

# Association Analysis of the Maize Gene *ZmYS1* with Kernel Mineral Concentrations

Zefeng Yang · Sijia Ma · Yunyun Hu · Enying Zhang · Zhengwen Xie · Shuhui Xu · Li Liu · Lele Deng · Chenwu Xu · Jinling Huang

Published online: 28 November 2014  
© Springer Science+Business Media New York 2014

**Abstract** Iron (Fe) is an essential micronutrient for humans, and iron deficiency is the most common micronutrient deficiency worldwide. Research on the genetic basis of iron concentration in maize kernels will provide guidance for the development of iron-rich crops, a major breeding goal to address iron deficiency. The maize *Yellow Stripe 1* (*ZmYS1*) gene encodes a specific transporter that takes up Fe(III)-phytosiderophore complexes into roots. Here, we re-sequenced *ZmYS1* gene in 88 elite maize inbred lines and detected the association between the nucleotide polymorphisms and micronutrient concentrations in maize kernels. Our analyses detected a total of 71 sequence variants in *ZmYS1*, including 61 single nucleotide polymorphisms (SNPs) and 10 insertions and deletions (indels), from the tested population. Fourteen haplotypes, which encode 13 different *ZmYS1* proteins, are classified based on the polymorphism in the coding region. Numerous polymorphic sites in maize *ZmYS1* locus were found in linkage disequilibrium (LD) that decay with increasing physical distance, and at least

14 recombination events have contributed to the LD and haplotype diversity of this gene. Additionally, our data show that a non-synonymous site in the *ZmYS1* gene is associated with the maize kernel Fe concentration, and two other non-synonymous sites with Zinc (Zn) concentration. These findings suggest that the polymorphism in maize *ZmYS1* locus may be used for biofortifying kernel mineral concentrations in maize breeding programs.

**Keywords** Maize · *ZmYS1* gene · Nucleotide polymorphism · Kernel iron and zinc concentrations · Association analysis

## Introduction

Micronutrient malnutrition (MNM) is not only common in industrialized nations but also serious in developing regions of the world. It is estimated that nearly one third of the world's population are affected by deficiencies of one or several micronutrients (Mayer et al. 2008). Iron (Fe) is a micronutrient essential to cells in the human body, especially red blood cells, which transport oxygen from the blood to tissues in the body. Iron deficiency, one of the most common forms of MNM worldwide, may cause anemia (Lucca et al. 2002; Tako et al. 2013). Zinc (Zn), another mineral micronutrient, is essential for over 300 enzymes and nearly 2000 transcription factors in the human body. Zinc deficiency will impact multiple body functions and result in a wide variety of symptoms (Prasad 2012). Biofortification, the process of breeding food crops rich in bioavailable micronutrients, is the most effective and economical way to increase the micronutrient quality of food (Bouis et al. 2011; Lung'aho et al. 2011). Thus, development of micronutrient-enriched staple foods is an important goal of crop breeding (Chakraborti et al. 2011).

Maize (*Zea mays* L.), one of the oldest human-domesticated plants, is among the most important cereal crops

**Electronic supplementary material** The online version of this article (doi:10.1007/s11105-014-0836-8) contains supplementary material, which is available to authorized users.

Z. Yang · S. Ma · Y. Hu · E. Zhang · Z. Xie · S. Xu · L. Liu · L. Deng · C. Xu (✉)

Jiangsu Key Laboratory of Crop Genetics and Physiology/  
Co-Innovation Center for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou 225009, China  
e-mail: qtls@yzu.edu.cn

Z. Yang · J. Huang (✉)  
Department of Biology, East Carolina University, Greenville,  
NC 27858, USA  
e-mail: huangj@ecu.edu

E. Zhang  
College of Agronomy and Plant Protection, Qingdao Agricultural University, Qingdao 266109, Shandong, China

in the world. In addition to its use in feed and industry, maize provides about one third of food calories to more than 4.5 billion people (Xue et al. 2014). Maize is also a major supplementary source of iron, magnesium, zinc, and other minerals for humans in many nations. Benefiting from genetic improvement of plant architecture and better management, the average maize yield has increased steadily over the years. However, to meet the demand of human diet, significant improvement is still needed in corn quality, especially micronutrient concentrations in kernels. Unlike the improvement of plant architecture, the genetic improvement of kernel nutrition traits is dependent much more on molecular marker-assisted selection (Prasanna et al. 2010; Yang et al. 2014a; Choudhary and Watson 2013). Therefore, identifying functional molecular markers for kernel nutrition traits has become increasingly important for maize breeding programs (Flint-Garcia et al. 2005; Yan et al. 2011).

Recently, much progress has been made in understanding primary Fe uptake from the soil (Waters et al. 2006; Curie and Briat 2003). The process of iron acquisition by gramineous plants includes four steps, namely biosynthesis of phytosiderophores (mugineic acids) in roots, secretion of phytosiderophores to the rhizosphere, solubilization of insoluble iron in soils by chelation of phytosiderophores, and uptake of the ferric-phytosiderophore complex by roots (Ueno et al. 2009; Ma 2005). The maize *Yellow Stripe 1* (*ZmYS1*) gene encodes a specific transporter that takes up ferric-phytosiderophore complexes into root (Yen et al. 2001; Le Jean et al. 2005; Curie et al. 2001). Expression of *ZmYS1* in heterologous systems has shown that *ZmYS1* functions as a proton-coupled symporter for phytosiderophore-chelated metals (Roberts et al. 2004). In addition to root, *ZmYS1* is also expressed in blades and sheaths of leaves as well as crowns of plants, suggesting that *ZmYS1* is involved in both primary iron acquisition and intracellular transport of Fe and other metals (Ueno et al. 2009). Moreover, members of the yellow stripe 1-like (YSL) family that transport Fe(III)-phytosiderophores were further identified in barley and rice (Araki et al. 2011; Murata et al. 2006; Lee et al. 2009; Inoue et al. 2009; Kakei et al. 2012).

Although the role of *ZmYS1* in Fe(III)-phytosiderophore uptake has been illustrated (Curie et al. 2001; Yen et al. 2001), the sequence polymorphism of this gene in natural populations has not been investigated. It is also unclear whether and how *ZmYS1* sequence variants are associated with changes in mineral concentrations in maize kernels. In the present study, we aimed to analyze the sequence variability of *ZmYS1* in natural populations and to test the association of identified sequence variants with mineral concentrations, including copper (Cu), Fe, Zn, calcium (Ca), potassium (K), magnesium (Mg), and phosphorus (P). The results of this work will lay the foundation of molecular marker-associated selection for the biofortification of Fe and Zn in maize kernels.

## Materials and Methods

### Plant Materials, Field Experiments, and Detection of Mineral Concentrations in Maize Kernels

A total of 88 elite maize inbred lines were used in this study. These inbred lines covered temperate germplasm from five heterotic groups, as well as tropic and waxy germplasm. They represented much of the genetic diversity available to breeding and research programs in China. In addition, some germplasm introduced from other countries, including the USA, Canada, France, and Germany, were also included in this study (Supplementary Table 1).

The inbred lines were grown in two-row plots with a randomized block design of two repetitions in a natural environment during 2013 in Sanya, Hainan province. All lines were self-pollinated, and ears were air-dried before manual shelling. Kernels from the middle of at least three ears in each replicate were harvested and further used for phenotypic analyses.

Maize kernels of each inbred line were oven dried at 70 °C until constant weight. The kernels were further grounded with a stainless steel grain crusher and oven dried at 60–70 °C for 2 h to get sample powder. Each test tube contained 0.5 g of the powder sample, 2 ml of HNO<sub>3</sub>, and two drops of H<sub>2</sub>O<sub>2</sub>. The digestion was performed using a microwave digestion system (MARS5, CEM, USA). The concentrations (mg/g) of Cu, Fe, Zn, Ca, K, Mg, and P were measured using an atomic absorption spectrometer (Solar S4+Graphite Furnace System 97, Thermo Elemental, USA).

### DNA Isolation, *ZmYS1* Re-sequencing, and Analysis

Genomic DNA was extracted from maize leaves at the four-leaf stage using the CTAB method (Murray and Thompson 1980). *ZmYS1* genes from the selected 88 inbred lines were re-sequenced by BGI Life Tech Co., Ltd. using the target sequence capture technology on the NimbleGen platform. The genomic sequences of *ZmYS1* (GRMZM2G156599) from the B73 inbred line were used as the reference sequences for target sequence capture.

Sequence alignment of *ZmYS1* in the 88 inbred lines was performed using the software Clustal X (Larkin et al. 2007), with the resulting alignment further edited manually. The software DNASP 5.0 (Librado and Rozas 2009) was used to detect the sequence nucleotide polymorphism, haplotype diversity, and recombination and to test the neutral mutation hypothesis. The linkage disequilibrium (LD) between any two polymorphic sites was estimated using TASSEL v4.0 (Bradbury et al. 2007). In addition, the decay of LD with physical distance in *ZmYS1* was evaluated by the nonlinear regression (PROC NLIN

in SAS software) following Remington's model (Remington et al. 2001).

### Population Structure and Association Analyses

To exclude the effect of population structure on association mapping, all inbred lines were genotyped with the single nucleotide polymorphism (SNP) chips that contained 3072 random SNP markers evenly covering the maize genome (Yang et al. 2014a). SNP genotyping was performed via the GoldenGate assay at the National Maize Improvement Centre of China, China Agricultural University. The population structure was evaluated based on above SNP markers, and resulting  $Q$  values were calculated using the STRUCTURE program (Falush et al. 2003). Five independent runs were performed setting the number of populations ( $k$ ) from 1 to 10, burn in time and Markov chain Monte Carlo (MCMC) replication number both to 100,000, and a model for admixture and correlated allele frequencies. The  $k$  value was determined by  $\text{LnP}(D)$  in STRUCTURE output and an ad hoc statistic  $\Delta k$  based on the rate of change in  $\text{LnP}(D)$  between successive  $k$ . According to the rate of change of log likelihood, the ad hoc statistic  $\Delta k$  was estimated using the STRUCTURE HARVESTER software (Earl 2012). In order to estimate the genetic relatedness among inbred lines, pairwise relatedness coefficients (kinship matrix) were calculated using the software SPAGeDi (Hardy and Vekemans 2002). Association mapping between seven mineral concentrations in maize kernels and the nucleotide diversity of *ZmYS1* was performed using TASSEL 4.0 (Bradbury et al. 2007), with a mixed linear model (MLM) controlling both population structure and relative kinship. Only variants with a minor allele frequency (MAF) higher than 0.05 were used in the association analyses.

## Results

### Nucleotide Diversity and Selection on the *ZmYS1* Gene

Sequence polymorphism of *ZmYS1* among 88 maize inbred lines across 2684 base pairs (bp) of sequence includes seven exons covering 2049 bp and six introns covering 635 bp. Nucleotide substitutions and insertion and deletion (indel) variations at the *ZmYS1* locus are summarized in Table 1 and Supplementary Tables 2 and 3. A total of 61 SNP sites were identified, among which four and 57 sites are singleton variable sites and parsimony informative sites, respectively. In addition, ten indels covering 76 bp were identified, all of which are located in introns. For all 88 inbred lines, the overall nucleotide diversity ( $\pi$ ) of the *ZmYS1* locus is 0.0068, but the coding regions are less diverse than the intron regions (Table 1 and Supplementary Table 3). On average, single nucleotide

changes occur every 52.5 bp in the coding regions, while this value is 28.9 bp for the intron regions. When a sliding window of 100 bp under a step size of 25 bp was used, we found different frequencies of polymorphic sites in 13 regions (7 exons and 6 introns) of the *ZmYS1* gene (Fig. 1). The highest nucleotide diversity was found in the 1–100-bp region of the first exon with  $\pi=0.0314$ . In the coding region,  $\pi$  often dropped to zero, suggesting no polymorphic site. Among the 13 exon and intron regions, only exon 04 does not possess any variants. The observed distribution of SNPs and indels was significantly different (for SNP,  $\chi^2=5.615$ ,  $P<0.05$ ) from an expected even distribution between exons and introns. This uneven distribution of polymorphisms might be particularly due to the low frequency of variants in the coding region.

The Tajima's  $D$  statistic is commonly used to identify sequences that do not fit the neutral theory model at equilibrium between mutation and genetic drift (Yang et al. 2014b; Tajima 1989). In our analyses of *ZmYS1*, all estimates of Tajima's  $D$  for exons, introns, and the entire sequence region were not statistically significant, indicating no significant selection in *ZmYS1* in the tested population. However, both Fu and Li's  $D^*$  and  $F^*$  (Fu and Li 1993) are significant for exons of this gene, but their estimates were not significant for intron regions (Table 1). In addition, the estimate of Fu and Li's  $F^*$  is also significant, when the entire sequence of this gene was used. Although these results could not reject the hypothesis of mutation drift equilibrium, a lack of a footprint of positive selection in *ZmYS1* was suggested. We also noticed that the directions for all of these statistics are positive, suggesting low levels of both low and high frequency polymorphisms, possibly resulting from balancing selection on the *ZmYS1* gene.

### Haplotype Diversity of the *ZmYS1* Gene

Analyses of full sequences of *ZmYS1* identified 17 haplotypes, with a haplotype diversity ( $Hd$ ) of 0.8443 (Supplementary Table 4). The tested inbred lines are unevenly distributed in these haplotypes. Among the identified haplotypes, seven contained only one inbred line. The most frequent haplotype is Hap\_1, which consists of 16 inbred lines. The other haplotypes with more than 10 lines include Hap\_4 and Hap\_17, consisting of 12 and 15 inbred lines, respectively.

The coding region of *ZmYS1* contained 2049 bp (682 aa). Although no indel was found in the coding regions of *ZmYS1*, 39 SNPs were detected. When only coding sequences were used to evaluate the haplotype diversity, 14 haplotypes were identified, with an  $Hd$  equal to 0.8289 (Supplementary Table 4). Among these 14 haplotypes, five contain only one inbred line. The most frequent CDS haplotype was CDS\_Hap\_1, which contains 29 inbred lines. The other haplotypes with more than 10 inbred lines are CDS\_Hap\_3 and CDS\_Hap\_15, containing 14 and 15 inbred lines, respectively.

**Table 1** Summary of parameters for the analysis of nucleotide polymorphisms of the maize *ZmYS1* gene

Parameters	Coding region	Intron region	Entire region
Total length of amplicons (bp)	2049	635	2684
Number of all of the sequence variants (SNPs and indels)	39	32	71
Frequency of all of the sequence variants	0.0190	0.0504	0.0266
Number of nucleotide substitutions (bp)	39	22	61
Frequency of polymorphic sites per bp	0.0190	0.0346	0.0227
Number of indels	0	10	10
Number of indel sites	0	76	76
Average indel length	–	7.6	7.6
Frequency of indels per bp	–	0.0157	0.0037
$\pi$	0.0059	0.0102	0.0068
$\theta$	0.0039	0.0078	0.0047
Tajima's <i>D</i>	1.6263	0.9049	1.4321
Fu and Li's <i>D</i> *	1.5402*	0.9715	1.5039
Fu and Li's <i>F</i> *	1.8858*	1.1305	1.7689*

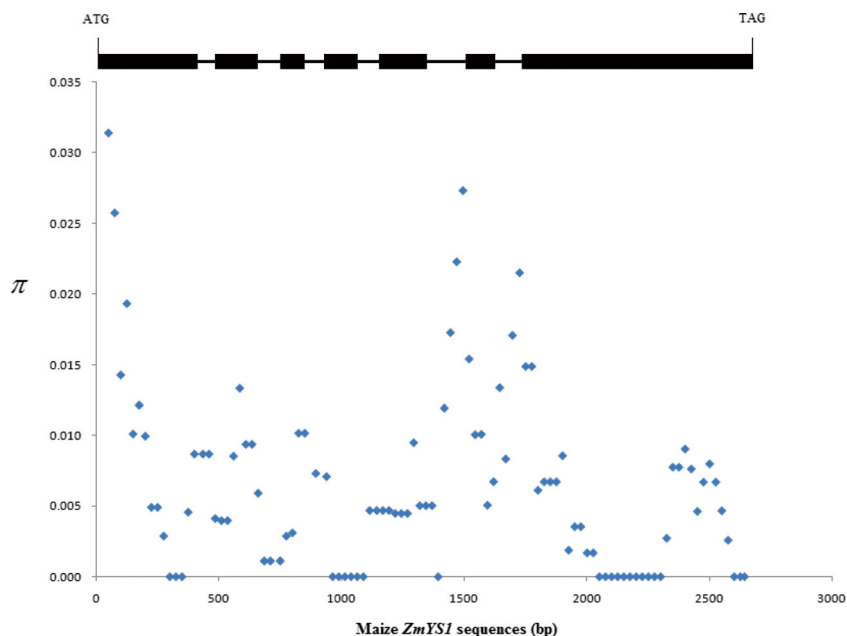
\* indicates a statistical significance at  $P < 0.05$  level

Although no indel was found in the coding region of *ZmYS1*, 13 non-synonymous sites were detected, which may cause heteromorphosis of ZmYS1 protein. Among these non-synonymous sites, the sites 1896 and 1898 belong to the same codon; therefore, these 13 non-synonymous sites will result in 12 amino acid changes. The site 44 possesses three variants encoding three different amino acids. When we translated the coding sequence (CDS) into amino acid sequences, 13 types of ZmYS1 proteins were found to be encoded (Supplementary Fig. 1). Haplotypes CDS\_Hap\_1 and CDS\_Hap\_4 encoded the most frequent type of ZmYS1 protein and contained 40 inbred lines.

### Linkage Disequilibrium and Recombination Events

Linkage disequilibrium (LD) was investigated between pairwise segregating sites in order to predict the expected resolution and marker density needed for candidate gene association mapping. In this analysis, all SNPs identified in *ZmYS1* were used to estimate the LD between two polymorphic sites, and  $r^2$  values were used as the index of LD (Supplementary Fig. 2). In addition, the decay of LD with increasing physical distance was also estimated according to Remington's model ( $F = 1123.40$ ,  $P < 0.0001$ ) (Remington et al. 2001). Our analyses found significant LD in most pairs

**Fig. 1** The nucleotide diversity ( $\pi$ ) estimated along the sequences of maize *ZmYS1* gene.  $\pi$  is calculated using the method of sliding windows of 100 bp with a step of 25 bp. The exons and introns were indexed on the top of the coordinate



(739 out of 326 for the tested LD possessed a value of  $r^2 < 0.1$ ; Supplementary Fig. 2). The LD decays rapidly with increasing physical distance. The predicted value of  $r^2$  declines to 0.1 within 1640 bp at the *ZmYS1* locus.

The patterns of the polymorphism identified in the tested inbred lines indicate a history of recombination at the *ZmYS1* locus. Under the algorithm of Hudson and Kaplan (Hudson and Kaplan 1985), at least 14 recombination events were responsible for the polymorphism of *ZmYS1* locus. The consequences of recombination events are evident in the pattern of polymorphisms when compared the sequence of one haplotype with others. For example, the sequence of the first exon of the Hap\_4 is the same as that of Hap\_5. However, across the first intron to the sixth intron regions, there are two variants between them. The seventh exon region of Hap\_4 was also virtually identical to Hap\_5. This result suggested that the *ZmYS1* sequence in Hap\_5 has resulted from at least two past recombination events relative to Hap\_4.

#### Phenotypic Variations and Association Analysis

The concentrations of Cu, Fe, Zn, Ca, K, Mg, and P in maize kernels were measured, and their descriptive statistics are presented in Table 2. Significant differences in all these mineral concentrations were observed among different maize inbred lines through one-way ANOVA (Table 2). These results suggest that the 88 inbred lines are representative in terms of maize mineral concentrations and are qualified for association analysis. To explore the relationship between these mineral concentrations, pairwise correlation analyses were performed, and the Pearson correlation coefficients ( $r$ ) between any two concentrations were calculated (Table 2). Interestingly, most correlations were statistically significant, and only five pairwise correlations did not reach the significant level of 0.05. The concentration of Cu in maize kernels was significantly correlated with those of Zn and Ca but not with those of Fe, K, Mg, and P. In addition, the concentration of Fe in maize kernels was significantly correlated with all other mineral contents except for those of Cu and Ca. It is worth mentioning that all the correlation coefficients were positive, suggesting that potentially similar genetic mechanisms are responsible for these mineral concentrations.

In order to determine the association of *ZmYS1* nucleotide polymorphism with maize kernel mineral concentrations, we also estimated the population structure of 88 inbred lines using 3072 SNPs. The  $k$  value was determined by LnP(D) in STRU CTURE output and an ad hoc statistic  $\Delta k$  that is based on the rate of change in LnP(D) between successive  $k$ . The  $\Delta k$  show a clear peak at the true value of  $k=3$  (Supplementary Fig. 3), suggesting that the total panel could be divided into three major subpopulations. The information of population structure was further used in candidate gene association analysis.

MLM of association analysis controlling the effects of population structure and individual relative relationship was used to identify the association between the mineral concentrations in maize kernels and the nucleotide polymorphism in *ZmYS1*. Five variation sites, including SNPs 17, 29, 401, and 443 and indel\_02, were found to be statistically associated with the Fe concentrations (Table 3) at the level of 0.01. In addition, three other SNPs, including SNPs 31, 43, and 98, were estimated to be statistically associated with Zn concentrations. No associations were identified for other mineral concentrations. Among the SNP sites showing association with Fe and Zn concentrations, three (SNPs 17, 31, and 43) are non-synonymous. These sites might cause phenotypic variation by causing changes in amino acids. Other associated sites might be the result of linkage with these non-synonymous SNPs.

We also estimated the allelic effects of the three non-synonymous SNPs associated with Fe and Zn concentrations. The SNP\_17 explains 8.01% of the variation of Fe concentration in maize kernels. The concentration of Fe in the inbred lines carrying allele *ZmYS1\_17A* is significantly lower than that of Fe in the inbred lines containing *ZmYS1\_17C* at the level of 0.05, when independent-samples  $t$  test was employed (Fig. 2). The SNP\_31 explained 8.65% of the variation of Zn concentration in maize kernels. The inbred lines carrying allele *ZmYS1\_31G* are significantly lower in Zn concentration than those carrying *ZmYS1\_31C*. The SNP\_43, explaining 12.63% of the variation of Zn concentration, can classify the inbred lines into three alleles, *ZmYS1\_43A*, *ZmYS1\_43G*, and *ZmYS1\_43T*. According to the result of one-way ANOVA, there is statistical significance among the three alleles caused by this site. In addition, we also noticed that the Zn concentration of inbred lines carrying *ZmYS1\_43A* was significantly lower than that of inbred lines carrying *ZmYS1\_43G* and *ZmYS1\_43T*. These results support the strength of the observed association between the three non-synonymous SNPs in *ZmYS1* and Fe and Zn concentrations in maize kernels.

#### Discussion

Significant phenotypic differences and nucleotide polymorphism are essential for genetic mapping via linkage or association analyses (Yang et al. 2014a; Zhang et al. 2013). Maize is a typical outcrossing crop with broad morphological variation, genetic diversity, and high effective frequency of recombination (Li et al. 2012; Yan et al. 2011). It was suggested that the divergence between two maize inbred lines is even greater than that between human and chimpanzees, which diverged about 3.5 million years ago (Buckler et al. 2009). In the present study, analyses of genomic sequences of the maize *ZmYS1* gene from 88 inbred lines identified 61 SNPs and 10

**Table 2** The result of descriptive statistics, one-way ANOVA, and pairwise correlation coefficients for kernel mineral concentrations among 88 maize inbred lines

Parameters	Cu	Fe	Zn	Ca	K	Mg	P
Mean	0.00135	0.04828	0.02416	0.32717	1.80289	1.58814	3.27952
SD	0.00099	0.01568	0.00548	0.09796	0.46915	0.43008	0.66131
Minimum	0.00005	0.01730	0.00756	0.14810	0.05270	0.26732	0.47258
Maximum	0.00661	0.11736	0.04027	0.62153	2.84598	2.68324	4.69292
<i>F</i>	3.6086	6.9693	8.1255	2.4521	17.6903	20.1378	12.1003
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Fe	0.0739						
Zn	0.2262*	0.3965**					
Ca	0.2999**	0.1302	0.4135**				
K	0.1350	0.3453**	0.4253**	0.3436**			
Mg	0.1097	0.3056**	0.5599**	0.3812**	0.3927**		
P	0.0940	0.3808**	0.5284**	0.2945**	0.7442**	0.7277**	

\* indicates a statistical significance at  $P < 0.05$  level

\*\* indicates a statistical significance at  $P < 0.01$  level

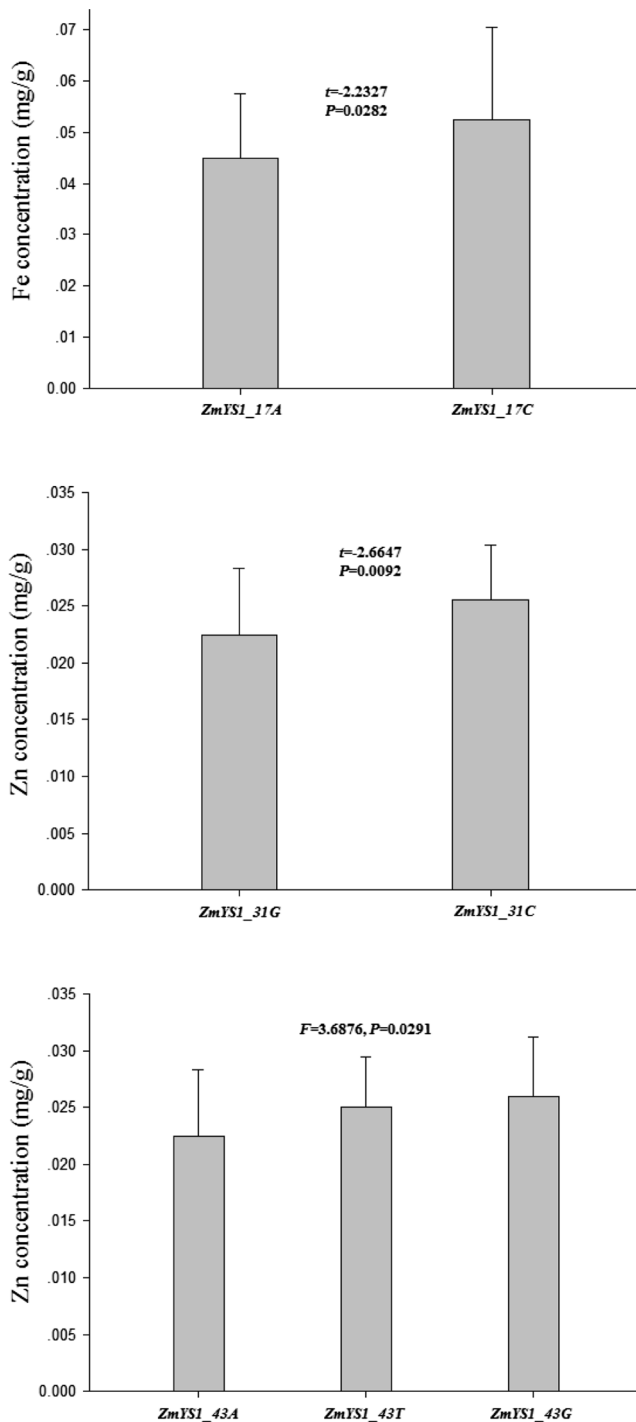
indels and an average SNP frequency of one SNP/44 bp in the entire region with a  $\pi = 0.0068$ . The nucleotide diversity of this gene is slightly higher than the overall nucleotide diversities estimated using 27 diverse inbred lines (Gore et al. 2009) and 278 temperate maize inbred lines (Jiao et al. 2012). Coding regions of functional genes tend to be relatively conserved, due to their specificity for and affinity with other types of molecules (Weng et al. 2013). In this study, the intron regions of *ZmYS1* also show higher frequency of nucleotide polymorphisms, especially indel variation. In addition, all the estimated values of Tajima's *D* in coding, intron, and entire regions were positive, indicating balancing selection with an excess of alleles of intermediate frequencies and a scarcity of rare alleles.

Association mapping through LD analysis is a powerful tool to dissect complex agronomic traits and identify alleles useful for biofortifying target traits (Yan et al. 2011). The

candidate gene method of association analysis is also a hypothesis-driven approach for complex trait dissection that aims to improve the chances of identifying the most important alleles (Hall et al. 2010). For example, the rice *GS3* gene was cloned using map-based cloning, and a mutant resulting in a stop codon in the second exon was detected as the contributory polymorphism for larger rice grain size (Fan et al. 2006, 2009). Based on the hypothesis that its orthologs in maize might possess similar roles, analyses of *GS3* orthologs in maize also revealed that several polymorphic sites are significantly associated with maize kernel size (Li et al. 2010a). Candidate gene association mapping is widely used to detect functional SNPs or alleles that are associated with gene-related agronomic traits in maize, such as *Dwarf8* (Thornsberry et al. 2001; Andersen et al. 2005) and *Vgt1* (Ducrocq et al. 2008) for maize flowering time; *ae1*, *bt2*, *sh1*, *sh2*, *sugary1*, and *waxy1* for kernel composition and

**Table 3** Association analysis between the sequence polymorphisms of the maize *ZmYS1* gene and the kernel mineral concentrations

Trait	Variant	Polymorphisms	Region	Proportion	Amino acid change	<i>F</i>	<i>P</i>	$r^2$
Fe	SNP_17	A/C	Exon01	49/39	R/S	6.9641	0.0099	0.0801
Fe	SNP_29	A/T	Exon01	49/39	–	6.9641	0.0099	0.0801
Fe	SNP_401	C/T	Exon01	63/25	–	8.8149	0.0039	0.1013
Fe	Indel_02	–/TCTGC	Intron01	62/26	–	8.3815	0.0048	0.0964
Fe	SNP_443	T/C	Intron01	63/25	–	8.8149	0.0039	0.1013
Zn	SNP_31	C/G	Exon01	48/40	A/G	7.3364	0.0082	0.0865
Zn	SNP_43	A/G/T	Exon01	40/25/23	E/G/V	8.7535	0.0044	0.1263
Zn	SNP_98	G/A	Exon01	48/40	–	7.3364	0.0082	0.0865



**Fig. 2** Comparisons of maize kernel Fe and Zn concentrations among groups carrying different *ZmYSI* alleles. *P* values for *t* test and one-way ANOVA comparing the groups carrying different alleles were indexed on the top

starch pasting properties (Wilson et al. 2004); *lcyE* (Harjes et al. 2008) and *crtRB1* (Yan et al. 2010) for carotenoid content; *GS3* (Li et al. 2010b) and *GW2* (Li et al. 2010a) for kernel shape and weight; and *Zmisa2* for starch pasting and gelatinization properties (Yang et al. 2014a).

Micronutrient malnutrition, especially zinc and iron deficiency, may seriously affect human health. Biofortification is the most viable method to tackle micronutrient malnutrition of humans in low-income countries (Bouis et al. 2011; Lung'aho et al. 2011). However, a precondition for biofortified breeding program of Fe and Zn concentrations in maize kernels is to detect the relevant genes (Lung'aho et al. 2011). In addition to association analysis, the other powerful and widely used tool for dissecting the genetic basis of micronutrient concentration in grains is quantitative trait locus (QTL) mapping (Jin et al. 2013). QTL analysis has been applied to Fe and Zn concentrations in maize kernels, and a few QTLs were detected (Qin et al. 2012; Lung'aho et al. 2011; Jin et al. 2013; Šimić et al. 2012). In this study, we showed that the maize *ZmYSI* gene possessed abundant nucleotide polymorphisms in the tested population. Our data based on association with mineral concentrations in kernels also show that the variations of this gene are significantly associated with Fe and Zn concentrations. Although both of the Fe and Zn concentrations show correlations with other mineral concentrations, no significant association was detected between the polymorphic sites and concentrations of Cu, Ca, K, Mg, and P. The uptake of Fe and Zn is a complex process that is likely influenced by multiple genes (Ueno et al. 2009; Ma 2005). Thus, our results need further verification since only *ZmYSI* was used in this study.

Many factors, including genetic diversity, rate of LD decay, sample size, and population structure, can influence the effect and power of candidate gene approach of association mapping (Yan et al. 2011). A larger population size will provide more power and precision for candidate-gene association analyses (Rafalski 2010). The present study used 88 maize inbred lines, including the representative lines of five temperate heterotic groups, tropical and waxy germplasm. This population represents much of the genetic diversity available to breeding and research programs in China. The relatively balanced sample with extensive genetic diversity in association mapping will reduce the possibility of false discovery (Xu et al. 2013). In addition, the validation of causal polymorphisms in candidate genes should be confirmed in distinct association mapping populations (Larsson et al. 2013; Kumar et al. 2014). Further research should focus on other population with adequate genetic diversity to confirm the causal variants of maize kernel Fe and Zn concentrations within *ZmYSI*.

**Acknowledgments** This work was supported by grants from the National Program on the Development of Basic Research (2011CB100100), the Priority Academic Program Development of Jiangsu Higher Education Institutions, the National Natural Science Foundations (31391632, 31200943, and 31171187), the Natural Science Foundations of Jiangsu Province (BK2012261), the Natural Science Foundation of the Jiangsu Higher Education Institutions (14KJA210005), and the Innovative Research Team of Universities in Jiangsu Province.

## References

- Andersen JR, Schrag T, Melchinger AE, Zein I, Lubberstedt T (2005) Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). Theoretical and applied genetics (TAG; Theoretische und Angewandte Genetik) 111(2):206–217. doi:10.1007/s00122-005-1996-6
- Araki R, Murata J, Murata Y (2011) A novel barley yellow stripe 1-like transporter (HvYSL2) localized to the root endodermis transports metal-phytosiderophore complexes. Plant Cell Physiol 52(11):1931–1940. doi:10.1093/pcp/pcr126
- Bouis HE, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH (2011) Biofortification: a new tool to reduce micronutrient malnutrition. Food Nutr Bull 32(1 Suppl):S31–S40
- 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(19):2633–2635. doi:10.1093/bioinformatics/btm308
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, 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, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadaya N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325(5941):714–718. doi:10.1126/science.1174276
- Chakraborti M, Prasanna B, Hossain F, Mazumdar S, Singh AM, Guleria S, Gupta H (2011) Identification of kernel iron-and zinc-rich maize inbreds and analysis of genetic diversity using microsatellite markers. J Plant Biochem Biotechnol 20(2):224–233
- Choudhary R, Watson DG (2013) Microwave drying kinetics and quality characteristics of corn. Int J Agric Biol Eng 6(1):90–99
- Curie C, Briat JF (2003) Iron transport and signaling in plants. Annu Rev Plant Biol 54:183–206. doi:10.1146/annurev.arplant.54.031902.135018
- Curie C, Panaviene Z, Loulergue C, Dellaporta SL, Briat JF, Walker EL (2001) Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. Nature 409(6818):346–349. doi:10.1038/35053080
- Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178(4):2433–2437. doi:10.1534/genetics.107.084830
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567–1587
- Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and applied genetics (TAG; Theoretische und Angewandte Genetik) 112(6):1164–1171. doi:10.1007/s00122-006-0218-1
- Fan C, Yu S, Wang C, Xing Y (2009) A causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. Theoretical and applied genetics (TAG; Theoretische und Angewandte Genetik) 118(3):465–472. doi:10.1007/s00122-008-0913-1
- Flint-Garcia SA, ThUILlet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J Cell Molec Biol 44(6):1054–1064. doi:10.1111/j.1365-313X.2005.02591.x
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133(3):693–709
- Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326(5956):1115–1117. doi:10.1126/science.1177837
- Hall D, Tegstrom C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Gen 9(2):157–165. doi:10.1093/bfgp/elp048
- 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(4):618–620
- Harjes CE, Rocheford TR, Bai L, Brumell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319(5861):330–333. doi:10.1126/science.1150255
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111(1):147–164
- Inoue H, Kobayashi T, Nozoye T, Takahashi M, Kakei Y, Suzuki K, Nakazono M, Nakanishi H, Mori S, Nishizawa NK (2009) Rice OsYSL15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. J Biol Chem 284(6):3470–3479. doi:10.1074/jbc.M806042200
- Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44(7):812–815. doi:10.1038/ng.2312
- Jin T, Zhou J, Chen J, Zhu L, Zhao Y, Huang Y (2013) The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. Breed Sci 63(3):317–324. doi:10.1270/jsbbs.63.317
- Kakei Y, Ishimaru Y, Kobayashi T, Yamakawa T, Nakanishi H, Nishizawa NK (2012) OsYSL16 plays a role in the allocation of iron. Plant Mol Biol 79(6):583–594. doi:10.1007/s11103-012-9930-1
- Kumar B, Abdel-Ghani AH, Pace J, Reyes-Matamoros J, Hochholdinger F, Lubberstedt T (2014) Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (*Zea mays* L.) seedlings. Plant Sci Int J Experiment Plant Biol 224:9–19. doi:10.1016/j.plantsci.2014.03.019
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948. doi:10.1093/bioinformatics/btm404
- Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. PLoS Genet 9(2):e1003246. doi:10.1371/journal.pgen.1003246
- Le Jean M, Schikora A, Mari S, Briat JF, Curie C (2005) A loss-of-function mutation in AtYSL1 reveals its role in iron and nicotianamine seed loading. Plant J Cell Molec Biol 44(5):769–782. doi:10.1111/j.1365-313X.2005.02569.x
- Lee S, Chiecko JC, Kim SA, Walker EL, Lee Y, Guerinot ML, An G (2009) Disruption of OsYSL15 leads to iron inefficiency in rice plants. Plant Physiol 150(2):786–800. doi:10.1104/pp.109.135418
- Li Q, Li L, Yang X, Warburton ML, Bai G, Dai J, Li J, Yan J (2010a) Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. BMC Plant Biol 10:143. doi:10.1186/1471-2229-10-143
- Li Q, Yang X, Bai G, Warburton ML, Mahuku G, Gore M, Dai J, Li J, Yan J (2010b) Cloning and characterization of a putative GS3



- ortholog involved in maize kernel development. Theoretical and applied genetics (TAG; Theoretische und Angewandte Genetik) 120(4):753–763. doi:10.1007/s00122-009-1196-x
- Li X, Zhu C, Yeh CT, Wu W, Takacs EM, Petsch KA, Tian F, Bai G, Buckler ES, Muehlbauer GJ, Timmermans MC, Scanlon MJ, Schnable PS, Yu J (2012) Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Res* 22(12):2436–2444. doi:10.1101/gr.140277.112
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452. doi:10.1093/bioinformatics/btp187
- Lucca P, Hurrell R, Potrykus I (2002) Fighting iron deficiency anemia with iron-rich rice. *J Am Coll Nutr* 21(3 Suppl):184S–190S
- Lung'aho MG, Mwaniki AM, Szalma SJ, Hart JJ, Rutzke MA, Kochian LV, Glahn RP, Hoekenga OA (2011) Genetic and physiological analysis of iron biofortification in maize kernels. *PLoS One* 6(6):e20429. doi:10.1371/journal.pone.0020429
- Ma JF (2005) Plant root responses to three abundant soil minerals: silicon, aluminum and iron. *Crit Rev Plant Sci* 24(4):267–281
- Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. *Curr Opin Plant Biol* 11(2):166–170. doi:10.1016/j.pbi.2008.01.007
- Murata Y, Ma JF, Yamaji N, Ueno D, Nomoto K, Iwashita T (2006) A specific transporter for iron(III)-phytosiderophore in barley roots. *Plant J Cell Molec Biol* 46(4):563–572. doi:10.1111/j.1365-313X.2006.02714.x
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8(19):4321–4325
- Prasad AS (2012) Discovery of human zinc deficiency: 50 years later. *J Trace Elem Med Biol Org Soc Min Trace Elem* 26(2–3):66–69. doi:10.1016/j.jtemb.2012.04.004
- Prasanna B, Pixley K, Warburton ML, Xie C-X (2010) Molecular marker-assisted breeding options for maize improvement in Asia. *Mol Breed* 26(2):339–356
- Qin H, Cai Y, Liu Z, Wang G, Wang J, Guo Y, Wang H (2012) Identification of QTL for zinc and iron concentration in maize kernel and cob. *Euphytica* 187(3):345–358
- Rafalski JA (2010) Association genetics in crop improvement. *Curr Opin Plant Biol* 13(2):174–180. doi:10.1016/j.pbi.2009.12.004
- Remington DL, Thomsberry 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 U S A* 98(20):11479–11484. doi:10.1073/pnas.201394398
- Roberts LA, Pierson AJ, Panaviene Z, Walker EL (2004) Yellow stripe1. Expanded roles for the maize iron-phytosiderophore transporter. *Plant Physiol* 135(1):112–120. doi:10.1104/pp.103.037572
- Šimić D, Drinić SM, Zdunić Z, Jambrović A, Ledenčan T, Brkić J, Brkić A, Brkić I (2012) Quantitative trait loci for biofortification traits in maize grain. *J Hered* 103(1):47–54
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595
- Tako E, Hoekenga OA, Kochian LV, Glahn RP (2013) High bioavailability iron maize (*Zea mays* L.) developed through molecular breeding provides more absorbable iron in vitro (Caco-2 model) and in vivo (*Gallus gallus*). *Nutr J* 12:3. doi:10.1186/1475-2891-12-3
- Thomsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28(3):286–289. doi:10.1038/90135
- Ueno D, Yamaji N, Ma JF (2009) Further characterization of ferric-phytosiderophore transporters ZmYS1 and HvYS1 in maize and barley. *J Exp Bot* 60(12):3513–3520. doi:10.1093/jxb/erp191
- Waters BM, Chu HH, Didonato RJ, Roberts LA, Eisleby RB, Lahner B, Salt DE, Walker EL (2006) Mutations in Arabidopsis yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. *Plant Physiol* 141(4):1446–1458. doi:10.1104/pp.106.082586
- Weng J, Li B, Liu C, Yang X, Wang H, Hao Z, Li M, Zhang D, Ci X, Li X, Zhang S (2013) A non-synonymous SNP within the isopenentenyl transferase 2 locus is associated with kernel weight in Chinese maize inbreds (*Zea mays* L.). *BMC Plant Biol* 13:98. doi:10.1186/1471-2229-13-98
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16(10):2719–2733. doi:10.1105/tpc.104.025700
- Xu F, Zhang G, Tong C, Sun X, Corke H, Sun M, Bao J (2013) Association mapping of starch physicochemical properties with starch biosynthesizing genes in waxy rice (*Oryza sativa* L.). *J Agric Food Chem* 61(42):10110–10117. doi:10.1021/jf4029688
- Xue Y, Yue S, Zhang W, Liu D, Cui Z, Chen X, Ye Y, Zou C (2014) Zinc, iron, manganese and copper uptake requirement in response to nitrogen supply and the increased grain yield of summer maize. *PLoS One* 9(4):e93895. doi:10.1371/journal.pone.0093895
- Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, Dellapenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. *Nat Genet* 42(4):322–327. doi:10.1038/ng.551
- Yan J, Warburton M, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Sci* 51(2):433–449
- Yang Z, Zhang E, Jiang Y, Xu S, Pan L, Chen Q, Xu C (2014a) Sequence polymorphisms in *Zmisa2* gene are significantly associated with starch pasting and gelatinization properties in maize (*Zea mays* L.). *Mol Breed*. doi:10.1007/s11032-014-0142-z
- Yang Z, Zhang E, Li J, Jiang Y, Wang Y, Hu Y, Xu C (2014b) Analyses of sequence polymorphism and haplotype diversity of LEAFY genes revealed post-domestication selection in the Chinese elite maize inbred lines. *Mol Biol Rep* 41(2):1117–1125. doi:10.1007/s11033-013-2958-8
- Yen MR, Tseng YH, Saier MH Jr (2001) Maize yellow stripe1, an iron-phytosiderophore uptake transporter, is a member of the oligopeptide transporter (OPT) family. *Microbiology* 147(Pt 11):2881–2883
- Zhang E, Yang Z, Wang Y, Hu Y, Song X, Xu C (2013) Nucleotide polymorphisms and haplotype diversity of RTCS gene in China elite maize inbred lines. *PLoS One* 8(2):e56495. doi:10.1371/journal.pone.0056495