the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **93,** 843–855 doi: 10.1111/tpj.13806

Largely unlinked gene sets targeted by selection for domestication syndrome phenotypes in maize and sorghum

Xianjun Lai[1,2], Lang Yan[1,3,4], Yanli Lu[2] and James C. Schnable[1,*]

[1]*Center for Plant Science Innovation and Department of Agronomy and Horticulture, University of Nebraska-Lincoln, NE 68588, USA,*

[2]*Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, China,*

[3]*Laboratory of Functional Genome and Application of Potato, Xichang College, Liangshan 615000, China, and*

[4]*College of Life Sciences, Sichuan University, Chengdu 610065, China*

Received 4 September 2017; revised 27 November 2017; accepted 4 December 2017; published online 19 December 2017.

*For correspondence (e-mail schnable@unl.edu).

Xianjun Lai and Lang Yan are the authors contributed equally to this work.

SUMMARY

The domestication of diverse grain crops from wild grasses was a result of artificial selection for a suite of overlapping traits producing changes referred to in aggregate as 'domestication syndrome'. Parallel phenotypic change can be accomplished by either selection on orthologous genes or selection on non-orthologous genes with parallel phenotypic effects. To determine how often artificial selection for domestication traits in the grasses targeted orthologous genes, we employed resequencing data from wild and domesticated accessions of *Zea* (maize) and *Sorghum* (sorghum). Many 'classic' domestication genes identified through quantitative trait locus mapping in populations resulting from wild/domesticated crosses indeed show signatures of parallel selection in both maize and sorghum. However, the overall number of genes showing signatures of parallel selection in both species is not significantly different from that expected by chance. This suggests that while a small number of genes will extremely large phenotypic effects have been targeted repeatedly by artificial selection during domestication, the optimization part of domestication targeted small and largely non-overlapping subsets of all possible genes which could produce equivalent phenotypic alterations.

Keywords: grasses, domestication, selection, parallel evolution, *Zea mays*, *Sorghum bicolor*.

INTRODUCTION

The characteristics of modern crops are the result of thousands of years of artificial selection applied consciously or unconsciously by farmers and plant breeders. An estimated 2500 plant species have experienced some degree of artificial selection, with approximately 10% of these being domesticated to the point that the species depends on humans for survival (Dirzo and Raven, 2003). Of the hundreds of crops domesticated by human civilizations, three species – rice, wheat and maize – provide more than half of all calories consumed around the world. These three crops all belong to the same family Poaceae (the grasses), a clade that has contributed a total of at least 48 domesticated crop species to human civilization, including at least 30 species domesticated as sources of grain (Glémin and Bataillon, 2009). Artificial selection for grain production produced a suite of shared phenotypic changes in grain crops referred to as 'the domestication syndrome' (Harlan *et al.*, 1973). Grain crop domestication syndrome includes loss of seed shattering, increased apical dominance, more uniform maturity across inflorescences and across tillers, increase in size and/or number of inflorescences, larger seeds, greater carbohydrate content and lower protein content per seed, and reduction or loss of seed dormancy (Harlan *et al.*, 1973). The genes involved in producing domestication syndrome phenotypes can be identified through two broad types of studies which are sometimes referred to as 'top-down' and 'bottom-up' approaches (Ross-Ibarra *et al.*, 2007).

Top-down approaches utilize quantitative genetic studies to identify large-effect genes involved in producing the changes associated with the domestication syndrome in crop species that are interfertile with their wild progenitors. In maize, an estimated five loci have large enough effects on domestication traits to be mapped using conventional quantitative trait locus (QTL) analysis (Doebley and Stec,

1993). Of these loci, several have been mapped including *teosinte branched1* (*tb1*), where the allele selected for significantly reduces the development of tillers (Clark *et al.*, 2004), *teosinte glume architecture1* (*tga1*), where the allele selected for abolishes the stony fruitcase surrounding teosinte seeds (Dorweiler *et al.*, 1993), and *grassy tillers1* (*gt1*), where the allele selected for results in many fewer ears per plant during domestication (Whipple *et al.*, 2011; Wills *et al.*, 2013). In rice, many functionally characterized genes that underlie phenotypic changes during the domestication process have also been identified through QTL mapping, such as *Seed dormancy 4* (*Sdr4*), where the allele selected for produces a reduction in seed dormancy (Sugimoto *et al.*, 2010), *Tiller Angle Control1* (*TAC1*), where the allele selected for decreases tiller angle, producing more photosynthetically efficient canopy architecture (Yu *et al.*, 2007), and *betaine aldehyde dehydrogenase2* (*BADH2*), a loss-of-function allele selected for during the domestication process that results in the accumulation of 2-acetyl-1-pyrroline in fragrant rice (Bradbury *et al.*, 2005).

In contrast, bottom-up approaches use changes in the diversity and frequency of haplotypes at particular regions of the genome between populations of a crop species and populations of wild relatives to identify loci which were targets of artificial selection. Notably, while quantitative genetic evaluation of recombinant populations generally identifies relatively small numbers of large-effect loci responsible for many of the differences observed between domesticated grain crops and their wild relatives, genome-wide population genetic approaches generally identify hundreds to thousands of loci as targets of selection during domestication in the same species. A similar orders of magnitude difference between candidate gene (top-down) and selection scans (bottom-up) has been noted in other systems, including studies of positive selection in humans (Akey, 2009). Hufford and co-workers used resequencing data from a set of 75 teosinte and maize lines to identify 484 regions of the maize genome likely to have experienced selection during transition from wild teosintes to maize landraces and another 695 regions likely to have experienced selection during transition from largely tropical landraces to largely temperate elite lines (Hufford *et al.*, 2012). Huang and co-workers used genome resequencing data of 1083 cultivated rice and 446 wild rice accessions to identify 55 selection regions encompassed 2547 candidate artificially selected genes during domestication from wild to cultivated rice (Huang *et al.*, 2012). In sorghum, a set of 725 genes which were likely the targets of artificial selection during domestication and/or crop improvement were identified from resequencing of 44 lines of sorghum and wild relatives (Mace *et al.*, 2013). However, one critical limitation of bottom-up approaches is that candidate genes identified using these techniques will not initially be linked to any specific phenotypic trait (Ross-Ibarra *et al.*, 2007).

Parallel phenotypic changes which are part of the domestication syndrome in grain crops could result from parallel or lineage-specific changes at a molecular level. A recently published study demonstrated the loss of seed shattering resulted from disruption of the same gene in maize, sorghum and rice (Lin *et al.*, 2012). *Heading Date1* is a major QTL controlling flowering time which shows evidence of being under parallel artificial selection during the process of domestication for sorghum, setaria and rice (Liu *et al.*, 2015). A flowering time QTL identified in a population of wild × domesticated *Setaria* lines co-localizes with a flowering time QTL identified at syntenic orthologous locations in the genomes of maize and sorghum (Mauro-Herrera *et al.*, 2013). A significant number of candidate genes associated with seed size exhibited signals of parallel selection during domestication in maize, rice and sorghum (Tao *et al.*, 2017). However, not all parallel phenotypic changes produced by artificial selection result from parallel evolution at the molecular level. Artificial selection for adaption to high altitudes in different maize populations targeted largely unrelated sets of genes in Mexican and Andean highland populations (Takuno *et al.*, 2015).

Here we focus on two grain crops, maize (*Zea mays* ssp. *mays*) and sorghum (*Sorghum bicolor* ssp. *bicolor*). Maize was domesticated from Balsas teosinte (*Zea mays* ssp *parviglumis*) in Mesoamerica, with a center of origin in the lowlands of southwestern Mexico (Van Heerwaarden *et al.*, 2011). Sorghum is believed to have first been domesticated from broomcorn (*Sorghum bicolor* ssp. *verticilliflorum*) in Ethiopia (Wendorf *et al.*, 1992), with a potential second independent domestication in west Africa (Sagnard *et al.*, 2011; Mace *et al.*, 2013). The wild ancestors of these two crops diverged approximately 12 million years ago (Swigoˇová *et al.*, 2004). Subsequent to the divergence of these two lineages, maize experienced a whole-genome duplication creating two functionally distinct subgenomes – maize1 and maize2 – each of which is, in principle orthologous to the entire genome of sorghum (Schnable *et al.*, 2011). In some cases orthologs of a single sorghum gene from both subgenomes are still present, creating a pair of maize genes which are co-orthologous to a single sorghum gene. In other cases the maize1 or maize2 gene copy was lost from the genome after the whole-genome duplication, restoring a 1:1 orthologous relationship between the two species. Maize2 gene copies have been lost from the genome more frequently than maize1 gene copies, tend to be expressed to lower mRNA levels and, on average, tend to explain less phenotypic variation than maize1 gene copies (Schnable and Freeling, 2011; Renny-Byfield *et al.*, 2017). The parallel set of phenotypic changes during domestication in these two species (Harlan *et al.*, 1973), the high degrees of conserved collinearity across grass genomes (Bennetzin and Freeling, 1993; Moore *et al.*, 1995) and the

bias towards genes with detectable phenotypic effects being conserved at syntenic locations across grass genomes (Schnable, 2015) offer an opportunity to test the hypothesis that the parallel phenotypic changes in maize and sorghum resulted from artificial selection acting on orthologous genes in both species.

We found that genes conserved at syntenic orthologous locations in maize and sorghum were significantly more likely to be targets of selection during domestication than non-syntenic genes unique to one species. In maize, domestication preferentially targeted genes on the dominant maize1 subgenome rather than their retained duplicates on the maize2 subgenome. Consistent with a much earlier study of maize and sorghum QTLs controlling domestication phenotypes (Paterson *et al.*, 1995), genes identified through quantitative genetic studies of domestication traits in one species were likely to show signatures of selection in the other species. However, the overall overlap between genes identified using population genetic methods in both species was only marginally greater than expected by chance.

RESULTS

Population genetic datasets for both species

Maize and sorghum accessions were sampled from published datasets (Chia *et al.*, 2012; Mace *et al.*, 2013; Luo *et al.*, 2016) (see Experimental Procedures and Table S1 in the Supporting Information). After quality filtering to remove low-quality single nucleotide polymorphisms (SNPs) and those potentially representing alignments of paralogous sequence elsewhere in the genome (see Experimental Procedures), a total of 10.3 million segregating SNPs in 56 maize accessions and 3.3 million segregating SNPs in 42 sorghum accessions remained. These proportions roughly correspond to the difference in genome size between the two species (approximately 2.0 Gb for maize and 700 Mb for sorghum); however, as much of the maize genome is repetitive and cannot be uniquely identified using short sequence reads, this is consistent with higher overall levels of nucleotide diversity in maize relative to sorghum. Also consistent with previous reports, wild relatives had higher levels of nucleotide diversity ($\pi = 0.00377$ in maize and $\pi = 0.00381$ in sorghum), than both landraces ($\pi = 0.00338$ in maize and $\pi = 0.00242$ in sorghum) and improved inbreds ($\pi = 0.00334$ in maize and $\pi = 0.00226$ in sorghum) (Table S3) (Hufford *et al.*, 2012; Mace *et al.*, 2013).

Maize accessions were primarily collected in the Western Hemisphere and sorghum accessions primarily from the Eastern Hemisphere, with some exceptions in both cases (Figure 1a). Wild relatives were primarily collected near the centers of domestication: southwestern Mexico for maize and central east Africa for sorghum. The sorghum dataset also included data for a forage sudangrass line (Greenleaf: sweet sorghum × Sudan grass) from North America. Among the maize lines, modern elite lines and wild relatives each formed distinct clades (Figure 1b). In sorghum, wild relatives formed a distinct clade; however, lines reported to be elite or landrace lines were intercalated, potentially as a result of distinct sorghum breeding efforts developing lines for different agroclimactic zones around the world (Figure 1c).

Genetic maps for both species were sourced from public datasets. For maize, a genetic map was employed that included 10 085 markers genotyped in a set of 232 recombinant inbred lines (RILs) from the maize IBM population using tGBS (Zou *et al.*, 2012; Ott *et al.*, 2017) while for sorghum a genetic map was employed which was constructed from a set of 3418 markers genotyped in a set of 244 RILs from a grain sorghum × sweet sorghum cross using resequencing (Zou *et al.*, 2012).

Genomic signals of selection in maize and sorghum

In each species, the identification of regions under selection was performed for three separate pairwise comparisons: landraces versus wild relatives (domestication), improved lines versus landraces (improvement), and improved lines versus wild relatives. The genome was scanned with XP-CLR using a window size of 0.05 cM and a step size of 1 kb (see Experimental Procedures) and each gene was assigned the XP-CLR score of the highest scoring bin that overlapped with the gene. Genes above the 90th percentile of XP-CLR scores for a given pairwise comparison were considered as candidate 'under selection' genes. The set of gene annotations employed in this analysis included 63 480 maize gene models and 34 027 sorghum gene models. Thus, for each of the three possible pairwise comparisons, 6348 genes in maize and 3403 genes in sorghum were identified as candidates for selection (Figure S1). In both species, estimated selection coefficients were higher during domestication (mean $s = 0.06$ in maize and 0.047 in sorghum) than improvement (mean $s = 0.045$ in maize and 0.024 in sorghum).

Since 10% of genes were identified as candidates for selection during domestication and 10% of genes were identified as candidates for selection during improvement, the overlap expected if these datasets are unrelated is 1%. In fact, approximately 1% of all annotated maize genes (620 genes) were in the top 10% in both comparisons, and approximately 1% of all annotated sorghum genes (345 genes) were in the top 10% in both comparisons (Figure 2a-b). As expected, the set of candidate genes identified in the comparison of wild relatives and improved lines showed significant overlap with both the domestication and crop improvement candidate gene sets (Figure 2a). Among a set of 112 classical maize mutants cloned using forward genetics (Schnable and Freeling, 2011), 23 were

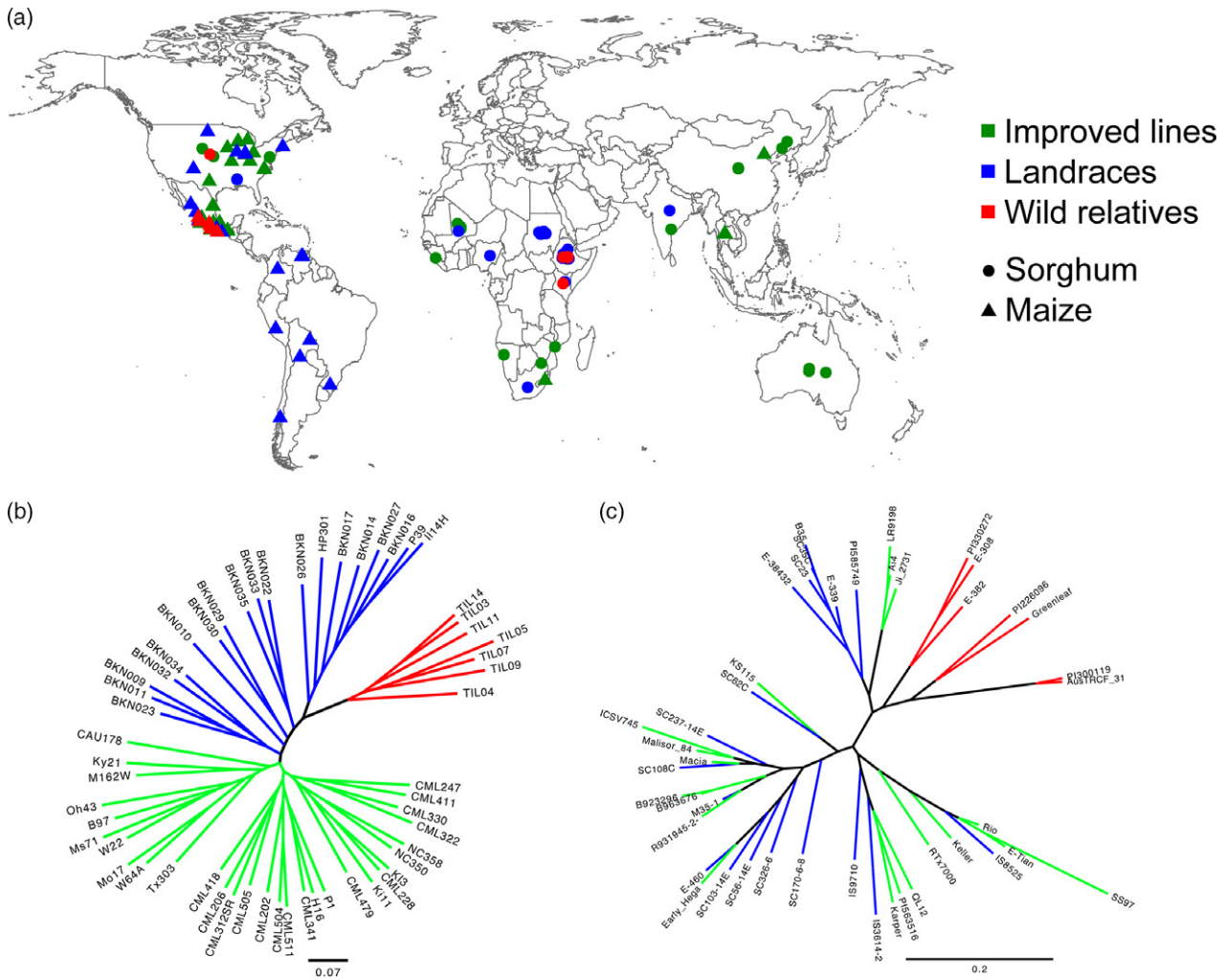

Figure 1. Geographic distribution and phylogenetic relationship of maize and sorghum accessions employed in this study.
(a) The geographic distribution of maize and sorghum accessions. Dots and triangles represent sorghum and maize, respectively. Colors represent membership in different populations: improved lines (green), landraces (blue) and wild relatives (red).
(b), (c) Neighbor-joining (NJ) trees for maize accessions (b) and sorghum accessions (c). Taxa in the NJ tree are color coded using the same system as in (a).

identified as candidate genes under selection during domestication and/or improvement (*P*-value = 0.0007, Fisher exact test).

The maize and sorghum genetic maps are largely collinear (Moore *et al.*, 1995); however, maize has experienced a whole-genome duplication relative to sorghum, producing two functionally distinct subgenomes with differing levels of gene expression, purifying selection and loss rates for both genes and conserved regulatory sequences (Schnable and Freeling, 2011; Pophaly and Tellier, 2015; Lai *et al.*, 2017). In a dataset of 14 433 genes conserved at syntenic locations in sorghum and maize, 7041 genes are conserved between sorghum and maize1 but lost from maize2 and 3031 genes are conserved between sorghum and maize2 but lost from maize1; the balance (4361 genes) are present in all three genomes (Table S2) (Schnable *et al.*, 2016). The balance of the gene complement of each species consists of non-syntenic genes, which are the majority (57.6% and 70.4% for sorghum and maize, respectively) of all annotated genes in both species. Genes with known mutant phenotypes tend to be syntenic (Schnable and Freeling, 2011), while non-syntenic genes tend to exhibit greater allelic variation in gene regulation, and it has been speculated that they may contribute to phenotypic plasticity and heterosis (Paschold *et al.*, 2014; Baldauf *et al.*, 2016; Li *et al.*, 2016).

In sorghum, genes conserved at syntenic orthologous locations between maize and sorghum were significantly more likely to be identified as targets of selection in all three pairwise comparisons (binomial testing *P*-value $< 1.0 \times 10^{-6}$ for the domestication and improvement, *P*-value = 0.0058 for the wild–improved comparison). In maize, syntenic genes were significantly more likely to be identified as targets of selection during domestication and

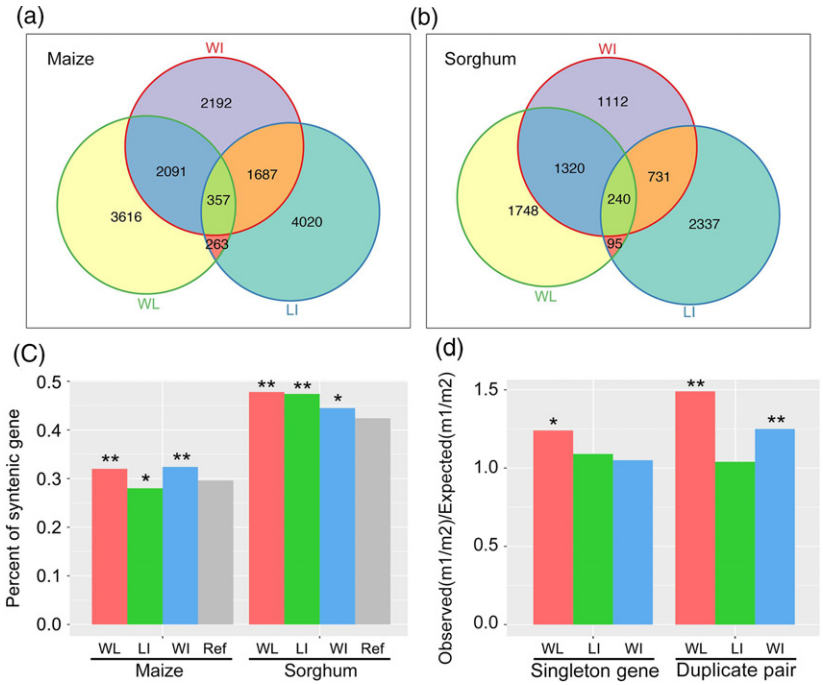

Figure 2. Summary information for candidate genes under selection.
(a), (b) The number of candidate genes shared among the three pairwise comparisons of populations in maize (a) and sorghum (b), respectively.
(c) The proportion of non-syntenic genes and syntenic genes under selection in pairwise comparison in maize and sorghum.
(d) Ratio of maize1:maize2 genes among genes identified as selection candidates.
In (c) and (d) analysis was conducted separately for singleton genes and duplicate genes. One asterisk denotes cases which are significantly different at a threshold of $P < 0.01$ and two asterisks denote cases which were significantly different at a threshold of $P < 0.001$. I, improved lines; L, landraces; W, wild relatives. [Colour figure can be viewed at wileyonlinelibrary.com].

in the wild relative improved line comparison (*P*-value = 0.00024 and *P*-value = 1.0×10^{-6}, respectively); however, genes identified as targets of selection during the crop improvement process were significantly more likely to be non-syntenic genes (*P*-value = 0.0026) (Figure 2c).

In maize, selection candidates were also unevenly distributed between subgenomes. Maize1 genes were more likely to be identified as candidates for selection during domestication, both among genes retained as duplicate pairs (1.49×) and genes which fractionated to single copy (1.24×) (*P*-value = 0.00013 and *P*-value < 1.0×10^{-6}, respectively, binomial test) (Figure 2d). Fewer singleton genes than duplicate gene pairs were identified as likely to be under selection during domestication, while the opposite pattern was observed for genes identified as likely to be under selection during improvement; however, these differences were not statistically significant (Figure S2).

Testing for parallel selection during domestication

In order to control for differences in gene content and biases towards syntenic genes, new sets of candidate genes were selected consisting of only genes above the 90th percentile for XP-CLR scores of syntenically conserved genes in each species. Based on the percentage of genes classified as domestication candidates in each species, in the absence of parallel selection on the same genes during domestication 189 gene pairs would be expected to be identified as gene candidates in both species. Among domestication candidate genes 196 gene pairs were identified independently in both species, slightly more than the 189 gene pairs expected in the absence of parallel selection (determined via permutation testing). This difference of seven genes was not statistically significant [false discovery rate (FDR) < 0.27, permutations] (Figure 3a). The gene exhibiting the strongest combined selection signal across the two species *GRMZM2G026024*/*Sobic.004G272100* encodes a phosphoribulokinase, an enzyme that catalyzes a key step in carbon fixation as part of the Calvin cycle.

Comparison of genes under apparent selection in the landrace versus elite comparison identified fewer pairs of syntenic genes under parallel selection than during the domestication (Figure 3b). This may be linked to a lower proportion of the syntenic genes being under selection in maize during the improvement process (Figure 2c). In the case of genes under selection during crop improvement, a total of 186 overlapping genes were expected but 174 were observed (FDR < 0.85, permutations) (Figure 3b). In the wild relatives versus elite comparison, 195 overlapping gene pairs were identified and 188 were expected (FDR < 0.31, permutations) (Figure S3c).

The cut-off of genes in the 90th percentile of XP-CLR scores was chosen somewhat arbitrarily. In order to test whether the lack of a greater than expected overlap between genes identified in maize and those identified in sorghum was an artifact of the threshold score employed, the analysis above was repeated using a range of percentile-based score thresholds from the 85th percentile to the 99th percentile. None of these thresholds identified a significant enrichment of gene pairs under selection in both species relative to the expectations of the null hypothesis (Figure S4a, b).

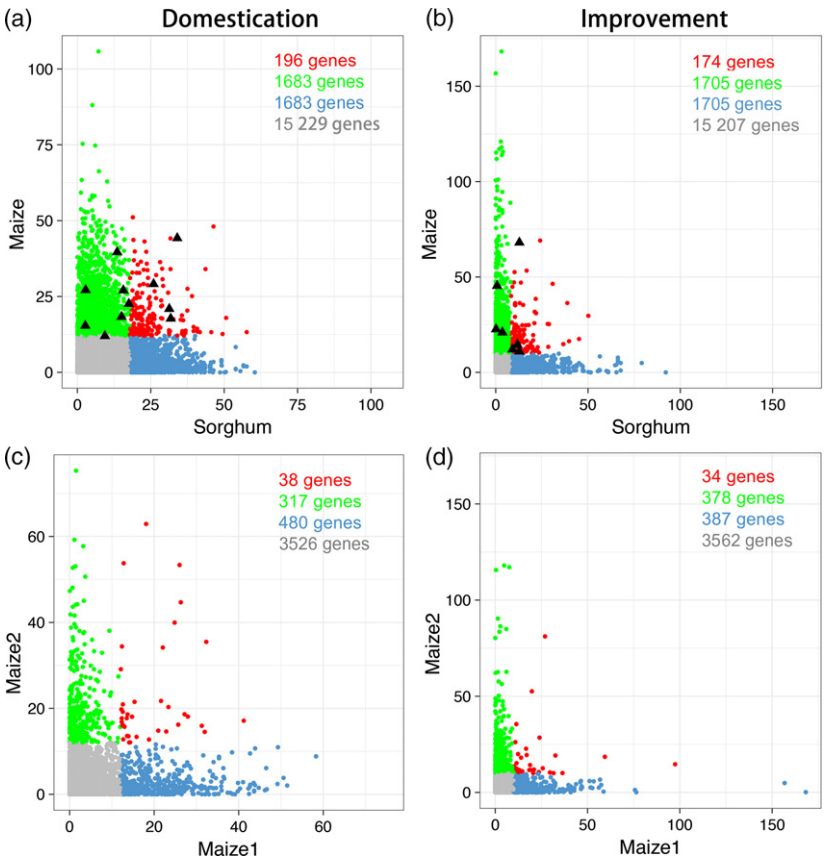

Figure 3. Comparison of scores for syntenic orthologous gene pairs. Comparison of scores for syntenic orthologous gene pairs in the wild relatives/landrace (a) and landrace/improved (b) lines XP-CLR analyses of maize and sorghum as well as the selection in duplicated maize genes during domestication (c) and improvement (d). Red, blue, orange and black dots (in (a) and (b)) mark gene pairs identified as putative selection candidates in both maize and sorghum, only in sorghum, only in maize, or in neither species, respectively. The triangles in (a) and (b) show the classic domestication genes of maize listed in Table 1. Red, blue, orange and black dots (in c and d) mark gene pairs identified as putative selection candidates in both maize1 and maize2, only in maize1, only in maize2, or in neither subgenome, respectively.

Another potential explanation is that the analysis above was partially confounded as a result of the partially paired data structure, with some sorghum genes paired with a single maize syntenic ortholog and others paired with two syntenic orthologs on opposite maize subgenomes. Separate permutation tests were conducted using only maize1–sorghum or maize2–sorghum gene pairs. Slightly more gene pairs were identified as likely under selection during improvement between maize1 and sorghum than expected under the null hypothesis, and slightly fewer gene pairs than expected were identified in the maize2 sorghum comparison. However, this bias was not large and was not replicated in the comparison of genes identified as likely under selection during domestication (Figure S4c-f).

While many phenotypic changes during domestication appear to be shared between sorghum and maize – the domestication syndrome referred to above – domestication probably also involved selection on some traits only in one species or the other. Therefore we also searched for signatures of parallel selection between homeologous gene pairs retained between the two subgenomes of maize, as these genes experienced identical whole-plant-level artificial selection during the domestication of maize. Thirty-eight duplicate pairs under selection during domestication were identified from 4362 pairs of retained maize duplicates tested (Figure 3c) and only 34 duplicate pairs under parallel selection during improvement were identified (Figure 3d). There were fewer duplicate genes under parallel selection during the improvement because of a smaller proportion of syntenic genes under selection during this period. In both cases, the number of gene pairs identified as likely under parallel selection was lower than the expectation for unlinked genes, but not by a statistically significant amount (FDR < 0.88 and 0.80, permutations, respectively).

Another potential explanation for the absence of significant overlap between genes appearing to have experienced selection in maize and sorghum, or between the two maize subgenomes, is simply that the dataset used or the analysis method employed was invalid in some way. To test this concern, we employed a positive control set of 16 maize genes with known and functionally validated links to domestication phenotypes in maize described above (Table 1). All 16 of these genes were indeed included in the set of maize gene candidates identified through XP-CLR analysis. In nine cases the sorghum orthologs of these target genes were also identified as likely targets of artificial selection (P-value < 1.0×10^{-6}, binomial test).

A set of 16 genes shown to exhibit functional variation between maize and teosinte or between maize landraces and improved lines for traits linked to domestication based

Table 1 Well-characterized domestication genes in maize and their orthologs/homeologs

Gene symbol	Gene name	Gene ID in sorghum	Gene ID in maize1	Gene ID in maize2
sh1	Seed shattering1	Sobic.001G152901	GRMZM2G085873	GRMZM2G074124
gt1	Grassy tillers1	Sobic.001G468400	GRMZM2G005624	NoGene
tga1	Teosinte glume architecture1	Sobic.007G193500	NoGene	GRMZM2G101511
tb1	Teosinte branched1	Sobic.001G121600	AC233950.1_FG002	AC190734.2_FG003
ra1	Ramosa1	Sobic.002G197700	GRMZM2G003927	NoGene
ELF4	Early flowering 4	Sobic.002G193000	GRMZM2G025646	NoGene
CCT	Flowering time related	Sobic.002G275100	GRMZM2G179024	NoGene
ohp2	Opaque2 zein storage protein synthesis	Sobic.001G056700	GRMZM2G007063	NoGene
yab14	Yabby14	Sobic.006G160800	GRMZM2G054795	GRMZM2G005353
GI	GIGANTEA	Sobic.003G040900	GRMZM5G844173	GRMZM2G107101
zag2	Zea agamous2	Sobic.008G072900	GRMZM2G010669	GRMZM2G160687
bif2	Barren inflorescence2	Sobic.008G170500	GRMZM2G171822	NoGene
zfl2	Zea floricaula/leafy2	Sobic.006G201600	GRMZM2G180190	GRMZM2G098813
gln2	glutamine synthetase2	Sobic.001G116400	GRMZM2G024104	NoGene
sbe3	starch branching enzyme3	Sobic.006G066800	GRMZM2G073054	NoGene
c2	colorless2	Sobic.005G136200	GRMZM2G422750	GRMZM2G151227

Genes in red, cyan and black represent the genes identified as likely under selection during domestication or improvement phases, or not showing evidence of selection in either process, respectively.

on single-gene or single-gene-family studies was assembled (Table 1). Characterized genes showing signatures of parallel selection in both maize and sorghum include the previously reported *sh1* gene involved in the loss of seed shattering (Lin *et al.*, 2012), genes involved in reshaping plant architecture such as *gt1*, identified as a controller of ear number in maize (Wills *et al.*, 2013), and genes involved in regulation of flowering time such as *ELF4* and *GI (Bendix et al., 2015)*, as well as two important genes in starch synthesis pathway, *ss1* and *sbe3 (Whitt et al., 2002; Campbell et al., 2016)*. The *tb1* gene, which is involved in the repression of axillary branching in both maize (Doebley *et al.*, 1997) and sorghum (Kebrom *et al.*, 2006), showed signatures of parallel selection, and was identified as a selection candidate in both maize and sorghum. However, *tb1* was not one of the strongest signals of selection, and was not even found to be a candidate locus for selection when *mexicana* teosinte lines were included as part of the wild population (Hufford *et al.*, 2012). A second TCP transcription factor, belonging to the same gene family as *tb1*, was identified as under parallel selection in sorghum, maize1 and maize2 (Figure 4).

Functional roles of genes selected in parallel

A total of 1014 maize/sorghum syntenic gene pairs were identified as under parallel selection, including both genes under parallel selection in the same comparison (i.e. wild versus landrance or landrace versus improved) or in opposite comparisons in different species. These genes were enriched in transcription factors relative to all syntenic gene pairs in both sorghum (P-value = 8.70×10^{-4}, Fisher exact test) and maize (P-value = 3.50×10^{-5}, Fisher exact test), although the absolute enrichment is modest (1.36× and 1.46× for maize and sorghum, respectively) (Table S4).

To test whether genes identified as targets of selection in both maize and sorghum exhibited parallel expression patterns, we utilized a set of RNA-seq data generated from homologous tissues in maize and sorghum, with a specific focus on reproductive tissues (Davidson *et al.*, 2011, 2012). A total of 44 genes under apparent parallel selection exhibited conserved patterns of reproductive tissue-specific expression, with anther, embryo and endosperm being represented at the highest frequency (Table S5). One gene which showed strong parallel selective signals in sorghum (*Sobic.009G203900*) and both maize subgenomes – maize1 (*GRMZM2G074361*) and maize2 (*GRMZM2G109842*) – exhibited identical and highly specific expression patterns in the anthers of both species (Figure S5). This gene is annotated as the *profilin 1* (*PRF1*) gene which encodes a core cell-wall structural protein. Overall, no strong biases towards either expression in specific reproductive tissues or greater conservation of tissue-specific expression among genes with statistical signatures of parallel selection were detected.

A total of 237 domestication candidate genes in sorghum were involved in 112 annotated biochemical pathways, and 270 domestication candidate genes from the 18 794 syntenic genes of maize were involved in 113 annotated biochemical pathways. A total of 69 pathways overlapped between these two datasets, which was similar to the 71 pathways predicted to overlap based on permutations of orthologous relationships (FDR < 0.606,

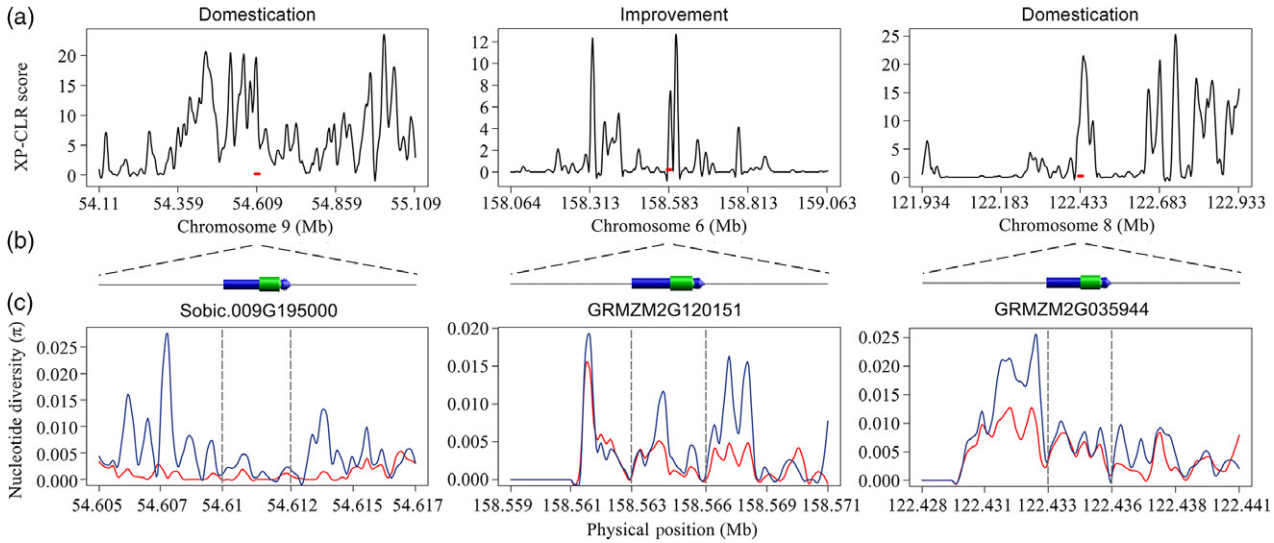

Figure 4. An example of evidence of selection on a *TCP* gene in maize and sorghum.
(a) Cross-population composite likelihood ratio test plot showing selection on partial regions of genomes in sorghum, maize1 and maize2. The red line represents the location of an orthologous *TCP* gene on the chromosomes.
(b) Gene model of syntenic *TCP* genes in sorghum, maize1 and maize2. The blue and green boxes represent the untranslated regions and exons.
(c) Level of nucleotide diversity (π) in the objective population (red) and background population (blue). [Colour figure can be viewed at wileyonlinelibrary.com].

permutations). For genes identified as candidates in the landrace–improved line comparison, a total of 231 candidate genes from sorghum were annotated as being part of 128 pathways and 258 selection candidate genes from maize were annotated as being part of 129 pathways. A total of 80 pathways had at least one maize gene and one sorghum gene under selection, which is moderately statistically significant compared with the expected number of overlapping pathways (FDR < 0.006, permutations). However, when analyzing gene pairs identified as under parallel selection, a set of 141 gene pairs were annotated as encoding enzymes which were involved in 89 different metabolic pathways (Table S4). This was a significantly larger number of metabolic pathways than would be expected given the overall number of gene pairs under parallel selection (expectation = 78 pathways, FDR < 0.028, permutations).

DISCUSSION

The high collinearity of genetic maps and gene content among related grasses (Moore *et al.*, 1995) makes it feasible to employ multiple grass species as a single genomic system. Here we sought to test whether the parallel phenotypic changes produced by artificial selection as part of the domestication syndrome in maize and sorghum resulted from parallel molecular changes targeting orthologous genes. However, as shown above, the number of genes showing parallel signatures of selection during domestication in maize and sorghum was not significantly different from the amount of overlap expected among random gene sets. This result stands in contrast to reports on individual large-effect genes where the same gene often appears to have been a target during independent domestication in different species (Lin *et al.*, 2012; Liu *et al.*, 2015), as well as the finding here that genes with validated links to domestication from single-gene studies in maize were significantly more likely to also be targets of selection in sorghum (Table 1).

A recent report that compared published results from analyses of domestication in maize and rice reached a similar result, with only 65 orthologous gene pairs being shared between a set 969 genes identified as candidate targets of selection in maize and 1526 gene pairs identified as candidate targets of selection in rice (Gaut, 2015). However, that study also identified a number of limitations in their analysis, including the relatively high linkage disequilibrium (LD) in rice, and the potential for selection for different traits during domestication in the two species given the large differences in growth habit between modern rice and maize cultivars. It was also noted that this analysis used candidate gene sets identified by different research groups using different sets of parameters, and even within the same species different scans for positive selection can identify different sets of candidate genes (Akey, 2009). Here we employed data from two more closely related species with similar, and low, levels of LD and greater similarities of plant architecture and growth habit as well as conducting a reanalysis starting from raw SNP calls in order to ensure balanced and equivalent approaches to identifying candidate genes in both species. However, we also found an absence of parallel selection on orthologous genes at a whole-genome level between maize and sorghum.

However, as with the previous comparison between maize and rice, the analyses presented above come with a number of important caveats. The first caveat comes from the differences in the domestication process between the two species. Maize was domesticated from teosinte approximately 9000 years ago in what appears to have been a single event (Piperno *et al.*, 2009; Van Heerwaarden *et al.*, 2011). Improved maize lines used in this study were largely drawn from temperate elite varieties adapted to North America. In contrast, sorghum appears to have been domesticated independently at least twice (Mace *et al.*, 2013), and the improved sorghum lines used here were drawn from separate breeding efforts aimed at developing improved cultivars for African, Australian and North American climates. Parallel domestication and independent crop improvement efforts in sorghum are likely to reduce the statistical power to identify genes which are targets of selection, both because different haplotypes of the same genes may have been targets of selection during distinct domestication or crop improvement efforts and because different domestication and crop improvement efforts may have targeted different genetic loci. While the statistical approach employed by XP-CLR to identify signatures of selective sweeps is a significant advance over previous approaches in that it can detect both 'hard' sweeps, where a single beneficial haplotype at a given locus rapidly increases in frequency as a result of selection, and 'soft' sweeps, where multiple pre-existing haplotypes provide the same fitness advantage, either as a result of recombination or independent origins (Hermisson and Pennings, 2005; Przeworski *et al.*, 2005), the statistical signatures of soft sweeps remain more difficult to detect. Sweeps which reflect selection in only a subset of a group of germplasm, for example genes under selection in elite North American sorghum lines but not in elite Australian germplasm, are also likely to be missed by the current analysis. The reproductive habits of the two species may also have altered the relative contributions of standing genetic variation, likely to contribute to soft sweeps, and novel mutations, likely to contribute to hard sweeps, to the domestication syndrome in maize and sorghum. Under field conditions outcrossing rates for sorghum have been reported to be in the range of 7–18% (Djè *et al.*, 2004; Barnaud *et al.*, 2008), while teosinte outcrossing rates can reach ~97% (Hufford *et al.*, 2011). The high outcrossing rate of wild teosinte may have allowed tolerance of alleles with a wider range of phenotypic consequences, producing a deeper pool of standing functional genetic variation than would have been present in more inbred wild sorghum plants.

The observation that in maize, but not in sorghum, non-syntenic genes were enriched among genes targeted for selection during crop improvement was unexpected. One potential explanation is the relative importance of hybrid breeding and heterosis for these two crops. The absolute size of the effect of heterosis for maize yield has remained constant while the inbred yield values have increased (Schnell, 1974). The effect of heterosis on yield in early maize single-cross hybrids was 300% in 1930s crosses, and in relative terms has slowly decreased to 100% in relative terms (Duvick, 2005a). In contrast, in sorghum, the increase in yield resulting from heterosis is generally of the order of 40% (Duvick, 1999; Mindaye *et al.*, 2016). Non-syntenic genes are more likely to exhibit presence–absence variation across different maize lines (Swanson-Wagner *et al.*, 2010; Schnable *et al.*, 2011) and to display non-additive expression (Paschold *et al.*, 2014). Both presence–absence variation and non-additive expression have been speculated to contribute to heterosis. The greater emphasis on heterosis and contribution to yield of heterosis in maize relative to sorghum may therefore have resulted in a greater proportion of artificial selection targeting non-syntenic genes in maize.

One final assumption in the analyses presented here is that selection during domestication truly did target the same phenotypic traits in both maize and sorghum. While this should generally be the case, selection during crop improvement has differed between these two species in at least one key phenotype: selection for higher yield in maize has resulted in indirect selection for decreased tassel size (Duvick, 2005b), while selection for decreased head size in sorghum would presumably be detrimental to yield. As a partial control for the potential explanation, namely that the lack of identified overlap between maize and sorghum resulted from selection for different traits during domestication in these two species, we also compared patterns of selection between conserved homeologous genes in the different maize subgenomes. These genes started out with equivalent functional roles prior to the maize whole-genome duplication, and experienced the same selective pressures during domestication and crop improvement. However, we also failed to identify any statistically significant correlation between genes identified as targets of selection between the two subgenomes. Another explanation, that selection targeted different genes in the same pathways, also failed to find support in this study; however, it should be noted that improved annotations of biochemical and transcriptional pathways may produce a different result in the future. Finally, the re-identification of many genes from a set of positive control genes previously identified through top-down approaches as playing a role in domestication provided a validation that the statistical methods, software implementations and genomic datasets employed do indeed have the power to identify genes which were targets of selection during domestication and crop improvement.

The observation that the few maize domestication genes characterized in conventional single-gene genetic studies were much more likely to also be identified as

domestication candidates in sorghum (Table 1) suggests that the genes involved in domestication may fall into a two-tier system. A few large-effect genes appear to have been repeatedly targeted to create the domestication syndrome in multiple grain crops (Lin *et al.*, 2012; Liu *et al.*, 2015). As described previously, strong and relatively recent selection should rapidly identify and fix a small number of large-effect alleles which may have pleiotropic consequences, while a range of different smaller-effect alleles at other genetic loci are then selected to fine tune the effect size and mitigate any negative pleiotropic effects of the initial large-effect alleles (Orr, 1998). Studies of the genetic architecture of different traits inferred to be under selection or largely neutral characters during domestication and crop improvement in maize and maize–teosinte RIL populations have produced findings consistent with this model (Wallace *et al.*, 2014; Xu *et al.*, 2017).

Here we propose that the initial, large-effect alleles selected for during selection for domestication syndrome traits in grain crops are drawn from a constrained pool of genes, and therefore orthologous genes are more likely to be selected for in parallel across multiple grain crops. In contrast, the set of small-effect genes which fine-tune domestication rates and mitigate potentially deleterious pleiotropic effects of large-effect alleles may be drawn from a much larger pool and would thus exhibit little repeat sampling of the same orthologous genes across different domesticated grasses. Alternatively, these fine-tuning genetic changes may more frequently be drawn from standing genetic variation, resulting in more soft sweeps or incomplete sweeps, reducing the statistical power to consistently identify these genes in repeated statistical trials across different species. Indeed, simulation studies suggest that small-effect loci are more likely to become fixed during selection if they originate from standing genetic variation than from novel mutations (Hermisson and Pennings, 2005). An additional potential confounding variable is that the mutational target space – defined as the number of potential mutations at a given locus which produce alleles with the same phenotypic outcome – may well vary between genes with larger and smaller phenotypic effects. A larger mutational target space at a given locus increases the probability that the response to artificial selection during domestication will result from a multiple-origin soft sweep and reduces the potential to identify the locus as a target of selection using bottom-up population genetic approaches (Hermisson and Pennings, 2017). The decreasing cost of whole-genome sequencing and resequencing, and the large number of different grain crops that have experienced parallel selection for domestication syndrome phenotypes, should enable more rigorous tests of this model in the near future incorporating data from syntenic orthologous genes across many different species.

EXPERIMENTAL PROCEDURES

Data collection and preliminary polishing

The maize and sorghum whole-genome resequencing data used in this study were taken from Hapmap2 (Chia *et al.*, 2012) and SorGSD (Luo *et al.*, 2016), respectively. SNPs that scored as heterozygous in 3.0% of accessions in the maize Hapmap2 and sorghum dataset were removed prior to analysis. A subset of maize accessions was selected and separated into three groups: 30 improved lines, 19 landraces and 7 wild relatives (Table S1). Data from a total of 42 sorghum accessions were obtained, including 17 improved lines, 18 landraces and 7 wild relatives (Table S1). SNPs with missing rates of >50% in either species, or with heterozygous calls in any of the remaining samples, were discarded, resulting in a final dataset consisting of 10.3 million SNPs in maize and 3.3 million SNPs in sorghum.

Population genetics analysis

The genetic distance between individuals was first calculated using a 0.1% subset of the total SNP set constructed by sampling every 1000th SNP position along each chromosome for a total of 10 286 SNPs in maize and using a 0.2% subset of the total SNP set constructed by sampling every 500th SNP position along each chromosome for a total of 6719 SNPs in sorghum.

Neighbor-joining trees were constructed for the accessions of each species using Phase (Jow *et al.*, 2002) and Phylip v3.696 (Felsenstein, 1981) with default parameters. The resulting phylogenetic trees were visualized using Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Nucleotide diversity (π) values were calculated for each species with non-overlapping windows of 10 kb using an in-house Perl script (https://figshare.com/articles/Tajima_D_pl/5544484). Reported π values are the average of all genomic windows.

Syntenic gene identification

A pan-grass syntenic gene set using the sorghum genes as reference was download from figshare (Schnable *et al.*, 2016). When multiple sorghum genes were identified as syntenic orthologs of the same gene in maize – a result that can be produced by tandem duplication events in sorghum – the tie was broken using a separate dataset of syntenic orthologous genes using the *Setaria italica* genome as a reference. This resulted in a final set of 14 433 sorghum genes paired with a syntenic ortholog in either the maize1 subgenome (11 402 gene pairs) and/or the maize2 subgenome (7392 gene pairs), including 4361 sorghum genes with syntenic co-orthologs on both maize subgenomes (Table S2).

Genome-wide scan for selection

To identify genes affected by selection during domestication in maize and sorghum, genome-wide scans for signals of selection were conducted using a cross-population composite likelihood approach (XP-CLR) [Chen *et al.*, 2010; updated by Hufford *et al.* (2012) to incorporate missing data], based on the allele frequency differentiation between target and reference populations. This approach was employed in three separate pair wise comparisons: wild relatives versus landraces, landraces versus improved lines and wild relatives versus improved lines.

Recombination rates in maize and sorghum were measured using high-density genetic maps constructed using RILs from biparental crosses in maize (Ott *et al.*, 2017) and sorghum (Zou *et al.*, 2012). Genetic maps were transferred to the more recent

versions of the maize and sorghum genomes used in this analysis (B73 RefGen v3 and v3.1, respectively). The transfer was performed using the two genes flanking the marker (when the marker was in a non-coding region) or the single gene the marker was located in. For each pseudomolecule in maize, a ninth-order polynomial curve was fitted to the genetic and physical coordinates of all markers presented on chromosomes, and genetic positions for each marker were reassigned based on the value predicted for the genetic and physical position of the marker and the polynomial formula.

The same parameters were employed for XP-CLR analysis for maize and sorghum. A 0.05-cM sliding window with 1000-bp steps across the whole-genome scan was used for scanning. Individual SNPs were assigned a position along the genetic map based on the polynomial fitting curves in maize and by assuming uniform recombination between pairs of genetic markers in sorghum. The number of SNPs assayed in each window was fixed at 100 and pairs of SNPs in high LD ($r^2 > 0.75$) were down-weighted.

To obtain XP-CLR scores for each gene, each gene was assigned a window starting 5 kb upstream of its annotated transcription start site and extending to 5 kb downstream of its annotated transcription stop site. The maximum XL-CLR score among all the XL-CLR intervals within this window was assigned to the gene.

Testing for enrichment of genes selected in parallel

Gene pairs were considered to be under selection if a sorghum gene and at least one maize syntenic ortholog were both identified as being under selection. To determine the optimal cut-off for testing the enrichment of syntenic genes under parallel selection, a series of cut-offs from 85% to 99% were used in the analysis. At each cut-off the number of gene pairs under selection in both species were recorded and compared with the number of gene pairs identified when orthologous relationships between maize and sorghum were shuffled using a permutation test repeated 100 times.

Gene annotation and enrichment analysis

Maize and sorghum GO annotations were retrieved from Phytozome (https://phytozome.jgi.doe.gov/). The maize transcription factor (TF) lists were downloaded from Grassius (http://grassius.org/grasstfdb.html). Metabolic pathway lists were downloaded from the Gramene (ftp://ftp.gramene.org/pub/gramene/pathways/). Annotated enzyme name and the corresponding pathways for these genes were obtained by searching the pathway list.

A set of 1000 permutations was used to calculate the expected number of pathways in the same number of random genes in the R package to examine whether maize and sorghum genes under selection were significantly more likely to be present in the same pathways than expected if selection was unlinked.

ACKNOWLEDGEMENTS

We thank Professor Edward Buckler (Cornell University) for advice on the Hapmap dataset and Professor Jeffrey Ross-Ibarra (UC Davis) for advice and access to an updated version of the XP-CLR software. This work was supported by a China Scholarship Council fellowship awarded to XL and a Science Foundation of Xichang College awarded to LY.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genome-wide scan for signatures of selection in maize and sorghum.

Figure S2. Ratio of singleton genes: duplicate pairs among genes identified as selection candidates.

Figure S3. Comparison of scores for syntenic orthologous gene pairs in the wild relative/improved line XP-CLR analyses of maize and sorghum.

Figure S4. Testing for enrichment of candidate selection genes for different proportions of genes with top XP-CLR scores.

Figure S5. Gene model and gene expression level visualization of *profilin1* in sorghum and maize.

Table S1. List of maize and sorghum accessions employed in this study, their data sources and their classifications as wild relative, landrace or improved line datasets.

Table S2. Set of high-confidence syntenic orthologous maize–sorghum gene pairs employed in this study.

Table S3. Population nucleotide diversity statistics for maize and sorghum.

Table S4. A list of the genes identified as likely under selection during the domestication and/or improvement process in maize and sorghum and functional annotations of each.

Table S5. Tissue-specific expression of genes under parallel selection between maize and sorghum.

REFERENCES

Akey, J.M. (2009) Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.*, **19**, 711–722.

Baldauf, J.A., Marcon, C., Paschold, A. and Hochholdinger, F. (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive, and allelic expression patterns in maize hybrids. *Plant Physiol.*, **171**, 1144–1155.

Barnaud, A., Trigueros, G., McKey, D. and Joly, H.I. (2008) High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? *Heredity*, **101**, 445.

Bendix, C., Marshall, C.M. and Harmon, F.G. (2015) Circadian clock genes universally control key agricultural traits. *Molecular plant*, **8**, 1135–1152.

Bennetzin, J.L. and Freeling, M. (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. *Trends Genet.*, **9**, 259–261.

Bradbury, L.M., Fitzgerald, T.L., Henry, R.J., Jin, Q. and Waters, D.L. (2005) The gene for fragrance in rice. *Plant Biotechnol. J.*, **3**, 363–370.

Campbell, B.C., Gilding, E.K., Mace, E.S., Tai, S., Tao, Y., Prentis, P.J., Thomelin, P., Jordan, D.R. and Godwin, I.D. (2016) Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol. J.*, **14**, 2240–2253.

Chen, H., Patterson, N. and Reich, D. (2010) Population differentiation as a test for selective sweeps. *Genome Res.*, **20**, 393–402.

Chia, J.-M., Song, C., Bradbury, P.J. *et al.* (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.*, **44**, 803–807.

Clark, R.M., Linton, E., Messing, J. and Doebley, J.F. (2004) Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proc. Natl Acad. Sci. USA*, **101**, 700–707.

Davidson, R.M., Hansey, C.N., Gowda, M. *et al.* (2011) Utility of RNA sequencing for analysis of maize reproductive transcriptomes. *The Plant Genome*, **4**, 191–203.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.-H., Jiang, N. and Robin Buell, C. (2012) Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J.*, **71**, 492–502.

Dirzo, R. and Raven, P.H. (2003) Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.*, **28**, 137–167.

Djè, Y., Heuertz, M., Ater, M., Lefèbvre, C. and Vekemans, X. (2004) In situ estimation of outcrossing rate in sorghum landraces using microsatellite markers. *Euphytica*, **138**, 205–212.

Doebley, J. and Stec, A. (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics*, **134**, 559–570.

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

Dorweiler, J., Stec, A., Kermicle, J. and Doebley, J. (1993) Teosinte glume architecture 1: a genetic locus controlling a key step in maize evolution. *Science*, **262**, 233–233.

Duvick, D.N. (1999) Heterosis: Feeding people and protecting natural resources. In *The Genetics and Exploitation of Heterosis in Crops* (Coors, J.G. and Pandey, S. eds). ASA, CSSA: Madison, WI, pp. 19–29.

Duvick, D.N. (2005a) The contribution of breeding to yield advances in maize (Zea mays L.). *Adv. Agron.*, **86**, 83–145.

Duvick, D.N. (2005b) Genetic progress in yield of United States maize (Zea mays L.). *Maydica*, **50**, 193.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.

Gaut, B.S. (2015) Evolution Is an Experiment: assessing Parallelism in Crop Domestication and Experimental Evolution. *Mol. Biol. Evol.*, **32**, 1661–1671.

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

Harlan, J.R., Wet, J.D. and Price, E.G. (1973) Comparative evolution of cereals. *Evolution*, **27**, 311–325.

Hermisson, J. and Pennings, P.S. (2005) Soft sweeps. *Genetics*, **169**, 2335–2352.

Hermisson, J. and Pennings, P.S. (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.*, **8**, 700–716.

Huang, X., Kurata, N., Wang, Z.-X. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497.

Hufford, M.B., Gepts, P. and ROSS-IBARRA, J. (2011) Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (Zea mays ssp. parviglumis). *Mol. Ecol.*, **20**, 46–55.

Hufford, M.B., Xu, X., van Heerwaarden, J. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.*, **44**, 808–811.

Jow, H., Hudelot, C., Rattray, M. and Higgs, P.G. (2002) Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Mol. Biol. Evol.*, **19**, 1591–1601.

Kebrom, T.H., Burson, B.L. and Finlayson, S.A. (2006) Phytochrome B represses Teosinte Branched1 expression and induces sorghum axillary bud outgrowth in response to light signals. *Plant Physiol.*, **140**, 1109–1117.

Lai, X., Behera, S., Liang, Z., Lu, Y., Deogun, J.S. and Schnable, J.C. (2017) STAG-CNS: an Order-Aware Conserved Non-coding Sequences Discovery Tool For Arbitrary Numbers of Species. *Molecular Plant*, **10**, 990–999.

Li, L., Briskine, R., Schaefer, R., Schnable, P.S., Myers, C.L., Flagel, L.E., Springer, N.M. and Muehlbauer, G.J. (2016) Co-expression network analysis of duplicate genes in maize (Zea mays L.) reveals no subgenome bias. *BMC Genom.*, **17**, 875.

Lin, Z., Li, X., Shannon, L.M. *et al.* (2012) Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.*, **44**, 720–724.

Liu, H., Liu, H., Zhou, L., Zhang, Z., Zhang, X., Wang, M., Li, H. and Lin, Z. (2015) Parallel domestication of the heading date 1 gene in cereals. *Mol. Biol. Evol.*, **32**, 2726–2737.

Luo, H., Zhao, W., Wang, Y. *et al.* (2016) SorGSD: a sorghum genome SNP database. *Biotechnol. Biofuels*, **9**, 1.

Mace, E.S., Tai, S., Gilding, E.K. *et al.* (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.*, **4**, 2320–2320.

Mauro-Herrera, M., Wang, X., Barbier, H., Brutnell, T.P., Devos, K.M. and Doust, A.N. (2013) Genetic control and comparative genomic analysis of flowering time in Setaria (Poaceae). *G3: Genes. Genomes, Genetics*, **3**, 283–295.

Mindaye, T.T., Mace, E.S., Godwin, I.D. and Jordan, D.R. (2016) Heterosis in locally adapted sorghum genotypes and potential of hybrids for increased productivity in contrasting environments in Ethiopia. *The Crop Journal*, **4**, 479–489.

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Cereal genome evolution: grasses, line up and form a circle. *Curr. Biol.*, **5**, 737–739.

Orr, H.A. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, **52**, 935–949.

Ott, A., Liu, S., Schnable, J. C., Yeh, C., Wang, K. and Schnable, P.S. (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Res.* **45**, e178.

Paschold, A., Larson, N.B., Marcon, C., Schnable, J.C., Yeh, C.-T., Lanz, C., Nettleton, D., Piepho, H.-P., Schnable, P.S. and Hochholdinger, F. (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell*, **26**, 3939–3948.

Paterson, A.H., Lin, Y.-R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R., Liu, S.-C., Stansel, J.W. and Irvine, J.E. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**, 1714–1718.

Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J. and Dickau, R. (2009) Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci.*, **106**, 5019–5024.

Pophaly, S.D. and Tellier, A. (2015) Population level purifying selection and gene expression shape subgenome evolution in maize. *Mol. Biol. Evol.*, **32**, 3226–3235.

Przeworski, M., Coop, G. and Wall, J.D. (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.

Renny-Byfield, S., Rodgers-Melnick, E. and Ross-Ibarra, J. (2017) Gene fractionation and function in the ancient subgenomes of maize. *Mol. Biol. Evol.*, **34**, 1825–1832.

Ross-Ibarra, J., Morrell, P.L. and Gaut, B.S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl Acad. Sci.*, **104**, 8641–8648.

Sagnard, F., Deu, M., Dembélé, D. *et al.* (2011) Genetic diversity, structure, gene flow and evolutionary relationships within the Sorghum bicolor wild?weedy?crop complex in a western African region. *Theoretical and applied genetics*, **123**, 1231.

Schnable, J.C. (2015) Genome evolution in maize: from genomes back to genes. *Annu. Rev. Plant Biol.*, **66**, 329–343.

Schnable, J.C. and Freeling, M. (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE*, **6**, 17855.

Schnable, J.C., Springer, N.M. and Freeling, M. (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci.*, **108**, 4069–4074.

Schnable, J., Zang, Y. and Ngu, W.C.D. (2016) Pan-Grass Syntenic Gene Set (sorghum referenced). *Figshare*, https://doi.org/10.6084/m9.figshare.3113488.v1.

Schnell, F.W. (1974) *Trends and problems in breeding methods for hybrid corn*. Birmingham, England: Proc. of the British Poultry Breeders Roundtable, 16th. pp. 86–98.

Sugimoto, K., Takeuchi, Y., Ebana, K. *et al.* (2010) Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl Acad. Sci.*, **107**, 5792–5797.

Swanson-Wagner, R.A., Eichten, S.R., Kumari, S., Tiffin, P., Stein, J.C., Ware, D. and Springer, N.M. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.*, **20**, 1689–1699.

Swigoňová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L. and Messing, J. (2004) Close split of sorghum and maize genome progenitors. *Genome Res.*, **14**, 1916–1923.

Takuno, S., Ralph, P., Swarts, K., Elshire, R.J., Glaubitz, J.C., Buckler, E.S., Hufford, M.B. and Ross-Ibarra, J. (2015) Independent molecular basis of convergent highland adaptation in maize. *Genetics*, **200**, 1297–1312.

Tao, Y., Mace, E.S., Tai, S., Cruickshank, A., Campbell, B.C., Zhao, X., van Oosterom, E.J., Godwin, I.D., Botella, J.R. and Jordan, D.R. (2017) Whole-genome analysis of candidate genes associated with seed size and weight in Sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. *Frontiers in Plant Science*, **8**, 1237.

Van Heerwaarden, J., Doebley, J., Briggs, W.H., Glaubitz, J.C., Goodman, M.M., Gonzalez, J.d.J.S. and Ross-Ibarra, J. (2011) Genetic signals of

origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl Acad. Sci.*, **108**, 1088–1092.

Wallace, J.G., Bradbury, P.J., Zhang, N., Gibon, Y., Stitt, M. and Buckler, E.S. (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.*, **10**, 1004845.

Wendorf, F., Close, A.E., Schild, R., Wasylikowa, K., Housley, R.A., Harlan, J.R. and Królik, H. (1992) Saharan exploitation of plants 8,000 years BP. *Nature*, **359**, 721–724.

Whipple, C.J., Kebrom, T.H., Weber, A.L., Yang, F., Hall, D., Meeley, R., Schmidt, R., Doebley, J., Brutnell, T.P. and Jackson, D.P. (2011) Grassy tillers 1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proc. Natl Acad. Sci.*, **108**, 506.

Whitt, S.R., Wilson, L.M., Tenaillon, M.I., Gaut, B.S. and Buckler, E.S. (2002) Genetic diversity and selection in the maize starch pathway. *Proc. Natl Acad. Sci.*, **99**, 12959–12962.

Wills, D.M., Whipple, C.J., Takuno, S., Kursel, L.E., Shannon, L.M., Ross-Ibarra, J. and Doebley, J.F. (2013) From many, one: genetic control of prolificacy during maize domestication. *PLoS Genet.*, **9**, 1003604.

Xu, G., Wang, X., Huang, C. *et al.* (2017) Complex genetic architecture underlies maize tassel domestication. *New Phytol.*, **214**, 852–864.

Yu, B., Lin, Z., Li, H. *et al.* (2007) TAC1, a major quantitative trait locus controlling tiller angle in rice. *Plant J.*, **52**, 891–898.

Zou, G., Zhai, G., Feng, Q. *et al.* (2012) Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *J. Exp. Bot.*, **63**, 5451–5462.

