Mol Breeding (2017) 37:12
DOI 10.1007/s11032-016-0602-8

Allele mining in the Argentine public maize inbred line collection of two paralogous genes encoding NAC domains

M. A. Estermann · A. Ugarte · N. Pretini · C. R. C. Vega · S. E. Olmos

Received: 20 May 2016 / Accepted: 24 November 2016
© Springer Science+Business Media Dordrecht 2017

Abstract NAC proteins are plant transcription factors involved in biotic and abiotic stress responses, with conserved sub-domain motifs. To our knowledge, no diversity studies have been carried out at the nucleotide level of candidate NAC genes of crops. Here, we investigated allele diversity and putative protein-coding sequences in two paralogous maize NAC genes, GRMZM2G179885 and GRMZM2G347043. The GRMZM2G179885 gene is the closest ortholog of *WNAC-B1*, responsible for enhancing nitrogen remobilization and senescence in wheat. The GRMZM2G347043 gene is the homolog to the characterized *ZmSNAC1* allele, which promotes the dehydration tolerance response in Arabidopsis. Nucleotide diversity for both genes was determined in a core set of the Argentine public maize collection of temperate inbred lines structured in four diverse subpopulations. The frequency of nucleotide changes at both exons and introns of GRMZM2G179885 and GRMZM2G347043 was lower (1/137 bp-1/83 bp and 1/472 bp-1/51 bp, respectively) than that of other maize genes. Tajima's D values were non-significant, indicating the absence of selection at these genes. The five sub-domains of NAC proteins were predicted at both genes, thus implying: (i) high conservation of these two genes at the genomic level, and (ii) scarce divergence between GRMZM2G179885 and GRMZM2G347043 protein-coding sequences, which confirms their relationships as ancestral paralogs. Although the functional activity of both genes in maize metabolism is still unclear, our findings contribute to demonstrating that their NAC proteins are common alleles and would serve as a source for abiotic stress improvement in the Argentine public maize inbred line collection.

Keywords *ZmSNAC1* · *WNAC-B1* · Orthologous · Paralogous · Population diversity · Synteny · Domain conservation

Electronic supplementary material The online version of this article (doi:10.1007/s11032-016-0602-8) contains supplementary material, which is available to authorized users.

M. Estermann · A. Ugarte · N. Pretini
Universidad Nacional del Noroeste de Buenos Aires (UNNOBA), Monteagudo 2772, CP 2700 Pergamino, Buenos Aires, Argentina

C. R. C. Vega
Instituto Nacional de Tecnología Agropecuaria EEA Manfredi, Ruta Nac. 9 Km 636, CP 5988 Manfredi, Córdoba, Argentina

S. E. Olmos (✉)
Instituto Nacional de Tecnología Agropecuaria EEA Pergamino, Ruta 32 Km 4,5, CP 2700 Pergamino, Buenos Aires, Argentina
e-mail: olmos.sofia@inta.gob.ar

Introduction

NAC (NAM/ATAF/CUC) proteins are plant transcription factors involved in resistance to biotic and abiotic stress (Puranik et al. 2012). At the phenotypic level, these transcription factors play essential roles in several physiological processes, such as senescence and N remobilization from vegetative tissues to grains (Uauy

Published online: 17 January 2017

et al. 2006b), as well as in tolerance to salt (Hu et al. 2006), dehydration (Lu et al. 2012), oxidative stress (You et al. 2013), low temperatures (Mao et al. 2014) and disease (Nakashima et al. 2007; Ayyanagouda and Bashashab 2013; Muthiah et al. 2013). These findings indicate that NAC transcription factors may be useful targets to improve genetic gains in crop plants via biotechnological approaches, particularly under abiotic stresses (Nuruzzaman et al. 2013; Shao et al. 2015; Wang et al. 2016b). Particularly in maize, several NAC members have been functionally characterized (Zimmermann and Werr 2005; Liu et al. 2009; Verza et al. 2011; Lu et al. 2012; Voitsik et al. 2013; Shiriga et al. 2014; Lu et al. 2015; Mao et al. 2015; Peng et al. 2015; Mao et al. 2016; Wang et al. 2016a). In contrast, diversity studies of functional NAC alleles in the maize-breeding population and studies about the contributions to maize yield and ecophysiological related traits are still scarce.

The phenotypic variation of a trait in a breeding population may be correlated with the genetic diversity at the genome or at the candidate gene level, which is the base of mapping association studies to identify quantitative trait loci (QTLs) (Flint-Garcia et al. 2005; Wallace et al. 2014). Diversity analysis of candidate genes by allele mining in breeding populations may facilitate identification of useful allelic variations to be exploited in breeding programs (Iyer-Pascuzzi et al. 2007). Also, allele mining can be effectively used to find superior and naturally occurring alleles by syntenic relationships among species (Kumar et al. 2010).

In crops, NAC transcription factor proteins seem to be conserved at the molecular level. NAC proteins contain a N-terminal dimerization and DNA binding domain, the NAC domain, and diverse C-terminal transcription regulatory domains involved in specific interactions with DNA and other proteins (Podzimska-Sroka et al. 2015). Phylogenetic analysis by whole-genome sequence analysis of nine major lineages of land plants has shown that most angiosperms have more than 100 NAC transcription factors. The traditional NAC region consists of five sub-domains: A, B, C, D, and E. Since the sub-domain E is not always present, it is not included in phylogenic studies (Zhu et al. 2012). In maize, the MEME server (http://meme-suite.org/) has allowed the detection of 32 putative motifs in the NAC family of maize, 13 of which have known functions, providing crucial information to predict the roles of these motifs in developing a particular phenotype (Shiriga et al. 2014).

Studies on synteny have revealed that all grass species sequenced to date share an ancient whole-genome duplication from the pre-grass ancestor (Schnable et al. 2012). In maize, relatively recent whole-genome duplication has been found just subsequent to the divergence of this lineage from the common ancestor of sorghum (Swigonova et al. 2004). As a result of the whole genome duplication in the common ancestor of grasses and maize, four genomes are predicted to be contained within the maize genome. Therefore, duplication of NAC genes in the maize genome is likely to be found (Fan et al. 2014). The CoGe comparative genomics web platform (Lyons et al. 2008) and the public web interface PlantSyntenyViewer (http://urgi.versailles.inra.fr/) may help to structurally compare wheat and other grasses and to incorporate transcriptome and epigenetic differences among cereals (Murat et al. 2014).

Syntenic relationships between maize and *Triticeae* crop species (Devos and Gale 1997) provide a great opportunity for the application of the allele mining strategy. Wild emmer wheat [*Triticum turgidum ssp. dicoccoides* (DIC)] is the ancestor of cultivated durum wheat (*T. turgidum* ssp. *durum*). The wild type *TtNAM-B1* (functional *WNAC-B1)* allele has been found in all 42 and in 17 of the 19 domesticated emmer accessions examined in previous studies; however, cultivated durum wheat carries a nonfunctional *WNAC-B1* allele (Uauy et al. 2006a; Uauy et al. 2006b). *WNAC-B1* is a NAC transcription factor associated both with leaf senescence and with a more efficient remobilization of nitrogen (N) from the vegetative tissues to the wheat grain (Uauy et al. 2006a). Natural allelic variants of the ortholog *WNAC-B1* have been suggested to increase or decrease protein content in the grain of small grain cereals such as rice, maize, and forage grasses (Dubcovsky et al. 2009). Other studies, however, have found that the rice ortholog *Os07g37920* is not involved in senescence but in anther fertility (Distelfeld et al. 2012). In maize, the potential contribution of the maize ortholog *WNAC-B1* associated with N metabolism is still unknown and hence studies on both diversity and functional analysis in contrasting phenotypes are needed. Since N is the most limiting macronutrient that determines grain yield of most cereals, the elucidation of the genetic basis of complex traits such as N-use efficiency (NUE) and N remobilization is of main interest. However, no NUE specific gene has been mapped to date (Han et al. 2015).

Transcription factors that regulate gene expression in response to various abiotic stresses, including water deficit conditions (drought), have been identified and the proof-of-the-concept validated (Mattoo et al. 2015). Among them, the *ZmSNAC1* gene allele from the maize inbred line CN165 seems to be a promising source of drought stress tolerance (Lu et al. 2012). Since overexpression of *ZmSNAC1* allows enhanced tolerance to dehydration of transgenic Arabidopsis, this gene is a potential target for the design of drought-tolerant crops. A similar approach in sorghum has demonstrated that its ortholog *SbSNAC1* is another candidate gene to improve stress tolerance in crops (Lu et al. 2013). Local maize-breeding programs would benefit from a functional *ZmSNAC1* allele that confers such a characteristic. However, little is known about the genetic diversity of *ZmSNAC1* in breeding populations.

The genetic diversity of the Argentine public temperate inbred maize collection has been previously assessed by microsatellite analysis (Olmos et al. 2014a; Olmos et al. 2016). This collection includes orange flint germplasm and a group of temperate inbred lines, and the population is structured into four main subpopulations: the P465-, Argentine × Caribbean Derived Stocks (ACDS)-, LP299–2-, and Iowa Stiff Stalk Synthetic (BS13-BSSS)-related lines. A set of inbred lines from this collection has been phenotyped under different soil N conditions (D'Andrea et al. 2006) for traits associated with canopy structure and biomass production. In the present study, the maize ortholog to the *WNAC-B1* gene was chosen to search for variability in six inbred lines that have shown contrasting N response (D'Andrea et al. 2009). To search for nucleotide variability of the *ZmSNAC1* gene, a mini-core set of 20 inbred lines representing the four main subpopulations of the Argentine public temperate inbred maize collection was screened.

The objectives of this study were: (i) to analyze the nucleotide diversity and linkage disequilibrium in two maize genes: GRMZM2G179885, the ortholog closest to *WNAC-B1*, and GRMZM2G347043, the homolog identical to *ZmSNAC1*, (ii) to predict the effect of polymorphisms at the protein level, in particular at the five conserved sub-domains of NAC proteins, and (iii) to discuss perspectives of functional analysis of these two genes in maize, according to the divergence of genetics and physiology of maize in terms of grain quality in comparison to other cereals.

Material and methods

Search for gene homology

The B73 RefGen_v3 reference sequence was used to search for sequence similarity using the Blastn and Blastp programs (Altschul et al. 1990) implemented at the Gramene interface (http://ensembl.gramene.org/tools.html). Identities and positives refer to the exact number of matches of amino acids with the same residues at the same positions in an alignment and to those that are similar according to their physiochemical property, respectively (Fassler and Cooper 2011).

GRMZM2G179885 was retrieved with 100 % full length identity using the query *Zea mays* ZmNAC (partial) GSS contig ZmGSStuc11–12-04.2640.2 described in Dubcovsky et al. (2009). GRMZM2G179885 was mapped at Bin 7.04 and spans from 7:158,275,103 to 158,277,300 bp reverse strand at the B73 RefGen_v3 sequence. The automated gene annotation at the Gramene interface predicted five splice variants: the first four alleles encoding NAC superfamily domains, and the last allele encoding an unrelated NAC protein. GRMZM2G179885_T02, which encodes a full NAC protein [1700 bp transcript length, three exons, two introns, a 1233-bp coding sequence (CDS), and 410 residues of translation length], was selected for further studies.

GRMZM2G347043 was retrieved with 100 % full length identity to the query *Zea mays* SNAC1 (*ZmSNAC1*) mRNA from the CN165 inbred line (GenBank ID: JQ217429). GRMZM2G347043 was mapped at Bin 1.11 and spans from 1: 292,162,441–292,164,227 forward strands at the B73 RefGen_v3 sequence. GRMZM2G347043 encodes a single transcript: NAC1–201 GRMZM2G347043_T01, and a NAC1 transcription factor called SNAC1 was selected for further studies. GRMZM2G347043_T01 has 1685 bp transcript length, two exons, one intron, a 939-bp CDS, and 312 residues of translation length.

Plant material and molecular characterization

An Argentine public collection of 111 temperate maize inbred lines (collection A) was provided by the maize-breeding program of the National Institute of Agricultural Technology (INTA, Argentina). This collection represents the most important public orange flint lines from Argentina and reference US lines, and

103 out of the 111 inbred lines have been previously genotyped with 50 simple sequence repeats (SSRs) (Olmos et al. 2014a). The current collection mostly comprises temperate inbred lines resulting from (i) inbreeding of a variety of source populations, such as intermated open-pollinated populations (composites), (ii) synthetics, (iii) planned crosses, and, to a much lesser extent, (iv) commercial hybrids. The Bayesian model-based method discriminated four main subpopulations: P465, ACDS, LP299–2, and BS13-BSSS, which represent the high genetic diversity within the Argentine public maize inbred lines. Variation can also be observed at the phenotypic level. This collection currently includes inbred founders of Recombinant Inbred Line populations for NUE, high-oleic acid, light interception, *Fusarium* sp. resistance and *Cercospora* sp. resistance studies (Olmos et al. 2016). ID and pedigree information are summarized in Supplementary Table 1.

Leaf tissue from five 6-day-old seedlings was collected to obtain a bulk for each line. Genomic DNA was isolated according to Kleinhofs et al. (1993). The quality and concentration of the genomic DNA bulks were assessed using electrophoretic analysis and a Nanodrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, MA, USA). Genomic DNA was normalized to 10 ng/μl before PCR amplification.

GRMZM2G179885 was sequenced in a set of six inbred lines (B100, ZN6, LP662, LP611, LP561, and LP2) which had shown contrasting responses to low-N stress (D'Andrea et al. 2006). For further diversity characterization at the population level, two InDel markers found within the GRMZM2G179885 region (InDel1,Ch7:158,275,507 and InDel2,Ch7:158,275,952) were selected for PCR-based marker development and genotyping in the entire Argentine public maize collection of 111 inbred lines.

GRMZM2G347043 was sequenced in a core set of five inbred lines representative of the four diverse subpopulations previously identified (Olmos et al. 2016), as follows: (i) P465, LP611, LP662, LP613, and LP168 in the P465 subpopulation; (ii) LP125-R, LP122, LP1032, LP199, and LP1044 in the ACDS subpopulation; (iii) LP299–2, LP197, LP223, LP304, and LP29 in the LP299–2 subpopulation; (iv) 2915xLP2541-A, 2915xLP2541-B, 2915xLP2541-C, (LP915x3125–2)-1–10, and the local accession of B73, # B73–05-6081 (Olmos et al. 2014b), in the BS13-BSSS subpopulation.

Sequencing primer design

The entire coding and non-coding regions of GRMZM2G179885 and GRMZM2G347043 were covered using primers designed with the Primer3Plus program (Untergasser et al. 2007). High-quality sequences were obtained by excluding the HapMap 1 (Gore et al. 2009) and HapMap 2 (Chia et al. 2012) variants from the B73 genomic sequence (http://ensembl.gramene.org/Zea_mays/). Design conditions were as follows: product size between 600 and 800 bases for sequencing and 200–300 bases for genotyping, primer size of ~20 bases, annealing temperature of 60 °C, GC content between 20 % and 80 %, and the remainder of the setting under standard parameters. Primers are listed in Supplementary Table 2.

Sequencing reaction

PCR reaction mixtures contained 30 ng of DNA, 250 nM each primer, 200 μM each dNTP, 1.5 mM Mg++, 0.5 unit Taq DNA polymerase Fermentas Cat. Num. EP0402 (Thermo Scientific), 1× PCR buffer and sterile double-distilled water to a final volume of 13 μl. The PCRs were carried out as follows: one cycle of 5 min at 94 °C; then, a touch-down cycling profile (annealing temperature 65–55 °C, 1 °C per cycle decrease) was used followed by 30 cycles of 1 min at 94 °C, 1 min at 55 °C, and 1.5 min at 72 °C; and then 10 min at 72 °C. PCR products were separated and isolated from a 1 % agarose gel. Several PCR repetitions were collected to obtain a final concentration of 30 ng/μl of purified DNA. Purification was carried out with a QIAquick Gel Extraction Kit Cat. num. 28,104 (Qiagen Inc., Valencia, CA, USA). Purified PCR products were quantified in a Nanodrop 1000 Spectrophotometer (Thermo Scientific). Sequencing was carried out by the Sanger method in an ABI3130XL capillary equipment (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were run with the forward and reverse primers to obtain the double-strand sequencing.

Sequence editing and gene annotation

Partial sequences of each allele were assembled to span the amplified sequences from the 5′-

untranslated region (UTR) to the 3′-UTR. Only the polymorphisms observed in both the forward and reverse sequencing reads of a particular line were recorded. Heterozygous single nucleotide polymorphisms (SNPs) were designated "N," whereas superimposed traces produced by the presence of heterozygous insertions/deletions (InDels) were decoded using the reverse sequence.

Sequences were annotated with the GENSCAN web server at MIT using the Maize.smat genome model (Burge and Karlin 1997). Genomic DNA and predicted proteins were aligned with ClustalW (Thompson et al. 1994) implemented in BioEdit ® version 7.0.9.0 (Hall 1999). The annotation of the B73 RefGen_v3 reference sequence was retrieved from the evidence-based gene build system (Liang et al. 2009) at the Gramene interface. All sequences were uploaded at GenBank (http://www.ncbi.nlm.nih.gov/genbank/), which provided the corresponding nucleotide accession number.

The HapMap 1 and HapMap 2 variations and new polymorphisms found by direct sequencing were projected on the B73 RefGen_v3 reference sequence, using the GRMZM2G179885_T02 and GRMZM2G347043_T01 templates. Variation consequences were defined according to the Sequence Ontology terms at http://www.sequenceontology.org/ (Eilbeck et al. 2005).

Development of InDel markers and genotyping

Two InDel markers found at GRMZM2G179885 (InDel1,Ch7:158,275,507 and InDel2,Ch7:158,275,952), expected to provide size differences to be resolved with agarose electrophoresis system, were selected for PCR-based marker development and genotyping in collection A. The Primer3Plus program was set to amplify between 200 and 300 bp DNA fragments, and the remaining conditions were kept similar to those of sequencing. PCR products were electrophoresed on 2 % agarose gel stained with 0.5 μg/ml ethidium bromide, run at 14 V/cm in TBE (Tris/Borate/EDTA) 1× Buffer for 120 min and examined under UV light on a Benchtop UV transilluminator UVP (Upland, CA , USA) provided with a digital camera, Canon PC 1089 (Canon Inc., Tokyo, Japan). Primers are listed in Supplementary Table 2.

Data analysis

Genetic diversity

The PowerMarker software V3.0 (Liu and Muse 2005) was used to calculate major allele frequencies, residual heterozygosity (observed heterozygosity), and average gene diversity (expected heterozygosity).

Nucleotide diversity

Sequence alignments from GRMZM2G179885 and GRMZM2G347043 were compared among inbred lines using the DnaSP v5.10.01 program (Librado and Rozas 2009). The nucleotide diversity among alleles was estimated using the number of polymorphic sites (S), the number of haplotypes (NHap) (Nei 1987), the average nucleotide difference (K), nucleotide diversity (π) (Nei 1987), the theta per site from S (w) (Watterson 1975; Nei 1987), parsimony informative sites (sites that have a minimum of two nucleotides that are present at least twice), and singleton variable sites (non-informative sites). Analyses were performed with both the entire nucleotide and the coding regions. Neutrality tests, Tajima's D (Tajima 1989) and Fu and Li's D* (Fu and Li 1993) were calculated to determine the nature of selection/departure from neutrality. In the case of GRMZM2G347043, it was not possible to obtain the full length alignment of LP662 and LP304; consequently, these inbred lines were discarded for the entire nucleotide analysis.

Linkage disequilibrium (LD)

LD was calculated with D, Lewontin's D', and the Pearson correlation coefficient (r^2), using the pairwise analysis of the PowerMarker software. The significance of LD was assessed with an exact *p*-value test and chi-square test using PowerMarker default settings; the significance level was set up at $P \leq 0.01$. In addition, LD (r^2) and haplotype block analysis were calculated and displayed using Haploview software version 4.2 (Barrett et al. 2005). Block structure was defined with the four-gamete rule, whose blocks are formed by consecutive markers where only three gametes are observed.

Phylogenetic tree

The genetic relatedness and evolutionary history among alleles were inferred for the entire nucleotide sequence of each gene, using the Neighbor-Joining (NJ) method (Saitou and Nei 1987) in the MEGA4 program (Tamura et al. 2007). The multiple aligned sequences were used as an input file to construct the tree. An unrooted linear NJ tree was plotted with an option of 10,000 bootstrap replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) is shown in the figure next to the branches (Felsenstein 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and expressed in the units of the number of base substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). A total of 1,835 and 1,249 positions, respectively, were observed in the final datasets for GRMZM2G179885 and GRMZM2G347043.

Results

Maize GRMZM2G179885 gene

Allele sequences of DNA 1802 bp length average, which partially covered 91 bp of the 5′-UTR and 9 bp of the 3′-UTR and covered the full length of the open reading frame (ORF) of GRMZM2G179885_T02 were amplified in all inbred lines: B100 (KP729886.1, 1799 bp), ZN6 (KP729887.1, 1783 bp), LP662 (KP729883.1, 1811 bp), LP611 (KP729882.1, 1811 bp), LP561 (KP729884.1, 1810 bp), and LP2 (KP729885.1, 1799 bp).

Direct sequencing revealed the occurrence of 28 polymorphisms (Supplementary Table 3). Sixteen variations had been previously reported at the HapMap1 and HapMap2 databases, whereas 12 variations are reported for the first time in the present study. The arrangement of variations along the full length of GRMZM2G179885 discriminated only one haplotype for each inbred line (Supplementary Table 4). The average major allele frequency, gene diversity, and heterozygosity were 0.72, 0.37, and 0.00, respectively. The overall frequency of SNPs and InDel variants was one every 65 bp.

Nucleotide diversity analysis of 1835 bp (full sites including gaps) revealed a total of 16 SNPs, six haplotypes, nine parsimony informative sites, and seven singleton variable sites within the entire sequenced region. The average number of nucleotide difference (K) was 6.857 for the entire sequence and 4.381 for the CDS (Table 1). Nine SNPs were found in CDS-1 and CDS-3 and six SNPs in introns, while no variants were found in CDS-2. The frequency of SNPs was one SNP every 115 bp for the entire sequence, one SNP every 137 bp in the CDS, and one SNP every 83 bp for the intron region.

The number of InDels in the entire region was 12. The average frequency of InDel polymorphism in the entire region was one InDel site every 153 bp. Two InDels were located in the CDS and seven InDels in introns. The frequency of InDels in the CDS and intronic regions was one every 616 bp and 71 bp, respectively. The K value was 5.238 for the entire sequence and 1.143 for the CDS (Table 2).

The nucleotide diversity π and w of the total sequence were 0.00385 and 0.00367, respectively. The CDS region had almost similar results (π = 0.00360; w =0.00303). The InDel diversity was much lower in

Table 1 Summary statistics of nucleotide diversity analysis of GRMZM2G347043 and GRMZM2G179885

Gene	Chr.	Region	No. of sites	S	Nhap	K	π	w	Fu and Li's D	Tajima's D	Pis	Svs
GRMZM2G179885	7	Entire nucleotide	1835	16	6	6.857	0.00385	0.00367	0.13037 ns	0.27745 ns	9	7
		Coding region	1232	9	6	4.381	0.00360	0.00303	0.81778 ns	1.02361 ns	7	2
GRMZM2G347043	1	Entire nucleotide	1249	6	5	1.567	0.00128	0.00141	-0.13002 ns	-0.28178 ns	4	2
		Coding region	945	2	3	0.848	0.00090	0.00059	0.85807 ns	1.20871 ns	2	0

S no. of polymorphic sites, *Nhap* no. of haplotypes, *K* average nucleotide difference, π nucleotide diversity, *w* theta (per site) from S, *Pis* parsimony informative site, *Svs* singleton variable sites

Statistical significance: ns, not significant, $P > 0.10$

Table 2 Summary statistics of InDel diversity analysis of GRMZM2G347043 and GRMZM2G179885

Gene	Chr.	Region	No. of sites	I	Nhap	K	π	w	Tajima's D
GRMZM2G179885	7	Entire nucleotide	1835	12	6	5.238	0.00285	4.898	0.37812 ns
		Coding region	1232	2	4	1.143	0.00093	0.816	1.64955 ns
GRMZM2G347043	1	Entire nucleotide	1249	5	5	1.731	0.00139	1.431	0.65310 ns
		Coding region	945	1	2	0.257	0.00027	0.278	-0.13252 ns

InDel Option: Model 1: Diallelic (non-overlapping), I = Total number of InDels events analyzed, Nhap = No. of haplotypes, K = InDel Diversity, k(i), π = InDel Diversity per site, Pi(i), w = Theta (per sequence) from I, Theta(i)-W. Statistical significance: ns, not significant, $P > 0.10$

the CDS (π =0.00093) than in the entire sequence (π =0.00285) and was substantially lower than the nucleotide diversity found within the CDS.

Tajima's test for nucleotide and InDel polymorphisms from coding and non-coding regions showed no significant differences between π and w ($P > 0.10$), showing consistency with the neutral theory. For the neutrality test of SNPs and InDel markers, positive near zero Tajima's D values were obtained for the entire sequence and CDS.

Multi-allelic r^2 for pair-wise consecutive markers ranged from 0.0083 to 1.0 (Fig. 1), whereas multi-allelic D′ values were all 1.0 except for the pair-wise PZE07152570454-PZE07152570275 markers (D′ = 0.125). The first seven markers that covered the beginning of the 5′-UTR and the entire CDS-1 were in complete LD. Besides, three blocks of LD were found: the first block expanded from the 5′-UTR5 to the intron 2 regions, whereas the other two blocks comprised the CDS-3 region, and from the CDS-3 to the beginning of the 3′-UTR, respectively. The LD plot of r^2 values as a function of pairwise distance between informative sites revealed a slight decay of the locus analyzed within 1835 bp (regression equation: Y = 0.7107–0.3639X). However, only pairwise comparison with $r^2 = 1$ was significant as determined by Fisher's exact test and the chi-square test.

Entire nucleotide sequences of GRMZM2G179885 were used to display the genetic relationship and evolutionary history of the set of inbred lines, using a NJ tree (Supplementary Fig. 1). Two major clades (I and II), supported by high bootstrap values shown next to the branches, were present (Supplementary Fig. 1). Clade I comprised two minor clades grouping flint lines. Within clade I, LP611, LP662, and LP561 clustered separately from LP2, LP611, and LP662, which derived from the historic P465 flint inbred line, shared identical haplotype along GRMZM2G179885 full length. Clade II also comprised two minor clades. Dent lines B73 and B100 were clustered separately from the historic flint ZN6, showing that all these alleles were highly distinct from the remaining alleles of flint inbred lines.

Large InDel markers found in GRMZM2G179885 (InDel1,Ch7:158,275,507 and InDel2,Ch7:158,275,952) were split in the set of seven inbred lines into two different LD blocks (blocks 2 and 3, respectively). In contrast, the genotyping of InDel1 and InDel2 PCR-based markers in the collection of 111 inbred lines resulted in a low extent but significant LD (D = 0.08; D′ = 0.65; $r^2 = 0.15$; $P = 0.000$).

The allele sizes of InDel1 and InDel2 were 245/257 bp and 219/234 bp, respectively (Supplementary Table 5). The average major allele frequency, gene diversity and heterozygosity of the two markers were 0.62, 0.45 and 0.06, respectively. InDel1 showed higher allele frequency (0.73) and lower gene diversity (0.39) than InDel2 (0.50 and 0.50, respectively). Haplotype frequency was as follows: 0.2197 (245/219 bp), 0.2803 (257/219 bp), 0.0464 (245/234 bp), and 0.4536 (257/234 bp). The most frequent haplotype (the double insertion 257/234) was shared by B73, BS13-BSSS-related lines, historic flint lines (code 104 and 105) and Cuarentín-related inbred lines (code 107 to 111). The low-frequency double deletion 245/219 bp was only present in flint inbred lines from the P465 and ACDS subpopulations, including the US B100 inbred line. Recombinants between the B73 allele double insertion at InDel1 and the double deletion from ACDS inbred lines at InDel2, 257/219 and the rare haplotype 245/234 bp were observed in P465 and ACDS and the historic inbred line A1 (code 105), respectively.

All GRMZM2G179885 alleles sequenced in the set of inbred lines showed an intact ORF, giving a full length protein which encodes NAC domains. Only two

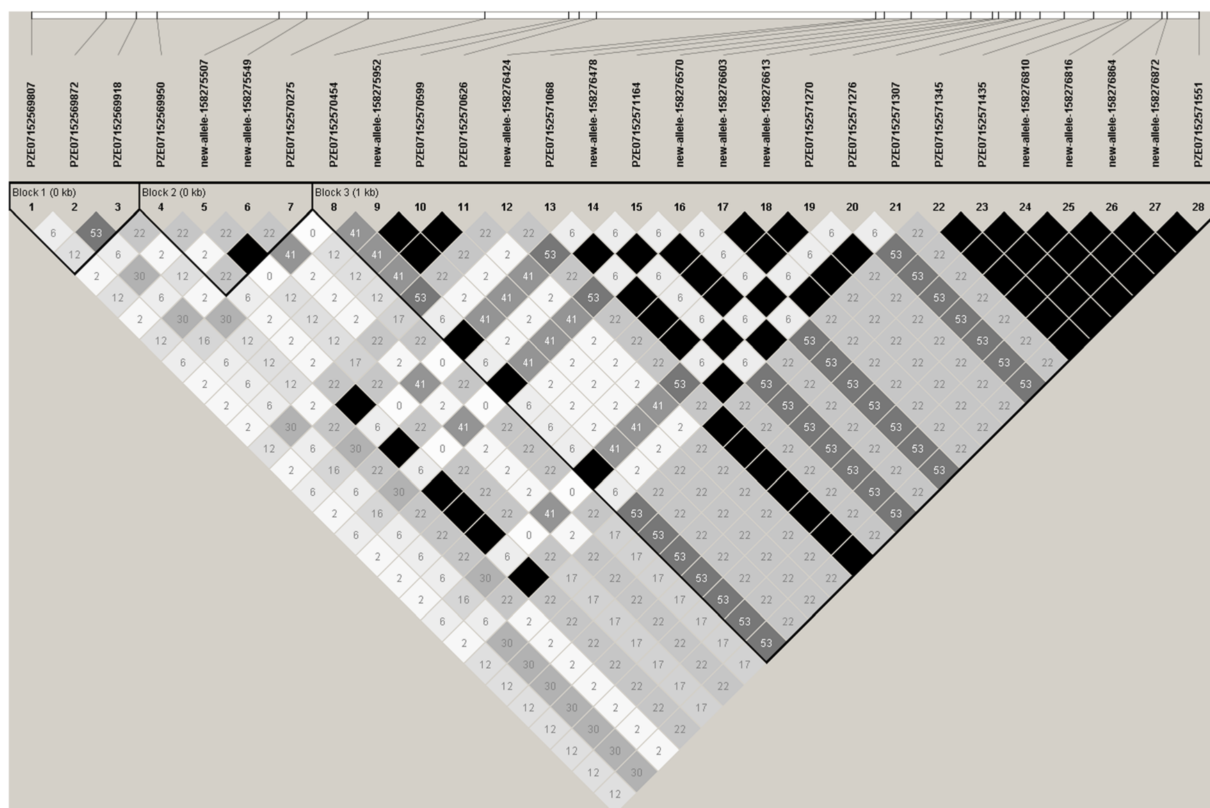

Fig. 1 Haplotype and LD analysis of the GRMZM2G179885 maize NAC gene. The numbers in each box correspond to the LD (r^2) values between markers. Strengths of the LD between markers are indicated by the following color scheme: $r^2 = 0$ (white), $0 < r^2 < 1$ (shades of gray), $r^2 = 1$ (black)

SNP-type polymorphisms were found within NAC domain coding sequences; however, these were synonymous (PZE07152571435 and PZE07152571345). InDels did not affect frame shifts: when present in the CDS, they were mostly multiple of three base pairs, producing amino acid deletions outside NAC domains. The set of inbred lines had all identical residues in the five conserved sub-domains (A, B, C, D, and E) and were identical to the WNAC-B1a (DIC) allele NAC domain of wheat (Table 3).

Maize GRMZM2G347043 gene

Sequences of 1242 bp average, which partially covered 104 bp of the 5′-UTR and 98 bp of the 3′-UTR and fully covered the ORF of GRMZM2G347043_T01, were amplified from the set of 20 inbred lines (PopSet 871,076,411: KP744840.1 to KP744859.1).

Direct sequencing revealed the occurrence of 12 polymorphisms (Supplementary Table 6), including an SNP variation within the InDels (CACC/——) at position 1: 292,162,610. Two variations had been previously reported at the HapMap1 and HapMap2 databases, whereas ten variations are reported for the first time in the present study. There were nine haplotypes along the full length of GRMZM2G347043. The average major allele frequency, gene diversity, and heterozygosity were 0.82, 0.28, and 0.00, respectively. The overall frequency of SNPs and InDel variants was one every 103 bp.

Nucleotide diversity analysis of 1249 bp revealed a total of six SNPs, five haplotypes, four parsimony informative sites, and two singleton variable sites. The overall average number of nucleotide difference (K) was 1.567 for the entire sequence and 0.848 for the CDS (Table 1). There were two SNPs in both the CDS and intronic regions. The frequency of SNPs was one SNP every 208 bp for the entire sequence, one SNP every 472 bp for the CDS, and one SNP every 51 bp for the intron.

The number of InDels in the entire region was five. The average frequency of InDel polymorphism in the

Table 3 Alignment of predicted NAC domains (A, B, C, D, and E) in the GRMZM2G179885 and GRMZM2G347043 maize genes. Sequences were aligned with ClustalW and compared with the durum wheat *WNAC-B1a* (DIC) and the allele of the CN165 maize inbred line, respectively. Alignment of NAC domains from GRMZM2G179885 and GRMZM2G347043 resulted in 81 % identities and 93 % positives.The asterisk indicates positions which have a single fully conserved residue. The colon indicates conservation between groups of strongly similar properties. The period indicates conservation between groups of weakly similar properties.

Sequence	A	B	C	D	E
WNAC-B1a (DIC) allele	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
GRMZM2G179885_T02	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
LP611	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
LP662	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
LP561	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
LP2	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
B100	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
ZN6	LPPGFRFHPTDEELVVHYLKKK	IAEVDLYKFDPWELP	GEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPI	GVKKALVFYRGKPPKGLKTNWIMHEYRL	RLDDWVLCRIYKK
	***********:***.***.:*	*******:****:**	*.:*****:*********:******.:*******.***:	*:*******.**.*:*:**:********	*********:*.*
"CN165"	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
GRMZM2G347043_T01	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
P465	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP611	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP662	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP613	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP168	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP125R	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP122	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP1032	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP199	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP1044	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP299-2	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP197	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP223	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP304	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP29	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
2915xLP2541-A	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
2915xLP2541-B	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
2915xLP2541-C	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
LP915x3125-2)-1-10	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK
INTA-Pergamino-B73	LPPGFRFHPTDDELVEHYLCRK	IAEVDLYRFDPWDLP	GAREWYFFTPRDRKYPNGSRPNRAAGNGYWKATGADKPV	GIKKALVFYAGKAPRGVKTDWIMHEYRL	RLDDWVLCRLYNK

entire sequence was one InDel site every 250 bp. Only one InDel was located in the CDSs (CDS-2); the other four were in the 5′-UTR whereas the intron had no InDels. The frequency of InDels in the CDS was one InDel every 945 bp. The K value was 1.731 for the entire sequence and 0.257 for the CDS (Table 2).

The nucleotide diversity π and w of the total sequence were 0.00128 and 0.00141, respectively. The exonic region had almost similar results (π = 0.00090; w = 0.00059). InDel diversity was much lower in the coding regions (π =0.00027) than in the entire sequence (π =0.00139) and was similar to the nucleotide diversity found within the entire sequence and coding regions.

Tajima's test for nucleotide and InDel polymorphisms from coding and non-coding regions showed no significant differences between π and w ($P > 0.10$), showing consistency with the neutral theory. However, negative Tajima's D values were obtained for SNPs within the entire sequence and for InDels within exons.

Multi-allelic r^2 for pair-wise consecutive markers ranged from 0.005 to 0.625 (Fig. 2), whereas multi-allelic D' values were 1.0 except for the pair-wise new alleles 292,162,610–292,162,619 (D′ = 0.72). Two blocks of LD were found. The first block comprised markers from the 5′-UTR, whereas the second block comprised the remaining markers. The LD plot of r^2 values as a function of pairwise distance between informative sites revealed an increase in LD within 1249 bp (regression equation: $Y = 0.0824 + 0.5480X$). However, only pairwise comparison with $r^2 = 1$ was significant as determined by Fisher's exact test and the chi-square test.

The phylogenetic NJ tree revealed the presence of two major clades (Supplementary Fig. 2). Clade I clustered two minor clades which included BS13-BSSS, LP299–2 and ACDS inbred lines. Inbreds from BS13-BSSS were less divergent and were allocated through minor clades in Clades I and II. The two sources of B73 were assigned at the minor clade within Clade I but at a different branch. LP29 and the B73 RefGen_v3 reference sequence shared an identical haplotype. The split of LP29 from the remaining LP299–2 lines is in correspondence to the genetic background of LP29 (Compuesto Colorado Precoz), which is different from the LP299–2-related lines (which derived from self-pollination of a commercial hybrid). Clade II clustered the P465-derived lines and two inbreds (2915xLP2541-A and 2915xLP2541-B) from the BS13-BSSS group at two minor clades. In Clade II, members of the minor clades shared identical haplotype along GRMZM2G347043 full length.

All GRMZM2G347043 alleles sequenced in the set of inbred lines showed an intact ORF, giving a full length protein which encodes NAC domains. Only two SNP-type polymorphisms were found within the NAC domain coding sequences; however, they were synonymous (PZE01291031431 and new allele 292,162,851). InDels did not affect frame shifts; when present in CDS-2, they were multiple of three base pairs, producing amino acid deletions outside NAC domains. The set of inbred lines had all identical residues in the five conserved sub-domains and were also identical to the CN165 inbred *ZmSNAC1* domain (Table 3).

Discussion

In the present study, a combination of the top-down approach (Andrade et al. 2015), synteny (Devos and Gale 1997) and the sequencing-based allele mining method (Ashkani et al. 2015) was applied to dissect the actual allelic variants of two NAC maize genes (Fig. 3).

Molecular diversity

To date, this is the first study on the molecular diversity of the NAC genes performed on a mini-core set of maize inbred lines from a breeding program. The frequency of nucleotide changes in GRMZM2G179885 and GRMZM2G347043 was lower than that previously found in a survey at 18 maize genes (Ching et al. 2002), where higher values, in the order of one polymorphism every 31 bp in noncoding regions and one polymorphism per 124 bp in coding regions, were obtained. Nucleotide and InDel diversities of the GRMZM2G179885 and GRMZM2G347043 genes in each respective set were low, particularly in the coding region, in agreement with diversity surveys in mini-core sets of maize inbred lines such as the phytoene synthase *PSY2* and *Y1* (Palaisa et al. 2003) and *Dwarf8* genes (Thornsberry et al. 2001). GRMZM2G179885 showed greater diversity than GRMZM2G347043. However, the diversity of GRMZM2G347043 was comparatively slightly higher than that of the *terminal ear1* (*te1*) gene of maize (White and Doebley 1999). Low nucleotide diversity may indicate that domestication bottleneck or natural or artificial selection have reduced

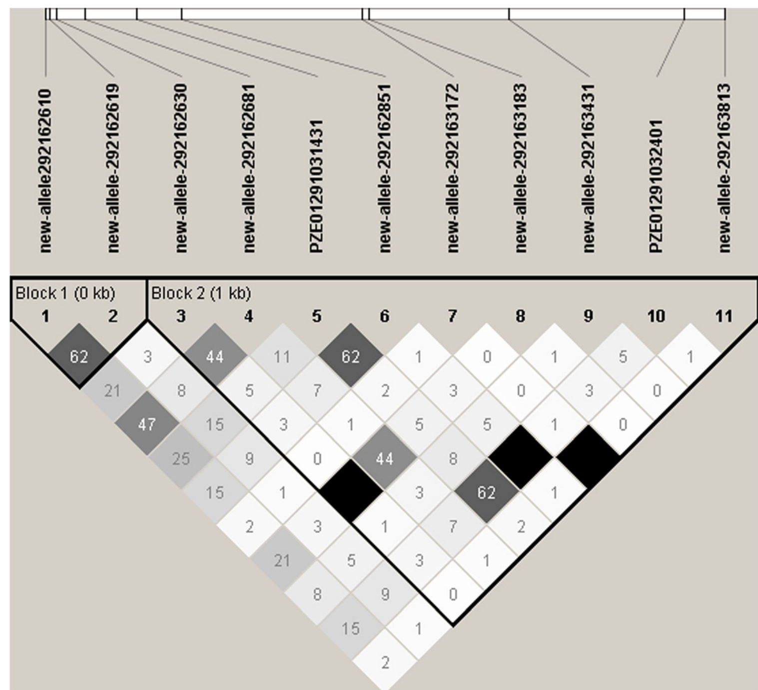

Fig. 2 Haplotype and LD analysis of the GRMZM2G347043 maize NAC gene. The numbers in each box correspond to the LD (r^2) values between markers. Strengths of the LD between markers are indicated by the following color scheme: $r^2 = 0$ (white), $0 < r^2 < 1$ (shades of gray), $r^2 = 1$ (black)

polymorphism at a locus (Buckler et al. 2001). To test for selection, Tajima's D statistics is widely applied. The maize gene analyzed in this study failed to show a significant Tajima's D, which suggests the absence of selection at the two loci and supports the assumptions of a neutral equilibrium model (Yamasaki et al. 2007). Low population size might lead to a non-significant Tajima's D test. This possibility, however, can be neglected since GRMZM2G347043 was screened in a mini-core set of inbred lines which captured maximum population structure.

Similarly, in the case of GRMZM2G179885, a low sample size of the inbred line set did not decrease the number of polymorphisms. This result also shows that the phenotypic variation for which those lines were selected (D'Andrea et al. 2006) was in correspondence to the contrasting genetic relatedness observed among them at the SSR level (Olmos et al. 2014a). The large number of singleton variable sites within the entire nucleotide of GRMZM2G179885, in particular in the non-coding region of the ZN6 inbred line, also confirms the divergent nature of the set of inbred lines. The occurrence of divergence of the ZN6 allele from the remaining flint inbred lines and its close relationship with dent lines B73 and B100 were evident from the phylogram. ZN6 is a historic flint grain type released in 1959. Microsatellite analysis of ZN6 revealed the admixture nature of ZN6 between Andean and Italian Flint groups (Camus-Kulandaivelu et al. 2006) and its close similarity with the Argentine ACDS subpopulation (Olmos et al. 2014a). We have previously speculated that the similarity between B100 and ZN6, and consequently that between B100 and ACDS, comes through the common US dent germplasm incorporated into inbred lines clustered in ACDS (Olmos et al. 2014a). Another interesting finding was that LP611 and LP662, derived from the historic P465 flint inbred line, had the same haplotype along GRMZM2G179885 length. Likewise, all P465 members shared the same haplotype in GRMZM2G347043, evidencing nucleotide conservation within this group and a large divergence from the remaining inbred lines. Variations at GRMZM2G347043 may also reveal the divergence among Argentine temperate inbred lines derived from different sources of the US B73 inbred line in the breeding history of Argentine germplasm. Since LP29 and the B73 RefGen_v3 sequence shared an identical haplotype, it is sound to hypothesize that LP29 (a semi-flint inbred line) was introgressed with a source of B73 different from that of the #B73–05-6081 currently used by the maize-breeding program of Argentina (Olmos et al. 2014b).

In this study, the genotyping of the two InDels from GRMZM2G179885 in collection A resulted in lower

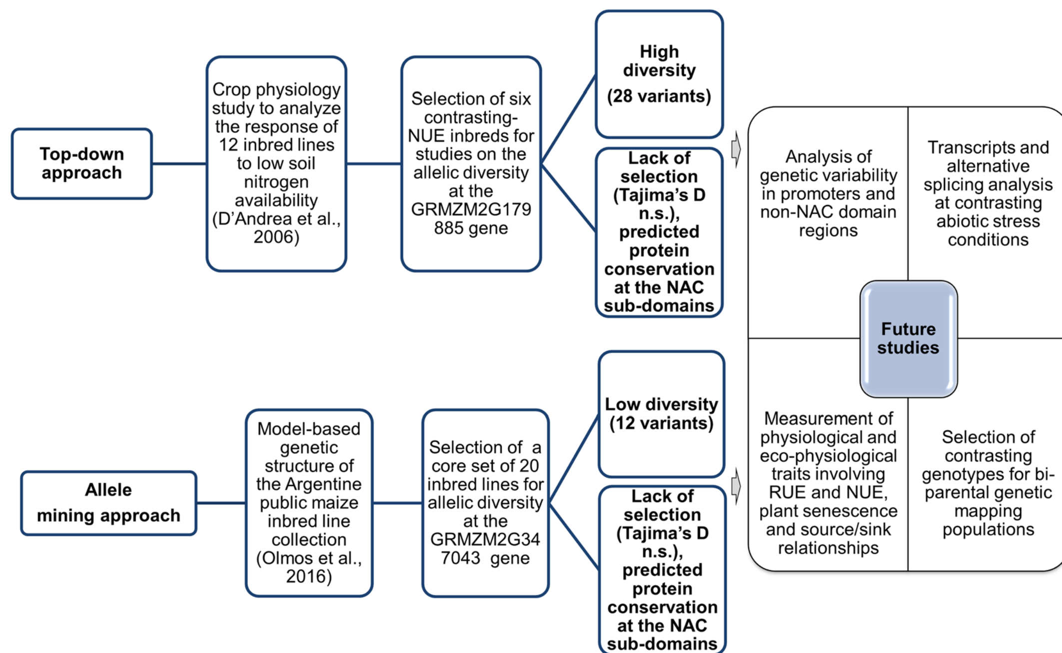

Fig. 3 Diagram linking methodology, main results and prospect studies for the GRMZM2G347043 and GRMZM2G179885 maize NAC genes

average gene diversity (0.45) than that obtained with SSRs (0.69) (Olmos et al. 2014a). Thus, a low information content of InDels is expected based on the biallelic nature of this marker type. However, the classification of the inbred line collection based on the haplotype of both InDels was almost in correspondence to the classification of the 103-inbred-line collection into four subpopulations (Olmos et al. 2014a). Unexpectedly, the most frequent haplotype, the double InDel insertion 257/234 bp, was shared between B73 and Cuarentín-related inbred lines. Cuarentín corn types have been classified into Cateto flint races from South America (Paterniani and Goodman 1977), a group that, according to (Blumenschein 1973), shows strong phenotypic similarity to Cuban flints. The haplotype sharing between B73 and Cuarentín lines elucidates the complex history from which these Caurentín lines were derived.

Protein conservation

Besides the high rate of SNPs and InDel polymorphisms found in GRMZM2G179885 and GRMZM2G347043, the characteristic five sub-domains were predicted in all alleles in these diverse sets of inbred lines, thus supporting that GRMZM2G179885 and GRMZM2G347043 functional NAC alleles are common in the subpopulations. According to the Gramene Database, GRMZM2G179885 and GRMZM2G347043 are paralogous genes, with different designations through the bibliography, for which the most common ancestor node is a duplication event (Vilella et al. 2009). The fact that identical residues of the *ZmSNAC1* CN165 inbred alleles from Lu et al. (2012) were found in collection A indicates that any of the GRMZM2G347043 alleles could be useful for crop improvement, particularly for tolerance to abiotic stresses. Phylogenetic analyses of maize NAC proteins have shown that GRMZM2G179885 (*ZmNAC88*) and GRMZM2G347043 (*ZmNAC15*) belong to the subfamily clade G, closely related to the *ZmNAC41* (GRMZM2G312201) and *ZmNAC100* (GRMZM2G068973) paralogous genes which are induced in response to *Colletotrichum graminicola* infection (Voitsik et al. 2013). However, transcriptome analysis of *ZmNAC15* (GRMZM2G347043) has exhibited a weak response during the necrotrophic stage of *C. graminicola* infection, but its promoter structure suggests a relationship with host defense regulation (Voitsik et al. 2013). Transcript analysis of maize NAC genes, including GRMZM2G347043 and GRMZM2G179885, has revealed that *ZmNAC18* (GRMZM2G347043) has one of the greatest up-regulation responses in a drought-tolerant genotype under drought stress, thus providing

additional evidence of the role of GRMZM2G347043 in drought tolerance (Shiriga et al. 2014). Similar results have been found in plantlets of the reference B73 inbred line grown under drought conditions (Peng et al. 2015). The above mentioned results do not support the possibility that GRMZM2G179885 is associated with differential responses to drought stress. In contrast, GRMZM2G179885 (designated as *ZmNAC AtVNI2-like*) colocalized with a QTL for lignin content (Barrière et al. 2015). It is known that the VND-INTERACTING2 (*VNI2*) gene integrates ABA-mediated abiotic stress during leaf aging (Yang et al. 2011), suggesting that GRMZM2G179885 is involved in the maize lignin pathway as well as in the plant responses to environmental stresses by modulation of leaf longevity.

Perspectives for future studies

The application of functional genomics to crop research needs the combination of physiological and genetic information (Fig. 3) in order to develop a more complete model of gene-to-phenotype relationships and genotype-by-environment interactions (Edmeades et al. 2004). Complementarily, new tools such as the GRASSIUS database (http://grassius.org/), aimed to improve the knowledge on transcription factors and cis-regulatory elements, will facilitate the annotation of discovered maize NAC genes (Burdo et al. 2014). Transcriptome sequencing (Tao et al. 2013) and quantitative transcript analysis (Lv et al. 2016) of crops grown under different types of stress (limited N, water or solar radiation) could also give insights into the expression of maize genes under unfavorable growing conditions. Integration of molecular, biochemical, and physiological responses (Shao et al. 2009) will allow the dissection of secondary (NUE, RUE) and primary components (Otegui et al. 2015) which limit maize yield, particularly under abiotic stress. The correlation between the phenotype and the genetic structure in inbred lines harboring useful genes including NAC genes would also give a better understanding both of the genetic variability within a maize-breeding population and of the human selection process involved in the breeding history of maize.

Although biotechnology and genotyping are important tools to explore gene resources for stress tolerance in various life forms, contributions to modern and sustainable agriculture must be evaluated in field crops using adequate phenotyping protocols and methods (Shao et al. 2009) which realistically represent the more frequent environment scenarios (Cooper et al. 2014; Cooper et al. 2016; Hammer et al. 2016). Recently, Shao et al. (2015) reviewed the progress and prospects on NAC transcription factors and pointed out their potential usefulness in stress tolerance for new emerging scenarios of climate change and complex stress environments. For that reason, we consider that the two aspects described below should be taken in account in future research aimed to link the genomic and crop physiological levels in these two NAC genes in maize (Fig. 3). First, the genome evolution in maize has led to the more competitive C_4 photosynthetic pathway to concentrate CO_2 in Rubisco-containing cells (Linder and Rudall 2005). As a consequence, maize has a higher photosynthetic capacity and NUE than C_3 species such as wheat and rice (Schmitt and Edwards 1981). Because modern maize generates a considerably high amount of grain and straw biomass with high carbon to N (C:N) ratio (USDA 2011), grain production is generally conditioned by NUE and NUE components, i.e. N uptake and N utilization efficiency (Uribelarrea et al. 2007), particularly in N-limiting environments. Because cereal crops are harvested for grain, the most relevant measure of NUE is the change in grain yield per unit change in N accumulation in aboveground biomass (Cassman et al. 2002). Although N accumulation in maize grains is low compared to wheat (e.g. around 13 g kg^{-1} and 18 g kg^{-1}, respectively) (Cassman et al. 2002), N is needed to maximize the C_4 photosynthetic pathway. Thus, N deficiency decreases phosphoenolpyruvate carboxylase (PEPCase) activity and affects the maximal efficiency of PSII photochemistry, radiation use efficiency (RUE), and yield production (Ding et al. 2005). In this manner, genetic studies in modern maize involving N limitation and closely related traits such as drought response (Bänziger et al. 2002) would need to account for the role of minor and major QTLs and candidate genes responsible of simple traits to finally cope with the understanding of complex traits such as NUE, RUE, and finally yield production at the crop level. Second, plant senescence is an essential process that allows mineral nutrients (particularly N) to be mobilized and translocated from storage to growing sink organs (Noodén 1988), particularly under abiotic or biotic stress (Thomas and Ougham 2015). Delayed leaf senescence has been postulated as a trait leading to high yields (Thomas and Stoddart 1980) and tolerance to drought (Valente et al. 2009). However, in cereal crops, there has been a long debate on whether grain yield is

determined by sink organs (i.e., the developing grain) or by source organs (i.e., the photosynthesizing vegetative tissues) (Gregersen et al. 2013), a key aspect that conditions the senescence process (Thomas and Ougham 2015). It has been demonstrated that senescence is controlled by source-sink relationships during the reproductive stage (Crafts-Brandner and Poneleit 1992; Borrás et al. 2003) under complex relationships between sugar sensing, autophagic mechanisms, and transcription factors (Thomas and Ougham 2015). The convenience of selecting for normal senescing (Hirel et al. 2001; Schippers et al. 2015) or stay-green (delayed senescence) genotypes (Ma and Dwyer 1998) needs to be further evaluated since debate still remains regarding the connection between these traits and crop yield (Thomas and Ougham 2014; Jagadish et al. 2015; Schippers et al. 2015). For instance, results of the historic durum wheat line and its near-isogenic lines contrasting in the occurrence of the *WNAC-B1* allele (Kade et al. 2005) have shown low harvest indexes (0.25) in both genotypes when grown at optimal conditions, suggesting a sink rather than a source limitation. As reviewed by Distelfeld et al. (2014), the introgression of the *WNAC-B1* allele into modern wheat cultivars results in slight yield penalties (due to reduced grain size) depending on the genetic background and the environment; however, when the *WNAC-B1* alleles are knocked out, the plants show delayed senescence, without affecting grain weight, suggesting that sink capacity may be limiting in this instance. It has been further demonstrated that wheat (*Triticum aestivum*) *NAM* RNAi plants with delayed senescence carry out 40 % more flag leaf photosynthesis after anthesis than control plants, but have the same grain-filling capacity (Borrill et al. 2015). Hence, when sink limitations condition potential yield, a rather limited/low contribution of the delayed leaf senescence trait would be expected in non-limiting environments. Because maize has a good balance of the source/sink relationship when grown in spring and under no limitation of water supply (Borrás et al. 2003; Borras et al. 2004; Bonelli et al. 2016), probably the stay-green trait would be more adequate for environments or management practices that lead to terminal stress (during kernel filling), which is also more likely to occur under the new scenarios of climate change. This would be the case for summer sowings in rainfed maize, which have become a frequent practice to avoid abiotic stress during flowering in stress-prone areas (Cerrudo et al. 2013; Caviglia et al. 2014; Tao et al. 2015; Giménez et al. 2015). However, this management practice positions the grain filling under environments with poor solar radiation and low temperatures, which can make the crop more source-limited (Tao et al. 2015; Bonelli et al. 2016). Thus, a better knowledge of the genetic basis controlling kernel filling, senescence, and C and N remobilization can contribute to improving the potential yield or to lowering its temporal variation (Thomas and Ougham 2014).

Acknowledgements This research was conducted at the National Institute of Agricultural Technology (INTA) Pergamino station, Argentina, and financed by INTA projects and PICT 2009 N° 0036 (ANPCyT). We thank all staff members who participated during laboratory analysis and Dr. Guillermo Eyherabide for providing seeds from the Argentine collection of temperate maize inbred lines.

References

Andrade FH, Sala RG, Pontaroli AC et al (2015) Chapter 19 - integration of biotechnology, plant breeding and crop physiology. Dealing with complex interactions from a physiological perspective A2 - Sadras, victor O. In: Calderini DF (ed) Crop physiology, Second edn. Academic Press, San Diego, pp. 487–503

Ashkani S, Yusop MR, Shabanimofrad M et al (2015) Allele mining strategies: principles and utilisation for blast resistance genes in rice (*Oryza sativa* L.). Curr Issues Mol Biol 17:57–74

Ayyanagouda P, Bashashab F (2013) Expressional quantitation of transcription factor genes under charcoal rot disease manifestation in *Sorghum bicolor* using qRT-PCR. Res J Biotechnol 8:67–76

Bänziger M, Edmeades GO, Lafitte HR (2002) Physiological mechanisms contributing to the increased N stress tolerance of tropical maize selected for drought tolerance. F Crop Res 75:223–233

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. doi:10.1093/bioinformatics/bth457

Barrière Y, Courtial A, Soler M, Grima-Pettenati J (2015) Toward the identification of genes underlying maize QTLs for lignin content, focusing on colocalizations with lignin biosynthetic genes and their regulatory MYB and NAC transcription factors. Mol Breed. doi:10.1007/s11032-015-0275-8

Blumenschein A (1973) Chromosome knob patterns in Latin American maize. Basic Life Sci 2:271–277

Bonelli LE, Monzon JP, Cerrudo A et al (2016) Maize grain yield components and source-sink relationship as affected by the delay in sowing date. F Crop Res 198:215–225. doi:10.1016/j.fcr.2016.09.003

Borras L, Slafer GA, Otegui ME (2004) Seed dry weight response to source-sink manipulations in wheat, maize and soybean: a quantitative reappraisal. F Crop Res 86:131–146

Borrás L, Westgate ME, Otegui ME (2003) Control of kernel weight and kernel water relations by post-flowering source–sink ratio in maize. Ann Bot 91:857–867. doi:10.1093/aob/mcg090

Borrill P, Fahy B, Smith AM, Uauy C (2015) Wheat grain filling is limited by grain filling capacity rather than the duration of flag leaf photosynthesis: a case study using NAM RNAi plants. PLoS One 10:e0134947. doi:10.1371/journal.pone.0134947

Buckler ES, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. Genet Res 77:213–218

Burdo B, Gray J, Goetting-Minesky MP et al (2014) The maize TFome–development of a transcription factor open reading frame collection for functional genomics. Plant J 80:356–366. doi:10.1111/tpj.12623

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94. doi:10.1006/jmbi.1997.0951

Camus-Kulandaivelu L, Veyrieras JB, Madur D, et al (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449-2463. doi:10.1534/genetics.105.048603

Cassman KG, Dobermann A, Walters DT (2002) Agroecosystems, nitrogen-use efficiency, and nitrogen management. Ambio 31:132–140

Caviglia OP, Melchiori RJM, Sadras VO (2014) Nitrogen utilization efficiency in maize as affected by hybrid and N rate in late-sown crops. F Crop Res 168:27–37. doi:10.1016/j.fcr.2014.08.005

Cerrudo A, Di Matteo J, Fernandez E et al (2013) Yield components of maize as affected by short shading periods and thinning. Crop Pasture Sci 64:580–587

Chia JM, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807. doi:10.1038/ng.2313

Ching A, Caldwell KS, Jung M et al (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Cooper M, Messina CD, Podlich D et al (2014) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65:311–336

Cooper M, Technow F, Messina C et al (2016) Use of crop growth models with whole-genome prediction: application to a maize Multienvironment trial. Crop Sci 56:2141–2156. doi:10.2135/cropsci2015.08.0512

Crafts-Brandner SJ, Poneleit CG (1992) Selection for seed growth characteristics: effect on leaf senescence in maize. Crop Sci. doi:10.2135/cropsci1992.0011183X003200010028x

D'Andrea KE, Otegui ME, Cirilo AG, Eyhérabide G (2006) Genotypic variability in morphological and physiological traits among maize inbred lines - nitrogen responses. Crop Sci 46:1266–1276. doi:10.2135/cropsci2005.07-0195

D'Andrea KE, Otegui ME, Cirilo AG, Eyhérabide GH (2009) Ecophysiological traits in maize hybrids and their parental inbred lines: phenotyping of responses to contrasting nitrogen supply levels. F Crop Res 114:147–158. doi:10.1016/j.fcr.2009.07.016

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Ding L, Wang KJ, Jiang GM et al (2005) Effects of nitrogen deficiency on photosynthetic traits of maize hybrids released in different years. Ann Bot 96:925–930. doi:10.1093/aob/mci244

Distelfeld A, Avni R, Fischer AM (2014) Senescence, nutrient remobilization, and yield in wheat and barley. J Exp Bot 65: 3783–3798. doi:10.1093/jxb/ert477

Distelfeld A, Pearce SP, Avni R et al (2012) Divergent functions of orthologous NAC transcription factors in wheat and rice. Plant Mol Biol 78:515–524. doi:10.1007/s11103-012-9881-6

Dubcovsky J, Fahima T, Uauy C, Distelfeld A (2009) High Grain Protein Content Gene. United States Pat. Appl. 2009020507

Edmeades GO, McMaster GS, White JW, Campos H (2004) Genomics and the physiologist: bridging the gap between genes and crop response. F Crop Res 90:5–18. doi:10.1016/j.fcr.2004.07.002

Eilbeck K, Lewis SE, Mungall CJ, et al. (2005) The Sequence Ontology: a tool for the unification of genome annotations.

Fan K, Wang M, Miao Y et al (2014) Molecular evolution and expansion analysis of the NAC transcription factor in *Zea mays*. PLoS One. doi:10.1371/journal.pone.0111837

Fassler F, Cooper P (2011) BLAST Glossary. BLAST® Help

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N Y) 39:783–791

Flint-Garcia SA, Thuillet AC, Yu J et al (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064. doi:10.1111/j.1365-313X.2005.02591.x

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

Giménez R, Mercau JL, Houspanossian J, Jobbágy EG (2015) Balancing agricultural and hydrologic risk in farming systems of the Chaco plains. J Arid Environ 123:81–92. doi:10.1016/j.jaridenv.2014.09.004

Gore MA, Chia JM, Elshire RJ et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117. doi:10.1126/science.1177837

Gregersen PL, Culetic A, Fau - Boschian L, Boschian L, Fau - Krupinska K, Krupinska K (2013) Plant senescence and crop productivity. Plant Mol Biol TA - Plant Mol, Biol 82:603–622

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hammer, G, Messina, C, van Oosterom, E, et al. (2016) Molecular breeding for complex adaptive traits: How integrating crop ecophysiology and modelling can enhance efficiency. In: Yin X, Struik P (eds) Crop Systems Biology. Springer International Publishing, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Qld, Australia DuPont-Pioneer, 7250 NW 62nd Avenue,552, Johnston, IA, United States CSIRO Plant Industry and Climate Adaptation Flag, pp 147–162

Han M, Okamoto M, Beatty PH et al (2015) The genetics of nitrogen use efficiency in crop plants. Annu Rev Genet 49:269–289

Hirel B, Bertin P, Quilleré I et al (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiol 125:1258–1270. doi:10.1104/pp.125.3.1258

Hu H, Dai M, Yao J et al (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought

resistance and salt tolerance in rice. Proc Natl Acad Sci U S A 103:12987–12992. doi:10.1073/pnas.0604882103

Iyer-Pascuzzi AS, Sweeney MT, Sarla N, McCouch SR (2007) Use of naturally occurring alleles for crop improvement. In: Rice functional genomics: challenges. Progress and Prospects. Springer New York, New York, NY, pp. 107–147

Jagadish KS, Kavi Kishor PB, Bahuguna RN et al (2015) Staying alive or going to die during terminal senescence-an enigma surrounding yield stability. Front Plant Sci 6:1070. doi:10.3389/fpls.2015.01070

Kade M, Barneix AJ, Olmos S, Dubcovsky J (2005) Nitrogen uptake and remobilization in tetraploid "Langdon" durum wheat and a recombinant substitution line with the high grain protein gene Gpc-B1. Plant Breed 124:343–349. doi:10.1111/j.1439-0523.2005.01110.x

Kleinhofs A, Kilian A, Saghai Maroof MA et al (1993) A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor Appl Genet 86:705–712. doi:10.1007/BF00222660

Kumar GR, Sakthivel K, Sundaram RM et al (2010) Allele mining in crops: prospects and potentials. Biotechnol Adv 28:451–461. doi:10.1016/j.biotechadv.2010.02.007

Liang C, Mao L, Ware D, Stein L (2009) Evidence-based gene predictions in plant genomes. Genome Res 19:1912–1923

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452. doi:10.1093/bioinformatics/btp187

Linder HP, Rudall PJ (2005) Evolutionary history of Poales. Annu Rev Ecol Evol Syst 36:107–124. doi:10.1146/annurev.ecolsys.36.102403.135635

Liu K, Muse SV (2005) PowerMaker: an integrated analysis environment for genetic maker analysis. Bioinformatics 21: 2128–2129. doi:10.1093/bioinformatics/bti282

Liu ZJ, Shao FX, Tang GY et al (2009) Cloning and characterization of a transcription factor ZmNAC1 in maize (*Zea mays*). Yi Chuan 31:199–205. doi:10.3724/SP.J.1005.2009.00199

Lu M, Sun QP, Zhang DF et al (2015) Identification of 7 stress-related NAC transcription factor members in maize (*Zea mays* L.) and characterization of the expression pattern of these genes. Biochem Biophys Res Commun 462:144–150. doi:10.1016/j.bbrc.2015.04.113

Lu M, Ying S, Zhang DF et al (2012) A maize stress-responsive NAC transcription factor, ZmSNAC1, confers enhanced tolerance to dehydration in transgenic Arabidopsis. Plant Cell Rep 31:1701–1711. doi:10.1007/s00299-012-1284-2

Lu M, Zhang DF, Shi YS et al (2013) Expression of SbSNAC1, a NAC transcription factor from sorghum, confers drought tolerance to transgenic Arabidopsis. Plant Cell Tissue Organ Cult 115:443–455. doi:10.1007/s11240-013-0375-2

Lv X, Lan S, Guy KM et al (2016) Global expressions landscape of NAC transcription factor family and their responses to abiotic stresses in *Citrullus lanatus*. Sci Rep 6:30574

Lyons E, Pedersen B, Kane J et al (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol 148: 1772–1781. doi:10.1104/pp.108.124867

Ma BL, Dwyer LM (1998) Nitrogen uptake and use of two contrasting maize hybrids differing in leaf senescence. Plant Soil 199:283–291. doi:10.1023/a:1004397219723

Mao H, Wang H, Liu S et al (2015) A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nat Commun. doi:10.1038/ncomms9326

Mao H, Yu L, Han R et al (2016) ZmNAC55, a maize stress-responsive NAC transcription factor, confers drought resistance in transgenic Arabidopsis. Plant Physiol Biochem 105: 55–66. doi:10.1016/j.plaphy.2016.04.018

Mao X, Chen S, Li A et al (2014) Novel NAC transcription factor TaNAC67 confers enhanced multi-abiotic stress tolerances in Arabidopsis. PLoS One. doi:10.1371/journal.pone.0084359

Mattoo AK, Upadhyay RK, Rudrabhatla S (2015) Abiotic stress in crops: candidate genes, osmolytes, polyamines, and biotechnological intervention. Elucidation of Abiotic Stress Signaling in Plants: Functional Genomics Perspectives, Volume 2:415–437

Murat F, Pont C, Salse J (2014) Paleogenomics in Triticeae for translational research. Curr Plant Biol 1:34–39. doi:10.1016/j.cpb.2014.08.003

Muthiah M, Ramadass A, Amalraj RS et al (2013) Expression profiling of transcription factors (TFs) in sugarcane X Colletotrichum falcatum interaction. J Plant Biochem Biotechnol 22:286–294. doi:10.1007/s13562-012-0157-7

Nakashima K, Tran LSP, Van Nguyen D et al (2007) Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. Plant J 51:617–630. doi:10.1111/j.1365-313X.2007.03168.x

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Noodén LD (1988) 1 - the phenomena of senescence and aging. Senescence and Aging in Plants. Academic Press, In, pp. 1–50

Nuruzzaman M, Sharoni AM, Kikuchi S (2013) Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front Microbiol 4:248

Olmos SE, Delucchi C, Ravera M et al (2014a) Genetic relatedness and population structure within the public Argentinean collection of maize inbred lines. Maydica 59:16–31

Olmos SE, Lia V V, Eyhérabide GH (2016) Genetic diversity and linkage disequilibrium in the Argentine public maize inbred line collection. Plant Genet Resour FirstView:1–12.

Olmos SE, Lorea R, Eyhérabide GH (2014b) Genetic variability within accessions of the B73 maize inbred line. Maydica 59: 298–305

Otegui ME, Borrás L, Maddonni GA (2015) Chapter 15 - crop phenotyping for physiological breeding in grain crops: a case study for maize A2 - Sadras, victor O. In: Second E (ed) Calderini DFBT-CP. Academic Press, San Diego, pp. 375–396

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806

Paterniani E, Goodman M (1977) Races of maize in Brazil and adjacent areas. **México**

Peng X, Zhao Y, Li X et al (2015) Genomewide identification, classification and analysis of NAC type gene family in maize. J Genet 94:377–390. doi:10.1007/s12041-015-0526-9

Podzimska-Sroka D, Shea C, Gregersen P, Skriver K (2015) NAC transcription factors in senescence: from molecular structure to function in crops. Plants 4:412

Puranik S, Sahu PP, Srivastava PS, Prasad M (2012) NAC proteins: regulation and role in stress tolerance. Trends Plant Sci 17:369–381. doi:10.1016/j.tplants.2012.02.004

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schippers JHM, Schmidt R, Wagstaff C, Jing H-C (2015) Living to die and dying to live: the survival strategy behind leaf senescence. Plant Physiol 169:914–930. doi:10.1104/pp.15.00498

Schmitt MR, Edwards GE (1981) Photosynthetic capacity and nitrogen use efficiency of maize, wheat, and rice: a comparison between C3 and C4 photosynthesis. J Exp Bot 32:459–466. doi:10.1093/jxb/32.3.459

Schnable JC, Freeling M, Lyons E (2012) Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol Evol 4:265–277. doi:10.1093/gbe/evs009

Shao H, Wang H, Tang X (2015) NAC transcription factors in plant multiple abiotic stress responses: progress and prospects. Front Plant Sci. doi:10.3389/fpls.2015.00902

Shao H-B, Chu L-Y, Jaleel CA et al (2009) Understanding water deficit stress-induced changes in the basic metabolism of higher plants - biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. Crit Rev Biotechnol 29:131–151. doi:10.1080/07388550902869792

Shiriga K, Sharma R, Kumar K et al (2014) Genome-wide identification and expression pattern of drought-responsive members of the NAC family in maize. Meta Gene 2:407–417. doi:10.1016/j.mgene.2014.05.001

Swigonova Z, Lai J, Ma J et al (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923. doi:10.1101/gr.2332504

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599. doi:10.1093/molbev/msm092

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101:11030–11035. doi:10.1073/pnas.0404206101

Tao H, Xia L, Xu L et al (2015) Poor post-silking kernel development limits summer maize yield in the North China plain. ScienceAsia 41:229–235

Tao T, Zhao L, Lv Y et al (2013) Transcriptome sequencing and differential Gene expression analysis of delayed gland morphogenesis in Gossypium Australe during seed germination. PLoS One. doi:10.1371/journal.pone.0075323

Thomas H, Ougham H (2014) The stay-green trait. J Exp Bot 65:3889–3900. doi:10.1093/jxb/eru037

Thomas H, Ougham H (2015) Chapter 10 - senescence and crop performance A2 - Sadras, victor O. In: Second E (ed) Calderini DFBT-CP. Academic Press, San Diego, pp. 223–249

Thomas H, Stoddart JL (1980) Leaf senescence. Annu Rev Plant Physiol 31:83–111. doi:10.1146/annurev.pp.31.060180.000503

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thornsberry JM, Goodman MM, Doebley J et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289. doi:10.1038/90135

Uauy C, Brevis JC, Dubcovsky J (2006a) The high grain protein content gene Gpc-B1 accelerates senescence and has pleiotropic effects on protein content in wheat. J Exp Bot 57:2785–2794. doi:10.1093/jxb/erl047

Uauy C, Distelfeld A, Fahima T et al (2006b) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301. doi:10.1126/science.1133649

Untergasser A, Nijveen H, Rao X et al (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71–W74. doi:10.1093/nar/gkm306

Uribelarrea M, Moose SP, Below FE (2007) Divergent selection for grain protein affects nitrogen use in maize hybrids. F Crop Res 100:82–90. doi:10.1016/j.fcr.2006.05.008

USDA (2011) Carbon to Nitrogen Ratios in Cropping Systems.

Valente MAS, Faria JAQA, Soares-Ramos JRL et al (2009) The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. J Exp Bot 60:533–546. doi:10.1093/jxb/ern296

Verza NC, Figueira TRS, Sousa SM, Arruda P (2011) Transcription factor profiling identifies an aleurone-preferred NAC family member involved in maize seed development. Ann Appl Biol 158:115–129. doi:10.1111/j.1744-7348.2010.00447.x

Vilella AJ, Severin J, Ureta-Vidal A et al (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19:327–335. doi:10.1101/gr.073585.107

Voitsik AM, Muench S, Deising HB, Voll LM (2013) Two recently duplicated maize NAC transcription factor paralogs are induced in response to Colletotrichum graminicola infection. BMC Plant Biol. doi:10.1186/1471-2229-13-85

Wallace JG, Bradbury PJ, Zhang N et al (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet 10:e1004845. doi:10.1371/journal.pgen.1004845

Wang D, Yu Y, Liu Z et al (2016a) Membrane-bound NAC transcription factors in maize and their contribution to the oxidative stress response. Plant Sci 250:30–39. doi:10.1016/j.plantsci.2016.05.019

Wang H, Shao H, Tang X (2016b) Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. Front Plant Sci. doi:10.3389/fpls.2016.00067

Watterson GA (1975) On the number of segregating sites in geneticalmodels without recombination. Theor Popul Biol 7:256–276

White SE, Doebley JF (1999) The molecular evolution of terminal ear1, a regulatory gene in the genus Zea. Genetics 153:1455–1462

Yamasaki M, Wright SI, McMullen MD (2007) Genomic screening for artificial selection during domestication and improvement in maize. Ann Bot 100:967–973. doi:10.1093/aob/mcm173

Yang SD, Seo PJ, Yoon HK, Park CM (2011) The Arabidopsis NAC transcription factor VNI2 integrates abscisic acid signals into leaf senescence via the COR/RD genes. Plant Cell 23:2155–2168. doi:10.1105/tpc.111.084913

You J, Zong W, Li X et al (2013) The SNAC1-targeted gene OsSRO1c modulates stomatal closure and oxidative stress tolerance by regulating hydrogen peroxide in rice. J Exp Bot 64:569–583. doi:10.1093/jxb/ers349

Zhu T, Nevo E, Sun D, Peng J (2012) Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. Evolution (N Y) 66:1833–1848

Zimmermann R, Werr W (2005) Pattern formation in the monocot embryo as revealed by NAM and CUC3 orthologues from *Zea mays* L. Plant Mol Biol 58:669–685. doi:10.1007/s11103-005-7702-x