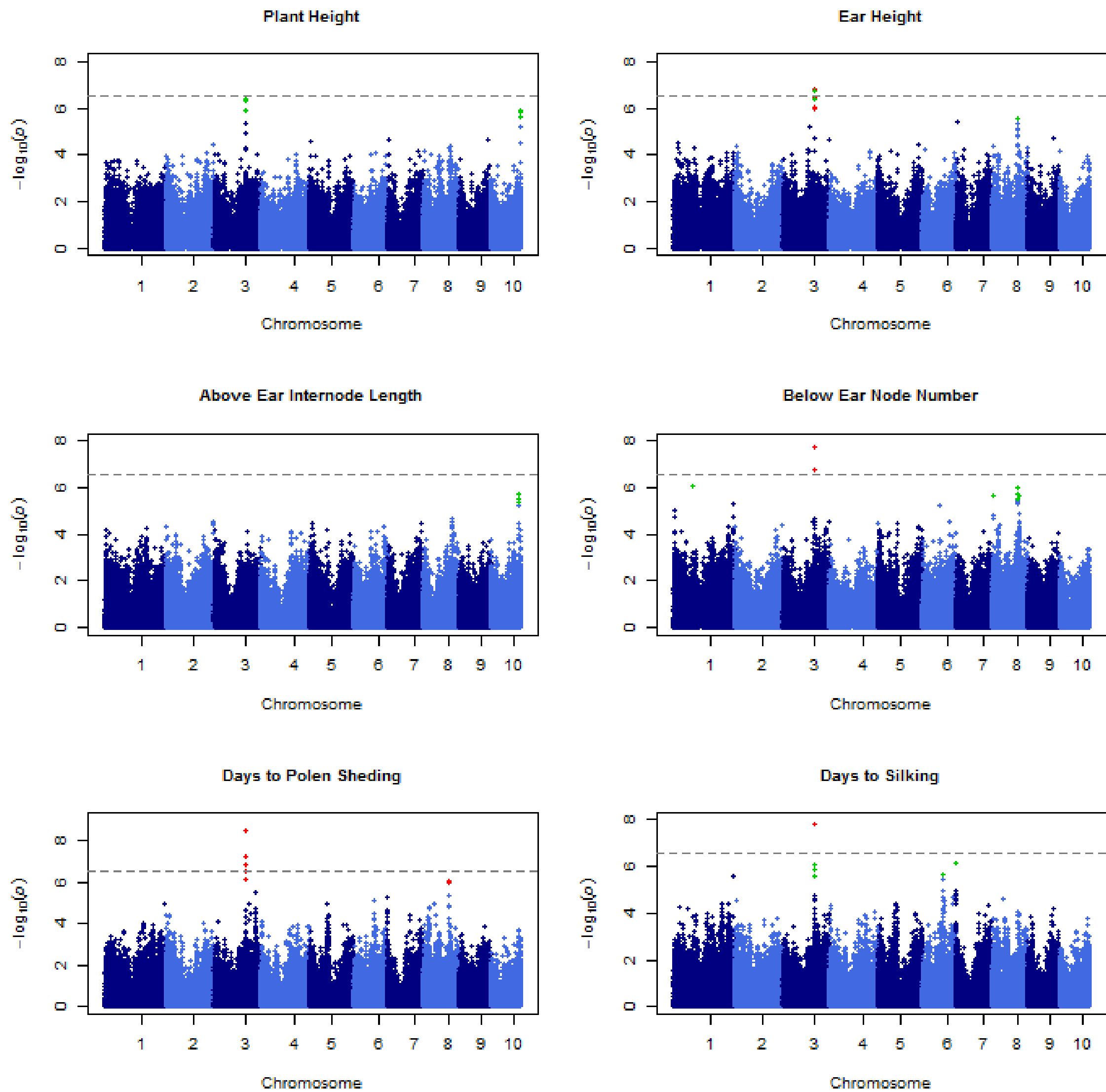


Figure 3. Manhattan plots for plant height, ear height, above ear internode length, below ear node number, days to pollen sheding and days to silking. *P*- values are shown on a $\log_{10}$ scale. The dashed horizontal line depicts the Bonferroni-adjusted correction based on the effective number of independent test (2.9 x 10^{-6}). Green dots are significant SNPs based on a chromosome-wise threshold also determined through simple*M* method. The SNPs with in red dots are significant with FDR ≤ 0.05 and based on a chromosome-wise threshold determined through simple method.

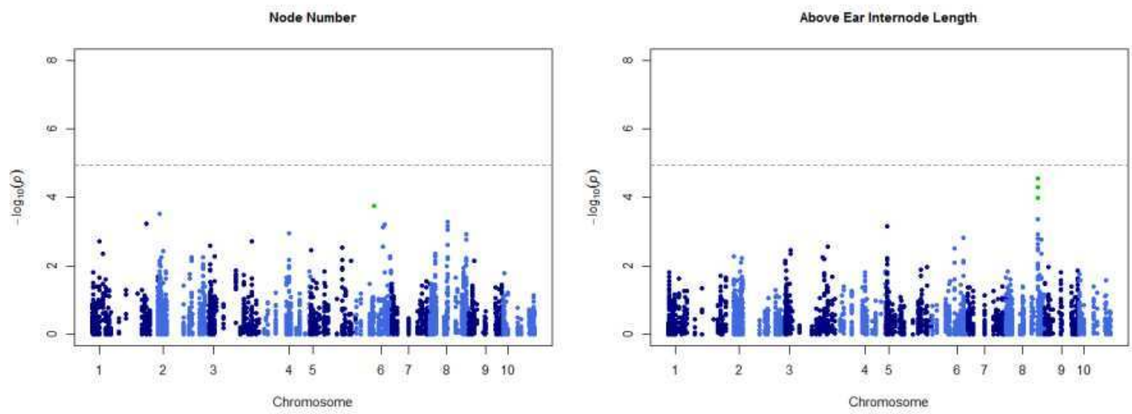


Figure 4. Manhattan plots for node number and above ear internode length using a set of SNP located within of region interval of the candidate gene for maize plant height-related traits. The dashed horizontal line depicts the Bonferroni-adjusted correction based on the effective number of independent test (1.16 x 10^{-4}). Green dots are significant SNPs based on a chromosome-wise threshold also determined through simple*M* method.

8. APPENDICES

8.1. Supplemental Figures

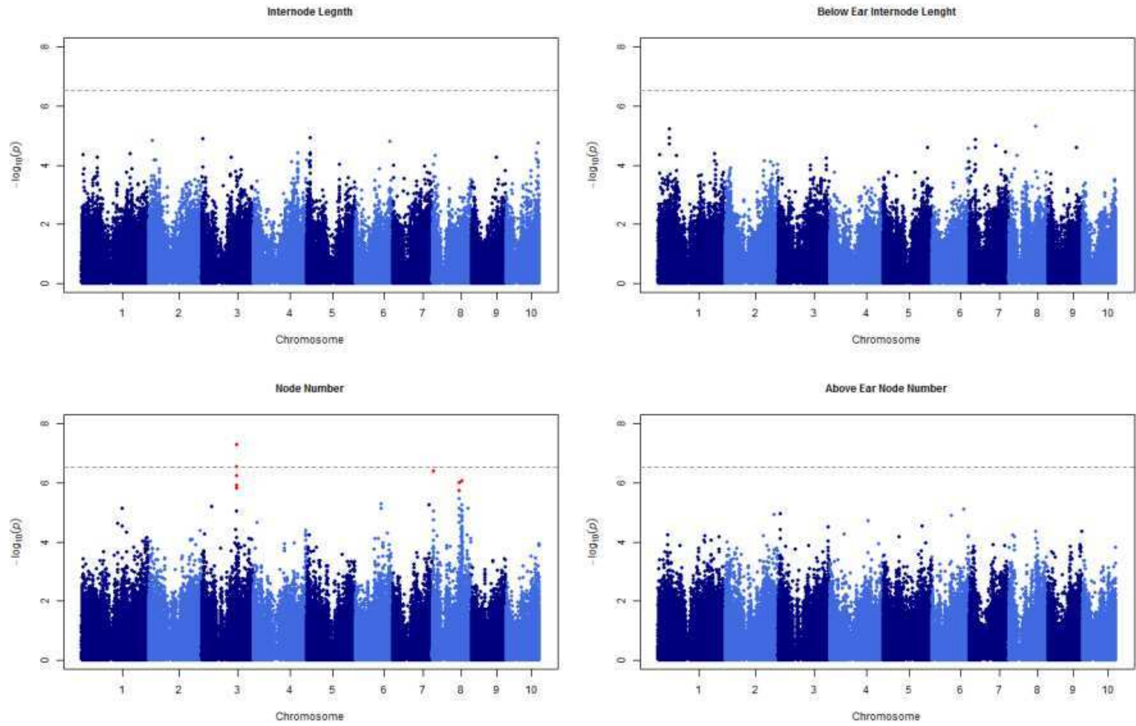


Fig. S1. Manhattan plots for internode length, below ear internode length, node number, above ear node number. *P* values are shown on a $\log_{10}$ scale. The dashed horizontal line depicts the Bonferroni-adjusted correction based on the effective number of independent test (2.9×10^{-6}). The SNPs with in red dots are significant with FDR ≤ 0.05 and based on a chromosome-wise threshold determined through simple*M* method.

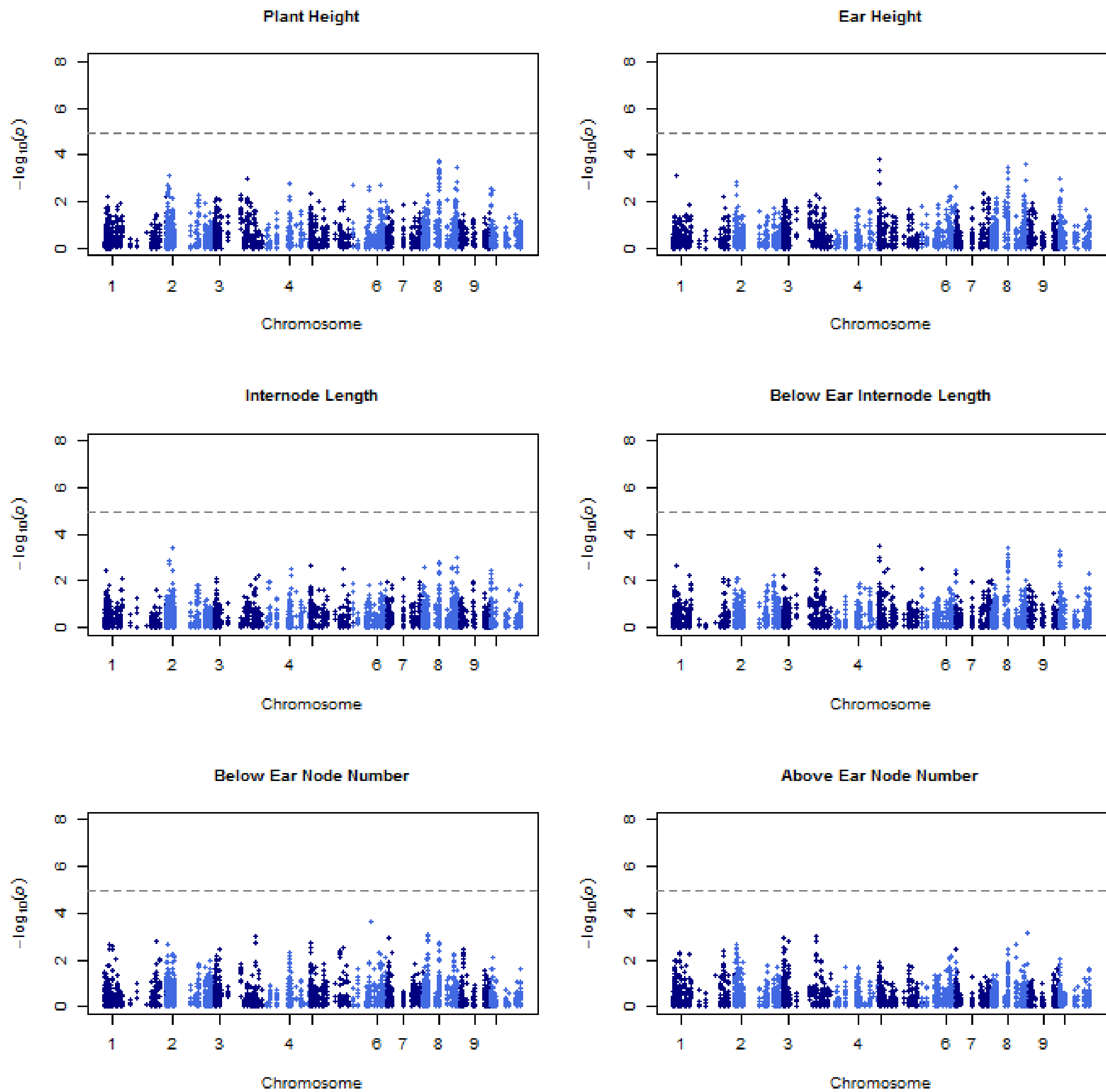


Fig. S2. Manhattan plots for plant height, ear height, internode length, below ear internode length, below ear node number, above ear node number using list of candidate genes for maize plant height traits. *P* values are shown on a $\log_{10}$ scale. The dashed horizontal line depicts the Bonferroni-adjusted correction based on the effective number of independent test (2.9×10^{-6}).

8.2. Supplemental Tables:

Table S1. Summary of imputation results, genome-wise and chromosome-wise effective number of tests for the WiDiV RNAseq genotype data

Chr. [1/]	No. SNPs	No. SNPs (>2 alleles)	% missing data (before imp.)	No. SNPS (after imp.)	% missing data (after imp.)	Effective No. Indep. tests
1	73,639	3,267	28.10	70,372	0.53	27,122
2	54,385	2,348	30.42	52,037	0.56	20,663
3	50,755	2,198	29.25	48,557	0.52	18,778
4	43,635	1,864	30.09	41,771	0.62	17,185
5	55,835	2,474	28.47	53,361	0.51	20,292
6	37,510	1,620	29.07	35,890	0.56	14,202
7	37,807	1,692	29.42	36,115	0.55	14,723
8	39,093	1,713	29.01	37,380	0.58	14,848
9	34,026	1,374	29.50	32,652	0.60	12,767
10	31,496	1,409	30.58	30,087	0.56	11,993
Total	458,181	19,959	29.39	438,222	0.56	172,573

Chr.: Chromosome, No.: numbers, imp.: imputation, Indep. Independent.

Table S2. Summary of the IBM mapping population genetic map. The IBM population consisted of 282 RILs

Chromosome	Number of Markers	Length (in cM)	Average spacing (in cM)	Maximum spacing (in cM)
1	1,320	939.5	0.71	5.9
2	876	666.7	0.76	10.5
3	1,000	713.1	0.71	7.1
4	903	627	0.69	7.1
5	859	627.2	0.73	6.3
6	621	459.1	0.73	10.9
7	663	593.5	0.89	33.8
8	761	571.8	0.75	8.4
9	636	524.6	0.82	5.1
10	536	421.6	0.79	5.9
Total	8,175	6,144.1	0.75	33.8

Table S3. Summary of the OW mapping population genetic map. The OW population consisted of 255 RILs

Chromosome	Number of Markers	Length (in cM)	Average spacing (in cM)	Maximum spacing (in cM)
1	913	472.6	0.52	2.7
2	638	353.3	0.55	5.1
3	587	328.0	0.56	2.7
4	610	331.1	0.54	4.3
5	576	335.2	0.58	9.2
6	475	262.0	0.55	3.5
7	528	289.8	0.55	3.5
8	497	269.7	0.54	3.3
9	440	253.5	0.58	3.2
10	419	271.5	0.65	6.5
Total	5,683	3,166.7	0.56	9.2

Table S4. Summary of the NyH mapping population genetic map. The NyH population consisted of 233 RILs

Chromosome	Number of Markers	Length (in cM)	Average spacing (in cM)	Maximum spacing (in cM)
1	851	498.9	0.59	4.7
2	674	380.9	0.57	6.5
3	610	352.1	0.58	4.8
4	547	292.7	0.53	2.6
5	574	314.4	0.54	2.2
6	376	218.5	0.58	3.0
7	480	262.4	0.55	3.5
8	407	232.0	0.57	2.8
9	448	264.6	0.59	7.1
10	321	189.9	0.59	3.0
Total	5,288	3,006.4	0.57	7.1

Table S5. Summary of the NAM22 mapping population genetic map. The NAM22 population consisted of 192 RILs

Chromosome	Number of Markers	Length (in cM)	Average spacing (in cM)	Maximum spacing (in cM)
1	175	202.4	1.16	9.1
2	127	155.7	1.23	8.9
3	130	155.5	1.20	15.0
4	111	141.1	1.27	7.2
5	139	153.4	1.10	14.3
6	78	109.7	1.41	8.6
7	78	135.0	1.73	14.8
8	106	129.8	1.22	10.6
9	85	114.8	1.35	12.6
10	77	101.9	1.32	8.2
Total	1,106	1,399.3	1.27	15.0

Table S6. QTL linkage mapping for plant height (PH), ear height (EH), internode number (IL), node number (NN), below ear internode length (BEIL), above ear internode length (AEIL), below ear node number (BENN) and above ear node number (AENN) evaluated in recombinants inbred lines from IBM, OW, NyH, and NAM populations

Pop.	Trait	Chr.	QTL	Closest Marker	Pos. (cM)	Position AGP_v2 (bp)	1.5-LOD Support Interval AGP_v2 (bp)	Interval Length (Mb)	LOD	*P*-value	A[a]	R^2 [b]	$LOD_{0.05}$ [c]
IBM	PH	1	qPH1a	bin_0441	351.4	85,150,000	79,950,000 – 88,050,000	8.10	7.8	3.62×10^{-9}	5.5	8.35	3.69
		1	qPH1b	bin_0703	469.8	183,850,000	173,050,000 -189,000,000	15.95	4.8	3.63×10^{-6}	-4.3	5.01	
		1	qPH1c	bin_0827	560.3	206,250,000	203,200,000 – 206,550,000	3.35	6.6	5.56×10^{-8}	4.9	7.00	
		3	qPH3a	bin_2740	322.8	145,550,000	127,500,000 – 146,850,000	19.35	5.5	7.03×10^{-7}	-4.6	5.78	
		3	qPH3b	bin_3186	670.4	223,950,000	221,650,000 – 225,200,000	3.55	4.2	1.60×10^{-5}	-3.9	4.32	
		6	qPH6	bin_5332	185.1	112,250,000	110,950,000 – 116,850,000	5.90	6.5	8.33×10^{-8}	5.0	6.81	
		9	qPH9	bin_7393	235.7	99,500,000	98,650,000 -105,050,000	6.40	11.1	1.89×10^{-12}	6.4	12.28	
IBM	EH	1	qEH1	bin_0867	590.3	213,750,000	211,950,000 – 215,400,000	3.45	6.0	1.90×10^{-7}	4.0	7.82	3.69
		3	qEH3a	bin_2294	90.7	7,800,000	5,750,000 – 9,550,000	3.80	5.4	8.72×10^{-7}	-3.7	6.93	
		3	qEH3b	bin_2937	454.6	183,300,000	179,250,000 – 183,550,000	4.30	5.5	5.94×10^{-7}	-3.9	7.15	
		9	qEH9	bin_7391	233.7	99,050,000	96,000,000 – 100,250,000	4.25	10.5	6.08×10^{-12}	5.3	14.21	
IBM	IL	1	qIL1	bin_1187	789.9	281,450,000	266,750,000 -283,550,000	16.80	5.3	1.10×10^{-6}	0.2	6.76	3.70
		9	qIL9a	bin_7084	31.0	5,250,000	4,650,000 – 6,400,000	1.75	6.4	7.40×10^{-8}	0.2	8.33	
		9	qIL9b	bin_7391	233.7	99,050,000	98,050,000 – 102,750,000	4.70	14.8	3.33×10^{-16}	0.4	20.78	
IBM	NN	1	qNN1a	bin_0118	126.8	16,150,000	15,250,000 – 17,050,000	1.80	5.9	2.84×10^{-7}	0.2	6.65	3.70
		1	qNN1b	bin_0704	471.0	183,950,000	183,150,000 – 295,450,000	112.30	4.6	6.67×10^{-6}	-0.2	5.06	
		3	qNN3	bin_2806	359.1	158,700,000	155,350,000 -160,150,000	4.80	4.0	2.56×10^{-5}	-0.2	4.39	
		4	qNN4	bin_3898	430.4	185,750,000	182,050,000 – 187,850,000	5.80	7.1	2.07×10^{-8}	-0.3	8.02	
		7	qNN7	bin_5910	204.8	101,050,000	13,900,000 – 109,150,000	95.25	5.5	8.44×10^{-7}	-0.2	6.09	
		8	qNN8	bin_6768	302.2	118,950,000	118,750,000 – 122,250,000	3.50	10.3	1.41×10^{-11}	-0.3	12.00	
		10	qNN10	bin_8180	346.2	145,050,000	144,650,000 – 146,150,000	1.50	5.1	1.69×10^{-6}	-0.2	5.74	
IBM	BEIL	1	qBEIL1a	bin_0443	352.0	85,900,000	75,850,000 -8,895,0000	13.10	4.5	7.55×10^{-6}	0.2	5.43	3.69
		1	qBEIL1b	bin_0862	587.7	212,850,000	211,950,000 – 214,200,000	2.25	5.5	6.30×10^{-7}	0.2	6.79	
		3	qBEIL3	bin_2305	103.8	9,450,000	7650,000 – 12,250,000	4.60	4.6	4.90×10^{-6}	-0.2	5.67	
		9	qBEIL9	bin_7391	233.7	99,050,000	96,000,000 – 105,350,000	9.35	11.8	3.41×10^{-13}	0.4	15.39	

Table S6: (continued… QTL linkage mapping)

Pop.	Trait	Chr.	QTL	Closest Marker	Position (cM)	Position AGP_v2 (bp)	1.5-LOD Support Interval AGP_v2(bp)	Interval Length (Mb)	LOD	*P*-value	A[a]	R^2 [b]	$LOD_{0.05}$ [c]
IBM	BENN	1	qBENN1	bin_0705	471.4	184,050,000	18,3150,000 – 295,450,000	112.3	5.1	1.92×10^{-6}	-0.2	6.74	3.67
		4	qBENN4	bin_3900	431.2	185,950,000	182,050,000 -187,850,000	5.80	4.9	3.00×10^{-6}	-0.2	6.48	
		8	qBENN8	bin_6770	303.2	119,150,000	118,750,000 – 123,900,000	5.15	4.2	1.31×10^{-5}	-0.2	5.61	
		9	qBENN9	bin_7663	486.3	152,750,000	151,750,000 – 153,650,000	1.90	3.9	3.25×10^{-5}	-0.2	5.08	
		10	qBENN10	bin_8180	346.2	145,050,000	142,850,000 – 146,250,000	3.40	4.6	5.94×10^{-6}	-0.2	6.07	
IBM	AENN	2	qAENN2	bin_1330	11.9	2,050,000	1,550,000 – 236,550,000	235.00	5.0	1.86×10^{-6}	-0.1	8.28	3.69
		3	qAENN3	bin_2932	447.4	182,700,000	177,650,000 -183,450,000	5.80	5.0	1.69×10^{-6}	0.1	8.34	
NAM22	PH	2	qPH2	L00268	77.1	124,057,780	40,503,046 – 147,310,459	106.81	3.5	8.75×10^{-5}	-2.8	5.0	3.00
		4	qPH4	L00763	61.8	147,915,924	35,918,862 – 167,190,374	131.27	8.6	6.11×10^{-10}	-4.7	13.24	
		5	qPH5	L00723	66.8	239,413,450	35,277,086 – 158,686,956	123.41	3.6	5.57×10^{-5}	-3.1	5.29	
		9	qPH9	L00599	74.2	140,970,753	135,267,973 – 14,534,7369	10.08	6.5	8.13×10^{-8}	4.1	9.7	
		10	qPH10	L00614	56.1	132,620,909	130,080,970 -139,978,756	9.90	3.5	7.82×10^{-5}	-2.9	5.08	
NAM22	EH	1	qEH1	L00187	116.4	203,947,082	195,557,990 – 212,843,739	17.29	4.1	1.83×10^{-5}	-2.5	6.42	3.00
		3	qEH3	L01161	25.4	6,739,219	4,862,341 – 12,132,193	7.27	3.9	3.28×10^{-5}	-2.4	6.01	
		4	qEH4	L00763	61.8	147,915,924	35,918,862 – 155,907,930	119.99	5.1	2.13×10^{-6}	-2.8	7.95	
		5	qEH5	L00214	66.8	75,935,660	33,285,364 – 145,869,341	112.58	3.8	3.90×10^{-5}	-2.4	5.89	
		10	qEH10	L00614	56.1	132,620,909	130,080,970 -139,978,756	9.90	6.6	6.58×10^{-8}	-3.2	10.51	
NyH	PH	6	qPH6	bin_3564	153.9	148,000,000	139,650,000 – 153,850,000	14.20	3.5	7.33×10^{-5}	-4.9	6.59	3.38
NyH	EH	2	qEH2	bin_1457	327.3	216,000,000	204,050,000 – 230,350,000	26.30	4.9	2.09×10^{-5}	3.0	6.42	3.37
		3	qEH3	bin_1598	45.7	12,300,000	4,300,000 – 184,600,000	180.3	3.8	3.49×10^{-5}	3.0	6.06	
		5	qEH5	bin_3184	240.8	200,250,000	193,950,000 – 204,900,000	10.95	5.4	7.70×10^{-7}	3.5	8.78	
OW	PH	1	qPH1a	bin_0095	64.8	16,900,000	13,750,000 -20,400,000	6.65	5.2	1.60×10^{-6}	-4.6	4.81	3.43
		1	qPH1b	bin_0273	151.6	56,150,000	46,200,000 – 58,850,000	12.65	5.7	5.67×10^{-7}	5.4	5.25	
		1	qPH1c	bin_0388	198.2	121,550,000	85,750,000 – 163,150,000	77.40	4.2	1.57×10^{-5}	4.7	3.86	
		1	qPH1d	bin_0591	289.6	219,400,000	217,100,000 – 229,350,000	12.25	5.3	1.39×10^{-6}	-4.6	4.87	
		3	qPH3	bin_1794	137.2	151,250,000	144,050,000 – 156,000,000	11.95	7.5	9.50×10^{-9}	-5.3	7.02	
		5	qPH5a	bin_2916	125.6	32,900,000	15,350,000 -45,700,000	30.35	6.7	5.84×10^{-8}	5.1	6.23	
		5	qPH5b	bin_3254	273	208,550,000	204,750,000 – 211,050,000	6.30	5.8	3.97×10^{-7}	4.7	5.4	
		6	qPH6	bin_3507	81.7	101,600,000	87,650,000 – 106,050,000	18.4	6.1	2.81×10^{-7}	-4.7	5.55	
		10	qPH10	bin_5439	108.1	68,650,000	65,600,000 – 70,750,000	5.15	9.5	1.05×10^{-10}	6.2	9.07	

Table S6: (continued…QTL linkage mapping)

Pop.	Trait	Chr.	QTL	Closest Marker	Position (cM)	Position AGP_v2 (bp)	1.5-LOD Support Interval AGP_v2 (bp)	Interval Length (Mb)	LOD	*P*-value	A[a]	R^{2} [b]	$LOD_{0.05}$ [c]
OW	EH	1	qEH1a	bin_0083	57.1	14,500,000	3,000,000 – 17,250,000	14.25	4.1	2.06×10^{-5}	-2.9	4.14	3.42
		1	qEH1b	bin_0358	187	93,100,000	85,750,000 -103,800,000	18.05	16.4	2.00×10^{-16}	6.1	18.86	
		2	qEH2	bin_1500	306.1	228,350,000	223,150,000 – 234,550,000	11.40	3.7	4.41×10^{-5}	2.8	3.80	
		5	qEH5a	bin_2938	135.2	45,000,000	24,800,000 – 80,050,000	55.25	9.2	1.45×10^{-10}	4.5	9.85	
		5	qEH5b	bin_3228	257.5	204,950,000	193,750,000 – 211,050,000	17.30	5.3	2.33×10^{-6}	3.3	5.14	
		10	qEH10	bin_5515	141.3	116,550,000	24,050,000 – 122 200,000	98.15	5.0	2.48×10^{-6}	3.2	5.11	
OW	IL	6	qIL6	bin_3506	81.3	100,750,000	90,500,000 -106,950,000	16.45	5.4	7.94×10^{-7}	-0.4	8.33	3.41
		10	qIL10	bin_5438	107.3	68,050,000	22,700,000 – 129,350,000	106.65	8.0	1.61×10^{-9}	0.5	12.76	
OW	NN	1	qNN1	bin_0344	182.1	86,150,000	82,800,000 – 93,500,000	10.70	12.7	3.77×10^{-14}	0.4	15.16	3.43
		3	qNN3	bin_1811	142.9	158,100,000	150,400,000 – 160,150,000	9.75	10.5	5.61×10^{-12}	-0.4	12.3	
		5	qNN5a	bin_2903	121.1	24,800,000	22,850,000 – 154,050,000	131.20	6.9	2.52×10^{-8}	0.3	7.77	
		5	qNN5b	bin_3270	284.3	210,650,000	208,150,000 -211,250,000	3.10	4.0	2.42×10^{-5}	0.2	4.34	
OW	BEIL	1	qBEIL1a	bin_0081	56.3	13,950,000	12,350,000 – 16,300,000	3.95	4.7	4.47×10^{-6}	-0.4	6.51	3.37
		1	qBEIL1b	bin_0313	171.3	72,850,000	62,200,000 – 128,100,000	65.90	5.8	3.45×10^{-7}	0.4	8.11	
		6	qBEIL6	bin_3704	186.8	154,950,000	91,900,000 – 162,850,000	70.95	3.4	1.10×10^{-4}	-0.3	4.57	
		10	qBEIL10	bin_5505	137.4	113,700,000	85,150,000 – 122,200,000	37.05	7.3	1.01×10^{-8}	0.5	10.40	
OW	AEIL	6	qAEIL6	bin_3506	81.3	100,750,000	78,450,000 -109,700,000	31.25	3.5	7.29×10^{-5}	-0.5	5.74	3.38
		10	qAEIL10	bin_5439	108.1	68,650,000	19,550,000 – 130,400,000	110.85	4.7	3.99×10^{-6}	0.6	7.85	
OW	BENN	1	qBENN1	bin_0353	185.6	89,650,000	84,950,000 – 97,950,000	13.00	8.8	2.90×10^{-10}	0.3	11.29	3.40
		3	qBENN3a	bin_1615	57.7	8,850,000	4,400,000 – 9,250,000	4.85	5.2	1.43×10^{-6}	0.2	6.38	
		3	qBENN3b	bin_1813	143.7	158,800,000	147,950,000 – 161,550,000	13.60	5.6	5.23×10^{-7}	-0.2	6.94	
		5	qBENN5	bin_2991	155.3	80,050,000	24,500,000 -159,000,000	134.50	7.3	1.06×10^{-8}	0.3	9.16	
OW	AENN	1	qAENN1	bin_0612	298.6	228,100,000	217,850,000 – 229,800,000	11.95	5.1	2.19×10^{-6}	-0.1	5.54	3.42
		2	qAENN2	bin_0967	45.1	7,450,000	5,750,000 – 236,750,000	231.0	4.5	7.68×10^{-6}	0.1	4.92	
		3	qAENN3	bin_1816	145.3	159,750,000	150,400,000 – 162,350,000	11.95	9.2	1.58×10^{-10}	-0.2	10.53	
		4	qAENN4	bin_2578	217.5	193,300,000	191,000,000 – 197,450,000	6.45	6.2	1.62×10^{-7}	-0.2	6.86	
		6	qAENN6a	bin_3679	170.3	151,250,000	147,950,000 -153,900,000	5.95	3.7	4.62×10^{-5}	-0.1	4.05	
		6	qAENN6b	bin_3790	253.3	167,350,000	165,650,000 – 167,650,000	2.00	8.5	9.19×10^{-10}	0.2	9.57	
		8	qAENN8	bin_4382	40.4	8,450,000	6,050,000 – 140,550,000	134.5	3.6	6.20×10^{-5}	0.1	3.91	

[a] Additive effect: positive values indicate that the parents alleles are increasing the traits and vice versa for negative values

[b] R^2: contribution rate

[c] $LOD_{0.05}$: logarithm of odds at threshold at $P < 0.05$ genome wide risk level.

Table S7: Overlap between QTL interval mapped for plant height-related traits in the OW, IBM, NAM22 and NyH populations

Trait	Pop.	Chr.	QTL	Closest Marker	Pos. (cM)	position (bp)	Distance between Overlap QTL (bp)	1.5-LOD Support Interval AGP_v2bp)	Interval Length (Mb)	LOD	*P*-value	A[a]	R^{2b}
PH	OW	3	qPH3a	bin_1794	137.20	151,250,000	5,700,000	144,050,000 – 156,000,000	11.95	7.5	9.50×10^{-9}	-5.3	7.02
	IBM	3	qPH3a	bin_2740	322.80	145,550,000		127,500,000 – 146,850,000	19.35	5.5	7.03×10^{-7}	-4.6	5.78
PH	OW	5	qPH5a	bin_2916	125.60	32,900,000	206,513,450	15,350,000 – 45,700,000	30.35	6.7	5.84×10^{-8}	5.1	6.23
	NAM22	5	qPH5	L00723	66.80	239,413,450		35,277,086 – 158,686,956	123.41	3.6	5.57×10^{-5}	-3.1	5.29
EH	NAM22	1	qEH1	L00187	116.40	203,947,082	9,802,918	195,557,990 – 212,843,739	17.29	4.1	1.83×10^{-5}	-2.5	6.42
	IBM	1	qEH1	bin_0867	590.30	213,750,000		211,950,000 – 215,400,000	3.45	6.0	1.90×10^{-7}	4.0	7.82
EH	NyH	2	qEH2	bin_1457	327.30	216,000,000	12,350,000	204050000 – 230,350,000	26.30	4.0	2.09×10^{-5}	3.0	6.42
	OW	2	qEH2	bin_1500	306.10	228,350,000		223,150,000 – 234,550,000	11.40	3.7	4.41×10^{-5}	2.8	3.80
EH	NAM22	3	qEH3	L01161	25.40	6,739,219	1,060,781	4,862,341 – 12,132,193	7.27	3.9	3.28×10^{-5}	-2.4	6.01
	IBM	3	qEH3a	bin_2294	90.70	7,800,000		5,750,000 – 9,550,000	3.80	5.4	8.72×10^{-7}	-3.7	6.93
EH	NyH	5	qEH5	bin_3184	240.80	200,250,000	4,700,000	193,950,000 – 204,900,000	10.95	5.4	7.70×10^{-7}	3.5	8.78
	OW	5	qEH5b	bin_3228	257.50	204,950,000		193,750,000 – 211,050,000	17.30	5.0	2.33×10^{-6}	3.3	5.14
BEIL	OW	1	qBEIL1b	bin_0313	171.30	72,850,000	13,050,000	62,200,000 – 128,100,000	65.9	5.8	3.45×10^{-7}	0.4	8.11
	IBM	1	qBEIL1a	bin_0443	352.00	85,900,000		75,850,000 – 88,950,000	13.1	4.5	7.55×10^{-6}	0.2	5.43

[a] Additive effect: positive values indicate that the parents alleles are increasing the traits and vice versa for negative values

[b] R^2: contribution rate

Table S8: Overlap between QTL interval mapped for plant height-related traits in the OW, IBM, NAM22 and NyH populations

Pop.	Trait	Chr.	QTL	Closest Marker	Position (cM)	Position AGP_v2 (bp)	1.5-LOD Support Interval AGP_v2(bp)	Interval Length (Mb)	LOD	*P*-value	A[a]	R^{2b}
OW	BEIL	1	qBEIL1a	bin_0081	56.3	13,950,000	12,350,000 – 16,300,000	3.95	4.7	4.47×10^{-6}	-0.4	6.51
OW	EH	1	qEH1a	bin_0083	57.1	14,500,000	3,000,000 – 17,250,000	14.25	4.1	2.06×10^{-5}	-2.9	4.14
IBM	NN	1	qNN1a	bin_0118	126.8	16,150,000	15,250,000 – 17,050,000	1.80	5.9	2.84×10^{-7}	0.2	6.65
OW	PH	1	qPH1a	bin_0095	64.8	16,900,000	13,750,000 -20,400,000	6.65	5.2	1.60×10^{-6}	-4.6	4.81
IBM	PH	1	qPH1a	bin_0441	351.4	85,150,000	79,950,000 – 88,050,000	8.10	7.8	3.62×10^{-9}	5.5	8.35
IBM	BEIL	1	qBEIL1a	bin_0443	352	85,900,000	75,850,000 -88,950,000	13.10	4.5	7.55×10^{-6}	0.2	5.43
OW	NN	1	qNN1a	bin_0344	182.1	86,150,000	82,800,000 – 93,500,000	10.70	12.7	3.77×10^{-14}	0.4	15.16
OW	BENN	1	qBENN1a	bin_0353	185.6	89,650,000	84,950,000 – 97,950,000	13.00	8.8	2.90×10^{-10}	0.3	11.29
OW	EH	1	qEH1b	bin_0358	187	93,100,000	85,750,000 -103,800,000	18.05	16.4	2.00×10^{-16}	6.1	18.86
NAM22	EH	1	qEH1	L00187	116.4	203,947,082	195,557,990 – 212,843,739	17.29	4.1	1.83×10^{-5}	-2.5	6.42
IBM	PH	1	qPH1c	bin_0827	560.3	206,250,000	203,200,000 – 206,550,000	3.35	6.6	5.56×10^{-8}	4.9	7.00
IBM	BEIL	1	qBEIL1b	bin_0862	587.7	212,850,000	211,950,000 – 214,200,000	2.25	5.5	6.30×10^{-7}	0.2	6.79
IBM	EH	1	qEH1	bin_0867	590.3	213,750,000	211,950,000 – 215,400,000	3.45	6.0	1.90×10^{-7}	4	7.82
NAM22	EH	3	qEH3	L01161	25.4	6,739,219	4,862,341 – 12,132,193	7.27	3.9	3.28×10^{-5}	-2.4	6.01
IBM	EH	3	qEH3a	bin_2294	90.7	7,800,000	5,750,000 – 9,550,000	3.80	5.4	8.72×10^{-7}	-3.7	6.93
OW	BENN	3	qBENN3b	bin_1615	57.7	8,850,000	4,400,000 – 9,250,000	4.85	5.2	1.43×10^{-6}	0.2	6.38
IBM	BEIL	3	qBEIL3	bin_2305	103.8	9,450,000	7,650,000 – 12,250,000	4.60	4.6	4.90×10^{-6}	-0.2	5.67
IBM	PH	3	qPH3a	bin_2740	322.8	145,550,000	127,500,000 – 146,850,000	19.35	5.5	7.03×10^{-7}	-4.6	5.78
OW	PH	3	qPH3	bin_1794	137.2	151,250,000	144,050,000 – 156,000,000	11.95	7.5	9.50×10^{-9}	-5.3	7.02
OW	NN	3	qNN3b	bin_1811	142.9	158,100,000	150,400,000 – 160,150,000	9.75	10.5	5.61×10^{-12}	-0.4	12.3
IBM	NN	3	qNN3	bin_2806	359.1	158,700,000	155,350,000 -160,150,000	4.80	4	2.56×10^{-5}	-0.2	4.39
OW	BENN	3	qBENN3c	bin_1813	143.7	158,800,000	147,950,000 – 161,550,000	13.60	5.6	5.23×10^{-7}	-0.2	6.94
OW	AENN	3	qAENN3c	bin_1816	145.3	159,750,000	150,400,000 – 162,350,000	11.95	9.2	1.58×10^{-10}	-0.2	10.53
IBM	AENN	3	qAENN3	bin_2932	447.4	182,700,000	177,650,000 - 183,450,000	5.80	5.0	1.69×10^{-6}	0.1	8.34
IBM	EH	3	qEH3b	bin_2937	454.6	183,300,000	179,250,000 – 183,550,000	4.30	5.5	5.94×10^{-7}	-3.9	7.15
IBM	NN	4	qNN4	bin_3898	430.4	185,750,000	182,050,000 – 187,850,000	5.80	7.1	2.07×10^{-8}	-0.3	8.02
IBM	BENN	4	qBENN4	bin_3900	431.2	185,950,000	182,050,000 -187,850,000	5.80	4.9	3.00×10^{-6}	-0.2	6.48
NyH	EH	5	qEH5	bin_3184	240.8	200,250,000	193,950,000 – 204,900,000	10.95	5.4	7.70×10^{-7}	3.5	8.78
OW	EH	5	qEH5b	bin_3228	257.5	204,950,000	193,750,000 – 211,050,000	17.30	5.3	2.33×10^{-6}	3.3	5.14
OW	PH	5	qPH5	bin_3254	273	208,550,000	204,750,000 – 211,050,000	6.30	5.8	3.97×10^{-7}	4.7	5.40
OW	NN	5	qNN5d	bin_3270	284.3	210,650,000	208,150,000 -211,250,000	3.10	4	2.42×10^{-5}	0.2	4.34
OW	IL	6	qIL6a	bin_3506	81.3	100,750,000	90,500,000 -106,950,000	16.45	5.4	7.94×10^{-7}	-0.4	8.33
OW	PH	6	qPH6	bin_3507	81.7	101,600,000	87,650,000 – 106,050,000	18.40	6.1	2.81×10^{-7}	-4.7	5.55
NyH	PH	6	qPH6	bin_3564	153.9	148,000,000	139,650,000 – 153,850,000	14.20	3.5	7.33×10^{-5}	-4.9	6.59
OW	AENN	6	qAENN6c	bin_3679	170.3	151,250,000	147,950,000 -153,900,000	5.95	3.7	4.62×10^{-5}	-0.1	4.05
IBM	NN	8	qNN8	bin_6768	302.2	118,950,000	118,750,000 – 122,250,000	3.50	10.3	1.41×10^{-11}	-0.3	12.00
IBM	BENN	8	qBENN8	bin_6770	303.2	119,150,000	118,750,000 – 123,900,000	5.15	4.2	1.31×10^{-5}	-0.2	5.61
IBM	BEIL	9	qBEIL9	bin_7391	233.7	99,050,000	96,000,000 – 105,350,000	9.35	11.8	3.41×10^{-13}	0.4	15.39
IBM	EH	9	qEH9	bin_7391	233.7	99,050,000	96,000,000 – 100,250,000	4.25	10.5	6.08×10^{-12}	5.3	14.21
IBM	IL	9	qIL9b	bin_7391	233.7	99,050,000	98,050,000 – 102,750,000	4.70	14.8	3.33×10^{-16}	0.4	20.78
IBM	PH	9	qPH9	bin_7393	235.7	99,500,000	98,650,000 -105,050,000	6.40	11.1	1.89×10^{-12}	6.4	12.28
NAM22	EH	10	qEH10	L00614	56.1	132,620,909	130,080,970 -139,978,756	9.90	6.6	6.58×10^{-8}	-3.2	10.51
NAM22	PH	10	qPH10	L00614	56.1	132,620,909	130,080,970 -139,978,756	9.90	3.5	7.82×10^{-5}	-2.9	5.08
IBM	BENN	10	qBENN10	bin_8180	346.2	145,050,000	142,850,000 – 146,250,000	3.40	4.6	5.94×10^{-6}	-0.2	6.07
IBM	NN	10	qNN10	bin_8180	346.2	145,050,000	144,650,000 – 146,150,000	1.50	5.1	1.69×10^{-6}	-0.2	5.74

Table S8: SNPs and genes modassociated significantly with maize plant height traits of the Wisconsin Diverse population

Gene Models	Trait	Chr.	Physical Position	MAF	*P*-value	Functional Description
GRMZM2G171622	PH	3	158,975,082	0.46	3.99×10^{-7}	CBS domain containing protein
GRMZM2G171622	PH	3	158,974,646	0.44	4.37×10^{-7}	CBS domain containing protein
GRMZM2G171622	PH	3	158,974,189	0.22	1.23×10^{-7}	CBS domain containing protein
GRMZM2G171622	EH	3	158,974,189	0.23	1.46×10^{-7}	CBS domain containing protein
GRMZM2G171622	EH	3	158975,082	0.46	1.88×10^{-7}	CBS domain containing protein
GRMZM2G171622	EH	3	158,974,360	0.50	3.63×10^{-7}	CBS domain containing protein
GRMZM2G171622	EH	3	158,974,646	0.44	3.94×10^{-7}	CBS domain containing protein
GRMZM2G171622	EH	3	158,974,409	0.26	1.03×10^{-6}	CBS domain containing protein
GRMZM2G171622	NN	3	158,974,409	0.26	5.01×10^{-8}	CBS domain containing protein
GRMZM2G171622	NN	3	158,975,082	0.46	2.89×10^{-7}	CBS domain containing protein
GRMZM2G171622	NN	3	158,974,360	0.50	5.96×10^{-7}	CBS domain containing protein
GRMZM2G171622	NN	3	158,974,646	0.44	1.51×10^{-6}	CBS domain containing protein
GRMZM2G082185	NN	8	6,467,823	0.38	4.20×10^{-7}	Hypothetical protein LOC100280321
GRMZM2G363460	NN	8	135,457,262	0.18	8.62×10^{-7}	LOC100283640
GRMZM2G171622	BENN	3	158,974,409	0.26	1.93×10^{-8}	CBS domain containing protein
GRMZM2G171622	BENN	3	158,974,189	0.22	1.66×10^{-7}	CBS domain containing protein
GRMZM2G082185	BENN	8	6,467,823	0.39	2.32×10^{-6}	Hypothetical protein LOC100280321
GRMZM2G070943	BENN	1	96,310,277	0.13	9.09×10^{-7}	Hypothetical protein LOC100381745
GRMZM2G363460	BENN	8	135,457,262	0.18	2.29×10^{-6}	LOC100283640
GRMZM2G129071	AEIL	10	141,019,997	0.16	2.03×10^{-6}	Hypothetical protein LOC100279590
GRMZM2G129071	AEIL	10	141,020,053	0.16	3.14×10^{-6}	Hypothetical protein LOC100279590
GRMZM2G415229	AEIL	10	141,021,407	0.16	4.06×10^{-6}	Hypothetical protein LOC100384464
GRMZM2G171622	DTP	3	158,975,082	0.46	3.44×10^{-9}	CBS domain containing protein
GRMZM2G171622	DTP	3	158,974,360	0.49	5.99×10^{-8}	CBS domain containing protein
GRMZM2G171622	DTP	3	158,974,646	0.45	2.93×10^{-7}	CBS domain containing protein
GRMZM2G171622	DTP	3	158,974,409	0.27	7.92×10^{-7}	CBS domain containing protein
GRMZM2G171622	DTS	3	158,975,082	0.46	1.55×10^{-8}	CBS domain containing protein
GRMZM2G171622	DTS	3	158,974,360	0.49	7.48×10^{-7}	CBS domain containing protein
GRMZM2G171622	DTS	3	158,974,646	0.46	1.24×10^{-6}	CBS domain containing protein
GRMZM2G169089	DTS	3	104,865,691	0.15	2.06×10^{-6}	Diacylglycerol acyltransferase