Identification of candidate gene-based markers (SNPs and SSRs) in the zinc and iron transporter sequences of maize (*Zea mays* L.)

Arti Sharma[1] and R. S. Chauhan[2,*]

[1]Advanced Centre of Hill Bioresources and Biotechnology, CSK HP Agricultural University, Palampur 176 062, India
[2]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, P.O. Dumehar Bani, Kandaghat, Solan 173 215, India

Molecular marker systems such as SSRs (simple sequence repeats) and SNPs (single nucleotide polymorphisms) derived from candidate genes are proving extremely useful in the genetic improvement of crop plants. The availability of maize genome sequence has opened up the possibility of identifying candidate genes involved in zinc and iron accumulation in grains. Scanning of maize genome sequence consisting of 153,370 contigs resulted in the identification of 48 candidate genes predicted to be involved in iron and zinc transport in maize. Thirteen genes belonging to the ZIP (zinc-regulated transporter/iron-regulated transporter proteins) family, 16 to the NRAMP (natural resistance associated macrophage protein) family, 17 to the YS (yellow stripe) family, 1 to the CE (cation efflux) family and 1 to the ferritin family were identified based on the presence of characteristic signature sequences in the respective gene families. Thirty-four SSRs of di-, tri- and tetranucleotide repeats were identified in exons, introns, 3′UTRs and 5′UTRs of 28 candidate genes for zinc and iron transporters of maize. SSRs of di- and trinucleotide repeats are more prevalent in the candidate genes identified in this study. Amplification of candidate gene SSRs on a set of 124 maize inbreds, which are part of the maize biofortification programme of Department of Biotechnology, Govt of India, resulted in varying allelic diversities. SNPs were detected in exons of candidate genes *ZmZIP7*, *ZmZIP11* and *ZmNRAMP11*. The SNPs showed restriction site differences for *Msp*1 in the exon of *ZmZIP11* and for *Dpn*1 in *ZmNRAMP11*. A total of 14 markers out of which seven were from introns of *ZmZIP1*, *ZmZIP2*, *ZmZIP3*, *ZmNRAMP1*, *ZmNRAMP3*, *ZmNRAMP5* and *ZmYS14* genes, three from 3′UTRs of *ZmZIP3*, *ZmZIP10* and *ZmYS2* genes, three from SSRs in *ZmYS1*, *ZmFer1* and *ZmYS3* genes and one SNP in the *ZmNRAMP11* gene showed size polymorphism among 124 maize inbreds. Allelic frequencies varied from 0.8 to 86.3% for SSR markers, 4 to 91.94% in intron-size markers and 5.6 to 89.52% in 3′UTR size markers.

Keywords: Candidate genes, maize, zinc and iron transporters.

ZINC and iron are the mineral micronutrients most frequently deficient in humans[1]. Approximately 40% of the world's population suffers from micronutrient deficiencies (the so called 'hidden hunger'), including Zn deficiency[2]. Utilization of genetically enriched food with these micronutrients has been suggested as the most cost-effective means of managing malnutrition[3]. Recently, a sustainable solution to mineral malnutrition termed as 'biofortification' has been proposed, wherein the bioavailable concentrations of essential micronutrients are enhanced in the edible portions of crop plants through agronomic intervention or genetic selection[4]. Molecular mechanisms effecting the accumulation of iron and zinc are being investigated in rice, which is accelerating the efforts at genetic manipulation of crop plants[5,6].

Candidate gene approach is becoming a widespread method for characterizing QTLs (quantitative trait loci) as well as Mendelian traits in both the animal and plant systems[7]. Candidate genes for economically important traits have been potentially useful in plant breeding[8]. Candidate genes have been successfully used in genetic and association mapping, molecular marker-assisted selection and development of transgenic plants for various traits in crop plants[9]. Candidate genes associated with carbon metabolism in leaves and mature grains have been identified in maize[10], wherein enzyme activities were measured in mature leaves and apparent co-locations were observed between a QTL for activity and structural locus for sucrose-phosphate-synthase on chromosome 8 and for acid-soluble invertase on chromosomes 2 and 5. Wilson *et al.*[11] used an association approach to evaluate six maize candidate genes involved in kernel starch biosynthesis: amylose extender1 (*ae1*), brittle endosperm 2 (*bt2*), shrunken1 (*sh1*), *sh2*,

*For correspondence. (e-mail: rajinder.chauhan@rediffmail.com)

sugary1 and waxy1. They showed that *bt2*, *sh1* and *sh2* were significantly associated for kernel composition traits, in which *ae1* and *sh2* showed significant associations for starch-pasting properties, and *ae1* and *sh1* were associated with amylose levels. Chagne *et al.*[12] identified the *Rni* locus as a candidate gene determining red foliage and red colour in the core of apple fruit. In a population segregating for the red flesh and foliage phenotype, they determined the inheritance of the *Rni* locus and DNA polymorphisms in candidate genes involved in anthocyanin biosynthesis regulation. Similarly, Ingvarsson *et al.*[13] identified SNPs at the putative candidate gene *phyB2* in four populations of European aspen (*Populus tremula*) for day length-induced growth cessation and bud set. Simple sequence repeats (SSR) markers were identified within the candidate genes, which were potentially responsible for variation in a range of traits such as heading date, biomass and stress tolerance[14]. A candidate gene approach was used to investigate the genetic control of principal compounds (terpenes) characterizing the aroma of the berries in an interspecific grape progeny. Based on grape ESTs (expressed sequence tags), candidate genes involved in the terpenoid biosynthesis in plants were tested for polymorphism and segregation, and putative QTLs for berry content of monoterpenes were identified[15].

The exponential increase in the available DNA sequences in the databases has made it possible to identify SSRs and SNPs (single nucleotide polymorphisms) by 'database mining'. SSRs are favoured for applications in plant breeding and genetics because they are abundant in plant genomes, highly polymorphic within species, relatively rapid and inexpensive to assay, and can be used to identify specific chromosomal regions consistently across populations[16,17]. Now SNPs are gaining interest, fueled by the ever-increasing sequence data available that have revealed their abundance. The current study was aimed at identifying candidate gene markers (SSRs and SNPs) in zinc and iron transporters of maize and testing their extent of polymorphism across a range of maize inbreds with the ultimate goal of utilizing these markers in association mapping.

Materials and methods

Plant material

Maize inbreds used in the study were obtained from Hill Agriculture Research and Extension Centre (HAREC), Bajaura, Himachal Pradesh; Vivekananda Parvatiya Krishi Anusandhan Sansthan (VPKAS), Almora, Uttrakhand; Indian Agriculture and Research Institute (IARI), New Delhi; Directorate of Maize Research (DMR), New Delhi and the International Maize and Wheat Improvement Center (CIMMYT), Mexico. In addition, 20 germplasm lines received from Harvest-plus programme were also included in the study.

Identification of candidate genes for zinc and iron transporters in maize

Database search (BLASTn) was performed using previously identified zinc and iron transporter sequences from *Arabidopsis thaliana*, *Oryza sativa* and metal ion transporter sequences from other organisms available at http://plantst.genomics.purdue.edu/. The nucleotide sequences were used as query sequences against maize genome sequence in the TIGR maize genome database version 5.0 (www.tigr.org) consisting of 153,370 contigs. Maize genome contig sequences showing identity to zinc and iron transporter gene sequences from other organisms were downloaded and annotated for open reading frames (ORFs), including the promoter regions and the 3′UTRs using gene prediction algorithms of FGenesH (http://sun1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind). The amino acid sequences of ORFs were subjected to BLASTp analysis to determine the identity of identified protein sequences to known zinc and iron transporters families at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST/). The identification of conserved signature sequences of ZIP, Nramp and YS protein families was done by multiple sequence alignment using clustalW algorithm in MegAlin module of DNASTAR. The identities of signature sequences for protein families were confirmed in Pfam (http://www.sanger.ac.uk/cgi-bin/Pfam/dql.pl).

Identification of SSRs and SNPs in candidate genes

SSR motifs were identified in exons, introns, 3′UTR and 5′UTR regions of candidate genes using simple sequence repeat identification tool (SSRIT) at (http://www.gramene.org/db/searches/ssrtool). SSRs with minimum repeat length of eight nucleotides were considered in this study (Table 1). For SNP detection, ESTs and genomic sequences corresponding to the candidate genes were downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/dbEST/index.html) and maize genome database (http://www.maizegdb.org/est.php). SNPs were detected in candidate genes by pairwise sequence alignment of target regions of candidate genes with either the ESTs or genomic sequences for the target regions from 3 to 5 genotypes of maize for which corresponding sequences were available in the GenBank. Restriction sites were detected in the SNP-containing sequences using MapDraw module of DNASTAR for converting into cleaved amplified polymorphic sequence (CAPS) analysis. Twenty-one primer pairs were designed from the sequences flanking the microsatellite repeat motifs using Primer 3.0 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer_3_www.cgi; Table 1). Primer pairs were also designed from introns, 3′UTR and 5′UTR regions of candidate genes for zinc and iron transport so as to detect size polymorphisms due to indels (Table 2).

Table 1. Primer sequences to amplify SSRs in maize genome sequence contigs

Gene	Contig number	Primer sequence	Repeat motif	Location of SSR with respect to gene	Primer position in the contig	Product size (bp)
ZmZIP3	AZM5_12866-1F	F: TCCTCCTCCTCCTCCTCCT	(GC)5	5′UTR	2802→	194
	AZM5_12866-1R	R: CCTTGATGAGGAAGAACACGTC			2996←	
ZmZIP3	AZM5_12866-2F	F: CCATCCTGCTAACTATTGCTTTTT	(TCC)6	Exon	525→	234
	AZM5_12866-2R	R: GAGGTCGATTTGTGCACTAGTATTT			759←	
ZmZIP2	AZM5_6047-1F	F: GTGTTCGTCGTGGCCATACTC	(CCT)5	3′UTR	3968→	175
	AZM5_6047-1R	R: GTGCAGCATGTGGACGAA			4143←	
ZmZIP5	AZM5_13565-1F	F: CTCTCAGGTGAGCACCTTCC	(AG)8	3′UTR	1708→	159
	AZM5_13565-1R	R: CCAAATCAAAGTAGCTCGTCATATC			1867←	
ZmZIP5	AZM5_13565-2F	F: AGTCACGGGGTTCAGCAG	(CCG)5	Exon	4867→	165
	AZM5_13565-2R	R: GCATTCTGCCTCCACGAC			5032←	
ZmZIP6	AZM5_27028-1F	F: GCGGTCACCAAGAGAAGC	(GAA)5	3′UTR	791→	197
	AZM5_27028-1R	R: AAGATGGAGTGGAAGCAGAGC			988←	
ZmYS3	AZM5_16199-1F	F: AGAATCATACCTATCGCGCTAAAC	(CT)8	5′UTR	1483→	217
	AZM5_16199-1R	R: CTAACCAAGAAAGGCATCTGGTA			1700←	
ZmYS4	AZM5_35346-1F	F: GATCTGCAGCAGCTTTTGGT	(GT)10	Intron	20→	166
	AZM5_35346-1R	R: GGGACAGATACTAGTAGCGTGAAAG			186←	
ZmYS4	AZM5_35346-2F	F: GGAATCACGATCGAAGACGAC	(AAC)7	Intron	1084→	212
	AZM5_35346-2R	R: ATCACGATGAAGCTCAGGAAC			1296←	
ZmYS6	AZM5_31963-1F	F: GTAGTGGCATGGTCTGGTTCTT	(GCC)5	Exon	1096→	200
	AZM5_31963-1R	R: GCCGTCTTCTCGATCTCGT			1296←	
ZmYS1	AZM5_13741-1F	F: AGAGTAGCATGAAAAGCTTATACGG	(TTC)31	Intron	3987→	168
	AZM5_13741-1R	R: CATGTTTACTGACATATGATCTCCA			4155←	
ZmYS14	AZM5_91540-1F	F: ATCTCACATGCTGGAAATCTTCTT	(AAAT)7	Intron	880→	180
	AZM5_91540-1R	R: GAACCCGTTTATTAGAACAGCAG			1060←	
ZmYS15	AZM5_13121-1F	F: CTGGCCATCTTCACGGTAG	(GCG)5	3′UTR	2926→	197
	AZM5_13121-1R	R: ATCACCTGGCTCACGAACAT			3123←	
ZmNRAMP6	AZM5_92795-1F	F: ATGGACACGATGACCTTCTCAG	(GC)5	3′UTR	1262→	202
	AZM5_92795-1R	R: GTCGTCCAGGTCCAAGGTC			1464←	
ZmNRAMP4	AZM5_54141-1F	F: AAGTAGGTGTTCACCACGATGAG	(AT)6	Intron	579→	119
	AZM5_54141-1R	R: GCTAGAGCAAACAGAAAACTGAACT			698←	
ZmNRAMP5	AZM5_89415-1F	F: TAGACATGTGCCAATAAGAAGTCAA	(ATT)5	5′UTR	2279→	204
	AZM5_89415-1R	R: ACTCTTCCTCCTTCCTCCTCAAG			2483←	
ZmNRAMP5	AZM5_89415-2F	F: GTATGCAGATGATCCTGTCCTTC	(CAG)5	Intron	1560→	99
	AZM5_89415-2R	R: ATGAGGAATGCTTACATAGATGGAG			1659←	
ZmNRAMP5	AZM5_89415-3F	F: TACTTGAAGAAGACAGGGAGAGTGT	(CGA)5	3′UTR	825→	226
	AZM5_89415-3R	R: CCGTCATCTACCTCACCTTCAG			1051←	
ZmFER1	AZM5_17530-1F	F: CATCAGTCAGCTCAACACTTGTTAC	(AG)8	Exon	2694→	104
	AZM5_17530-1R	R: CATGCATGCTGTCTCAAAGAA			2798←	
ZmFER1	AZM5_17530-2F	F: GGTGGCCAACCATCTCTCT	(CCG)6	3′UTR	2177→	165
	AZM5_17530-2R	R: AGCTCCCCCTTGATCTCCT			2342←	

F: Forward; R: Reverse.

Genotyping of SSRs and SNPs

Genomic DNA was isolated from young leaves using CTAB method of Murray and Thompson[18]. DNA amplification for detection of SSRs and SNPs was carried out in a 25 µl reaction volume containing 20 ng template DNA, 0.2 µM of each dNTP, 1.5 mM $MgCl_2$, 1× PCR buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3), 1 unit *Taq* polymerase (MBI, Fermentas) and 0.5 µl of 10 mM forward and 0.5 µl of 10 mM reverse primer. PCR amplification was carried out in a Thermocycler (Biometra, Germany) using the following temperature profile: initial denaturation at 94°C for 3 min followed by 35 cycles at 94°C for 1 min, 58°C for 1 min, 72°C for 1 min and a final extension at 72°C for 5 min followed by cooling to 4°C. Next 10 µl of each PCR product was mixed with 2 µl of 10× gel-loading dye (0.2% bromophenol blue, 0.2% xylene cyanol dye and 30% glycerol in a TE buffer) and electrophoresed in a 3.5% agarose gel prepared in 1× Tris borate-EDTA (TBE) buffer (0.05 M Tris, 0.05 M boric acid,

Table 2. List of 5′ and 3′UTRs and intron-specific primers from candidate genes

Gene region	Primer sequence	Primer position	Product size (bp)
Zip 2_3′UTR-F	GCCCTCTCGATTCCAGTTTT	2266→	106
Zip 2_3′UTR-R	ATTTTGCCTTCAAGGACACG	2540←	
Zip 3_3′UTR-F	GGCAAAGATTTGAGAGTGCTG	5077→	476
Zip 3_3′UTR-R	GGCTTGCGAAAAATTCTTGT	5553←	
Zip 5_3′UTR-F	ATCACCGTCAGCGAGAATTT		
Zip 5_3′UTR-R	TGGTTTCCATCCTCGTTTTC		
Zip 6_3′UTR-F	TCCAGCAATATGGACACGAG	1435→	169
Zip 6_3′UTR-R	CAACAAACTGCTTCAGTTCATGT	1604←	
Zip 7_3′UTR-F	ATAAAAGGTGAAGCAGTGTCCA	124→	499
Zip 7_3′UTR-R	CATAGTTATACATGTCCAAATGTGCT	623←	
Zip 8_3′UTR-F	GGTGCCGTCAAGTAACCACT	2134→	946
Zip 8_3′UTR-R	CCTCACACAAATGTGCCAAA	3080←	
Zip 9_3′UTR-F	GCTCCTTGATGTAAAGGAATGG	4→	106
Zip 9_3′UTR-R	GCTCACAAACTGTGGTCGTG	110←	
YS 8_3′UTR-F	TGCTTGTCTGTTCGGTCAAA	4322→	169
YS 8_3′UTR-R	TTTGTTTTGTTGCACCTAGCC	4503←	
YS 1_3′UTR-F	CAGAAATGGAGAATGCGAGA	5352→	932
YS 1_3′UTR-R	GGTTGAACCCAACATGACAA	5540←	
YS 2_3′UTR-F	CCCTCTCAACATGCTTGGTC	174→	208
YS 2_3′UTR-R	CCTCTTTGGAATTGCTCTGC	389←	
YS 2_5′UTR-F	CGAGGAATGAGAATGGATTG	2908→	202
YS 2_5′UTR-R	ACTATAGCAGGAGCTAAATGTTTGT	3109←	
YS 5_3′UTR-F	GCAAGTAGACACGCCCTCAT	224→	201
YS 5_3′UTR-R	TGAACGTTTCACATTTCAAAAA	441←	
YS 7_3′UTR-F	GGGACAAGTGTTATCCATTTTCT	1153→	188
YS 7_3′UTR-R	CCTATGTCCCACGCAAGAGT	1276←	
YS 11_3′UTR-F	GTTGAAATCGTGACGTGCAT	1677→	178
YS 11_3′UTR-R	TGAAATTTGAACCCAGGTAACA	1893←	
YS 12_3′UTR-F	AGCTTCAGCAGTCACCCATT	3029→	274
YS 12_3′UTR-R	CCCATGCAGATTTCCTTTTG	3237←	
YS 13_3′UTR-F	ATCATCTGCTGCGTCCTTTT	1575→	215
YS 13_3′UTR-R	GAAGCCCTAGACATCCGAGA	643←	
YS 14_3′UTR-F	CACCTGGAAGCTAGATAAATGG	2884→	216
YS 14_3′UTR-R	CAGAAGGCAAAATCGAAAGAA	3086←	
YS 15_3′UTR-F	TCCTGATGGAGCAGGAGACT	3677→	123
YS 15_3′UTR-R	CAGCACTGCTCGTACTTGCT	3891←	
YS 16_3′UTR-F	CGAGTGATAGTTTGATCGGAGT	1597→	181
YS 16_3′UTR-R	TAATGGAGGAGCAACCATCG	1775←	
NRAMP1_3′UTR-F	TTGGGTAGAAGCAGCTTGAAA	73→	217
NRAMP1_3′UTR-R	GGGAAAAGATTATTCATACTGAGAAGG	237←	
NRAMP3_3′UTR-F	CAGCTGCCAGGAAAATAAATG	642→	214
NRAMP3_3′UTR-R	CCCTTCCAGGATTCAGTTCA	856←	
NRAMP4_5′UTR-F	TTGAGATCATTTCTTAATGCCAAA	783→	133
NRAMP4_5′UTR-R	AAGAAGCAAACGCTCAAAGAA	916←	
NRAMP6_3′UTR-F	TTGGGGAAAACATACCATCG	384→	303
NRAMP6_3′UTR-R	TATGCCCACTGTCAAAGCAG	687←	
NRAMP5_3′UTR-F	TGATTGCTCGTAGCCGTAGA	84→	649
NRAMP5_3′UTR-R	CGTAAGGACTAACTCTGAATTTTGG	733←	

1 mM, EDTA pH 8.0). The gel was run at a constant voltage of 100 V for 1.5–2 h.

Sixteen primer pairs designed from intron, 3′UTR and 5′UTR-specific amplicons of candidate genes were subjected to CAPS analysis after identifying the restriction sites for different enzymes in those regions of the candidate genes (Table 3). The amplicons were digested with various restriction enzymes *Hha*I, *Mva*I, *BamH*1, *Dpn*1, *Msp*1, *Rsa*1, *Mbo*1, *Hpa*II for CAPS analysis. For restriction digestion, 1 µl of the appropriate restriction enzyme

Table 3. PCR primers to amplify candidate gene regions for CAPS analysis

Gene	Gene region	Primer sequence	Primer position	Product size (bp)
ZmZIP1	Exon + intron	F: GCCTCCTCTGCCTCTC R: CTCTTGTTATAACAACAGCTGTGAA	712→ 1190←	478
ZmZIP2	Exon + intron	F: ACACGGTTATAACAACAGCTGTGAA R: TTCTTCTCCATCCTCGTGTGC	1257→ 653←	604
ZmZIP3	Exon ntron	F: ATATGTACGAGCGCAAGAAGCAC R: AAATCTTACTGTCCACCACCAACAT	3103→ 5241←	2138
ZmZIP4	3′UTR + exon	F: TATTGAAGCCTCCTCTGGTGAACT R: TTCAACGATCATTAAGCATCAACAT	1767→ 869←	898
ZmZIP6	3′UTR + exon	F: GGTCCTCTGTGTTTATTGCCATCTA R: GCTCTGCTTCCACTCCATCTTC	1637→ 967←	670
ZmZIP5	Exon + intron	F: ATTCCACGAAGAATGAAAGTGAAAG R: ACATCAAGGAGCAATACCTCTGAAG	2374→ 3153←	779
ZmYS1	5′UTR + exon	F: CAAGCTCGCCATCTTCATCTT R: TAAACAATCCATTCTCATTCCTCGT	2122→ 2931←	809
ZmYS2	Exon + intron	F: GTGTTTAGCGCGATAGGTATGATTC R: GGTGACGCTGGGGATGTAAT	2874→ 3710←	836
ZmYS6	5′UTR + exon	F: CCGAACCCTATGGTGTAGCAG R: GGTTTTCAATTCAGGATTGGATATTT	1616→ 862←	754
ZmYS7	5′UTR + exon	F: GATCGATCGAGAGAATAACACCTG R: GTTCACCTTCTCCCAGACAAAGAC	4351→ 5191←	840
ZmYS11	Exon + intron	F: TCACCGACCTGGACAAGAACT R: AGCTCAAGTGGTACTACATCCTGGT	1639→ 848←	791
ZmYS12	Exon	F: TACGATCCAAGAAAAACATCCACTT R: CATATCTATCGGTACTGTCCCACAA	1841→ 985←	856
ZmYS14	Exon	F: GACGAGGTCTTCAACCGAGACT R: GACGTCCTTCATCACGTTGG	2010→ 2627←	617
ZmNRAMP5	5′UTR + exon	F: GGCCTCCTTCTAATAATCCACAGTT R: CGTGCCGGTTCTTCCTCTAC	1270→ 488←	782
ZmNRAMP11	Exon + intron	F: AAGAACTGGAAGATGGAAGAATAGA R: GGCTGAGAAGGTCATCGTGTC	1111→ 562←	549
ZmZIP7	Exon + intron	F: TTTACCACGACAAATATTTAAGAACA R: ATGGGCATGAAACTAGCTAACAGTC	2017→ 2793←	776

buffer and 1 unit of appropriate restriction enzyme were added to each sample tube containing 8 µl of amplified product, incubated at 37°C for 8 h and heated to 65°C for 5 min. Digested products were mixed with 1 µl of gel-loading dye and separated on 3.5% agarose gel. The resolved PCR products were visualized on an UV trans-illuminator.

Allele frequency study

The proportion of primer pairs revealing polymorphism was computed for each class of marker. Allele frequencies were calculated for each primer pair as:

$$\text{Allele frequency} = \frac{\text{No. of inbreds in which a particular allele amplified}}{\text{Total no. of inbreds}} \times 100.$$

Results

A total of 405 contigs showing the presence of cation transporter genes were identified in the maize genome using various gene-identification parameters. Annotation of maize genome sequence contigs revealed that 46 contigs harbour ORFs for zinc and iron transporters. Forty-eight candidate genes were predicted as zinc and iron transporters in the maize genome. The zinc and iron transporters were categorized into specific protein families based on the presence of signature domains/motifs of a particular protein family as given in Pfam (http://www.sanger.ac.uk/cgi-bin/Pfam/dql.pl). Thirteen genes belonged to ZIP (zinc-regulated transporter/iron regulated transporter proteins) family, 16 to NRAMP (natural resistance associated macrophage protein) family, 17 to YS (yellow stripe) family, 1 to CE (cation efflux) family and 1 to ferritin family. Candidate genes identified in the present study were named in the same manner as proposed by Chau-

han[19] for *ZIP*, *NRAMP*, *YS* and ferritin families respectively.

Out of 48 candidate genes, 28 showed the presence of SSRs of varying repeat unit lengths. Thirty-four SSRs were identified in exons, introns, 3′UTRs and 5′UTRs of 28 candidate genes for zinc and iron transporters of maize. SSRs of di-, tri- and tetra-nucleotide repeats with repeat numbers varying from 4 to 31 were identified in the candidate genes. Out of 34 SSRs identified in this study, 19 were di-nucleotide repeats, 13 were tri-nucleotide repeats and the remaining two were tetra-nucleotide repeats. SSRs were identified in the 3′UTRs of all the *ZmZIP* genes, except *ZmZIP3* and *ZmZIP8* genes, for which tri-nucleotide repeats (CCT)5 and (CGA)6 were identified in the exons. SSR motifs were found in the 3′UTR and 5′UTR regions and introns of 13 *ZmYS* genes with the exception of *ZmYS5*, which had a (GCG)5 SSR in the exon. SSRs were detected in eight *Nramp* genes out of which the *ZmNramp6* gene had a (GC)5 motif in the exon. In the remaining seven *ZmNRAMP* genes, *ZmNRAMP3*, *ZmNRAMP4*, *ZmNRAMP5*, *ZmNRAMP7*, *ZmNRAMP9*, *ZmNRAMP10* and *ZmNRAMP15 SSR* motifs were found in the intron and 3′UTR and 5′UTR regions. SSRs were also identified in the 3′UTR of *ZmCef1* and *ZmFer1* genes. SSR locus AZM5_13741 of *ZmYS1* located in the intron region of the gene had an exceptionally long repeat unit length of $(TTC)_{31}$. The SSR loci identified in the present study were also checked for whether they were new or the same loci which had been identified in the previously reported maize SSR analysis or genome maps. For this purpose, the primer sequences of the SSR loci identified in the previous studies were searched against the contig sequences encompassing the zinc and iron transporter genes reported in the current investigation. Only one SSR locus *umc1796*, reported previously (http://www.maizegdb.org/cgi-bin/displaylocusrecord.cgi?id=292734) showed identity to a SSR locus which has been identified in the candidate gene *ZmZIP3*. No primer pair sequence for other SSR loci identified in this study showed identity to previously mapped SSR loci in the maize genome.

From among 34 SSRs identified in the target candidate gene sequences, primers were designed for 21 of them based on repeat motif length, out of which three SSR loci, AZM5_1374, AZM5_17530 and AZM5_16199, derived from *ZmYS1*, *ZmFer1* and *ZmYS3* genes respectively, showed polymorphisms on 124 maize inbred lines and the rest were monomorphic. Different size amplicons were observed from three polymorphic SSR markers. SSR locus AZM5_17530 derived from *ZmFer1* gene showed six alleles with amplicon sizes of 250, 235, 220, 215, 195 and 180 bp in 4, 4, 6, 107, 2 and 1 in maize inbred lines respectively (Figure 1 *a*). The other SSR locus, AZM5_13741 identified in the *ZmYS1* gene amplified four alleles of 200 bp in 19 lines, 190 bp in six lines, 170 bp in 97 lines and 130 bp in two inbred lines respectively (Figure 1 *b*). SSR locus AZM5_16199 identified in *ZmZIP3* gene amplified four alleles of 125, 115, 100 and 90 bp in 12, 11, 97 and 4 maize inbreds respectively (Figure 1 *c*).

The introns and 5′UTR and 3′UTR regions of candidate genes were amplified to check for size polymorphisms on 124 maize inbreds. Twelve primer pairs were derived from intron sequences and ten primer pairs were derived from either the 3′UTR or 5′UTR regions of candidate genes. Seven intron primer pairs from *ZmZIP1*, *ZmZIP2*, *ZmZIP3*, *ZmNRAMP1*, *ZmNRAMP3*, *ZmNRAMP5* and *ZmYS14* genes showed amplicon size polymorphisms in 124 maize inbreds. Three 3′UTR amplicons from *ZmZIP3*, *ZmZIP10* and *ZmYS2* genes were polymorphic in 124 maize inbreds. Allele frequencies of polymorphisms for intron and 3′UTR amplicons varied from 4 to 91.94% (Table 4). *ZmZIP1_intron* amplified five alleles with fragment sizes of 540, 520, 500, 480 and 465 bp in 5, 3, 105, 5 and 6 maize inbred lines respectively. *ZmZIP2_intron* amplified two alleles of size 435 bp in 11 inbreds and 410 bp in 113 lines (Figure 2 *a*). *ZmZIP3_intron* amplified three alleles of 240 bp in 109 lines, 225 bp in eight lines and 210 bp in seven lines. *ZmNRAMP1_intron* amplified four alleles of 230 bp amplicon in nine lines, 215 bp amplicon in 110 lines and 185 bp amplicon in five lines.

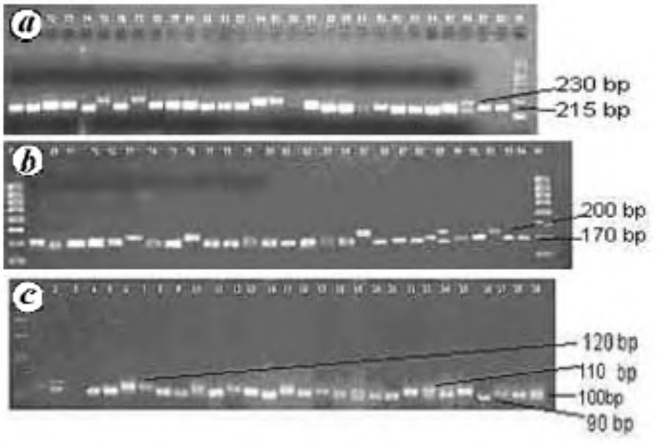

Figure 1. Polymorphism at (***a***) SSR locus AZM5_17530 in *ZmFer1* gene, (***b***) SSR locus AZM5_13741 in *ZmYS1* gene and (***c***) SSR locus AZM5_16199 in *ZmYS3* gene.

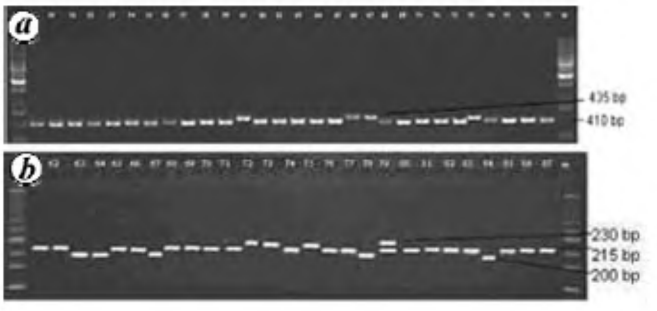

Figure 2. Polymorphism in amplicon size of (***a***) second intron of *ZmZIP2* gene and (***b***) 3′UTR of *ZmYS2* gene.

Table 4. Allele frequency and extent of polymorphism in various candidate gene markers (SSRs, SNPs and CAPS)

Molecular marker	Number of alleles	Sizes of alleles (bp)	Allele frequency (%)
AZM5_17530 (ZmFer1) (SSR)	6	250	3.22
		230	3.22
		220	4.84
		215	86.3
		195	1.6
		180	0.8
AZM5_13741 (ZmYS1) (SSR)	4	200	15.32
		190	4.84
		170	78.22
		130	1.6
AZM5_16199 (ZmYS3) (SSR)	4	120	9.68
		110	8.87
		100	78.22
		90	3.22
ZmZIP1 (intron-specific)	5	540	4.0
		520	2.4
		500	84.68
		480	4.8
		465	4.0
ZmZIP2 (intron-specific)	2	435	8.87
		410	91.13
ZmZIP3 (intron-specific)	3	240	87.9
		225	6.45
		215	5.64
ZmNRAMP1 (intron-specific)	3	230	7.26
		215	88.7
		185	4.0
ZmNRAMP3 (intron-specific)	3	940	9.68
		900	86.29
		880	4.0
ZmNRAMP5 (intron-specific)	3	295	91.94
		280	4.0
		265	4.0
ZmYS14 (intron-specific)	6	285	6.45
		165	6.45
		150	5.65
		135	64.52
		120	12.9
ZmZIP3 (3′UTR-specific)	2	475	89.52
		360	10.48
ZmZIP10 (3′UTR-specific)	3	182	5.6
		172	89.52
		162	4.8
ZmYS2 (3′UTR-specific)	3	230	5.6
		215	87.9
		200	6.45

The *ZmNRAMP3_intron* amplified three alleles of 937 bp in 12 lines, 906 bp in 107 lines and 889 bp in five lines. *ZmZIP3_3′UTR* amplified two alleles of 480 bp in 111 lines and 360 bp in 46 lines in the 3′UTR region of the gene. *ZmZIP10_3′UTR* amplified four alleles of size 185 bp in seven lines, 165 bp in 111 lines and 150 bp in six lines. *ZmYS2_3′UTR* amplified alleles of 230 bp in seven lines, 215 bp in 109 lines and 200 bp in eight lines in the 3′UTR region of the gene (Figure 2 *b*).

SNPs were detected in the exons of *ZmZIP7*, *ZmZIP11* and *ZmNRAMP11* genes. The SNPs showed restriction site differences, e.g. a SNP in *ZmZIP11* created a restriction site for *Msp*1 (Figure 3 *a*) and a SNP in *ZmNRAMP11* created a restriction site for *Dnp*1 (Figure 3 *b*). The SNP encompassing the coding region (exon) of *ZmNRAMP11* gene was amplified and subsequently digested with *Dpn*1 to reveal polymorphism (Figure 4). Sixteen primer pairs from exons and introns of zinc and iron transporter can-

didate genes from various families of zinc and iron transporters were amplified to detect SNPs in the candidate genes by CAPS analysis.

Discussion

The present study reports identification of zinc and iron transporter gene sequences in the maize genome with the objective of identification of candidate gene markers for their ultimate usage in the molecular marker-assisted selection and association mapping of zinc and iron content traits in maize. The association mapping is being projected as an alternative to linkage mapping to correlate genotypic variation to a phenotypic trait in human genetics for various disease genes[20] and in plants[21]. Forty-eight candidate genes involved in zinc and iron transport were identified in the available maize genome sequences.

Different classes of molecular markers such as SSRs, SNPs, CAPS, etc. were identified in the candidate genes encoding zinc and iron transporters. Different classes of SSR motifs such as di-, tri- and tetra-nucleotide repeats of repeat units varying from 4 to 31 were identified in most of the candidate genes. The proportion of SSRs identified in this investigation is higher in the 3′UTR region (44.12%) followed by introns (29.41%), exons (14.71%) and 5′UTR region (11.76%). As of today, there are 1800 SSR loci genetically mapped in the maize genome[22], which are accessible at http://www.maizegdb.org/ssr.php. However, for most of those SSR loci, there is no information on their association with a particular candidate gene. Twenty-one primer pairs were designed from SSRs and tested for amplification and polymorphism on a range of maize inbreds. The set of maize inbreds used in this study to identify molecular markers in zinc and iron transporter genes, is a collection of maize genotypes being evaluated for zinc and iron content under a Department of Biotechnology, Government of India and HarvestPlus programme on maize biofortification. Therefore, molecular markers identified in this study can be utilized in the biofortification programme being pursued by various governments. SSR locus AZM5_17530 (*ZmFer1*) from the 3′UTR region and SSR loci AZM5_13741 (*ZmYS1*) and AZM5_16199 (*ZmYS3*) from introns of candidate genes showed pronounced polymorphisms. Allele frequencies of intron- and 3′UTR-specific amplicons varied from 4.0 to 91.94%. A study on *sh1* locus in maize had also shown higher levels of polymorphism in the 3′UTR region of maize genes, wherein sixteen nucleotide changes were detected in 540 bp of coding regions compared to ten changes in 270 bp of the 3′UTR region[23]. Identification of SNPs is extensively pursued, particularly in those species where sequences for target genomic regions are available from multiple genotypes. Using the same analogy, SNPs were detected in the exons of *ZmZIP7*, *ZmZIP11* and *ZmNRAMP11*. SNPs in *ZmZIP11* and *ZmNRAMP11* created restriction site differences for *Msp*1 and *Dpn*1 respectively. SNP in *ZmNRAMP11* revealed size polymorphisms through CAPS analysis in maize inbreds. Rafalski[24] reported one SNP approximately every 31 bp in non-coding regions compared to one SNP in 124 bp in the coding regions of maize genes. More polymorphisms were observed in the introns of maize genes[25]. The present study has provided a useful resource of candidate gene SSRs and SNPs for zinc and iron transporters in maize, which are expected to be of potential use in genetic and association mapping, molecular marker-assisted selection and development of transgenic plants for micronutrient enrichment traits in maize.

```
ZmZIP11    ACGCTGGCCTACTGCCGGCGTTCGCGGTGTCCAGAGGGCAG
TC309791   ACGCTGGCCTACTCGCGGCGTTCGCGGTGTCCAGAGGGCAG

ZmNRAMP11  TACTGACACAGAATAGCTGATAGATTAAAAA
TC304386   TACTGCCACAGAATAGCCGATAGATTAAAAA
```

Figure 3. SNPs in the exon sequence of *ZmZIP11* gene showing difference for *Msp*1 site and *ZmNRAMP11* gene showing difference for *Dpn*1 site.

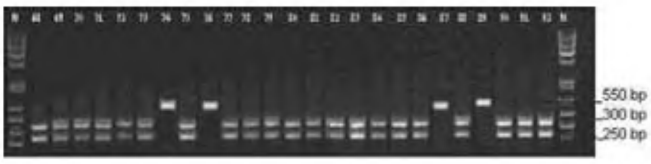

Figure 4. Polymorphism in *ZmNRAMP11* gene detected by CAPS analysis. The amplicon (550 bp) was digested with *Dpn*1.

1. Bouis, H. E., Enrichment of food staples through plant breeding: A new strategy for fighting micronutrient malnutrition. *Nutrition*, 2000, **16**, 701–704.
2. Graham, R. D. and Welch, R. M., Breeding for staple-food crops with high micronutrient density. In Agricultural Strategies for Micronutrients Working Paper 3, pp. 1–72.
3. Guerinot, M. L., The ZIP family of metal transporters. *Biochim. Biophys. Acta*, 2000, **1465**, 190–198.
4. Bouis, H. E., Genetically modified food crops and their contribution to human nutrition and food quality. *Trends Food Sci. Technol.*, 2003, **14**, 191–209.
5. Gross, J., Stein, R. J., Fett-Neto, A. G. and Fett, J. P., Iron homeostasis-related genes in rice. *Genet. Mol. Biol.*, 2003, **26**, 477–497.
6. Hacisalihoglu, G. and Kochian, L. V., How do some plants tolerate low levels of soil zinc? Mechanisms of zinc efficiency in crop plants. *New Phytol.*, 2003, **159**, 341–350.
7. Remington, D. L. and Purugganan, M. D., Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sci.*, 2003, **164**, 7–20.
8. Pflieger, S., Lefebvre, V. and Causse, M., The candidate gene approach in plant genetics: A review. *Mol. Breed.*, 2004, **7**, 275–291.
9. Barret, P., Brinkman, M., Dufour, P., Murigneux, A. and Beckert, M., Identification of candidate genes for *in vitro* androgenesis induction in maize. *Theor. Appl. Genet.*, 2004, **109**, 1660–1668.
10. Prioul, J. L., Pelleschi, S., Se´ne, M., The´venot, C., Causse, M., de Vienne, D. and Leonardi, A., From QTLs for enzyme activity to candidate genes in maize. *J. Exp. Bot.*, 1999, **50**, 1281–1288.
11. Wilson, L. M., Whitt, S. R., Ibanez, A. M., Rocheford, T. R., Goodman, M. M. and Buckler, E. S., Dissection of maize kernel

composition and starch production by candidate gene association. *Plant Cell*, 2004, **16**, 2719–2733.
12. Chagne, D. *et al.*, Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. *BMC Genomics*, 2007, **8**, 212–223.
13. Ingvarsson, P. K., Garcia, M. V., Hall, D., Lequez, V. and Jansson, S., Clinal variation in *phyB2*, a candidate gene for day length induced growth cessation and bud set, across a latitudinal gradient in European Aspen (*Populus tremula*). *Genetics*, 2006, **172**, 1845–1853.
14. Gill, G., Bryant, C., Whittaker, D., Puthigae, S., Jones, C. and Elborough, K., Mapping and haplotyping candidate genes for forage quality in perennial ryegrass. In Plant and Animal Genomes XI Conference, San Diego, California, USA, 14–18 January 2006.
15. Sevini, F., Marino, R., Grando, M. S., Moser, S. and Versini, G., Mapping candidate genes and QTLS for aroma content in grape. *Acta Hortic.*, 2004, **652**, 439–446.
16. Chen, X., Temnykh, S., Xu, Y., Cho, Y. G. and McCouch, S. R., Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor. Appl. Genet.*, 1997, **95**, 553–567.
17. Chin, E. C. L., Senior, M. L., Shu, H. and Smith, J. S. C., Maize simple sequence repetitive DNA sequences: Abundance and allele variation. *Genome*, 1996, **39**, 866–873.
18. Murray, M. G. and Thompson, W. F., Rapid isolation of high molecular weight plant DNA. *Nucl. Acids Res.*, 1980, **8**, 4321–4325.
19. Chauhan, R. S., Bioinformatics approach towards the identification of candidate genes for zinc and iron transporters in maize. *Curr. Sci.*, 2005, **91**, 510–515.
20. Li, K. C. *et al.*, Finding disease candidate genes by liquid association. *Genome Biol.*, 2007, **8**, 1–20.
21. Zeng, H. *et al.*, PlantQTL-GE: A database system for identifying candidate genes in rice and *Arabidopsis* by gene expression and QTL information. *Nucl. Acids Res.*, 2007, **35**, 879–882.
22. Sharopova, N. *et al.*, Development and mapping of SSR markers for maize. *Plant Mol. Biol.*, 2002, **48**, 463–481.
23. Shattuck-Eidens, D. M., Bell, R. N., Neuhausen, S. L. and Helentjaris, T., DNA sequence variation within maize and melon: Observations from polymerase chain reaction amplification and direct sequencing. *Genetics*, 1990, **126**, 207–217.
24. Rafalski, A., Applications of single nucleotide polymorphisms in crop genetics and breeding. *Curr. Opin. Plant Biol.*, 2002, **5**, 94–100.
25. Holland, J. B., Helland, S. J., Sharopova, N. and Rhyne, D. C., Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. *Genome*, 2001, **44**, 1065–1076.

ACKNOWLEDGEMENTS. We thank the Department of Biotechnology, Govt of India for providing financial support to R.S.C. The maize inbreds were provided by Dr Satish Guleria, HAREC, Bajaura; Dr B. M. Prasanna, IARI, New Delhi; Dr H. S. Gupta, VPKAS, Almora, and CIMMYT, Mexico.

Received 25 February 2008; revised accepted 1 August 2008