Evolutionary History of Pearl Millet (*Pennisetum glaucum* [L.] R. Br.) and Selection on Flowering Genes since Its Domestication

Jérémy Clotault,†,[1,2] Anne-Céline Thuillet,†,[1] Marylène Buiron,[1] Stéphane De Mita,[1] Marie Couderc,[1] Bettina I.G. Haussmann,[3] Cédric Mariac,[1] and Yves Vigouroux*,[1]

[1]Institut de Recherche pour le Développement, Unité Mixte de Recherche Diversité Adaptation et Développement des Plantes (UMR DIADE), Montpellier, France

[2]Université Montpellier 2, UMR DIADE, Montpellier, France

[3]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Niamey, Niger

†These authors contributed equally to this work.

***Corresponding author:** E-mail: yves.vigouroux@ird.fr.

Associate editor: Hideki Innan

Abstract

The plant domestication process is associated with considerable modifications of plant phenotype. The identification of the genetic basis of this adaptation is of great interest for evolutionary biology. One of the methods used to identify such genes is the detection of signatures of selection. However, domestication is generally associated with major demographic effects. It is therefore crucial to disentangle the effects of demography and selection on diversity. In this study, we investigated selection in a flowering time pathway during domestication of pearl millet. We first used a random set of 20 genes to model pearl millet domestication using approximate Bayesian computation. This analysis showed that a model with exponential growth and wild–cultivated gene flow was well supported by our data set. Under this model, the domestication date of pearl millet is estimated at around 4,800 years ago. We assessed selection in 15 pearl millet DNA sequences homologous to flowering time genes and showed that these genes underwent selection more frequently than expected. We highlighted significant signatures of selection in six pearl millet flowering time genes associated with domestication or improvement of pearl millet. Moreover, higher deviations from neutrality were found for circadian clock–associated genes. Our study provides new insights into the domestication process of pearl millet and shows that a category of genes of the flowering pathway were preferentially selected during pearl millet domestication.

Key words: domestication bottleneck, adaptation, flowering time, signatures of selection, approximate Bayesian computation, pearl millet.

Introduction

Plant domestication is associated with major morphological modifications to fit human needs. Some studies support the hypothesis that the very strong morphological differentiation we observe today was progressively selected over hundreds to thousand of years (Smith 2001; Tanno and Willcox 2006; Fuller 2007) and surprisingly that the speed of selection by humans was similar to that of natural selection (Purugganan and Fuller 2011). Consequently, the study of the domestication process provides an interesting glimpse of selection acting on morphological traits at a relatively small timescale. Over the last 20 years, the study of the domestication process led to the identification of several key domestication genes using quantitative trait loci (QTL) mapping experiments and positional cloning. Genes associated with plant and inflorescence architecture were cloned in maize (Doebley et al. 1995), rice (Jin et al. 2008), and barley (Komatsuda et al. 2007). Another very important trait associated with cultivated crops is the absence of seed shattering. This trait was selected over hundreds of years (Fuller et al. 2009). The genetic bases of this trait were recently identified in rice (Konishi et al. 2006; Li et al. 2006) and wheat (Simons et al. 2006). One of the genes conferring seed casing, another important trait associated with the domestication process, was also recently cloned in maize (Wang et al. 2005).

Another method used to understand the genetic impact of human selection is genome selection scanning (Vigouroux, McMullen, et al. 2002; Casa et al. 2005; Vigouroux et al. 2005; Wright et al. 2005; Yamasaki et al. 2005; Papa et al. 2007; Chapman et al. 2008; Wang, Shen, et al. 2010). This method relies on the analysis of natural genetic diversity to identify genomic regions shaped by selection. However, the identification of genes under selection during or after the domestication process is considerably hampered by the demographic associated processes. In particular, genetic bottlenecks associated with domestication mimic positive selection signature. This problem was taken into account by using demographic models to simulate diversity evolution in crops (Vigouroux, McMullen, et al. 2002; Tenaillon et al. 2004). Inferences of scenarios of demographic history have been facilitated by the development of approximate Bayesian computation (ABC). This approach enables complex demographic

scenarios to be built and tested (Beaumont et al. 2002; Csilléry et al. 2010). The scenario retained is then used as a benchmark to measure deviation from the expected diversity pattern (Tenaillon et al. 2004; De Mita et al. 2007).

Some morphological and biochemical changes linked to the domestication process may have involved genes from regulatory pathways. To better understand how evolution acted, it is interesting to know whether some genes of the considered pathway have been preferentially selected or not (Cork and Purugganan 2004). To our knowledge, very few studies have focused on selection acting on genes networks during domestication (but see Whitt et al. 2002; Yu et al. 2011). One reason is that not all developmental pathways are sufficiently well described. One important pathway for crop domestication, adaptation, and improvement that has been sufficiently well studied from a functional point of view is the pathway governing flowering time. Numerous studies performed in *Arabidopsis* have elucidated this developmental signaling network (Blázquez 2000; Blázquez et al. 2001; Komeda 2004). The light-dependent pathway includes photoreceptors that perceive light exposure (Mockler et al. 2003). Photoperiod information is integrated by the pathway controlling the circadian clock (Hayama and Coupland 2003). The response to temperature is influenced by the vernalization pathway (Michaels and Amasino 2000). Plants can be categorized according to their response to vernalization and day length: short-day plants, for example, rice and maize, require no vernalization, whereas for long-day plants, for example, *Arabidopsis*, wheat, barley, vernalization represents an important flowering signal (Cockram et al. 2007). Two other developmental pathways also influence floral initiation: the autonomous and gibberellin pathways (Wilson et al. 1992; Simpson 2004). Environmental and developmental stimuli are then conveyed to floral integrators, which stimulate (or not) floral meristem identity genes (Blázquez et al. 2001; Mouradov et al. 2002). Some studies have focused on specific genes of this pathway for their role during domestication or crop improvement (Purugganan et al. 2000; Thornsberry et al. 2001; Ducrocq et al. 2008; Blackman et al. 2011). In wheat and barley, vernalization (*Vrn*) and photoperiod (*Ppd*) genes are involved in variations in flowering time (Cockram et al. 2007). In rice, several different QTLs for flowering time were found and cloned in early and late flowering cultivars and may have been involved in adaptation to cold regions (Izawa 2007; Xue et al. 2008). *Dwarf8*, a gene involved in the gibberellin pathway, is associated with flowering time in maize and may have been targeted by diversifying selection in contrasted climates during maize history (Thornsberry et al. 2001; Camus-Kulandaivelu et al. 2006). However, a more thorough analysis of the impact of human selection on the flowering pathway is still lacking.

In the present study, we performed an analysis of the evolution of the flowering pathway of pearl millet (*Pennisetum glaucum* [L.] R. Br.). Pearl millet is a major cereal crop in West Africa. It is also widely grown in eastern and southern Africa and in semiarid to arid areas of India. The current hypothesis concerning its domestication suggests it originated in the northern–central Sahel in West Africa (Oumar et al. 2008). The oldest archaeobotanical evidence of pearl millet cultivation was found in Mali and dated at around 4,500 BP (Manning et al. 2011). The domestication process of pearl millet is associated with common morphological changes among cereals: suppression of spikelet shedding, reduction in the size of bristles and bracts, increase in seed size, increase in spikelet pedicel length, loss of dormancy, reduction in the number of basal tillers, and an increase in spike length (Poncet et al. 1998). Interestingly, although the origin of cultivation is hypothesized to have occurred in the dry areas of the Sahel, pearl millet is also cultivated further south in more humid areas. Consequently, in West Africa, pearl millet displays a wide range of flowering times. Varieties from Sahelian areas flower very early (as early as 40 days after planting), whereas varieties from the tropical coast may flower very late (up to 150 days after planting) (Haussmann et al. 2006). Interestingly, there is a correlation between time to flowering and annual rainfall (Haussmann et al. 2006), and we can thus assume that control of flowering time was a major trait for the adaptation of cultivated pearl millet to wetter areas. Some genes associated with flowering time in pearl millet have been recently discovered (Saïdou et al. 2009; Mariac et al. 2011), but the role of flowering genes in pearl millet's adaptation to climate during domestication remains unknown.

In this paper, we first document the evolutionary history of pearl millet domestication by studying a set of random genes in wild and cultivated pearl millet populations. The model built is then used to evaluate the contribution of several genes in the flowering pathway to domestication and adaptation in pearl millet, another short-day plant model, besides rice and maize.

Materials and Methods

Plant Materials and DNA Sequencing

Seeds of 33 cultivated pearl millet individuals (*P. glaucum* subsp. *glaucum*) and 13 wild pearl millet individuals (*P. glaucum* subsp. *monodii*) were collected in West Africa. Sampling was designed to cover the geographical distribution of wild and cultivated pearl millet in West Africa (supplementary table S1, Supplementary Material online). Plants were self-pollinated from three to six generations to reduce heterozygosity in this primarily outcrossing species and also to allow direct sequencing without cloning. Two samples of the related species *P. polystachion* were used to orientate ancestral versus derived polymorphisms. DNA was extracted using previously published methods (Mariac et al. 2006).

We amplified 20 random genes to model the evolutionary history of the sample. These genes were chosen regardless of their biological function (Feltus et al. 2006). Details are provided as supplementary data in supplementary table S2 (Supplementary Material online). Seventeen genes were directly sequenced using the primers designed by Feltus et al. (2006). Three primer pairs gave unspecific

amplification, and gene sequences were obtained after amplicon cloning with newly designed primers.

Fifteen primer pairs were designed to amplify potential homologs of flowering genes. Primer design was based on rice, maize, sorghum, and/or *Arabidopsis* alignments for genes involved in the light-dependent pathway, the autonomous pathway, circadian clock control, or the integration pathway (Higgins et al. 2010). The involvement of each gene in one or other of these pathways is listed in supplementary table S3 (Supplementary Material online). Both gDNA and cDNA sequences were used to design primers. The obtained fragments were checked for their sequence homology with targeted genes. For the gene *PgPHYC*, after the initial sequencing of one fragment (named *PgPHYC*), we designed primers for five more fragments (supplementary table S3, Supplementary Material online) of this gene (named *PgPHYC6*, *PgPHYC6bF*, *PgPHYC7*, *PgPHYC9*, *PgPHYC10*).

Polymerase chain reactions (PCR) were performed as previously described (Saïdou et al. 2009). Primers and Tm are shown in supplementary table S2 (Supplementary Material online). Two independent PCR reactions were performed for each gene. PCR products were purified using Ampure kits (Agencourt Bioscience). An independent reaction sequence was performed using each PCR reaction. Sequence reactions were performed using BigDye v3.1 Terminator kit (Applied Biosystems, Foster City, CA). Sequence reactions were purified using CleanSeq kits (Agencourt Bioscience). Sequences were run on an ABI 3130 XL automated sequencer (Applied Biosystems). Sequence checking and alignments were made using Geneious (Drummond et al. 2010). Haplotypes for heterozygous sequences were solved using PHASE (Stephens et al. 2001).

Gene Diversity Statistics

Diversity statistics were computed using DnaSP 5.10 (Rozas et al. 2003). Analyses were first performed considering only single nucleotide polymorphisms (SNPs). A second analysis was run considering both SNPs and simple insertions/deletions (indels), that is, excluding microsatellites or nested indels. For each gene, nucleotide polymorphism θ_w (Watterson 1975), nucleotide diversity π (Nei 1987), Tajima's *D* (Tajima 1989), and Fay and Wu's *H* (Fay and Wu 2000) were calculated for both wild and cultivated pearl millet groups. The differentiation index F_{ST} (Hudson et al. 1992) was calculated between wild and cultivated pearl millet. The Mann–Whitney test was used to compare the rank of these statistics for flowering and random genes or cultivated and wild populations (Mann and Whitney 1947). A chi-squared goodness of fit test was used to compare Tajima's *D* or Fay and Wu's *H* observed for random genes with expectations under the Wright–Fisher model. Under the null hypothesis, a quarter of random genes would show a *D* or *H* included in each quartile of the distribution according to the Wright–Fisher model. *P* values for chi-squared goodness of fit test were obtained by computing 10^4 Markov chain Monte Carlo (MCMC) simulations. The population structure of the sample was checked using the Bayesian approach implemented in the STRUCTURE 2.3 software

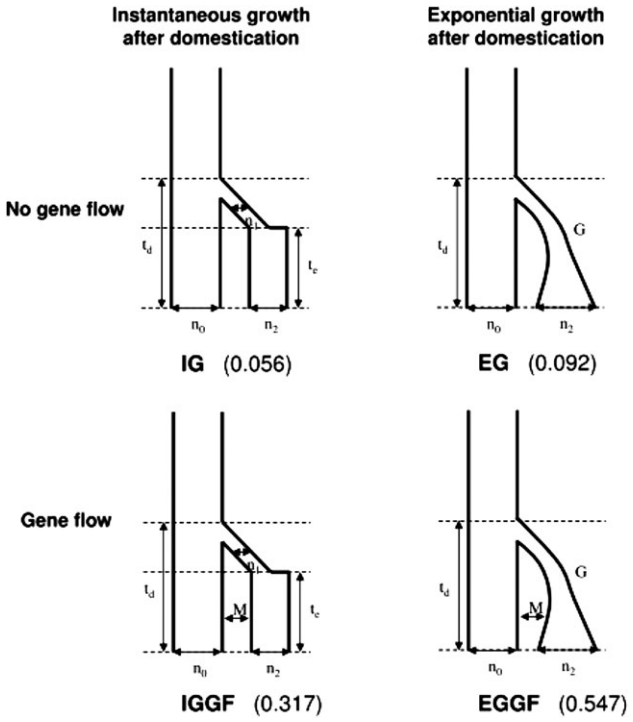

FIG. 1. Alternative demographic models of the evolution of wild and domesticated pearl millet. We considered four models describing the domestication of pearl millet. A model assuming instantaneous growth after domestication, without gene flow between cultivated and wild pearl millet (IG); a model assuming exponential growth after domestication, without gene flow between cultivated and wild pearl millet (EG); a model assuming instantaneous growth after domestication, with gene flow between cultivated and wild pearl millet (IGGF); a model assuming exponential growth after domestication, with gene flow between cultivated and wild pearl millet (EGGF). The posterior probability for each model is given in parentheses. For a more detailed description of these models, see Materials and Methods.

(Pritchard et al. 2000). The analysis was based on haplotypes for the random genes. The admixture model was used with a burn-in period of 30,000 steps and 10^6 MCMC replicates. Ten independent runs were performed for different numbers of assumed populations ($K = 1$–10). The most probable value of *K* was determined by calculating ΔK (Evanno et al. 2005).

Demographic Modeling

Demographic events during domestication, such as bottlenecks or gene flow, may have impacted nucleotide patterns in the genome and, if not taken into account, may lead to false negatives and false positives in neutrality tests. In order to disentangle the effect of demographic events on candidate gene polymorphisms, neutrality statistics observed for candidate genes were compared with expectations predicted from random gene polymorphisms on a simulation basis.

The demographic history of the sample was simulated using ABC methodology (Beaumont et al. 2002). Four models describing the domestication of pearl millet were compared (fig. 1). These models differ in the occurrence of gene

Table 1. Prior Distributions of Parameter Values in the Demographic Models Used During ABC Analysis.

Parameter Name	Parameter Definition	Parameter Prior
θ	$4N_0\mu$	U (3, 10)
ρ	$4N_0r$	U (0, 10)
$n_2>$	N_2/N_0	U (0.01, 5)
n_1	N_1/N_0	U (0.01, 0.4)
t_e	$T_e/(4N_0)$	U (0.0043, 0.0673)
t_d	$T_d/(4N_0)$	U (0.0043, 0.0673)
G	$-\ln(N_1/N_2)/t_d$	U (−50, 150)
M	$4N_0m$	U (0, 50)

NOTE.—All parameters are expressed relative to the constant effective population size of the wild population N_0. The different parameters are listed and explained in the text. For each model parameter, priors followed a uniform distribution.

flow between cultivated and wild pearl millet and in instantaneous/exponential growth after domestication. The instantaneous growth after domestication model (IG) assumes two populations, that is, wild and cultivated populations that diverged from an ancestral population at time T_d. T_d represents the time since the start of domestication. Effective population size N_0 of the wild population remains constant. Until the time since population growth after domestication T_e, the effective population size for cultivated population is N_1. After T_e, the effective population size for the cultivated population is instantaneously increased to N_2. Another model we studied assumes exponential growth after domestication (EG). We also considered two models assuming gene flow after domestication: a model with instantaneous growth after domestication followed by gene flow between cultivated and wild populations (IGGF) and a model with exponential growth after domestication followed by gene flow between cultivated and wild populations (EGGF). The last two models assume bidirectional gene flow between wild and cultivated populations at a rate m.

The IG model assumes six parameters: two associated with time ($t_d = T_d/4N_0$; $t_e = T_e/4N_0$), two associated with the ratio of effective size ($n_2 = N_2/N_0$; $n_1 = N_1/N_0$), and two associated with the gene diversity or recombination rate ($\theta = 4N_0\mu$; $\rho = 4N_0r$). The IGGF model has a supplementary parameter $M = 4N_0m$ associated with postdomestication gene flow. In the EG model, parameters are associated with domestication timing (t_d), ratio of effective size (n_2), exponential growth ($G = -\ln(N_1/N_2)/t_d$), and two associated with gene diversity or recombination rate (θ; ρ). The EGGF model has a supplementary parameter M associated with postdomestication gene flow. Noninformative prior bounds were arbitrarily chosen to avoid being restrictive for posterior evaluation (table 1) except for T_d and T_e for which we chose prior bounds between the introduction of pearl millet from Africa into India estimated to have occurred around 3,000 years ago (Purseglove 1976) and the most ancient estimation for cereal domestication, that is, 12,000 years ago (Glémin and Bataillon 2009). Values are expressed relative to effective population size, assuming a nucleotide mutation rate of 3.30×10^{-8} (Clark et al. 2005) or 7.90×10^{-9} (Gaut and Clegg 1991) substitutions per site per generation. We assumed $T_d > T_e$. For each model, 10^6 ABC simulations were run using msABC (Pavlidis et al. 2010).

Posterior Parameter Estimation

The choice of summary statistics to estimate posterior parameters is a crucial step in ABC methodology (Csilléry et al. 2010). We used two sets of summary statistics. The first set was chosen based on relevant values in the literature and was made up of nine summary statistics: nucleotide diversity π (Nei 1987), nucleotide polymorphism θ_w (Watterson 1975), Tajima's D (Tajima 1989), ZnS (Kelly 1997) for cultivated and wild populations and F_{ST} (Hudson et al. 1992) between cultivated and wild populations. The second set was chosen according to the informativeness of summary statistics for summarizing model parameters. The informativeness of summary statistics was assessed by calculating the correlation between summary statistics and model parameters (Pavlidis et al. 2010; supplementary table S4, Supplementary Material online). The second set was made up of seven summary statistics: nucleotide polymorphism θ_w in the wild population, Tajima's D in the cultivated population, ZnS in the wild and pooled population, F_{ST} between wild and cultivated populations, the percentage of private polymorphisms in wild and cultivated populations, haplotype diversity in the cultivated population.

Posterior distribution of parameters were estimated by the rejection–regression procedure described in Beaumont et al. (2002), adapted by a log–tangent transformation of parameters (Hamilton et al. 2005). This transformation of parameters ensures that the posterior distribution is contained within the bounds of the priors. The rejection–regression step was realized with a threshold of 10^{-3}, using abcreg software (Thornton 2009). Prior and posterior distribution of parameters were drawn using R (R Development Core Team 2009).

Model Validation and Comparison

Fay and Wu's H requires estimation of the ancestral and derived states of polymorphisms. Homoplasy in phylogenetic branches linking pearl millet to the outgroup species leads to misorientations and could bias Fay and Wu's H (De Mita et al. 2007). For random genes, we estimated the back mutation rate P_M, that is, the rate of misinferred sites (Baudry and Depaulis 2003). Sequence simulations were then run using a similar error rate when estimating the ancestral/derived allele state, in order to include the bias in the expected distribution of Fay and Wu's H under a neutral model.

To check each model, 10^4 simulations per gene were run for a given model. Each simulation was run using combinations of parameters resampled from posterior parameter distributions. As parameters might be correlated, we used the following algorithm to sample in the posterior distribution. Each posterior parameter is a vector that could be interpreted as a point in the multiparameter space. For each combination of parameters, we 1) randomly picked one of the posterior parameter points, 2) identified the closest neighbor in the multiparameter space, 3) drew a random point on the continuous line connecting these two points, and 4) recorded the coordinates of this point (vector of parameter). This parameter vector was then used

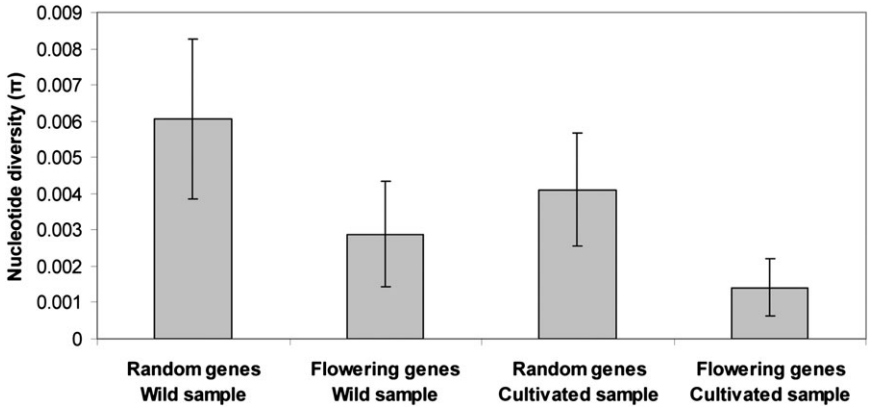

FIG. 2. Comparison of nucleotide diversity in random and flowering genes and in wild and cultivated samples. Per base nucleotide diversity (π) was calculated, based on segregating sites, for random and flowering genes, by distinguishing wild and cultivated samples. In the wild sample, flowering genes showed lower nucleotide diversity than random genes (Mann–Whitney test, $P = 1.03 \times 10^{-2}$). In the cultivated sample, flowering genes showed lower nucleotide diversity than random genes (Mann–Whitney test, $P = 4.02 \times 10^{-3}$). The nucleotide diversity of random genes in the wild sample was greater than that of flowering genes in the cultivated sample (Mann–Whitney test, $P = 1.69 \times 10^{-4}$). Random and flowering genes considered together showed higher nucleotide diversity values in the wild sample than in the cultivated sample (Mann–Whitney test; $P = 2.50 \times 10^{-2}$).

for simulation. Simulated data were then processed to calculate the same summary statistics as the ones used to estimate posterior parameters. A chi-squared goodness of fit test was used to compare calculated and expected distribution of summary statistics. This test assumes that ideal models must lead to a quarter of simulated genes exhibiting a summary statistic in each quartile of the distribution of summary statistic for simulations. *P* values for the chi-squared goodness of fit test were obtained by computing 10^4 MCMC simulations.

To compare the four models, the posterior probability of the models was calculated using the R script "calmod.r" provided by Beaumont (2008). The rejection–regression method was used on 10^5 simulations per model.

Identification of Outliers

Tajima's *D* (Tajima 1989), Fay and Wu's *H* (Fay and Wu 2000), and F_{ST} (Hudson et al. 1992) were calculated for the simulations used for model validation and comparison, for both cultivated and wild pearl millet samples. For each of the candidate genes, the rank of observed Tajima's *D* and Fay and Wu's *H* in their respective expected distribution according to the selected model was calculated. The rank of F_{ST} observed for a given gene was calculated in comparison with the expected distribution of F_{ST} for simulated data sets sharing similar θ_w per gene ± 0.1, as F_{ST} is influenced by the mutation rate (Kronholm et al. 2010). The entire selection of flowering genes was tested using Fisher's combining probability test (Sokal and Rohlf 1995).

Results

Nucleotide Variation at Random and Flowering Genes in Cultivated and Wild Samples

Twenty random genes and 15 flowering genes (Genbank accessions JQ269840–JQ271534) were amplified in 33 cultivated and 13 wild pearl millet individuals (supplementary table S5, Supplementary Material online). All the flowering genes showed high identity with library sequences identified as our target genes using BLAST (supplementary table S6, Supplementary Material online). A total of 9,649 bp for random genes (per gene average: 482 bp) and 8,174 bp for flowering genes (per gene average: 545 bp) were aligned. We detected an average of 12.1 and 7.7 SNPs in random genes and flowering genes, respectively. At least one of the two outgroups was amplified for 16 random genes and 11 flowering genes. The best estimation of group number using STRUCTURE was found for $K = 2$ (supplementary fig. S1A, Supplementary Material online). These groups correspond to the cultivated and wild populations (supplementary fig. S1B, Supplementary Material online).

We compared nucleotide diversity π for random and flowering genes, in wild and cultivated samples (fig. 2; supplementary table S7, Supplementary Material online). In the wild sample, we found a 2.1-fold higher π for random genes (average $\pi = 6.04 \times 10^{-3}$) than for flowering genes (average $\pi = 2.88 \times 10^{-3}$; Mann–Whitney test, $P = 0.01$). In the cultivated sample, we found a 2.9-fold higher π for random genes ($\pi = 4.11 \times 10^{-3}$) than for flowering genes ($\pi = 1.42 \times 10^{-3}$; Mann–Whitney test, $P < 0.005$). Differences in π between wild and cultivated samples were nonsignificant when compared separately in each gene class (Mann–Whitney test, $P > 0.05$). Random and flowering genes considered together had a 1.6-fold higher π in the wild sample ($\pi = 4.69 \times 10^{-3}$) than in the cultivated sample ($\pi = 2.96 \times 10^{-3}$; Mann–Whitney test; $P < 0.03$). Differences were still significant if we considered π estimated using synonymous or silent sites but not if we considered nonsynonymous sites (supplementary table S8, Supplementary Material online). In conclusion, our data showed that flowering genes had less genetic variability than random genes and that genetic variability in the cultivated sample was lower than in the wild sample.

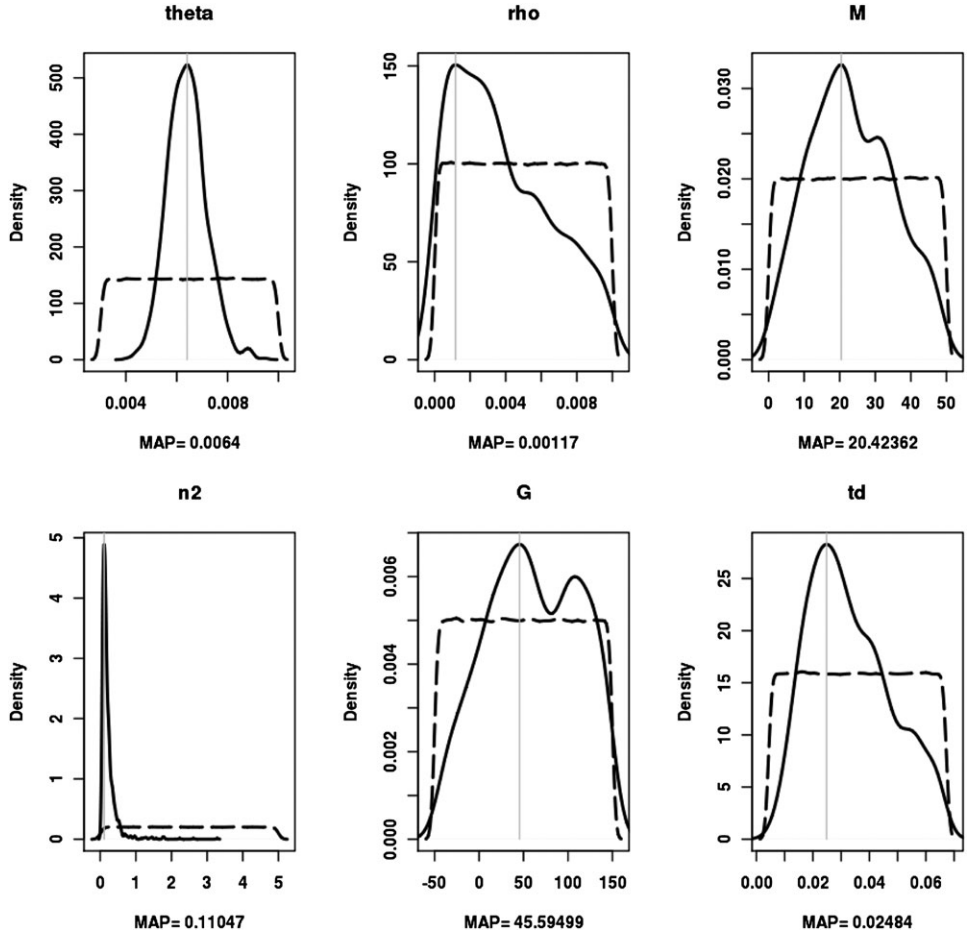

FIG. 3. Prior and posterior distribution of parameters in the EGGF model. For the EGGF model, prior (dashed line) and posterior (solid line) distribution of model parameters are shown. Maximum a posteriori is indicated below each graph and symbolized by the gray line. Parameters describing models: θ: mutation parameter; ρ: recombination parameter; n_2: effective size of the cultivated population now; M: bidirectional gene flow rate between cultivated and wild populations; t_d: time since the beginning of pearl millet domestication.

Tajima's D and Fay and Wu's H observed in random genes were compared with expected values under the Wright–Fischer model (supplementary table S7, Supplementary Material online). We observed no deviations from expected values for Tajima's D in the wild sample (average $D = -0.15$; chi-squared goodness of fit test: $P = 0.502$) or in the cultivated sample (average $D = 0.34$; chi-squared goodness of fit test: $P = 0.843$). Fay and Wu's H was also not different than expected in the wild sample (average $H = -0.38$; chi-squared goodness of fit test: $P = 0.152$). However, random genes in the cultivated sample showed a significant deviation from expected Fay and Wu's H under standard neutral model (average $H = -1.19$; chi-squared goodness of fit test: $P = 0.003$). This deviation may have been caused by demographic events during the evolutionary history of pearl millet. Disentangling the effect of demographic history in the search for signatures of selection in flowering genes requires modeling the history of pearl millet domestication.

Modeling Pearl Millet Demographic History

Twenty random genes were used to build four models describing pearl millet domestication (fig. 1). The rate of undetected back mutations for random genes was $P_M = 0.055$. Therefore, when estimating the ancestral state of polymorphisms, we were mistaken at a rate of 5.5%. This rate was taken into account in the simulations.

Models were first checked by comparing observed and expected distribution of summary statistics. None of the four models showed a distribution of summary statistics that statistically deviated from observed data (chi-squared goodness of fit test; $P > 0.09$). Consequently, all four models could be used to model the demographic history of the sample.

The posterior probability of the models was computed to compare the four models (supplementary fig. S3, Supplementary Material online). For all thresholds tested, models exhibiting gene flow between cultivated and wild pearl millet (IGGF, $P = 0.317$ and EGGF, $P = 0.547$ for $\alpha = 0.01$) showed a higher posterior probability than models assuming no gene flow (IG, $P = 0.056$ and EG, $P = 0.092$). The EGGF model was the most probable model, whatever the threshold used.

Using the EGGF model, we obtained the distribution of posteriors for model parameters (fig. 3; supplementary table S9, Supplementary Material online). The estimated

Table 2. Pearl Millet Population Genetics Parameters.

Parameters	Maximum a posteriori Estimate and 95% CI
T_d	4821 (1665; 15370)
N_0	48514 (38029; 60564)
N_1	1726 (0; 63672)
N_2	5359 (2020; 43268)
r ($\times 10^{10}$)	61 (9; 393)
m ($\times 10^{6}$)	105 (21; 191)

NOTE.—Parameter estimates were inferred from the posterior distributions of the EGGF model (fig. 1). The effective population size of the wild population (N_0), the current cultivated population effective size (N_2), as well as the effective size during the bottleneck (N_1) are reported. Effective population sizes are expressed in numbers of individuals. T_d is the time since the beginning of domestication in years before present. r is the estimated crossover between adjacent base pairs per generation. m is the fraction of each subpopulation made up of new migrants in each generation. The estimation is based on nucleotide mutation rate $\mu = 3.30 \times 10^{-8}$ substitutions per site per generation (Clark et al. 2005).

mutation parameter θ in the wild population was 6.40×10^{-3} (95% confidence interval [CI]: 5.02×10^{-3}–7.99×10^{-3}). Estimates for all the different parameters (recombination parameter ρ, n_2 describing the effective size of the domesticated population, t_d expressing the time since the beginning of domestication, G the growth parameter describing how the cultivated population has grown since domestication) and their CI were reported (supplementary table S9, Supplementary Material online). Transforming the value of these parameters into numbers of individuals or time in years requires a generally unknown quantification of the mutation rate. We used a recently published mutation rate $\mu = 3.30 \times 10^{-8}$ substitutions per site per generation for maize (Clark et al. 2005). The effective population size (table 2) of the wild population $N_0 = \theta/4\mu$ was 48,514 individuals (95% CI: 38,029–60,564). The estimated effective population size of the cultivated population $N_2 = n_2 \times N_0$ was 5,359 individuals (95% CI: 2,020–43,268). The initial domesticated population, prior exponential growth $N_1 = N_2 \times \exp^{-Gtd}$ was 1,726 individuals (95% CI: 0–63,672). The estimated crossover between adjacent base pairs per generation was $r = \rho/4N_0 = 6.1 \times 10^{-9}$ (95% CI: 0.9×10^{-9}–3.93×10^{-8}). The fraction of each subpopulation made up of new migrants in each generation was $m = M/4N_0 = 1.05 \times 10^{-4}$ (95% CI: 2.1×10^{-5} –1.91×10^{-4}). This estimation means that for each generation, 0.01% of wild or cultivated population was composed of individuals coming from the other population. The estimated time since domestication $T_d = t_d \times 4N_0$ was 4,821 years (95% CI: 1,665–15,370). All these estimations should be interpreted with caution as posteriors can vary between models, and the nucleotide mutation rate μ remains unknown for pearl millet (supplementary table S10, Supplementary Material online). These different parameters were used to test flowering genes for deviations from expectations under neutrality.

Detection of Signatures of Selection in the Flowering Network

Values of Tajima's D, Fay and Wu's H, and F_{ST} for candidate genes were then compared with their expected distribution according to the EGGF model (table 3). We found that the repartition of these statistics for observed random genes within the statistics quartiles predicted by the EGGF model (chi-squared goodness of fit test; $P > 0.05$) was correct. For the cultivated population, Tajima's D appeared to shift toward positive values (median: 0.268; 95% CI: −1.82 to 2.74, supplementary fig. S4, Supplementary Material online), whereas D was closer to $N(0,1)$ for the wild population (median: −0.11; 95% CI: −1.73 to 1.82). This result indicates a higher proportion of intermediate frequency polymorphisms in the cultivated population. This pattern coincides

Table 3. Neutrality Tests Applied to Flowering Genes in Cultivated and Wild Pearl Millet.

	Tajima's D				Fay and Wu's H				F_{ST}			
	Cultivated		Wild		Cultivated		Wild					
Gene	D	P	D	P	H	P	H	P	θ_w	n	F_{ST}	P
PgEMF2	1.030	0.704	−1.041	0.160	NA[b]	NA[b]	NA[b]	NA[b]	0.592	4857	0.047	0.520
PgFY	0.132	0.462	−0.200	0.463	−4.146	0.106	−0.529	0.242	3.155	8140	0.372	0.975*
PgGI	−0.148	0.384	−0.119	0.497	−1.851	0.289	0.714	0.881	2.786	10479	0.430	0.988*
PgHD1	NA[a]	NA[a]	−0.685	0.273	NA[a]	NA[a]	0.366	0.598	0.410	2613	0.059	0.599
PgHD3a	0.312	0.512	−0.195	0.465	NA[b]	NA[b]	NA[b]	NA[b]	1.026	9729	0.385	0.977*
PgHD6	−1.902	0.018*	−0.207	0.460	0.096	0.663	−4.261	0.007**	3.998	3717	0.115	0.648
PgLFL1	−0.564	0.273	−1.856	0.016*	0.108	0.668	−0.357	0.286	0.982	9729	0.025	0.324
PgMADS11	−1.315	0.096	NA[a]	NA[a]	NA[b]	NA[b]	NA[a,b]	NA[a,b]	0.394	2613	0.010	0.340
PgPHYA	1.119	0.725	−0.052	0.523	0.170	0.694	−0.320	0.297	0.785	7356	0.134	0.766
PgPHYB	−0.976	0.172	−0.766	0.245	0.522	0.857	0.935	0.983**	1.577	14426	0.024	0.267
PgPHYC	2.845	0.980*	−0.617	0.298	−0.091	0.605	−0.638	0.219	2.179	13534	0.311	0.954*
PgPIPK1	NA[a]	NA[a]	−1.006	0.170	NA[a]	NA[a]	−0.464	0.258	1.194	11841	0.081	0.574
PgPRR73	−1.378	0.086	−0.624	0.296	−1.709	0.307	−1.935	0.060	1.624	14109	0.542	0.997**
PgPRR95	2.646	0.968	0.012	0.548	0.084	0.658	1.268	1.000***	1.571	14426	0.102	0.637
PgTFL1	−1.198	0.119	−0.714	0.262	NA[b]	NA[b]	NA[b]	NA[b]	0.589	4857	0.029	0.420

NOTE.—Rank of statistics for a given gene was calculated relative to the expected distribution obtained by the EGGF model. The number of simulations (n) used to calculate the rank of F_{ST} for a given gene is shown. Retained simulations are those that share similar θ_w per gene ±0.1. NA: not available.

[a] No polymorphism in the sample.

[b] Outgroup sequence not available.

For Tajima's D and Fay and Wu's H, bilateral tests were used: *$P < 0.025$ or $P > 0.975$, **$P < 0.005$ or $P > 0.995$, ***$P < 0.0005$ or $P > 0.9995$. For F_{ST} a unilateral test was used: *$P > 0.95$, **$P > 0.99$. P value was estimated using the rank of the observed value in the expected distribution, divided by the number of simulations.

with a Fay and Wu's *H* distribution deviated toward lower values in the cultivated sample (median: −0.57; 95% CI: −9.21 to 0.89) than in the wild sample (median: 0.21; 95% CI: −2.80 to 0.90), suggesting more high frequency–derived alleles in the cultivated than in the wild population (supplementary fig. S5, Supplementary Material online).

The 15 flowering genes were considered successively to test for deviations from neutral expectations under the EGGF model. Several genes showed a neutrality statistic outside the 95% CI defined based on the expected distribution (table 3). *PgHD6* ($D = -1.902$, $P = 0.018$) and *PgLFL1* ($D = -1.856$, $P = 0.016$) showed a significant negative Tajima's *D*, respectively, for cultivated and wild samples, suggesting an excess of low-frequency polymorphisms. This result may indicate positive selection and/or selective sweep, even if Fay and Wu's *H* was not significant for these genes. *PgHD6* showed a significantly lower *H* for the wild sample, indicating an excess of high frequency–derived polymorphisms ($H = -4.261$, $P = 0.007$). A significant positive Tajima's *D* was found for *PgPHYC* ($D = 2.845$, $P = 0.980$) in the cultivated sample. *PgPRR95* ($H = 1.268$, $P = 1$) and *PgPHYB* ($H = 0.935$, $P = 0.983$) showed a significantly higher Fay and Wu's *H* in the wild sample, indicating an excess of high frequency–derived polymorphisms. A high differentiation between wild and cultivated samples was found for *PgPRR73* ($F_{ST} = 0.542$, $P = 0.997$), *PgGI* ($F_{ST} = 0.372$, $P = 0.988$), *PgHD3a* ($F_{ST} = 0.385$, $P = 0.977$), *PgFY* ($F_{ST} = 0.372$, $P = 0.975$), and *PgPHYC* ($F_{ST} = 0.311$, $P = 0.954$; supplementary fig. S6, Supplementary Material online). Considering both simple indels and SNPs in the analysis gave similar results, except the loss of significance for Tajima's *D* for *PgHD6* in the cultivated sample (supplementary table S11, Supplementary Material online). For the Phytochrome C gene, we further sequenced several fragments along the gene (supplementary fig. S7, Supplementary Material online). Significant neutrality tests were only observed at the 3′ end of the gene, for *PgPHYC10* gene fragment and the initial sequenced fragment *PgPHYC* (supplementary table S12, Supplementary Material online). *PgPHYC10* showed a significant positive *H* for the cultivated polymorphism ($H = 0.940$, $P = 0.982$). *PgPHYC* fragment showed the only significant positive Tajima's *D* ($D = 2.845$, $P = 0.980$) and the only significant F_{ST} between the wild and the cultivated populations ($F_{ST} = 0.311$, $P = 0.954$). Considering both SNPs and indels did not change results (data not shown).

To test for selection in the flowering gene network as a whole, we compared the test-associated probabilities for individual genes with the expected distribution of the neutrality statistic using Fisher's combining probability test (Sokal and Rohlf 1995). Flowering genes did not show more significant Tajima's *D* for cultivated (supplementary table S13, Supplementary Material online; Fisher combining probability test, $X^2 = 34.7$; $P = 0.117$) and wild (supplementary table S13, Supplementary Material online; $X^2 = 18.2$; $P = 0.922$) samples or Fay and Wu's *H* for cultivated samples (supplementary table S13, Supplementary Material online; $X^2 = 11.5$; $P = 0.872$) than expected under the EGGF model. Flowering genes showed higher F_{ST} than expected under the EGGF model (supplementary table S13, Supplementary Material online; Fisher's combining probability test, $X^2 = 56.9$; $P = 0.002$). Fay and Wu's *H* deviated significantly from expectations under the EGGF model for the wild sample (supplementary table S13, Supplementary Material online; Fisher's combining probability test, $X^2 = 43.3$; $P = 0.004$). The same test applied to the other models gave similar results, except Tajima's *D* for the cultivated sample which gave significant results for IG, EG, and EGGF models (supplementary table S13, Supplementary Material online; $X^2 = 40–49$; $P < 0.04$). In conclusion, flowering genes showed a higher differentiation between cultivated and wild samples than expected based on random genes and a tendency toward extreme Tajima's *D* values in the cultivated sample and toward extreme Fay and Wu's *H* values in the wild sample.

Several references in the literature suggest circadian clock–related genes may have undergone more intense selection pressure (Hall et al. 2011) or that their genetic variations are more closely associated with flowering time than the rest of the flowering pathway (Brachi et al. 2010). We tested this hypothesis with our data set. The four genes involved in the circadian clock pathway, that is, *PgGI*, *PgHD6*, *PgPRR73*, and *PgPRR95*, showed higher F_{ST} values between cultivated and wild samples than random genes (Mann–Whitney test; $P = 0.01$) and other flowering genes (Mann–Whitney test; $P = 0.04$). Therefore, the set of circadian clock–related genes showed an excess of differentiation between the cultivated and the wild samples. No differences for Fay and Wu's *H* and Tajima's *D* were observed between circadian clock and other flowering time genes or random genes.

It is commonly considered that demographic estimations by ABC are largely influenced by the choice of summary statistics (Joyce and Marjoram 2008). We used another set of summary statistics and repeated the analysis between the posterior estimation and the search for selection outliers. We compared ranks of *D*, *H*, and F_{ST} observed for flowering genes by comparing them with the distribution expected according to the model used and found a high correlation between ranks of observed values in the expected distribution whatever the model used (for all comparisons, Pearson's correlation test: $r > 0.99$; $P < 0.001$). This indicates that our results are robust concerning the choice of summary statistics.

Discussion

We first built a demographic model describing the domestication of pearl millet by using polymorphisms for 20 random genes. This demographic model was then used to obtain information about the evolutionary history of pearl millet and to test for signatures of selection in the flowering pathway.

Effect of the Domestication on Pearl Millet Diversity

Domestication is expected to lead to a reduction in genetic diversity in cultivated populations compared with wild

populations (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005; Liu and Burke 2006; Haudry et al. 2007). To study the effect of this domestication bottleneck on pearl millet diversity, we compared the nucleotide polymorphism in random genes between wild and cultivated samples. The wild sample showed 1.5-fold higher nucleotide diversity in random genes than the cultivated sample. Cultivated pearl millets thus conserved 68% of the nucleotide diversity found in wild pearl millets in West Africa. This ratio is in agreement with the 67% found for the *Adh1* gene (Gaut and Clegg 1993) but lower than the 74% of gene diversity from wild populations conserved in cultivated pearl millet, assessed using microsatellites (Oumar et al. 2008). The mutation rate for microsatellite loci is higher than for DNA sequences (Thuillet et al. 2002; Vigouroux, Jaqueth, et al. 2002). Together with the larger sample used by Oumar et al. (2008) as in our study, particularly for cultivated pearl millet, this may explain the difference.

When compared with other species, the loss of 32% of nucleotide diversity in cultivated pearl millet compared with wild pearl millet is in the same order of magnitude as the 35% loss of diversity estimated for maize (Wright et al. 2005) or 34% for soybean (Hyten et al. 2006) but is substantially lower than the estimated loss of 49% in foxtail millet (Wang, Chen, et al. 2010), 62% in barley (Kilian et al. 2006), and 70% in wheat (Haudry et al. 2007) compared with their wild relatives. The three last species are self-pollinating, whereas two of the three former species are cross-pollinating. Selfing would increase the loss of diversity in cultivated plants by increasing the impact of selection in the genome and by preventing wild ×cultivated crosses (Glémin and Bataillon 2009).

To investigate alternative models for domestication, we compared four models differing in instantaneous or exponential growth after domestication and by the occurrence or absence of gene flow between wild and cultivated populations (fig. 1). All these models gave a satisfying fit with observed random genes. However, models assuming gene flow between wild and cultivated populations were the most probable (supplementary fig. S3, Supplementary Material online). This result suggests that gene flow between wild and cultivated millets had a significant impact on pearl millet genetic diversity since the beginning of the domestication process. In the whole of West Africa, we found 0.01% of migration among the two types of population in each generation (table 2). Experimental studies showed that the presence of both wild and cultivated pearl millets in a field led to 8% of hybrids in the progeny of the wild plants and 45% in the progeny of the cultivated plants (Renno et al. 1997). Differences in phenology (Renno and Winkel 1996), pollen competition (Sarr et al. 1988; Robert et al. 1991), and reproductive barriers (Amoukou and Marchais 1993) between wild and cultivated pearl millets may explain why gene flow remains limited even in areas where wild and cultivated pearl millets live in sympatry. In fact, a microsatellite analysis of natural pearl millet populations from Niger found 4.2% of wild plants introgressed by cultivated alleles and 1.4% of cultivated plants introgressed by wild alleles (Mariac et al. 2006). The larger distribution area of cultivated pearl millet than of wild relatives, the longer flowering period of wild plants, and the elimination of wild ×cultivated hybrids in cultivated progeny may cause asymmetrical gene flow. The models we used assume a symmetric bidirectional migration rate between wild and cultivated samples. A larger data set will probably be required to test differences in wild-to-cultivated or cultivated-to-wild migration rates because it would make the model more complex.

Among models assuming gene flow, the model with exponential growth was shown to be more likely than the model with instantaneous growth after the bottleneck (supplementary fig. S3, Supplementary Material online). Most previous domestication models assumed instantaneous growth following domestication (Eyre-Walker et al. 1998; Vigouroux, McMullen, et al. 2002; Innan and Kim 2004; Gao and Innan 2008). This model is probably not biologically satisfactory (Glémin and Bataillon 2009). Archaeological studies on the evolution of domestication traits tend to favor the protracted domestication model (Allaby et al. 2008). The protracted model assumes a slow rate of domestication and a long period of predomestication before domestication traits are fully fixed in the cultivated pool (Allaby 2010). Selection for traits involved in the domestication syndrome may have occurred at different times. For example, large seed size would have been selected a long time after nonshattering in pearl millet (Fuller 2007). We hypothesize that the selection rate varied during domestication and that the size of the bottlenecked population varied too. Modeling domestication by exponential growth of the effective population size of cultivated pearl millet instead of instantaneous growth following domestication assumes that selection pressure for domestication traits decreased during the domestication process, as more and more domestication traits would have been fixed. The biological reality of domestication was probably more complex. However, strong confidence in the EGGF model compared with other models makes domestication models that assume exponential growth an interesting alternative to the classical model.

The EGGF model estimates the start of pearl millet domestication at 4,821 years ago (95% CI: 1,665–15,370 BP). This estimation is in accordance with the oldest archaeological remains of cultivated pearl millets found in Mali dated at around 4,500 BP (Manning et al. 2011). The effective population size of wild and bottlenecked populations is considered to be, respectively, 48,514 and 1,726 individuals, which indicates a 28-fold decrease in effective population size during domestication. These estimations suggest that 3.6% of the wild population contributed to the genetic diversity observed in cultivated pearl millet. Similarly, only 2% of wild soybean ancestors are reported to have contributed to the soybean bottlenecked population (Guo et al. 2010). Fewer than 10% of the teosinte population is reported to have contributed to the diversity observed in maize (Wright et al. 2005). In rice, 40% of wild rice contributed to the domestication of rice (Zhu et al. 2007). Because

there is a positive correlation between the duration of the bottleneck $d = T_d - T_e$ and the size of the bottlenecked population N_1 (Tenaillon et al. 2004; Wright et al. 2005), some authors prefer to use the ratio $k = N_1/d$ to estimate the severity of bottlenecks. It was only possible to calculate this ratio with the IG and IGGF models. The time of divergence and bottlenecks calculated by the IGGF model was not precise and mirrored the prior information distribution (supplementary fig. S2, Supplementary Material online; supplementary table S9, Supplementary Material online). We consequently estimated the bottleneck severity using the IG model and found $k = 1.9$. This figure is similar to results found in maize $k = 2.45$ (Wright et al. 2005) and soybean $k = 2$ (Guo et al. 2010) but is higher than in rice $k = 0.2$ and 0.5 for domestication of *japonica* and *indica*, respectively (Zhu et al. 2007), or foxtail millet $k = 0.61$ (Wang, Chen, et al. 2010). The bottleneck severity observed in pearl millet implies that the domestication was not especially severe in this species. However, this estimation should be interpreted with caution because of the low posterior probability of the IG model.

Apart from the domestication bottleneck, population growth of cultivated pearl millet and gene flow between wild and cultivated pearl millets and other demographic or selective processes may have affected the observed nucleotide diversity. A significant deviation of Fay and Wu's *H* for cultivated pearl millet implies an excess of genes with high frequency–derived alleles in the genome of cultivated pearl millet. This pattern could have been created by genetic drift during the bottleneck or may indicate the occurrence of many large selective sweeps in the cultivated population. A similar pattern was observed in rice (Caicedo et al. 2007). A bottleneck model that incorporates selective sweeps explains this pattern in rice more plausibly than a single domestication model (Caicedo et al. 2007). In pearl millet, genes controlling the domestication syndrome have been shown to be located on several linkage groups (Poncet et al. 2000), allowing selective sweeps to influence several genomic regions. Genome-wide genetic studies will probably provide information about the role of selection in the nucleotide patterns observed at the genome scale.

Signatures of Selection for Flowering Genes

Due to the spread of pearl millet to different agroecological areas in West Africa after domestication and climate changes that have impacted this region for millennia, genetic changes affecting flowering time may have been involved in adaptation during and after the domestication of pearl millet.

Genes in the flowering gene network may have been targeted by selection, especially during pearl millet domestication. Nucleotide diversity was significantly reduced in flowering genes compared with random genes in pearl millet (table 2). This result is expected in the presence of either positive selection or background selection (Charlesworth et al. 1995; Kim and Stephan 2000). In *A. thaliana*, flowering genes have undergone purifying selection (Flowers et al. 2009). We showed that *P* values associated with F_{ST}, Fay and Wu's *H* in the wild sample, and for some models, Tajima's *D* in the cultivated sample for flowering genes were not distributed according to random gene expected distributions (supplementary table S12, Supplementary Material online). This shows a shift of some neutrality tests toward the tails of the distribution expected under neutrality for flowering genes. This means that more flowering genes than expected deviated from neutrality in pearl millet suggesting selection on these genes. From resequencing data on the phytochrome C gene, we show that signature of selection could be localized in the gene. Similarly for maize, signature of selection in the *tga1* gene was restricted to a small gene portion (Wang et al. 2005). This result suggests that we might underestimate the number of flowering time genes under selection.

The flowering time gene network comprises several interacting pathways that may have undergone contrasting selection pressures during evolution (Blázquez 2000; Blázquez et al. 2001). For example, in *A. thaliana*, protein evolution is faster in genes acting on the vernalization pathway than integrator, photoperiod and gibberellic acid genes, probably because of different pleiotropic levels (Flowers et al. 2009). We did not study enough genes to systematically compare selection pressures in different interacting pathways of the flowering time network. However, the genes involved in the circadian clock showed a significantly higher F_{ST} in wild and cultivated samples than random genes or other flowering genes (table 3). The excess of genetic differentiation between wild and cultivated pearl millet pools in circadian clock–related genes would be expected under positive selection for these genes in one or the other population (Lewontin and Krakauer 1973). Of the 11 statistically significant neutrality tests observed for Tajima's *D*, Fay and Wu's *H*, and F_{ST}, five concerned genes associated with the circadian clock. In *A. thaliana*, a genome-wide association analysis revealed a prevalence of circadian clock–related genes among genes associated with variations in flowering time in natural conditions (Brachi et al. 2010). In *Populus tremula*, with the exception of one gene, all photoperiod genes targeted by positive selection were involved in the circadian clock pathway (Hall et al. 2011). These references and our results suggest that circadian clock–related genes may represent an important target for adaptation of the flowering pathway. The circadian clock is involved in synchronizing most biological functions with the day/night cycle and the seasonal cycle by integrating light and temperature signals (Más and Yanovsky 2009). It is especially involved in several developmental stages in *A. thaliana*, including seed germination, seedling growth, stress responses, and the flowering induction (de Montaigu et al. 2010). The role of circadian clock–related genes in the synchronization of such important developmental traits for adaptation capability probably explains the specific selection pattern in this part of the flowering network.

Among the signatures of selection detected, the positive Tajima's *D* in the cultivated sample ($D = 2.84$; $P = 0.980$, table 3) confirmed the signal of balancing selection in

cultivated pearl millet of *PgPHYC* found in a set of inbred lines of worldwide origin (Saïdou et al. 2009). Polymorphisms at this gene were associated with flowering time (Saïdou et al. 2009). As our sampling covered a wide area of West Africa, we can hypothesize that *PgPHYC* has been involved in adaptation to flowering time in contrasting climates. This hypothesis is reinforced by the absence of selection in the wild sample ($D = -0.62$; $P = 0.298$, table 3), which was confined to the Sahelian region, and by the large significant differentiation between the wild and cultivated samples ($F_{ST} = 0.311$; $P = 0.954$, table 3). Interestingly, the signature of selection is observed in the 3′ end of the gene only, suggesting targets of selection are on this part of the gene or in the downstream region. *PHYC* is one of the genes influencing flowering time in *A. thaliana*, and its alleles are correlated with latitudinal and longitudinal clines in *A. thaliana* (Balasubramanian et al. 2006; Samis et al. 2008). In the present study, no signature of selection was found for *PgMADS11*, but we previously found evidence of selection for this gene in a very large cultivated sample (Mariac et al. 2011). This result is certainly the consequence of the very small sequence fragment amplified for this gene (around 200 bp) and a lack of diversity (supplementary table S7, Supplementary Material online).

A negative Tajima's *D* was found in the cultivated sample for *PgHD6* ($D = -1.902$; $P = 0.018$, table 3), even though the test was nonsignificant when indels were taken into account (supplementary table S11, Supplementary Material online). *PgHD6* also showed 25.6-fold lower nucleotide diversity in the cultivated sample, whereas the reduction was on average 1.6 for random and other flowering genes (supplementary table S7, Supplementary Material online). This reduction in diversity and the negative Tajima's *D* are consistent with a selective sweep at *PgHD6* (Braverman et al. 1995). In rice, a homolog gene *Hd6* is a QTL that controls flowering time (Yamamoto et al. 2000). Further studies will check if the observed pattern was obtained by positive selection at *PgHD6* or a close gene during domestication or by variations in the mutation rate in the *PgHD6* region.

Besides *PgPHYC*, four flowering genes showed a significant F_{ST} in both cultivated and wild samples: *PgPRR73*, *PgGI*, *PgHD3a*, and *PgFY* (table 3). Flowering genes are not only associated with variations in flowering time but also with different development traits (Tienderen et al. 1996). In pearl millet, Saïdou et al. (2009) found an association between polymorphisms at *PgHD3a* and the diameter and length of the primary spike. Polymorphisms at *PgGI* were associated with the same two traits and with the number of basal tillers at head emergence (Saïdou et al. 2009). These two genes were not statistically associated with flowering time per se (Saïdou et al. 2009). Cultivated pearl millet has fewer basal tillers and a higher spike diameter and length than wild pearl millet (Poncet et al. 2000). We can thus hypothesize that *PgHD3a* and *PgGI* were targeted by selection for these traits during pearl millet domestication. Comparison of the genomic localization of these two genes and that of QTLs involved in the number of basal tillers or spike diameter and length in a wild ×cultivated cross progeny will allow this hypothesis to be tested. Our study identified very interesting candidates for genes associated with pearl millet domestication and subsequent improvement. Further studies are needed to decipher causal polymorphisms associated with the selection observed in these different genes.

Supplementary Material

Supplementary tables S1–S13 and figures S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the French Agence Nationale de la Recherche (ANR-07-JCJC-0116-01 to Y.V.).

References

Allaby R. 2010. Integrating the processes in the evolutionary system of domestication. *J Exp Bot.* 61:935–944.

Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proc Natl Acad Sci U S A.* 105:13982–13986.

Amoukou AI, Marchais L. 1993. Evidence of a partial reproductive barrier between wild and cultivated pearl millets (*Pennisetum glaucum*). *Euphytica* 67:19–26.

Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D. 2006. The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat Genet.* 38:711–715.

Baudry E, Depaulis F. 2003. Effect of misoriented sites on neutrality tests with outgroup. *Genetics* 165:1619–1622.

Beaumont MA. 2008. Joint determination of topology, divergence time and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, editors. Simulations, genetics and human prehistory. Cambridge: McDonald Institute for Archaeological Research. p. 135–154.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

Blackman BK, Rasmussen DA, Strasburg JL, Raduski AR, Burke JM, Knapp SJ, Michaels SD, Rieseberg LH. 2011. Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187:271–287.

Blázquez M, Koornneef M, Putterill J. 2001. Flowering on time: genes that regulate the floral transition. *EMBO Rep.* 2:1078–1082.

Blázquez MA. 2000. Flower development pathways. *J Cell Sci.* 113:3547–3548.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 6:e1000940.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* 140:783–796.

Caicedo AL, Williamson SH, Hernandez RD, et al. (12 co-authors). 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* 3:e163.

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

Casa A, Mitchell S, Hamblin M, Sun H, Bowers J, Paterson A, Aquadro C, Kresovich S. 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor Appl Genet.* 111:23–30.

Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, Knapp SJ, Burke JM. 2008. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 20:2931–2945.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141:1619–1632.

Clark RM, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol Biol Evol.* 22:2304–2312.

Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *J Exp Bot.* 58:1231–1244.

Cork JM, Purugganan MD. 2004. The evolution of molecular genetic pathways and networks. *BioEssays* 26:479–484.

Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol.* 25:410–418.

De Mita S, Ronfort J, McKhann HI, Poncet C, El Malki R, Bataillon T. 2007. Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in Nod factor signaling in *Medicago truncatula*. *Genetics* 177:2123–2133.

de Montaigu A, Tóth R, Coupland G. 2010. Plant development goes like clockwork. *Trends Genet.* 26:296–306.

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.

Drummond AJ, Ashton B, Buxton S, et al. (15 co-authors). 2010. Geneious v4.8. Available from: http://www.geneious.com

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

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14:2611–2620.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci U S A.* 95:4441–4446.

Fay JC, Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH. 2006. A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. *Plant Physiol.* 140:1183–1191.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the *Arabidopsis thaliana* flowering time gene network. *Mol Biol Evol.* 26:2475–2486.

Fuller DQ. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the old world. *Ann Bot.* 100:903–924.

Fuller DQ, Qin L, Zheng Y, Zhao Z, Chen X, Hosoya LA, Sun G-P. 2009. The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* 323:1607–1610.

Gao L-z, Innan H. 2008. Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* 179:965–976.

Gaut BS, Clegg MT. 1991. Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. *Proc Natl Acad Sci U S A.* 88:2060–2064.

Gaut BS, Clegg MT. 1993. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). *Genetics* 135:1091–1097.

Glémin S, Bataillon T. 2009. A comparative view of the evolution of grasses under domestication. *New Phytol.* 183:273–290.

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y. 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann Bot.* 106:505–514.

Hall D, Ma X-F, Ingvarsson PK. 2011. Adaptive evolution of the *Populus tremula* photoperiod pathway. *Mol Ecol.* 20:1463–1474.

Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc Natl Acad Sci U S A.* 102:7476.

Haudry A, Cenci A, Ravel C, et al. (11 co-authors). 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol.* 24:1506–1517.

Haussmann BIG, Boubacar A, Boureima SS, Vigouroux Y. 2006. Multiplication and preliminary characterization of West and Central African pearl millet landraces. *Int Sorghum Millet Newsl.* 47:110–112.

Hayama R, Coupland G. 2003. Shedding light on the circadian clock and the photoperiodic control of flowering. *Curr Opin Plant Biol.* 6:13–19.

Higgins JA, Bailey PC, Laurie DA. 2010. Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. *PLoS One.* 5:e10065.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci U S A.* 103:16666.

Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci U S A.* 101:10667–10672.

Izawa T. 2007. Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice. *J Exp Bot.* 58:3091–3097.

Jin J, Huang W, Gao J-P, Yang J, Shi M, Zhu M-Z, Luo D, Lin H-X. 2008. Genetic control of rice plant architecture under domestication. *Nat Genet.* 40:1365–1369.

Joyce P, Marjoram P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat Appl Genet Mol Biol.* 7:26.

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* 146:1197–1206.

Kilian B, Özkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Mol Genet Genomics.* 276:230–241.

Kim Y, Stephan W. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* 155:1415–1427.

Komatsuda T, Pourkheirandish M, He C, et al. (15 co-authors). 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci U S A.* 104:1424–1429.

Komeda Y. 2004. Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annu Rev Plant Biol.* 55:521–535.

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312:1392–1396.

Kronholm I, Loudet O, de Meaux J. 2010. Influence of mutation rate on estimators of genetic differentiation—lessons from *Arabidopsis thaliana*. *BMC Genet.* 11:33.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195.

Li C, Zhou A, Sang T. 2006. Rice domestication by reducing shattering. *Science* 311:1936–1939.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173:321–330.

Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist.* 18:50–60.

Manning K, Pelling R, Higham T, Schwenniger J-L, Fuller DQ. 2011. 4500-year old domesticated pearl millet (*Pennisetum glaucum*) from the Tilemsi Valley, Mali: new insights into an alternative cereal domestication pathway. *J Archaeol Sci.* 38:312–322.

Mariac C, Jehin L, Saïdou A-A, et al. (11 co-authors). 2011. Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Mol Ecol.* 20:80–91.

Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezancon G. 2006. Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. *Theor Appl Genet.* 114:49–58.

Más P, Yanovsky MJ. 2009. Time for circadian rhythms: plants get synchronized. *Curr Opin Plant Biol.* 12:574–579.

Michaels SD, Amasino RM. 2000. Memories of winter: vernalization and the competence to flower. *Plant Cell Environ.* 23:1145–1153.

Mockler T, Yang H, Yu X, Parikh D, Cheng Y-c, Dolan S, Lin S. 2003. Regulation of photoperiodic flowering by Arabidopsis photoreceptors. *Proc Natl Acad Sci U S A.* 100:2140–2145.

Mouradov A, Cremer F, Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14:S111–S130.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Oumar I, Mariac C, Pham J-L, Vigouroux Y. 2008. Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br) as revealed by microsatellite loci. *Theor Appl Genet.* 117:489–497.

Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L, Attene G. 2007. Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Ann Bot.* 100:1039–1051.

Pavlidis P, Laurent S, Stephan W. 2010. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol Ecol Resour.* 10:723–727.

Poncet V, Lamy F, Devos KM, Gale MD, Sarr A, Robert T. 2000. Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L., *Poaceae*). *Theor Appl Genet.* 100:147–159.

Poncet V, Lamy F, Enjalbert J, Joly H, Sarr A, Robert T. 1998. Genetic analysis of the domestication syndrome in pearl millet (*Pennisetum glaucum* L., *Poaceae*): inheritance of the major characters. *Heredity* 81:648–658.

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

Purseglove JW. 1976. Millets. In: Simmonds NW, editor. Evolution of crop plants. London: Longmans.

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.

Purugganan MD, Fuller DQ. 2011. Archeological data reveal slow rates of evolution during plant domestication. *Evolution* 65:171–183.

R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Renno JF, Winkel T. 1996. Phenology and reproductive effort of cultivated and wild forms of *Pennisetum glaucum* under experimental conditions in the Sahel: implications for the maintenance of polymorphism in the species. *Can J Bot.* 74:959–964.

Renno JF, Winkel T, Bonnefous F, Bezançon G. 1997. Experimental study of gene flow between wild and cultivated *Pennisetum glaucum*. *Can J Bot.* 75:925–931.

Robert T, Lespinasse R, Pernes J, Sarr A. 1991. Gametophytic competition as influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoides*). *Genome* 34:195–200.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

Saïdou AA, Mariac C, Luong V, Pham JL, Bezancon G, Vigouroux Y. 2009. Association studies identify natural variation at *PHYC* linked to flowering time and morphological variation in pearl millet. *Genetics* 182:899–910.

Samis KE, Heath KD, Stinchcombe JR. 2008. Discordant longitudinal clines in flowering time and *PHYTOCHROME C* in *Arabidopsis thaliana*. *Evolution* 62:2971–2983.

Sarr A, Sandmeier M, Pernes J. 1988. Gametophytic competition in pearl millet, *Pennisetum typhoides* (Stapf et Hubb.). *Genome* 30:924–929.

Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, Gill BS, Faris JD. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547–555.

Simpson GG. 2004. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr Opin Plant Biol.* 7:570–574.

Smith BD. 2001. Documenting plant domestication: the consilience of biological and archaeological approaches. *Proc Natl Acad Sci U S A.* 98:1324–1326.

Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. New York: W.H. Freeman & Co.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 68:978–989.

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

Tanno K-i, Willcox G. 2006. How fast was wild wheat domesticated? *Science* 311:1886.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol Biol Evol.* 21:1214–1225.

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.

Thornton K. 2009. Automating approximate Bayesian computation by local linear regression. *BMC Genet.* 10:35.

Thuillet A-C, Bru D, David J, Roumet P, Santoni S, Sourdille P, Bataillon T. 2002. Direct estimation of mutation rate for 10 microsatellite loci in Durum Wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Mol Biol Evol.* 19:122–125.

Tienderen PHV, Hammad I, Zwaal FC. 1996. Pleiotropic effects of flowering time genes in the annual crucifer *Arabidopsis thaliana* (*Brassicaceae*). *Am J Bot.* 83:169–174.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol.* 19:1251–1260.

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc Natl Acad Sci U S A.* 99:9650–9655.

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J. 2005. An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169:1617–1630.

Wang C, Chen J, Zhi H, Yang L, Li W, Wang Y, Li H, Zhao B, Chen M, Diao X. 2010. Population genetics of foxtail millet and its wild ancestor. *BMC Genet.* 11:90.

Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF. 2005. The origin of the naked grains of maize. *Nature* 436:714–719.

Wang Y, Shen D, Bo S, Chen H, Zheng J, Zhu Q-H, Cai D, Helliwell C, Fan L. 2010. Sequence variation and selection of small RNAs in domesticated rice. *BMC Evol Biol.* 10:119.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci U S A.* 99:12959–12962.

Wilson RN, Heckman JW, Somerville CR. 1992. Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. *Plant Physiol.* 100:403–408.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* 308:1310–1314.

Xue W, Xing Y, Weng X, et al. (11 co-authors). 2008. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat Genet.* 40:761–767.

Yamamoto T, Lin H, Sasaki T, Yano M. 2000. Identification of heading date quantitative trait locus *Hd6* and characterization of its epistatic interactions with *Hd2* in rice using advanced backcross progeny. *Genetics* 154:885–891.

Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17:2859–2872.

Yu G, Olsen KM, Schaal BA. 2011. Molecular evolution of the endosperm starch synthesis pathway genes in rice (*Oryza sativa* L.) and its wild ancestor, *O. rufipogon* L. *Mol Biol Evol.* 28:659–671.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol.* 24:875–888.