Mol Breeding
DOI 10.1007/s11032-011-9631-5

Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize

Yanli Lu · Jie Xu · Zhimin Yuan · Zhuanfang Hao · Chuanxiao Xie · Xinhai Li · Trushar Shah · Hai Lan · Shihuang Zhang · Tingzhao Rong · Yunbi Xu

Received: 5 June 2011 / Accepted: 26 August 2011

Abstract Drought often delays developmental events so that plant height and above-ground biomass are reduced, resulting in yield loss due to inadequate photosynthate. In this study, plant height and biomass measured by the Normalized Difference Vegetation Index (NDVI) were used as criteria for drought tolerance. A total of 305 lines representing temperate, tropical and subtropical maize germplasm were genotyped using two single nucleotide polymorphism (SNP) chips each containing 1536 markers, from which 2052 informative SNPs and 386 haplotypes each constructed with two or more SNPs were used for linkage disequilibrium (LD) or association mapping. Single SNP- and haplotype-based LD mapping identified two significant SNPs and three haplotype loci [a total of four quantitative trait loci (QTL)] for plant height under well-watered and water-stressed conditions. For biomass, 32 SNPs and 12 haplotype loci (30 QTL) were identified using NDVIs measured at seven stages under the two water regimes. Some significant SNP and haplotype loci for NDVI were shared by different stages. Comparing significant loci identified by single SNP- and haplotype-based LD mapping, we found that six out of the 14 chromosomal regions defined by haplotype loci each included at least one significant SNP for the same trait. Significant SNP haplotype loci explained much higher phenotypic variation than individual SNPs. Moreover, we found that two significant SNPs (two QTL) and one haplotype locus were shared by plant height and NDVI. The results indicate the power of comparative LD mapping using single SNPs and SNP haplotypes with QTL shared by plant height and biomass as secondary traits for drought tolerance in maize.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-011-9631-5) contains supplementary material, which is available to authorized users.

Y. Lu · J. Xu · Z. Yuan · H. Lan · T. Rong (✉)
Maize Research Institute, Sichuan Agricultural University, Wenjiang 611130, Sichuan, China
e-mail: rongtz@sicau.edu.cn

Y. Lu · Y. Xu
CIMMYT, Km 45, Carretera, Mexico-Veracruz, El Batan, Texcoco, Mexico

Z. Hao · C. Xie · X. Li · S. Zhang
Institue of Crop Science, National Key Facilities for Crop Genetic Resources and Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China

T. Shah
International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Hyderabad 502324, Andhra Pradesh, India

Y. Xu (✉)
Institute of Crop Science/International Maize and Wheat Improvement Center (CIMMYT), National Key Facilities for Crop Genetic Resources and Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: y.xu@cgiar.org

Published online: 22 September 2011

Keywords Maize · Drought tolerance · Secondary traits · Plant height · Biomass · Single nucleotide polymorphism (SNP) · Haplotype · LD mapping · Association mapping

Abbreviations

abph1	Aberrant phyllotaxy1
ALDH	Aldehyde dehydrogenase
ASI	Anthesis–silking interval
CAAS	Chinese Academy of Agricultural Sciences
CIMMYT	International Maize and Wheat Improvement Center
DAP	Days after planting
DT chip	Drought tolerance candidate gene based chip
GWAS	Genome-wide association study
HP	Haplotype
IL	Introgression line
LD	Linkage disequilibrium
MAF	Minor allele frequency
MLM	Mixed linear model
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared region
OPA	Oligo pool assay
PC	Principal component
PCA	Principal component analysis
PH	Plant height
PIC	Polymorphism information content
RA chip	Random candidate gene based chip
SNP	Single nucleotide polymorphism
VIS	Visible region
WS	Water-stressed
WW	Well-watered

Introduction

Drought is a serious agronomic problem, and one of the most important factors limiting maize biomass and seed production in almost all areas where it is grown. Morphological, physiological and molecular responses to water deficits in maize have been associated with productivity under drought conditions. Plant shoots and roots respond to water deficits differently. To maintain root growth under water stress, plant roots, as an important organ for water absorption, redirect their growth and dry matter accumulation away from the shoot to the root, which allows an increase in cell wall extensibility in roots, mediated by an increased expression level of expansions, xyloglucan endotransglucosylase/hydrolases and other wall-loosening factors at the root tip (Ribaut et al. 2009). On the contrary, the shoot growth, which can be measured by plant height and biomass above ground, is inhibited. Under severe drought conditions, the growth fully stops and the shoots can be damaged.

Plant growth inhibited by a limited water supply has a serious impact on many physiological processes through the reduction in photosynthesis associated with leaf expansion, leaf rolling (Bolanos et al. 1993) and foliar senescence (Wolfe et al. 1988), and reduction in carbon fixation per unit leaf area because of stomatal closure or a decline in carboxylation capacity (Bruce et al. 2002). An easily measured increase in the anthesis–silking interval (ASI) as silk emergence is delayed is shown most directly and obviously under stress, resulting in grain abortion (Edmeades et al. 1993, 2000). Drought often delays developmental events so that plant height and above-ground biomass are reduced. Consequently, yield is also affected since the plant needs to reach a sufficient stature to have adequate photosynthate. Therefore, continuous increase of plant height and biomass can be used as criteria for drought tolerance.

The Normalized Difference Vegetation Index (NDVI) is a simple numerical indicator that can be used to analyze remote sensing measurements and assess whether the target being observed contains live green vegetation or not. The measurements of reflectance taken in the visible region (VIS; band: 0.58–0.68 μm) and the near infrared region (NIR; band: 0.725–1.0 μm) are used to calculate the index: NDVI = (NIR − VIS)/(NIR + VIS). The differential reflectance provides a means of monitoring density and vigour of green vegetation growth. Green leaves commonly have larger reflectance in the near infrared than in the visible range. Vegetation NDVI typically ranges from 0.1 up to 0.6, with higher values associated with greater density and greenness of the plant canopy (http://www.bom.gov.au/sat/NDVI/NDVI2.shtml). Under water stress, leaves become more yellow or die back, and reflect significantly less in the near infrared range. Biomass as measured nondestructively by NDVI is basically proportional to the leaf area of the population, which is a functional

part of photosynthesis and a determinant of the final grain yield (Lu et al. 2011a). In this study, we used NDVI as one of the criteria for drought tolerance.

Genetic mapping has been an important approach to identifying genes and genomic regions for complex traits including drought tolerance. Linkage disequilibrium (LD) mapping, also known as association mapping, has been used in quantitative trait locus (QTL) mapping for plants (Thornsberry et al. 2001; Aranzana et al. 2005; Malosetti et al. 2007; Casa et al. 2008; Zhu et al. 2008; Huang et al. 2010). In maize, associations between *Dwarf8* polymorphisms and flowering time were identified (Thornsberry et al. 2001) and the genetic basis of important leaf architecture traits was determined with some key genes identified through a genome-wide association study (GWAS) of the maize nested association mapping panel (Tian et al. 2011). In rice, association analysis between microsatellite markers and multiple agronomic traits was conducted (Wen et al. 2009; Yan et al. 2009b), and GWAS for 14 agronomic traits using 373 *indica* lines and 671,355 single nucleotide polymorphisms (SNPs) was performed (Huang et al. 2010).

Due to high genome density, low mutation rate, and amenability to high-throughput detection systems, SNPs have recently become the marker of choice for genetics and breeding applications. With the advent of next-generation sequencing technology, a large number of genotypes from each plant species including maize can be resequenced, and thus SNP detection and allele discovery throughout the whole genome become possible. In this study, two 1536-SNP chips were used to genotype a mapping panel containing a large set of diverse maize inbred lines. One chip was designed based on random candidate genes (RA chip), which were chosen for study without any prior knowledge or consideration of the function of the proteins (or RNAs) that they encode. It provides a random sample of genes across the genome for background control. The other chip (DT chip) was designed based on genes of known or suspected function from putative drought-related metabolic pathways as described by Yan et al. (2009a) and Lu et al. (2010). The candidate gene approach targets genes with known functions related to the traits of interest to increase the likelihood of finding meaningful trait associations (Setter et al. 2010), providing a great opportunity to conduct candidate-gene-based LD mapping for drought resistance. The mapping panel used in this study contains tropical/subtropical and temperate maize germplasm from diverse breeding programs, so the allelic variation identified is more likely to be useful to maize improvement. To overcome the limited resolution at each marker locus caused by the biallelic variation of SNP markers and to increase the power of SNP -based genetic mapping, we developed a strategy using haplotypes constructed with several consecutive SNPs, comparing with single SNP-based mapping. The objectives of this study were to (1) identify genomic loci or genes that are significantly associated with plant height and biomass (NDVI) under well-watered (WW) and water-stressed (WS) conditions, and (2) compare significant loci identified by single SNP- and haplotype-based LD mapping to understand their mapping efficiency.

Materials and methods

Plant materials

A total of 305 maize lines representing temperate, tropical and subtropical maize germplasm were selected from diverse maize breeding programs; most of them are tolerant to biotic or abiotic stresses as identified in previous experiments during 2005–2007 at the International Maize and Wheat Improvement Center (CIMMYT), Mexico and the Chinese Academy of Agricultural Sciences (CAAS), Beijing. The tested lines included 189 lines mainly from CIMMYT maize breeding programs based in Mexico, Zimbabwe and Kenya, and 116 lines from CAAS and Sichuan Agricultural University of China. All the CIMMYT maize lines and six of the Chinese lines tested were tropical/subtropical type, and the remaining 110 lines from China were temperate type. Of 305 lines, 105 were introgression lines (ILs) from the CAAS drought tolerance breeding program. Development of the 105 ILs was described in detail by Hao et al. (2009). All the maize inbred lines used in this study are detailed in Electronic Supplementary Material Table S1.

Field experiment and phenotyping

Experimental design and trait measurement in the field trial were implemented during the dry winter season

(November 2007–April 2008) at Tlaltizapan Field Station of CIMMYT, Mexico (18°41′N lat.; 940 masl), when the precipitation is very limited and water supply is fully dependent on irrigation. For evaluation of drought tolerance, all entries were tested under two water regimes (treatments) (well-watered, WW and water-stressed, WS) using an alpha (0,1) lattice field design, each with two replications. The inbred lines were planted in one-row plots with rows 5 m long, 26 plants per row/plot. All the plants received the first three irrigations [1, 18 and 41 days after planting (DAP)]. After this period, irrigation was applied every 2 weeks to the WW regime. The water stress at both vegetative and reproductive stages was imposed by reduced irrigation (once at 65 DAP) and then no irrigation was applied until flowering was completed. Immediately after the flowering period (around 101 DAP), all the trials were well irrigated to encourage adequate development of the kernels that had been set. More information about field experimental design and managed drought trials is given in our recent drought evaluation report (Lu et al. 2011a). In this study, plant height and the NDVIs representing biomass at different stages were scored for both water regimes. For plant height, the average height of six plants from ground to the tassel tip was scored randomly in each plot. Measurements for NDVI were taken at time intervals of about 2 weeks on 43, 51, 64, 78, and 92 DAP during the period between January 9 and February 27, 2008, which were named as stages I, II, III, IV and V, respectively. Multiple measurements of NDVI were taken with a portable spectroradiometer (GreenSeeker). NDVI measurements at 92 DAP were taken twice (in the morning–before leaf rolling, and afternoon–after leaf rolling), which were named as stages Va and Vb. The difference measured between morning and afternoon was shown as V(a–b) and used as an indicator of leaf rolling.

SNP genotyping and haplotype construction

Genomic DNA for each sample was extracted from the seedling leaves by the CTAB method (CIMMYT Applied Molecular Genetics Laboratory, 2003). All samples were genotyped with two 1536-SNP chips via Illumina GoldenGate assay (Fan et al. 2006; Lu et al. 2009) established at the Cornell University Life Sciences Core Laboratories Center. The SNP dataset was analyzed using Illumina BeadStudio genotyping software, which can cluster and call the allele for each locus automatically. Data not automatically dealt with by the software were manually scored. SNP markers that were monomorphic, duplicated in two chips or had more than 20% missing data points in the mapping panel of 305 lines were excluded from further analyses. As a result, a total of 2052 informative markers were scored with high quality, which included 1035 markers from the RA chip and 1017 markers from the DT chip. A detailed list of all SNPs from two chips can be found at http://www.panzea.org/.

As biallelic markers, SNPs per se cannot fully use the high level of genetic diversity existing within a diverse mapping panel, which has been considered less informative compared to simple sequence repeat (SSR) markers with multiple alleles. Using haplotypes derived from multiple SNPs provides a high level of allele diversity, thus increasing genotyping resolution. Of 2052 SNP markers, 29 SNPs had multiple hits to the genome and 32 had unknown positions. The remaining 1991 SNPs that had unique positions were mapped in silico onto the maize physical map and therefore used for haplotype scoring. Yan et al. (2009a) reported that the average distance of LD decay across the whole genome in maize was 5–10 kb within a highly diverse global maize collection of 632 inbred lines from temperate, tropical and subtropical breeding programs. Moreover, SNP sequence annotation analysis indicated that the size of gene harboring several SNPs did not exceed 10 kb. Using 447 diverse maize inbreds, Lu et al. (2011b) found that the average LD decay distances in the temperate and tropical germplasm collections were 10–100 kb and 5–10 kb, respectively. Based on these estimates of LD decay distance, the relative distances between SNPs were added up and SNPs within a total of 10 kb distance, which were used to construct haplotypes and assigned to the same locus (Yan et al. 2009a; Lu et al. 2010; 2011b), were also used in this study. Haplotype variation within such a locus was recorded as alleles. As a result, 1003 SNPs more than 10 kb away from any other SNPs were considered as ‘unlinked’ markers and excluded from the haplotype analysis. From the remaining SNPs, 386 haplotypic loci were identified, each consisting of two or more SNPs. If the genotype of any SNP at a haplotypic locus was missing in an individual, then the entire locus was treated as missing for that individual.

Genetic diversity and structure analysis

PowerMarker V3.25 software (Liu and Muse, 2005) were used to calculate observed heterozygosity, allele number, allele frequency and polymorphism information content (PIC). Although the tested maize inbreds were normally pure lines, heterozygous genotypes were still observed (less than 0.5%), which were considered missing in the study. For haplotype loci, alleles with frequency less than 5% of the most frequent allele were defined as rare alleles and excluded from further analysis.

In order to control spurious associations, a composite approach that integrates population structure and relatedness between individuals was used in LD mapping. Principal component analysis (PCA) was conducted to visualize genetic structure and generate principal component (PC) matrices for 305 maize lines using TASSEL 2.1 (Yu et al. 2006; Bradbury et al. 2007). According to the same set of marker data, family relatedness within populations (kinship matrix) was calculated using TASSEL 2.1.

Linkage disequilibrium and trait mapping

LD extents in the 305 maize lines were evaluated using the software package TASSEL2.1 (Yu et al. 2006; Bradbury et al. 2007). The squared correlation coefficient (r^2), representing the correlation between alleles at two loci, and its corresponding P-value, was estimated for each pair of SNP markers. The SNPs with minor allele frequency (MAF) less than 0.05 were excluded from the analysis.

Marker–trait association was calculated using the mixed linear model (MLM) that combines both population structure (PC-matrix) and level of kinship (K-matrix) in TASSEL2.0 (Yu et al. 2006). The first five PCs derived from 305 lines were included in the MLM analysis to replace the commonly used Q-matrix following Yu et al. (2006) and Lu et al. (2010). Marker data for LD mapping included the 2052 informative SNPs and 386 SNP haplotypes. Both PC- and K-matrices were used in LD mapping to control spurious associations. Correction for multiple comparisons followed Setter et al. (2010), i.e., a family-wise probability level, $\alpha < 0.10$, was chosen, and a Bonferroni-corrected threshold probability based on individual tests, β, was calculated as $\beta < 0.10/n$, where n was the number of individual trait–SNP combinations tested.

Results

Allele diversity as revealed by SNPs and SNP haplotypes

Of 1536 maize SNPs present in each custom random oligo pool assay (OPA, RA chip) and candidate gene based OPA (DT chip), 1035 (67%) and 1017 (66%) SNPs showed two distinct alleles with high quality in the 305 lines. The average PIC values for these 1035 and 1017 markers were 0.256 and 0.240, respectively, with an average of 0.248. Heterozygosity was observed with an average of 5.12% across 305 lines. Only 13.65% (280/2052) showed MAF lower than 5% in 305 lines. Approximately 52.44% (1076/2052) of the SNPs had a MAF higher than 20%, and 8.09% (166/2052) showed almost equal allele frequencies (close to 50%) for two alternative alleles.

To make better use of the information content of the 2052 SNPs, haplotypes were constructed. For the 386 haplotype loci, each contained two or more SNPs with an average of 2.6 SNPs per haplotype locus. A total of 1623 alleles were detected with 4.2 alleles on average, ranging from two to 23 alleles per locus. PICs for the 386 haplotypic loci ranged from 0.026 to 0.877 with an average of 0.433. As a significantly higher level of allele diversity was found for haplotypes than for SNP markers per se, it can be expected that haplotypes would be more powerful in genetic diversity analysis and gene mapping.

Genetic structure of the tested maize lines

The population structure based on PCA showed three genetically distinct groups, which matched well with previously reported heterotic groups identified in maize. All Chinese ILs were well clustered into groups that were consistent with their recurrent parent origins. The first two PCs, PC1 and PC2, explained 13.10 and 8.2% of the total variation. In conclusion, structure analysis separated maize lines very well into different populations and the lines closely related in pedigree usually did cluster together.

LD in the entire population

The extent of genome-wide LD was evaluated among 1772 SNPs with MAF > 0.05 using 305 maize lines. The LD pattern expressed by r^2 averaged 0.0318, ranging from 0 to 1.0. LD was significant at the $P = 0.01$ level for 33.6% of the pairwise SNPs, and 6.5% of the pairwise SNPs had a r^2 higher than 0.2. The extent of LD showed distinct differences across 10 chromosomes. Chromosome 10 had the highest percentage of pairwise SNPs showing significant LD (43.1%), with 10.5% of pairwise SNPs with a r^2 value exceeding 0.2. Chromosomes 1 and 2 had the lowest percentage of pairwise SNPs in LD (25.8 and 28.9%). The remaining seven chromosomes had 31.0–38.0% of pairwise SNPs in LD.

SNP markers associated with NDVI and plant height

Two target traits, plant height and NDVI, showed almost normal distribution among the 305 inbreds. Figure 1 provides the histograms for plant height and the NDVI measured in the morning and afternoon at a representative stage (92 DAP). A mixed linear model was used with the P and K matrices to perform association analysis between the trait (plant height and NDVI) and marker (2052 SNPs and 386 haplotype loci). Given the large number of trait–SNP combinations tested, it was necessary to use a family-wise error rate. The relatively conservative Bonferroni method of controlling the family-wise error rate was used to increase the likelihood that claimed associations are true positives; the error rate was set at 10%, a commonly used value. The SNP–trait associations which exceeded the threshold are assumed to be true associations.

Using 305 lines, single-marker-based analysis identified two SNPs with MAF ≥ 0.05 on chromosomes 4 and 8 that were significantly associated with plant height ($P \leq 4.87 \times 10^{-5}$, $-\log P \geq 4.31$) under WS condition, and no significant SNP was detected under WW condition (Table 1). Haplotype-based LD mapping showed that three haplotype loci on chromosomes 7, 8 and 10 were significantly associated with plant height ($P \leq 2.59 \times 10^{-4}$, $-\log P \geq 3.59$), with one for WW and two for WS conditions (Table 1).

Single-marker-based LD mapping identified 37 marker–trait combinations (23 QTL) involving 32 SNPs for different stages of NDVI, of which 5 and 29 SNPs (4 and 21 QTL) were associated with different stages of NDVI under WW and WS conditions, respectively (Table 2). Two significant SNPs were detected under both water regimes. The numbers of QTL identified at stages of IV, Va and Vb were 10, 16 and 6, with 10, 17 and 10 SNPs involved, respectively. No significant SNPs were found to be associated with the NDVI values measured at the first three stages or with V(a–b), the difference of NDVI values taken between morning and afternoon at the stage V that represented the degree of leaf rolling. Identification of none or few QTL at the earlier stages indicates that the genetic difference for the biomass was not well expressed at these stages. Some unique SNPs were identified for each stage: SNP PZB02194.1 on bin 2.03 only for stage IV under WS, SNP PZA01238.1 on bin 1.11 only for stage Va under WS, abph1.15 on bin 2.03 only for stage Vb under WS. However, some significant SNPs were shared by different stages of NDVI,

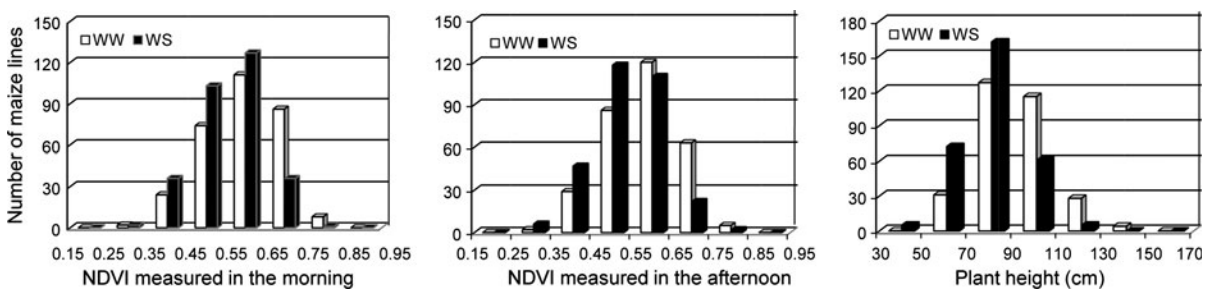

Fig. 1 Distribution of NDVI measured in the morning and afternoon at 92 days after planting and plant height under well-watered (WW) and water-stressed (WS) conditions for 305 inbred lines

Table 1 Single SNPs and haplotype loci significantly associated with plant height (PH)

Trait	Loci	−log (*P*)	R^2 (%)	Chromosome	Bin	SNP within HP loci	Position (bp)
PH	HP291	7.67	2.80	7	7.05	PHM1912.23	155970264
WW						PHM1912.20	155970323
PH	HP331	8.03	3.97	8	8.07	PZA03592.1	161817310
WS						PZA03592.3	161817408
						PZA03593.1	161819706
	HP383	3.70	2.86	10	10.07	PZA03132.3	139830278
						PZA03134.1	139832920
	Single SNP	4.51	1.51	4	4.05	PHM3112.9	32968646
	Single SNP	5.47	1.60	8	8.07	PZA03592.3	161817408

WW well-watered, *WS* water-stressed. R^2, explained phenotypic variance. *HP* haplotype

The threshold probabilities that deviate from the null distribution for multiple comparisons by $\alpha \leq 0.10$ are $P \leq 4.87 \times 10^{-5}$ ($-\log P \geq 4.31$) and $P \leq 2.59 \times 10^{-4}$ ($-\log P \geq 3.59$), for 2052 SNPs and 386 haplotype loci respectively. SNP positions were based on maize B73 genome (http://www.maizesequence.org)

with three and one significant SNPs shared by two and three stages, respectively.

A total of 12 significant haplotype loci were identified for NDVI of different stages, each corresponding to a QTL, of which six and eight haplotype loci were identified under WW and WS conditions, respectively (Table 3). Two loci were found for both water regimes, but no haplotype locus was detected for stages I, II or V(a–b) (Table 3). Haplotype locus 377 was shared by two stages and haplotype locus 200 was shared by three stages. Moreover, some specific haplotype loci were identified for only one of the three stages, IV, Va and Vb. These results suggested that the same trait was controlled by multiple gene loci and some loci showed different levels of expression across developmental stages.

Comparison of significant loci identified by single SNP- and haplotype-based LD mapping

Using haplotype-based LD mapping, three significant haplotype loci/QTL were identified for plant height. For haplotype locus 331, there was a SNP PZA03592.3 within the locus that was also significantly associated with plant height under WS condition identified by single SNP-based LD mapping, which explained relatively low phenotypic variation (1.60%). However, its corresponding haplotype locus including this SNP plus other two SNPs nearby could explain higher phenotypic variation (3.97%) (Table 1).

For the 12 haplotype loci/QTL identified in this study for different stages of NDVI, five contained at least one SNP that was also significantly associated with NDVI identified by single SNP-based LD mapping (Table S2). For example, SNP PHM3112.9 at HP156, PZA01715.2 at HP363, and PZA03570.1 at HP382 were significantly associated with the NDVI values taken at stages IV and Va under WS. Two SNPs, PHM1912.23 and PHM1912.20, at HP291were associated with the NDVI taken at stages IV and Va under WW condition. Four of the 12 SNPs at HP377 (PZA03710.2, PZA03710.3, PZA03710.4 and PZA03710. 5) were associated with the NDVI at stage Vb under WS. In conclusion, single SNP-based LD mapping provided results about the locations almost consistent with haplotype-based analysis. Some chromosomal regions defined by haplotype loci were closely linked to the specific traits where significant single SNP markers were identified. Furthermore, significant SNP haplotype loci explained much higher phenotypic variation than single SNPs.

Putative candidate genes associated with significant loci for plant height and NDVI

The putative candidate genes for plant height and NDVI were identified using associated SNP markers revealed by single-SNP and SNP haplotype-based LD mapping. The identified genes include those involved in phytochrome, transcription factor (MYB-related protein Zm1, MADS-box), SET domain, tyrosine protein kinase, aberrant phyllotaxy1 (*abph1*), aldehyde

Table 2 SNPs significantly associated with NDVIs at different stages

QTL no.	Bin	SNPs	Treatment	−log (*P*)		
				IV	Va	Vb
1	1.11	PZA01238.1	WS		8.47	
2	2.03	PZB02194.1	WS	5.55		
		abph1.15	WS			8.05
3	2.10	PZD00022.5	WS		6.32	5.96
4	3.03	PZA00508.2	WS	6.11		
5	3.04	PZB01964.5	WS		4.31	
		PZA01114.2	WS		7.24	
6	3.05	PHM15449.10	WS	7.51		
7	3.06	PHM1675.29	WS		8.39	
8[a]	4.04	PHM3112.9	WW		5.20	
			WS	6.01		
9	4.05	PZA03367.1	WS		9.07	
10	5.01	PZA02653.12	WS	8.89		
11	5.04	PZA03049.24	WS	7.67		
12	5.05	PZA03704.1	WS	7.92	5.52	6.27
		PZB02005.1	WS			7.11
13	5.06	PZB00765.1	WS		7.89	
14	6.01	PZA03389.1	WW		4.47	
15	6.06	PZA01672.1	WS		5.25	
16[a]	7.04	PHM1912.23	WW	5.23		
		PHM1912.20	WW		5.63	
17	8.06	PZA03592.3	WS	6.06		
18	8.08	PZB01057.4	WS	5.04		
		PZA03202.2	WS			6.18
19[a]	9.06	PZA01715.2	WS		10.91	
20	10.02	PZA01451.1	WS		7.65	
21	10.03	PZA03157.2	WS		6.80	4.37
22[a]	10.04	PZA00814.1	WS		7.56	
		PZA03710.2	WS			6.68
		PZA03710.4	WS			6.13
		PZA03710.5	WS			7.12
		PZA03710.3	WS			7.34
23[a]	10.06	PZA03570.1	WW/WS		10.02	

WW well-watered, *WS* water-stressed. The threshold for Bonferroni-corrected multiple comparisons is $P \leq 4.87 \times 10^{-5}$ ($-\log P \geq 4.31$) for 2052 SNPs

[a] QTL regions associated with NDVI identified by both single SNP- and haplotype-based linkage disequilibrium mapping

dehydrogenase (ALDH), histone acetyltransferase HD2 (*HDT104*), etc. (Table S2). *abph1* encodes a cytokinin-inducible type A response regulator and regulates leaf initiation and morphogenesis in the embryo, and plays a role in determining the phyllotaxy of the shoot (Lee et al. 2009). ALDH represents a group of enzymes catalyzing the conversion of aldehydes to the corresponding acids. It is reported that ALDH-null mutants in maize and *Arabidopsis* and ALDH-reduced mutants in rice (*Oryza sativa*) showed less tolerance to anaerobic conditions (Wei et al. 2009). Several significant loci were shared by plant height and NDVI, including two SNPs PHM3112.9 and PZA03592.3 and one haplotype locus HP291.

Table 3 Haplotype loci significantly associated with NDVI at different stages

HP loci	Bin	Treatment	−log (*P*)				SNPs within haplotype loci
			III	IV	Va	Vb	
HP63	1.12	WW		3.60			PZB01937.2 PZB01937.3
HP65	2.00	WS			4.68		PZA00732.5 PHM4951.8
HP156[a]	4.04	WS			4.23		PHM3112.9
		WW			4.24		PHM3112.5
HP200	5.01	WS		3.81	3.84	3.73	PZB00054.3 PZB00102.2 PZB02107.1 PZB00102.1
HP239	5.07	WW		4.19			PZA03339.1 PZA03339.2 PZA03335.2
HP269	7.01	WW				3.81	PZA03617.1 PZB02542.2
HP285	7.03	WS			9.31		PZA00655.1 PZA00111.10
HP291[a]	7.04	WW			5.83		PHM1912.23 PHM1912.20
HP350	9.03	WS		4.21			PZA03759.2 PZA03759.1
HP363[a]	9.06	WS			7.64		PZA01715.2 PZA01715.1
HP377[a]	10.04	WS	4.11			6.51	PZA03713.1 PZA03711.2 PZA03711.3 PZA03711.1 PZA03710.2 PZA03710.4 PZA03710.1 PZA03710.5 PZA03710.3 PZA03710.6 PZA03709.1 PZA03709.2
HP382[a]	10.06	WS			6.79		PZA03570.2
		WW			5.55		PZA03570.1

WW well-watered, *WS* water-stressed, *HP* haplotype. The threshold for Bonferroni-corrected multiple comparisons is $P \leq 2.59 \times 10^{-4}$ ($-\log P \geq 3.59$) for 386 haplotype loci

[a] QTL regions associated with NDVI identified by both single SNP- and haplotype-based linkage disequilibrium mapping

PHM3112.9 was associated with major facilitator superfamily transporter (MFS-1), which is a secondary carrier transporting small solutes in response to chemiosmotic ion gradients (Paulsen et al. 1998). SNP PZA03592.3 was located on histone acetyltransferase gene *HD2* (HDT104), which is involved in chromatin remodeling. Casati et al. (2006) demonstrated that HDT104 was differentially expressed before and after UV-B treatment and is crucial for UV-B acclimation in maize. The putative related genes for these significant loci and their functions are listed in Table S2.

Discussion

Comparison of QTL for plant height with previous reports

We compared our QTL mapping results for plant height with those identified in linkage mapping using different segregation populations and tested in different environments with results available through the Gramene database (http://www.gramene.org/). In this study, three haplotype loci and two SNPs were identified for plant height under two water regimes using single-marker-based and haplotype-based LD mapping. These QTL were mainly located on bins 4.05, 7.05, 8.07 and 10.07, which have been also revealed by different studies. The significant chromosomal region on bin 4.05 for plant height was also identified by Yan et al. (2003) using the maize $F_{2:3}$ population derived from an elite hybrid (Zong3 × 87-1). On the nearby regions, two QTL on bins 4.04 and 4.06 for plant height under two water regimes were also detected by Messmer et al. (2009). On chromosome 7, there is one plant height QTL located on bin 7.05 (between 91.5 and 116.3 cM) identified by CIMMYT (1994) using linkage mapping with two RILs (Ki3 × CML139 and CML131 × CML67) and by Veldboom and Lee (1996) using a population of 150 $F_{2:3}$ lines produced from the cross of elite inbreds Mo17 and H99 in stress and non-stress environments. This QTL was also identified by LD mapping in our study. Moreover, two QTL located on bins 8.07 (between 105.5 and 128.6 cM) and 10.07 (between 97.7 and 114.1 cM) for

plant height identified by Koester et al. (1993) and Berke and Rocheford (1995) were also confirmed by our results.

Secondary traits for drought tolerance in maize

As one of the most important abiotic stresses that strongly affect the production of cultivated crops, drought has received great attention, particularly in genetics and breeding research. The genetic basis of the molecular, cellular and developmental responses to drought involves many gene functions regulated by water availability (Tuberosa and Salvi 2006). An ideal secondary trait that can be used for drought tolerance evaluation should be genetically correlated with grain yield in the target environment, genetically variable, and highly inheritable, be simple, cheap, non-destructive and fast to assay, be stable throughout the measurement period, but should not be associated with any yield loss under non-stressed conditions (Edmeades et al. 1997; Lafitte et al. 2003). The flowering and phenological parameters, plant height, ear number, photosynthesis, chlorophyll fluorescence and leaf abscisic acid concentration, and root traits have been commonly used as the important secondary traits for evaluation of drought tolerance in maize. A large QTL mapping effort has been conducted at CIMMYT over the last 12 years to map grain yield and its secondary traits (Ribaut et al. 2009). The QTL analysis was first conducted individually on each of the 56 trials (44 WS and 12 WW), identifying 1080 QTL covering nine traits. Besides, the biomass as measured by NDVI was introduced as a new trait for drought tolerance evaluation (Lu et al. 2011a), which has also been proved in this study to be a good index for evaluating plant drought tolerance, particularly at the vegetative stage. A total of 30 significant chromosomal regions were identified in this study for seven stages of NDVI under WW and WS conditions, and three of them were shared with the loci affecting plant height.

LD mapping based on single SNPs and SNP haplotypes

The number of alleles per locus is one of the important factors that affect the power of LD mapping. As the best marker type for high-throughput genome-wide genetic analysis, SNPs have a limited number of alleles at each locus, resulting in much lower PIC estimates compared to SSR markers (Hamblin et al. 2007; Lu et al. 2009). Based on Laval et al. (2002), ten times or more SNP markers will be needed in order to obtain the same amount of information as SSR markers, although the method by which PIC is constructed and evaluated is questionable (Lu et al. 2009). The results from this study have indicated that the biallelic problem associated with SNP markers can be overcome using haplotypes that are constructed using multiple SNP markers. SNP haplotypes can be defined to include SNPs from the same genes, a specific genome region, a chromosome, and a specific slide window or physical distance. In this study, SNP markers used to generate haplotypes were selected within a 10-kb window, an estimate of the average size for LD blocks in maize (Yan et al. 2009a; Lu et al. 2010). Using haplotypes derived from multiple SNPs provides a high level of allele diversity, thus increased genotyping resolution. The number of alleles per locus increased from two for single SNPs to up to 23 for a haplotypic locus. Accordingly, the average PIC increased from 0.248 to 0.433 in 305 maize inbred lines. Using haplotype-based LD mapping, three and 12 significant haplotype loci were identified for plant height and different stages of NDVI, respectively, with one haplotype locus shared by the two tested traits. For the six haplotype loci, each contained at least one SNP that was also significantly associated with the specific trait revealed by single SNP-based LD mapping. However, significant SNP haplotype loci explained much higher phenotypic variation than single SNPs, which increased from 1.37 to 2.89% on average. When multiple SNPs from the same genes become available, haplotype-based mapping can be used to establish the association using within-gene variation. Comparison of single SNP- and haplotype-based analyses will provide more insights about the effect of combined SNPs and haplotypic diversity on LD mapping. This study provides an example of using both single markers and their combinations–haplotypes–for improvement of LD mapping.

Breeding applications

Plant breeding through phenotypic selection under drought conditions has resulted in major progress (Bänziger and Araus 2007), but it is time-consuming and laborious. Linkage and LD mapping are the two

most commonly used association analysis tools for dissecting drought tolerance and identifying actual genes underlying QTL. As whole genome sequences become available for many crop species including maize and for multiple genotypes of the same species through resequencing, along with cost-effective high-throughput genotyping systems and the next generation of sequencing technologies, GWAS becomes practical and its use in plant breeding will allow the manipulation of many traits at the whole-genome level. LD mapping using a set of global diverse breeding germplasm and high-throughput SNP chips, as shown in this study, provides high-resolution genetic mapping, and a GWAS with SNPs covering all candidate genes can narrow down the associated chromosomal regions to specific genes. Genes that are found to be associated with complex traits such as drought tolerance are highly useful and can be exploited to improve our knowledge of the pathways and mechanisms that plants employ to deal with the stress. This knowledge in turn will be useful not only for designing marker-assisted selection strategies but also for optimizing conventional breeding systems.

Acknowledgments The SNP chips used in this study were developed through a collaborative project between Cornell University (Ed Buckler and Tim Setter) and CIMMYT (Marilyn Warburton and Jianbing Yan). Maize drought tolerance work was funded by the Rockefeller Foundation, Bill and Melinda Gates Foundation, Generation Challenge Programme, European Community, National High Technology Research and Development Program of China (No. 2009AA10AA03-2) and National Basic Research Program of China (No.2007CB108900).

References

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Bänziger M, Araus J (2007) Recent advances in breeding maize for drought and salinity stress tolerance. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer, Netherlands, pp 587–601

Berke T, Rocheford T (1995) Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. Crop Sci 35:1542–1549

Bolanos J, Edmeades GO, Martinez L (1993) Eight cycles of selection for drought tolerance in tropical maize. III. Responses in drought-adaptive physiological and morphological traits. Field Crops Res 31:269–286

Bruce WB, Edmeades GA, Barker TC (2002) Molecular and physiological approaches to maize improvement for drought tolerance. J Exp Bot 53:13–25

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Casa AM, Pressoira G, Brown PJ, Mitchell SE, Rooney WL, Tuinstrac MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Casati P, Stapleton AE, Blum JE, Walbot V (2006) Genome-wide analysis of high-altitude maize and gene knockdown stocks implicates chromatin remodeling proteins in response to UV-B. Plant J 46:613–627

CIMMYT (1994) QTL data for populations Ki3 x CML139 and CML131 x CML67. http://www.agron.missouri.edu/locus.html

CIMMYT Applied Molecular Genetics Laboratory (2003) Laboratory protocols, 3rd edn. Mexico, DF, pp 7–11

Edmeades GO, Bolaños J, Chapman SC (1997) Value of secondary traits in selecting for drought tolerance in tropical maize. In: Edmeades GO, Bänziger M, Mickelson HR, Peña-Valdivia CB (eds) Developing drought and low-N tolerant maize. CIMMYT, Mexico, pp 222–234

Edmeades GO, Bolaños J, Elings A, Ribaut JM, Bänziger M, Westgate ME (2000) The role and regulation of the anthesis-silking interval in maize. In: Westgate M, Boote K (eds) Physiology and modeling kernel set in maize. Proceedings of ASA-CSSA-SSSA Meeting, Baltimore, USA. 17–22 October 1998, CSSA Special Publication Number 29, 43–75

Edmeades GO, Bolaños J, Hernandez M, Bello S (1993) Causes for silk delay in a lowland tropical maize population. Crop Sci 33:1029–1035

Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Garcia EW, Lebruska LL, Laurent M, Shen R, Barker D (2006) Illumina universal bead arrays. Methods Enzymol 410:57–73

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 12:e1367

Hao Z, Liu X, Li X, Xie C, Li M, Zhang D, Zhang S, Xu Y (2009) Identification of quantitative trait loci for drought tolerance at seedling stage by screening a large number of introgression lines in maize. Plant Breed 128:337–341

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Koester RP, Sisco PH, Stuber CW (1993) Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines of maize. Crop Sci 33:1209–1216

Lafitte R, Blum A, Atlin G (2003) Using secondary traits to help identify drought-tolerant genotypes. In: Fischer KS, Lafitte

R, Fukai S, Atlin G, Hardy B (eds) Breeding rice for drought-prone environments. IRRI, Manila

Laval G, San Cristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genet Sel Evol 34:481–507

Lee BH, Johnston R, Yang Y, Gallavotti A, Kojima M, Travencolo BAN, Costa LF, Sakakibara H, Jackson D (2009) Studies of aberrant *phyllotaxy1* mutants of maize indicate complex interactions between auxin and cytokinin signaling in the shoot apical meristem. Plant Physiol 150:205–216

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Lu Y, Yan J, Guimarães CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Mugo S, Makumbi D, Shah T, Rong T, Crouch JH, Xu Y (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120: 93–115

Lu Y, Zhang SH, Shah T, Xie C, Hao Z, Li X, Farkhari M, Ribaut JM, Cao M, Rong T, Xu Y (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Lu Y, Hao Z, Xie C, Crossa J, Araus JL, Gao S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Taba S, Pan G, Li X, Rong T, Zhang S, Xu Y (2011a) Comparison of multiple selection criteria for drought resistance through large-scale screening under water-stressed and well-watered environments in maize. Field Crops Res 124:37–45

Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Prakash AB, Rong T, Xu Y (2011b) Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS ONE 6(9):e24861

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879–889

Messmer R, Fracheboud Y, Bänziger M, Vargas M, Stamp P, Ribaut JM (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theor Appl Genet 119:913–930

Paulsen IT, Pao SS, Saier MH Jr (1998) Major facilitator superfamily. Microbiol Mol Biol Rev 62:1–34

Ribaut JM, Betran J, Monneveux P, Setter T (2009) Drought tolerance in maize. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, New York, pp 324–325

Setter TL, Yan J, Warburton M, Ribaut JM, Xu Y, Sawkins M, Buckler ES, Zhang Z, Gore MA (2010) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J Exp Bot 62:701–716

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

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

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trend Plant Sci 11:405–412

Veldboom LR, Lee M (1996) Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: II. Plant height and flowering. Crop Sci 36:1320–1327

Wei Y, Lin M, Oliver DJ, Schnable PS (2009) The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of *Arabidopsis*. BMC Biochem 10:7

Wen W, Mei H, Feng F, Yu S, Huang Z, Wu J, Chen L, Xu X, Luo L (2009) Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). Theor Appl Genet 119: 459–470

Wolfe DW, Henderson DW, Hsiao TC, Alvino A (1988) Interactive water and nitrogen effects on senescence of maize. I. Leaf area duration nitrogen distribution and yield. Agron J 80:859–864

Yan J, Tang H, Huang YQ, Shi YG, Zheng YL, Li JS (2003) Genomic analysis of plant height in maize through molecular marker. Agric Sci Chin 2(10):1069–1075

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009a) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Yan WG, Li Y, Agrama HA, Luo D, Gao F, Lu X, Ren G (2009b) Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). Mol Breed 24: 277–292

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

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