Theor Appl Genet (2012) 124:1521–1537
DOI 10.1007/s00122-012-1807-9

ORIGINAL PAPER

Maize genetic diversity and association mapping using transposable element insertion polymorphisms

Tatiana Zerjal · Agnès Rousselet · Corinne Mhiri · Valérie Combes · Delphine Madur · Marie-Angèle Grandbastien · Alain Charcosset · Maud I. Tenaillon

Received: 20 July 2011 / Accepted: 31 January 2012 / Published online: 18 February 2012

Abstract Transposable elements are the major component of the maize genome and presumably highly polymorphic yet they have not been used in population genetics and association analyses. Using the Transposon Display method, we isolated and converted into PCR-based markers 33 Miniature Inverted Repeat Transposable Elements (MITE) polymorphic insertions. These polymorphisms were genotyped on a population-based sample of 26 American landraces for a total of 322 plants. Genetic diversity was high and partitioned within and among landraces. The genetic groups identified using Bayesian clustering were in agreement with published data based on SNPs and SSRs, indicating that MITE polymorphisms reflect maize genetic history. To explore the contribution of MITEs to phenotypic variation, we undertook an association mapping approach in a panel of 367 maize lines phenotyped for 26 traits. We found a highly significant association between the marker *ZmV1-9*, on chromosome 1, and male flowering time. The variance explained by this association is consistent with a flowering delay of +123 degree-days. This MITE insertion is located at only 289 nucleotides from the 3′ end of a Cytochrome P450-like gene, a region that was never identified in previous association mapping or QTL surveys. Interestingly, we found (i) a non-synonymous mutation located in the exon 2 of the gene in strong linkage disequilibrium with the MITE polymorphism, and (ii) a perfect sequence homology between the MITE sequence and a maize siRNA that could therefore potentially interfere with the expression of the Cytochrome P450-like gene. Those two observations among others offer exciting perspectives to validate functionally the role of this region on phenotypic variation.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1807-9) contains supplementary material, which is available to authorized users.

T. Zerjal · M. I. Tenaillon
CNRS, UMR 0320/UMR 8120 Génétique Végétale,
Ferme Du Moulon, 91190 Gif sur Yvette, France

T. Zerjal (✉)
UMR1313 Génétique Animale et Biologie Intégrative
INRA-AgroParisTech, Domaine de Vilvert,
78352 Jouy en Josas, France
e-mail: tatiana.zerjal@jouy.inra.fr

A. Rousselet · V. Combes · D. Madur · A. Charcosset
INRA, UMR 0320/UMR 8120 Génétique Végétale,
Ferme Du Moulon, 91190 Gif sur Yvette, France

C. Mhiri · M.-A. Grandbastien
Institut Jean-Pierre Bourgin, UMR 1318 INRA-AgroParisTech,
78026 Versailles Cedex, France

Abbreviations

TE	Transposable element
TD	Transposon display
MITE	Miniature inverted repeat transposable element
siRNA	Small interfering RNA
MAF	Minor allele frequency

Introduction

Maize (*Zea mays* ssp. *mays*) was domesticated in the lowlands of southwest Mexico from the annual teosinte *Zea mays* ssp. *parviglumis* (van Heerwaarden et al. 2011; Matsuoka et al. 2002) around 9,000 years BP (Matsuoka et al. 2002; Piperno et al. 2009). Subsequent diffusion from its center of origin has been established by two broad microsatellite surveys (Matsuoka et al. 2002; Vigouroux et al. 2008) revealing a southward expansion through Central America and Columbian lowlands followed by a two-route expansion: (i) a spread through South America, along the Andes and the east coast of South America; and (ii) a northward expansion through northern Mexico and southwestern USA to northern USA and Canada (for a review, see Tenaillon and Charcosset 2011).

Drift accompanying maize migration has left only subtle footprints on the structure of maize diversity. For example, a genome-wide SNP survey of 1,127 landraces has revealed poor distinction among 10 geographic groups, with a principal component axis explaining only 4.8% of the variance (van Heerwaarden et al. 2011). Similarly a genome wide survey of 310 landraces based on microsatellite markers, which probably evolve faster than SNPs and thereby provide better resolution of recent history, has revealed that 44% of maize landraces share an ancestry in more than one of the four well-defined genetic clusters representing the Mexican highlands, Tropical lowlands, Andeans and Northern flints. Both surveys have relied on species samples—i.e., a single plant per accession—rather than a population-based sample. As a result, the level of intra-landrace diversity has not yet been characterized. However, intra-landrace diversity is presumably high given the extensive morphological and genetic variability at isozyme markers (Sanchez et al. 2000). Nevertheless, both SNP studies reveal a complex history of maize accompanied by recurrent germplasm exchange, admixture within and among subspecies (Matsuoka et al. 2002; Ross-Ibarra et al. 2009; van Heerwaarden et al. 2011) and potential long distance gene flow (Pressoir and Berthaud 2004).

Transposable elements (TEs) are the major component of the maize genome, comprising more than 85% of its sequence (Schnable et al. 2009). Although little is known about TE mutation rates, large scale sequence comparisons of allelic regions among modern maize inbred lines have revealed that genomic sequence differs from 25 up to 84%, due largely to differences in TE content (Brunner et al. 2005; Wang and Dooner 2006). These results have recently been confirmed by a comparative genomic hybridization assay between maize inbred lines B73 and Mo17, revealing an unprecedented amount of structural variation (Springer et al. 2009). These observations suggest that TE-based markers may be particularly relevant for describing the genetic diversity and studying the evolutionary factors driving its structuring.

In addition, TE polymorphisms near genes have been associated with adaptation and phenotypic variation in a variety of organisms. In *E. coli*, for example, Chao and McBroom (1985) have found that the insertion of a TE in the vicinity of the tetracycline-resistance determinant confers a fitness advantage. In *Drosophila melanogaster* repeated adaptive insertion of TEs in the 5′ end of the *Cyp6g1* gene leads to its over-transcription and increasing pesticide resistance (Schmidt et al. 2010). In *Arabidopsis thaliana* the insertion of a TE in the first intron of the *FLC* locus generates a weak allele that suppresses the late flowering phenotype of FRI by rendering the FLC locus subject to repressive chromatin modifications mediated by short interfering RNAs (Michaels et al. 2003; Liu et al. 2004). In maize, two elements from the *ZmV1* and the *Zead8* MITE families located in the *Vgt1* locus and in the *d8* gene respectively are associated with flowering time variation (Ducrocq et al. 2008; Salvi et al. 2007; Thornsberry et al. 2001). In theory, TE-based markers could therefore be used advantageously in genome wide association mapping studies.

TE-based markers were first developed in maize by Casa et al. (2000) using transposon display (TD). This technique is a modification of the AFLP technique (Vos et al. 1995) and permits the detection of many TE- based polymorphisms simultaneously. Casa et al. (2000) have developed TD using the *Heartbreaker* (*Hbr*) MITE family. *Hbr* family contains 3,000–4,000 copies that share a high sequence identity suggesting a recent spread of the family. In addition *Hbr* displays a strong preference of insertion towards genic regions (Casa et al. 2000). Similar properties were described recently for the *ZmV1* MITE family (Zerjal et al. 2009) that has likely undergone a recent amplification, as suggested by patterns of sequence diversity and the conservation of TIR (Terminal Inverted Repeat) sequences. *ZmV1* is also found preferentially near gene, i.e. 42.5% of the copies were found within 1 kb from an annotated gene in the reference genome (Zerjal et al. 2009).

In this study we have employed three MITE families (*Hbr*, *ZmV1* and *Ins2*) to assess, for the first time, the reliability of TD for studying maize genetic diversity. We converted TD information into PCR-based markers to genotype MITE insertion polymorphisms in a population sample of American landraces and in inbred lines. The purpose was, on one hand, to compare the genetic structuring revealed by these markers with microsatellites and SNPs data, and, on the other hand, to look for potentially adaptive insertions by performing an association study on a large association panel of inbred lines (Camus-Kulandaivelu et al. 2006). The genetic structuring pattern revealed that the largest fraction of genetic diversity was captured within landraces but a significant portion was captured among landraces. Our results were overall highly consistent with maize known genetic history. Finally, the association study revealed, on chromosome 1, a tandem-MITE insertion strongly associated with male flowering time, which does not map to any previously characterized QTL for maize flowering time (Chardon et al. 2004, Salvi et al. 2009).

Materials and methods

Plant material

Three panels of plant material were used in this study. A first panel of 26 maize inbred lines was used for the Transposon Display (TD) analysis and band extraction. These inbreds were half of American and half of European origin and were chosen both for having contrasting flowering time and for being representative of maize major genetic groups (Camus-Kulandaivelu et al. 2006). The second set of material consisted of 26 maize American landraces, for a total of 322 plants, analysed in the population genetic study (Table 1; Online Resource 1). These populations were selected to represent the largest fraction of genetic diversity among 275 American landraces, based on published SSR data (Camus-Kulandaivelu et al. 2006). Twenty of these landraces were from Central and North America and 6 from South America. Finally, for the association study we used a panel of 367 inbred lines representative of American, European, and Tropical maize, all phenotyped for 26 different characters including kernel quality traits and male flowering time (expressed in degree-days). A full description of this association panel is available in Camus-Kulandaivelu et al. (2006) and Manicacci et al. (2009). In the following, we will refer to the full landrace panel (26 landraces), the north landrace panel (a subset of 20 landraces from the full landrace panel) and the association panel.

Transposon display

Three MITE families were used for the TD analysis and band extraction. Two of them, *ZmV1* and *Hbr*, are Tourist-like MITEs and have already being described (Casa et al. 2000; Zerjal et al. 2009; Zhang et al. 2000). The third one, called *Ins2*, was originally identified as a small insertion into the bronze gene (Ralston et al. 1988) and corresponds to a 367 bp MITE, probably derived from a class II hAT DNA transposon. We used this MITE sequence to retrieve similar sequences from the maize genome and design the *Ins2* specific primer used in this study (see below).

For the Transposon Display assay we used a modified version of the one originally proposed by Waugh et al. (1997) following Zerjal et al. (2009). Genomic DNA (250 ng) was digested with *Msp*I and *Eco*RI enzymes and

Table 1 List of the 26 American maize landraces analyzed

Number ID	Population name	Origin[a]	Longitude[b]	Latitude[b]
404	Wakefields Rhode Island Flint	North America-NF	−72.567	41.9
405	Quicks Rhode Island Flint	North America-NF	−71.35	42.45
406	Grays Rhode Island Flint	North America-NF	−72.017	42.41
409	Canada Yellow Flint	North America-NF	−106	52
427	King Philip	North America-NF	−70.667	45
430	Sioux Tribe	North America-NF	−102.55	43.017
431	Gaspe Flint	North America-NF	−64.25	48.817
447	Minnesota 13	North America-CBD	−94.467	45.583
449	Gourdseed Dent	North America-SD	−96	37.5
453	Hickory King	North America-SD	−84.7	34
459	Tesuque Pueblo	North America-SW	−105.983	35.8
604	Conico	Mexico	−99.65	19.283
619	Apachito	Mexico	−110.067	27.8
717	Harinoso de Ocho	Mexico	−105.3	21.95
727	Chapalote	Mexico	−107.4	24.8
732	Zapalote Chico	Mexico	−101.15	18.383
627	Oloton	Guatemala	−90.517	14.633
666	Chandelle	Cuba	−76.25	20.9
694	Flint Cuba	Cuba	−83.733	22.417
810	Chandelle	Guadalupe	−61.383	16.333
637	San Jeronimo	Peru	−73.217	−13.35
640	Confite Morocho	Peru	−75.683	−12.45
650	Canguil	Ecuador	−78.25	−0.217
166	Choshuernco.Aesa	Chile	−72.017	−40.233
175	Cateto Amarillo	Argentina	−65.5	−26.033
179	Cuarento Cateto	Argentina	−57.95	−34.9

[a] North American types: *NF* Northern Flint, *CBD* Corn Belt Dent, *SD* Southern Dent, *SW* South West

[b] Geographical coordinates from Alain Charcosset personal maize database

ligated to the corresponding adaptor as previously described (Zerjal et al. 2009). For *ZmV1* and *Hbr* pre-amplification was performed using a primer complementary to the *Msp*I adapter and a first MITE specific primer as described in Casa et al. (2000) and Zerjal et al. (2009) (*Msp*I primer: 5′-GATGAGTCTAGAACGG-3′; *ZmV1*_int: 5′-CRATCC CRCTCAATCCAC-3′; *Hbr*Int5-E: 5′-GATTCTCCCCAC AGCCAGATTC-3′). For *Ins2* the pre-amplification was performed instead using a couple of primers complementary to the *Msp*I and the *Eco*RI adapters enriched, at the 3′ end, by an extra C and an extra A respectively (*Msp*I primer: 5′-GATGAGTCTAGAACGGC-3′; *Eco*RI 5′-GACT GCTACCAATTCA-3′). PCR conditions were as described in Zerjal et al. (2009) with annealing temperature at 56°C for *Hbr*, at 58°C for *ZmV1* and at 60°C for *Ins2*.

Selective amplification was performed using the non-labelled *Msp*I primer and the corresponding ^{33}P-labelled MITE-specific primer (*ZmV1*_est 5′-TCCACATGGA TTGAGAGCTAA-3′, *HbrInt5*-F 5′-GAGCCAGATTTTCA GAAAAGCTG-3′ and *Ins2* 5′-CCCGTTTAGCACGAAA AA-3′). A first assay of selective amplification was performed on a restricted sample of 10 individuals using MspI adapter primers enriched with one selective base (A, C, G, T) for *ZmV1* and *Hbr* as described in Casa et al. (2000). For *Ins2*, we tested 4 combinations of 3 selective bases (CGG, CTA, CGC, CGT). For each TE, the selective base combination that produced the clearest pattern of bands was chosen for screening the full data set (A for *ZmV1* and *Hbr* and CGG for Ins2). Three microlitres of amplified products were separated on 6% denaturing (7.5 M urea) acrylamide-bisacrylamide (19:1) gels in 1× TBE buffer and exposed, after drying, to an X-ray film for 24 h.

Recovery of fragments from TD gels and conversion into PCR markers

One hundred and sixty-six DNA polymorphic fragments were excised from TD radioactive gels eluted in 50 µl water and left at 37°C overnight followed by 5 min at 95°C. Fragments were re-amplified in 20 µl of volume, using 5 µl of eluted DNA, using non-labelled MITE specific-primers and the same PCR conditions described above for the selective amplification. Reactions were resolved in 1.5% agarose gels, and cloned (pGEM®-T vector, Promega and DH5α™ competent cells, Invitrogen) using between 0.5 and 1 µl of PCR product. Cloning performance was tested by PCR, checking 10 different colonies from each cloning reaction, using pUC/M13 forward and reverse primers and standard PCR conditions. The PCR products obtained from three independent colonies were sent for sequencing (Genoscreen). Sequences were then aligned (ClustalW2, EMBL-EBI) to verify the level of identity among them and were blasted against the high-throughput genomic sequence (HTGS) database to identify maize chromosomal regions of homology. We discarded all sequences that corresponded to *Hbr*, *ZmV1* or Ins2 insertions into repetitive portions of the genome. We transformed other sequences into dominant PCR markers by choosing one primer in the sequenced region flanking the TE insertion and one inside the transposable element. Primers were designed using Primer3 (Rozen and Skaletsky 2000). Sequences for which we were able to retrieve both the upstream and the downstream sequence flanking the MITE insertion were transformed into co-dominant PCR markers. The list of PCR primers is available in Online Resource 2. Each PCR system (dominant and co-dominant) was tested using standard PCR conditions on 4 out of the 26 inbred lines used for the TD, chosen for being polymorphic for the corresponding TD fragment.

Genetic diversity, population differentiation and Hardy–Weinberg equilibrium

All population summary statistics were calculated using the program ARLEQUIN 3.11 (Excoffier et al. 2005). We estimated allele frequencies and population observed (Ho) and non-biased expected heterozygosity (He) averaged across loci and tested for Hardy–Weinberg equilibrium (HWE) at each locus within each population using an analog to Fisher's exact test (Guo and Thompson 1992). We also calculated the fixation index (F_{IS}) across loci following Weir and Cockerham (1984) and tested for a significant deficit or excess of heterozygotes (when compared with HWE expectations) based on 20,000 permutations. The same software was also used to perform hierarchical analysis of molecular variance (AMOVA) and pairwise-F_{ST} (Slatkin 1995). Locus by locus analysis was employed to estimate marker heterozygosity and F_{ST}.

Multilocus test of neutrality

A coalescent simulation-based method developed by Beaumont and Nichols (1996) and implemented in the FDIST2 software (http://www.rubic.rdg.ac.uk/;mab/software.html (Beaumont and Balding 2004)) was applied to test for potential signatures of selection at MITE-derived markers in the full landrace panel. For comparison, we also used published data for 15 SSR markers and 25 RFLP markers available for the same populations (but different plants) (Camus-Kulandaivelu et al. 2006; Rebourg et al. 2003). The FDIST2 method was based on a symmetrical island model of population structure and generated data sets with mean F_{ST} similar to the empirical distribution. For each locus, allele frequencies were used to compute F_{ST} values, conditional on heterozygosity, which were compared with the distribution of simulated F_{ST} values (based on 20,000 simulated loci) to identify putative outliers deviating from the neutral expectations. Sample sizes were set to 24 alleles per population in all simulations, i.e. 12 individuals per population. We assumed an infinite allele mutation model for the RFLP and for the TE-based markers and a stepwise mutation model for the SSRs.

Structure analyses of genetic variation

Spatial structuring of genetic variation was investigated considering population geographic locations and individual multilocus genotypes. The six southern American landraces were excluded from the analysis because we had too few to make inferences on their dispersion. In total, 20 landraces of central and North America consisting of 249 plants were used. The contribution of gene flow and drift in determining genetic structure, were tested by plotting linearly transformed population pairwise F_{ST} values, $F_{ST}/(1 - F_{ST})$, against pairwise geographical distances (Hutchison and Templeton 1999; Rousset 1997; Slatkin 1993). Log transformed genetic distances were correlated with Euclidian geographic distances. The strength and significance of the relationship were evaluated using a reduced major axis regression analysis and a Mantel test (30,000 randomizations) as implemented in the program IBDWS program 3.16 (http://ibdws.sdsu.edu/~ibdws/ (Jensen et al. 2005)). In order to investigate spatial structuring of genetic variation at interval distances, we used spatial autocorrelation analysis of multi-alleles using Smouse and Peakall's (1999) method as implemented in GenAlEx version 6 (Peakall and Smouse 2006). This method differs from classical spatial autocorrelation analysis in that it employs a multivariate approach to simultaneously assess the spatial signal generated by multiple loci and alleles, and generates an autocorrelation coefficient '*r*', related to Moran's I, that ranges from −1 to +1 (Peakall et al. 2003). The number and width of distance classes were chosen to obtain an accurate spatial resolution without compromising the number of data points per interval. Geographical distances between plants from the same populations were set to zero. This resulted in the choice of 6 distance classes. Tests for statistical significance were performed by 1,000 permutations and the same number of bootstraps was used to estimate the 95% confidence interval (CI). A significant departure from the null hypothesis of no spatial genetic structure ($r = 0$) is obtained when a positive *r* value falls outside the 95% CI.

The genetic structure of populations was assessed using the Bayesian clustering method implemented in the program STRUCTURE (v2.1) (Pritchard et al. 2000). A burn-in period of 100,000 interactions followed by 500,000 MCMC repetitions and a model with admixture and correlated allele frequencies was applied. Five replicates were performed for each cluster K, from $K = 2$ to $K = 11$. STRUCTURE outputs were processed using the Greedy algorithm of CLUMPP (http://rosenberglab.bioinformatics.med.umich.edu/clumpp.html (Jakobsson and Rosenberg 2007)) and for each pair of runs with a given K, we used the function G as a global measurement of similarity among replicates. Graphical representations of population structure were produced using DISTRUCT (http://rosenberglab.bioinformatics.med.umich.edu/distruct.html (Rosenberg 2004)).

The phylogenetic network showing the relationships of pairwise F_{ST} distances were constructed with SplitsTree version 4.6 (http://www.splitstree.org/ (Huson and Bryant 2006)) using the Neighbor-Net algorithm (Bryant and Moulton 2004).

Association study

The tests of associations between 32 TE-based markers and 26 phenotypic traits measured on 367 inbred lines (association panel) were computed using the software package TASSEL (Bradbury et al. 2007). Two models were used: a General Linear Model (GLM) using the percentages of admixture of each accession (Q matrix) as covariates to take population structure into account and a Mixed Linear Model (MLM) using both population structure and the kinship coefficients as covariates ($Q + K$). The population Structure (Q matrix) and Ritland's kinship coefficients, K (Ritland 1996), for the 367 inbred lines were estimated from 55 genome-wide microsatellites (SSR) as described in Camus-Kulandaivelu et al. (2006) and in Manicacci et al. (2009).

The GLM model was run with 10,000 permutations and a site-wise P value was estimated for each TD marker. For the MLM model, the P values were adjusted for multiple testing using the false discovery rate (FDR) (Benjamini and Hochberg 1995) implemented in the R software package QVALUE (Storey 2003). The Mixed Linear Model (MLM) is the most appropriate model for phenotypes for which population structure explains more than 10% of the phenotypic variance (Manicacci et al. 2009; Yu et al. 2006).

To calculate the among-group marker phenotypes divergence, we estimated the group specific average phenotype value for each allele by summing the individual phenotypic values, weighted by the individual group membership and normalized by the group individual equivalent number (obtained by adding the group individual Q values). The same procedure was applied to calculate the group allele frequency. We estimated the effects of the genotype, Q and the (genotype × Q) interaction on male flowering time through ANOVA using the lm(stats) function in the R package (http://www.r-project.org/).

Sequence analysis of the ZmV1-9 insertion site

The proximal and distal regions of the *ZmV1-9* insertion site were amplified by PCR on a sample of 43 inbred lines from the association panel with the primer sets *ZmV1*-9F2-*ZmV1*-9R2 (5′CCACATTAGGGAGCAGGA3′, 5′TTAAC TTTCCACCCCTGCTG3′ respectively) and *ZmV1*-9L-*ZmV1*Rb (5′ GGAAGGGACCCCAGGTACT3′, 5′ CGGT TCCTTATATTTCAGGACC 3′ respectively). Reactions were performed in 25 μl containing 1× PCR buffer, 50 ng DNA, 0.4 μM of each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 1 unit of Taq DNA polymerase. Cycling parameters were: 94°C/2 min followed by 33 cycles of 94°C/30 s, 58°C/1 min, 72°C/1 min (1.30 min for the *ZmV1*-9F2- *ZmV1*-9R2 primer set) and a final cycle of 72°C/3 min. PCR products were of 1,100 bp for the *ZmV1*-9F2- *ZmV1*-9R2 primer set and between 500 and 800 bp for the *ZmV1*-9L- *ZmV1*Rb depending on the presence or absence of the MITE insertion. PCR products were sequenced on both strands with the above-mentioned primers and results were aligned and manually verified using the program Bioedit (Hall 1999). Nei's measure (Nei 1987) of pairwise nucleotide diversity, π, and neutrality tests were calculated using the program DnaSP version 3.51 (Rozas and Rozas 1999). To compare sequence diversity among the different *ZmV1-9* alleles we constructed a median-joining (MJ) network using the network algorithm implemented in the program Network 4.6.0.0 (www.fluxus-engineering.com/sharenet.htm (Bandelt et al. 1999)).We used the software package TASSEL (Bradbury et al. 2007) to estimate pairwise linkage disequilibrium (r^2, Weir 1996) between all pairs of polymorphic sites of allele-1 and allele-2 sequences and performed Fisher's exact tests to test its significance.

Results

Transposon display and band sequencing

We employed the TD technique to screen 26 maize inbred lines to verify its reliability and to isolate genomic fragments containing MITE insertions. Those fragments were converted into single locus PCR markers. Although we obtained profiles for *Hbr* that resemble those produced by Casa et al. (2000) and overall clear TD profiles for *ZmV1* and *Ins2* (Online Resource 3), band extraction of TD gels and conversion into PCR markers was a process only partly successful. Out of the 166 bands cut from the gels, only 124 were successfully re-amplified. Of these, 106 were cloned and sequence information was obtained for 102. Out of 102 sequences: 30 MITE insertions (30%) were non-allelic and corresponded to MITE insertions into different genomic regions, i.e. cloning and sequencing of several bands migrating at the same level size in the TD gel revealed multiple sequences extremely close in size but different in nucleotide composition. Sixteen additional MITE insertions (15%) were allelic but corresponded to bands migrating at different size in the TD gel, due to small indels in the region flanking the MITE. Overall, almost half of the bands that were successfully sequenced either were non-allelic but of similar size or were allelic but migrated at different positions.

All 102 sequences were blasted to identify their chromosomal location and to select those inserted into unique regions. We designed primers to convert 84 MITE insertion polymorphisms into both dominant and co-dominant PCR markers. Thirty-three markers (25 co-dominant and 8 dominant), 2, 11, and 20 for the Ins2, the *Hbr*, and the *ZmV1* families, respectively, were retained for this study because of their high polymorphism and their low percentage of missing data among maize landraces (missing data <10%). All 33 were used to genotype the 367 plants from the association panel but only the co-dominant markers (25) were used to screen the 26 American landraces. The majority of markers had a classical biallelic pattern: presence or absence of the TE. However, for three markers, *ZmV1*-14, *ZmV1*-16 and, *ZmV1*-9, we identified three possible alleles due to the presence of a tandem insertion of two almost identical MITE copies. For these markers in the landraces we scored only the most frequent MITE insertion allele, the tandem insertion for *ZmV1*-14 and the single MITE insertion for *ZmV1*-16 and *ZmV1*-9. In the association panel instead we scored the three alleles.

Differentiation and structure of landraces revealed by TE-based markers

We genotyped 26 American landraces (the full landrace panel) for a total of 322 plants (Table 1) using 25 co-dominant TE-based markers. Two of the markers did not amplify in two populations and were therefore discarded from further analyses. For each marker the number of landraces displaying polymorphism was variable (Table 2), ranging from 5 (*ZmV1*-4) to 22 (*ZmV1*-9). Observed and expected population-heterozygosities averaged across all markers, ranged from 0.07 to 0.24 and from 0.12 to 0.30, respectively (Table 3). F_{IS} values were comprised between −0.21 and 0.4 (Table 3). A departure from Hardy–Weinberg equilibrium, after Bonferroni correction, was found in 9 populations for one marker, and in 1 population for 2 markers (Table 3). In all cases, the H–W departure was due to an excess of homozygosity, in line with results obtained from classical markers (Dubreuil and Charcosset 1998). F_{ST} values were rather high, with average values of 0.3 and 0.4 respectively (Table 2). To test whether any TE-based

Table 2 TE-based marker frequencies and properties

Landraces	TE-based markers[a]											
	Hbr-1	Hbr-2	Hbr-3	Hbr-4	Hbr-5	Hbr-6	Ins-1	ZmV1-1	ZmV1-2	ZmV1-3	ZmV1-4	ZmV1-5
166	0.92	0.00	0.79	0.96	0.04	0.75	0.50	1.00	0.00	0.29	0.00	0.25
175	0.42	0.83	0.42	0.50	0.00	0.08	0.29	1.00	0.08	0.17	0.00	0.67
179	0.42	0.25	0.00	0.94	0.00	0.17	0.71	1.00	0.00	0.21	0.33	0.00
404	0.15	0.19	0.40	0.92	0.00	0.15	0.96	1.00	0.19	0.00	0.00	0.73
405	1.00	0.64	1.00	0.55	0.32	0.18	0.00	1.00	0.55	0.36	0.00	0.09
406	1.00	0.00	1.00	1.00	0.63	0.04	0.08	1.00	0.67	0.08	0.00	0.00
409	0.96	0.04	1.00	1.00	0.71	0.00	0.00	1.00	0.07	0.68	0.04	0.11
427	1.00	0.04	0.54	1.00	0.04	0.00	0.19	1.00	0.58	0.65	0.00	0.00
430	0.67	0.00	0.75	0.33	0.42	0.00	0.71	0.83	0.11	0.54	0.00	0.00
431	1.00	0.00	0.96	0.88	0.92	0.00	0.77	1.00	0.46	0.88	0.00	0.23
447	0.75	0.08	0.46	0.50	0.38	0.75	0.58	1.00	0.00	0.00	0.00	0.04
449	0.00	0.00	0.25	0.08	0.54	0.08	0.08	1.00	0.00	0.00	0.00	0.50
453	0.08	0.17	0.17	0.00	0.25	0.17	0.58	1.00	0.17	0.00	0.00	0.00
459	0.88	0.19	0.04	0.54	0.00	0.42	0.92	1.00	0.35	0.35	0.00	0.00
604	0.81	0.00	0.00	0.62	0.00	0.00	0.58	0.73	0.21	0.00	0.00	0.00
619	0.96	0.04	0.38	0.63	0.00	0.27	0.63	0.83	0.17	0.00	0.00	0.04
627	0.33	0.83	0.38	1.00	0.00	0.00	0.13	0.83	0.00	0.00	0.00	0.00
637	0.54	0.00	0.88	0.33	0.04	0.21	0.17	0.96	0.04	0.17	0.00	0.88
640	0.50	0.00	0.71	0.71	0.00	0.42	0.17	0.88	0.00	0.00	0.00	0.83
650	0.08	0.00	0.71	0.58	0.00	0.00	0.33	0.42	0.00	0.00	0.00	0.00
666	0.00	0.85	0.27	0.96	0.00	0.00	1.00	0.35	0.00	0.00	0.00	0.08
694	0.23	0.35	0.00	0.86	0.00	0.15	1.00	0.42	0.00	0.00	0.00	0.00
717	0.79	0.00	0.54	0.83	0.00	0.50	0.88	1.00	0.29	0.00	0.63	0.00
727	0.42	0.00	0.31	0.96	0.00	0.00	0.46	0.92	0.00	0.00	0.69	0.42
732	1.00	0.04	0.00	0.95	0.00	0.88	0.75	1.00	0.00	0.00	0.00	0.00
810	0.42	0.65	0.23	1.00	0.00	0.62	0.92	0.62	0.00	0.00	0.77	0.73
# polymorphic populations[d]	19	15	19	20	11	17	22	11	14	11	5	14
He[d]	0.48	0.32	0.50	0.41	0.28	0.35	0.50	0.22	0.26	0.29	0.18	0.34
F_{ST}^{d}	0.49	0.52	0.44	0.39	0.51	0.37	0.43	0.34	0.28	0.45	0.60	0.51

Table 2 continued

Landraces	TE-based markers[a]												
	ZmV1-6	ZmV1-7	ZmV1-8[b]	ZmV1-9	ZmV1-10	ZmV1-11	ZmV1-12	ZmV1-13	ZmV1-14	ZmV1-15	ZmV1-16	ZmV1-17[c]	ZmV1-18[c]
166	0.21	1.00	0.00	0.17	0.38	0.71	0.83	0.29	0.25	0.13	0.04	1.00	1.00
175	0.00	1.00	0.04	0.46	0.17	0.00	1.00	0.08	0.10	0.00	0.13	1.00	1.00
179	0.54	0.88	0.00	0.38	0.00	0.25	1.00	0.00	0.00	0.50	0.00	1.00	1.00
404	0.62	0.58	0.00	0.27	0.50	0.12	0.38	0.04	0.00	0.00	0.00	1.00	1.00
405	0.00	1.00	0.00	0.95	0.27	1.00	1.00	1.00	0.45	0.00	0.00	1.00	1.00
406	0.00	1.00	0.00	0.63	0.54	1.00	1.00	0.21	0.00	0.00	0.00	1.00	1.00
409	0.61	1.00	0.04	0.00	1.00	1.00	0.46	0.39	0.93	0.61	0.00	0.68	1.00
427	0.04	1.00	0.00	0.31	0.04	0.31	0.85	0.42	0.69	0.23	0.00	0.77	1.00
430	0.00	1.00	0.00	0.50	0.33	0.71	0.71	0.83	0.08	0.54	0.00	1.00	0.79
431	0.00	1.00	0.00	0.50	0.38	0.88	0.85	0.46	0.96	0.00	0.00	0.92	1.00
447	0.08	1.00	0.00	0.17	0.42	0.33	0.58	0.63	0.75	0.42	0.63	1.00	1.00
449	0.67	1.00	0.38	0.25	0.08	0.33	0.33	0.33	0.71	0.08	0.00	0.63	1.00
453	0.50	1.00	0.17	1.00	0.05	0.46	0.83	0.08	0.29	0.17	0.38	1.00	1.00
459	0.00	0.85	0.00	0.73	0.00	0.73	0.15	0.15	0.23	0.00	0.00	0.62	0.42
604	0.38	1.00	0.00	0.12	0.00	0.46	0.54	0.15	0.12	0.19	0.08	0.42	X
619	0.38	0.83	0.00	0.42	0.38	0.75	1.00	0.38	0.00	0.00	0.00	0.29	0.96
627	0.79	0.42	0.17	0.83	0.55	0.54	0.46	0.00	0.00	0.00	0.21	X	1.00
637	0.50	1.00	0.00	0.00	0.04	0.08	0.96	0.04	0.50	0.00	0.29	1.00	1.00
640	0.21	1.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.04	0.00	1.00	1.00
650	1.00	1.00	0.00	0.63	0.00	0.71	1.00	0.00	0.29	0.17	0.33	1.00	1.00
666	0.81	1.00	0.15	0.55	0.00	0.00	1.00	0.00	0.25	0.00	0.15	1.00	1.00
694	0.96	1.00	0.08	0.38	0.04	0.00	0.88	0.00	0.00	0.00	0.19	1.00	1.00
717	0.00	0.75	0.00	0.00	0.00	0.04	0.83	0.00	0.00	0.00	0.00	0.67	1.00
727	0.12	1.00	0.00	0.79	0.00	0.81	0.09	0.46	0.73	0.08	0.00	0.00	1.00
732	0.04	0.21	0.00	0.58	0.00	0.00	0.63	0.50	0.00	0.00	0.42	1.00	1.00
810	0.96	1.00	0.00	0.23	0.00	0.00	0.81	0.00	0.00	0.00	0.54	1.00	1.00
# polymorphic populations[d]	18	7	7	22	15	17	18	17	16	12	12	/	/
He[d]	0.47	0.17	0.07	0.49	0.32	0.49	0.43	0.37	0.42	0.21	0.22	/	/
F_{ST}^{d}	0.51	0.44	0.17	0.29	0.40	0.51	0.40	0.37	0.49	0.30	0.28	/	/

[a] For each locus, the frequency of the allele with the MITE insert is given

[b] Excluded from the association study, i.e. MAF <5%

[c] Excluded from the population genetic analyses, i.e. 100% missing data in one population

[d] Number of polymorphic populations, expected heterozygosity and F_{ST} estimated for each TE-based marker

Table 3 Diversity and H–W equilibrium at TE-based markers

Population #	Polymorphic loci	F_{IS}	Ho	He	HW[a]	Marker
404	17	−0.01	0.23	0.23	0	
405	10	0.01	0.17	0.17	0	
406	8	0.14	0.11	0.12	1	*ZmV1-2*
409	13	−0.06	0.17	0.16	0	
427	14	0.00	0.20	0.20	1	*ZmV1-15*
430	14	0.05	0.24	0.26	0	
431	13	0.22	0.07	0.17	1	*ZmV1-13*
447	17	0.23	0.24	0.30	2	*Hbr-4,* *Hbr-6*
449	15	0.03	0.23	0.23	1	*Hbr-3*
453	16	0.18	0.18	0.23	0	
459	14	0.05	0.20	0.21	0	
604	13	−0.01	0.19	0.21	0	
619	15	0.23	0.18	0.23	0	
717	10	0.07	0.14	0.16	1	*Hbr-6*
727	13	0.09	0.18	0.21	0	
732	10	0.07	0.13	0.14	0	
627	13	0.14	0.19	0.22	0	
666	9	0.28	0.10	0.14	1	*ZmV1-1*
694	10	0.40	0.08	0.14	1	*ZmV1-9*
810	12	0.07	0.18	0.20	1	*ZmV1-9*
637	17	0.20	0.18	0.20	0	
640	10	0.16	0.13	0.16	1	*Hbr-6*
650	10	−0.21	0.22	0.18	0	
166	14	0.19	0.17	0.22	0	
175	14	0.09	0.19	0.21	0	
179	11	0.07	0.19	0.20	0	

[a] Number of loci per population deviating from H–W equilibrium after Bonferroni correction

F_{IS} Fixation Index, *Ho* Averaged observed heterozygosity, *He* Averaged expected heterozygosity

Marker name of the markers for which a deviation from H–W equilibrium was observed

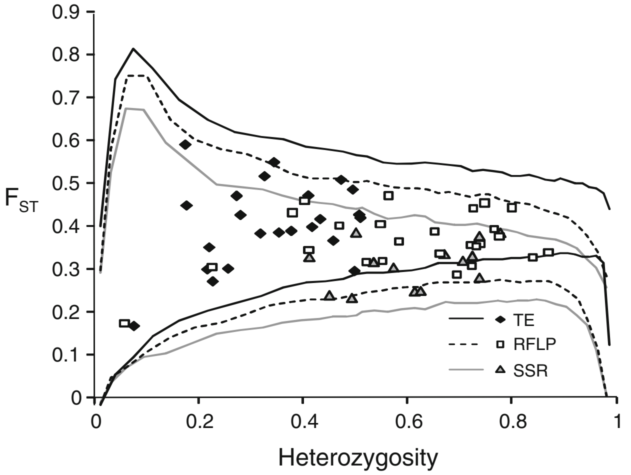

Fig. 1 Distribution of observed F_{ST} values for 23 TE-based markers, 25 RFLPs and 14 SSRs, as a function of their mean heterozygosity across 26 maize landraces (*full landrace panel*). The 5 and 95% quantiles of the simulated neutral envelopes are reported for each type of marker and are represented by *black lines* for TEs, *dotted lined* for the RFLPs and *gray lines* for the SSRs

markers had F_{ST} values larger than what expected under neutrality, we used the coalescent simulation-based method implemented in the FDIST2 software, which has already been used to identify outlier markers in other species (Eveno et al. 2008; Moen et al. 2008; Pariset et al. 2009). Because none of the markers exhibited F_{ST} values outside the neutral envelope, we considered the TE-based markers as neutral in all further population genetic analyses.

We compared the structuring revealed by TE-based markers with published results obtained by SSR and RFLP markers (Camus-Kulandaivelu et al. 2006; Rebourg et al. 2003) available for the same set of landraces (but different samples). The results are summarised in Fig. 1. High F_{ST} values were found for the SSR (0.31) and the RFLP markers (0.36) supporting the high level of population differentiation identified by the TE-based markers (0.4). Indeed, significant pairwise F_{ST} were obtained for all population comparisons pointing to high differentiation among populations. The analysis of molecular variance, however, revealed that the largest fraction of variation was captured within landraces, with 57% of variation among plants within landraces, and 42% among landraces.

The genetic distance among landraces correlated positively with geographical distance ($r = 0.407$, $P < 0.001$) (Fig. 2a) in the North landrace panel (20 landraces). Genetic correlations between individuals were high and significantly greater than zero at distance smaller than 3,000 km (Fig. 2b). Beyond this distance, autocorrelation were either zero or negative. The shape of the correlogram was consistent with a clinal pattern of TE allele frequencies in central and North American landraces. However, the ~5-fold reduction of '*r*' within 500 km supported the evidence that most of the differentiation occurred at short distances.

The program STRUCTURE uses the Bayesian clustering method to infer the most likely number of clusters (*K*) of individuals by maximizing H–W and linkage equilibrium at each *K* (Pritchard et al. 2000). It has previously been applied to maize landraces and inbred lines using SSR data sets (Camus-Kulandaivelu et al. 2006; Vigouroux et al. 2008) and SNPs (van Heerwaarden et al. 2011) but was never been tested on maize TE insertion polymorphisms. We analyzed the genetic structure of the full landrace panel (322 plants) using the 23 MITE insertion polymorphisms applying an increasing value of *K* (from 2 to 11) with five replicates for each *K*.

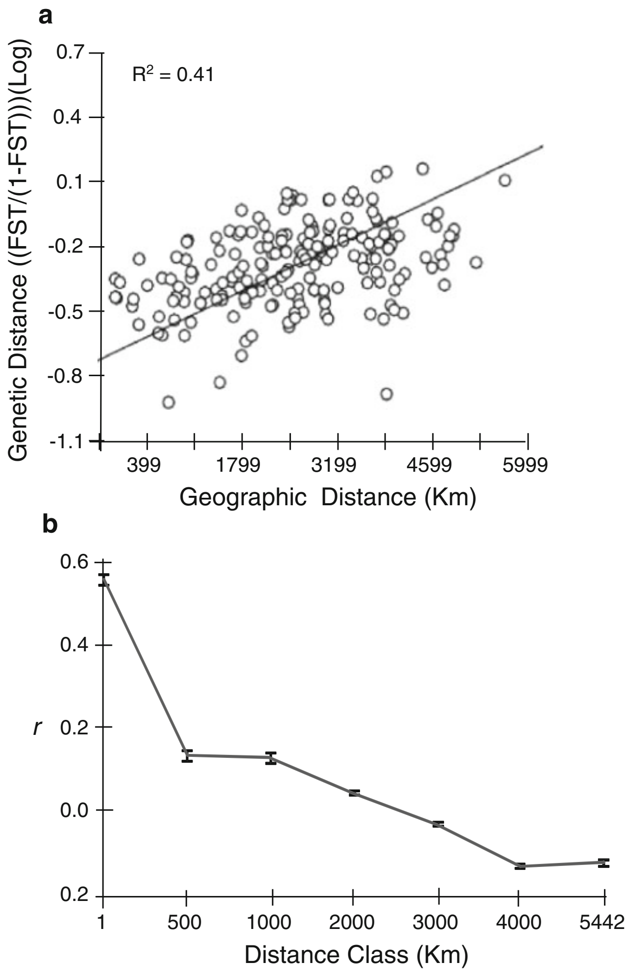

Fig. 2 Isolation by distance in the north landrace panel. **a** Correlation between genetic distances inferred from TE-based markers and geographical distances. **b** *Correlogram* showing the spatial autocorrelation analysis where genetic correlation (*r*) is plotted as a function of distance class

Interpreting the plot of the log-likelihood values at increasing *K* was not straightforward because there was no optimal *K* value: log-likelihood values increased steadily with increasing *K* values (Online Resource 4). The mean similarity coefficient *G* as obtained by CLUMPP was however comprised between 0.99 and 0.98 from $K = 2$ to $K = 5$ revealing a high consistency among replicates but dropped to 0.5 for the other *K* values tested ($K > 5$). We therefore chose to compare STRUCTURE output for several *K* values (2, 4 and 5) based on *K* values employed in two previous analyses (Camus-Kulandaivelu et al. 2006; Vigouroux et al. 2008) and geographical concordance of individuals' ancestry.

At $K = 2$ the full landrace panel separated the landraces from the Caribbean Islands (referred thereafter as Tropical) from all other landraces (Fig. 3a). A number of landraces had average membership coefficients in both clusters (graphically displayed as vertical lines broken in *K* colored segments). This result supported evidence of an allele frequency gradient, with the Northern Flint and the Tropical landraces representing the two allelic extremes in our data set. At $K = 4$ the admixed landraces from $K = 2$ formed two new genetic groups characterized mainly by Mexican landraces on one hand and Andean-Southern Dent landraces on the other hand. This last group was further split at $K = 5$ separating the Southern Dent and one Andean landrace from the previous Andean-Southern Dent landraces.

Interestingly, the Tropical landraces constituted a robust cluster that did not split throughout increasing *K* values, while the Northern Flint cluster harbored several landraces with admixed origins including the Canadian Northern Flint that separated from the US Northern Flint at $K > 6$ (data not shown). The known admixed origin of some landraces, such as the Corn Belt Dents and the South West landraces emerged early in the analysis. The Mexican cluster was overall less supported than the others, harboring a high level of admixture. The intermediate position of the Mexican landraces also emerged in the NeighborNet tree based on pairwise F_{ST} (Fig. 3b). While the two extremes of the tree were occupied by the Northern Flints, and the South American and Tropical groups, respectively, Mexican landraces lay between the two.

Association mapping

In order to test the association between 33 MITE polymorphisms and variation of phenotypic traits, we undertook an association analysis using an association panel of 367 inbred lines characterized for 26 phenotypes, including kernel quality traits and male flowering time. STRUCTURE analysis on this material was performed by Camus-Kulandaivelu et al. (2006) and identified 5 genetic groups: Tropical, Northern Flint, European Flint, Corn Belt Dent and Stiff Stalk. In the association study, the General Linear Model (GLM) that controls for population structure (*Q* model) and the Mixed Linear Model (MLM) that controls for both population structure and kinship ($Q + K$ model) yielded similar results. Both models identified a highly significant association between male flowering time and the MITE insertion *ZmV1*-9 (Table 4). The *ZmV1*-9 is a marker with 3 allelic states called thereafter allele-0, allele-1 and allele-2 according to the number of MITE insertion present at the *ZmV1*-9 locus (0, 1 or 2 insertions, respectively, see below).

Alleles 0 and 1 were frequent in the 26 landraces described previously (Table 2) and in the association panel (Table 4). Allele-2 was instead very rare among the landraces (2.7%) but it was more common in the association panel (14.5%), exhibiting highly significant differences in

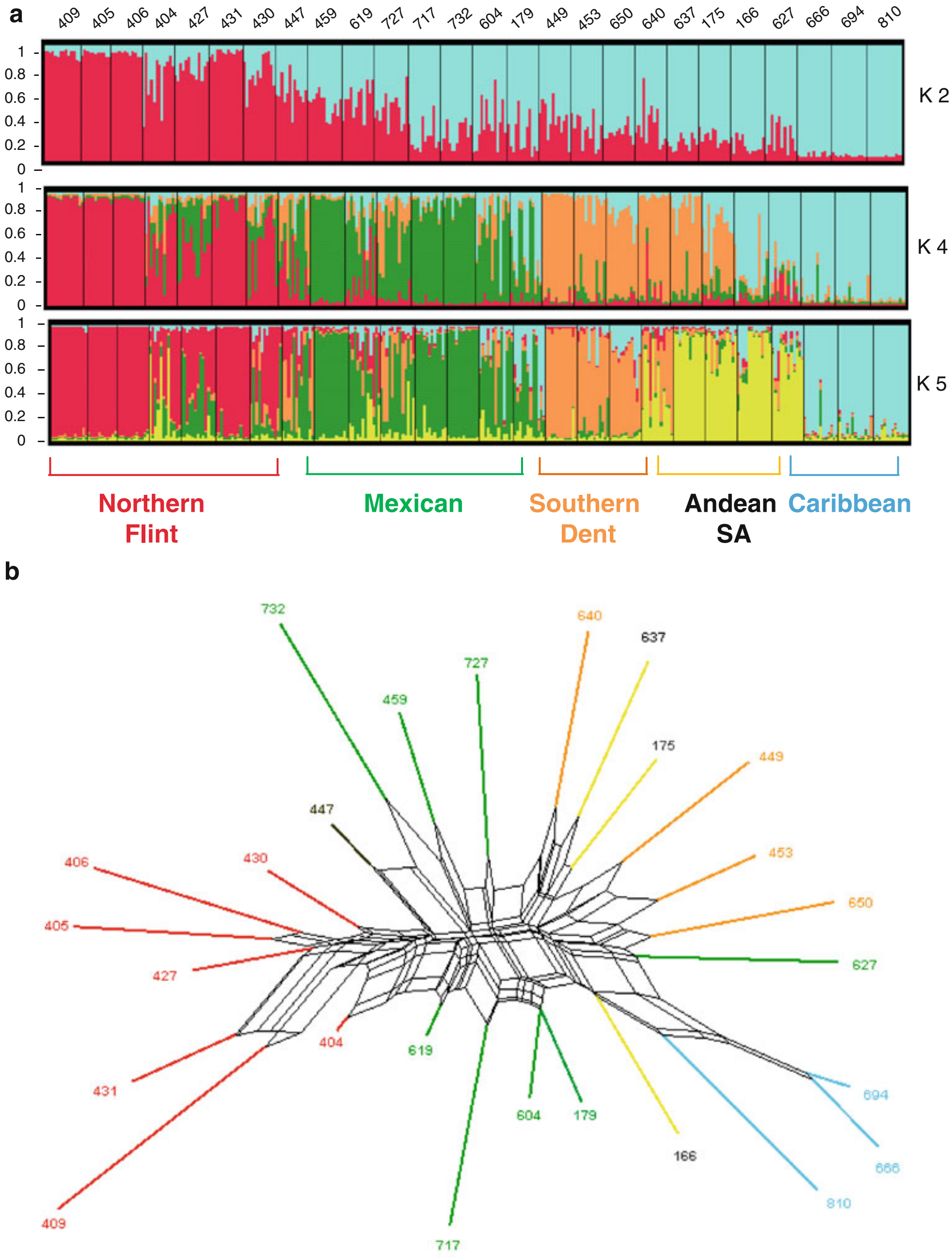

Fig. 3 Population structure analysis of full landrace panel using 23 MITE insertion polymorphisms. **a** Posterior probability assignment to each genetic cluster is represented by proportional membership of each individual assuming 2, 4 and 5 clusters (*from top to bottom*). **b** Neighbour-Net based on pairwise F_{ST} genetic distances considering the 5 genetic groups ($K = 5$) identified in the STRUCTURE analysis

allele frequency among the five genetic groups in the panel (χ^2 test, $P < 0.0001$). Among inbred lines, Corn Belt dents and Tropical inbreds had the highest allele-2 frequency (22 and 15%, respectively), while it was rather rare in the Northern and European Flints (4%). We estimated the fraction of total variation in flowering time that is explained by variation at the *ZmV1-9* marker (R^2 marker), after correcting for population structure, to 4% (Table 4). The allelic effect of allele 2 estimated after accounting for population structure was of 123 degree-days, which increased slightly when no correction for *Q* was applied (128 degree-days). No allelic effect was observed for allele 0 and 1. As shown in Fig. 4, for each genetic group plants carrying the allele 0 or 1 had a very similar male flowering time. Qualitatively, the effect of allele-2 on flowering time was consistent among groups (Fig. 4), as confirmed by the fact that the effect of interaction (genotype × *Q*) was not significant ($P = 0.79$).

Sequence analysis of the *ZmV1*-9 insertion site

The MITE insertion polymorphism *ZmV1*-9 resides on chromosome 1 (bin 1.04 between the markers bnlg1811 and csu3 based on the IBM genetic map (www.maizegdb.org)) at

Table 4 *ZmV1-9* allele frequencies and significant association with male flowering time

Marker	Allele frequency	Trait	P (Q model)[a]	R^2 marker[b] (%)	R^2 model[c] (%)	P ($Q + K$ model)[d]	FDR Q value[e]
ZmV1-9	0 = 0.55 1 = 0.29 2 = 0.14	Male flowering time	$<10^{-5}$	0.04	0.55	$<10^{-6}$	0.0044

[a] P value for the Q model estimated after 100,000 permutations

[b] R^2 of the marker

[c] R^2 of the model

[d] P value of the $Q + K$ model

[e] FDR Q value based on the P value ($Q + K$ model)

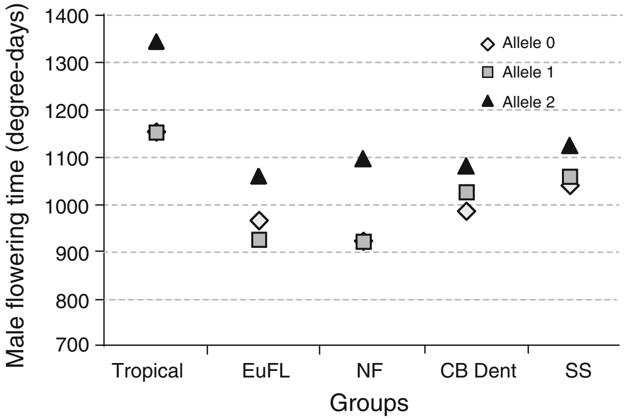

Fig. 4 Group specific allelic effect on male flowering time in the association panel. The alleles 0, 1 and 2 correspond, respectively, to 0, 1 and 2 MITE copies of the *ZmV1-9* element. Groups are defined according to Camus-Kulandaivelu et al. (2006) with *EuFl* European Flint, *NF* Northern Flint, *CB Dent* Corn Belt Dents and *SS* Stiff stalks

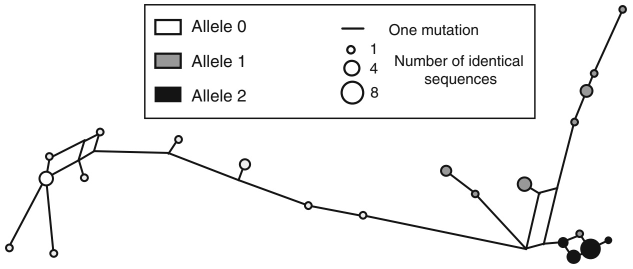

Fig. 5 Network analysis of 43 inbred lines from the association panel chosen to represent one of the three *ZmV1-9* alleles: 0, 1 or 2. MITE sequences were omitted from the analysis, but gaps were included. *Circles* represent one haplotype and the circle size is proportional to the haplotype frequency; the *shortest line* represents one mutational step

289 nucleotides from the 3′ end of the AC214524.3_FG002 gene, a Cytochrome P450-like gene, and at 31 Kb from the GRMZM5G867996 gene, whose function is unknown. To examine in greater details the genomic context of the *ZmV1*-9 insertion, we sequenced 1,600–1,900 nucleotides surrounding *ZmV1-9* in 43 inbred lines from the association panel. The inbred lines from Tropical and Corn Belt Dent origin were chosen to represent an equal (or almost) sample of one of the 3 alleles: 14 lines the allele-0, 14 the allele-1 and 15 the allele-2. The sequenced region covered exon 2 of the P450-like gene, the MITE insertion-deletion and 340 nucleotides downstream from the MITE insertion site. We analyzed the sequence diversity of the region flanking the insertion(s). Allele-2 likely occurred in an allele-1 background as demonstrated by the network analysis with closely related alleles-2 deriving from allele-1 (Fig. 5).

We estimated the average pairwise nucleotide diversity (π/bp) for the region (excluding gaps and omitting the MITE sequences) as 0.00535, with 25 polymorphic sites. This value compared well with average genome wide estimates of π/bp in coding region of 0.0068, obtained by analyzing 580 ESTs issued from 14 American maize lines (Corbi et al. 2011). Slightly lower values of π/bp were obtained when considering the allele-0 and 1 separately (0.00379 and 0.00326 respectively). However, we observed a strikingly different pattern in the allele-2 that was characterized by a 14-fold reduction in π/bp estimate (0.00024) with only 3 polymorphic sites (2 SNPs and 1 indel) spanning the entire region and defining 4 haplotypes. The lack of polymorphism information in the allele-2 offered little power to perform neutrality tests (Tajima's *D*, Fu and Li's *D* and *F* values) and none of them were significant. Linkage disequilibrium analysis revealed that the tandem-MITE insertion of allele-2 was in strong ($r^2 = 0.76$) and significant LD ($P < 0.0001$) with a non-synonymous mutation located 1 kb away from the MITE insertion.

Discussion

Even though Transposable Elements are the major component of the maize genome, little is known about their contribution to phenotypic variation and adaptation. In maize, association mapping studies have repeatedly pointed to specific TE insertions that may contribute to flowering time variation (Andersen et al. 2005; Ducrocq et al.

2008; Salvi et al. 2007; Thornsberry et al. 2001). Interestingly these insertions belong to one specific category of DNA Transposable Elements, the MITEs (Miniature Inverted repeat Transposable Elements). MITEs are very abundant in the genome of many eukaryotes and have often been found close to genes (Santiago et al. 2002; Tenaillon et al. 2010a; Zerjal et al. 2009). Although non-autonomous, MITEs participate actively to genome reshaping as shown in rice (Jiang et al. 2003; Naito et al. 2006; Yang et al. 2007). Hence, in that species Yang et al. (2009) have demonstrated that MITEs can be cross-mobilized by transposases produced by several autonomous partners and that they carry internal motives enhancing their excision ability. They may also escape silencing more efficiently than other TEs (Tenaillon et al. 2010b). Overall MITEs therefore offer an interesting source of genomic variation, with potential phenotypic impact, that could be used advantageously in population genetics and association analyses.

We employed the transposon display (TD) technique in 3 MITE families (*Hbr*, *ZmV1* and *Ins2*) to assess the reliability of this method in maize, and to transform MITE insertion polymorphisms into PCR-based markers for application in population genetics and association analyses. The sequencing of 102 TD bands revealed a rather unexpected result: 30% of bands resulted from homoplasy, encompassing a population of insertions at different genomic location, and 16% of allelic bands migrated at different locations in the TD gel owing to small insertion-deletions in the MITE flanking sequence. These results indicate that care must be taken when using TD in maize, at least for the MITE families used in this analysis. They contrast with other plant systems in which sequencing verification of TD bands revealed little homoplasy. These systems include *Arabidopsis* species (Wright et al. 2001), pea (Ellis et al. 1998) and tobacco (Melayah et al. 2001; Petit et al. 2009). Perhaps the difference among plant systems is due to the great fluidity of the maize genome, but fluidity may be more the rule than the exception among angiosperms since most of their genomes are composed of TEs (Wessler et al. 1995; Messing and Bennetzen 2008; Tenaillon et al. 2010a). Moreover, the large number of almost identical MITE copies in the maize genome (Casa et al. 2000; Zerjal et al. 2009) may exceed the discrimination capacity of TD analyses. The diminishing cost of next generation sequencing methods may soon allow the direct sequencing of TD products, offering a reliable genotyping platform. We are in the process of experimenting this technique in maize.

Overall, the patterns of diversity and differentiation revealed by our MITE-based markers were in good agreement with previous reports based on other types of markers (Rebourg et al. 2003; Vigouroux et al. 2008; van Heerwaarden et al. 2011). First, we found high values of F_{ST} revealing the capacity of MITE-based markers to detect population differentiation. However, the largest fraction of the variation (57%) was captured within landraces. Second, our markers revealed a high level of genetic diversity in most landraces (mean value = 0.193, Table 3) but the majority of Mexican landraces had a genetic diversity above average. Third, the pattern of structuring pointed to the existence of an allelic cline between two genetic groups: the North American flint and the Caribbean landraces, which presented contrasting allele frequencies (Fig. 3). Groups were further divided with increasing K values into Mexican, Andean-South American, and Southern Dent landraces ($K = 5$, Fig. 3a). This clustering pattern agreed with previous Structure results (Camus-Kulandaivelu et al. 2006; Vigouroux et al. 2008) with the exception that the Southern Dent cluster was not previously identified as a separate one, but rather part of the Tropical lowland group. It is noteworthy that the Southern Dent cluster identified in our study, both by Structure and network analyses, corresponds to the Central US genetic group recently characterized in a large SNP study (van Heerwaarden et al. 2011) and a worldwide SSR analysis of maize variation (Charcosset, unpublished data). Finally, we also observed a strong geographic influence on the partitioning of MITE-based genetic diversity. Hence, we found a significant correlation between geographic distances and pairwise genetic distances (F_{ST}) (Fig. 1a) consistent with a pattern of isolation by distance that extended up to 3,000 km but was particularly strong at short distances (<500 km). Likewise, SSRs have revealed both a fine (50 km) and a large (4,000 km) scale effect of geographical distance on the decay of genetic similarity (Vigouroux et al. 2008). Altogether our data support the classical scenario with a center of maize origin in Mexico (Matsuoka et al. 2002; van Heerwaarden et al. 2011) and a loss of genetic diversity “out of Mexico” caused by demographic and adaptive events accompanying maize northward and southward expansion into southern and northern America (Tenaillon and Charcosset 2011; Vigouroux et al. 2008). Both the strong geographic component in the organization of maize diversity and the pattern of isolation by distance revealed by our analyses are consistent with a simple diffusion model when a species spread on a broad geographic range and has been established long enough to allow recombination and migration to combine existing haplotypes (Platt et al. 2010).

The second goal of our study was to investigate the contribution of our MITE-based markers to phenotypic variation in an inbred line association panel. Transposable elements can contribute to phenotypic variation in several ways. For instance, they can modulate gene expression through epigenetic control (Feschotte 2008; Hollister and

Gaut 2009; Lippman et al. 2004) and through the introduction of new *cis*-regulatory elements (Chung et al. 2007). To explore this aspect we undertook an association mapping approach using 33 insertions in a panel of 367 maize lines that has been measured for various phenotypic traits including kernel quality traits and male flowering time (Manicacci et al. 2009). The genetic structure of this panel has been previously investigated using genome-wide neutral SSRs (Camus-Kulandaivelu et al. 2006). Variation at the marker *ZmV1-9* was significantly associated with male flowering time (MFT). More specifically, one allele at this marker characterized by the presence of two almost identical tandem *ZmV1* MITE insertions (i.e. allele-2) was associated to late flowering phenotypes with an average effect of +123 degree days (after accounting for population structure). Flowering time is a complex trait that reflects plant adaptation to local climatic conditions. It is controlled by more than 60 QTLs spread over all chromosomes (Buckler et al. 2009). Comparison with QTL consensus maps for flowering time (Chardon et al. 2004) including the most updated one summarizing the results of 29 independent studies (Salvi et al. 2009) did not reveal any overlap between the *ZmV1-9* insertion and a region of high QTL density.

We estimated the *ZmV1-9* contribution to flowering-time variation to 4% (R^2 marker, Table 4). This value was similar to that found for the *Vgt1* region (Ducrocq et al. 2008) which was so far the highest observed for male flowering time in our association panel. The allele-2 frequency varied significantly among the inbred genetic groups, but its allelic effect was consistent within groups (Fig. 4) with no significant marker × Q interaction suggesting an allelic effect independent from the genetic background. Likewise, Buckler et al. (2009) reported little evidence for epistatic interactions in a recent nested association mapping study for flowering time variation in maize. Interestingly the allele-2 was associated with a "very-late flowering" phenotype. For example among the Caribbean inbreds, which are known to be late flowering, the allele-2 genotypes had an average flowering time of +181 degree days compared with the genotypes bearing the alleles −0 or −1 (i.e. no insertion and 1 MITE insertion, respectively).

The *ZmV1-9* polymorphism maps on chromosome 1 at the position 77.966.696 (maize sequence: Release 5b.60, http://www.maizesequence.org/index.html), 289 nucleotides from the 3′ end of the AC214524.3_FG002 gene, a Cytochrome P450-like gene. Cytochrome P450s form a diverse gene superfamily and are essential for catalyzing step in the synthesis of many compounds including pigments, defense compounds, hormones, and signaling molecules. In maize, function has been assigned to some P450 genes; four are part of the defense DIBOA pathway (Schuler and Werck-Reichhart 2003) and the *dwarf3* gene is part of the gibberellin pathway (Winkler and Helentjaris 1995). Interestingly, a gene with strong similarity with the *Zea dwarf3* gene was found in a QTL region involved in the inflorescence development pathway of *Arabidopsis* (Ungerer et al. 2002). It is therefore very likely that the gene identified influences maize flowering time. We propose that the tandem-MITE insertion or a mutation in close linkage disequilibrium could be the causative variant. In order to investigate those two alternative hypotheses, we first estimated the Linkage Disequilibrium (LD) and its significance in the region surrounding the insertion. We found a strong and significant LD ($r^2 = 0.76$, $P < 0.0001$) between the tandem-MITE insertion of allele-2 and a non-synonymous mutation located 1 kb away from the MITE insertion. This mutation is in the exon 2 of the Cytochrome P450-like gene and transforms a Serine residue (TCT) into a Cysteine (TGT) in the allele-2 background. A comparative analysis of the homologous locus in rice and *Arabidopsis* has revealed that this mutation also segregates in these species. Whether this mutation alters the structure and functionality needs, however, yet to be determined. One alternative hypothesis is that the tandem MITE located in the vicinity of the 3′ end of the Cytochrome P450-like gene could modulate its expression and ultimately influence phenotypic variation. Modulation of expression via TEs are common in other model organisms such as *Drosophila* where most of the adaptive TE insertions so far identified are involved in regulatory rather than in coding changes (Gonzalez et al. 2008; Gonzalez and Petrov 2009). In sorghum (*Sorghum bicolor*), a potentially quantitative allelic effect due to a variable number of MITE copy insertions was described for an aluminum tolerance gene (Magalhaes et al. 2007). In our case, gene regulation may be altered because MITEs of the *ZmV1* family, as most other MITEs, are able to produce stable hairpin (stem–loop) secondary structures (Bureau and Wessler 1992). The size of this secondary structure increases with the number of insertions and may alter the accessibility of the 3′ region of the gene and explain why tandem-MITE insertions (allele-2) are associated with a different phenotype than single insertion. An alternative possibility involves siRNA regulation. Hence, screening maize siRNA databases (CSRDB http://sundarlab.ucdavis.edu/smrnas/) we found one motif with 100% homology between the 24 nucleotide maize siRNA zma-smRNA14581 and the *ZmV1-9* MITE insertion, allele-1 (single insertion) bearing a single motif instead of two in allele-2 (tandem insertion). This MITE insertion could therefore well be a target site for this specific siRNA and in turn affect the expression of the gene nearby. Finally, the MITE may be itself a source of siRNA. This function of MITEs was recently demonstrated in *Solanaceae* (Kuang et al. 2009) where several MITE families were shown to generate small RNAs of primarily 24 nt in length and similar observations exist for *Arabidopsis* and rice (Piriyapongsa and Jordan 2008).

In conclusion, first we have demonstrated that while MITE-based polymorphisms accurately reflect the genetic history of the maize landraces, the Transposon Display method must be used with great caution and alternative methods such as nextgen sequencing of TD reactions need to be examined. Second, we have identified a new candidate region associated with flowering time variation in maize that does not map to any previously characterized QTL, stressing the importance of considering structural polymorphism in association studies. This region encompasses a Cytocrome P450-like gene with a triallelic polymorphism of a MITE insertion in its 3′ region. Whether a non-synonymous mutation within the gene or the MITE polymorphism itself influences phenotypic variation remains to be determined. Finally, we propose a number of exciting although speculative hypotheses for explaining how the MITE polymorphism could modulate the expression of the Cytocrome P450-like gene. These include formation of hairpin structure limiting the gene 3′ accessibility and siRNA regulation.

Acknowledgments We are very grateful to Domenica Manicacci for her advice on the association analysis and insightful discussions, to Brandon Gaut and Domenica Manicacci for providing comments on the manuscript. We thank Jean-Baptiste Veyrieras for helping in choosing maize landraces, Celine Mir who shared unpublished data, Pauline Garnier-Géré who brought valuable insights and help with Fdist2, Fabrice Dumas and Xavier Raffoux who provided technical assistance. This work was funded by the Agence Nationale de la Recherche (ANR-05-JCJC-0067-01 to MIT) and by Promaïs members: Caussade Semences, Euralis semences, Limagrain Verneuil holding, Maïsadour Semences, Monsanto France SAS, Syngenta seeds SAS, RAGT 2n and KWS France. We are particularly thankful to Jean Beigbeder for stimulating coordination of Genetic Diversity programs within Promaïs.

References

Andersen JR, Schrag T, Melchinger AE, Zein I, Lubberstedt T (2005) Validation of Dwarf 8 polymorphisms associated with flowering time in elite European inbred lines of maize (Zea mays L.). Theor Appl Genet 111:206–217

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 17:3425–3427

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc London B263:1619–1626

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B Methodological 57:289–300

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

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Bureau TE, Wessler SR (1992) Tourist: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294

Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) The MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Chao L, McBroom S (1985) Evolution of transposable elements: an IS10 insertion increases fitness in Escherichia coli. Mol Biol Evol 2:359–369

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ (2007) Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. Genetics 175:1071–1077

Corbi J, Debieu M, Rousselet A, Montalent P, Le Guilloux M, Manicacci D, Tenaillon MI (2011) Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme. Theor Appl Genet 122:705–722

Dubreuil P, Charcosset A (1998) Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor Appl Genet 96:577–587

Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433–2437

Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Eveno E, Collada C, Guevara MA, Leger V, Soto A, Diaz L, Leger P, Gonzalez-Martinez SC, Cervera MT, Plomion C, Garnier-Gere PH (2008) Contrasting patterns of selection at Pinus pinaster Ait. Drought stress candidate genes as revealed by genetic differentiation analyses. Mol Biol Evol 25:417–437

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397–405

Gonzalez J, Petrov DA (2009) The adaptive role of transposable elements in the Drosophila genome. Gene 448:124–133

Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol 6:e251

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

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

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898–1914

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:13

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421:163–167

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res 19:42–56

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476

Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes Dev 18:2873–2878

Magalhaes JV, Liu J, Guimaraes CT, Lana UGP, Alves VMC, Wang Y-H, Schaffert RE, Hoekenga OA, Pineros MA, Shaff JE, Klein PE, Carneiro NP, Coelho CM, Trick HN, Kochian LV (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nat Genet 39:1156–1161

Manicacci D, Camus-Kulandaivelu L, Fourmann M, Arar C, Barrault S, Rousselet A, Feminias N, Consoli L, Frances L, Mechin V, Murigneux A, Prioul JL, Charcosset A, Damerval C (2009) Epistatic interactions between Opaque2 transcriptional activator and its target gene CyPPDK1 control kernel trait variation in maize. Plant Physiol 150:506–520

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien M-A (2001) The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. Plant J 28:159–168

Messing J, Bennetzen JL (2008) Grass genome structure and evolution. Genome Dyn 4:41–56

Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. Proc Natl Acad Sci USA 100:10102–10107

Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden SE, Berg PR, Lien S (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. BMC Genet 9:18

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR (2006) Dramatic amplification of a rice transposable element during recent domestication. Proc Natl Acad Sci USA 103:17620–17625

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

Pariset L, Joost S, Marsan PA, Valentini A (2009) Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. BMC Genet 10:7

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian Bush Rat, *Rattus fuscipes*. Evolution 57:1182–1195

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien M-A, Mhiri C (2009) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186:135–147

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc Natl Acad Sci USA 106:5019–5024

Piriyapongsa J, Jordan IK (2008) Dual coding of siRNAs and miRNAs by plant transposable elements. RNA 14:814–821

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K, Dunning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO (2010) The scale of population structure in Arabidopsis thaliana. PLoS Genet 6:e1000843

Pressoir G, Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92:95–101

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

Ralston EJ, English JJ, Dooner HK (1988) Sequence of three bronze alleles of maize and correlation with the genetic fine structure. Genetics 119:185–197

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res 67:175–185

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus Zea. Genetics 181:1399–1413

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Rozas J, Rozas R (1999) DnaSP version 3: an integreted program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao GH, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104:11376–11381

Salvi S, Castelletti S, Tuberosa R (2009) An updated consensus map for flowering time QTLs in maize. Maydica 54:501–512

Sanchez GJJ, Goodman MM, Stuber CW (2000) Isozymatic and morphological diversity in the races of maize of Mexico. Econ Bot 54:43–59

Santiago N, Herraiz C, Goni JR, Messeguer X, Casacuberta JM (2002) Genome-wide analysis of the Emigrant family of MITEs of Arabidopsis thaliana. Mol Biol Evol 19:2285–2293

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, Robin C (2010) Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Genet 6:e1000998

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol 54:629–667

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82(Pt 5):561–573

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734

Storey JD (2003) The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Stat 31:2013–2035

Tenaillon MI, Charcosset A (2011) A European perspective on maize history. C R Biol 334:221–228

Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J (2010a) Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians. Genome Biol Evol 3:219–229

Tenaillon MI, Hollister J, Gaut BS (2010b) A triptych of the evolution of plant transposable elements. Trends Plant Sci 15:471–478

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

Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TF, Purugganan MD (2002) Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics 160:1133–1151

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA 108:1088–1092

Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sanchez GJ, Doebley J (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci USA 103:17644–17649

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Inc., Sunderland, pp 91–138

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. Evolution 38:1358–1370

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Winkler RG, Helentjaris T (1995) The maize Dwarf3 gene encodes a cytochrome P450-mediated early step in Gibberellin biosynthesis. Plant Cell 7:1307–1317

Wright SI, Le QH, Schoen DJ, Bureau TE (2001) Population dynamics of an Ac-like transposable element in self- and cross-pollinating arabidopsis. Genetics 158:1279–1288

Yang G, Zhang F, Hancock CN, Wessler SR (2007) Transposition of the rice miniature inverted repeat transposable element mPing in Arabidopsis thaliana. Proc Natl Acad Sci USA 104:10962–10967

Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR (2009) Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science 325:1391–1394

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

Zerjal T, Joets J, Alix K, Grandbastien MA, Tenaillon MI (2009) Contrasting evolutionary patterns and target specificities among three Tourist-like MITE families in the maize genome. Plant Mol Biol 71:99–114

Zhang Q, Arbuckle J, Wessler SR (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. Proc Natl Acad Sci USA 97:1160–1165