

S. J. Szalma · E. S. Buckler IV
M. E. Snook · M. D. McMullen

Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks

Received: 22 February 2004 / Accepted: 19 February 2005 / Published online: 2 April 2005
© Springer-Verlag 2005

Abstract Two compounds, the C-glycosyl flavone maysin and the phenylpropanoid product chlorogenic acid (CGA), have been implicated in corn earworm (*Helicoverpa zea* Boddie) resistance in maize (*Zea mays* L.). Previous quantitative trait locus (QTL) analyses identified the *pericarp color* (*p*) locus, which encodes a transcription factor, as the major QTL for maysin and CGA. QTL analysis has also implicated the dihydroflavanol reductase (DFR; E.C. no. 1.1.1.219) locus *anthocyaninless1* (*al*) and the duplicate chalcone synthase (CHS; E.C. no. 2.3.1.74) loci *colorless2* (*c2*) and *white pollen1* (*whp1*) as genes underlying QTL for maysin and/or CGA synthesis. Epistatic interactions between *p* and *al* and between *p* and *c2* were also defined. CHS catalyzes the first step in the flavonoid pathway and represents

one of the first enzyme steps following the branch off the general phenylpropanoid pathway towards CGA synthesis. In maize, the reduction of dihydroflavanol to leucoanthocyanin by DFR immediately follows the pathway branch leading to C-glycosyl flavone production. The detection of QTLs for maysin and CGA concentration at loci encoding enzyme steps following the pathway branch points implicates alterations in the flow of biochemical intermediates as the biological basis of the QTL effects. To examine if sequence variation among alleles of *al*, *c2*, and *whp1* affect maysin and CGA synthesis in maize silks, we performed an association analysis. Because the *p* locus has often been a major QTL for maysin and CGA and has exhibited epistatic interactions with *al*, *c2*, and *whp1*, association analysis was conditioned on the *p* genotype. A highly significant association of two sequence polymorphisms in the promoter of *al* with maysin synthesis was demonstrated. Additional conditioning on the genotype of the significant *al* polymorphism allowed the detection of a significant polymorphism within the *whp1* promoter. Our analyses demonstrate that conditioning for epistatic factors greatly increases the power of association testing.

Electronic Supplementary Material Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00122-005-1973-0>

Communicated by S.J. Knapp

S. J. Szalma
Genetics Area Program, University of Missouri,
Columbia, MO 65211, USA

E. S. Buckler IV
USDA-ARS, Plant, Soil, and Nutrition Research Unit,
Cornell University, Ithaca, NY 14853, USA

M. E. Snook
Richard B. Russell Research Center,
University of Georgia, Athens, GA 31793, USA

M. D. McMullen (✉)
Plant Genetics Research Unit, USDA-ARS,
and Department of Agronomy, University of Missouri,
Curtis Hall, Columbia, MO 65211, USA
E-mail: mcmullenm@missouri.edu
Tel.: +1-573-8827606
Fax: +1-573-8847850

Present address: S. J. Szalma
USDA-ARS, Department of Genetics,
North Carolina State University, Gardner Hall,
Raleigh, NC 27695, USA

Introduction

The phenylpropanoid and flavonoid pathways have been widely studied in plants, including maize (*Zea mays* L.) (Coe et al. 1988; Harbourne 1988; Grotewold et al. 1998; McMullen et al. 1998), *Arabidopsis thaliana* (Burbulis and Winkel-Shirley 1999; Xie et al. 2003), snapdragon (*Antirrhinum majus*) (Moyano et al. 1996; Tamagnone et al. 1998; Schwarz-Sommer et al. 2003), and the *Solanaceae* (Kroon et al. 1994; De Jong et al. 2003). A wide array of regulatory and structural genes of flavonoid biosynthesis have been cloned in these species. The phenylpropanoid compound chlorogenic acid (CGA) and the C-glycosyl flavone maysin have been implicated in corn earworm (*Helicoverpa zea* Boddie) antibiosis (Waiss et al. 1979; Elliger et al. 1980; Isman

and Duffy 1983). By determining the genetic basis of the variation in the accumulation of maysin and CGA in maize silks, we advance both our understanding of the flavonoid pathway as a model genetic system and promote crop improvement.

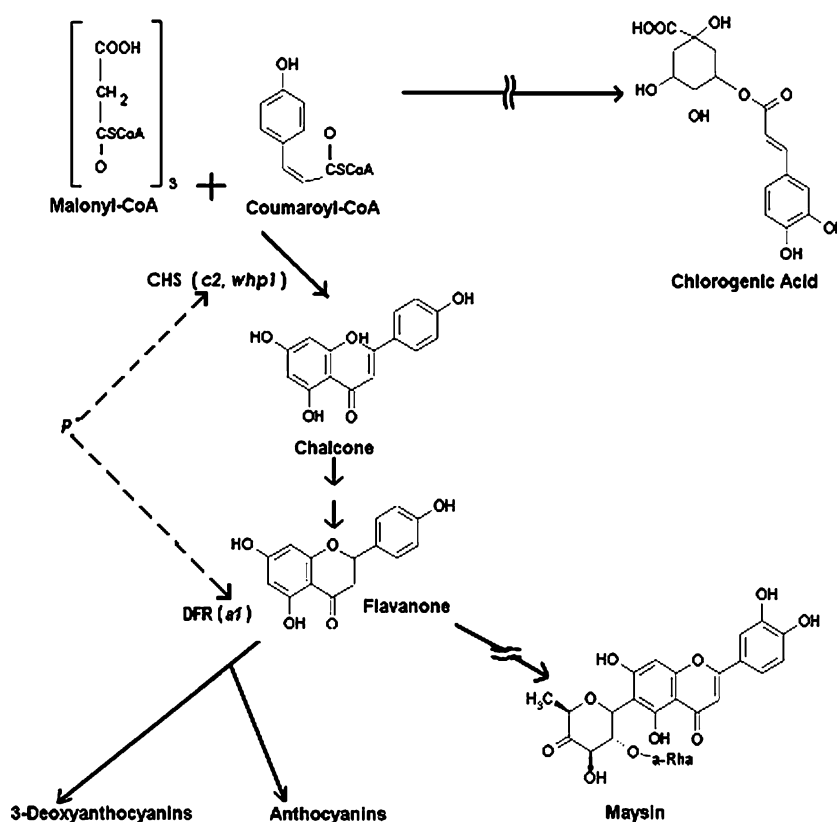
Maysin and CGA share precursors prior to the condensation of the nine-carbon coumaroyl-CoA with three malonyl-CoA moieties to form the 15-carbon compound naringenin chalcone by chalcone synthase (CHS, E.C. no. 2.3.1.74) (Fig. 1). Two maize loci encode CHS, *white pollen1* (*whp1*) and *colorless2* (*c2*). Further downstream in flavonoid biosynthesis, the *anthocyaninless1* (*a1*) gene encodes a nicotinamide adenine dinucleotide phosphate-dependent dihydroflavanol reductase (DFR, E.C. no. 1.1.1.219). This enzyme is responsible for the reduction of dihydroflavanol to flavan-3,4-diol en route to the anthocyanin and 3-deoxyanthocyanin synthesis. The relationship between the synthesis of maysin and other branches of the flavonoid pathway has been a focus of our investigation of the genetics behind phenylpropanoid and flavonoid biosynthesis (McMullen et al. 2001b; Bushman et al. 2002; Szalma et al. 2002).

Quantitative trait locus (QTL) analysis has been employed using populations constructed to maximize phenotypic variance for C-glycosyl flavone and/or CGA accumulation (Byrne et al. 1996b, 1998; Lee et al. 1998; McMullen et al. 1998; Bushman et al. 2002). One striking discovery was that regulatory loci explained a tremendous amount of phenotypic variation. A region on chromosome 1 containing the *pericarp color* (*p*) locus

was identified in several populations as the QTL explaining the largest amount of phenotypic variation for both maysin and CGA accumulation (McMullen et al. 1998, 2001a). The *p* locus encodes the duplicate *myb*-like transcription factors *p1* and *p2* (Zhang et al. 2000). Additional support for the effect of *p* on flavone and phenylpropanoid accumulation has come from transformation studies (Grotewold et al. 1998; Dong et al. 2001). Any investigation of the expression of the structural genes within this pathway requires careful monitoring of the *p* genotype.

The role that genetic variation of structural genes in the flavonoid pathway has on maysin and CGA synthesis remains a question. Byrne et al. (1998) reported that the *whp1* region was a QTL for maysin and corn earworm antibiosis. Szalma et al. (2002) demonstrated that increased dosage of functional alleles at *c2* and *whp1* had a positive effect on maysin accumulation and a negative effect on CGA, presumably by modulating substrate flow through alternative pathways. McMullen et al. (2001b) also showed a similar genetic effect for the interaction of the flavone and 3-deoxyanthocyanin pathways, as the elimination of DFR function by a mutation at *a1* resulted in increased maysin at the expense of 3-deoxyanthocyanins, in a *p*-dependent manner. The *BANYULS* gene in *Arabidopsis thaliana* and an equivalent gene in *Medicago truncatula*, an anthocyanidin reductase involved in the production of condensed tannins, is also a locus at a branch point of the flavonoid pathway, and its expression was shown to

Fig. 1 Schematic of the flavonoid and phenylpropanoid pathways in maize. The relative positions of CHS [encoded by *colorless2* (*c2*) and *white pollen1* (*whp1*)], and DFR [encoded by *anthocyaninless1* (*a1*)] within the pathway are indicated. CHS catalyzes the transition from 9-carbon to 15-carbon compounds, which is the first step in the pathway of all flavonoid compounds. DFR catalyzes the first step in the branch from the flavone pathway towards 3-deoxyanthocyanin and 3-hydroxyanthocyanin biosynthesis. Relative positions of the structural enzymes are presented. *Hatched arrows* indicate the regulation of *c2* and *a1* by the *pericarp color* (*p*) locus



alter the relative amounts of anthocyanins and condensed tannins (Xie et al. 2003). Although *al*, *whp1*, and *c2* can all be made into QTLs by the use of non-functional alleles, we have never seen QTLs of comparable magnitude for these loci in QTL experiments with standard inbred lines. Consequently, the question remains—does the natural sequence variation present at *c2*, *whp1*, and *al* in diverse maize lines affect maysin and/or CGA synthesis?

Association mapping, also referred to as linkage disequilibrium (LD) mapping, is a method that can test the relationship of specific sequence polymorphisms in candidate genes to phenotypic variation (Thornsberry et al. 2001). The non-random association of polymorphisms within a locus that defines alleles in a population constitutes LD (Flint-Garcia et al. 2003). The rate of LD decay in maize is rapid, generally within 1,000–1,500 bp (Remington et al. 2001; Tenaillon et al. 2001). Therefore, this within-gene resolution of association mapping represents a much greater precision than QTL mapping and provides an independent approach to test candidate genes identified in standard QTL experiments. In addition, because association analysis links specific nucleotide polymorphisms to trait variation, one may often hypothesize specific biological effects for significant polymorphisms. A potential obstacle in association studies is the spurious association of polymorphisms with traits due to relatedness rather than sequence function. This is especially true in maize because lines have varying degrees of shared pedigree histories. False positives due to population structure can be reduced by including in the association model a vector quantity for sub-population membership derived from simple sequence repeat (SSR) information (Pritchard and Rosenberg 1999; Pritchard et al. 2000a, 2000b; Thornsberry et al. 2001). For this study investigating the natural allelic variation at *al*, *c2*, and *whp1*, the association model was further enhanced by the inclusion of genotype-class variables for epistatic factors.

The sequence variation present in the *al* promoter and the *c2* and *whp1* loci was characterized. After determining the constitution of, and correcting for the effect of the various *p* alleles, association tests were performed to examine the correlation between discrete sequence polymorphisms in *c2*, *whp1*, and *al*, with maysin and CGA accumulation.

Materials and methods

Plant material

The 86 maize inbred lines used in this study represent a broad spectrum of the available maize diversity (Remington et al. 2001). On the basis of SSR marker information, this collection of lines (see electronic supplemental material, ESM-S1) has previously been divided into three sub-populations: Stiff Stalk (SS), non-Stiff Stalk (NSS), and sub-tropical/tropical (ST)

(Thornsberry et al. 2001). Only this set of lines with sub-population definition was used in structured association tests involving *c2*, *whp1*, and *al* with maysin and CGA. Twenty-five near-isogenic lines containing different alleles of *p* in the 4Co63 background (Brink and Styles 1966) (ESM-S2) and six additional lines utilized by our group in previous QTL studies (ESM-S3) were included in experiments to define the relationship of *p* gene structure with phenotype. Plants were grown during the summer of 2001 at the University of Missouri Genetics Research Farm, Columbia, Mo., USA. Two replications of all lines were grown in a common field to increase the sample size and ensure the collection of multiple silk samples from each line. Leaf tissue for DNA extraction was collected from whorl leaves 1 month after planting.

Silk collection and analysis

Primary ear shoots were covered prior to silk emergence to prevent pollination. Silks were collected 2 days post-emergence from five plants of each line in each replication for an average of nine silks per line (low = 5 silks, high = 10 silks). Silk browning (Byrne et al. 1996a), pericarp pigmentation, and cob coloration, which are under the influence of the *p* locus, were noted when applicable (ESM-S1). Chemical analysis of the silks was performed using reversed-phase high-performance liquid chromatography on individual samples following extraction of the silks in methanol at 0°C for 14 days (Snook et al. 1989, 1993). The concentrations of maysin and CGA were determined and expressed as the percentage of fresh silk weight for each silk mass.

Molecular structure classification of the *p* allele

PCR primers EP5-8, EP3-13, and P2-5 were utilized to amplify either *p2* specifically or both *p1* and *p2* (Zhang et al. 2000). The combination of primers EP5-8 and EP3-13 yields an approximately 380-bp fragment for *p1* and an approximately 300-bp fragment for *p2*. The combination of the P2-5 and EP3-13 primers yields an approximately 240-bp fragment for *p2* and no fragment for *p1*. Using these two sets of primers, we classified lines as ‘*p1* only’, ‘*p2* only’, or ‘*p1* and *p2*’. A fourth, novel amplification pattern was seen with many *p-www* alleles; no amplification with P2-5 and EP3-13 and a single, large (> 2 kb) fragment produced with the EP5-8 and EP3-13. This pattern was designated as the ‘A619 type’.

DNA extraction and sequencing

Extraction, purification, and quantification of DNA were performed according to methods published in the University of Missouri-Columbia RFLP Procedures Manual (<http://www.maizemap.org>, confirmed August, 2004). Primers specific to regions of the genes to be se-

quenced were designed using PRIMER3 software (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi/>, confirmed August, 2004). The fragments to be sequenced ranged between 500 bp and 1,400 bp. To ensure specificity in amplification between the duplicate CHS loci, we anchored at least one primer for each fragment in an intron. The quality of the amplification products was checked on 1% (w/v) agarose gels. Generally, amplification conditions for the sequencing templates followed those described at <http://www.maizegdb.org/ssr.php> (confirmed August, 2004), or they were modified to extension times of 3 min.

The precipitation of PCR products was performed prior to and following the sequencing reactions and consisted of 1 volume 2 M MgCl₂ and 2.5 volumes 95% (v/v) EtOH and a 15- to 30-min incubation followed by centrifugation at 1,924 g for 30 min. The pellets were washed with 5 volumes 70% (v/v) EtOH for 10 min and centrifuged for 15 min at 1,924 g. All stages of precipitation and the reagents were at room temperature. Sequencing reactions were performed with the dRhodamine Terminator Ready Reaction Sequencing kit according to manufacturer's recommendations [Applied Biosystems (ABI), Foster City, Calif.]. The samples were denatured in 10 µl deionized formamide at 95°C for 4 min and then cooled to 4°C before analysis on the ABI 3100 sequencer. Two or three replications of sequencing were performed in both forward and reverse directions for each fragment to provide a twofold to sixfold redundancy for each fragment in each line. Sequence quality was assessed manually with the SEQMAN application of the DNASTAR software suite (DNASTAR, Madison, Wis.) and with the assistance of PHRED software (Ewing and Green 1998; Ewing et al. 1998). Prior GenBank sequence information was used as a framework upon which to assemble data (*c2*, no. X60205; *whp1*, no. X60204; *a1*, no. X05068). Contiguous fragments were assembled manually with SEQMAN or PHRAP (Phil Green, University of Washington, 2000). Several large insertions in the intron of *c2* relative to the reported sequence hindered assembly in some lines. The sequence of the *c2-Idf* allele, provided by C. Della-Vedova, University of Missouri (personal communication), proved useful in completing sequence assembly. Initial sequence alignments were obtained with CLUSTALX for the Microsoft Windows platform and the CLUSTALW protocol (Thompson et al. 1994) within the MEGALIGN application of the DNASTAR software suite. Alignments were edited manually with sequence quality scores provided by PHRED and by visual examination of the original trace files for each line. The *a1* promoter was sequenced in 86 lines (ESM-S1) for the region (1,358–1,761 bp) of the reported sequence (no. X05068). The promoter and coding regions of *whp1* corresponding to the regions 29–908 of no. X60204 were sequenced in 77 lines, and the second exon (3,007–3,810 bp) was obtained in 64 lines (ESM-S1). A total of 3.4 kb of sequence was obtained for the *c2* locus corresponding to base pairs 331–3,778 of the GenBank accession (no. X60205) in 84 lines (ESM-

S1). The sequences for the *a1*, *c2*, and *whp1* alleles determined in this study have been deposited in GenBank with the accession numbers (*c2*, AY728808–AY728890; *a1*, AY730781–AY730865; *whp1*, AY731288–AY731363).

Statistical analysis

The SAS FOR WINDOWS VER. 8.2 (SAS Institute, Cary, N.C.) was used to conduct the analysis of variance (ANOVA) for the preliminary statistical analysis of trait data. The TASSEL software package (<http://www.maizegenetics.net/bioinformatics/tasselindex.htm>, confirmed August, 2004) was used to identify single nucleotide polymorphisms (SNP) and insertions/deletions (indels). Polymorphisms were tested by association analysis if high-quality sequences were available for at least 90% of the lines sequenced for that gene and if there was an allele frequency of at least 10%. Tests for LD and associations of polymorphisms with trait values were also conducted with TASSEL. Mean maysin and CGA contents of each line were used in the association analyses. Association analyses were conditioned on population structure estimates from SSR data, the functional versus non-functional classification of the *p* region, and significant polymorphisms at structural loci (Thornsberry et al. 2001). The test statistic has been described by Thornsberry et al. (2001). To select an appropriate significance threshold, data were permuted 1,000 times. A polymorphic site was deemed to have a significant association if the logistic regression *P*-value was below the 5% empirically derived value.

Association analysis implements likelihood estimation, thereby incorporating more information into the statistical model through estimates of conditional posterior probability and, consequently, resulting in increased sensitivity of the test (Kao 2000). PROC GLM in SAS allows for computationally straightforward detection of significant differences in phenotypic variation relative to genotype classes using an ANOVA approach, although differences in the estimation of parameters and potentially biased estimation of residual variance exist relative to the likelihood approach (Kao 2000). Under appropriate conditions, the ANOVA and likelihood approaches should provide similar results, making it possible to use ANOVA to confirm the results of association analysis. Indels and SNPs identified in the *a1* and *whp1* sequences were tested with PROC GLM in SAS. Since a priori knowledge existed that the *p* region and genetic background of the lines used in this study accounted for significant differences in silk maysin concentrations, genetic background classification (SS, NSS, or ST) and functional versus non-functional classification of the *p* region were used as class variables when testing the significance of the polymorphism genotypes of *a1* and *whp1*. A multiple-locus model for maysin accumulation in maize silks was created with PROC GLM in SAS testing the functional versus non-functional classification of the *p* region, the

genetic background, the most significant sequence polymorphisms from *al* and *whp1* [*al*, position 1,369(A/G); *whp1*, position 261(13/15)], and interaction terms. Loci and interaction terms satisfying the selection criterion of a Type-III sum of squares ($P < 0.01$) were retained in the model along with the main-effect components of the significant interaction terms.

Results

Silk maysin and CGA concentrations

From ANOVA, neither individual plants within lines nor lines between replications were significant for maysin and/or CGA concentration. Therefore, data from replications were combined to provide line means for association tests. Mean maysin values ranged from 0.00 to 0.87% ($\pm 0.019\%$, experiment-wide) of fresh silk weight, and mean CGA concentrations ranged from 0.00 to 0.18% ($\pm 0.007\%$, experiment-wide) of fresh silk weight (ESM-S1).

Sequence polymorphisms

In the *al* promoter, 11 SNPs and five indels with a frequency of at least 10% were identified. Eighty-five SNPs and 92 indels were identified throughout the *c2* sequence. Thirty-five SNPs and 18 indels were identified in the *whp1* promoter, 3'-untranslated region, and first exon, and two indels and 19 SNPs were found in the second exon of *whp1*. In addition, the T nucleotide at a rare T/G SNP at position 799 in the *whp1* sequence would result in a premature stop codon within the first intron. K55, the source of the original mutant (non-functional) *whp1* allele (Coe et al. 1981), was one of five lines with the T base, suggesting this site as the cause of the *whp1-ref* mutant allele. Unfortunately, the sample size of only five lines harboring the position-799 T SNP was too small to test this polymorphism in structured association analysis.

Structured association testing

Pericarp color (*p*)

Functional alleles at *p* induce the expression of genes within the phenylpropanoid and flavone pathways to allow the production of maysin and CGA, while non-functional alleles result in the accumulation of lower quantities of CGA and negligible amounts of maysin (Byrne et al. 1996b; Bushman et al. 2002; Szalma et al. 2002). Both *al* and *c2* have been previously demonstrated to be under *p* regulation (Dooner et al. 1980; Grotewold et al. 1994). Therefore, it was necessary to account for the effect of specific *p* alleles when testing polymorphisms in *al*, *c2*, and *whp1*.

The PCR-based assay to classify alleles of the *p* region resulted in four distinct amplification patterns: '*p1* only', '*p2* only', '*p1* and *p2*', and 'A619 type'. A strong relationship between the molecular classification with pericarp and cob pigmentation and silk phenotype was evident (Table 1). With one exception, lines characterized as '*p2*-only' were of the colorless pericarp, colorless cob, and silk-browning phenotype (*p-www*). Lines in the '*p1*-only' and '*p1* and *p2*' classes had pigmented cobs and/or pericarps, but all had browning silks (*p-wrb*, *p-rwb*, and *p-rrb*) (Table 1). The 'A619-type' amplification product was observed only in lines with a colorless pericarp, white cob, and non-browning silks (*p-www*). Therefore, this type was classified as non-functional for silk expression. Based on silk browning and the presence of maysin in silks of all other *p* allele classes (*p-wrb*, *p-rwb*, and *p-rrb*), these three phenotypic classes of *p* were all classified as functional for silk maysin expression.

Single-factor ANOVA was tested for statistical significance of the molecular *p*-allele classification for maysin and CGA accumulation. The analysis was performed on a combined set of all maize lines to ensure adequate representation of each *p* class. Similar patterns were evident in the 86 diverse inbred lines and 25 near-isogenic lines analyzed separately (data not shown). Genetic variation at the *p* locus was significant for both maysin and CGA accumulation at the $P < 0.0001$ level. A progression of maysin and CGA accumulation from the 'A619-type' *p* allele to the '*p1* and *p2*' class was observed (Fig. 2). The '*p2* only' allele resulted in elevated CGA accumulation and moderate maysin, while the '*p1* only' class resulted in high concentrations of maysin and moderate CGA. The union of *p1* and *p2* in the '*p1* and *p2*' class resulted in an accumulation of maysin and CGA equal to the maximal amounts observed in the '*p1* only' and '*p2* only' classes, respectively.

A strong relationship can be seen to exist between *p* classification and SSR-based population structure (ESM-S1). Lines classified as SS were primarily of the '*p1* only' class, and the ST lines were mostly '*p2* only'.

Table 1 The relationship of molecular *pericarp color* (*p*) classification to the visual *p* phenotype^a($n = 76$)

	<i>p-www</i> ^b	<i>p-www</i> ^c	<i>p-wrb</i> ^d	<i>p-rwb</i> ^e	<i>p-rrb</i> ^f
A619 type	9	–	–	–	–
<i>p2</i> only	–	26	–	–	1
<i>p1</i> only	1	2	16	1	7
<i>p1</i> and <i>p2</i>	–	3	–	3	9

^aVisual phenotype due to *pericarp color* (*p*) was scored in all lines used in this study and compared with lines for which the molecular classification of *p* could be obtained

^b*p-www*, Colorless pericarp, white cob, and non-browning silks

^c*p-www*, Colorless pericarp, white cob, and browning silks

^d*p-wrb*, Colorless pericarp, red cob, and browning silks

^e*p-rwb*, Colored pericarp, white cob, and browning silks

^f*p-rrb*, Colored pericarp, red cob, and browning silks

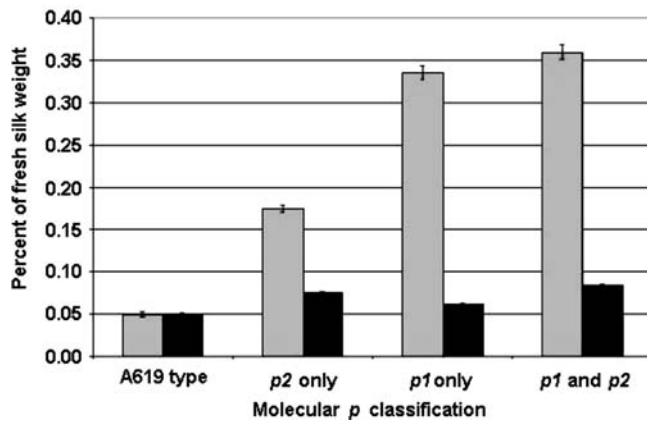


Fig. 2 Molecular classification of the *pericarp color* locus and silk maysin and chlorogenic acid (CGA) accumulation. Means for maysin and CGA were calculated for all lines for different *p* alleles as classified molecularly. Light bars Maysin concentrations, dark bars CGA concentrations, error bars the standard error calculated to two standard deviations. The number of lines in each class is: A619 type—14, *p1* only—40, *p2* only—43, *p1* and *p2*—14

		Maysin	SEM +/-
A	<i>a1</i> 1369(A) - TGCTCGAGA[CAAT]ACGTTTCTG	0.285	0.053
	<i>a1</i> 1369(G) - TGCTCGAGA[CAGT]ACGTTTCTG	0.157	0.019
B	<i>a1</i> 1600(+) - CCCAGC[TTGGTTTGTG]AAAAAA	0.246	0.049
	<i>a1</i> 1600(-) - CCCAGC[TT-----]AAAAAA	0.157	0.023
C	<i>whp1</i> 261(13) - AAC[GAGCGAGAACTAG--]CATA	0.225	0.042
	<i>whp1</i> 261(15) - AAC[TAGCGCAGCGCATAC]CATA	0.142	0.020

Fig. 3 Significant sequence polymorphisms and corresponding maysin values. Significant polymorphisms detected by structured association testing were: *a1* 1,369(A/G) SNP (a), *a1* 1,600(+/-) 8-bp indel (b), and *whp1* 261(13/15) sequence divergence (c). Polymorphisms are indicated by nucleotides in bold enclosed within boxes. Brackets indicate regions within the *a1* promoter of potential biological significance. In a, the bracketed region indicates the CAAT box in the region 1,367–1,370 of *a1*. The bracketed region in b includes an *a1* enhancer located in the region 1,598–1,607. Mean values, with standard errors, for percentage of fresh silk weight of maysin accumulation are to the right of each polymorphism

The alleles ‘*p1* only’ and ‘*p2* only’ were approximately equally represented within the NSS subpopulation. Therefore, SSR population structure largely accounts for differences among the functional *p* classes within the SS and ST groups. Based on these results, structured association analysis was performed with SSR-based population structure estimates and the classification of functional versus non-functional *p* alleles.

Anthocyaninless1 (a1)

Structured association analysis was performed with SSR-based population structure estimates and with classification based on the *p*-allele constitution

Table 2 Significant sites identified with structured association analysis

Locus ^a	Site ^b	Type ^c	SSR-based ^d	<i>p</i> ^e	<i>a1</i> ^f
<i>a1</i>	1,369	SNP(A/G)	0.018	0.003	–
<i>a1</i>	1,600	Indel(8)	0.015	0.007	–
<i>whp1</i>	261	Diverg (13/15)	0.112	0.054	0.033

^aStructured association tests identified significant associations with polymorphic sites in the *anthocyaninless1 (a1)* promoter and *white pollen1 (whp1)* with TASSEL

^bNucleotide positions of the most significant sites identified by structured association tests. Nucleotide positions are reported relative to GenBank accessions for *a1* (no. X05068) and *whp1* (no. X60204)

^cSignificant sites were classified as single nucleotide polymorphisms (SNP), insertions/deletions (indels), or divergent sequences (diverg) based on the type of polymorphism present. The number in parenthesis indicates the length of the indel or divergence, or the type of SNP present

^dPermuted *P* values for structured association testing with SSR-based population structure were taken into account

^ePermuted *P* values for structured association testing with SSR-based population structure and functional versus non-functional classification of the *p* region

^fPermuted *P* values for structured association testing with the SSR-based population structure, functional versus non-functional classification of the *p* region, and the genotype of the *a1* 1,369(A/G) polymorphism

(Table 2). Two polymorphisms in *a1* were identified as being significant for maysin accumulation. These polymorphisms were an A/G SNP at position 1,369 [*a1* 1,369(A/G)] and an 8-bp indel at position 1,600 [*a1* 1,600(+/-)]. The significance of each polymorphism was greatly enhanced by conditioning on *p*, as expected if corresponding alleles are involved in epistatic interaction with the *p* genotype. This result confirms the importance of the *p* classification, in addition to the background classification, in determining maysin concentration. The two significant *a1* sites were in significant LD with one another ($0.50 \leq r^2 \leq 0.59$, $P < 0.0001$); the *a1* 1,369(A) allele occurs with the *a1* 1,600(+) allele in 42 of 43 cases, and the *a1* 1,369(G) allele is in association with the *a1* 1,600(-) allele in 21 of 35 cases (ESM-S1). The *a1* 1,369(A) and *a1* 1,600(+) alleles synthesized appreciably higher amounts of maysin. Lines with the *a1* 1,369(G) allele averaged 0.157% of the fresh silk weight of maysin, whereas the *a1* 1,369(A) allele resulted in mean silk maysin concentrations of 0.285% of the fresh silk weight, an increase of 81.5% (Fig. 3). A similar 60.1% increase in maysin was observed for the *a1* 1,600(+) allele. No significant difference in maysin accumulation was detected between the *a1* 1,369(G)/*a1* 1,600(-) and the *a1* 1369(G)/*a1* 1600(+) combinations of alleles (data not shown). This result suggests that the *a1* 1,369(A/G) polymorphism may be the more important polymorphism affecting maysin accumulation. Because significant sites within the *a1* promoter were in LD with one another, only the *a1* 1,369(A/G) SNP was used for epistatic correction in structured association testing of *c2* and *whp1*.

white pollen1 (whp1)

Population structure estimates from SSR data and corrections for the *p* and *al* 1,369(A/G) genotypes were used in structured association tests of *whp1* polymorphisms. One significant association was detected between a 13-bp and a 15-bp divergent sequence and maysin accumulation at position 261 [*whp1* 261(13/15)] (Table 2, Fig. 3). Lines with the *whp1* 261(13) allele synthesized 0.225% of the fresh silk weight of maysin versus 0.142% in lines with the *whp1* 261(15) allele. This site was only significant in TASSEL after conditioning the association tests on both the *p*-allele class and the *al* 1,369(A/G) polymorphism. No significant associations were detected between polymorphisms of the second exon of *whp1* and either silk maysin or CGA accumulation.

Colorless2 (c2)

Structured association tests were conducted with and without correction for the *p* genotype. No significant association was found between any polymorphism of *c2* and either the maysin or CGA phenotype. Additional correction for the genotype of the *al* 1,369(A/G) SNP did not alter this result.

Model for maysin synthesis

Similar to the standard practice of building a multiple locus model within a QTL experiment we used PROC GLM to construct a genetic model for maysin synthesis among the inbred lines. In agreement with the results from TASSEL, sub-population membership, functional versus nonfunctional *p* classification, and the *al* 1,369 (A/G) SNP were significant as main effects in the model (Table 3). The *whp1* 261(13/15) polymorphism was significant as a main effect in the model before interaction terms were included. Several interaction terms representing the epistatic interaction of *p* with *al*, and background with *p*, *al*, and *whp1* were also significant and improved the model for maysin accumulation (Table 3). The model explained 36.0% of the variation for maysin concentration.

Discussion

Epistatic interactions in association testing

Epistasis is the phenomenon where phenotypic expression from an allele of one gene is specifically altered depending on the allele(s) present at another locus. In previous QTL experiments (McMullen et al. 2001b; Szalma et al. 2002), the *p* gene exhibited highly significant epistatic interactions with alleles at *al* and *c2*. This epistatic interaction is due to *p* controlling the expres-

sion of these genes. The study reported here is the first in which epistatic interactions have been intentionally accounted for in association analysis. Using the molecular classification of *p* as a guide, it was possible to test sequence variation of *c2*, *whp1*, and the promoter region of *al* for association with CGA and maysin accumulation, while accounting for epistasis with *p* alleles. Conditioning on the *p* phenotype greatly increased the significance of the two *al* polymorphisms. The *whp1* polymorphism was only significant in structured association tests after including both the *p* and *al* genotypes in the model (Table 2). However, this site was significant in SAS PROC GLM when the molecular *p* classification and genetic background (SS, NSS, and ST) were included in the model. Maysin concentrations increased with the accumulation of the maysin-increasing alleles at *al* and *whp1*, suggesting a cumulative effect of these two loci. However, because *p* is required for overall expression of the flavone pathway, the effects of *al* and *whp1* are not observed unless a functional *p* allele is present (Fig. 4).

The multiple-locus model for maysin synthesis provides further support for the significance of the *al* 1,396(A/G) and *whp1* 261(13/16) polymorphisms and for the epistatic interaction of *p* and *al* (Table 3). The importance of conditioning the association analysis by genetic background is verified by the significance of the interaction terms of background with *whp1*, and with *p* classification. In our study, a large portion of the “relatedness” effect in background is due to the non-random distribution of specific functional *p* alleles among the three maize subpopulations (ESM-S1). The retention of significant interaction terms between individual loci and genetic background in the multiple-locus model suggests the presence of other epistatic factors, in addition to *p* classification, for maysin synthesis. Presumably, epistatic interactions play a major role in explaining the background effect seen in testing other candidate genes and traits.

Biological mechanisms of polymorphism effects

Anthocyaninless1 (al)

The two significant sites within *al* were in substantial LD with one another, making it difficult to separate an individual polymorphism's effect on phenotype. Because no significant difference was detected between mean maysin concentrations of the *al* 1,369(G)/*al* 1,600(–) and the *al* 1,369(G)/*al* 1,600(+) combinations of alleles, the *al* 1,369(A/G) polymorphism is likely the more important polymorphism affecting maysin accumulation. However, the low number of lines in this comparison limits the strength of this conclusion. Both significant polymorphic sites are in positions identified as potentially affecting transcription (Schwarz-Sommer et al. 1987). The *al* 1,369(A/G) SNP is in a proposed CAAT box (1,367–1,370 bp) within the upstream *al* promoter (1,367–1,734 bp). Individuals with the

Table 3 Multiple locus model of maysin synthesis. Total R^2 of model = 0.360

Source ^a	$Pr > F$
p^b	<0.0001
al 1369(A/G) ^c	<0.0001
$whp1$ 261(13/15) ^d	0.4306
Bkgd ^e	<0.0001
$p^* al$ 1369(A/G) ^f	0.0026
$p^* Bkgd$	0.0031
$whp1$ 261(13/15)*Bkgd	<0.0001
al 1369(A/G)*Bkgd	0.0071

^aParameters included in models were selected using a $P < 0.01$ Type-III sums of squares significance threshold. Terms involved in significant interactions were retained regardless of the significance of the main effect

^b p was represented by the functional versus non-functional classification of the p region

^c $whp1$ (261) was defined by genotype at position 261 of the $whp1$ sequence as the presence of a 13-bp or 15-bp sequence at that location

^d al (1,369) was defined by the genotype at position 1,369 of the al sequence as the presence of an 'A' or 'G' bp at that position

^eBackground (Bkgd) was defined as Stiff Stalk, non-Stiff Stalk, or sub-tropical/tropical

^fThe * indicates interaction

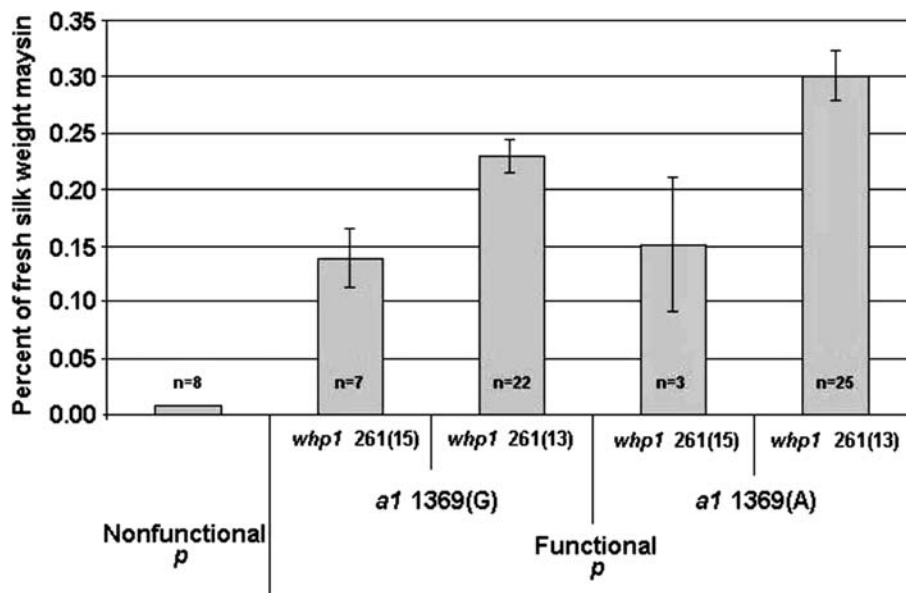
al 1,369(A) allele at this position synthesized more maysin on average than individuals with al 1,369(G), therefore disruption of the CAAT box at 1,369 bp results in a significant decrease in silk maysin. In the context of our current model of flavone biosynthesis, a decreased function at al should result in increased maysin accumulation (Fig. 1) (McMullen et al. 2001b). The initial methionine residue in the DFR protein is coded for by the ATG sequence at position 1,843. Several ATGs that could be used as start sites (positions 1,563, 1,631, and 1,667) exist between the upstream promoter and the start site at position 1,843.

Translation initiation from any ATG site prior to position 1,843 would result in the early termination of protein synthesis. We propose that the al 1,369(G) polymorphism weakens the upstream promoter, thereby enhancing transcription initiation at the downstream promoter immediately before the ATG for the start of the DFR protein.

The deletion at al 1,600(+/-), resulting in reduced maysin accumulation, includes an 8-bp section of the al promoter identified as an enhancer (Schwarz-Sommer 1987). Again, disruption of a sequence element postulated to support al transcription results in enhanced expression of maysin from an alternative pathway. Turck and Fromm (1994) performed a partial deletion analysis of the al promoter, however they did not analyze the sequence upstream of position 1,620. All regulatory elements proposed within the region they investigated are conserved throughout all the lines in this study.

Our data suggest that the natural genetic variation within the al promoter results in phenotypic variation for maysin synthesis. We favor the hypothesis in which the change of an A to a G in an upstream CAAT box enhances transcription initiation at the downstream regulatory sequence. The increased transcription of al would promote the synthesis of 3-deoxyanthocyanins and anthocyanins and decrease maysin synthesis (McMullen et al. 2001b). We cannot rule out the possible additional involvement of the enhancer element deletion. Additionally, we recognize that we cannot rule out potential alternative linked sites located in flanking regions not sequenced in this study. An examination of relative transcript abundances between the promoter types is needed to determine if our hypotheses are true. Our results support earlier findings from QTL studies that al is a QTL for maysin accumulation.

Fig. 4 Effect of *pericarp color*, *anthocyaninless1*, and *white pollen1* on maysin accumulation in maize silks. Mean maysin values were calculated for the different segregating classes of *pericarp color* (p) (functional vs. non-functional), al 1,369(A/G), and $whp1$ 261(13/15) alleles. No significant differences in maysin accumulation were detected within the non-functional p class due to segregation at the al 1,369(A/G) or $whp1$ 261(13/15). Numbers indicated by 'n' indicate the number of lines within each segregating class. Error bars represent the standard error calculated to two standard deviations



Chalcone synthase loci

No significant associations were detected between *c2* and either maysin or CGA concentration. One significant association was detected at the $P < 0.05$ level in the *whp1* promoter region for maysin accumulation when association tests were conditioned on *p* and the *a1* 1,369(A/G) polymorphism. The significant polymorphism detected in the *whp1* promoter began at 261 bp and continued through either 273 bp or 275 bp depending on whether the *whp1* 261(13) or *whp1* 261(15) allele was present, respectively. No important regions of the *whp1* promoter before 289 bp, where the first CAAT signal is reported, are identified in the literature (Franken et al. 1991), and the biological basis for the effect of this *whp1* polymorphism is unknown.

The CHS function is required for the production of viable pollen. The flavonoids quercetin (3,4-dihydroxyflavonol) and kaempferol are needed in either pollen or silk for pollen-tube function. The lack of functional alleles at both *c2* and *whp1* results in conditional male fertility in maize (Mo et al. 1992; Pollak et al. 1995). Functional diversity in *c2* was not observed, and only a single significant polymorphism in *whp1* was detected. Functional diversity within a candidate gene may be reduced due to natural and human selection against deleterious effects of polymorphisms and can only accumulate to the level allowed by the most essential process affected by that gene.

These observations help explain why *p*, rather than the CHS loci, is commonly the major QTL for maysin. The *p* gene is the major QTL for maysin synthesis not only because it controls expression of the pathway, but also because allelic variation is generally not selected against in modern maize inbreds, as it controls non-essential pathways. Therefore, non-functional alleles of *p*, along with an extensive array of alleles expressing tissue-specific variation, can be maintained in inbred maize germplasm. The expression of CHS loci *c2* and *whp1* is a required function that can occur independently of *p* for the synthesis of the flavonols and is maintained by selection.

Summary

Prior QTL analyses have demonstrated that *a1*, *c2*, and *whp1* are all candidate genes for QTLs for maysin synthesis. Analysis of these loci by association analysis is complicated by the fact that the major genetic factor for maysin synthesis is the genotype of the *p* locus. All of the lines in this study were classified based on *p*-allele genotype, the effect of the different *p*-allele classes on maysin synthesis determined, and association analysis conditioned on both population structure and the *p*-allele classification. Two sequence polymorphisms in the *a1* promoter were significant for maysin synthesis. Both polymorphisms alter sequences implicated in transcription regulation. By including an *a1* polymorphism as an

additional factor in association analysis, a significant sequence polymorphism was detected in *whp1*. These results expand the application of association techniques beyond single-locus analysis and emphasize the need for understanding the relative roles of loci with major effects on trait expression. Association analysis can be used to test secondary QTL *if*, and probably often *only if*, the primary QTL(s) for the trait is/are known and are considered in the analysis model.

Acknowledgements The authors would like to thank Katherine Houchins and Chris Browne for technical assistance and Sherry Flint-Garcia and Jim Holland for reviewing the manuscript and for helpful suggestions. This research was supported by USDA-National Research Initiative, Plant Genome Grant no. 2001-35301-10581 and funds provided by USDA-Agricultural Research Service. SJS was supported by a University of Missouri-Molecular Biology Program Predoctoral Fellowship. The names of products are necessary to report factually on available data; however, neither the USDA nor the University of Missouri guarantees or warrants the standard of the product, and the use of the name does not imply approval of the product to the exclusion of others that may also be suitable.

References

- Brink RA, Styles D (1966) A collection of pericarp factors. *Maize Genet Coop Newsl* 40:149–160
- Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proc Natl Acad Sci USA* 96:12929–12934
- Bushman SB, Gerke JP, Szalma SJ, Snook ME, Berhow MA, Houchins KE, McMullen MD (2002) Two loci exert major effects on chlorogenic acid synthesis in maize silks. *Crop Sci* 42:1669–1678
- Byrne PF, Darrah LL, Snook ME, Wiseman BR, Widstrom NW, Moellenbeck DJ, Barry BD (1996a) Maize silk-browning, maysin content, and antibiosis to the corn earworm *Helicoverpa zea* (Boddie). *Maydica* 41:13–18
- Byrne PF, McMullen MD, Snook ME, Musket TA, Theuri JM, Widstrom NW, Wiseman BR, Coe EH (1996b) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc Natl Acad Sci USA* 93:8820–8825
- Byrne PF, McMullen MD, Snook ME, Musket TA, Theuri JM, Widstrom NW (1998) Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. *Crop Sci* 38:461–471
- Coe EH, McCormick S, Modena S (1981) White pollen in maize. *J Hered* 72:318–320
- Coe EH, Neuffer MG, Hoisington DA (1998) The genetics of corn. In: Sprague GF, Dudley JW (eds) *Corn and corn improvement*. ASA, CSSA, SSSA, Madison, pp 81–258
- De Jong WS, Eannetta NT, De Jong DM, Bodis M (2003) Candidate gene analysis of anthocyanin pigmentation loci in the *Solanaceae*. *Theor Appl Genet* 108:423–432
- Dong X, Braun EL, Grotewold E (2001) Functional conservation of plant secondary metabolic enzymes revealed by complementation of *Arabidopsis* flavonoid mutants with maize genes. *Plant Physiol* 127:46–57
- Dooner HK (1980) Genetic regulation of anthocyanin biosynthetic genes in maize. *Genetics* 94:s29
- Elliger CA, Chan BG, Waiss AC, Lundin RE, Haddon WF (1980) C-glycosylflavones from *Zea mays* that inhibit insect development. *Phytochemistry* 19:293–297
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using PHRED, 2. Error probabilities. *Genome Res* 8:186–194

- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred, 1. Accuracy assessment. *Genome Res* 8:175–185
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Franken P, Niesbach-Klosgen U, Weydemann U, Marechal-Drouard L, Saedler H, Wienand U (1991) The duplicated chalcone synthase genes *C2* and *Whp* (*white pollen*) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene *in1*. *EMBO J* 10:2605–2612
- Grotewold E, Drummond BJ, Bowen B, Peterson T (1994) The *myb*-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* 76:543–553
- Grotewold E, Chamberlin M, Snook ME, Siame B, Butler L, Swenson J, Maddock S, St. Clair G, Bowen B (1998) Engineering secondary metabolism in maize cells by ectopic expression of transcription factor. *Plant Cell* 10:721–740
- Harbourne JB (1988) The flavonoids: advances in research since 1980. Chapman & Hall, London
- Isman MB, Duffy SS (1983) Pharmacokinetics of chlorogenic acid and rutin in larvae of *Heliothis zea*. *J Insect Physiol* 29:295–300
- Kao CH (2000) On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* 156:855–865
- Kroon J, Souer E, de Graaff A, Xue Y, Mol J, Koes R (1994) Cloning and structural analysis of the anthocyanin pigmentation locus *Rt* of *Petunia hybrida*: characterization of insertion sequences in two mutant alleles. *Plant J* 5:69–80
- Lee EA, Byrne PF, McMullen MD, Snook ME, Wiseman BR, Widstrom NW, Coe EH (1998) Genetic mechanisms underlying apimaysin and maysin synthesis and corn earworm antibiosis in maize (*Zea mays* L.). *Genetics* 149:1997–2006
- McMullen MD, Byrne PF, Snook ME, Wiseman BR, Lee EA, Widstrom NW, Coe EH (1998) Quantitative trait loci and metabolic pathways. *Proc Natl Acad Sci USA* 95:1996–2000
- McMullen MD, Lee EA, Szalma S, Bushman S, Snook ME (2001a) The role of quantitative trait locus analysis in gene discovery. In: *Proc 56th Ann Corn Sorghum Res Conf. American Seed Trade Association, Madison*, pp 237–245
- McMullen MD, Snook ME, Lee EA, Byrne P, Kross H, Musket T, Houchins KE, Coe EH (2001b) The biological basis of epistasis between quantitative trait loci for flavone and 3-deoxyanthocyanin synthesis in maize (*Zea mays* L.). *Genome* 44:667–676
- Mo Y, Nagel C, Taylor LP (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proc Natl Acad Sci USA* 89:7213–7217
- Moyano E, Martinez-Garcia F, Martin C (1996) Apparent redundancy in *myb* gene function provides gearing for the control of flavonoid biosynthesis in *Antirrhinum* flowers. *Plant Cell* 8:1519–1532
- Pollak PE, Hansen K, Astwood JD, Taylor LP (1995) Conditional male fertility in maize. *Sex Plant Reprod* 8:231–241
- Pritchard JK, Rosenberg NA (1999) The use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65:220–228
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 98:11479–11484
- Schwarz-Sommer ZS, Shepherd NS, Take E, Gierl A, Rohde W, Leclercq L, Mattes M, Berndtgen R, Peterson PA, Saedler H (1987) Influence of transposable elements on the structure and function of the *Al* gene of *Zea mays*. *EMBO J* 6:287–294
- Schwarz-Sommer Z, Davies B, Hudson D (2003) An everlasting pioneer: the story of *Antirrhinum* research. *Nat Rev Genet* 4:655–664
- Snook ME, Widstrom NW, Gueldner RC (1989) Reversed-phase high-performance liquid chromatographic procedure for the determination of maysin in corn silks. *J Chromatogr* 477:439–447
- Snook ME, Gueldner RC, Widstrom NW, Wiseman BR, Himmelsbach DS, Harwood JS, Costello CE (1993) Levels of maysin and maysin analogs in silks of maize germplasm. *J Agric Food Chem* 41:1481–1485
- Szalma SJ, Snook ME, Bushman BS, Houchins KE, McMullen MD (2002) Duplicate loci as QTL: the role of chalcone synthase loci in flavone and phenylpropanoid biosynthesis in maize. *Crop Sci* 42:1679–1687
- Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martin C (1998) The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* 10:135–154
- Tenaillon MA, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* 98:9161–9166
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nat Genet* 28:286–289
- Tuerck JA, Fromm ME (1994) Elements of the maize *al* promoter required for transactivation by the anthocyanin *B/C1* or phlobaphene *P* regulatory genes. *Plant Cell* 6:1655–1663
- Waiss AC, Chan BG, Elliger CA, Wiseman BR, McMillian WW, Widstrom NW, Zuber MS, Keaster AJ (1979) Maysin, a flavone glycoside from corn silks with antibiotic activity toward corn earworm. *J Econ Entomol* 72:256–258
- Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* 299:396–399
- Zhang P, Chopra S, Peterson T (2000) A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. *Plant Cell* 12:2311–2322