Association Analysis of the *nced* and *rab28* Genes with Phenotypic Traits Under Water Stress in Maize

Zhijun Su · Xinhai Li · Zhuanfang Hao · Chuanxiao Xie · Mingshun Li · Jianfeng Weng · Degui Zhang · Xiaoling Liang · Zhigang Wang · Julin Gao · Shihuang Zhang

Abstract As a very complex quantitative trait, drought tolerance has always been suspended with questions at the molecular level. Abscisic acid (ABA) is the main drought-induced hormone that regulates the expression of many genes related to drought tolerance. 9-cis-epoxycarotenoid dioxygenase (NCED) and ABA-responsive gene protein 28 (RAB28) are key enzymes in ABA biosynthesis and regulating drought tolerance induced by ABA under water stress, respectively. In this study, a total of 22 phenotypic traits including morphological traits, grain yield, and its components were investigated under water stress among 196 maize inbred lines majorly collected from five growing zones of China, and they were further genetically sequenced based on *nced* and *rab28* genes and analyzed sequence polymorphism. The phenotypic analysis results showed that ten traits were highly related to variations of drought tolerance. The sequencing results showed low genetic diversity and evidence of neutral selection in both genes and high linkage disequilibrium level among pairwise polymorphic sites. Nucleotide diversity was significantly lower in the coding region than in non-coding regions. By mixed linear model, 13 and 11 polymorphisms of the *nced* and *rab28* genes, respectively, were associating with phenotypic traits under water stress in two different environments. These results will be useful for further functional molecular breeding in improvement of drought tolerance.

Keywords Maize (*Zea mays* L.) · Association analysis · Drought tolerance · *nced* and *rab28* genes

Introduction

Maize is one of the important crops for its food, feed, and industry use. Maize has now become the most important grain crop in the world (FAO statistics data). In its growth periods, maize is almost variously affected by biotic and abiotic stress. Among these, drought stress is the major abiotic factor limiting grain yield and seed production in maize. Yield losses resulting from drought stress annually exceed the sum of other abiotic stresses. In developing countries, drought is especially severe because of a lack of irrigation facilities so that rainfall is the main source of water available to the crop growing (Banziger et al. 1999). Drought tolerance is a complex quantitative trait which is influenced by numerous genes (Wang et al. 2010) and regulated by several physiological and molecular pathways (Champoux et al. 1995; Lilley et al. 1996; Li et al. 2009). Selection of drought-tolerant varieties allows us to mitigate the yield loss caused by water stress and to reduce agricultural use of fresh water resources. The maize genome is rich in allelic variants and discovering these alleles from genes related to drought tolerance would be useful for maize breeding, as well as understanding the complex

Z. Su, X. Li, and Z. Hao contributed equally to this paper.

Z. Su · X. Li · Z. Hao · C. Xie · M. Li · J. Weng · D. Zhang · S. Zhang (✉)
Institute of Crop Science,
Chinese Academy of Agricultural Sciences,
National Key Facilities for Crop Genetic Resources and Improvement,
No. 12 Zhongguancun, South Street, Haidian District,
Beijing 100081, China
e-mail: cshzhang2000@yahoo.com.cn

X. Liang
Institute of Food Crops Science,
Xinjiang Academy of Agricultural Sciences,
Urumqi, Xinjiang 830000, China

Z. Su · Z. Wang · J. Gao
Agronomy College, Inner Mongolia Agricultural University,
Hohhot, Inner Mongolia 010019, China

genetic and molecular mechanism underlying drought tolerance. Drought tolerance research focuses on morphological, physiological, and molecular features that influence the adaptive response of plants to drought stress (Bolanos and Edmeades 1993; Tuberosa and Salvi 2006; Messmer et al. 2009; Liu et al. 2009). Studies on maize drought tolerance through linkage mapping have made major progress for agriculture. Association analysis based on linkage disequilibrium (LD) is an effective method to identify functional allelic variation related to drought-tolerant traits.

As a well-known growth inhibitor, abscisic acid (ABA) modulates many key growth and physiological processes in plants, such as seed germination suppression, seed dormancy maintenance by inhibiting cell growth, stomatal closure induction to minimize transpiration for reducing water loss, and acceleration of abscission and senescence (Finkelstein et al. 2002; Fujita et al. 2005; Meng et al. 2009; Gao et al. 2010). Under water-limited conditions, there is evidence that ABA production is enhanced and it can effectively protect plants against drought stress (Shinozaki and Yamaguch-Shinozaki 2000; Xiong et al. 2002). ABA is mainly synthesized in an indirect way in which 9-cis-epoxycartenoid dioxygenase (NCED) encoded by the *nced* gene is the key regulatory enzyme of ABA biosynthesis in higher plants (Seo and Koshiba 2002). In maize, *Vp14* (the same name as *nced*) was firstly cloned from the ABA-deficient mutant (Schwartz et al. 1997; Tan et al. 1997). The *nced* expression is induced in response to water deficit (Nambara and Marion-Poll 2005). Additionally, the maize *rab28* gene (ABA-responsive gene), which encodes one of the late embryogenesis abundant proteins and is regulated by two bZIP transcription factors (Nieva et al. 2005) induced by ABA and water stress has been identified (Pla et al. 1991). The promoter region of the *rab28* gene contains a cis-acting element CCACGTGG (ABRE, involved in ABA dependent pathway), which is induced by ABA induction (Pla et al. 1993). Therefore, the *nced* and *rab28* genes are interesting candidate genes for further study of association analysis between nucleotide polymorphisms with phenotypic traits under water stress.

Based on the LD, association analysis can test the relationship between specific nucleotide sequence polymorphisms in candidate genes and phenotypic variation (Thornsberry et al. 2001). In maize, LD decays very rapidly (Remington et al. 2001; Tenaillon et al. 2001) so association mapping within gene resolution has a much greater precision than quantitative trait loci (QTL) mapping and provides an independent approach to test candidate genes identified in standard QTL experiments. At the same time, specific biological effects for significant polymorphisms can be hypothesized when association analysis links specific polymorphisms to trait variation. Although the spurious association of polymorphisms with traits is a potential obstacle due to relatedness rather than sequence variation function in association studies, there are now two ways to decrease the effect of false positive association. One is the structure population analysis which can be effectively estimated with the Structure 2.0 program using several models for linked and unlinked markers; another is the kinship analysis which includes the level of pairwise relatedness coefficients by SPAGeDi 1.2 software (Hardy and Vekemans 2002). Recently, Yu et al. (2005) developed a mixed linear model (MLM) which combines population structure information and kinship in the association analysis.

The objectives of this study were to investigate the natural variation in a collection of maize inbred lines and drought-tolerant candidate genes in order to identify functional alleles and haplotypes in the coding sequence of the *nced* and *rab28* genes (including 5′-UTR, 2 exons, one intron). Association analysis was performed with the MLM program to examine the correlation between sequence polymorphisms in candidate genes with phenotypic traits related to drought tolerance.

Materials and Methods

Plant Materials

A total of 196 maize inbreds (*Zea mays* L.) were used in this study mostly collected from five major maize-growing regions in the corn belt in China while others were from the USA. They represented most of the genetic diversity available to breeding and research programs in China. The seeds of the maize inbred lines were acquired from the Institute of Crop Science, Chinese Academy of Agricultural Science, China; detailed information on these materials is reported in the paper of Xie et al. (2008). These inbred lines were divided into six subpopulations based on 70 simple sequence repeat (SSR) marker data described by Xie et al. (2008): BSSS (American BSSS including Reid), PA (group A germplasm derived from a modern U.S. hybrid in China), PB (group B germplasm derived from a modern US hybrid in China), Lan (Lancaster Surecrop), LRC (derivative lines from Lvda Reb Cob, a Chinese landrace), and SPT (derivative lines from Sipingtou, a Chinese landrace).

The inbred lines were grown in one-row plots with an alpha (0,1) lattice design (Barreto et al. 1997) at two different field sites of Sanya, HaiNan province (winter, 2007, abbr 2007HN), and Urumqi, XinJiang province (summer, 2008, abbr. 2008XJ) in China, All materials in a random order were evaluated under two water regimes (well-watered, WW and water-stressed, WS), each with two replications. Plants were open-pollinated. The morphological and physiological traits were investigated in flowering periods of each line as a unit under water-stressed and

well-watered situations, and then grain yield and yield component traits were acquired 20 drying days after harvesting. The following 22 traits were studied: grain weight per ear (GraW), single ear weight, grain yield (GY), ear weight, hundred kernal weight (HKW), total plant number (PlaN), total ear number (EarN), ears per plant (EPP), row number per ear (RowN), kernal number per row (KerN), ear length (EarL), ear diameter (EarD), bald tip length (TipL), plant height (PlaH), ear height (EarH), tassel length (TasL), days to anthesis (DtoA), days to silking (DtoS), anthesis-silking interval (ASI), stay green (StayG), leaf curling (LeafC), and water content of kernel (WatC). Of these, only GraW, SEarW, EPP, and ASI were indirectly calculated; all other traits were directly investigated from a minimum of five plants in each line.

With the average value of at least five plants to each line, phenotypic traits variance effects of genotype, location, replicate within location, and genotype–location interaction were evaluated through the general linear model (GLM) program in SAS (SAS 1989). Since many traits were highly correlated, we performed principal component analysis on the correlation matrix from adjusted means (FACTOR program in SAS). The mean value of two replicates for phenotypic trait in WW and WS regimes of each environment was used for association mapping.

Population Structure and Individual Kinship

Population structure is a major bias factor leading to false-positive associations (Flint-Garcia et al. 2003). To alleviate the effect of population structure, all inbred lines were genotyped with 70 genome-wide SSR molecular markers, and six groups were defined (Xie et al. 2008) using Structure version 2 software (Pritchard et al. 2000a). These independent group memberships were used as covariates in the genotype–phenotype association analyses. Based on recent studies (Yu et al. 2005), kinship coefficient matrices of 196 inbreds were calculated using the SPAGeDi version 1.2 software.

DNA Isolation, PCR Amplification, and DNA Sequencing

Genomic DNA samples were extracted from maize leaves using the CTAB method (Saghai-Maroof et al. 1984). Then, two PCR products were amplified using Taq DNA polymerase and primers based on the sequence of the maize *vp14* mRNA (Genebank accession number, U95953) and the sequence of the maize *rab28* gene (Genebank accession number, X59138), respectively. The combinations of forward- and reverse-primers were: (1) 5′-TCCAATTCCGT CAGGTTCTC-3′ and 5′-TGATGAAGGTGCCGTGGAA-3′ for *nced* gene, (2) 5′-ACCATCAAGTCTCAAGAGGC-3′ and 5′-ACATGGACGTAGGTAAAGCG-3′ for *rab28* gene,

respectively. Finally, the high-quality sequences of 162 inbred lines in the *nced* gene (with a length of 223–1,736 bp, the first nucleotide of U95953 marked +1 bp) and the high-quality sequences of 131 inbred lines in *rab28* gene (with a length of 433–1,501 bp, the first nucleotide of X59138 marked +1 bp) were acquired and aligned.

PCR conditions were as follows: 200 ng of genomic DNA, 0.2 mM of each dNTP, 1 μM of each primer, 2× GC buffer (2 mM mg^{2+}), and 2.5 U TransTaq high fidelity DNA polymerase (Transgen Biotechnology Corporation, Beijing, China), adjusted to a final volume of 25 μL. An initial denaturation of PCR reaction was at 94°C for 3 min, followed by 35 cycles at 94°C for 45 s, primer annealing at 58–60°C for 1 min (the annealing temperature were 58°C and 60°C for *nced* and *rab28*, respectively), then 72°C for 1 min, and the final extension at 72°C for 10 min.

Most PCR products were sequenced directly by ABI 3000. The others were cloned by ligating to pMD18-T vector using *Escherichia coli* strain DH5a as a host and then sequenced (Sambrook and Russell 2001). All sequences were sequenced by the public laboratory of National Key Facility of Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agriculture Science. The DNA was sequenced twice in both directions. Nucleic acid sequences were aligned using the DNAMAN software and adjusted by manual work.

The TASSEL software (http://www.maizegenetics.net/bioinformatics/tassel.index) was used to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). Nucleotide diversity was calculated as the average number of pairwise differences among sequences (Tajima 1989) and polymorphic sites. Haplotype diversities were calculated among different gene structure regions based on their position (favoring those in exons rather than introns) of polymorphic sites, potential functional role (favoring non-synonymous rather than synonymous changes), frequency (favoring balanced allele frequencies rather than rare alleles), and complementarity (avoiding redundancy among polymorphisms and favoring those that allow characterization of the highest number of haplotypes observed among the sequenced inbred lines).

Association Analyses

The mixed linear model pattern in TASSEL software was used to identify significant nucleotide variations associated with phenotypic traits under water-stressed regime. In addition, coefficient of drought resistance (CD) from GY under water-stressed regime divided by GY under well-watered regime was also used as a trait for association analysis. Both population structure and individual kinship relatedness were considered in association studies of all phenotypic traits with candidate genes. Consequently,

association studies presented here are based on the Q+K model (mixed linear model) that is more complex but performs better for some phenotypic traits. Afterwards, the GLM pattern in the TASSEL software based on the Q model was used to do multiple testing of significant associations between the polymorphism and phenotype traits.

Results

Statistics of Phenotypic Traits Under WW and WS Conditions

ANOVA results showed that the maize inbred lines showed significant variance among all the investigated traits, which indicated that inbred lines hold diverse characteristics for genotyping under water deficit. Besides, there was no significant variation found between the WW and WS treatments for traits, including WatC, HKW, TipL, EarL, and EarD. For some traits including LeafC, WatC, PlaN, EarN, EarH, SearN, GY, and EarD, there were significant variations found between replicates, which might be caused by interaction between environment and genotype. Finally, StayG, PlaH, TasL, ASI, DtoA, DtoS, GraW, RowN, KerN, and EPP were significantly influenced by inbred lines,

treatments, and interaction between environment and inbred lines (Table 1). Therefore, phenotypic data of the above 10 traits with two replicates were averaged for factor analysis using SAS software, whereas five factors represented more than 85% phenotypic variation. This indicated that several traits would be commonly regulated by the same variation underlying drought tolerance. To acquire accurate results, the above ten traits were used for the next association analysis between sequence variance of the *nced* and *rab28* genes and phenotypic traits under water stress.

Sequence Polymorphisms and LD Analysis of *nced* and *rab28* Genes

For the coding region of the *nced* gene, we identified 20 SNPs with a frequency of more than 10% of the rare alleles among the 162 lines and no indels. Among SNP, four variants were non-synonymous resulting in changes in amino acids, while the remaining ones were synonymous. Nucleotide diversity showed that the highest allelic variation frequency occurred around 600–750 bp among sequenced region of the *nced* gene with an average *Pi* of 0.00741 (Table 2). LD was estimated between all pairs of polymorphic sites in the sequenced region of the *nced* gene. A plot of r^2 against physical distance for polymorphic pairs indicated that the LD level remained high ($r^2 > 0.2$) along

Table 1 Combined variance analysis (GLM) of 22 traits and their significance levels

Source	StayG	LeafC	WatC	PlaN	EarN	PlaH	EarH	TasL	ASI
Environment	517.13*	939.23*	9,006.74*	5,067.79*	1,124.60*	1,405.8*	2,369.8*	66.43*	6.53*
Treatment	345.07*	1,406.46*	52.11	2,125.099*	4,191.74*	2,363.27*	1,776.21*	156.35*	344.12*
Block	2.36	6.44**	15.60**	82.92*	70.71**	2.09	9.33*	0	0.15
Lines	1.87*	1.39*	11.46*	21.88*	33.31*	5.18*	10.12*	2.1*	1.65*
Environment*line	1.85*	1.13	8.51*	19.54*	18.95*	2.42*	4.5*	1.39*	1.57*

Source	DtoA	DtoS	KerN	TipL	EarL	EarD	EPP	SearW
Environment	12,679.26*	10,912.54*	232.73*	0.15	3,685,617.58*	279,275.02*	0.02	108,055,897.50*
Treatment	1,239.56*	9,309.10*	689.66*	3.12	431.86	30.54	7.89*	9,850,621.20*
Block	34.78	5.19348	0.24	0.09	11.42	59.38**	0	1,085,640.40*
Lines	148.04*	164.00*	3.27*	2.5*	1,838.16*	55.09*	0.16*	147,482.80*
Environment*line	82.99*	101.02*	2.24*	1.38*	1,617.41*	45.49*	0.11*	147,775.20*

Source	GraW	HKW	GY	EarW	RowN
Environment	1,909.17*	1,476.77*	70,947,620.37*	58,047,314.11*	0.41
Treatment	404.75*	0.18	49,188,332.24*	13,536,446.99*	63.74*
Block	2.33	0.85	901,633.87*	58,879.30	0.42
Lines	38.79*	3.46*	281,543.37*	6,829,535.23*	1.45*
Environment*line	29.17*	1.97*	221,178.24*	6,745,593.59	1.21**

Numbers in table were mean squares of each trait

*SearW* single ear weight, *EarW* ear weight

*P<0.01; **P<0.05

Table 2 Nucleotide polymorphisms of the sequenced regions of *nced* and *rab28* genes

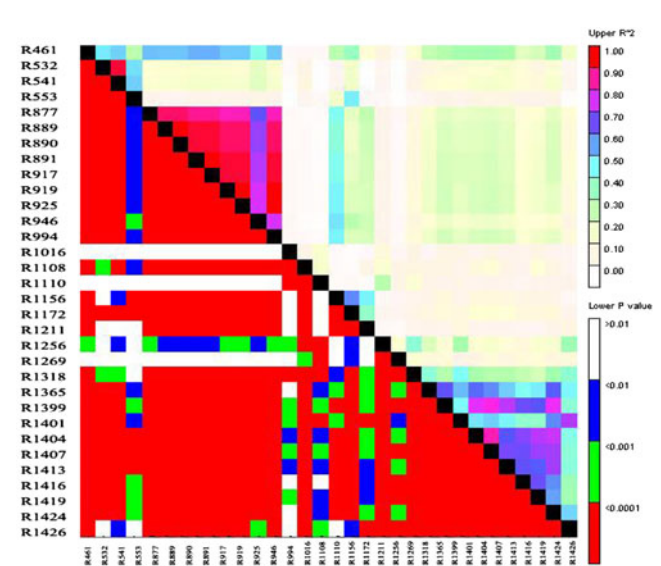| *nced* gene	Coding regions	*rab28* gene	Coding regions		Non-coding regions		Total
	Exon		Exon-1	Exon-2	5′UTR	Introns	
SNP	20	SNP	13	11	0	2	26
Indel		Indel	3	0	1	2	6
Pi (π)	0.00502	Pi (π)	0.00614	0.00781	0.00697	0.00487	0.01126
θ	0.00712	θ	0.00685	0.01033	0.01845	0.01917	0.00939
Tajima's D	−0.9493	Tajima's D	0.477	1.0823	−1.1227	−1.0419	0.3548
Fu and Li's D*	−4.4053	Fu and Li's D*	−1.8091	−0.3427	−1.1348	−1.2777	−1.7003
Fu and Li's F*	−3.4627	Fu and Li's F*	−1.0782	−0.2372	−1.3637	−1.4206	−0.9943

the entire sequenced region of the *nced* gene (Fig. 1). However, LD was not evenly distributed among these loci. The highest level of LD was identified by the detection of two recombination events: one hot spot between sites 234 and 310, and the other between sites 609 and 831 of all the alignments ($r^2 > 0.8$). Moreover, sites 420 and 672, 498 and 672, 672 and 1,554, and sites 720 and 831 showed higher LD ($r^2 > 0.5$). A relatively high degree of LD in the polymorphic sites of *nced* gene indicated its conserved evolution process of maize inbred lines.

For the *rab28* gene, we identified 26 SNPs and six indels in the sequenced region of the *rab28* gene with a frequency above 10% for the rare alleles among 131 inbred lines. Of these, one indel (five nucleotides) was found in the 5′UTR region, two SNPs, and two indels (four nucleotide deletion) were found in the intron region of the *rab28* gene, 11 SNPs were in exon-2, while 13 SNPs and three indels (two

include 18 nucleotides and one includes three nucleotides) were in exon-1. In the coding sequence of the *rab28* gene, there were 12 synonymous variances (SNP) and 12 SNPs were non-synonymous. At the same time, three indels in the coding sequence of the *rab28* gene did not result in frame shift mutation. Nucleotide diversity showed that the highest allelic variation frequency occurred around 600–950 bp among the *rab28* gene sequenced region with an average *Pi* of 0.03054. LD was estimated between all pairs of polymorphic sites in the sequenced region *rab28* gene. The same result with the *nced* gene, a plot of r^2 against physical distance for polymorphism pairs indicated that LD level remained high ($r^2 > 0.2$) for the entire length of the *rab28* gene-sequenced region. As is indicated in Fig. 2, LD is not evenly distributed along the locus. The highest level of LD was supported by the detection of two regions; one from sites 877 to 946, and the other from sites 1,399 and
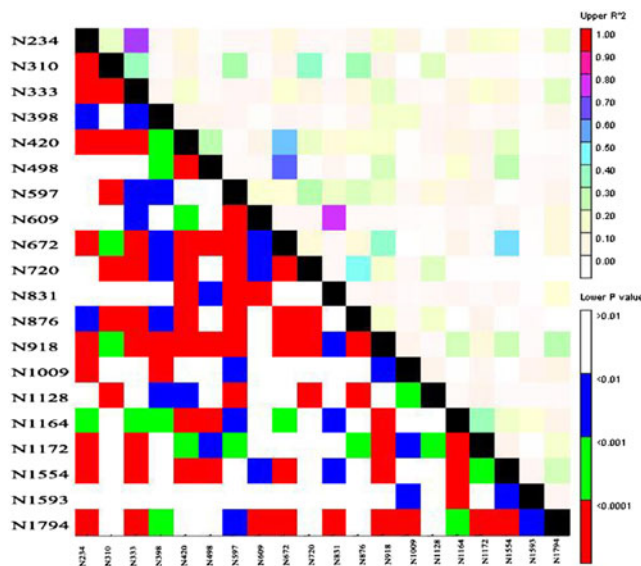
Fig. 1 LD level among pairwise polymorphisms in the alignment of *nced* gene. *Lower left triangle*, *P* values derived from Fisher's exact test. *Upper right triangle*, r^2 values. The *left and lower number* indicated the polymorphic sits of *nced* gene

Fig. 2 LD level among pairwise polymorphisms in the alignment of *rab28* gene. *Lower left triangle*, *P* values derived from Fisher's exact test. *Upper right triangle*, r^2 values. The *left and lower number* indicated the polymorphic sits of *rab28* gene

1,424 of the alignment ($r^2 > 0.5$). Between sites 532 and 541, there was higher LD ($r^2 = 1$), these two sites were thought to be significant polymorphisms. Other regions showed different levels of LD.

The Tajima'D test is a widely used test of neutrality in population genetics by measuring the allele frequency distribution of nucleotide sequence data. Tajima'D test result of both gene in this study showed no significant natural selection existed in their entire sequenced regions ($P < 0.05$). The non-coding regions of both genes gave negative values for selection, while the two exon regions of the *rab28* gene gave positive values. Fu and Li's tests showed similar results with Tajima's D test in sequenced regions of both genes at the significant level of $P < 0.05$ (Table 2). Low genetic diversity and weak evidence of selection could result in low nucleotide polymorphisms and high r^2 in LD analysis.

Phenotype–Genotype Associations of the *nced* and *rab28* Genes

MLM of association analysis that controls the effects of population structure and individual relative relationship were used to identify the association of phenotypic traits under water stress and genotypic variation data in the *nced* and *rab28* genes. All nucleotide polymorphisms, including SNPs and indels, with a frequency of more than 10% of the rare alleles were considered in the association analysis of phenotype–genotype in both genes.

By MLM analysis and multiple testing, seven polymorphisms in the *nced* gene-sequenced region (positions 498; 672; 876; 918; 1,009; and 1,554) were significantly associated ($P < 0.05$) with traits KerN, RowN, EPP, DtoS, and CD. There were no associations identified for GraW, TasL, PlaH, DtoA, ASI, and StayG in the 2007HN experiment. In the 2008XJ experiment, eight polymorphisms in the *nced* gene sequenced region (positions 310; 420; 597; 720; 918; 1,128; 1,164; and 1,554) were significantly associated ($P < 0.05$) with GraW, KerN, RowN, EPP, DtoS, DtoA, ASI, TasL, and PlaH. No associations were identified for StayG and CD. It indicated that grain yield components such as KerN, RowN, and morphological traits such as PlaH, DtoA, and ASI were highly associated with the *nced* gene under water stress. Based on the associated polymorphisms in 2007HN and 2008XJ, best allele at each site was identified for *nced* gene in all sequenced maize inbred lines, which was thought to be the superior haplotype of *nced* gene related to drought tolerance (Table 3).

By MLM analysis and multiple testing, four polymorphisms in the *rab28* gene sequenced region (positions 532; 541; 889; and 1,269) were significantly associated ($P < 0.05$) with RowN and PlaH. There were no associations identified for GraW, EPP, KerN, TasL, DtoA, DtoS, ASI, StayG, and CD in the 2007HN experiment. In the 2008XJ experiment, seven polymorphisms in the *rab28* gene-sequenced region (positions 553; 1,108; 1,172; 1,404; 1,407; 1,416; and 1,424) were significantly associated ($P < 0.05$) with EPP, RowN, and StayG. There were no associations identified for GraW,

Table 3 Candidate-gene association analysis of phenotypic traits related to drought tolerance in both environments based on polymorphic sites in the *nced* sequenced coding region of 162 maize inbred lines

Traits	Sites	234	310	420	**498**	597	672	720	876	918	**1,009**	1,128	**1,164**	1,554
Superior haplotype		G	T	G	C	C	T	C	G	T	T	T	C	C
2007HN	KerN				5.28		6.13			5.99				3.75
	RowN						3.54			3.83				2.95
	EPP							3.35						
	DtoS										2.48			
	CD	2.47					2.96							
2008XJ	GraW		3.71			4.55		8.12						
	KerN		3.29	2.54		3.30		6.96					2.88	
	RowN		5.58	3.70		5.51		7.65						
	EPP							4.70						
	DtoS		2.59			3.91				2.96				
	DtoA					3.76				2.94				
	ASI							4.07						
	TasL											2.81		
	PlaH							4.07					2.79	3.53

The data shows the phenotypic contribution (%) of these polymorphic sites toward the variation for corresponding traits. Superior haplotype related to drought tolerance was showed based on the 13 polymorphisms of *nced* gene. The non-synonymous polymorphisms were marked in bold in the table

KerN, TasL, DtoA, DtoS, ASI, PlaH, and CD. It indicated that grain yield component RowN was highly associated with the *rab28* gene under water-stressed conditions. Based on the associated polymorphisms, best allele at each site were identified for the *rab28* gene in all sequenced maize inbred lines, which was thought to be the superior haplotype of the *rab28* gene related to drought tolerance (Table 4).

Discussion

Nucleotide Diversities of the *nced* and *rab28* Genes Among Chinese Germplasms

In the study, there were 20 SNPs identified in the sequenced region of *nced* gene, four of which were non-synonymous and the others were synonymous. However, unlike the *nced* gene, there were 15 synonymous polymorphisms in the sequenced region of the *rab28* gene, and the other 12 nucleotide variances were non-synonymous (including SNPs and indels). Compared to other functional genes in maize previously reported, nucleotide diversity of the sequenced regions of the *nced* and *rab28* genes were slightly higher. In maize, SNP has been detected even more frequently because of its characteristic of cross pollinator. It has been reported that there is one SNP approximately every 48 bp (Tenaillon et al. 2001) and every 130 bp (Rafalski 2002) in un-translated and coding regions, respectively. In this study, we detected on average one SNP every 80 bp in the *nced* gene in the coding region and one SNP every 41 bp and one indel every 178 bp in the *rab28* gene. This result was quite similar or a little higher than the average level previously reported for other maize genes.

Ching et al. (2002) and Flint-Garcia et al. (2003) found nucleotide diversity decreased while LD increased at a given locus because of population bottlenecks and selection. In the study, there was no distinct positive evidence of

selection for *nced* and *rab28* genes through Tajima's (1989) and Fu and Li's (1993) neutrality tests (Table 2). Both genes had similar results, which were in agreement with results previously reported for maize functional genes. For example, Tenaillon et al. (2001) found that there was a rare departure from neutrality in random maize loci. Our research proved that negative or weak selection occurred for both genes.

Remington et al. (2001) and Tenaillon et al. (2001) reported a rapid breakdown of LD ($r^2 < 0.1$ within few hundred base pairs) for several loci in diverse sets of maize germplasm. Our research obtained the same result where there was a rapid decay of LD for more than 750 bp in the *nced* gene and for more than 700 bp in the *rab28* gene. It has been shown that LD decay level varies in different materials in maize. Ching et al. (2002) and Jung et al. (2004) reported extended LD might be up to hundreds of kilobase in sets of inbred lines. In contrast, the *rab28* gene had a more rapid decay than the *nced* gene.

Association Analysis of the *nced* and *rab28* Genes with Drought Tolerance

Association analysis based on linkage disequilibrium (LD) has recently emerged as an alternative approach to mapping QTL and genes associated with some comprehensive traits. Compared with the QTL mapping, association analysis identified the allele variations of gene sequence with no need for construction of mapping populations. In this study, 13 significant polymorphisms in the sequenced region of *nced* gene and 11 significant polymorphisms in the sequenced region of *rab28* gene were identified associated with drought tolerance through the MLM procedure and multiple testing in TASSEL software. Although the association analysis method is an effective way to find true polymorphisms with phenotypic traits in candidate gene, population structure is one of the major reasons that lead to

Table 4 Candidate-gene association analysis of phenotypic traits related to drought tolerance in both environments based on polymorphic sites in the *rab28* sequenced coding and non-coding regions of 131 maize inbred lines

Traits	Sites	**532**	541	553	889	**1,108**	1,172	1,269	**1,404**	**1,407**	**1,416**	**1,424**
Gene structure		Exon-1				Intron			Exon-2			
Superior haplotype	G	0	A	C	A	4	T	C	C	C	G	
2007HN	RowN	4.33	3.95		3.15							
	PlaH							4.23				
2008XJ	RowN								3.19	3.55	4.00	4.94
	EPP			5.59			4.81					
	StayG					3.82						

The data shows the phenotypic contribution (%) of these polymorphic sites toward the variation for corresponding traits. Superior haplotype related to drought tolerance was showed based on the 13 polymorphisms of *rab28* gene. The non-synonymous polymorphisms were marked in bold in the table

a false positive association (Knowler et al. 1988; Pritchard et al. 2000a, b). For example, population structure among inbred lines may cause candidate genes to lose their significance because of non-functional polymorphism and the related functional polymorphism coinciding with population structure. The latter can lead to a false negative result. In this study, the MLM procedure based on Q+K model decreased the chance of a positive association and then the GLM procedure based on Q model tested the significance of detected associations. The highly significant associations with lower P value were always existed. However, there were two associations with traits StayG (P= 0.044) and GraW (P=0.049) in 2008XJ for *nced* gene, two associations with trait RowN (P=0.048 and 0.036) in 2008XJ for *rab28* gene, which were not detected through multiple test and were deleted in the final result. It indicated that the multiple testing could adjust the accuracy of association analysis or the P threshold value we had chosen could be lower. In this study, two environments of 2007HN and 2008XJ located different latitudes of 18°14′N and 43°54′N, which suitable planting maize cultivars were greatly different in both locations. It caused the field phenotyping result highly varied, especially for GraW. So, the complex interaction of genotype by environment should be considered and studied in the future program.

Functional markers were frequently used for molecular-assisted breeding in recent years, in particular that association analysis accelerated their deployment for breeding purposes. Generally, they are derived from causative polymorphisms. For this research, the non-synonymous SNPs in the *nced* gene-sequenced region and the *rab28* gene significant polymorphisms could be firstly considered candidate variation sites related to drought tolerance. Allelic effects of these polymorphisms could be further elucidated in a gene expression and enzyme activity study. Consequently, further research should focus on the large number of genes involved in the drought tolerance network combining several genes acting on metabolism pathways related to drought tolerance. Additionally, by using TILLING technique, a series of drought tolerance gene mutants could be produced for comparisons of single polymorphisms, especially in complete LD with isogenic backgrounds.

References

Banziger M, Edmeades GO, Lafitte HR (1999) Selection for drought tolerance increases maize yields across a range of nitrogen levels. Crop Sci 39:1035–1040

Barreto H, Edmeades GO, Chapman SC, Crossa J (1997) PC-based system for using alpha lattice designs in plant breeding and agronomy. In: Edmeades GO, Bänziger M, Mickelson HR, Peña-Valdivia CB (eds) Developing drought and low N-tolerant maize. Proceedings of a Symposium, 25–29 March 1996, CIMMYT, El Batán, Mexico. Mexico D.F., CIMMYT

Bolanos J, Edmeades GO (1993) Eight cycles of selection for drought tolerance in lowland tropical maize. II. Responses in reproductive behavior. Field Crop Res 31:253–268

Champoux MC, Wang G, Sarkarung S, Mackill DJ, O'Toole JC, Huang N, McCouch SR (1995) Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor Appl Genet 90:969–981

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell 14(Suppl):S15–S45

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K (2005) AREB1 Is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. Plant Cell 17:3470–3488

Gao Y, Zhao Y, Li T, Ren C, Liu Y, Wang M (2010) Cloning and characterization of a G protein *β* subunit gene responsive to plant hormones and abiotic stresses in *brassica napus*. Plant Mol Biol Rep 28:450–459

Hardy OJ, Vekemans X (2002) Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theor Appl Genet 109:681–689

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) Gm3;5:13:14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43:520–526

Li F, Fu F, Sha L, He L, Li W (2009) Differential expression of serine/threonine protein phosphatase ype-2C under drought stress in maize. Plant Mol Biol Rep 27:29–37

Lilley JM, Ludlow MM, McCouch SR, O'Toole JC (1996) Locating QTL for osmotic adjustment and dehydration tolerance in rice. J Exp Bot 47:1427–1436

Liu Y, Li H, Shi Y, Song Y, Wang T, Li Y (2009) A maize early responsive to dehydration gene, *ZmERD4*, provides enhanced drought and salt tolerance in *arabidopsis*. Plant Mol Biol Rep 27:542–548

Meng C, Zhang T, Guo W (2009) Molecular cloning and characterization of a novel *Gossypium hirsutum* L. *bHLH* gene in response to ABA and drought stresses. Plant Mol Biol Rep 27:381–387

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

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56:165–185

Nieva C, Busk P, Domínguez-Puigjaner E, Lumbreras V, Testillano P, Risueño MC, Pagès M (2005) Isolation and functional characterisation of two new bZIP maize regulators of the ABA responsive gene rab28. Plant Mol Biol 58:899–914

Pla M, Gomez J, Goday A, Pages M (1991) Regulation of the abscisic acid-responsive gene rab28 in maize viviparous mutants. Mol Gen Genet 230:394–400

Pla M, Vilardell J, Guiltinan MJ, Marcotte WR, Niogret MF, Quatrano RS, Pages M (1993) The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. Plant Mol Biol 21:259–266

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. New York, Cold Spring Harbor Laboratory

Schwartz SH, Tan BC, Gage DA, Zeevaart JA, McCarty DR (1997) Specific oxidative cleavage of carotenoids by VP14 of maize. Science 276:1872–1874

Seo M, Koshiba T (2002) Complex regulation of ABA biosynthesis in plants. Trends Plant Sci 7:41–48

Shinozaki K, Yamaguch-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217–223

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tan BC, Schwartz SH, Zeevaart JA, McCarty DR (1997) Genetic control of abscisic acid biosynthesis in maize. Proc Natl Acad Sci USA 94:12235–12240

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci USA 98:9161–9166

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

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

Wang S, Liang D, Shi S, Ma F, Shu H, Wang R (2010) Isolation and characterization of a novel drought responsive gene encoding a glycine-rich RNA-binding protein in malus prunifolia (Willd.) Borkh. Plant Mol Biol Rep. doi:10.1007/s11105-010-0221-1

Xie C, Warburton M, Li M, Li X, Xiao M, Hao Z, Zhao Q, Zhang S (2008) An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. Mol Breeding 21:407–418

Xiong L, Lee H, Ishitani M, Zhu JK (2002) Regulation of osmotic stress-responsive gene expression by the LOS6/ABA1 locus in Arabidopsis. J Biol Chem 277:8588–8596

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208