MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 1296–1311 doi: 10.1111/j.1365-294X.2010.04562.x

Evidence of selection at the *ramosa1* locus during maize domestication

BRANDI SIGMON and ERIK VOLLBRECHT

Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA

Abstract

Modern maize was domesticated from *Zea mays parviglumis*, a teosinte, about 9000 years ago in Mexico. Genes thought to have been selected upon during the domestication of crops are commonly known as domestication loci. The *ramosa1* (*ra1*) gene encodes a putative transcription factor that controls branching architecture in the maize tassel and ear. Previous work demonstrated reduced nucleotide diversity in a segment of the *ra1* gene in a survey of modern maize inbreds, indicating that positive selection occurred at some point in time since maize diverged from its common ancestor with the sister species *Tripsacum dactyloides* and prompting the hypothesis that *ra1* may be a domestication gene. To investigate this hypothesis, we examined ear phenotypes resulting from minor changes in *ra1* activity and sampled nucleotide diversity of *ra1* across the phylogenetic spectrum between tripsacum and maize, including a broad panel of teosintes and unimproved maize landraces. Weak mutant alleles of *ra1* showed subtle effects in the ear, including crooked rows of kernels due to the occasional formation of extra spikelets, correlating a plausible, selected trait with subtle variations in gene activity. Nucleotide diversity was significantly reduced for maize landraces but not for teosintes, and statistical tests implied directional selection on *ra1* consistent with the hypothesis that *ra1* is a domestication locus. In maize landraces, a noncoding 3′-segment contained almost no genetic diversity and 5′-flanking diversity was greatly reduced, suggesting that a regulatory element may have been a target of selection.

Keywords: maize, plant domestication, positive selection, *ramosa1*

Received 7 July 2009; revision received 13 November 2009; accepted 17 November 2009

Introduction

Agriculture began some 10 000 years ago, when prehistoric farmers in both the New and Old World began domesticating plants and animals through selection on desirable traits. In plants such characters included seed or grain quality, yield and ease of harvest among numerous other traits (Smith 1995). The domestication process typically spans hundreds to thousands of years during which time preferred traits might be selected on singly or in combination with others. Due to continued selection after or in concert with initial domestication events, cultural and ecological adaptation results in locally adapted, diversified crops (Purugganan & Fuller 2009). For example, in many modern crops, accelerated selection of preferable traits has occurred within the last few 100 years, in a process known as crop improvement that is temporally well separated from domestication. Darwin recognized the artificial selection that occurred during the domestication of plants and animals as analogous to the process of natural selection that drives the evolution of species, and studies of domesticated species inform our understanding of the genetic basis of evolutionary diversification (Doebley *et al.* 2006).

In the context of either natural or artificial forces, different modes of selection on phenotypes lead to different outcomes in allele frequencies. For instance, directional selection on a preferred phenotype as may occur in domestication and/or cultural adaptation leads to an increase in frequency of the corresponding allele(s), whereas balancing selection maintains multiple

Correspondence: Erik Vollbrecht, Fax: +1 515 294 6755; E-mail: vollbrec@iastate.edu

alleles. For genes that are associated with positively selected traits in domestication and improvement, the process may leave behind so-called signatures of selection in the form of distinctive patterns of reduced nucleotide diversity (Tanksley & McCouch 1997). When a wild progenitor and intermediate cultivars are extant and can be sampled, appropriate statistical tests (Sabeti *et al.* 2006) may identify and differentiate these signatures of selection in the genome, and pinpoint a selective event to a particular phase of crop evolution.

Molecular, archaeobotanical and palaeoecological evidence suggest a single domestication event for maize, around 9000 years ago in Mexico (Matsuoka *et al.* 2002; Piperno *et al.* 2009). Maize was domesticated from its wild ancestor *Zea mays* ssp. *parviglumis* (hereafter referred to as teosinte) (Doebley 2004). Maize and teosinte have many similarities but exhibit striking differences in plant architecture and ear morphology. Phenotype-driven approaches such as mutant and QTL studies have shown that the molecular basis for some of these major phenotypic differences may be explained by selection on a few domestication loci of major effect. In particular, the *teosinte branched1* (*tb1*) gene explains most of the difference in plant architecture (Doebley *et al.* 1995, 1997; Clark *et al.* 2004, 2006). Other domestication and diversification loci may include genes involved in biochemical traits, such as those involved in the production of starch (Whitt *et al.* 2002) or those responsible for the carotenoid-rich yellow endosperm (Palaisa *et al.* 2004). These examples are among a few putative domestication loci in maize for which an accompanying, selected trait is clearly discernible.

Candidate domestication loci may also be identified by genotype-driven approaches, such as strictly molecular, large-scale screens for genes with signatures of artificial selection (Casa *et al.* 2005; Chapman *et al.* 2008). Such studies in maize suggest that more than 1000 genes may have been selected on during the derivation of modern maize (Wright *et al.* 2005; Yamasaki *et al.* 2005). A disadvantage of this approach is that when candidate genes are defined by strictly molecular criteria, the corresponding gene functions are *a priori* unknown. Thus, while these approaches hold great promise for identifying genes that contribute to adaptive traits (Ross-Ibarra *et al.* 2007), relating such domestication loci to a selected trait is challenging. Notably, in maize many of these putative domestication loci identified by genomics are hypothesized to be expressed preferentially in the ear (Hufford *et al.* 2007), and to perhaps be related to selection on ear traits, because among the grasses the prolific maize ear is a unique and highly derived organ. Moreover, in the ear the basis of maize-teosinte phenotype differences is poorly understood. Thus, a modified, genotype-driven approach in which a candidate gene(s) known to function in the morphological development of the maize ear is (are) examined for evidence of selection may prove more fruitful.

ramosa1 is a mutant of maize that has been studied by geneticists for many years (Gernert 1912) but only recently come to be understood at the molecular level. The *ra1* gene encodes a plant-specific C2H2, EPF-subclass zinc finger transcription factor. It is expressed in a boundary domain near the base of particular meristems in the inflorescences (the ear and tassel), thereby regulating fate of the adjacent meristem (Vollbrecht *et al.* 2005). *ra1* is a component of a genetic pathway, termed the *ramosa* pathway, that imposes a spikelet pair or short branch identity as branch meristems are initiated during tassel and ear development (Vollbrecht *et al.* 2005; Bortiri *et al.* 2006; Satoh-Nagasawa *et al.* 2006). Morphologically, normal *ra1* gene function results in the unbranched appearance of the ear and the upper part of the tassel, by causing both structures to produce short, determinate spikelet pair branches (Fig. 1a–d). In plants that contain strong mutant *ra1* alleles and therefore lack *ra1* gene function, the ears and tassels have additional long branches, leading to a highly disorganized ear (Fig. 1c and f). By contrast, in weak *ra1* mutants the tassels contain just a few extra branches (Fig. 1b, arrowhead) and ears have crooked rows but are otherwise relatively normal (Fig. 1e). Crooked rows probably affect grain yield in maize and are common to other mutants, including *ramosa2* mutants where they form as a consequence of spikelet triplets being produced in place of spikelet pairs (Bortiri *et al.* 2006). Thus, it is reasonable to hypothesize that the phenotype of weak *ra1* mutants has a similar developmental basis, but this has not been examined.

Existing molecular data concerning *ra1* function beg the question of whether or not *ra1* may have been a target of selection given the gene's role in ear development. Previously, nucleotide diversity in a ~700-bp segment of the *ra1* gene in maize was examined in a panel of diverse, modern inbred lines and a signature of selection was detected relative to the neighbouring genus *Tripsacum* (Vollbrecht *et al.* 2005). These data suggested an event of positive selection somewhere between the present and the point in time when maize and tripsacum diverged from their common ancestor several million years ago, with caveats. First, studies have shown that sampling a small region of a locus may lead to erroneous conclusions about positive selection (Yamasaki *et al.* 2008). Second, previous work could not distinguish during which time period the putative selection occurred: the evolution of teosinte, maize domestication and/or modern maize improvement. To find evidence of selection at *ra1* during the

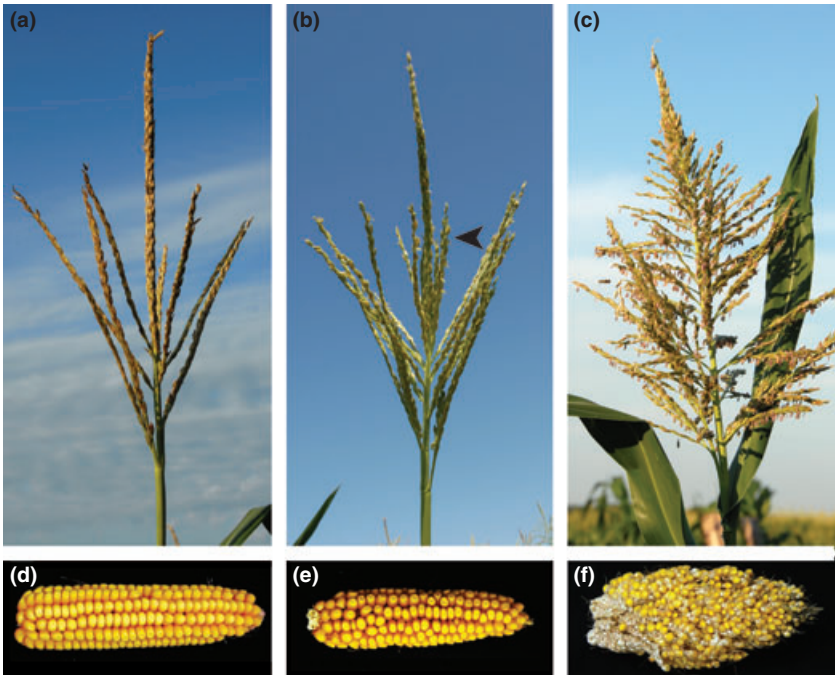

Fig. 1 Maize *ramosa1* mutant phenotypes in mature tassels and ears. Wild-type (inbred Mo17) tassel (a) and ear (d), weak mutant (*ra1-63.3359* in Mo17) tassel (b) and ear (e), and strong mutant (*ra1-R* in a hybrid background) tassel (c) and ear (f). Weak mutant tassels have a few additional long branches (arrowhead) when compared with wild type, whereas strong mutant tassels have long branches extending up the central rachis. Wild-type ears have straight rows, whereas weak mutant ears have crooked rows. Strong mutants exhibit branched ears.

domestication process would not be altogether surprising, as it is intuitive that inflorescence development genes may have been subjected to artificial selection because their role in organizing inflorescence architecture may impact grain yields. We hypothesize that if slight changes in *ra1* function alter the packing of kernels into straight rows on the ear then particular *ra1* alleles may have been selected during maize domestication and/or improvement for increasing grain yields. To address this question, we first determined the developmental basis of the crooked row morphology in the ears of weak *ra1* mutants to better understand this putative, selected trait. We also examined the nucleotide diversity around the *ra1* locus in a diverse panel of unimproved maize landraces and throughout the teosintes. Our statistical and phylogenetic analyses provide compelling evidence that positive selection on *ra1* occurred during the domestication of maize from teosinte, and we discuss the implications of *ra1* as a candidate domestication locus.

Materials and methods

Sampling strategy and plant materials

To examine the developmental basis of crooked rows in weak *ra1* mutants, the recessive *ra1-63.3359* weak mutant allele (Vollbrecht *et al.* 2005) was obtained from the Maize Genetics Cooperation Stock Center and crossed six times to the inbreds Mo17 and B73. Similarly, the recessive *ra1-RS* weak mutant allele, obtained from Robert Schmidt, UC-San Diego, was crossed four to six times to the inbreds. Field-grown mutant (crossed six times to inbreds) and normal plants from our summer nursery in Ames, IA, were sampled at various developmental stages. Scanning electron microscopy (SEM) was used as previously described (Vollbrecht *et al.* 2005) to obtain micrographs of younger ears from greenhouse-grown weak *ra1-RS* mutants (crossed four times to B73).

For nucleotide diversity analysis, accessions of maize and teosinte were chosen to optimize geographic distribution and limit bias toward North American varieties. Accessions of more distant teosintes including *Zea mays* ssp. *mexicana*, *Zea mays* ssp. *huehuetenangensis*, *Zea perennis*, *Zea diploperennis*, *Zea luxurians* and *Tripsacum dactyloides*, which were sampled for either phylogenetic or outgroup purposes, were provided by the USDA Agricultural Research Station, Iowa State University, Ames, IA.

PCR and sequencing

Regions of *ra1* were amplified from genomic DNA by PCR using *Ex Taq* polymerase (TaKaRa). Primers were designed to amplify a region of approximately 2400 bp centred on the approximately 560-bp *ra1* coding region. Primers were designed from known maize sequences at GenBank and the PCR was used over 35 cycles, with appropriate annealing temperatures and standard PCR conditions. Amplicons were then purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and cloned into pCR 2.1-TOPO (Invitrogen, Carlsbad,

CA, USA) to separate haplotypes in heterozygous samples. Eight clones from each PCR reaction were then sequenced to control for PCR errors. Sequence data were collected from both strands to minimize ambiguities. Sequences were assembled and edited using the PREGAP and GAP4 software from the Staden package (Staden 1996).

Sequence analysis

Sequences were initially aligned using ClustalW (Larkin *et al.* 2007) as part of the MacVector version 8.1 software (http://www.macvector.com) and further adjusted by hand. All diversity statistics were estimated using DnaSP, version 3.51 (Rozas & Rozas 1999). To test for selection, the Hudson–Kreitman–Aguade (HKA) test (Hudson *et al.* 1987) was employed using the three neutral loci *adh1*, *bz2* and *glb1* (Eyre-Walker *et al.* 1998; Hilton & Gaut 1998; Tenaillon *et al.* 2001). *P*-values from multiple tests were combined as previously described (Whitt *et al.* 2002). Maximum likelihood HKA (MLHKA) tests (Wright & Charlesworth 2004) were conducted using *Zea luxurians* as the outgroup with the previously mentioned three neutral loci. Three separate runs starting from different random seeds were performed, each using a Markov chain length of 1×10^6 simulations. The tests were conducted to compare the fit of *ra1* in a population with a neutral vs. a selected model. A hypothesis of selection at domestication is considered supported if the landrace likelihoods differ significantly, but the teosinte likelihoods do not. Significance was assessed using the likelihood ratio test.

Tajima's *D* (Tajima 1989), Fu and Li's *D* (Fu & Li 1993) and the minimum number of recombination events were estimated using DnaSP, version 3.51 (Rozas & Rozas 1999). The test statistics Tajima's *D* and Fu and Li's *D* were conducted on total sites and significance was assessed using critical values. The minimum number of recombination events (R_m) was obtained using the four-gamete test (Hudson & Kaplan 1985). Fay and Wu's *H*-statistic (Fay & Wu 2000) was calculated and significance assessed using a publicly available web interface (http://www.genetics.wustl.edu/jflab/htest.html). The *H*-statistic determines the level of high-frequency variants (ancestral polymorphisms) to detect hitchhiking. Recombination was examined with the pairwise module in LDhat version 2.1 (http://www.stats.ox.ac.uk/~mcvean/LDhat/).

Phylogenetic analysis

Phylogenies were reconstructed using PAUP* version 4.0b10 (Swofford 2003) and MrBayes version 3.1.2 (Ronquist & Huelsenbeck 2003). Bootstrap support was assessed utilizing a full heuristic search with 1000 bootstrap replicates. Phylogenies made using MrBayes were reconstructed using a Markov chain Monte Carlo (MCMC) algorithm under a general time reversal (GTR) model of evolution. Parameters and priors were set to account for nucleotide frequencies, substitution rates as well as transition/transversion ratios. The branch length prior was set to be uniform and unconstrained. All other parameters and priors were set at default. Two independent runs of four chains each (one cold and three heated) were run for 1 000 000 generations and the first 25 000 trees were removed. All phylogenies were viewed using the software TreeView X version 0.4.1 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

Results

Ear morphology in weak ra1 *mutants*

Maize ears and tassels (inflorescences) develop similarly, by producing a series of different meristem types on the main inflorescence axis (Vollbrecht & Schmidt 2009). The primary inflorescence meristem, located at the growing tip of the inflorescence, produces the axis and initiates second-order meristems on its flanks (Fig. 2a, black box). A few second-order meristems at the base of the tassel produce long branches; the rest of the second-order meristems in the tassel and all of the second-order meristems in the ear, are determinate in that they quickly cease growing. Thus, most second-order meristems produce short, spikelet pair branches that in turn bear third- and fourth-order meristems. The strong *ra1* mutant phenotype is due to indeterminacy of spikelet pair meristems, resulting in an outgrowth of long branches that leads to a highly disorganized ear (Fig. 1f) (Vollbrecht *et al.* 2005).

Ears of weak mutants, such as plants homozygous for the *ra1-63.3359* or *ra1-RS* allele, typically have crooked rows (Fig. 1e), and occasionally produce long branches at the base of the ear in some genetic backgrounds. The developmental basis of this phenotype has not been described. The *ra1-63.3359* allele, which arose spontaneously, has a 4-bp insertion into the stop codon that presumably results in the addition of 17 amino acids to the carboxyl terminus of the RA1 protein (Vollbrecht *et al.* 2005). On the other hand, the *ra1-RS* lesion is predicted to eliminate nine amino acids from the amino terminus of the RA1 protein and kernel rows are crooked on ears from *ra1-RS* mutants (Vollbrecht *et al.* 2005), slightly more so than from *ra1-63.3359* plants. The *ra1-63.3359* and *ra1-RS* mutant phenotypes vary in different genetic backgrounds, but are relatively weak in the Mo17 inbred background (R. Weeks and E. Vollbrecht, unpublished data). After

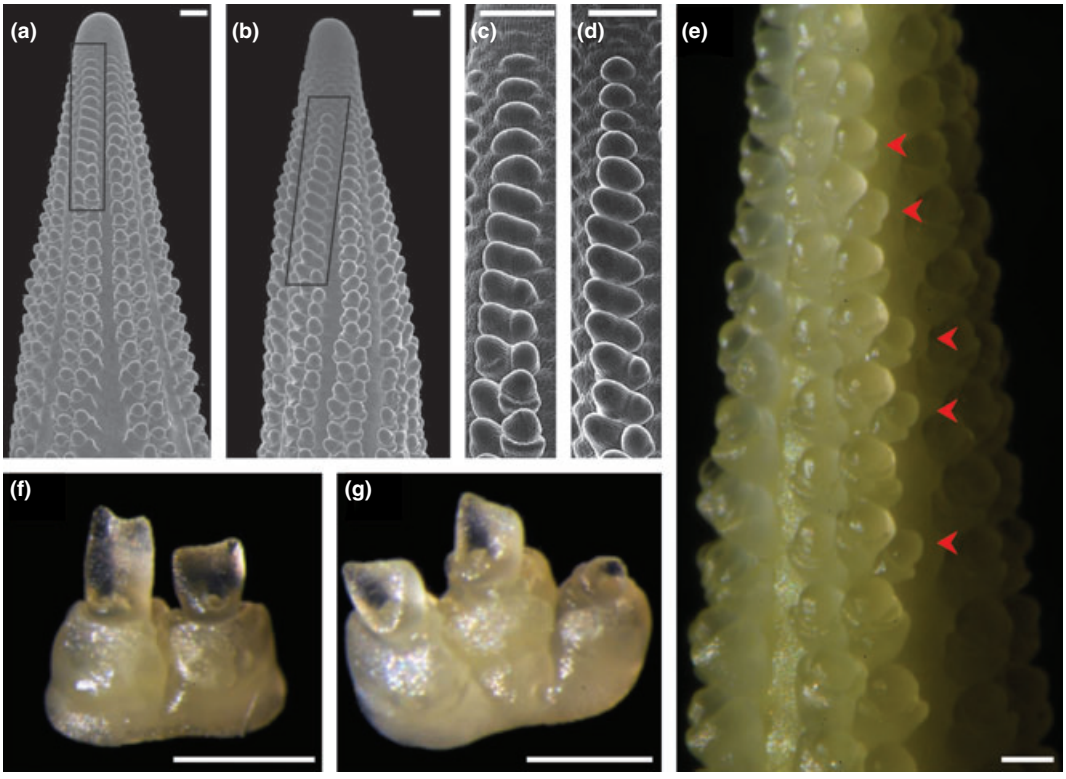

Fig. 2 Ear development of weak *ramosa1* mutants. (a–d) SEMs of developing, 0.5-cm ears from the inbred B73 (a) and the *ra1-RS* weak mutant in B73 (b). Boxed regions are examined in close-up views of a row of spikelet pairs from the B73 ear (c) and the *ra1-RS* ear (d), revealing delayed timing of spikelet pair development and elongated spikelet branches in mutants. (e–f) Portion of a developing (~1.5 cm) ear from a *ra1-63.3359* weak mutant. Red arrows point to developing spikelet triplets (e). Some positions on mutant ears, and all positions on ears from inbreds, bear spikelet pairs (f) but only mutants form spikelet triplets, which go on to develop triplets of florets (g). Scale bars: 200 μm.

introgressing these alleles into inbred backgrounds, we examined ears at various stages of development (Fig. 2). At very early stages (length 0.5 cm, Fig. 2a and b), *ra1-RS* ears showed normal behaviour of the primary inflorescence meristem. Initiation of second-order meristems was perhaps slightly delayed but otherwise normal. However, second-order meristems became abnormally elongated due to a delay in the initiation and differentiation of third-order meristems (Fig. 2d) when compared with wild type (Fig. 2c). This delay in the initiation of third-order meristems resulted in the beginnings of row disorganization in the ear. At this developmental stage, spikelets still appeared strictly as pairs, as in normal ear development (Vollbrecht & Schmidt 2009), but these positions eventually produced spikelets in groups of three (i.e. triplets) in the mutants (Fig. 2e). For example, of developing top ears from 12 *ra1-63.3359* mutants in the weaker Mo17 background (length 0.8–2.1 cm), five ears had in addition to many paired spikelets an average of 10 spikelet triplets per ear (Fig. 2e–g), and seven ears had no spikelet triplets (Table 1). When the *ra1-63.3359* mutant was examined in the B73 background, spikelet triplets were present in all 12 ears, and with greater frequency (Table 1). For *ra1-RS*, the same pattern was observed where more spikelet triplets were present in the B73 vs. the Mo17 genetic background. By contrast, 12 ears of each control (normal inbred) contained no spikelet triplets. In each inbred background *ra1-RS* mutants consistently produced more spikelet triplets, and had relatively more disturbed kernel rows at maturity, than *ra1-63.3359*, indicating that *ra1-RS* is a stronger mutant allele. Thus, the delayed initiation of third-order meristems was sometimes accompanied by initiation of an extra (third) spikelet. The presence of spikelet triplets in weak *ra1* mutants and their absence in wild-type ears strongly suggests these extra spikelets create disorganized rows. These data imply weak loss of function alleles of *ra1* condition crooked rows, supporting the hypothesis that natural variation in the *ramosa* pathway could also manifest as altered kernel rowing, and serve as the basis for a visibly selectable trait during domestication.

Table 1 Frequency of triplets in weak *ra1* mutants in Mo17 and B73 backgrounds

		Quartile (ears)[b]			
Genotype	*N*[a]	0	I	II	III
ra1-63 > Mo17	12	4	7	1	0
ra1-63 > B73	12	0	0	8	4
ra1-RS > Mo17	16	1	5	9	1
ra1-RS > B73	12	0	0	2	10

[a]Total number of ears sampled.
[b]Number of ears in each quartile. Quartile 0 = ears with no spikelet triplets, I = ears with 1–25% of total spikelets as triplets, II = 26–50% and III = 51–75% respectively. No ears had >75% of total spikelets as triplets.

Nucleotide diversity of the ra1 *locus*

To characterize nucleotide diversity at *ra1* within the genus *Zea*, we isolated and sequenced alleles of the *ra1* gene from 43 different plant accessions including 22 maize landraces, 11 *Z. m. parviglumis*, five *Z. m. mexicana* and one each of *Z. m. huehuetenangensis*, *Z. perennis*, *Z. diploperennis*, *Z. luxurians* and *T. dactyloides* (Table 2). Accessions were chosen across the geographic range of maize and teosinte focusing on central and South America in order for the sampling to be representative of both teosinte and maize landrace diversity. Sequence analysis of the 2.4-kb amplicon, which in addition to the *ra1* coding sequence includes over 1200 bp of 5′-noncoding sequence and approximately 600 bp of 3′-noncoding sequence, identified 18 distinct haplotypes within the landrace sampling and 18 within the *Z. m. parviglumis* (teosinte) sampling. For landraces, two haplotypes were represented multiple times within the sampling comprising 24% and 12% of the alleles, whereas all other alleles were unique. For the teosinte population, all haplotypes were unique. There were no shared haplotypes between maize landraces and teosintes. Nucleotide polymorphism (θ) (Watterson 1975) and nucleotide diversity (π) (Nei 1987) were estimated (Table 3). To visualize variation in polymorphism throughout the 2.4-kb *ra1* region, a sliding window analysis of π was performed (Fig. 3).

In the maize landraces, nucleotide diversity is unusually low across the entire region, within the range of $0 \leq \pi < 0.005$ as expected for domestication loci (Hufford *et al.* 2007), and differs significantly from the expected values of $\pi_{average} = 0.0087$ for neutral genes (*t*-test, $P << 0.01$). Within the known functional motifs including the zinc finger DNA-binding domain, a post-zinc finger, putative EAR repression motif (B. Sigmon and E. Vollbrecht, unpublished data), and the previously identified terminal EAR motif, the landraces have virtually no nucleotide diversity. When the nucleotide

Table 2 Names and origins of plant materials

Sample	Origin	Accession	GenBank
Sallu-yah (Cherokee)	USA	PI 213744	GQ891946
Bear Island Chippewa	USA	PI 213801	GQ891944–GQ891945
6 Nations Obsweken	USA	Ames 2355	GQ891943
Eagle Corn	USA	PI 222285	GQ891942
Celaya	Mexico	NSL 2839	GQ891923–GQ891924
Jala	Mexico	NSL 2834	GQ891936
Cariaco	Mexico	PI 260381	GQ891927
Conico	Mexico	NSL 2837	GQ891940
Chapalote	Mexico	PI 420245	GQ891939
Michoacan 286	Mexico	PI 629226	GQ891929
Costa Rica 370	Costa Rica	PI 498486	GQ891932–GQ891933
Boyaca 476	Columbia	PI 444174	GQ891926
Magdalena 399	Columbia	PI 444954	GQ891925
Magdalena 469	Columbia	PI 445007	GQ891935
Cajamarca 15	Peru	PI 571896	GQ891941
Puno 26	Peru	PI 571808	GQ891938
Cuzco 63	Peru	PI 485333	GQ891928
ARG 2334	Argentina	NSL 6539	GQ891922
Argentine Pop	Argentina	PI 451691	GQ891937
Uruguay 756A	Uruguay	PI 477694	GQ891930
Uruguay 293A	Uruguay	PI 479149	GQ891931
White soft corn, Lenha	Brazil	NSL 20134	GQ891934
Zea mays ssp. *parviglumis*	Mexico	Ames 21797	GQ891915
Zea mays ssp. *parviglumis*	Mexico	PI 566686	GQ891921
Zea mays ssp. *parviglumis*	Mexico	PI 566688	GQ891916
Zea mays ssp. *parviglumis*	Mexico	PI 566691	GQ891917–GQ891918
Zea mays ssp. *parviglumis*	Mexico	PI 566692	GQ891919–GQ891920
Zea mays ssp. *parviglumis*	Mexico	Ames 21889	GQ891896–GQ891898
Zea mays ssp. *parviglumis*	Mexico	PI 384062	GQ891899–GQ891902
Zea mays ssp. *parviglumis*	Mexico	PI 384063	GQ891903–GQ891904
Zea mays ssp. *parviglumis*	Mexico	PI 384064	GQ891905–GQ891906
Zea mays ssp. *mexicana*	Mexico	PI 566674	GQ891913
Zea mays ssp. *mexicana*	Mexico	PI 566680	GQ891914
Zea mays ssp. *mexicana*	Mexico	PI 566687	GQ891891–GQ891892
Zea mays ssp. *mexicana*	Mexico	PI 566691	GQ891893–GQ891895
Zea mays ssp. *mexicana*	Mexico	PI 566697	GQ891911
Zea mays ssp. *huehuetenangensis*	Guatemala	PI 441934	GQ891910
Zea perennis	Mexico	Ames 21882	GQ891907
Zea diploperennis	Mexico	PI 441931	GQ891908–GQ891909
Zea luxurians	Guatemala	PI 441933	GQ891912
Tripsacum dactyloides	USA	PI 421612	GQ891890

diversity analysis is partitioned into 5′-coding and 3′-components separately, the immediate 3′-sequence has the lowest nucleotide diversity. The π value for this region is 0.00065, approximately fourfold lower than values for both the coding and immediate 5′-regions of the gene. This π-value is among some of the lowest values found to date in surveys of maize landraces (Yamasaki *et al.* 2005, 2008; Hufford *et al.* 2007).

By contrast, *ra1* nucleotide diversity in the teosintes is 0.013 across the entire region and does not significantly vary among 5′-coding and 3′-regions of the gene. These levels of diversity do not differ significantly from that expected for the average teosinte gene of $\pi = 0.012$ (Hufford *et al.* 2007) (*t*-test, $P > 0.5$). There is higher diversity in known functional regions of the gene (zinc finger and two EAR motifs) compared with the landraces, but the changes are all synonymous. In summary, these results of lower than average nucleotide diversity in maize landraces and average diversity in teosinte indicate that genetic diversity at *ra1* was significantly reduced when the landraces were derived from teosinte during the initial domestication process.

Similarly, patterns of retention of nucleotide diversity at *ra1* also match expectations for a domestication locus. Low gene diversity in maize landraces is consistent with selection, but diversity can also vary due to differential functional constraints on sequence evolution. The ratio of landrace nucleotide diversity to teosinte diversity (π_{lr}/π_{teo}) has been used to measure retention of genetic diversity correcting for functional constraints (Clark *et al.* 2004). For neutral genes a range of 60–80% retention is expected, whereas a lower ratio is indicative of selection (Zhang *et al.* 2002; Hufford *et al.* 2007). For *ra1*, the low diversity 3′-noncoding region in the landraces retained only 5% of the diversity found in the teosinte panel, but the 5′-noncoding and coding regions retained approximately 20% and 17% respectively. These low levels of genetic retention reinforce the notion that much of the reduction in diversity, especially for the 3′-region, occurred during the domestication of maize from its progenitor teosinte.

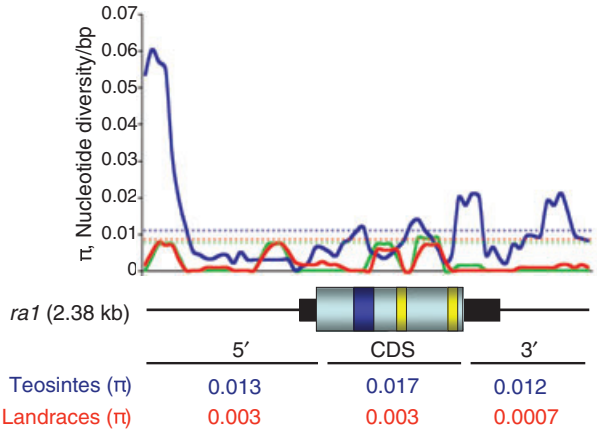

Fig. 3 Patterns of nucleotide diversity at *ramosa1*. A sliding window analysis of nucleotide diversity (π) at *ra1* is shown for maize inbreds (solid green line), landraces (solid red line) and teosintes (solid blue line) across both noncoding and coding sequence of the *ra1* locus compared with π for a corresponding average gene (dotted lines) (Hufford *et al.* 2007). Within the gene cartoon, the blue box represents the zinc finger and the yellow boxes represent putative EAR repression motifs; narrower black rectangles indicate UTRs. The lengths of the multiple sequence alignment for each region are 1236 bp for 5′-noncoding sequence (including the 60 bp, 5′-UTR), 566 bp for CDS, and 578 bp for 3′ (including the 164 bp, 3′-UTR). Significantly reduced diversity in both the 5′- and 3′-noncoding sequences for maize landraces is indicative of past selection. Step size = 100 bases, window size = 25 bases.

Tests of selection

Although a number of selection tests are commonly used, the HKA test (Hudson *et al.* 1987) is among the most widely used for maize candidate domestication loci due to its high statistical power to detect positive

Table 3 Nucleotide diversity statistics for *ramosa1*

		Landraces					Teosintes				
	bp	n^a	S^b	π^c	θ^d	R^e_m	n^a	S^b	π^c	θ^d	R^e_m
5′	1270	25	13	0.002	0.004	2	18	51	0.013	0.022	7
CDS	560	25	6 (3)	0.003	0.004	0	18	22 (6)	0.017	—	0
3′	576	25	4	0.0007	0.002	0	18	37	0.013	0.022	0
Total	2408	25	23 (3)	0.002	0.003	2	18	110 (6)	0.013	—	8

[a]Total number of sequences.
[b]Number of segregating silent plus synonymous, and nonsynonymous (in parentheses), sites.
[c]Nucleotide diversity per site.
[d]Nucleotide polymorphism per site.
[e]Minimum number of recombination events.

selection by combining information from comparative and population genetic data (Zhai *et al.* 2009). The HKA selection test is a stringent test for departure from neutrality that examines whether selection has significantly altered diversity at a locus relative to changes in diversity of neutral, control loci. We applied this test to our *ra1* data using *adh1*, *bz2* and *glb1* as the neutral genes (Tenaillon *et al.* 2001; Tiffin & Gaut 2001) and both *T. dactyloides* and *Z. luxurians* as outgroups (Table 4). Tripsacum is preferable as an outgroup for tests of selection because it is in the sister genus to *Zea* and does not naturally interbreed with maize, but if tripsacum sequence cannot be recovered for a portion of a gene then a distant outgroup within genus *Zea* may be utilized (Hanson *et al.* 1996; Hilton & Gaut 1998; Clark *et al.* 2004). Thus, for some tests we used *Z. luxurians* as an outgroup, with controls as described below. *Zea luxurians* is indigenous to southern Guatemala and therefore geographically isolated from most other teosintes, and there has been only minimal, historical gene flow between *Z. luxurians* and maize (Fukunaga *et al.* 2005; Ross-Ibarra *et al.* 2009). Initial HKA tests using tripsacum as the outgroup (Table 4) and a ~1.3-kb fragment that contains the *ra1* coding sequence plus some 5′ and flanking sequence and a small portion of the 3′-UTR showed a departure from neutrality for maize landraces. The departure was due to reduced diversity at *ramosa1*. By contrast, HKA tests performed on the same fragment for teosintes were not significant. These results strongly support the hypothesis that the *ra1* region experienced positive directional selection specifically during domestication.

In order to investigate whether or not a particular region of *ra1* may have been the target of selection, pairwise HKA tests were performed individually on the 5′-noncoding, coding and 3′-noncoding sequence. As we did not recover much 3′-noncoding sequence from tripsacum despite attempts using a variety of approaches, and to enable analyses of more extensive flanking sequence in the tests, *Z. luxurians* was used as the outgroup. The full 2.4-kb region (see Fig. 3) was easily recovered from *Z. luxurians*. As a control we first used *Z. luxurians* as the outgroup in HKA tests for the same 1.3-kb region previously analysed relative to tripsacum and compared the results with the two different outgroups. As for tests relative to tripsacum, control tests relative to *Z. luxurians* were similarly significant for the landraces but not for teosinte. Thus, results did not differ according to outgroup, validating the use of *Z. luxurians* to analyse the more extensive region. For the larger, 2.4-kb region, HKA tests for maize landraces were significant when the whole region was tested, or when either 5′- or 3′-noncoding sequence were tested, but not when coding sequence was analysed alone (Table 4). All of the analogous HKA tests for teosinte did not differ significantly from neutrality (Table 4), indicating no evidence for selection on *ra1* in *Z. m. parviglumis* as it diverged from its most recent shared ancestor with *Z. luxurians*. When two more divergent teosinte haplotypes, evident from the phylogenetic analysis (Fig. 3), were removed from the HKA analysis the tests were still not significant (data not shown). Therefore, even the remaining teosinte haplotypes, although more closely related to landrace haplotypes, are divergent enough to reveal selection in the landraces. These results imply that nucleotide diversity decreased in derivation of the landraces, i.e. during domestication. Moreover, the

Table 4 Tests of selection for *ramosa1* for all columns

			HKA test *P*-values					Statistics			
	bp	Outgroup	Sites	*adh1*	*bz2*	*glb1*	Aggregate	MLHKA	Tajima's *D*	Fu & Li's *D*	Fay & Wu's *H*
Landraces											
5′	1270	*Zea luxurians*	829	**2.4E-03**	**1.8E-04**	**3.0E-05**	**4.5E-09**	**2.4E-03**	−1.42	−1.62	0.76
CDS	560	*Zea luxurians*	127.9	0.58	0.43	0.25	0.48	0.34	−0.32	−0.26	0.85
3′	576	*Zea luxurians*	495	**1.6E-03**	**3.3E-04**	**3.6E-04**	**5.3E-08**	**5.4E-04**	**−1.89**	**−2.83**	0.32
Total	2408	*Zea luxurians*	1451.9	**9.3E-04**	**5.0E-05**	**3.5E-04**	**5.4E-09**	**1.9E-03**	−1.40	−1.84	0.83
Total	1324	*Tripsacum dactyloides*	634.6	0.07	**1.8E-02**	**5.1E-03**	**5.4E-04**	—	—	—	—
Teosintes											
5′	1270	*Zea luxurians*	691	0.60	0.35	0.37	0.53	0.76	−1.55	−1.02	0.97
CDS	560	*Zea luxurians*	128.1	0.78	0.95	0.91	0.99	0.57	−1.69	−1.70	0.82
3′	576	*Zea luxurians*	492	0.49	0.29	0.30	0.39	0.54	**−1.83**	−1.39	0.86
Total	2408	*Zea luxurians*	1311.1	0.64	0.35	0.40	0.56	0.62	−1.72	−1.34	0.96
Total	1324	*Tripsacum dactyloides*	620.8	0.84	0.62	0.38	0.78	—	—	—	—

Statistically significant values are in bold.

extremely low nucleotide diversity of the 3′-region and lack of a detectable signature in the coding region, together suggest that a target of selection may be located in a regulatory region, perhaps in the 3′-direction of the *ra1* coding sequence.

In addition to pairwise HKA tests we used an MLHKA test that also compares polymorphism within species and divergence between species but allows for an explicit test of selection at a locus using multilocus data. Departure from neutrality is assessed by the likelihood ratio test of a locus compared with neutral reference genes (Wright & Charlesworth 2004). MLHKA tests also provided strong evidence of selection in the landraces for both 5′- and 3′-noncoding regions of *ra1*, but not the coding sequence, and no evidence of selection on the *ra1* locus in teosinte (Table 4).

The maximum likelihood estimate of the selection parameter (k) measures the degree of reduction caused by selection (Wright & Charlesworth 2004). Neutral genes are expected to have $k = 1$, whereas those under strong selection are expected to have values of $k > 2$ or $k < 0.5$ (Moeller & Tiffin 2005). For *ra1*, the landraces had $k < 0.5$, but the 5′- and 3′-noncoding regions had significantly lower values ($k = 0.1$ and 0.06) than the coding region ($k = 0.42$), indicating an extreme loss of diversity in these regions (Table 4). By contrast, under the model that assumes *ra1* is under selection for all regions in teosinte, no regions have values of $k < 0.5$ (Table 4), indicating that the region does not have a significant reduction in diversity in teosinte. Compared with other putatively selected maize genes, k values for the 5′- and 3′-noncoding regions of *ra1* are comparable with those for other regions in the maize genome that exhibit evidence of a selective sweep (Moeller & Tiffin 2005; Wright *et al.* 2005; Camus-Kulandaivelu *et al.* 2008; Tian *et al.* 2009). These results suggest that the nucleotide diversity, especially in noncoding regions of *ra1*, has been greatly reduced by selection.

In addition to performing HKA tests, we assessed non-neutral evolution by calculating Tajima's *D*, Fu and Li's *D* and Fay and Wu's *H*-test statistics. A significantly negative Tajima's *D*-test statistic indicates an excess of low frequency of polymorphism, which is consistent with directional selection or population expansion (Tajima 1989). Similarly, a significantly negative Fu and Li's *D* also indicates directional selection, although this test statistic is based on the number of singletons in a sample (Fu & Li 1993). For *ra1*, Tajima's *D* was negative for both landraces and teosinte, however, only the value for the 3′-region was significant. Fu and Li's *D* was also negative for both populations but only significant for the 3′-region in maize landraces. Fay and Wu's *H*-statistic (Fay & Wu 2000) was calculated for the whole *ra1* 2.4-kb sequence and for partitions of the gene, in both the teosinte and maize landrace populations. The *H*-statistic detects the prevalence of high-frequency variants in order to detect hitchhiking. For teosintes all of the calculated *H*-statistics were similar in value, and none were significant (Table 4). Similarly, in the landraces none of the *H*-statistic values were significant enough to suggest evidence of hitchhiking, although the value for the 3′-region in the landraces was the lowest (Table 4). As this region only has four low-frequency variants (singletons) and no high-frequency variants (ancestral polymorphisms), there may not be enough sequence polymorphism in this partition of the gene to detect hitchhiking. Alternatively, a significant Tajima's *D*, which detects low-frequency variants, and a nonsignificant value for the *H*-statistic may indicate a region recovering from a recent bottleneck where all the ancestral polymorphisms were removed from the population (Fay & Wu 2000).

These test statistics have low power to detect selection especially if few segregating sites are considered, as is the case for *ra1* ($S_{\text{maize}} = 23$, $S_{\text{teosinte}} = 110$); demographic issues also influence results. Thus, Tajima's *D*, Fu and Li's *D*, and Fay and Wu's *H*-test statistics often cannot stand alone, to definitively determine if a signature of selection is indeed due to positive directional selection or due to population bottlenecks or expansions (Sabeti *et al.* 2006). However, when combined with the HKA tests for selection our analyses of Tajima's *D*, Fu and Li's *D* and Fay and Wu's *H* support the inference that the 3′-region of *ra1* was targeted by selection in derivation of maize landraces.

Phylogenetic analysis

For a domestication locus, if a single preferred allele becomes fixed at domestication and there is little subsequent change, then in phylogenetic analysis, the extant maize haplotypes are hypothesized to be confined to a single clade (Wang *et al.* 1999; Clark *et al.* 2004). For neutral loci, on the other hand, maize alleles are expected to scatter among teosinte alleles resulting in a more dispersed tree topology (Goloubinoff *et al.* 1993; Hanson *et al.* 1996; Hilton & Gaut 1998). Thus, although homoplasy, recombination, and population structure considerations may hinder accurate reconstruction of the relationships between teosinte and maize haplotypes, phylogenetic analysis may provide patterns consistent with a hypothesis of selection at domestication. Given the differences we observed in patterns of diversity at *ra1* and the precedent of intralocus variability at the domestication locus *tb1* (Clark *et al.* 2004), we separated our *ra1* data set to assay for differing phylogenetic signals within the data, by reconstructing phylogenies in three separate analyses,

of 5′-noncoding sequence, coding sequence and 3′-noncoding sequence.

The 5′-noncoding sequence data set contained the most haplotype diversity consisting of 43 different haplotypes over 1271 nucleotide sites of which 71 were parsimony informative. The Bayesian tree topology (Fig. 4a) shows all landrace and most teosinte haplotypes to be monophyletic with 100% posterior probability. These results illustrate the close relationship between maize and the subspecies *Z. m. parviglumis* and *Z. m. mexicana* due to their recent divergence. Within this clade, a secondary clade of four landrace haplotypes cluster with a *Z. m. mexicana* haplotype, which suggests that these landrace haplotypes may be more closely related to extant *Z m. mexicana* than to *Z. m. parviglumis*. In addition to the more distant members of *Zea* (ssp. *luxurians*, ssp. *perennis*, ssp. *diploperennis* and ssp. *mays huehuetenangensis*), four *Z. m. parviglumis* and two *Z. m. mexicana* haplotypes were basal to the main clade. Two of these *Z. m. parviglumis* haplotypes clustered with *Z. m. huehuetenangensis*, which may attest to the age of some *Z. m. parviglumis* allelic lineages (Hilton & Gaut 1998).

By contrast, the coding sequence data set consisted of only 29 differing haplotypes over 561 nucleotide sites of which 16 were parsimony informative. The 5′-noncoding and coding sequence trees (Fig. 4b) are similar in that most landrace and teosinte haplotypes are monophyletic and the secondary clade of landrace and *Z. m. mexicana* haplotypes is present in both. In addition, the same two divergent *Z. m. parviglumis* haplotypes in the 5′-tree fall to the base of the coding sequence tree with one clustering with *Z. m. huehuetenangensis*. However, support for this tree topology is low compared with that of the 5′-sequence tree; thus, these results should be interpreted with caution.

The 3′-noncoding sequence data set contained the least haplotype diversity with only 19 haplotypes over 577 total nucleotide sites of which 23 were parsimony informative. The topology of this tree is somewhat different from that of both the 5′-noncoding and coding sequence trees due to the reduction in haplotype diversity in this region (Fig. 4c). This reduction in haplotype diversity can partially be attributed to the absence of the secondary clade of landrace and *Z. m. mexicana* haplotypes. Interestingly, this region also exhibits a reduction in haplotype diversity for all the teosintes, but also exhibits basal placement of the same two *Z. m parviglumis* haplotypes. The extreme sequence conservation and low haplotype diversity of this region

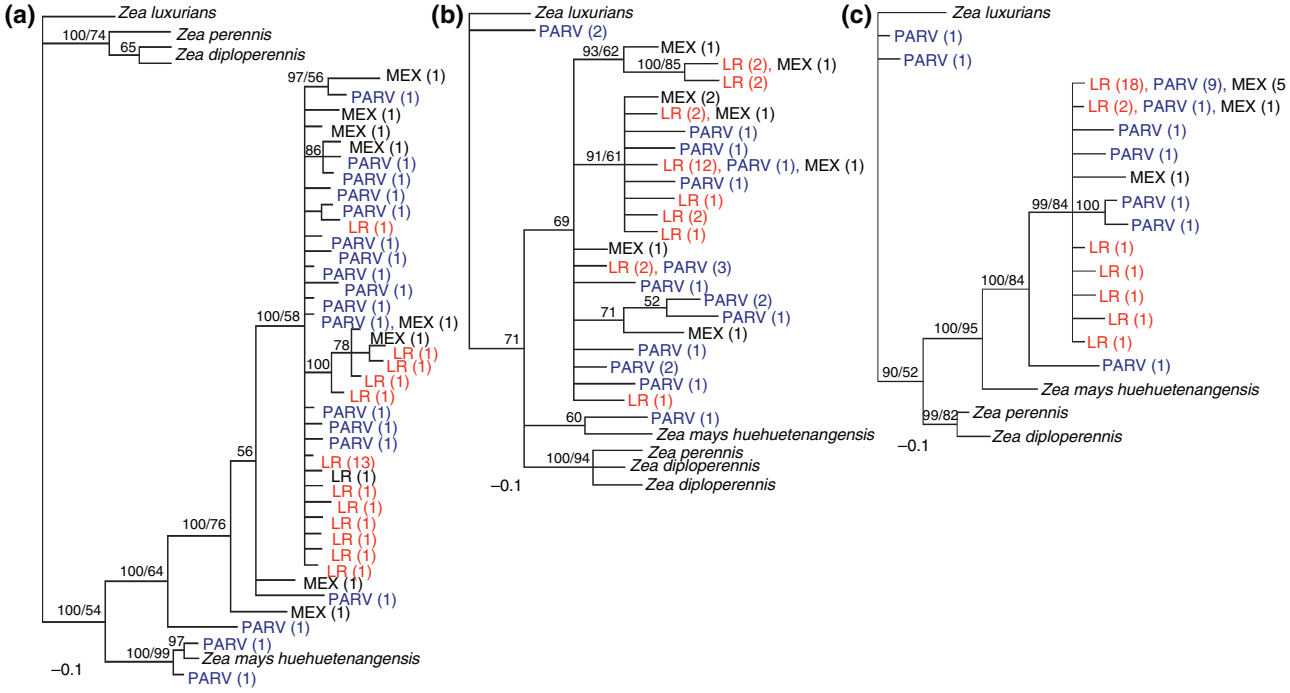

Fig. 4 *ramosa1* haplotype trees. Bayesian phylogenies shown for the (a) 1271-bp 5′-noncoding sequence, (b) 561-bp coding sequence, and (c) 577-bp 3′-noncoding sequence of *ra1*. Posterior probabilities and bootstrap support are shown for each node when >50%. Single values represent posterior probabilities signifying no bootstrap support for that node. *Zea luxurians* sequence was used as the outgroup to orient the tree. Landrace samples are abbreviated as LR, whereas teosintes are abbreviated as PARV (subspecies *parviglumis*) or MEX (subspecies *mexicana*) respectively. Other teosintes are indicated by their full name. The frequency of each haplotype within each population is in parentheses for each taxon.

for both landraces and teosintes suggest that it may have an essential role in *ra1* function and thus be under some additional functional constraint. However, functional analyses in maize have not detected such a function for this region at this time.

Phylogenetic analyses indicated that both the 5′-noncoding and the coding sequence have more haplotype diversity than the 3′-noncoding tree. Despite differing levels of polymorphism, for both regions the maize alleles fall predominantly into one clade with a few forming a second clade; this topology is more neutral than strictly expected for a domestication locus. However, for the 3′-noncoding region, one haplotype predominates in both the maize landrace and teosinte samples. This result is common among probable plant domestication loci, where the hypothesized selected allele is found frequently in the wild progenitor population (Purugganan *et al.* 2000; Nesbitt & Tanksley 2002; Clark *et al.* 2004).

Recombination and linkage disequilibrium

Recombination rates across a region can vary significantly and thus have variable roles in generating haplotype diversity. In maize, the minimum number of recombination events averages about 2.1 in genes and in teosinte the average is slightly higher at 2.7 (Ross-Ibarra *et al.* 2009). In a previous study, no evidence of recombination was found at the *ra1* locus in a diverse inbred population (Vollbrecht *et al.* 2005). For maize landraces we found evidence of a minimum number of two possible recombination events (R_m) using the four-gamete test (Hudson & Kaplan 1985) (Table 3). Eight possible recombination events were estimated for the teosinte population. For both populations, the recombination events are putatively located in the 5′-noncoding regions. However, accurate estimation and comparison of the recombination rate at the *ra1* locus and/or in partitions of it was precluded by the low level of polymorphism (only 10 informative SNPs) in the landrace populations. Similarly, linkage disequilibrium (LD), an estimate of the correlation between different polymorphisms due to shared mutations and recombination histories, was estimated for teosintes only. The analysis suggested two separate LD blocks, located in the 5′- and 3′-noncoding regions (data not shown). Using a cut-off value of 0.1 (Flint-Garcia *et al.* 2003; Palaisa *et al.* 2004) for r^2 (Hudson & Kaplan 1985), LD decays at ~500 bp in this region for the teosintes. This estimate corresponds to the observation that in both maize and teosinte, LD tends to decay within genes (Tenaillon *et al.* 2001; Flint-Garcia *et al.* 2003; Clark *et al.* 2004; Weber *et al.* 2007).

Discussion

Identifying the target of selection

Both the 5′- and 3′-noncoding sequences of *ra1* show evidence of selection in the form of reduced nucleotide diversity, suggesting that a target of selection lies outside the *ra1* coding sequence. The 3′-region of the gene, which includes the 3′-UTR and downstream sequences, has some of the lowest nucleotide diversity found in maize and the lowest *k* values from our MLHKA analysis, which would be consistent with selection occurring in or near it. However, from these analyses we cannot exclude that a target of selection lies in the 5′-region or even outside the analysed region. Extensive sequence and expression analysis provides no evidence of alternative splicing for *ra1* (E. Vollbrecht, unpublished data); so, the 3′-region does not contain coding sequences. BLAST searches show this 3′-sequence to be unique in the maize genome. Hence, while genetic studies to date have not revealed a function for this unique genomic sequence, if the conserved region harbours a target of selection then the target may be a *cis*-regulatory element, or involved in mRNA metabolism. In such a case, altering this sequence would be predicted to have some consequence on gene expression and potentially on phenotype in domesticated maize. Ongoing molecular mutageneses of *ra1*, for example using transposon (Ahern *et al.* 2009), TILLING (Till *et al.* 2004) or related chemical approaches, should prove useful in querying the functional significance of this highly conserved sequence.

A selected locus may be contained within a selective sweep, wherein DNA that neighbours a target of positive selection also contains reduced nucleotide diversity due to hitchhiking. Selective sweeps in maize can be limited or quite extensive. For example, the unidirectional selective sweep around the *teosinte branched1* (*tb1*) locus extends 60–90 kb in the 5′-direction of the gene (Clark *et al.* 2004), whereas an asymmetric sweep at the *yellow1* (*y1*) locus includes a much larger region of up to 600 kb (Palaisa *et al.* 2004). In the case of *y1*, the selective sweep is hypothesized to be so extensive because the locus underwent strong selection more recently than did *tb1*. Recently, a selective sweep in maize spanning 1100 kb and more than 15 genes was identified; however, the target of selection in this region is unknown (Tian *et al.* 2009). The extent of a selective sweep can be measured by sampling nucleotide diversity of low-copy genomic regions nearby and by estimating LD in the region of selection. In the *ra1* region analysed here, levels of polymorphism are too low for comparative analysis of LD or recombination. However,

regional patterns of nucleotide diversity will soon be accessible on a genome-wide scale using genome resequencing techniques and it will be interesting to see patterns of nucleotide diversity and LD in the region surrounding *ra1*. Preliminary analysis indicates *ra1* is imbedded in a region of overall low diversity, although such regions are not uncommon in the maize genome (E.S. Buckler, personal communication). The *ra1* gene is near the centromere; so, a selective sweep around it may be physically quite extensive due to low rates of recombination per kilobase (Fengler *et al.* 2007), aside from the strong LD that may accompany positive selection. Thus, an understanding of genomic patterns of nucleotide diversity around *ra1* may elucidate the boundaries of the selective sweep and therefore the strength and timing of selection at the locus (Olsen *et al.* 2006).

ra1 *as a candidate domestication locus*

Several lines of evidence suggest that the *ra1* region was a target of selection during maize domestication. This hypothesis is chiefly supported by the low nucleotide diversity present at the *ra1* locus and the significant HKA and MLHKA tests for maize landraces but not teosintes. Compared with nucleotide diversity levels in teosinte, the reduction in *ra1* diversity for maize landraces is much greater than expected from the population bottleneck occurring at domestication (Tenaillon *et al.* 2004). Previously, nucleotide diversity of *ra1* was reported for a diverse panel of modern inbred lines (Vollbrecht *et al.* 2005). In those data, the inbreds retained approximately 52% of the diversity found in maize landraces. In the present study, the landraces retained only 5–20% of the teosinte diversity, compared with an expected 60–80% for neutral genes (Zhang *et al.* 2002; Hufford *et al.* 2007). These data suggest that most of the reduction in *ra1* genetic diversity is due to selection during domestication from teosinte, with some further reduction following an improvement bottleneck, conclusions that are also consistent with the *D*- and the *H*-test statistics. Significantly low values for *D* are consistent with recovery from a recent bottleneck and the lack of high-frequency variants estimated by *H* suggests that a strong bottleneck has removed ancestral polymorphisms from the landrace population (Tajima 1989; Fay & Wu 2000). Therefore, it is unlikely that significant amounts of selection occurred during the evolution of teosinte or during the improvement process.

To date, molecular genetic approaches in plants have identified in the order of 10 genes as domestication loci, while another two dozens or so may be classified as post-domestication, crop-diversification genes; almost all of the domestication genes encode transcriptional regulators, while roughly half of the diversification genes encode structural genes like enzymes (Purugganan & Fuller 2009). Thus, a transcription factor like *ramosa1* that controls branching architecture of inflorescences is intuitively a good selection candidate. The weakest known mutant alleles of the *ra1* gene were analysed here, and shown to result in disordered rows on the maize ear. Therefore, we speculate that prehistoric farmers may have selected for straight rows on the ear for purposes of aesthetics and/or effects on grain yield, that in doing so they selected particular, relatively high-activity alleles of *ra1*, and that this artificial selection resulted in reduced genetic diversity for the *ra1* locus. Interestingly, in a study of present-day maize farmers in central Mexico, row straightness was ranked as a desirable or necessary criterion when selecting landrace seeds for propagation (Perales *et al.* 2003). As expected given the great genetic diversity of teosinte, we documented a large variety of *ra1* alleles within the population that could exhibit varying effects on inflorescence branching and, therefore, row formation. The function of *ra1* in teosinte has not been studied. As teosinte ears produce two ranks of solitary spikelets (Sundberg & Orr 1990), a trait such as crooked rows would have been inconsequential in teosinte, but may have become important as ear diameter and number of rows of spikelet pairs per ear increased during the domestication process. In an association mapping study in teosinte, markers for the *zea apetala homolog1* (*zap1*) gene showed a significant association with inflorescence branch number but accounted for only 2.7% of the phenotypic variance (Weber *et al.* 2007). As the frequency of the maize-like allele was found to be at a higher frequency in landraces than in teosintes, *zap1* may have been selected upon during the domestication process (Weber *et al.* 2007). The additive nature of putative domestication loci involving inflorescence architecture traits, like *zap1*, reinforces the notion that many genes responsible for the phenotypic variance of these traits in both teosinte and maize remain undiscovered. In any case, the absence of evidence for selection on *ra1* in teosinte suggests that its coupled trait was not subject to natural selection during the evolution of teosinte and its single-rowed inflorescence, but was subject to it as prehistoric farmers began domesticating teosinte and increasing row number, such that straight rows became important. This hypothesis could be tested by observing the phenotypic consequences in the ear for introgression into maize of various teosinte alleles of *ra1*, by themselves and in complementation tests with mutant maize alleles. Artificial selection on standing variation within the progenitor population is a feature common for plant domestication loci. Other examples include *tb1* in maize (Clark *et al.* 2004), *fw2.2* in tomato (Nesbitt & Tanksley 2002) and *BoCal* in cauliflower and broccoli

(Purugganan *et al.* 2000). To date, *tga1* in maize is the only domestication locus where the cultivated allele was not found in the wild progenitor population (Dorweiler *et al.* 1993; Dorweiler & Doebley 1997; Wang *et al.* 2005).

Our finding that two *ra1* alleles were prevalent within the landrace population at 24% and 12% frequency suggests that more than one allele could have made it through the domestication bottleneck. Phylogenetics could hypothetically be used to address that question, but phylogenetic reconstruction for domestication loci is problematic due both to the fact that low nucleotide diversity regions must be used to build trees, and to past occurrences of introgression and hybridization events between maize and teosinte populations, which complicate accurate reconstruction.

Evolution of the ramosa *pathway and maize inflorescence morphology*

ra1 regulates meristem function as part of a molecular genetic pathway that also includes *ramosa2*, which encodes a LOB domain transcription factor (Bortiri *et al.* 2006), and *ramosa3*, which encodes a trehalose-6-phosphate phosphatase (Satoh-Nagasawa *et al.* 2006). Thus, both *ra1* and *ra2* probably function as regulatory transcription factors in this pathway, while the biochemical role of *ra3* is less clear at this point in time. It has been shown that some genes identified as developmental regulators in maize do affect the natural variation of complex traits in extant teosinte, although the *ramosa* genes were not among the regulators tested, and that this variation may serve as a basis for directional selection during domestication (Weber *et al.* 2007). As originally pointed out by Darwin, artificial selection during domestication has many parallels with natural selection during evolution. Given that the *ramosa* genes regulate branching architecture in maize, it is tempting to speculate that expression differences among these genes or functional differences among their gene products may contribute to branching architecture variation in grass inflorescences (Vollbrecht & Sigmon 2005; McSteen 2006). Both *ra2* and *ra3* are conserved by purifying selection in the grasses (Bortiri *et al.* 2006; Satoh-Nagasawa *et al.* 2006), but putative *ra1* orthologues have only been identified from Panicoid grass species (B. Sigmon, E. Vollbrecht and E. Kellogg, unpublished data), which includes the cereal crops maize, sorghum and foxtail millet, and the Andropogoneae tribe as a subgroup that includes maize, sorghum and others (GPWG, Grass Phylogeny Working Group 2000, 2001). Notably, spikelet pairs are a defining morphological character of the Andropogoneae as most other grasses have spikelets as singlets (Kellogg 2000). As *ra1* acts as a switch from long indeterminate branches to short determinate spikelet pairs in maize, one possibility is that *ra1* was co-opted for this role in spikelet pair development within Andropogoneae grasses.

Little is known about the genetic basis behind the morphological transformation of the maize ear following domestication. Because for *ra1*, noncoding sequences have the lowest nucleotide diversity and exhibit evidence of selection, it is possible that regulation of the gene may have been more important for domestication than protein composition. These observations are consistent with the hypothesis that altered gene regulation can be a principal genetic basis of morphological differences arising from plant domestication (Doebley & Lukens 1998), although there is perhaps equivocal evidence for the importance of protein changes as well (Hoekstra & Coyne 2007; Takeda & Matsuoka 2008). It has been suggested that genes targeted by selection are more likely to be expressed in tissues that underwent drastic modifications during periods of artificial selection (Hufford *et al.* 2007; Zhao *et al.* 2008). To date, the domestication locus *tga1*, which is responsible for the development of 'naked' kernels in maize, is perhaps the only described gene that may help explain a portion of the mystery behind the dramatic changes implicit in the development of the modern maize ear. *tga1* encodes a transcription factor that is only expressed in the ear; thus, simple changes in this gene can have dramatic phenotypic consequence in the ear without further deleterious pleiotropic effects in other tissues (Dorweiler *et al.* 1993; Dorweiler & Doebley 1997; Wang *et al.* 2005). Similarly, *ra1* expression is restricted to developing inflorescences (Vollbrecht *et al.* 2005) which may minimize the likelihood of changes at the locus affecting other tissues in the plant. Given the examples of *tga1* and *ra1*, it seems reasonable that many other domestication loci that have modified the morphology of the maize ear will be transcription factors or other developmental genes with expression limited to inflorescence tissues, and that identifying and characterizing these domestication loci will further our understanding of the evolution of the maize ear. Such information would be potentially very useful for plant breeding projects aimed at modifying maize and other cereal inflorescences to affect grain yields.

Acknowledgments

We thank Thomas Peterson, Erica Unger-Wallace and two anonymous reviewers for constructive comments on the manuscript, and Jeffrey Ross-Ibarra for help with recombination analysis. This research was funded by Iowa State University and National Science Foundation DBI grant 0604923 (to E.V.).

References

Ahern KR, Deewatthanawong P, Schares J *et al.* (2009) Regional mutagenesis using dissociation in maize. *Methods*, **49**, 248–254.

Bortiri E, Chuck G, Vollbrecht E *et al.* (2006) *ramosa2* encodes a lateral organ boundary domain protein that determines the fate of stem cells in branch meristems of maize. *Plant Cell*, **18**, 574–585.

Camus-Kulandaivelu L, Chevin LM, Tollon-Cordet C *et al.* (2008) Patterns of molecular evolution associated with two selective sweeps in the *Tb1-Dwarf8* region in maize. *Genetics*, **180**, 1107–1121.

Casa AM, Mitchell SE, Hamblin MT *et al.* (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theoretical and Applied Genetics*, **111**, 23–30.

Chapman MA, Pashley CH, Wenzler J *et al.* (2008) A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell*, **20**, 2931–2945.

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proceedings of the National Academy of Sciences of the USA*, **101**, 700–707.

Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genetics*, **38**, 594–597.

Doebley J (2004) The genetics of maize evolution. *Annual Review of Genetics*, **38**, 37–59.

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell*, **10**, 1075–1082.

Doebley J, Stec A, Gustus C (1995) *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, **141**, 333–346.

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488.

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.

Dorweiler JE, Doebley J (1997) Developmental analysis of *teosinte glume architecture1*, a key locus in the evolution of maize (Poaceae). *American Journal of Botany*, **84**, 1313–1322.

Dorweiler J, Stec A, Kermicle J, Doebley J (1993) *teosinte glume architecture1*: a genetic locus controlling a key step in maize evolution. *Science*, **262**, 233–235.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences of the USA*, **95**, 4441–4446.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.

Fengler K, Allen SM, Li B, Rafalski A (2007) Distribution of genes, recombination, and repetitive elements in the maize genome. *Crop Science*, **47**, S-83–S-95.

Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, **54**, 357–374.

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

Fukunaga K, Hill J, Vigouroux Y *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics*, **169**, 2241–2254.

Gernert W (1912) A new subspecies of *Zea mays* L. *American Naturalist*, **46**, 616–622.

Goloubinoff P, Paabo S, Wilson AC (1993) Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proceedings of the National Academy of Sciences of the USA*, **90**, 1997–2001.

GPWG (Grass Phylogeny Working Group) (2000) A phylogeny of the grass family (Poaceae), as inferred from eight character sets. In: *Grasses: Systematics and Evolution* (eds Jacobs S, Everett J), pp. 3–7. Commonwealth Scientific and Industrial Research Organization (CSIRO), Collingwood, Vic.

GPWG (Grass Phylogeny Working Group) (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden*, **88**, 373–457.

Hanson MA, Gaut BS, Stec AO *et al.* (1996) Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics*, **143**, 1395–1407.

Hilton H, Gaut BS (1998) Speciation and domestication in maize and its wild relatives: evidence from the *globulin-1* gene. *Genetics*, **150**, 863–872.

Hoekstra HE, Coyne JA (2007) The locus of evolution: *evo devo* and the genetics of adaptation. *Evolution*, **61**, 995–1016.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Hufford KM, Canaran P, Ware DH, McMullen MD, Gaut BS (2007) Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. *Plant Physiology*, **144**, 1642–1653.

Kellogg E (2000) Molecular and morphological evolution in the Andropogoneae. In: *Grasses: Systematics and Evolution* (eds Jacobs S, Everett J), pp. 149–158. Commonwealth Scientific and Industrial Research Organization (CSIRO), Collingwood, Vic.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) ClustalW and ClustalX version 2.0. *Bioinformatics*, **23**, 2947–2948.

Matsuoka Y, Vigouroux Y, Goodman MM *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the USA*, **99**, 6080–6084.

McSteen P (2006) Branching out: the ramosa pathway and the evolution of grass inflorescence morphology. *Plant Cell*, **18**, 518–522.

Moeller DA, Tiffin P (2005) Genetic diversity and the evolutionary history of plant immunity genes in two species of Zea. *Molecular Biology and Evolution*, **22**, 2480–2490.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus *Lycopersicon*. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics*, **162**, 365–379.

Olsen KM, Caicedo AL, Polato N *et al.* (2006) Selection under domestication: evidence for a sweep in the rice *waxy* genomic region. *Genetics*, **173**, 975–983.

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective

sweep. *Proceedings of the National Academy of Sciences of the USA*, **101**, 9885–9890.
Perales RH, Brush SB, Qualset CO (2003) Dynamic management of maize landraces in central Mexico. *Economic Botany*, **57**, 21–34.
Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the USA*, **106**, 5019–5024.
Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848.
Purugganan MD, Boyles AL, Suddith JI (2000) Variation and selection at the *cauliflower* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics*, **155**, 855–862.
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the USA*, **104**(Suppl. 1), 8641–8648.
Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics*, **181**, 1399–1413.
Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.
Sabeti PC, Schaffner SF, Fry B *et al.* (2006) Positive natural selection in the human lineage. *Science*, **312**, 1614–1620.
Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature*, **441**, 227–230.
Smith B (1995) *The Emergence of Agriculture*. Freeman, New York.
Staden R (1996) The Staden sequence analysis package. *Molecular Biotechnology*, **5**, 233–241.
Sundberg MD, Orr AR (1990) Inflorescence development in two annual teosintes: *Zea mays* subsp. *mexicana* and *Z. mays* subsp. *parviglumis*. *American Journal of Botany*, **77**, 141–152.
Swofford D (2003) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Takeda S, Matsuoka M (2008) Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics*, **9**, 444–457.
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, **277**, 1063–1066.
Tenaillon MI, Sawkins MC, Long AD *et al.* (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences of the USA*, **98**, 9161–9166.
Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution*, **21**, 1214–1225.
Tian F, Stevens NM, Buckler ESt (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proceedings of the National Academy of Sciences of the USA*, **106**(Suppl. 1), 9979–9986.
Tiffin P, Gaut BS (2001) Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid Z. diploperennis: insights from four nuclear loci. *Genetics*, **158**, 401–412.
Till BJ, Reynolds SH, Weil C *et al.* (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology*, **4**, 12.
Vollbrecht E, Schmidt RJ (2009) Development of the inflorescences. In: *Handbook of Maize: Its Biology*, pp. 13–40. Bennetzen, Jeff L. and Hake, Sarah C., Springer, New York.
Vollbrecht E, Sigmon B (2005) Amazing grass: developmental genetics of maize domestication. *Biochemical Society Transactions*, **33**, 1502–1506.
Vollbrecht E, Springer PS, Goh L, Buckler ES, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. *Nature*, **436**, 1119–1126.
Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature*, **398**, 236–239.
Wang H, Nussbaum-Wagler T, Li B *et al.* (2005) The origin of the naked grains of maize. *Nature*, **436**, 714–719.
Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.
Weber A, Clark RM, Vaughn L *et al.* (2007) Major regulatory genes in maize contribute to standing variation in Teosinte (*Zea mays* ssp. *parviglumis*). *Genetics*, **177**, 2349–2359.
Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences of the USA*, **99**, 12959–12962.
Wright SI, Charlesworth B (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics*, **168**, 1071–1076.
Wright SI, Bi IV, Schroeder SG *et al.* (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.
Yamasaki M, Tenaillon MI, Bi IV *et al.* (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell*, **17**, 2859–2872.
Yamasaki M, Schroeder SG, Sanchez-Villeda H, Gaut BS, McMullen MD (2008) Empirical analysis of selection screens for domestication and improvement loci in maize by extended DNA sequencing. *The Plant Genome*, **1**, 33.
Zhai W, Nielsen R, Slatkin M (2009) An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Molecular Biology and Evolution*, **26**, 273–283.
Zhang L, Peek AS, Dunams D, Gaut BS (2002) Population genetics of duplicated disease-defense genes, hm1 and hm2, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics*, **162**, 851–860.
Zhao Q, Thuillet A-C, Uhlmann NK *et al.* (2008) The role of regulatory genes during maize domestication: evidence from nucleotide polymorphism and gene expression. *Genetics*, **178**, 2133–2143.

B.S., currently a graduate student at Iowa State University, is interested in the evolution of plant development focusing on

grass inflorescence branching as a system. This paper constitutes a chapter of her dissertation. E.V. is currently an Assistant Professor at Iowa State University. One of his main research interests is to elucidate molecular genetic pathways that regulate shoot branching using the floral shoots of the maize tassel and ear as an experimental system. Some of that research occurs as part of a large collaborative effort, known as the Maize Inflorescence Project. The group uses a variety of approaches to identify genes and pathways involved in all aspects of maize inflorescence development and examine gene function within the comparative context of the diversity of grass inflorescences.