

Natural Diversity in Flowering Responses of *Arabidopsis thaliana* Caused by Variation in a Tandem Gene Array

Sarah Marie Rosloski,* Sathya Sheela Jali,* Sureshkumar Balasubramanian,^{†,‡}
Detlef Weigel[†] and Vojislava Grbic^{*,†,1}

*Department of Biology, University of Western Ontario, London, Ontario N6A 5B8, Canada, [†]Department of Molecular Biology, Max-Planck Institute for Developmental Biology, D-72076, Tuebingen, Germany and [‡]The University of Queensland, School of Biological Sciences, St. Lucia, Queensland 4072, Australia

Manuscript received March 7, 2010

Accepted for publication May 24, 2010

ABSTRACT

Tandemly arrayed genes that belong to gene families characterize genomes of many organisms. Gene duplication and subsequent relaxation of selection can lead to the establishment of paralogous cluster members that may evolve along different trajectories. Here, we report on the structural variation in *MADS AFFECTING FLOWERING 2* (*MAF2*) gene, one member of the tandemly duplicated cluster of MADS-box-containing transcription factors in *Arabidopsis thaliana*. The altered gene structure at the *MAF2* locus is present as a moderate-frequency polymorphism in *Arabidopsis* and leads to the extensive diversity in transcript patterns due to alternative splicing. Rearrangements at the *MAF2* locus are associated with an early flowering phenotype in BC₅ lines. The lack of suppression of flowering time in a *MAF2*-insertion line expressing the *MAF2*-specific artificial miRNA suggests that these *MAF2* variants are behaving as loss-of-function alleles. The variation in gene architecture is also associated with segregation distortion, which may have facilitated the spread and the establishment of the corresponding alleles throughout the Eurasian range of the *A. thaliana* population.

A single, ancestral gene can give rise to several copies in the genome through the process of gene duplication. Gene duplicates can be found either as dispersed copies or in tandem arrays. While dispersed copies of gene duplicates arise commonly from segmental or whole genome duplication events or retroposition, tandem duplicates can be initiated from single-copy genes by repair of double-stranded DNA breakage and multiplied by unequal crossover events (STURTEVANT 1925; JELESKO *et al.* 1999; BLANC *et al.* 2003; SOLTIS and SOLTIS 2003; NARAYANAN *et al.* 2006; SLACK *et al.* 2006; YANDEAU-NELSON *et al.* 2006; KONG *et al.* 2007; FREELING *et al.* 2008).

Gene duplication leads to redundancy, which relaxes selection pressure and thereby creates a substrate for functional evolution (GU and GU 2003; SHAKHNOVICH and KOONIN 2006; WAGNER 2008). Due to redundancy, the capacity to sustain nonlethal mutations is enhanced in gene families compared to singleton genes (CLARK *et al.* 2007; ARMISEN *et al.* 2008). In plants, successful

retention of new gene duplicates is biased toward genes that are involved in plant's ability to respond to environmental cues (HARRISON and GERSTEIN 2002; SAKURAI *et al.* 2007; KORBEL *et al.* 2008). Pathogen response genes that belong to the nucleotide-binding site-leucine-rich repeat (NBS-LRR) gene family show case functional diversification (MONDRAGON-PALOMINO *et al.* 2002). Although long-lived resistance gene polymorphisms have been noted (STAHL *et al.* 1999), the large size of gene families associated with pathogen defense in plants may facilitate a response to the rapid evolution of resistance to plant defense mechanisms in pathogens (BAKKER *et al.* 2006; CLARK *et al.* 2007).

Gene families frequently display variation between individuals within a species (REDON *et al.* 2006; CUTLER *et al.* 2007). This is especially true for tandemly arrayed genes (TAGs). Compared to dispersed paralogs, TAG members are more susceptible to copy number expansion and contraction via unequal crossing over (JELESKO *et al.* 1999; LEISTER 2004), sequence exchange via gene conversion (GAO and INNAN 2004; MONDRAGON-PALOMINO and GAUT 2005; NEI and ROONEY 2005; GANLEY and KOBAYASHI 2007; XU *et al.* 2008), and the formation of chimeric genes (JELESKO *et al.* 2004; KUANG *et al.* 2006), thereby creating lineage-specific diversity in TAG regions (KUANG *et al.* 2004, 2008). Approximately 1 of 700 seeds produced per plant is estimated to be polymorphic in 1 of the ~1500 TAGs observed in *Arabidopsis thaliana* (ZHANG and GAUT 2003; GAUT *et al.* 2007).

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.116392/DC1>.

Available freely online through the author-supported open access option.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. HM487066–HM487102.

¹Corresponding author: Department of Biology, University of Western Ontario, 1151 Richmond St., London, Ontario N6A 5B8, Canada. E-mail: vgrbic@uwo.ca

In plants, development is highly dependent on the environment. One well-known class of developmental regulators is encoded by the MADS-box gene family. One hundred and seven MADS-box family members in *A. thaliana* are involved in the morphological diversification of floral and root tissues, development of fruit and seeds, and the transition to flowering (BECKER and THEISSEN 2003; PARENICOVA *et al.* 2003). Several of these genes affect flowering time, and among these, *FLOWERING LOCUS C (FLC)* and a small clade of closely related genes called *MADS AFFECTING FLOWERING (MAF)* encode repressors of flowering. *FLC* locus is located at the top of chromosome 5, and *MAF1*, also known as *FLOWERING LOCUS M (FLM)* or *AGAMOUS-LIKE 27 (AGL27)* is on chromosome 1. The remaining members of this group, *MAF2* through *MAF5*, are found in a cluster on the bottom of chromosome 5 (BECKER and THEISSEN 2003; KOFUJI *et al.* 2003; PARENICOVA *et al.* 2003). The products of these genes share between 55 and 88% amino acid identity and all clade members are alternatively spliced (RATCLIFFE *et al.* 2003; SCORTECCI *et al.* 2003; LEMPE *et al.* 2005; CAICEDO *et al.* 2004). The relevance of alternative splicing for normal function of these transcription factors is currently not known.

FLC and the *MAFs* mediate responsiveness to the environment (MICHAELS and AMASINO 1999; RATCLIFFE *et al.* 2003; BALASUBRAMANIAN *et al.* 2006b; SUNG *et al.* 2006). *FLC* expression is repressed by a long, cold treatment comparable to winter (vernalization), thereby promoting flowering after winter has passed (MICHAELS and AMASINO 1999). Variation in *FLC* expression is responsible for ~40% of variation in flowering time in a survey of global *A. thaliana* accessions (LEMPE *et al.* 2005; SHINDO *et al.* 2005). *MAF1* acts predominately to repress flowering under short-day conditions (SCORTECCI *et al.* 2003; WERNER *et al.* 2005; BALASUBRAMANIAN *et al.* 2006b). *MAF2* prevents early flowering in response to short periods of cold, which may be an adaptation to brief cold episodes followed by transient warmer temperatures in the autumn, thereby avoiding initiation of flowering prior to the onset of winter (RATCLIFFE *et al.* 2003). A T-DNA knockout of *MAF2* is early flowering, indicating that *MAF2* is not completely redundant to other TAG members, *MAF3-5*. Sequence analysis at the population level suggested purifying selection at *MAF4* and *MAF5* in contrast to *MAF2* and *MAF3*, which are characterized by a high degree of polymorphism (CAICEDO *et al.* 2009). One common polymorphism at *MAF2* locus has been described as structural variation that associates with early flowering in a set of *A. thaliana* accessions (CAICEDO *et al.* 2009). The extent and functional significance of this structural variation at *MAF2* is currently not well understood.

Here, we address the natural diversity, geographic distribution, and functional consequences of the widespread, natural *MAF2* allelic series. By screening a large accession population for both gene expression and

structural variation, we demonstrate extensive and previously unknown, natural variation at *MAF2*. Using a backcrossing strategy, we demonstrate that a series of rearrangements at the *MAF2* locus are associated with an early flowering phenotype under long-day conditions. Using *MAF2*-specific, artificial miRNA (amiRNA) knock-down lines, we show that early flowering associated with a rearranged *MAF2* allele is conferred by a loss of *MAF2* function. In addition, consistent with the widespread geographical distribution, we show that segregation distortion, linked to the *MAF* locus, favors chimeric *MAF2* alleles. Association of *MAF2* to a locus capable of distorting the segregation ratio provides an alternative explanation for the widespread distribution of *MAF2* structural variants.

MATERIALS AND METHODS

Plant material: Accessions were obtained from the Arabidopsis Biological Resource Center and are listed in the supporting information, Tables S1 and S2, respectively. *maf2*, a T-DNA knock-out allele in the Col background derived from the Salk collection (SALK-045625, s206432) was kindly provided by Oliver Ratcliffe, Mendel Biotechnology.

Growth conditions: Seeds were sterilized with 20% bleach, followed by 70% ethanol, washed in sterile water, plated on half-strength MS medium (Sigma) with 0.7% agarose, and stratified for 3 days at 4°. Seeds were then resuspended in half-strength MS medium and were planted in the corners of square 85-mm pots containing PRO-MIX soil (Plant Products) using a Pasteur pipette. Four seedlings were grown in each pot. Plants were grown under 100- to 150- $\mu\text{mol m}^{-2} \text{s}^{-1}$ cool-white fluorescent light at 22° in 16 hr light/8 hr dark long-day (LD) or 8 hr light/16 hr dark short-day (SD) cycles in controlled growth chambers. Plants for RNA extraction were stratified at 4° for 3 days and harvested after growth for 8 days at 21° on half-strength MS medium with 0.7% agarose and 1% sucrose under LD.

Genotyping backcrossed and segregating lines: Backcrossed (BC) lines were generated by crossing UWO, Tu-1, Sha, or Kas-1 accessions to *Ler* five times. Progeny of successive crosses were grown and their genotypes at the *MAF2* locus were determined in each generation by PCR (Table S6 lists primers used in this study). DNA was extracted using the method of EDWARDS *et al.* (1991). Cosegregation of the flowering-time phenotype with *MAF2* alleles was determined in a BC₅ F₂ segregating population. After counting leaves, DNA was extracted from individual plants and 2 μl of this suspension was subjected to PCR. A set of three unique primers was designed to diagnostically amplify each *MAF2* insertion allele in conjunction with *MAF2-Ler* in backcross and segregating lines. Control amplifications were conducted with plants of known genotype.

cDNA cloning, sequencing, and predicted protein analysis: RT-PCR products were cloned into pGEM-T-easy TA vector (Promega) using manufacturer's instructions. Plasmids were introduced into *Escherichia coli* DH5 α by the heat-shock method. Plasmid DNA was isolated from individual clones and sequenced. Sequences were aligned to the genomic region using SIM4 (FLOREA *et al.* 1998) and visualized in LalnView (DURET *et al.* 1996). Protein sequences were predicted with Transeq (RICE *et al.* 2000) at EBI (LABARGA *et al.* 2007).

Generation of amiRNA lines: The Web MicroRNA Designer platform (WMD2, <http://wmd.weigelworld.org/bin/>

mirnatools.pl) (SCHWAB *et al.* 2006) was used to design an amiRNA specific to *MAF2* (At5g65050) transcripts from *MAF2-Ler* and the *MAF2* insertion alleles. The plasmid pRS300 containing mir319a precursor was used as a template to replace the endogenous miRNA and miRNA* sequences by the 21mers of *MAF2*-amiRNA and *MAF2*-miRNA* (SCHWAB *et al.* 2006). The three overlapping PCR products were generated by using primers 57, 58, 59, and 60. The three PCR products were joined in a single PCR reaction by using the miR319a backbone specific primers (61 and 62). The resulting fragment containing amiRNA and the backbone of miRNA 319a were digested with *EcoRI* and *BamHI* and cloned into pJLBlue (-) entry vector. An LR reaction was performed to move this fragment into the binary vector pEarlyGate101. Transgenic plants were selected with phosphinothricine (BASTA).

Linkage analysis of *MAF2*-UWO transcript pattern: The *MAF2*-UWO transcript pattern was mapped in the F₂ generation of a UWO × Col cross. RNA and DNA were collected from 35 individual plants and subjected to either RT-PCR or PCR, as described above. Plants were genotyped at *MAF2* locus using primers 44 and 45, at ATTED2 marker using primers 47 and 48, and at EG7F2 using primers 49 and 50. Map distances were determined by MAPMAKER version 3.0 (LANDER *et al.* 1987; LINCOLN *et al.* 1992).

Sequencing of *MAF2* insertion alleles: In the genomic survey, accessions were first analyzed for the 3' of the UWO insertion allele using primers 7, 8, and 9. The 5' border of positive accessions was sequenced by amplifying the region of interest with Pfu polymerase (Promega) using primers 15 and 16. The PCR fragments were purified and directly sequenced.

The entire *MAF2* region was sequenced on both strands from six accessions representing different insertion classes after PCR amplification using Pfu polymerase (Promega) with overlapping primer sets. Because of sequence similarity with the downstream *MAF3* gene, the entire *MAF3* insertion region within *MAF2* was amplified. This fragment was sequenced directly using primers and PCR conditions reported in Table S6.

Gene expression experiments: Eight-day-old seedlings were harvested and pulverized in liquid nitrogen. RNA was extracted using the RNeasy kit (Qiagen). RNA (2 µg) was subjected to RT-PCR using the Superscript II reverse transcriptase (Invitrogen) in a 20-µl reaction using manufacturer's instructions. A 2-µl sample of this reaction was subjected to RT-PCR. Amplification of genomic DNA was undetectable using all primer combinations.

Phylogenetic analysis: A 1128-bp DNA sequence of the *MAF3* gene was derived from 20 accessions that tested negatively for the *MAF2* insertion allele using PCR primers 55 and 56 or from sequence information previously deposited in GENBANK (PopSet: EU980614-EU980630). The same region of the *MAF3* gene was sequenced or acquired from preexisting *MAF* GENBANK PopSet data from seven accessions bearing the *MAF2* insertion allele. Additionally, the paralogous *MAF3*-derived sequence was acquired from the *MAF3* insertion region from eight accessions using sequencing primers reported above or from the *MAF* GENBANK PopSet. These sequence data were used to generate a UPGMA dendrogram in PHYLIP version 3.66 using 10,000 replicates (FELSENSTEIN 2005).

Statistical analysis: Allelic Aggregation Index Analysis was performed using the Alleles in Space (AIS) software (MILLER 2005). Only coordinates of Eurasian accessions from the genomic survey were used. Differences in rosette leaf number in the BC₅ segregating populations were analyzed by ANOVA implemented in PAST version 1.81 (<http://folk.uio.no/ohammer/past/index.html>). Segregation distortion in BC₅ populations was assessed by chi-square analysis.

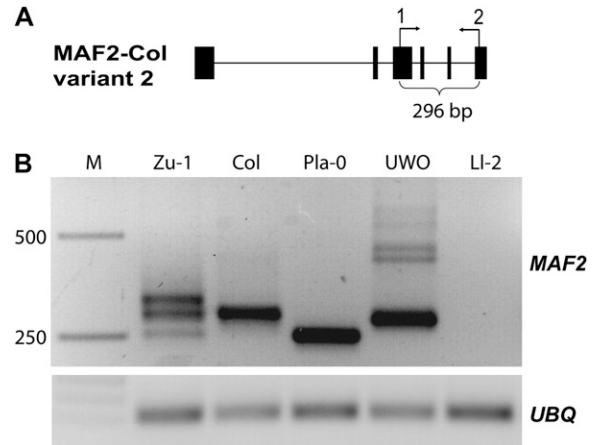


FIGURE 1.—Expression survey for transcript variation at *MAF2*. (A) Diagram of the *MAF2*-Col variant 2 transcript with position of primers used. Primer 1 is specific for the alternatively spliced exon present only in this single *MAF2* transcript. Solid boxes, exons; lines, introns. (B) Expression classes observed in the cDNA screen of Arabidopsis accessions. RT-PCR products for *MAF2* (top) and ubiquitin, UBQ, (bottom). M, molecular marker; Zu-1; Col (reference); Pla-0, UWO, and LI-2 are shown as class representatives.

RESULTS

Natural variation in transcript patterns of *MAF2*: To assess the spectrum and the frequency of natural variation in transcript patterns at *MAF2* locus we surveyed the transcript variation across 147 accessions using RT-PCR with primers 1 and 2 (Figure 1A). Approximately half (72/147) of accessions had a banding pattern characteristic of the Col reference strain (RATCLIFFE *et al.* 2003). We grouped the remaining accessions into four additional classes that are named after a representative accession: 16% (23/147) had an additional larger RT-PCR product (exemplified by Zu-1), 9% (13/147) had a single, smaller product (exemplified by Pla-0), 24% (36/147) of accessions had a slightly smaller product accompanied by a series of larger products (exemplified by UWO), and in three accessions (exemplified by LI-2) we failed to obtain RT-PCR products (Table S1). Upon sequencing the corresponding PCR products in accessions representative for each class, we were able to confirm in all cases, except Zu-1, that these novel bands contain *MAF2* sequence (data not shown). Therefore, in 52/147 accessions tested, the major transcript encoded by the *MAF2* locus deviates in size and sequence from the *MAF2*-Col reference allele, indicating that *MAF2* harbors extensive natural transcript variation.

The divergent RT-PCR banding pattern in the UWO class is associated with the production of chimeric transcripts containing both *MAF2* and *MAF3* sequences. As the UWO class had the greatest number of accessions, we further investigated the nature of *MAF2* gene expression characteristic for this set. To ensure that extra bands generated in this class are derived from

MAF2 cDNAs, we used another set of primers that specifically amplify two cDNAs in the *Col* and *Ler* backgrounds (Figure 2A). As previously seen, numerous extra fragments can be amplified from accessions of this class, suggesting that there is variation at *MAF2* transcripts in these accessions (Figure 2A). To determine the identity of amplified products, we cloned and sequenced bands amplified from a representative strain, UWO (Figure 2B). Sequencing revealed that bands in the higher-molecular-weight range, 800–1100 bp, are derived from a chimeric cDNAs containing exonic regions of *MAF3* inserted between the penultimate and last *MAF2* exons, relative to *Col* reference sequence (Figure 2B). Alternative splicing, characteristic of both *MAF2* and *MAF3*, resulted in the recovery of five *MAF2*–UWO splice variants. Surprisingly, the lower-molecular-weight bands, 600 and 750 bp, recovered from the UWO strain were derived from the *MAF3* locus, even though the *MAF3* cDNAs are not amplified with these primers in *Col* or *Ler*. Sequence analysis, however, revealed that the final exon containing the 3' UTR of the *MAF3*–UWO is derived from *MAF2* (Figure 2B) allowing for the non-specificity of the PCR reaction in this genotype. Therefore, the divergent RT–PCR banding pattern in this class of accessions is associated with the production of chimeric transcripts containing both *MAF2* and *MAF3* sequences. In addition to size differences of cDNAs recovered from the *MAF2*–UWO allele, it is worth noting that the band intensity of RT–PCR products is considerably lighter than that observed in *MAF2*–*Ler*. Therefore, a reduced expression level of individual transcripts also accompanies the polymorphism in *MAF2*–UWO transcripts.

***MAF2*–*MAF3* chimeric transcripts are generated from a rearranged locus in the UWO class:** Rearrangements at the *MAF2* and *MAF3* loci have been recently described in accessions of *A. thaliana* (CAICEDO *et al.* 2009). Therefore, we wondered whether the transcript pattern we observed could be due to a rearrangement of *MAF2* and *MAF3* in the UWO strain. Sequencing revealed a 1371-bp insertion of *MAF3* genomic sequences into the final intron of *MAF2* (Figure 3). In addition, consistent with the chimeric sequence of *MAF3*–UWO cDNAs, the final exon and the 3' UTR of *MAF3*–UWO were replaced by the corresponding *MAF2* sequences (Figure 3). While *MAF2*–UWO is similar in its structure to the *MAF2* rearranged alleles previously reported by CAICEDO *et al.* (2009), *MAF2*–UWO does not correspond to any of the six *MAF2* allelic groups described. Thus, *MAF2*–UWO allele is a newly identified form of the *MAF2* insertion allele.

Chimeric transcripts map at the *MAF* locus: The configuration of *MAF2*–UWO and *MAF3*–UWO genomic sequences can account for *all* cDNAs detected in this genotype through a genomic rearrangement combined with alternative splicing. However, both this study and previous study of genetic variations associated with

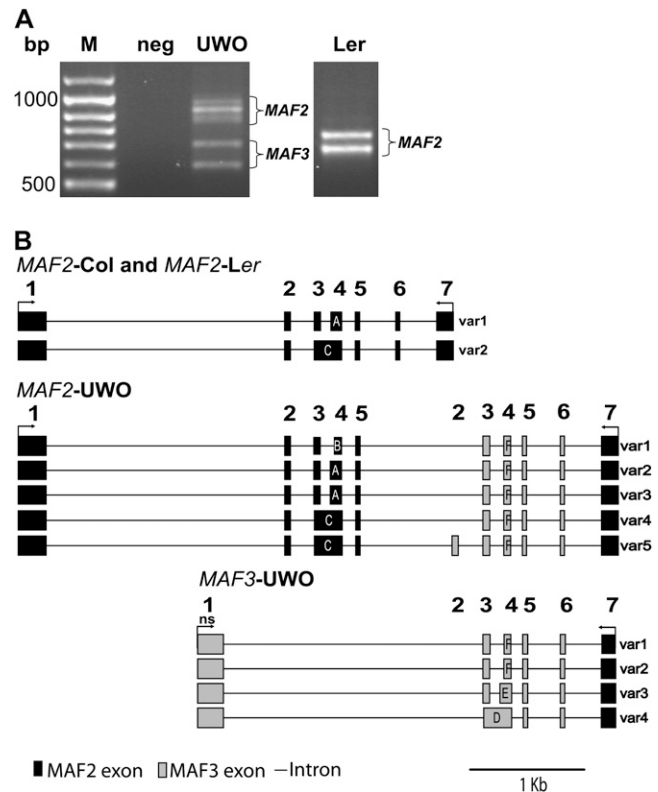


FIGURE 2.—Chimeric nature of *MAF2* and *MAF3* cDNAs from the UWO accession. (A) RT–PCR product profiles of UWO and *Ler* lines using primers 44 and 45. (B) Exon and intron structure of cDNAs sequenced from *MAF2*–*Col*, *MAF2*–*Ler*, *MAF2*–UWO, and *MAF3*–UWO. Primers 44 and 45 are shown as arrows; ns, nonspecific primer binding site; solid boxes, *MAF2* exons; shaded boxes, *MAF3* exons. In alternatively spliced regions, exons labeled with the same letter have the same 5' and 3' boundaries.

the *MAF* locus are based on the sequence of contiguous PCR products, which are untethered to a chromosome region (CAICEDO *et al.* 2009). Furthermore, a chimeric transcript containing *MAF2* and *MAF3* sequences has been reported in the *Col*-0 background (RATCLIFFE *et al.* 2003). Therefore, in principle, a newly formed *MAF2* paralog situated elsewhere in the *A. thaliana* genome, or a *trans*-splicing event, could lead to the observed transcription profile. To ensure that the observed pattern is due to the rearrangement at *MAF2* locus, we mapped the novel transcript pattern. Analysis of UWO × *Col*-0 F₁ plants revealed that this transcript pattern is codominant (Figure S1). We next tested linkage between the chimeric banding pattern and *MAF2* using ATTED2 and EG7F2 markers that are linked to the *MAF* locus (Figure S2). The RT–PCR profile was linked to these markers, indicating that transcript variation is controlled in *cis*, arising as a result of the rearrangement at the *MAF2* locus.

A series of rearranged *MAF2* alleles in *A. thaliana* populations: Discovery of the novel *MAF2* insertion allele in a screen of 147 accessions suggests that there may be even greater genetic diversity than previously reported

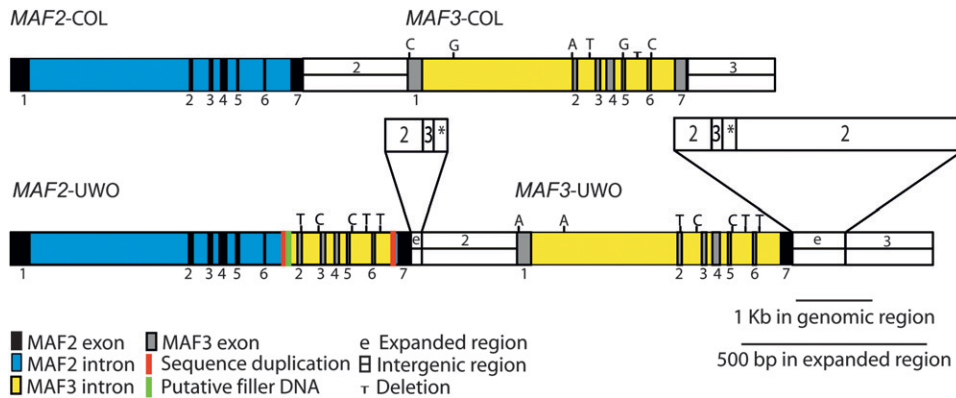


FIGURE 3.—Configuration of the *MAF2*-UWO and *MAF3*-UWO loci. *MAF2*-UWO has an insertion of 1371 bp of *MAF3* genomic DNA into the last intron. Polymorphisms differentiating *MAF3*-Col from *MAF2*-UWO, and *MAF3*-UWO are shown above. The expanded regions reveal rearrangements downstream of *MAF2*-UWO and *MAF3*-UWO. Numbers 2 and 3 refer to DNA sequence homologous to the downstream region of *MAF2*-Col or *MAF3*-Col, respectively; * refers to a region that cannot be definitively assigned to *MAF2* or *MAF3* because of sequence identity.

(CAICEDO *et al.* 2009). To determine the frequency of the novel *MAF2*-UWO insertion allele and to discover additional insertion alleles, we surveyed 323 accessions for diagnostic PCR fragments at the 5' and 3' boundary of the insertion. We used primers 7, 8, and 9 to amplify DNA at the 3' boundary of the insertion (Figure 4A). These primers are expected to generate a 639-bp fragment at *MAF2*-UWO and a 739-bp fragment in *MAF2*-Col. All 31 accessions initially identified in the expression screen that had a similar RT-PCR pattern as UWO tested positively for the 639-bp, 3' boundary fragment. Sixty-one accessions belonging to other RT-PCR classes were also retested; they all amplified the 739-bp, Col-like fragment. The congruence between the genomic and RT-PCR results indicates that our assay is diagnostic for the *MAF2* insertion allele. Of 225 accessions not previously tested, 127 accessions amplified the 739-bp Col-like fragment and 98 amplified the 639-bp fragment at the 3' insertion boundary. Therefore, a total of 129 out of 317 accessions, or 41%, had the insertion allele.

The identified 129 accessions were further tested for the presence of the *MAF3* insertion by amplification and sequencing of the fragment generated with primers 15 and 16 at the 5' boundary of the insertion (Figure 4A). All 129 accessions also contained this fragment, suggesting that these accessions contain insertions of *MAF3* into the *MAF2* genomic region. However, the size of amplified genomic fragments was not uniform. Eighteen accessions yielded the fragment at the 5' boundary of the insertion that was identical in size and sequence to the *MAF2*-UWO. Sequencing PCR fragments from the remaining accessions revealed the presence of different *MAF3* inserts into *MAF2*, sequestering the 129 accessions into six subclasses (Tables S2 and S3). All accessions within a single subclass have an identical sequence of the PCR fragment amplified at the 5' boundary of the insertion.

Subsequent sequencing of the entire *MAF2* genomic sequence of a representative accession from each of the six groups provided insight into the topology of these

MAF2 alleles (Figure 4B). The five additional alleles have longer insertions of *MAF3* into the last intron of *MAF2* compared to *MAF2*-UWO. The *MAF3* insertion is accompanied by deletions of varying lengths of the *MAF2* sequences adjacent to the 5' insertion boundary (Table S3). The coincident insertions of *MAF3* sequences and deletions of *MAF2* sequences create divergent product sizes with primers 15 and 16 (Figure 4B). Two of six identified alleles, from the Sha and Gr-3 subgroups, were previously identified as the S2 and S4 insertion types, respectively, by CAICEDO *et al.* (2009). However, the UWO, Sg-1, Tu-1, and KZ10 alleles identified four new *MAF2* allelic subclasses.

Geographic distribution of the *MAF2* insertion allele class and subclasses: Inclusion of accessions collected throughout the Eurasian range of *Arabidopsis* allowed us to study the allelic distribution of both allelic classes and subclasses. *MAF2* insertion alleles were found predominantly in accessions from the central and eastern Eurasian range (Figure 5A, shaded circles). Interestingly, each subclass also had a distinct geographical distribution (Figure 5B). The Sha subclass is ubiquitous across Eurasia with greater representation among accessions in the eastern Eurasian range (Figure 5B). In contrast to the wide distribution of the Sha subclass, the Sg-1 subclass is confined to the Alps region, while the KZ-10 and Gr-3 subclasses appear to be confined between 45° and 55° north latitude. The UWO subclass is mostly present in central Europe, and the Tu-1 subclass has a disjunct distribution with population centers in the Alps and Northern Scandinavia. Allelic Aggregation Index Analysis (MILLER 2005) showed significant clustering of all allele subclasses, except the Sha class ($R_{ave} = 17.3122$, $P \leq 0.048^*$; $R_{Sha} = 26.307$, $P = 0.207$). Thus, *MAF2* structural variation has a strong tendency to exhibit population structure, suggesting that these alleles may be phylogenetically related.

To test the phylogenetic relatedness among *MAF2* insertion alleles, we compared *MAF3* sequences originat-

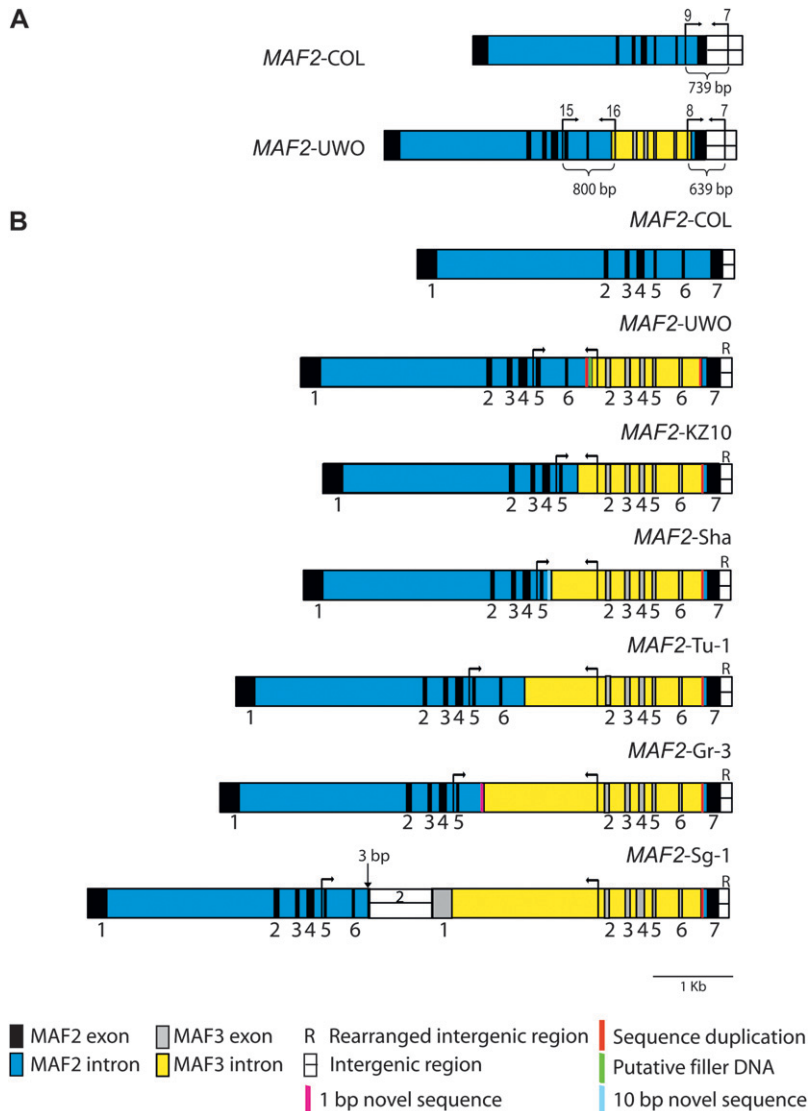


FIGURE 4.—A natural series of *MAF2* insertion alleles. (A) Schematic of *MAF2*-Col and *MAF2*-UWO genomic regions showing primers used in the genomic survey. (B) Natural *MAF2* allelic series based on the genomic DNA sequence of insertion allele subclasses. Genomic regions of the subclasses are aligned at the 3' end to demonstrate length variability in *MAF3* insertion region. Binding sites of primers 15 and 16 are depicted as arrows. Exons of *MAF2* variant 1 (black boxes) and *MAF3* variant 1 (gray boxes) are depicted for reference purposes.

ing from the *MAF3* locus itself and/or from the paralogous *MAF3* sequences inserted into the *MAF2* (Figure 6). Accessions were chosen randomly and, combined with preexisting accession sequence data, represent a set with a broad distribution across the *A. thaliana* range.

All *MAF3* sequences from noninsertion allele accessions were found in a single, well-supported clade. The *MAF3* sequences from the *MAF3* region of the insertion allele or the linked *MAF3* gene, clustered in another well-supported clade (Figure 6B), suggesting that these accessions share a common ancestor and overlapping geographic distribution of the allele subclasses. The relatedness of these alleles suggest that formation of the diverse *MAF2* alleles was likely a reiterative process resulting in the expansion, via gene conversion, or contraction, via repetitive deletion, of the *MAF3* insertion from a single precursor allele.

Polymorphism in *MAF3* gene is linked to some *MAF2* insertion alleles: The Sha subclass of the *MAF2* insertion alleles is the most numerous, and thus we characterized it

further. Using primers 44 and 45, a canonical RT-PCR banding pattern can be obtained for an accession in this subclass, as exemplified by the BI-1 (Figure 7). However, Sha and Kas-1 accessions displayed variation in their RT-PCR bands. Sha is lacking bands corresponding to rearranged *MAF3* RT-PCR products, and Kas-1 expresses neither *MAF2* nor *MAF3* RT-PCR bands. Further analysis established that the *MAF3* locus is deleted in both Sha and Kas-1 (Figure S3A). The deletion of this region, which is supported by RT-PCR data, was not detected in the published sequence of *MAF2*-Sha (CAICEDO *et al.* 2009). In addition, *MAF2*-Kas-1 has a 466-bp internal deletion (Figure S3B). It is unclear if this deletion prevents the expression of the *MAF2*-Kas-1 RNA; however, the RT-PCR primer sequences are intact. *MAF2*-Sha and *MAF2*-Kas-1 demonstrate that there may be further sequence diversity within the allele subclasses. Thus, putative loss-of-function *MAF3* alleles are linked to some *MAF2* insertion alleles.

Proteins encoded by the *MAF2* insertion alleles: To determine the predicted proteins encoded by these

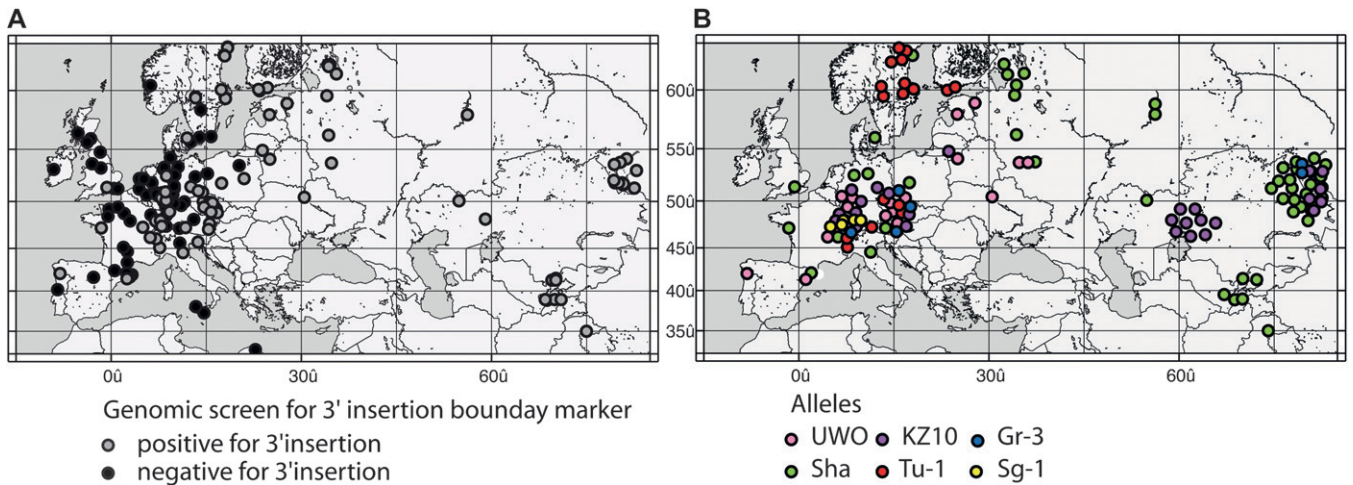


FIGURE 5.—Geographic distribution of accessions surveyed. (A) Location of accessions testing negatively (black circles) and positively (gray circles) for the *MAF3* insertion in *MAF2*. One dot often signifies more than one accession due to proximity of collection locations. (B) Subclasses of the *MAF2* insertion allele class. Accessions are grouped on the basis of sequence data from the 5' insertion boundary and mapped. Circles are slightly shifted where accession locations overlap.

novel alleles, we cloned and sequenced cDNAs encoded by *MAF2*–UWO, *MAF2*–Tu-1 and *MAF2*–Sha alleles. Alternative splicing is characteristic for the *MAF2* locus, and four splice variants have been reported for the *MAF2*–Col allele. Of these four transcripts, only two are amplified with the primers 44 and 45 used for the RT–PCR reaction. Corresponding transcripts of *MAF2* insertion alleles are all chimeras between *MAF2* and *MAF3* (Figure S4). Insertion of *MAF3* sequences, that are also alternatively spliced, contributed to the proliferation of transcripts generated by *MAF2* insertion alleles. Some of the canonical splice sites appear inefficient in the context of the insertion alleles, leading to the synthesis of new splice variants that include changes in transcript composition such as reduction in size of *MAF2* exon 4, skipping of *MAF2* exon 6, and frequently *MAF3* exon 2. However, despite a large transcript sequence dissimilarity of the *MAF2* insertion alleles to *MAF2*–Col, the protein sequences appear more similar due to the incorporation of the stop codon upstream of the *MAF3* insertion in most transcripts. *MAF2*–Col variant 1 is a transcript that encodes the full-length protein (Figure S5). Even though some of the predicted proteins encoded by the *MAF2* insertion alleles appear to contain both MADS-box and K-box conserved domains, there is either a truncation or change in the sequence within the K-box domain that may affect the functionality of these proteins even if they are synthesized and stably retained in the cell. Other *MAF2* insertion transcripts encode proteins identical or similar to *MAF2*–Col variants 2, 3, and 4 that encode a truncated protein containing only MADS-box domain. The function of these proteins is currently unknown.

***MAF2* insertion alleles confer an early flowering phenotype in BC₅ lines:** Given the transcript and predicted protein sequence variation in *MAF2* insertion

alleles, and its known association with flowering time, we asked whether insertion alleles are linked to variation in flowering time. To assess the effects of these alleles in a common background, we examined the flowering time of BC₅ lines, in which the *MAF2* alleles of UWO, Tu-1, Sha, and Kas-1 lines had been introgressed into *Ler*. Plants homozygous for *MAF2* insertion alleles flowered with about two leaves less than *MAF2*–*Ler* homozygote siblings (Figure 8). Thus, despite differences between the insertion alleles, all BC₅ lines displayed an early flowering phenotype under long days. This early flowering phenotype is similar to that observed in the *maf2* loss-of-function allele in Col, indicating that *MAF2* insertion alleles may not support the normal function of *MAF2*.

Association of *MAF2* insertion alleles with early flowering indicates that they are either loss-of-function alleles or that they are linked to another locus that confers early flowering phenotype. To distinguish between these possibilities we used the accessions Wassilewskija (WS), bearing the *MAF2*–WS insertion allele belonging to the *MAF2*–UWO subclass, Tu-1, bearing the *MAF2*–Tu-1 allele, and *Ler*, bearing the *MAF2*–*Ler* allele, as representatives of the *MAF2* insertion and wild-type alleles. Accessions were transformed with an amiRNA construct designed to specifically knock down *MAF2* expression. The amiRNAs targeted a cDNA region present in the full-length *MAF2*–*Ler* transcript and all transcripts detected in *MAF2* insertion allele accessions (see MATERIAL AND METHODS for details). All T₁ amiRNA lines were grown under long- and short-day conditions (Figure 9). The *Ler* transformants expressing the amiRNA–*MAF2* construct had reduced expression of the full-length *MAF2* transcript and flowered earlier than the empty vector control line under both LD and SD conditions (LD, $P = 0.005^{**}$; SD, $P = 0.0001^{***}$).

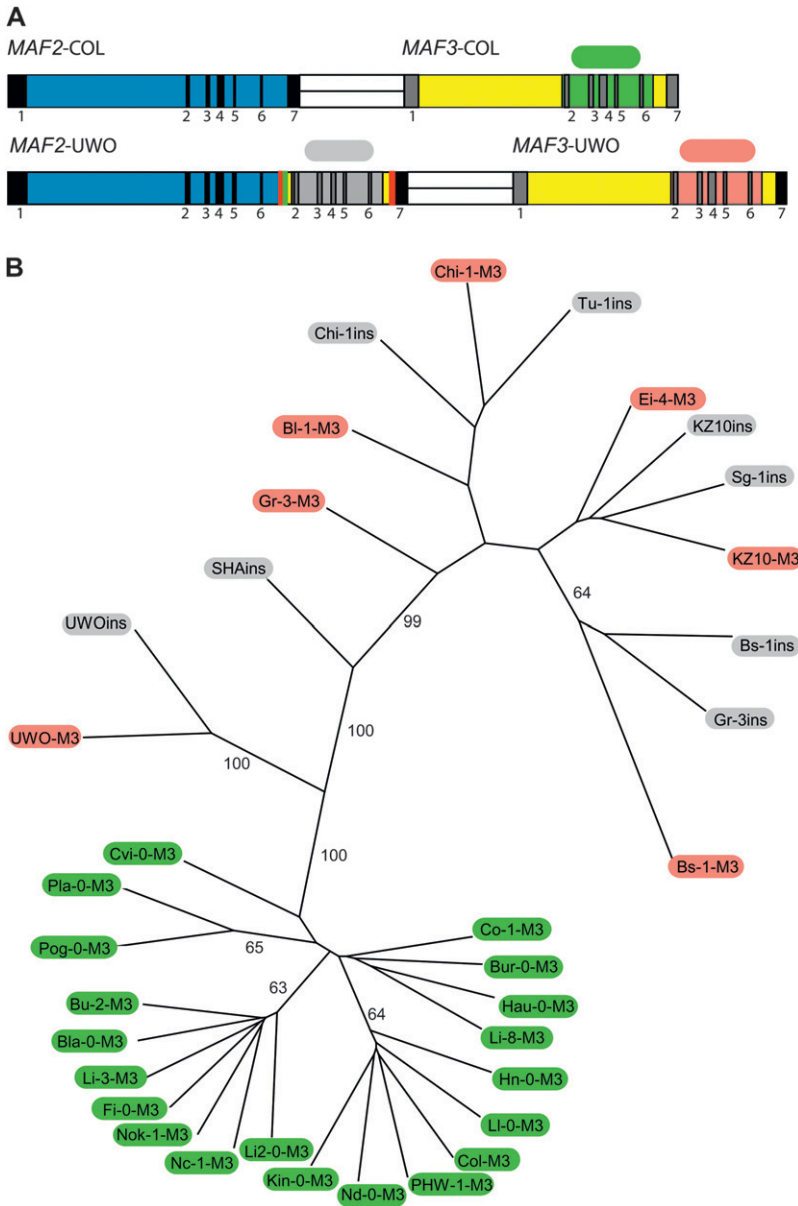


FIGURE 6.—The *MAF2* insertion alleles are phylogenetically related. (A) Location of *MAF3* or *MAF3*-homologous regions sequenced to generate an UPGMA dendrogram. Sequence sites are designated by a colored oval: green, *MAF3* gene sequence proper, from accessions that do not have an insertion of *MAF3* into *MAF2*; pink, *MAF3* gene sequence proper, from accessions containing *MAF2* insertion allele; gray, *MAF3*-homologous region derived from the insertion of *MAF3* into the *MAF2* gene. (B) UPGMA dendrogram of *MAF3* and *MAF3*-homologous sequences. Bootstrap values are reported when they exceed 50%. Colored ovals on branch termini reflect the origin of the *MAF3* sequence, as depicted in A.

However, in the *MAF2*-WS amiRNA or *MAF2*-Tu-1 amiRNA lines, despite the demonstrated ability of the amiRNA to downregulate *MAF2* expression (Figure 9B), the flowering time was indistinguishable from the control plants expressing an empty vector under both LD and SD conditions (WS—LD, $P = 0.986$ NS; SD, $P = 0.140$ NS; Tu-1—LD, $P = 0.44$ NS).

Therefore, the *MAF2*-WS and *MAF2*-Tu-1 are loss-of-function alleles.

Segregation distortion at *MAF2*: Several studies have reported a flowering-time QTL in the vicinity of the *MAF* locus, as well as a segregation distortion favoring the early flowering allele (LOUDET *et al.* 2002; EL-LITHY *et al.* 2004, 2006; SIMON *et al.* 2008). To determine the

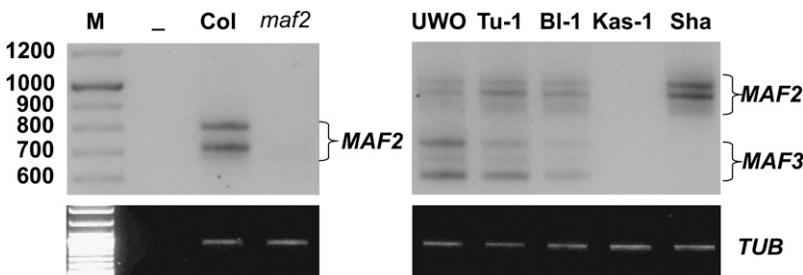


FIGURE 7.—Expression of *MAF2* and *MAF3* across insertion allele subclasses. *MAF2* RT-PCR showing expression divergence among Col, UWO, Tu-1, BI-1, Kas-1, and Sha accessions. RT-PCR was performed with primers 44 and 45.

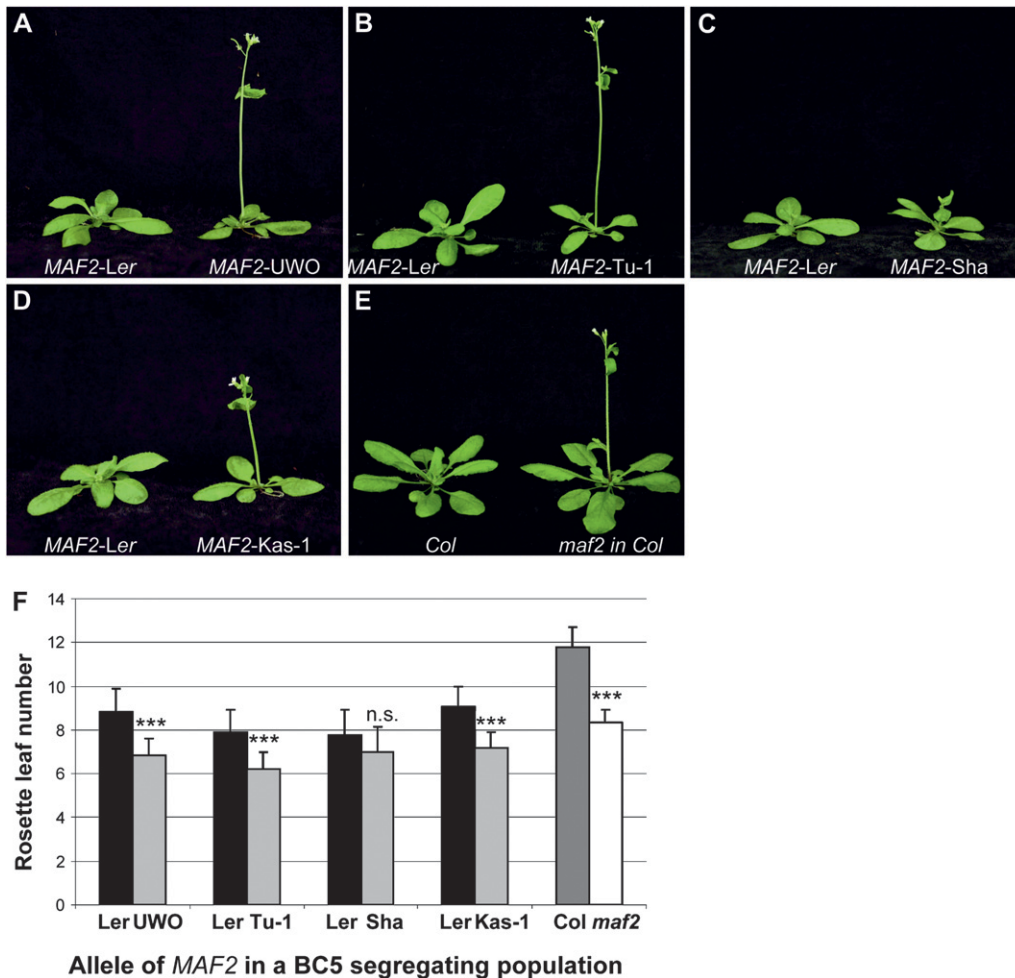


FIGURE 8.—Flowering-time comparison of BC₅ homozygous lines. Each *MAF2* insertion allele line is compared to a *MAF2-Ler* derived from the same BC₅ F₂ population. All photos were taken at an equivalent age. (A) *MAF2-Ler*, *MAF2-UWO*; (B) *MAF2-Ler*, *MAF2-Tu-1*; (C) *MAF2-Ler*, *MAF2-Sha*; (D) *MAF2-Ler*, *MAF2-Kas-1*; (E) Col, *maf2* in Col; (F) flowering time of homozygous plants derived from the BC₅ segregating population. Plants were scored for flowering time by counting leaves produced prior to bolting and subsequently genotyped. Homozygous *MAF2-Ler* progeny, solid bars; *MAF2* insertion allele homozygous progeny, light shaded bars; homozygous *MAF2-Col* plants, dark shaded bars; *maf2* in Col plants, open bars. Statistical difference in flowering time of *MAF2* insertion allele homozygotes compared to *MAF2-Ler* homozygotes is indicated as *** $P < 0.01$ with \pm SD.

frequency of the insertion alleles relative to *MAF2-Ler*, we analyzed the BC₅ F₂ segregating populations. All populations had a higher frequency of insertion alleles than the *MAF2-Ler* allele (Figure 10A and Table S4). Significant deviations from the predicted 1:2:1 ratio was observed in lines segregating for *MAF2-UWO*, *MAF2-Sha*, and *MAF2-Kas-1*. Segregation distortion at *MAF2* was also observed in an independent F₂ population derived from the cross *MAF2-UWO/MAF2-UWO*, *FLC-Ler/FLC-Ler* in *Ler* \times *MAF2-Ler/MAF2-Ler*, *FLC-Col* in *Ler/FLC-Col* in *Ler*. The degree of segregation distortion was similar to that previously observed for the *MAF2-UWO* allele (Figure 10B). However, the segregation ratio for *FLC-Col/FLC-Ler* in this progeny did not deviate from the expected 1:2:1 ratio (Figure 10B, Table S5), suggesting that *MAF2* and *FLC* segregate independently, as expected, and that segregation distortion is associated with the bottom of chromosome V and not the entire chromosome.

DISCUSSION

Common allelic variation underlies flowering-time variation in *A. thaliana*: Although several genes that are

associated with flowering-time variation have been described for *A. thaliana*, common allelic variation has been reported only for *FRI*, *FLC*, and *PHYC* (JOHANSON *et al.* 2000; GAZZANI *et al.* 2003; LE CORRE 2005; LEMPE *et al.* 2005; BALASUBRAMANIAN *et al.* 2006a). Most other naturally occurring alleles of flowering-time genes have been found only in single accessions (AUKERMAN *et al.* 1997; EL-DIN EL-ASSAL *et al.* 2001; WERNER *et al.* 2005; WANG *et al.* 2007). We found that the *MAF2* locus harbors great genetic variability resulting in the generation of transcript complements that differ from the canonical patterns present in Col. Of 147 *MAF2* alleles initially tested, 52 displayed alteration in *MAF2* transcript sizes. The chimeric alleles at the *MAF2* locus were the most numerous. These alleles are associated with early flowering and are present in at least 41% of the strains, providing another example for common alleles contributing to flowering-time variation in *A. thaliana*. Similar to *FRI* and *FLC*, we find extensive heterogeneity at the insertion alleles of *MAF2*. Although several *MAF2* insertion alleles were identified by CAICEDO *et al.* (2009), the unique set of PCR primers used for screening in this study resulted in the identification of additional *MAF2* insertion alleles, showing that variation at this locus is more extensive than

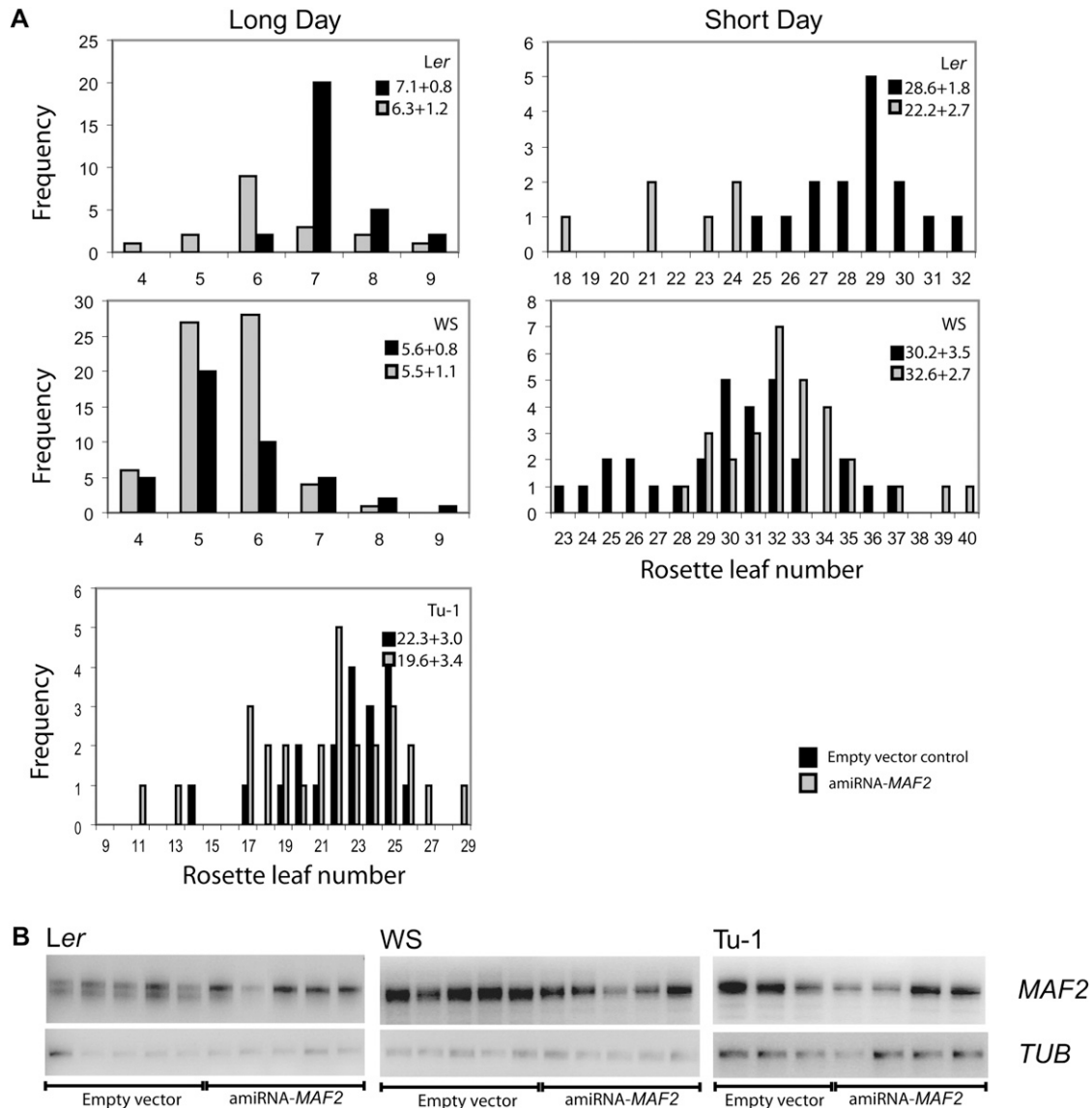


FIGURE 9.—*MAF2* artificial miRNA (*MAF2* amiRNA) does not reduce time to flowering in the WS or Tu-1 accessions. (A) The accession *Ler*, which bears an active allele of *MAF2*, and WS and Tu-1, which bear *MAF2* insertion alleles, were transformed with an empty vector control construct (solid bars) or the *MAF2* amiRNA construct (shaded bars). Flowering time was recorded under long- and short-day conditions as rosette leaf number at flowering. (B) RT-PCR of *MAF2* expression, using primers 44 and 45 in *Ler*, WS, and Tu-1 lines transformed with the empty vector or the *MAF2* amiRNA construct.

previously reported. In the case of the *MAF2* insertion alleles, similar to allelic effects of *FR1* and *FLC*, a common phenotypic effect is observed in spite of the heterogeneity.

The *MAF* locus on chromosome V has been suggested as a cause for natural variation in flowering time in several QTL studies that used parents falling into the *MAF2*-Sha allele subclass. QTL were found in the following populations: *Ler* × *Kas-2*, *Ler* × *Kondara* (EL-LITHY *et al.* 2006), *Sorbo* × *Gy-0* (O'NEILL *et al.* 2008), *Col* × *Sha* (SIMON *et al.* 2008), and *Bay-0* × *Sha* (BOTTO and COLUCCIO 2007). Our BC₅ data confirm the association of a flowering-time QTL in this region. However, the discovery of putative loss-of-function polymorphisms at the *MAF3* gene linked to *MAF2* insertion alleles, or a possibility of mutation in another locus linked to the

MAF region, confounds the assignment of this effect to *MAF2* insertion alleles. To unequivocally associate the early flowering phenotype to changes at the *MAF2* locus, we expressed an amiRNA construct designed to specifically knock down *MAF2* transcripts. While *MAF2* amiRNA constructs showed capability of downregulating *MAF2* gene expression, it conferred early flowering phenotype in *Ler* but not in the WS accession carrying a *MAF2* insertion allele. Failure of the *MAF2* amiRNA construct to alter flowering time in the WS and Tu-1 amiRNA transgenic lines shows that these insertion alleles are loss-of-function, as was previously suggested in the BC₅ lines. The loss-of-function in these alleles could be due to the *MAF3* insertion or to the lower abundance of individual transcript splice variants.

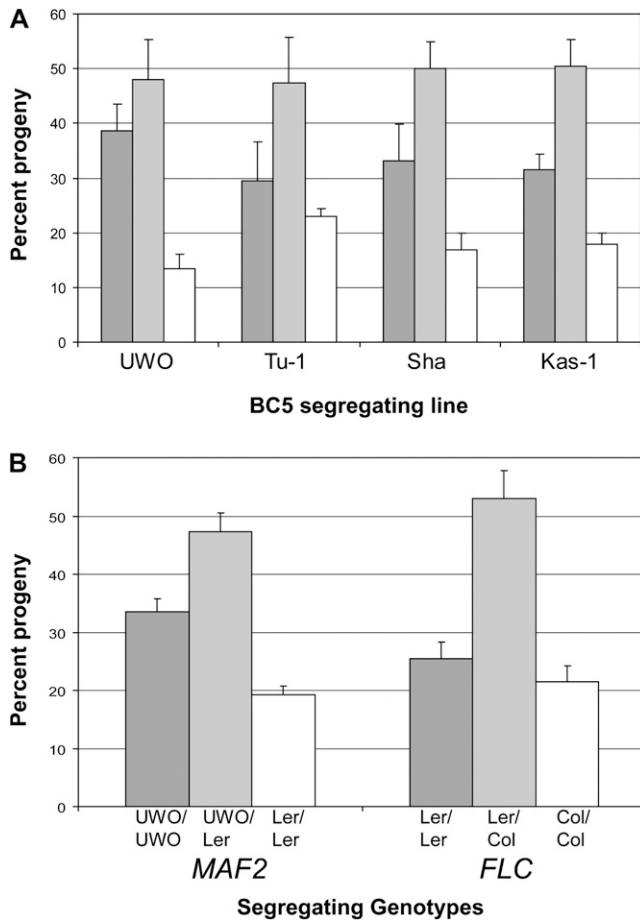


FIGURE 10.—Segregation distortion increases the frequency of *MAF2* insertion alleles in the progeny ratio. (A) Segregation distortion observed in BC₅ F₂ populations segregating for various *MAF2* insertion alleles and *MAF2-Ler*. *MAF2* insertion allele homozygotes are dark shaded bars, plants heterozygous for a *MAF2* insertion allele and *MAF2-Ler* are light shaded bars, and *MAF2-Ler* homozygotes are shown as open bars. (B) Segregation distortion of *MAF2* and *FLC* alleles observed in the BC₄ F₂ progeny of *MAF2-UWO* in *Ler* × *FLC-Col* in *Ler*. *MAF2-UWO* homozygotes are dark shaded bars, *MAF2-Ler* homozygotes are shown as open bars, and heterozygous plants are light shaded bars. *FLC-Ler* homozygotes are dark shaded bars, *FLC-Col* homozygotes are open shaded bars, and heterozygous plants are light shaded bars.

Segregation distortion in *A. thaliana*: In addition to the association of the early flowering-time phenotype to *MAF2* insertion alleles, we observed segregation distortion favoring the *MAF2* insertion alleles in BC₅ F₂ segregating lines. Segregation distortion has been widely reported in *A. thaliana* (LOUDET *et al.* 2002; BIKARD *et al.* 2009) and other plant species (LU *et al.* 2002; PARK *et al.* 2005; KOIDE *et al.* 2008). In *A. thaliana*, chimeric genes, as well as duplicated genes that undergo divergent evolution, have been shown to lead to genetic incompatibilities (BOMBLIES *et al.* 2007; ALCAZAR *et al.* 2009; BIKARD *et al.* 2009). We have repeatedly observed that segregation distortion favors the *MAF2* insertion allele in the progeny of a single, self-

fertilized BC₅ heterozygote. Therefore, segregation distortion is not caused by inadvertent selection during the generation of BC₅ lines, but must have been induced by some other mechanisms that would result in segregation distortion in the subsequent generation.

Since BC₅ lines are expected to be greater than 90% isogenic, the factor disrupting segregation likely resides at or close to the *MAF2* locus. Segregation distortion has been identified in several RILs studies. It coincided with the QTL for early flowering in *Ler* × *Kas-2* RILs, and favored the *Kas-2* allele (EL-LITHY *et al.* 2006), as it did with *Kas-1* in this study. In the *Sha* × *Bay-0* study, a significant segregation distortion favoring the *Sha* genomic region was observed in the region of the *MAF2* locus at the end of chromosome V (LOUDET *et al.* 2002). Significant linkage disequilibrium between a QTL at the *MAF2* locus and another region on chromosome IV was observed in a *Sha* × *Col* RIL population (SIMON *et al.* 2008). These studies support our assertion that a factor capable of distorting the expected Mendelian ratio is linked to the *MAF2* locus in accessions with *MAF2* insertion alleles. Although the identity and precise location of this factor is unknown, linkage of the *MAF2* insertion alleles to a segregation distortion factor may have contributed to the broad geographical distribution of the *MAF2* insertion alleles regardless of the phenotypic effect of these alleles on flowering time.

Evolution of tandemly duplicated genes: Discovery of novel chimeric alleles at the *MAF2* locus is consistent with the mode of TAG sequence evolution observed in other organisms. Our analysis revealed that all *MAF2* alleles with *MAF3* insertions produced chimeric transcripts that were usually not observed in *Col*. However, alternative splicing of exons 3 and 4 in the *MAF2* pre-mRNA created some transcripts that are lacking the *MAF3* insertion sequences. Similar conditional penetrance of chimeric exons through alternative splicing has been observed previously (CUSACK and WOLFE 2007a,b), demonstrating that gene rearrangements coupled with changes in alternative splicing provide a good substrate for functional evolution of genes.

Geographic clustering of *MAF2* alleles: Most chimeric *MAF2* allele forms have a clustered geographical distribution, suggesting that each allele spread from the site of its origin. Distinct population structure at a single locus in *A. thaliana* is unusual (SHARBEL *et al.* 2000; HOFFMANN *et al.* 2003; SCHMUTHS *et al.* 2004), but has been occasionally observed in multilocus studies (NORDBORG *et al.* 2005; SCHMID *et al.* 2005; BECK *et al.* 2008; FRANCOIS *et al.* 2008). For example, SCHMUTHS *et al.* (2004) found distinct population structure at only 2 of 67 CAPS polymorphisms. Like in this study, the identified markers displayed an east–west dichotomy. The east–west dichotomy is considered to be the result of postglacial expansion from eastern and western refugial populations after the last glaciation, possibly

assisted by the expansion of Neolithic agriculture (BECK *et al.* 2008; FRANCOIS *et al.* 2008). Population expansion may partially explain the distribution and fixation of the *MAF2* insertion alleles in eastern Eurasia.

It is possible that segregation distortion also contributed to numerical prominence of these alleles in eastern Eurasia during population expansion. A haplotype that dominates the segregation ratio will have a biased representation in progeny. Segregation bias may ultimately lead to the segregation distortion of linked genes in a population. This “genic meiotic drive” is considered a form of selection (LYTTLE 1991). If segregation distortion acted in the natural population in the same way as it operated in the lab (this study), then it could have contributed to the fixation of these alleles in eastern Eurasia (SCHIERUP *et al.* 2008; SCHIERUP and VEKEMANS 2008). Coupling a segregation distorter to a population expansion would provide a potent evolutionary force (MORITA *et al.* 1992).

We thank Josip Perkovic for help with plant work and Richard Clark, Kirsten Bomblies, Daniel Ortiz-Barrientos, and Patrice Salomé for comments and critical reading of the manuscript. We thank the National Science Foundation-supported Arabidopsis Biological Resource Centre (ABRC), the European Arabidopsis Stock Centres (NASC) for seeds, and Oliver Ratcliffe for the gift of *maf2* seeds. This work was supported by an European Molecular Biology Organization (EMBO) Long Term fellowship (S.B.), a Gottfried Wilhelm Leibniz Award of the DFG and the Max Planck Society (D.W.), and the Natural Sciences and Engineering Research Council of Canada (NSERC) (V.G.). Sequence data from this article have been deposited with the GenBank Data Libraries under consecutive accession numbers from HM487066 to HM487102.

LITERATURE CITED

- ALCAZAR, R., A. V. GARCIA, J. E. PARKER and M. REYMOND, 2009 Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc. Natl. Acad. Sci. USA* **106**: 334–339.
- ARMISEN, D., A. LECHARNY and S. AUBOURG, 2008 Unique genes in plants: specificities and conserved features throughout evolution. *BMC Evol. Biol.* **8**: 280.
- AUKERMAN, M. J., M. HIRSCHFELD, L. WESTER, M. WEAVER, T. CLACK *et al.*, 1997 A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant Cell* **9**: 1317–1326.
- BAKKER, E. G., C. TOOMAJIAN, M. KREITMAN and J. BERGELSON, 2006 A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell* **18**: 1803–1818.
- BALASUBRAMANIAN, S., S. SURESHKUMAR, M. AGRAWAL, T. P. MICHAEL, C. WESSINGER *et al.*, 2006a The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat. Genet.* **38**: 711–715.
- BALASUBRAMANIAN, S., S. SURESHKUMAR, J. LEMPE and D. WEIGEL, 2006b Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet.* **2**: e106.
- BECK, J. B., H. SCHMUTHS and B. A. SCHAAL, 2008 Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.* **17**: 902–915.
- BECKER, A., and G. THEISSEN, 2003 The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **29**: 464–489.
- BIKARD, D., D. PATEL, C. LE METTE, V. GIORGI, C. CAMILLERI *et al.*, 2009 Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**: 623–626.
- BLANC, G., K. HOKAMP and K. H. WOLFE, 2003 A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**: 137–144.
- BOMBLIES, K., J. LEMPE, P. EPPLE, N. WARTHMAN, C. LANZ *et al.*, 2007 Autoimmune response as a mechanism for a Dobzhansky–Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5**: e236.
- BOTTO, J. F., and M. P. COLUCCIO, 2007 Seasonal and plant-density dependency for quantitative trait loci affecting flowering time in multiple populations of *Arabidopsis thaliana*. *Plant Cell Environ.* **30**: 1465–1479.
- CAICEDO, A. L., J. R. STINCHCOMBE, K. M. OLSEN, J. SCHMITT and M. D. PURUGGANAN, 2004 Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci. USA* **101**: 15670–15675.
- CAICEDO, A. L., C. RICHARDS, I. M. EHRENREICH and M. D. PURUGGANAN, 2009 Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana MAF2-5* flowering time gene cluster. *Mol. Biol. Evol.* **26**: 699–711.
- CLARK, R. M., G. SCHWEIKERT, C. TOOMAJIAN, S. OSSOWSKI, G. ZELLER *et al.*, 2007 Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.
- CUSACK, B. P., and K. H. WOLFE, 2007a Not born equal: increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. *Mol. Biol. Evol.* **24**: 679–686.
- CUSACK, B. P., and K. H. WOLFE, 2007b When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* **23**: 270–272.
- CUTLER, G., L. A. MARSHALL, N. CHIN, H. BARIBAULT and P. D. KASSNER, 2007 Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* **17**: 1743–1754.
- DURET, L., E. GASTEIGER and G. PERRIERE, 1996 LALNVIEW: a graphical viewer for pairwise sequence alignments. *Comput. Appl. Biosci.* **12**: 507–510.
- EDWARDS, K., C. JOHNSTONE and C. THOMPSON, 1991 A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* **19**: 1349.
- EL-DIN EL-ASSAL, S., C. ALONSO-BLANCO, A. J. PEETERS, V. RAZ and M. KOORNNEEF, 2001 A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nat. Genet.* **29**: 435–440.
- EL-LITHY, M. E., E. J. CLERKX, G. J. RUYLS, M. KOORNNEEF and D. VREUGDENHIL, 2004 Quantitative trait locus analysis of growth-related traits in a new *Arabidopsis* recombinant inbred population. *Plant Physiol.* **135**: 444–458.
- EL-LITHY, M. E., L. BENTSINK, C. J. HANHART, G. J. RUYLS, D. ROVITO *et al.*, 2006 New *Arabidopsis* recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics* **172**: 1867–1876.
- FELSENSTEIN, J., 2005 *PHYLIP (Phylogeny Inference Package)*, version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- FLOREA, L., G. HARTZELL, Z. ZHANG, G. M. RUBIN and W. MILLER, 1998 A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**: 967–974.
- FRANCOIS, O., M. G. BLUM, M. JAKOBSSON and N. A. ROSENBERG, 2008 Demographic history of european populations of *Arabidopsis thaliana*. *PLoS Genet.* **4**: e1000075.
- FREELING, M., E. LYONS, B. PEDERSEN, M. ALAM, R. MING *et al.*, 2008 Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Res.* **18**: 1924–1937.
- GANLEY, A. R., and T. KOBAYASHI, 2007 Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* **17**: 184–191.
- GAO, L. Z., and H. INMAN, 2004 Very low gene duplication rate in the yeast genome. *Science* **306**: 1367–1370.
- GAUT, B. S., S. I. WRIGHT, C. RIZZON, J. DVORAK and L. K. ANDERSON, 2007 Recombination: an underappreciated factor in the evolution of plant genomes. *Nat. Rev. Genet.* **8**: 77–84.
- GAZZANI, S., A. R. GENDALL, C. LISTER and C. DEAN, 2003 Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* **132**: 1107–1114.
- GU, J., and X. GU, 2003 Natural history and functional divergence of protein tyrosine kinases. *Gene* **317**: 49–57.

- HARRISON, P. M., and M. GERSTEIN, 2002 Studying genomes through the acons: protein families, pseudogenes and proteome evolution. *J. Mol. Biol.* **318**: 1155–1174.
- HOFFMANN, M. H., A. S. GLASS, J. TOMIUK, H. SCHMUTHS, R. M. FRITSCH *et al.*, 2003 Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). *Mol. Ecol.* **12**: 1007–1019.
- JELESKO, J. G., R. HARPER, M. FURUYA and W. GRUISSEM, 1999 Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**: 10302–10307.
- JELESKO, J. G., K. CARTER, W. THOMPSON, Y. KINOSHITA and W. GRUISSEM, 2004 Meiotic recombination between paralogous RBCSB genes on sister chromatids of *Arabidopsis thaliana*. *Genetics* **166**: 947–957.
- JOHANSON, U., J. WEST, C. LISTER, S. MICHAELS, R. AMASINO *et al.*, 2000 Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**: 344–347.
- KOBUJI, R., N. SUMIKAWA, M. YAMASAKI, K. KONDO, K. UEDA *et al.*, 2003 Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol. Biol. Evol.* **20**: 1963–1977.
- KOIDE, Y., M. IKENAGA, N. SAWAMURA, D. NISHIMOTO, K. MATSUBARA *et al.*, 2008 The evolution of sex-independent transmission ratio distortion involving multiple allelic interactions at a single locus in rice. *Genetics* **180**: 409–420.
- KONG, H., L. L. LANDHERR, M. W. FROHLICH, J. LEEBENS-MACK, H. MA *et al.*, 2007 Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J.* **50**: 873–885.
- KORBEL, J. O., P. M. KIM, X. CHEN, A. E. URBAN, S. WEISSMAN *et al.*, 2008 The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Curr. Opin. Struct. Biol.* **18**: 366–374.
- KUANG, H., S. S. WOO, B. C. MEYERS, E. NEVO and R. W. MICHELMORE, 2004 Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**: 2870–2894.
- KUANG, H., O. E. OCHOA, E. NEVO and R. W. MICHELMORE, 2006 The disease resistance gene *Dm3* is infrequent in natural populations of *Lactuca serriola* due to deletions and frequent gene conversions at the RGC2 locus. *Plant J.* **47**: 38–48.
- KUANG, H., K. S. CALDWELL, B. C. MEYERS and R. W. MICHELMORE, 2008 Frequent sequence exchanges between homologs of *RPP8* in *Arabidopsis* are not necessarily associated with genomic proximity. *Plant J.* **54**: 69–80.
- LABARGA, A., F. VALENTIN, M. ANDERSON and R. LOPEZ, 2007 Web services at the European bioinformatics institute. *Nucleic Acids Res.* **35**: W6–W11.
- LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALEY *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- LE CORRE, V., 2005 Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* **14**: 4181–4192.
- LEISTER, D., 2004 Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.* **20**: 116–122.
- LEMPE, J., S. BALASUBRAMANIAN, S. SURESHKUMAR, A. SINGH, M. SCHMID *et al.*, 2005 Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet.* **1**: 109–118.
- LINCOLN, S., M. DALY and E. LANDER, 1992 Constructing genetic maps with MAPMAKER/EXP, version 3.0. Whitehead Institute Technical Report, Whitehead Institute, Cambridge, MA.
- LOUDET, O., S. CHAILLOU, C. CAMILLERI, D. BOUCHEZ and F. DANIEL-VEDELE, 2002 Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor. Appl. Genet.* **104**: 1173–1184.
- LYTTLE, T.W., 1991 Segregation distorters. *Annu. Rev. Genet.* **25**: 511–557.
- MICHAELS, S. D., and R. M. AMASINO, 1999 *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949–956.
- MILLER, M. P., 2005 Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *J. Hered.* **96**: 722–724.
- MONDRAGON-PALOMINO, M., and B. S. GAUT, 2005 Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **22**: 2444–2456.
- MONDRAGON-PALOMINO, M., B. C. MEYERS, R. W. MICHELMORE and B. S. GAUT, 2002 Patterns of positive selection in the complete *NBS-LRR* gene family of *Arabidopsis thaliana*. *Genome Res.* **12**: 1305–1315.
- MORITA, T., H. KUBOTA, K. MURATA, M. NOZAKI, C. DELARBRE *et al.*, 1992 Evolution of the mouse t haplotype: recent and worldwide introgression to *Mus musculus*. *Proc. Natl. Acad. Sci. USA* **89**: 6851–6855.
- NARAYANAN, V., P. A. MIECZKOWSKI, H. M. KIM, T. D. PETES and K. S. LOBACHEV, 2006 The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* **125**: 1283–1296.
- NEI, M., and A. P. ROONEY, 2005 Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**: 121–152.
- NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**: e196.
- O'NEILL, C. M., C. MORGAN, J. KIRBY, H. TSCHOEP, P. X. DENG *et al.*, 2008 Six new recombinant inbred populations for the study of quantitative traits in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **116**: 623–634.
- PARENICOVA, L., S. DE FOLTER, M. KIEFFER, D. S. HORNER, C. FAVALLI *et al.*, 2003 Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* **15**: 1538–1551.
- PARK, Y.-H., M. ALABADY, M. ULLOA, B. SICKLER, T. WILKINS *et al.*, 2005 Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol. Genet. Genomics* **274**: 428–441.
- RATCLIFFE, O. J., R. W. KUMIMOTO, B. J. WONG and J. L. RIECHMANN, 2003 Analysis of the *Arabidopsis* MADS *AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. *Plant Cell* **15**: 1159–1169.
- REDON, R., S. ISHIKAWA, K. R. FITCH, L. FEUK, G. H. PERRY *et al.*, 2006 Global variation in copy number in the human genome. *Nature* **444**: 444–454.
- RICE, P., I. LONGDEN and A. BLEASBY, 2000 EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**: 276–277.
- SAKURAI, T., G. PLATA, F. RODRIGUEZ-ZAPATA, M. SEKI, A. SALCEDO *et al.*, 2007 Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biol.* **7**: 66.
- SCHIERUP, M. H., and X. VEKEMANS, 2008 Genomic consequences of selection on self-incompatibility genes. *Curr. Opin. Plant Biol.* **11**: 116–122.
- SCHIERUP, M. H., J. S. BECHSGAARD and F. B. CHRISTIANSEN, 2008 Selection at work in self-incompatible *Arabidopsis lyrata*. II. Spatial distribution of S haplotypes in Iceland. *Genetics* **180**: 1051–1059.
- SCHMID, K. J., S. RAMOS-ONSINS, H. RINGYS-BECKSTEIN, B. WEISSHAAR and T. MITCHELL-OLDS, 2005 A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169**: 1601–1615.
- SCHMUTHS, H., M. H. HOFFMANN and K. BACHMANN, 2004 Geographic distribution and recombination of genomic fragments on the short arm of chromosome 2 of *Arabidopsis thaliana*. *Plant Biol.* **6**: 128–139.
- SCHWAB, R., S. OSSOWSKI, M. RIESTER, N. WARTHMAN and D. WEIGEL, 2006 Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**: 1121–1133.
- SCORTECCI, K., S. D. MICHAELS and R. M. AMASINO, 2003 Genetic interactions between FLM and other flowering-time genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* **52**: 915–922.
- SHAKHNOVICH, B. E., and E. V. KOONIN, 2006 Origins and impact of constraints in evolution of gene families. *Genome Res.* **16**: 1529–1536.
- SHARBEL, T. F., B. HAUBOLD and T. MITCHELL-OLDS, 2000 Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol.* **9**: 2109–2118.

- SHINDO, C., M. J. ARANZANA, C. LISTER, C. BAXTER, C. NICHOLLS *et al.*, 2005 Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* **138**: 1163–1173.
- SIMON, M., O. LOUDET, S. DURAND, A. BERARD, D. BRUNEL *et al.*, 2008 Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* **178**: 2253–2264.
- SLACK, A., P. C. THORNTON, D. B. MAGNER, S. M. ROSENBERG and P. J. HASTINGS, 2006 On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.* **2**: e48.
- SOLTIS, D. E., and P. S. SOLTIS, 2003 The role of phylogenetics in comparative genetics. *Plant Physiol.* **132**: 1790–1800.
- STAHL, E. A., G. DWYER, R. MAURICIO, M. KREITMAN and J. BERGELSON, 1999 Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis*. *Nature* **400**: 667–671.
- STURTEVANT, A. H., 1925 The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics* **10**: 117–147.
- SUNG, S., Y. HE, T. W. ESHOO, Y. TAMADA, L. JOHNSON *et al.*, 2006 Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires *LIKE HETEROCHROMATIN PROTEIN 1*. *Nat. Genet.* **38**: 706–710.
- WAGNER, A., 2008 Gene duplications, robustness and evolutionary innovations. *Bioessays* **30**: 367–373.
- WANG, Q., U. SAJJA, S. ROSLOSKI, T. HUMPHREY, M. C. KIM *et al.*, 2007 HUA2 caused natural variation in shoot morphology of *A. thaliana*. *Curr. Biol.* **17**: 1513–1519.
- WERNER, J. D., J. O. BOREVITZ, N. WARTHMAN, G. T. TRAINER, J. R. ECKER *et al.*, 2005 Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci. USA* **102**: 2460–2465.
- XU, S., T. CLARK, H. ZHENG, S. VANG, R. LI *et al.*, 2008 Gene conversion in the rice genome. *BMC Genomics* **9**: 93.
- YANDEAU-NELSON, M. D., Y. XIA, J. LI, M. G. NEUFFER and P. S. SCHNABLE, 2006 Unequal sister chromatid and homolog recombination at a tandem duplication of the *a1* locus in maize. *Genetics* **173**: 2211–2226.
- ZHANG, L., and B. S. GAUT, 2003 Does recombination shape the distribution and evolution of tandemly arrayed genes (TAGs) in the *Arabidopsis thaliana* genome? *Genome Res.* **13**: 2533–2540.

Communicating editor: B. BARTEL

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.116392/DC1>

Natural Diversity in Flowering Responses of *Arabidopsis thaliana* Caused by Variation in a Tandem Gene Array

**Sarah Marie Rosloski, Sathya Sheela Jali, Sureshkumar Balasubramanian,
Detlef Weigel and Vojislava Grbic**

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.110.116392

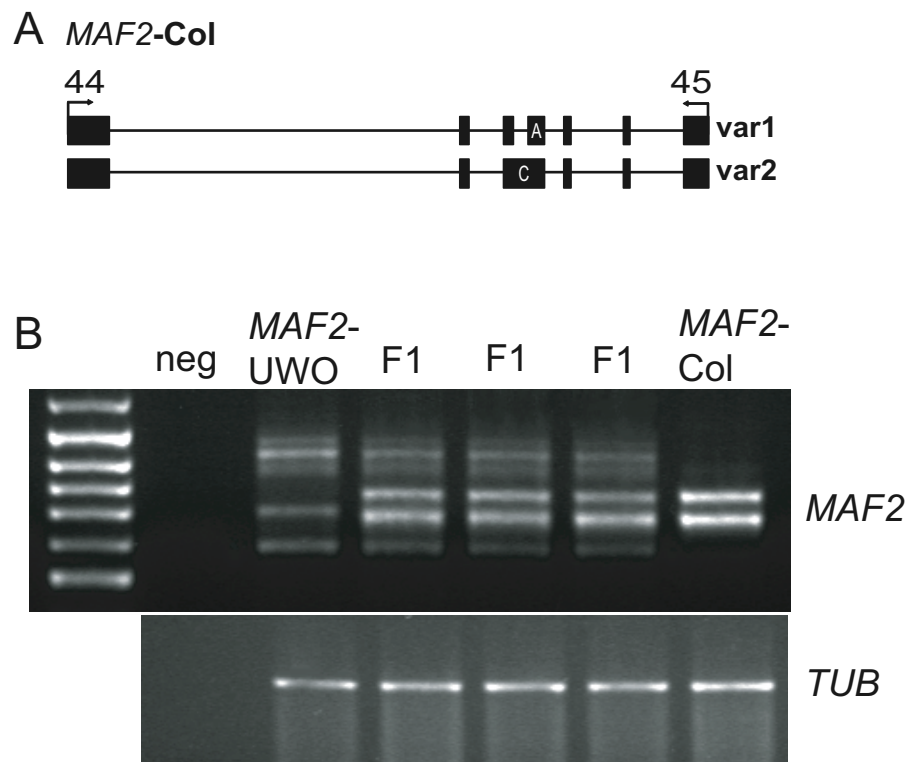


FIGURE S1.—The banding profile of the *MAF2-UWO* variant is heritable and co-dominant. (A) Primer 44 and 45 sites and RT-PCR products produced in *MAF2-Col*. (B) RT-PCR banding profile of the *MAF2-UWO* variant, *MAF2-Col* and three independent F₁ progeny of these parents.

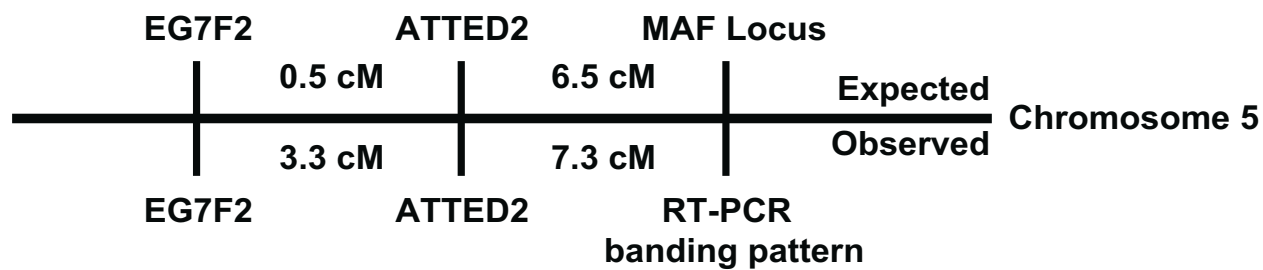


FIGURE S2.—The RT-PCR banding pattern of the *MAF2-UWO* variant maps in the vicinity of the *MAF* cluster. Expected distances were obtained from the Lister and Dean map (<http://www.arabidopsis.org>).

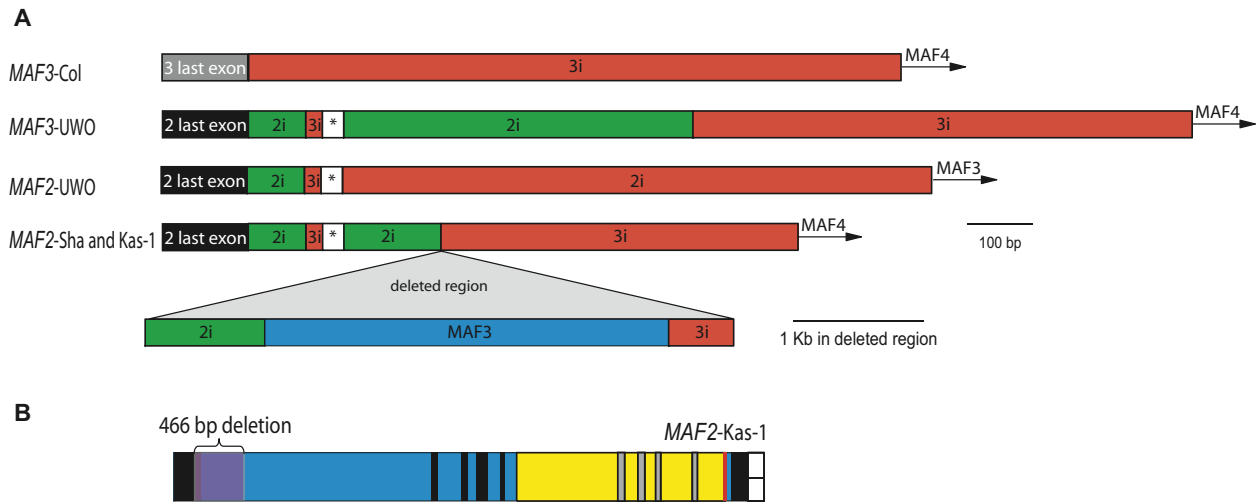


FIGURE S3.—(A) Sha and Kas-1 DNA sequence downstream of the *MAF2* coding region shows that *MAF3* has been deleted in these two accessions. 2i and 3i, are sequences found downstream of the *MAF2* or *MAF3* loci in Col, respectively; *, refers to a region that cannot be definitively assigned to *MAF2* or *MAF3* because of the shared sequence identity. (B) *MAF2*-Kas-1 allele showing the location of the 466 bp deletion.

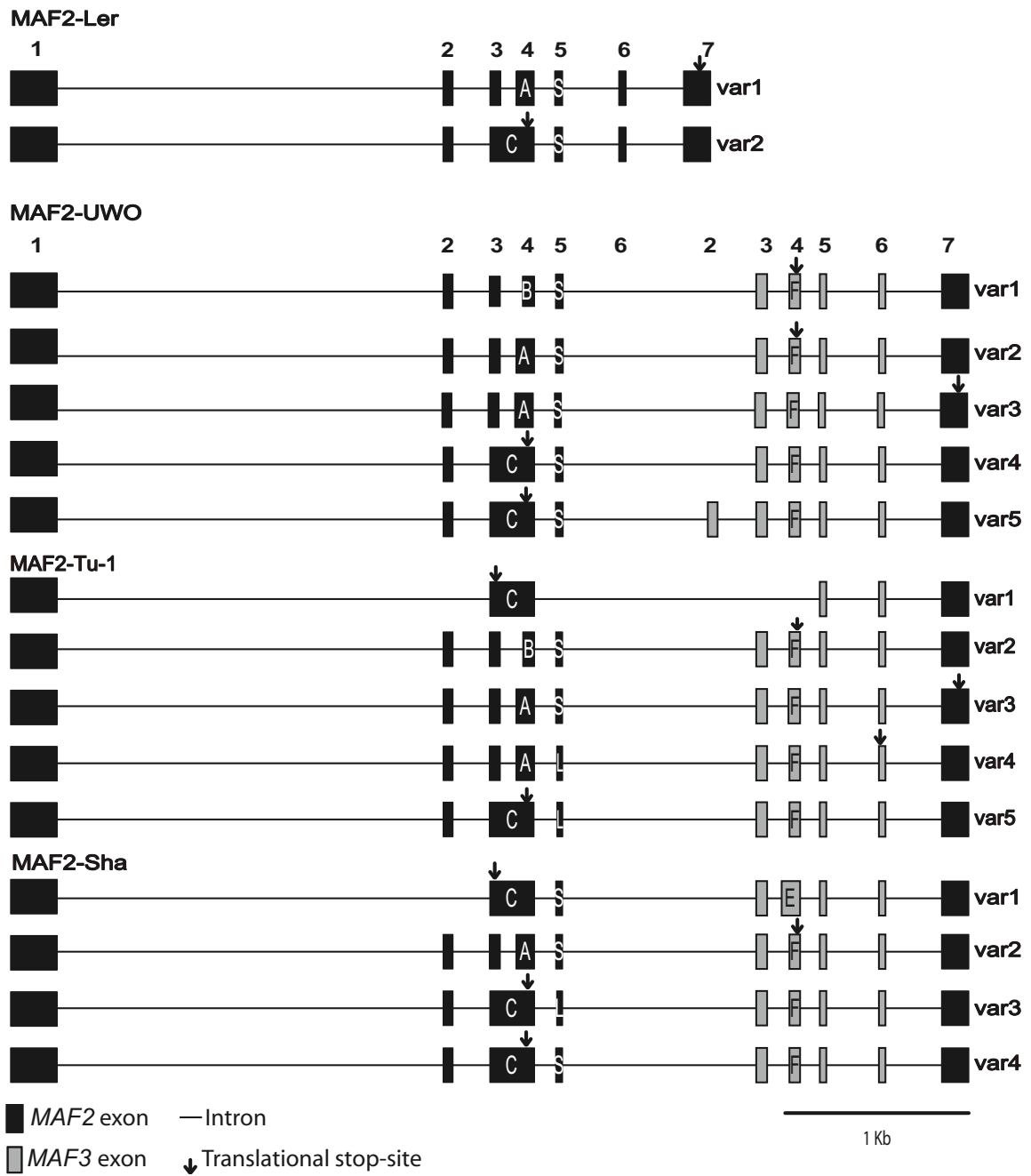


FIGURE S4.—cDNA sequence comparison between *MAF2-Col*, *MAF2-Ler*, *MAF2-UWO*, *MAF2-Sha* and *MAF2-Tu-1*. Both *MAF2-Sha* and *MAF2-Tu-1* cDNA sequences were aligned to the genomic region of *MAF2-UWO* to facilitate comparison of splice-site selection amongst the alleles.

Col var1-like proteins

Col var1 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Tu-1 var2 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 UWO var1 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 UWO var2 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Tu-1 var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Sha var2 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Tu-1 var3 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 UWO var3 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG

Col var1 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 Tu-1 var2 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQS-----VDTLISL
 UWO var1 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQS-----VDTLISL
 UWO var2 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 Tu-1 var4 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 Sha var2 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 Tu-1 var3 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 UWO var3 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL

Col var1 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 Tu-1 var2 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 UWO var1 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 UWO var2 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 Tu-1 var4 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 Sha var2 EEQLETALSVTRARKTELMMGVEVKS LQKT-----
 Tu-1 var3 EEQLETALSVTRARKTELMMGVEVKS LQKTDLAEKIRNYLPHKELLEIVQSVDSLISMEEQ
 UWO var3 EEQLETALSVTRARKTELMMGVEVKS LQKTDLAEKIRNYLPHKELLEIVQSVDSLISMEEQ

Col var1 -----ENLLREENQTLASQV GKKTF LVI EGDRGMSWENG
 Tu-1 var2 -----DLAEKIRNYLPHKELLEIVQR---FSNI
 UWO var1 -----DLAEKIRNYLPHKELLEIVQR---FSNI
 UWO var2 -----DLAEKIRNYLPHKELLEIVQR---FSNI
 Tu-1 var4 -----DLAEKIRNYLPHKELLEIVQR---FSNI
 Sha var2 -----DLAEKIRNYLPHKELLEIVQR---FSNI
 Tu-1 var3 LETALSVIRAKKTELMMEDMKSLQEREKLLI EENQTLASQV GKKTF LVI EGDRGMSWENG
 UWO var3 LETALSVIRAKKTELMMEDMKSLQEREKLLI EENQTLASQV GKKTF LVI EGDRGMSWENG

Col var1 SGNKVRETLPLLK
 Tu-1 var2 YGGTARDCSVSN-
 UWO var1 YGGTARDCSVSN-
 UWO var2 YGGTARDCSVSN-
 Tu-1 var4 YGGTARDCSVSN-
 Sha var2 YGGTARDCSVSN-
 Tu-1 var3 SGNKVRETLPLLK
 UWO var3 SGNKVRETLPLLK

Col var2-like proteins

Col var2 MGRKKVEIKRI NKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Sha var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Sha var3 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 UWO var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 UWO var5 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG
 Tu-1 var5 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVS GSGKLYKSASG

Col var2 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF
 Sha var4 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF
 Sha var3 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF
 UWO var4 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF
 UWO var5 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF
 Tu-1 var5 DNMSKIIDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLP LLLMKNTFF

Col var2 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Sha var4 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Sha var3 FLFFWRIMNTASLKNQMSIMQVWIL-----
 UWO var4 FLFFWRIMNTASLKNQMSIMQVWILQFLWRNSSRLLCP
 UWO var5 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Tu-1 var5 FFSFGEL-----

Col var3-like proteins

Col var3 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Col var2 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Sha var3 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Sha var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 UWO var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 UWO var5 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Tu-1 var5 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG


Col var2 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 Col var3 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 Sha var3 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 Sha var4 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 UWO var4 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 UWO var5 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF
 Tu-1 var5 DNMSKI IDRYEIHHADELEALDLAEKTRNYLPLKELLEIVQRLAQRHFYLPLLLMKNTFF

Col var2 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Col var3 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Sha var3 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Sha var4 FLFFWRIMNTASLKNQMSIMQVWIL-----
 UWO var4 FLFFWRIMNTASLKNQMSIMQVWILQFLWRNSSRLLCPL
 UWO var5 FLFFWRIMNTASLKNQMSIMQVWIL-----
 Tu-1 var5 FFSFCELE-----

Col var4-like proteins

Col var4 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Sha var1 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
 Tu-1 var1 MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG

Col var4 DNMSK---IIDR---Y
 Sha var1 DKILQKKLGIICHKSY
 Tu-1 var1 DKILQKKLGIICHTRSY

MADS-box domain 


K-box domain 

FIGURE S5.—Predicted protein products of *MAF2*-Col, *MAF2*-UWO, *MAF2*-Sha, *MAF2*-Tu-1, *MAF3*-Col and *MAF3*-UWO. Yellow box, MADS-box domain; Orange box, K-box domain.

TABLE S1**List of *Arabidopsis* accessions used in the *MAF2* Expression Screen, n=147.****Group 1: 23 accessions**

Ang-1, BG1, BG4, BG6, CEN-0, CIBC1 CIBC10, CSHL1, FM10, FM11, Gre-0, HS1, Kin-0, NFC1, NFC10, NFE1, REN1, Ri-0, RP1, RP10, Sf-2, Sf-2e, **Zu-0**

Group 2: 72 accessions

Bch-3, Bd-0, Bla-4, Bla-12, Br-0, Bs-1, Bs-2, Bu-11, Bur-0, C24, Cal-0, Co-1, **Col**, Cvi-0, Db-0, Ei-6, EDI-0, El-0, Er-0, Est-1, Et-0, Fi-0, Fr-2, Fr-6, Ga-0, Gd-1, Ge-1, Got1, Got10, Gr-1, Gr-6, Gu-0, Hh-0, Hl-0, Hl-3, , Hs-0, Is-0, Kil-0, Kl-0, La-1, *Ler*, Li-2, Li-2:1, Ll-11, M3385S, M7943s, Mc-0, Mc-1, Mh-1, Mz-0, Nc-1, Nd-1, Nok-0, Nok-1, Nok-2, Nok-3, Np-0, Nw-1, Ob-1, Old-2, Ove-0, Pa-2, Pi-0, SQ4, Ste-0, Tsu-1, Ty-0, UK3, UK4, Wc-1 Wl-0, Wu-0

Group 3: 13 accessions

Bla-1, Bla-2, Bla-5, Bla-6, Bla-11, **Pla-0**, Pla-2, Pog-0, Ra-0, Se-0, Ts-1, Ts-5, Ts-6

Group 4: 36 accessions

Ak-1, Blh-1, Blh-2, Bs-5, Cha-0, Chi-0, Chi-2, Dr-0, Dra-1, Fe-1, Ge-2, Gr-3, Hodja-Obi, In-0, Jl-1, Jm-1, Kn-0, Kondara, KZ10, Lip-0, Lo-1, M73235, Mir-0, Nw-3, Ost-0, Per-2, Per-3, RLD1, Rsch-0, Stw-0, Sn(5)-1, Ta-0, Te-0, UK2, **UWO**, Wei-0, Wil-1

Group 5: 3 accessions

Fr-4, **Ll-2**, Mv-0

Representative accessions from Figure 1 are in bold text.

TABLE S2**List of *Arabidopsis* accessions and their geographical coordinates used in the genomic DNA screen**

Table S2 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.110.116392/DC1>.

TABLE S3**Characteristics of *MAF2* insertion allele subclasses**

<i>MAF2</i> Insertion Allele Subclass	Number of Accessions in Genomic Screen	Length of <i>MAF3</i> Insertion, bp	Length of <i>MAF2</i> deletion ^a , bp
Sg-1	5	4155	834
Gr-3	6	2728	1336
UWO	18	1371	0
Tu-1	20	2220	834
KZ10	29	1557	172
Sha	54	1883	496

^aAll insertion allele subclasses, except *MAF2*-UWO, have both an insertion of the *MAF3* gene sequence into the last exon of *MAF2* and a deletion of *MAF2* genomic sequence adjacent to the *MAF3* insertion.

TABLE S4

Segregation distortion observed in the BC₅ F₂ progeny of crosses between UWO, Tu-1, Sha and Kas-1 accessions and *Ler*

	Observed	Expected	n	x ²	p-value
	1:2:1	1:2:1			
UWO-1	22:18:8	12:24:12	48	11.167	0.00376**
UWO-2	43:53:17	28:57:28	113	12.398	0.0020312**
UWO-3	25:34:10	17:35:17	69	6.5362	0.038078*
UWO-4	25:34:8	17:34:17	67	8.6418	0.013288*
UWO					
total	115:139:43	74:149:74	297	36.125	0.00000001431***
Tu-1-1	23:24:15	16:31:16	62	5.2258	0.073321 ns
Tu-1-2	15:36:14	16:33:16	65	0.78462	0.6755 ns
Tu-1-3	16:27:13	14:28:14	56	0.39286	0.82166 ns
Tu-1					
total	54:87:42	46:92:45	183	3.2951	0.19252 ns
Sha-1	25:29:11	16:33:16	65	6.7846	0.033631*
Sha-2	23:33:9	16:33:16	65	6.0462	0.048651*
Sha-3	14:30:11	14:18:14	55	0.78182	0.67644 ns
Sha-total	62:92:31	46:93:46	185	10.395	0.0055315**
Kas-1-1	18:35:10	16:32:16	63	2.8095	0.24543 ns
Kas-1-2	18:28:10	14:28:14	56	2.2857	0.31891 ns
Kas-1-3	17:23:10	13:25:13	50	2.28	0.31982 ns
Kas-1					
total	53:86:30	42:85:42	169	6.3136	0.042562*

TABLE S5**Segregation distortion observed in the BC₄F₂ progeny of a cross between *MAF2-UWO* in *Ler* x *FLC-Col* in *Ler***

Segregating marker	n		progeny ratio			x ²	p-value
			UWO/UWO	UWO/ <i>Ler</i>	<i>Ler</i> / <i>Ler</i>		
<i>MAF2-UWO/MAF2-Ler</i>	142	obs	48	69	25	7.56338	0.0228*
		exp 1:2:1	35.5	71	35.5		
	222	obs	79	97	46	13.34234	0.0013**
		exp 1:2:1	55.5	111	55.5		
	296	obs	92	147	57	8.290541	0.0158*
		exp 1:2:1	74	148	74		
total	660	obs sum	219	313	128	26.84545	0**
		exp 1:2:1	165	330	165		
			<i>Ler</i> / <i>Ler</i>	<i>Ler</i> / Col	Col/Col		
<i>flc-Ler/FLC-Col</i>	142	obs	32	83	27	4.408451	0.1104 n.s.
		exp 1:2:1	35.5	71	35.5		
	222	obs	63	112	47	2.324324	0.3129 n.s.
		exp 1:2:1	55.5	111	55.5		
	296	obs	75	149	72	0.074324	0.9635 n.s.
		exp 1:2:1	74	148	74		
total	660	obs sum	170	344	146	2.933333	0.2238 n.s.
		exp 1:2:1	165	330	165		

TABLE S6
Primers used in this study

Primer #	Primer sequence 5'> 3'	Purpose	PCR specifications*
1	TAGCACAAAGACACTT TTATCTCCCTC	cDNA screen	1 and 2, 296 bp in Col, 55°C, 2µM, 1.4%
2	CTATAACCAGAAACGT CTTCTTCCC	cDNA screen	
3	TAAAACTTTCTCTCA ATTCTCTCT	Ubiquitin, cDNA screen	3 and 4, 415 bp, 55°C, 2µM, 32x, 1%
4	TTG TCG ATG GTG TCG GAG CTT	Ubiquitin, cDNA screen	
5	AACTAATGATGGGGGA AGTGA	Sequencing <i>MAF2</i> alleles	5 and 6, 1167 bp in UWO, 56°C, 1.2µM, 32x, 1%
6	CTTTGGACTATTTCTA GTAACCTCCTTG	Sequencing <i>MAF2</i> alleles	
7	CTTGGAAAAGGAAAA ATCACTATG	Genomic DNA screen, 3' end of insertion	7, 8 and 9, 7 and 9 amplify 739 bp in Col, 7 and 8 amplify 638 in UWO, 56°C, 2µM, 32x, 1%
8	CAACACAGTTTTGAGT GAAACTCAC	Genomic DNA screen, 3' end of insertion	
9	GAAGGACTTTGATTGA TGTTAGGC	Genomic DNA screen, 3' end of insertion	
10	AAACAAAACGAAGCTC TTTTCTTT	Sequencing <i>MAF2</i> -Sha and <i>MAF2</i> -Kas, <i>MAF3</i> deletion	10 and 11, 405 bp in Sha, 54°C, 2µM, 32x, 1%
11	TAAGCCCGTTTTGATT GGAC	Sequencing <i>MAF2</i> -Sha and <i>MAF2</i> -Kas, <i>MAF3</i> deletion	
12	AAGTTGAAGGACTTTG ATTGATG	Genotyping <i>MAF2</i> -UWO BC5 F2 populations	12, 13 and 14, 12 and 14 amplify a 212 bp fragment in <i>Ler</i> , 12 and 13 amplify a 279 bp fragment in UWO, 54°C, 3µM, 32x, 1%
13	CGGTTGGAGGAATTTA TAGAGTG	Genotyping <i>MAF2</i> -UWO BC5 F2 populations	
14	ATTATTTTCTTACAGC TAACAA	Genotyping <i>MAF2</i> -UWO BC5 F2 populations	
15	GTCTTTTGATATTTTC GTAATGTCTTGTG	Genomic DNA screen, 5' end of insertion, Genotyping <i>MAF2</i> -Sha BC5 F2 populations	15 and 16, 749 bp in Sha, 58°C, 2µM, 32x, 1%
16	ATATCGGTTGGAGGAA TTTATAGAGTG	Genomic DNA screen, 5' end of insertion	
17	TCCACTGATATGTTGA TGCAAT	Genotyping <i>MAF2</i> -Sha BC5 F2 populations	15, 17 and 18, 15 and 18 amplify a 319 bp fragment in Col, 15 and 17 amplify a 299 bp fragment in Sha and Kas-1, 56°C, 2µM, 32x, 2%
18	TCATCCAATTCGGTGA TTCAT	Genotyping <i>MAF2</i> -Sha BC5 F2 populations	
19	TCCTTTCAAAAATTCAA ATTTCTTGT	Genotyping <i>MAF2</i> -Tu-1 BC5 F2 populations	19, 20 and 21, 19 and 21 amplify a 215 bp fragment in <i>Ler</i> , 19 and 20 amplify a 244 bp fragment in Tu-1, 56°C, 2µM, 32x, 2%
20	GGTTAATTTTCGTTTGG TTCAGTG	Genotyping <i>MAF2</i> -Tu-1 BC5 F2 populations	
21	CCATTTTCCCATGACA TTCC	Genotyping <i>MAF2</i> -Tu-1 BC5 F2 populations	
20	GAAGCACTGGTCAAC CTGA	Sequencing <i>MAF2</i> alleles	20 and 21, 850 bp, annealing, 2µM, 32x, 1 %
21	TGTTTGGACAAGATCA TAAGGTCA	Sequencing <i>MAF2</i> alleles	
24	ATTTCTGTTCGGTGGG GATG	Sequencing <i>MAF2</i> alleles	24 and 25, 1111 bp, 58°C, 2µM, 32 x, 1%
25	TTCCGGATCAGTAATT CCAC	Sequencing <i>MAF2</i> alleles	
26	CCTCCGGTGAATTTGT TTAGAG	Sequencing <i>MAF2</i> alleles	26 and 27, 1174 bp, 58°C, 2µM, 32 x, 1%
27	AACGACTCAGCAGGGA ATTG	Sequencing <i>MAF2</i> alleles	
28	AAGCCAAACCCATTTT CTTG	Sequencing <i>MAF2</i> alleles	28 and 29, 993 bp, 58°C, 2µM, 32x, 1%
29	GTCTCGAGCTGTTCCCT CCAG	Sequencing <i>MAF2</i> alleles	
30	ACAAGTCACGAGCGGA ATATC	Sequencing <i>MAF3</i> -UWO allele	30 and 31, 1005 bp, 58°C, 2µM,

31	CATATCTTGGCCACCT CAAAG	Sequencing <i>MAF3</i> -UWO allele	32x, 1%
32	CGGAAAACCTCTACGAC TCTGC	Sequencing <i>MAF3</i> -UWO allele	32 and 33, 1063 bp, 58°C, 2µM, 32x, 1%
33	GCAAAAAGGTCATGTGG TTAGG	Sequencing <i>MAF3</i> -UWO allele	
34	TTCGACCGATAGGGTG AAAC	Sequencing <i>MAF3</i> -UWO allele	34 and 35, 1155 bp, 58°C, 2µM, 32x, 1%
35	CTCAGGCTCGACCAGT AATTC	Sequencing <i>MAF3</i> -UWO allele	
36	TCATCAAAATTTCTCTG GAATGC	Sequencing <i>MAF3</i> -UWO allele	36 and 39, 1746 bp, 58°C, 2µM, 32x, 1%
37	TTTGGACACGATCAAA AGGTC	Sequencing <i>MAF3</i> -UWO allele	37 and 38, band size, 58°C, 2µM, 32x, 1%
38	GCCATATTGTAGGGTA ATGCTG	Sequencing <i>MAF3</i> -UWO allele	
39	TACGGACAGTACGGTT GAAGC	Sequencing <i>MAF3</i> -UWO allele	36 and 39, 1746 bp, 58°C, 2µM, 32x, 1%
40	TTTTCTCCCTCGATGA ATCAC	Sequencing <i>MAF2</i> alleles	40 and 41, 2354 bp in UWO, 56°C, 2µM, 30x, 0.8%
41	TCACGTGCAAGTAATC AAATAAC	Sequencing <i>MAF2</i> alleles	
42	TGTCTCCAAGG GTTCCAGGTT	<i>TUB2</i> , Loading control for expression analysis of <i>MAF2</i> -insertion alleles	42 and 43, 920 bp, 60°C, 1µM, 22x, 0.8%
43	TCACCTTCTTCATCCG CAGTT	<i>TUB2</i> , Loading control for expression analysis of <i>MAF2</i> -insertion alleles	
44	ACATTGTGGGTCTCCG GTGATTAGGATC-	Expression analysis of <i>MAF2</i> -insertion alleles	44 and 45, band sizes ranging from 600 to 1100 bp, 60°C, 1.5µM, 27x, 1.2-1.4%
45	AATCAGGCTGTAAGTT TAAGGTGAAAGC	Expression analysis of <i>MAF2</i> -insertion alleles	
47	CGTAGACAAGGTACTG TCAACC	Mapping <i>MAF2</i> -UWO	47 and 48 + EcoRV digest, 3 fragments of 0.35, 0.23, 0.08 kb in Col and 2 fragments of 0.43 0.23 kb in UWO 53°C, 1.5µM, 32x, 1%
48	GATAATCTCGTCTCCA AGTGTCC	Mapping <i>MAF2</i> -UWO	
49	GCATAGAATTTGACGA TAACGAGC	Mapping <i>MAF2</i> -UWO	49 and 50 + XbaI digest, a single 1.2 Kb fragment is seen in Col and a 0.7 and 0.5 Kb fragment is seen in UWO 53°C, 1.5µM, 32x, 1%
50	GATCTGTGTAGGACTA CGAGAC	Mapping <i>MAF2</i> -UWO	
51	GCTTTCACCTTAAACT TACAGCCTGATT	Sequencing <i>MAF2</i> -UWO 3' intergenic region	51 and 52, 947 bp in Sha and Kas-1, 56°C, 2µM, 32x, 1%
52	CCGGAGACACTGAACG TTTT	Sequencing <i>MAF2</i> -Sha and <i>MAF2</i> -Kas, <i>MAF3</i> deletion	
53	GATTTATTCGTGTGTT TGTCTTTTTG	Sequencing <i>MAF2</i> -UWO 3' intergenic region	53 and 54, 1181 bp, 56°C, 3µM, 32x, 1%
54	GGCTTCTTCTCCGAT AAGGTT	Sequencing <i>MAF2</i> -UWO 3' intergenic region	
55	TGATCAAACCAATTAT CCCAA	Sequencing 1128 bp region of <i>MAF3</i>	55 and 56, 1128 bp, 55°C, 2 µM, 32x, 1%
56	CAAGACATTACGGATT ATGTCAGA	Sequencing 1128 bp region of <i>MAF3</i>	
57	GATTAAGTTTTAAGGTG AAGGCCCTCTCTCTTT TGTATTCC	I miR-s	57 and 62, 300 bp, 55°C, 2.5µM, 32x, 1%
58	GAGGGCCTTCACCTTA AACTTAATCAAAGAGA ATCAATGA	II miR-a	58 and 59, 176 bp, 55°C, 2.5µM, 32x, 1%
59	GAGGACCTTCACCTTT AACTTATTCACAGGTC GTGATATG	III miR*s	58 and 59, 176 bp, 55°C, 2.5µM, 32x, 1%
60	GAATAAGTTAAAGGTG AAGGTCTCTACATAT ATATTCTT	IV miR*a	60 and 61, 274 bp, 55°C, 2.5µM, 32x, 1%

61	CTGCAAGGCGATTAAG TTGGGTAAC	miR319a backbone specific primers	61 and 62, 699 bp, 55°C, 2.5µM, 32x, 1%
62	GCGGATAACAATTTCA CACAGGAAACAG	miR319a backbone specific primers	61 and 62, 699 bp, 55°C, 2.5µM, 32x, 1%

Primer details from Materials and Methods

* PCR information includes: primer combination, expected band size (bp), annealing temperature, MgCl₂ concentration, cycle number, % agarose for band resolution.