Theor Appl Genet (2015) 128:1669–1683
DOI 10.1007/s00122-015-2523-z

ORIGINAL PAPER

Allelic variants in the *PRR37* gene and the human-mediated dispersal and diversification of sorghum

Robert R. Klein[1] · Frederick R. Miller[2] · Diana V. Dugas[1] · Patrick J. Brown[3] · A. Millie Burrell[4] · Patricia E. Klein[4]

Received: 3 December 2014 / Accepted: 17 April 2015 / Published online: 16 May 2015

Abstract

Key message **Allele phylogenetic analysis of the sorghum flowering-time gene *PRR37* provided new insight into the human-mediated selection of a key adaptive gene that occurred during sorghum's diversification and worldwide dispersal.**

Abstract The domestication and spread of the tropical cereal sorghum is associated with the historic movement of humans. We show that an allelic series at *PRR37* (pseudo-response regulator 37), a circadian clock-associated transcription factor, was selected in long-day ecosystems worldwide to permit floral initiation and grain production. We identified a series of loss-of-function (photoperiod-insensitive) alleles encoding truncated PRR37 proteins, alleles with key amino acid substitutions in the pseudo-receiver domain, and a novel splice variant in which the pseudo-receiver domain is truncated. Each *PRR37* allelic variant was traced to a specific geographic location or specialized agronomic type. We present a graphical model that shows evidence of human selection and gene flow of the *PRR37* allelic variants during the global dispersal and agronomic diversification of sorghum. With the recent identification of the *Ghd7* gene as an important regulator of flowering date in sorghum, we briefly examine whether loss-of-function *Ghd7* allelic variants were selected prior to the human-mediated movement of sorghum from its equatorial center of origin to temperate climates worldwide.

Communicated by H.-C. Jing.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2523-z) contains supplementary material, which is available to authorized users.

✉ Patricia E. Klein
pklein@tamu.edu

[1] United States Department of Agriculture-Agricultural Research Service, and Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX 77843, USA

[2] MMR Genetics, L.L.C, 3095 County Road 26, PO Box 60, Vega, TX 79092, USA

[3] Department of Crop Science, University of Illinois, Urbana, IL 61801, USA

[4] Department of Horticultural Sciences, and Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX 77843, USA

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a tropical cereal with its first appearance at about 6000 years before present (BP) originating in the savanna region of what is present-day Sudan (Rosenow and Dahlberg 2001; Wendorf et al. 1992). Sorghum was a key component of the African food complex that also included finger millet, banana, and Asian yam, and this food complex was critical for the rapid increase and dispersion of the human population (Mann et al. 1983; Russell et al. 2014). Expansion from its equatorial center of origin exposed primitive sorghum to climatic and ecological constraints that would significantly impact floral initiation and hence, grain production. Of particular importance were biologically significant increases in day length that would delay or prevent flowering as different groups of early humans migrated or traded with peoples north and south of the equator. During this global diffusion of sorghum that commenced about 5000 years BP, parallel selection of photoperiod-insensitive floral initiation alleles was apparently occurring at independent sites across Africa and the Middle East (Quinby 1967). Therefore, by the time

Benjamin Franklin introduced sorghum (Chinese broomcorn) to the USA in 1757 CE (Berenji et al. 2011; Martin and Leonard 1949), sorghum ecotypes had already undergone diversifying selection in tropical and temperate climates across the continents of Africa and Asia. The result of this human-mediated selection is a domesticated cereal with thousands of distinct tropical and temperate-adapted ecotypes with specialized uses such as a gluten-free grain, livestock forage, raw material for household implements, and more recently a cellulosic- and ethanol-based biofuel (Rooney et al. 2007).

Sorghum geneticists dating back to the 1960s CE had concluded that the presence of a recessive photoperiod-insensitive allele at the maturity locus ma_1 was critical for the development of early-flowering sorghum that would produce grain in long-day environments of the higher latitudes (Quinby 1974; Rosenow and Dahlberg 2001; Stephens et al. 1967). We have cloned the sorghum maturity gene ma_1, which was identified as a member of the pseudo-response regulator 37 gene family (*PRR37*), a central component of the biological circadian clock in grasses (Murphy et al. 2011). *PRR37* (also known as ma_1) in sorghum is a central floral repressor that blocks the transition of vegetative apices to floral meristems under non-permissive day lengths (>12 h) by repressing the expression of *FT* (*FLOWERING LOCUS T)* through its enhancement of the expression of *CO* (*CONSTANS*), a repressor of *FT*, and by inhibiting *Ehd1* (*EARLY HEADING DATE 1*) a grass-specific inducer of *FT* (Doi et al. 2004). An initial examination of a limited number of temperate-zone sorghum cultivars revealed several photoperiod-insensitive *PRR37* allelic variants, each with a unique loss-of-function mutation. This discovery of an allelic series at the *PRR37* gene supported the assertion of sorghum geneticist Roy Quinby (1974) that selection at the first maturity locus (ma_1) occurred multiple times in parallel in diverse temperate-zone ecosystems during sorghum's diversification by early humans and more recently by nineteenth century farmers in the Southern Great Plains of the USA.

In this study, we have examined the allele phylogeny of the sorghum flowering-time gene *PRR37* to provide a better understanding of the selection history and gene flow of key adaptive *PRR37* alleles. Due to the poor (charred seed) quality and limited quantities of archeological remains of sorghum, we have examined the allelic and genomic diversity in both extant landraces and a series of historical founder cultivars from sorghum's centers of origin in Africa and from diverse geographical locations where human migration and trade have transported sorghum. We found that human-meditated selection of *PRR37* allelic variants was evident during latitudinal transitions associated with human movements dating from about 5000 years BP to the early twentieth century. We detected unique *PRR37* alleles associated with the human-mediated development of specialized agronomic types of sorghum including forage, sweet-stem, and Chinese broomcorn cultivars. We also describe how the discovery of a rogue early-flowering plant in a field of sorghum by a nineteenth century farmer in the Southern Great Plains of the USA, in conjunction with modern breeding and international commerce, has led to the global dispersal of a once rare photoperiod-insensitive *PRR37* allele. Finally, with the recent identification of *Ghd7* as an important regulator of flowering date in sorghum (Murphy et al. 2014), we examined whether natural photoperiod-insensitive *Ghd7* alleles were central to the human-mediated movement of sorghum cultivation to temperate climates worldwide.

Materials and methods

Historic sorghum cultivars and landraces

To select the set of landraces and historic cultivars included in this study, we consulted sorghum historians and scientists, and the resulting list of historical cultivars and landraces examined are listed in Online Supplementary material 1. The integrity of the ecotype samples was checked by comparing phenotypic descriptors of the accessions to reference descriptors detailed in the Germplasm resources information network (GRIN, http://www.ars-grin.gov/npgs/orders.html); those accessions where major discrepancies were observed were eliminated from the study. Sorghum accessions were grown in greenhouses under short-day (10-h light/14-h dark) and long-day (14-h light/10-h dark) conditions or in winter nurseries in Puerto Vallarta, Mexico, to permit examination of the panicle and seed for confirmation of race classification and cultivar identity. The historic cultivars selected represent genotypes with known pedigrees whose development predates the industrialization of agriculture and extensive crop exchange beginning about 1965 CE. Each landrace is historically associated with a specific geographic location where the various specialized races of sorghum were domesticated. We have included wild/weedy ecotypes from the multiple centers of sorghum diversification across Africa to sample the depth of genetic diversity present in these geographically and culturally isolated regions. Since sorghum ecotypes have been historically selected by humans for use as forage grasses and for molasses syrup production (sweet sorghums), we have examined historic cultivars and landraces developed for these specialized agronomic purposes. To provide a more accurate estimate of the frequency and distribution of the different photoperiod-insensitive *PRR37* alleles, we sequenced the genic regions spanning the different *PRR37* functional mutation candidates in an additional panel of 253 supplemental landraces and historic sorghum cultivars as listed in Online Supplementary material 3.

DNA and RNA extraction

We purified DNA from 6-day-old seedlings (constant 29 °C, 10-h light/14-h dark) using the FastDNA Spin Kit (MP Biomedicals, Inc.). Total RNA was isolated from leaves of 20-day-old plants (14-h light/10-h dark, 30–34 °C/21–25 °C) 4 h after the start of the light period. Leaf tissue was ground with a mortar and pestle under $LiqN_2$, and RNA extracted using the miRNeasy Kit (Qiagen, Inc.). Total RNA was treated with TURBO DNase (Life Technologies, Inc.), and DNase-treated RNA was used in first-strand cDNA synthesis with SuperScript III (Life Technologies) primed with random hexamers.

Sequence analysis of *PRR37* alleles

Three regions spanning the *PRR37* gene space were sequenced for detection of mutations and for gene haplotype analysis. For each of the 48 sorghum accessions listed in Online Supplementary material 1, we sequenced a 4356 nucleotide region of the *PRR37* gene space that included the following; 1143 nucleotides upstream of the *PRR37* transcription start site (TSS), the complete 5′ UTR and the first three introns that reside between the TSS and the translation start site, and the entire coding region of the predicted protein. Each region was amplified by PCR prior to sequencing and subsequently concatenated in silico. Introns 4–10 within the *PRR37* protein-coding region were not resequenced since these regions were determined to be mostly monomorphic in four divergent genotypes that we sequenced prior to the present investigation. We utilized genomic DNA to sequence the region upstream of the translation start site and intronic sequences, and first-strand cDNA to sequence the coding region of the protein. We sequenced cDNA spanning the coding region of *PRR37* to reveal alternative splicing events that would not necessarily be evident (e.g., a mutation in consensus intron/exon splice sites) in sequencing genomic DNA. PCR-amplified products (Phusion High-fidelity DNA Polymerase, New England BioLabs, Inc.; or GoTaq, Promega, Inc.) were treated with ExoSAP-IT (Affymetrix) and both DNA strands were directly sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Invitrogen). The results were assembled and analyzed using Phred/Phrap and Consed (http://www.phrap.org/phredphrapconsed.html). When agarose gel electrophoresis of PCR products revealed alternative splice events in cDNA (i.e., multiple PCR bands), each DNA band was gel excised and purified (Gel Extraction Kit, Qiagen) prior to direct sequencing as detailed above.

For the supplemental accessions listed in Online Supplementary material 3, we assessed the frequency of the different photoperiod-insensitive alleles by sequencing PCR products that spanned the functional mutations of each *PRR37* allele in each of the 253 accessions. To estimate the frequency of the day-neutral allele $prr37^{Kafir-1}$ and the related *PRR37* haplotype observed in guinea margaritiferum accessions, we sequenced additional regions of the gene that encode mutations specific to these alleles (see Online Supplementary material 3, rightmost columns). Allele frequencies of supplemental Asian kaoliangs were not determined due to the repeated failure of PCR amplification spanning the $prr37^{Durra-Kaoliang}$ functional mutation from genomic DNA and cDNA template.

Allele genealogy relationships

To construct a median-joining network of *PRR37* gene sequences, we utilized the software package NETWORK 4.6.1.1 (Bandelt et al. 1999). We identified polymorphic positions based on DNA sequence alignments in ClustalW (http://workbench.sdsc.edu/) of the complete 4356 nucleotide alignment of the 48 historic sorghum cultivars and landraces examined (Online Supplementary material 1, 2). In the absence of ancient sorghum collections from which PCR-grade DNA can be obtained, we utilized a set of landraces (PI 302118, PI 329250, PI 329251, PI 532565) from sorghum's center of origin in Northeastern Africa as the wild-type *PRR37* (photoperiod responsive) reference sequence. SNPs and INDELs were extracted from the sequences using Tassel 3.0 (Bradbury et al. 2007), and both SNPs and INDELs were considered in the analysis; rare polymorphisms that were only observed in a single genotype were excluded to prevent excessive impact of sequencing/PCR errors. Simple sequence repeats (SSRs) were also excluded from the analysis. Utilizing the median-joining option of NETWORK, we assigned weights to different polymorphisms as follows: weight 20 for functional mutations (i.e., nonsense mutations, frameshift mutations, and nonsynonymous mutations in the pseudo-receiver domain); weight 10 for all other SNPs and INDEL nucleotides in coding and noncoding regions. A ε-parameter of 20 was used and the MP calculation was run as detailed in the NETWORK 4.6.1.1 User Guide (http://www.fluxus-engineering.com/sharenet.htm). To confirm the *PRR37* gene network obtained with NETWORK 4.6.1.1, we also examined the genealogical relationships among *PRR37* sequences using TCS 2.1 (Clement et al. 2000). Equal weighting was assigned to all polymorphisms except SSRs and rare polymorphisms that were excluded (weight of 0) from the analysis. Networks were exported for graphical editing in Illustrator CS6 (Adobe Systems Inc., San Jose, CA).

Sequence analysis of *Ghd7* alleles

Based on the *Ghd7* gene sequences reported by Murphy et al. (2014), *Ghd7* gene space was sequenced for detection of allelic variants. PCR primers were designed that spanned

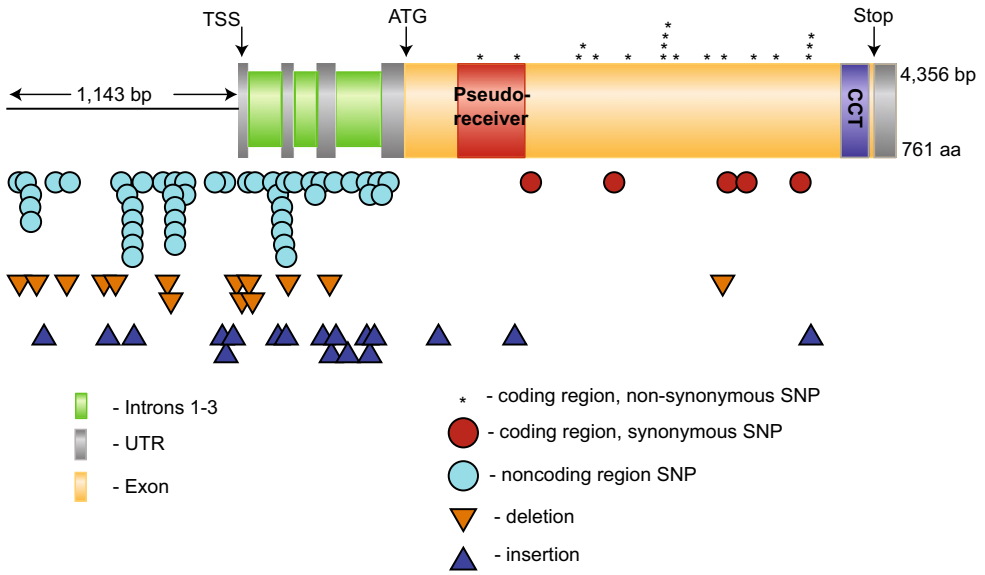

Fig. 1 Mutations in *PRR37* gene sequences. Nucleotide sequences that span the *PRR37* gene space were concatenated including 1143 nucleotides upstream of the transcription start site and 3213 nucleotides from the transcription start site through the 3′ UTR. Exons and sequenced introns are *color-coded*, as are protein functional domains (pseudo-receiver domain, CCT motif). *TSS* refers to the transcription start site, *ATG* refers to the predicted start methionine of the wild-type PRR37 protein

the functional mutation of allele *ghd7-1* (5′TCAGGACAACG ATGACCACCAAGA, 5′ATCAACCTCAAAGGTGAGCC CGTT), and the presence or absence of the functional mutation was determined by sequence analysis. For detection of the *ghd7-2* allele, which harbors an allele-specific intron insertion, allele-specific primers were designed that spanned a portion of the *ghd7-2*-specific intronic sequence (5′TTCTTCTTGCGA TCCTTCTCTC, 5′TGGATTGTGCTTCCAATCTCTA). The presence or absence of the *ghd7-2* allele was determined by fragment analysis of PCR products.

Statement of human and animal rights

This article does not contain any studies with human participants or animals performed by any of the authors.

GenBank sequence submission

The nucleotide sequences of *PRR37* alleles for each sorghum accession listed in Online Supplementary material 1 were registered under GenBank accession numbers JX512468-JX512553.

Results

Loss-of-function photoperiod-insensitive *PRR37* alleles

To search for evidence of human-mediated selection of *PRR37* photoperiod-insensitive alleles during the expansion of sorghum cultivation to temperate climates, we sequenced 4356 nucleotides of the *PRR37* locus in a panel of 48 sorghum historic cultivars and ecotypes (Online Supplementary material 1). The sorghum *PRR37* gene structure includes a pseudo-receiver domain near its N-terminal end and a characteristic CCT-motif at its C-terminus (Murphy et al. 2011) (Fig. 1). By sequencing our sorghum panel, we detected 142 variant nucleotides (Fig. 1), and while the biological importance of most of the sequence variants remains unknown, a series of loss-of-function mutation candidates were identified (Fig. 2). The phenotypic and genetic evidence that each of these proposed loss-of-function *PRR37* alleles confers an earlier-flowering date phenotype under long days is based on extensive genetic evidence of Roy Quinby (1967, 1974) and F. R. Miller (personal communication). The functional mutation candidates within the different *PRR37* alleles share a common biochemical phenotype; each leads to either the truncation of the sorghum PRR37 protein or to the substitution of conserved amino acid residues in the pseudo-receiver domain (Fig. 2a). One particular *PRR37* loss-of-function mutation (p.K184N) alters the conserved DDK-motif that is required for phosphotransfer activity of receiver domain proteins (Farré and Liu 2013). Additionally, in sequencing full-length cDNAs of *PRR37* we detected a unique RNA splice variant isoform in which exon 4 is spliced out of the mature *PRR37* transcript (Fig. 2b). This exon-skipped isoform lacks a portion of the 5′ UTR and 316 nucleotides of the protein-coding region including the predicted N-terminus of

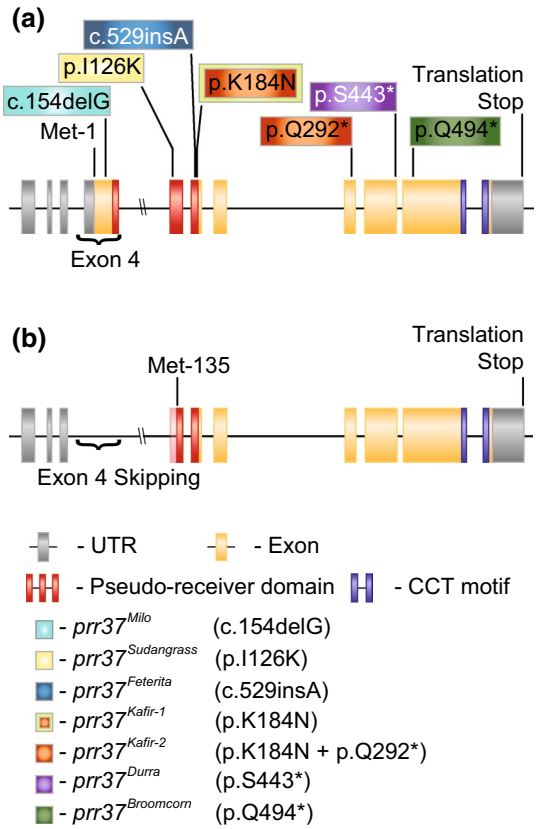

Fig. 2 *PRR37* photoperiod-insensitive alleles. **a** Schematic diagram of the *PRR37* gene with the nature and location of loss-of-function mutational candidates denoted above the gene schematic. Exons are *color-coded* to the corresponding protein functional domain (pseudo-receiver domain, CCT motif) and to the corresponding translated or untranslated region (UTR). Met-1, wild-type start Methionine**. b** Schematic diagram of day-neutral allele $prr37^{White\ African}$ in which exon #4 is spliced out of the mature transcript resulting in the truncation of the pseudo-receiver domain and the elimination of Met-1. Met-135 is a proposed alternative start methionine of the deduced $prr37^{White\ African}$ protein open reading frame (ORF). Descriptions of sequence variants are based on the nomenclature of the Human Genome Variation Society as follows: Prefix c., coding region nucleotide sequence where c.1 is the A of the ATG-translation initiation codon; prefix p., deduced amino acid sequence where p.1 is Methionine-1

the wild-type protein and the first 25 amino acids of the pseudo-receiver domain (see cultivars Della and White African Sorgo in Online Supplementary material 1, 2). Whether this exon-skipped transcript produces a stable, truncated PRR37 protein in vivo remains to be determined, but the exon-skipped isoform was the predominant *PRR37* transcript of several South African cultivars. In total, our survey of these historical (and wild/weedy) sorghums revealed 8 photoperiod-insensitive *PRR37* alleles with each allele being associated with a geographic location(s), sorghum race, or specialized agronomic type (see results below).

Phylogenetic relationship of *PRR37* gene haplotypes

Resequencing of the *PRR37* gene in our panel of historical cultivars, African landraces, and African wild accessions allowed us to organize the *PRR37* sequences into clusters (groups) of gene haplotypes that provide relevant information to the human-mediated selection of photoperiod-insensitive *PRR37* phenotypes during the dispersal of sorghum cultivation to long-day ecosystems. The phylogenetic relationship among the sorghum *PRR37* haplotypes is depicted in median-joining and statistical-parsimony networks with the resulting topologies being nearly identical (Fig. 3 and Online Supplementary material 5). The sorghum *PRR37* haplotype network clustered into a series of 7 branches with 4 branches (haplogroups) comprised of sorghum accessions with photoperiod-insensitive *PRR37* phenotypes, and 3 groups comprised of both photoperiod-insensitive and -sensitive *PRR37* phenotypes. At the center of the haplotype networks are four wild photoperiod-sensitive sorghum accessions (PI 302118, PI 329250, PI 329251, PI 532565) from the northeastern quadrant of Africa from which sorghum originated. Each of these wild sorghum accessions possesses a photoperiod-sensitive (wild type) *PRR37* phenotype. Group A consists of race kafir cultivars that encode the photoperiod-insensitive alleles $prr37^{Kafir-1}$ (functional mutation candidate p.K184N) and $prr37^{Kafir-2}$ (functional mutation candidates p.K184N plus p.Q292*). Also included in Group A are guinea margaritiferum cultivar PI 660565 (IS3620) from West Africa and a wild accession from southern Africa. The guinea margaritiferum and wild accession differ from the kafir haplotype by lacking the *PRR37* functional mutation candidates and thus, exhibit a photoperiod-sensitive *PRR37* phenotype. Group B is a cluster of pasture and forage sorghums (races bicolor and verticilliflorum) collected worldwide, which encode a photoperiod-insensitive $prr37^{Sudangrass}$ allele. This allele differs from the *PRR37* (photoperiod-responsive) allele by a limited number (2 or 3) of polymorphisms that include the functional mutation candidate p.I126K believed to underlie the functional variation of this allele. Group C consists of sorghum accessions of race guinea from West Africa plus races caudatum and durra from the northeastern quadrant of Africa. With the exceptions of the caudatum cultivars collectively referred to as "feteritas" that harbor the photoperiod-insensitive functional mutation candidate c.529insA, the remainder of the cultivars within this group encoded a photoperiod-sensitive *PRR37* allele. Groups D and E each contain a single rare haplotype found in historical cultivars that exhibit a photoperiod-insensitive *PRR37* phenotype. Group D consists of historic cultivar Early White Milo (race durra) and maturity genetic stock SM100, while Group E is comprised of Chinese kaoliangs (race bicolor) and cultivar Gahan (race durra) from Ethiopia. Group F

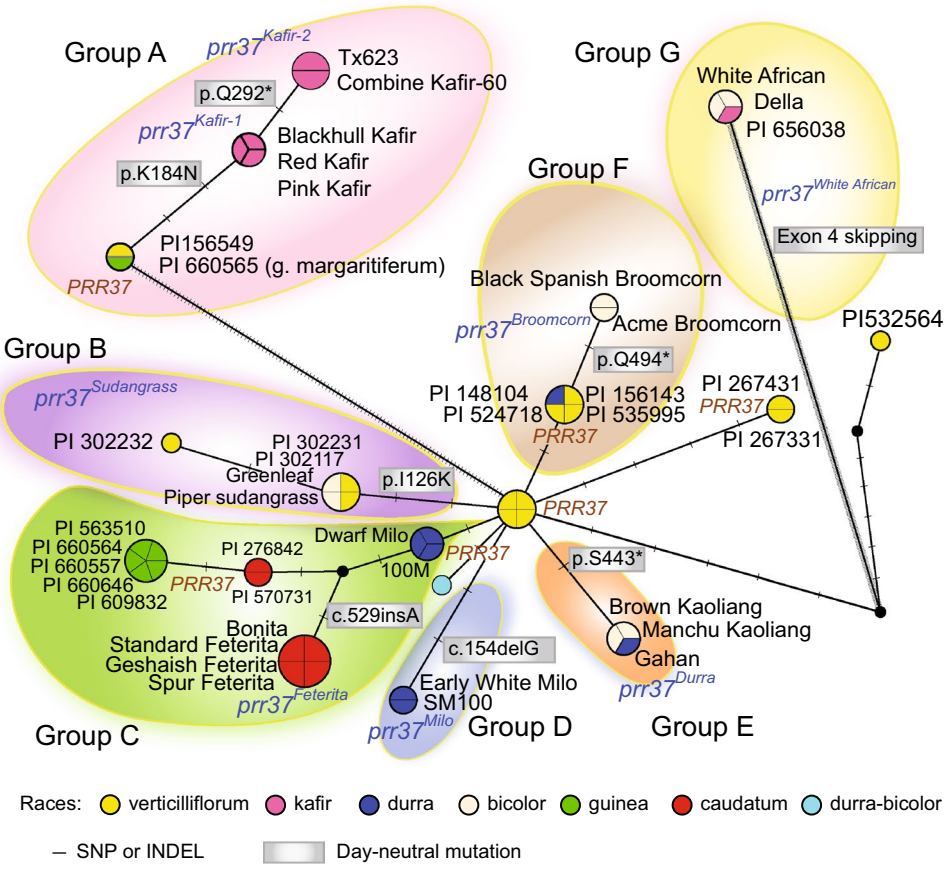

Fig. 3 Neighbor-joining network of *PRR37* alleles. Gene haplotype network analysis was conducted by comparing 4356 nucleotides spanning the *PRR37* gene of historical cultivars, specialized agronomic types, and wild/weedy accessions listed in Online Supplementary material 1. *Circle size* is proportional to the number of accessions sharing the haplotype and *circle color* denotes sorghum racial classification. *Lines* between *circles* represent proposed evolutionary pathways between gene haplotypes reconstructed by the median-joining algorithm. Each *crosshatch* in *line segments* represents a sequence variant between haplotypes; *line segment* lengths are not proportional to the number of sequence differences. Day-neutral mutations are highlighted in steel *gray boxes*, and *shaded areas* highlight groups (haplogroups) of related haplotypes. At the *center of the diagram* is the haplotype of four wild accessions (PI 302118, PI 329250, PI 329251, PI 532565) from sorghum's center of origin in northeastern Africa that was used as the *PRR37* reference sequence

contains several African photoperiod-sensitive (dominant *PRR37*) accessions (races verticilliflorum and durra) and the Chinese broomcorns that encode a photoperiod-insensitive *PRR37* phenotype (functional mutation candidate p.Q494*). *PRR37* haplotype Groups D, E, and several cultivars of F differ from the reference *PRR37* sequence by a limited number of polymorphisms, but in each case those polymorphisms include a loss-of-function mutation that prematurely terminates the PRR37 protein prior to the CCT motif (see Fig. 2). Finally, Group G includes a single haplotype resulting from an exon-skipping *PRR37* transcript isoform that was detected in 3 cultivars including historic sweet sorghum cultivars White African sorgo and Della. There are also several outlier wild/weedy (PI 267331, PI 267431, PI 532564) accessions including a perennial (*S. halepense*) sorghum, and each of the outlier accessions encoded a photoperiod-sensitive *PRR37* allele.

When placed in the context of the historical accounts of the movement of early humans and the spread of sorghum domestication (Kimber 2001; Mann et al. 1983), the haplotypes of the *PRR37* gene provide insight into the selection history and gene flow of this key adaptive gene. Below, we present a series of hypotheses for the origin and gene flow of specific *PRR37* alleles based on allele genealogy and historical accounts of the movement of early humans and the domesticated crops they carried for sustenance and trade (Kimber 2001; Mann et al. 1983).

The $prr37^{Kafir-1}$ day-neutral allele and South African sorghum domestication

In sequencing historical South African cultivars of race kafir, we detected one predominant *PRR37* variant ($prr37^{Kafir-1}$) in kafir cultivars domesticated for grain

production (Fig. 3 and Online Supplementary material 1). This recessive allele encodes a PRR37 protein in which an Asn replaces the Lys of the conserved DDK-motif that is required for phosphotransfer activity of receiver domain proteins (Farré and Liu 2013). The day-neutral allele $prr37^{Kafir-1}$ was shared by the historic grain cultivars Blackhull, Red Kafir, and Pink Kafir; and each of these cultivars trace their geographic origin to the long-day ecosystems of southern Africa (Maunder 2001; Vinall et al. 1936). This $prr37^{Kafir-1}$ allele is highly divergent from the dominant *PRR37* gene encoded by wild accessions from sorghum's center of origin in the northeastern quadrant of Africa (Fig. 3 and Online Supplementary material 1, 2). We resequenced genic regions spanning kafir-specific mutations in a panel of 253 supplemental ecotypes (Online Supplementary material 3), which revealed a high-frequency (0.86) of the $prr37^{Kafir-1}$ allele in race kafir or kafir-caudatum admixtures (results summarized in Table 1). The present allele genealogy study led us to conclude that the $prr37^{Kafir-1}$ allele was possibly selected during a foundation event of race kafir domestication, and this selection likely occurred during the transport of sorghum to the southern latitudes of Africa. A second kafir allele, $prr37^{Kafir-2}$, which differs from the foundational $prr37^{Kafir-1}$ allele by one additional nonsense mutation in the $prr37^{Kafir-1}$ sequence (Fig. 3, Online Supplementary material 1, 2), was traced through the pedigree of a modern USA inbred to the historic cultivar Combine Kafir-60; a foundational cultivar developed for mechanized harvest by mid-twentieth century breeders in the USA (Karper et al. 1951). The absence of the $prr37^{Kafir-2}$-specific mutation (p.Q292*) in our supplemental panel of sorghum ecotypes (Table 1 and Online Supplementary material 3; allele frequency 0.0) led us to propose that this allele was fixed (by drift or human-mediated selection) in a kafir cultivar after introduction from South Africa to the USA in 1876 CE (Maunder 2001).

Genetic link between western and southern African sorghums related to Bantu expansion

In examining the *PRR37* haplotypes of historical South African kafir cultivars and specific landraces of western Africa race guinea, we gained further insight into the genetic link between genotypes in distant geographic locations and the transport of sorghum by early humans. Race kafir was domesticated about ~2000 years BP through the transport of primitive sorghum by Bantu-speaking people migrating from the Guinea Coast of western Africa to isolated geographic regions of eastern and southern Africa (Oliver 1966; Posnansky 1968; Russell et al. 2014). In support of this, the *PRR37* gene haplotype of historic kafir grain cultivars (Blackhull, Red Kafir, and Pink Kafir) showed a deep connection to specific landraces of race guinea from western Africa (see Fig. 3). Specifically, western Africa guineas from the subgroup margaritiferum (Online Supplementary material 3, $n = 27$) shared a common *PRR37* gene haplotype (132 shared variant nucleotides) with southern African grain kafirs, with the notable sequence difference being that the western Africa guinea margaritiferum ecotypes lacked the photoperiod-insensitive mutation candidate p.K184N of the southern Africa $prr37^{kafir-1}$ allele (Fig. 3 and Online Supplementary material 3, subgroup margaritiferum allele frequency 0.7).

Exon-skipped allelic isoform in South African sweet sorghums

Sorghums have been selected throughout history for sweet juicy stems, and these specialized agronomic types are collectively referred to as saccharine or sweet sorghums (Hitchcock 1921; Maunder 2001). In sequencing the *PRR37* transcript of the historic sweet sorghum cultivar Della, a race bicolor cultivar bred in the USA for syrup production (Broadhead and Coleman 1973), we detected the novel exon-skipped isoform displayed in Fig. 2b. This novel *PRR37* transcript is truncated such that the predicted PRR37 protein would lack the N-terminus of the wild-type protein and the first 25 amino acids of the pseudo-receiver domain. Although the pedigree of Della is complex, we were able to trace this novel allele to historic South African cultivar White African, a sweet "Sorgo" cultivar bred in South Africa and subsequently introduced into the USA in 1857 CE (Maunder 2001; Vinall et al. 1936). As our supplemental panel of sorghum ecotypes were composed mostly of grain sorghums with only a limited number of the specialized sweet sorghums ($n = 3$, race bicolor, subgroups dochna, dochna-leoti), the frequency of the $prr37^{White\ African}$ allele was predictably low in our survey, although the allele was detected in another South African cultivar PI 656038, a photoperiod-insensitive grain sorghum with juicy stems. Despite originating in South Africa, the sequence of the $prr37^{White\ African}$ allele is highly divergent from the $prr37^{Kafir-1}$ allele of South African grain cultivars (Fig. 3, Online Supplementary material 1, 2) suggesting that the two alleles may have arisen independently in the same geographical region of Africa. Thus, while domesticating sorghums for grain production or for sweet juicy stems, selection of two unique photoperiod-insensitive *PRR37* allelic variants occurred in the southern latitudes of Africa.

Strategies for photoperiod-insensitive Chinese kaoliangs and broomcorns

The history of sorghum in Asia is as diverse and expansive as Asia itself (He and Bonjean 2010; Kimber 2001), which implies that human-mediated selection for *PRR37* allelic

Table 1 Historical sorghum cultivars and maturity alleles in sorghum races

Sorghum cultivar	Geographic origin	*PRR37* allele	*PRR37* allele freq.	*Ghd7* allele	*Ghd7* allele freq.
Race Kafir	South Africa				
Red Kafir	South Africa	$prr37^{Kafir-1}$	–	*ghd7-1*	–
Pink Kafir	South Africa	$prr37^{Kafir-1}$	–	*Ghd7*	–
Blackhull Kafir	USA selection	$prr37^{Kafir-1}$	–	*ghd7-1*	–
Combine Kafir-60	USA bred	$prr37^{Kafir-2}$	–	*ghd7-1*	–
Kafir other ($n = 14$)		$prr37^{Kafir-1}$ *PRR37*	0.86 0.14	*ghd7-1* *Ghd7*	0.93 0.07
Race Durra	Ethiopia, India				
Gahan	Sudan	$prr37^{Durra}$	–	*ghd7-1*	–
Std. Yellow Milo	Egypt	*PRR37*	–	*ghd7-1*	–
Early White Milo	USA selection	$prr37^{Milo}$	–	*ghd7-1*	–
Dwarf Milo	USA selection	*PRR37*	–	*ghd7-1*	–
Dbl. Dwarf Milo	USA selection	*PRR37*	–	*ghd7-1*	–
Durra other ($n = 84$)		$prr37^{Feterita}$ *PRR37*	0.02 0.98	*ghd7-1* *ghd7-2* *Ghd7*	0.84 0.04 0.12
Race Caudatum	Sudan				
Standard Feterita	Sudan	$prr37^{Feterita}$/*PRR37*	–	*ghd7-1/Ghd7*	–
Spur Feterita	USA selection	$prr37^{Feterita}$	–	*ghd7-1*	–
Bonita	USA bred	$prr37^{Feterita}$	–	*ghd7-2*	–
Geshaish Feterita	Sudan	$prr37^{Feterita}$	–	*ghd7-1*	–
Caudatum other ($n = 67$)		$prr37^{Feterita}$ $prr37^{Kafir-1}$ $prr37^{Broomcorn}$ *PRR37*	0.02 0.01 0.01 0.96	*ghd7-1* *ghd7-2* *Ghd7*	0.06 0.72 0.22
Race Guinea	Western Africa				
PI 660565	Nigeria	*PRR37*	–	*Ghd7*	–
PI 660557	Burkina Faso	*PRR37*	–	*Ghd7*	–
Guinea other ($n = 63$)		*PRR37*	1.00	*ghd7-2* *Ghd7*	0.02 0.98
Races Bicolor and Verticilliflorum	Worldwide				
Piper sudangrass	USA bred	$prr37^{Sudangrass}$	–	*ghd7-1*	–
Greenleaf	USA bred	$prr37^{Sudangrass}$	–	*ghd7-1/Ghd7*	–
Grass types other ($n = 9$)		$prr37^{Sudangrass}$ *PRR37*	0.55 0.44	*ghd7-1* *ghd7-1/ghd7-2* *Ghd7*	0.44 0.22 0.33
White African	South Africa	$prr37^{White\ African}$	–	*ghd7-1*	–
Della	USA bred	$prr37^{White\ African}$	–	*ghd7-1*	–
Wild, weedy ($n = 11$)		*PRR37*	1.00	*ghd7-1/ghd7-2* *Ghd7*	0.18 0.82
Chinese sorghums	Northeast China				
Brown Kaoliang	Beijing	$prr37^{Durra}$		*ghd7-1*	–
Manchu Brown	Manchuria	$prr37^{Durra}$		*ghd7-1*	–
Acme Broomcorn	USA selection	$prr37^{Broomcorn}$		*Ghd7*	–
Black Spanish Broomcorn	USA selection	$prr37^{Broomcorn}$		*Ghd7*	–
Broomcorn other ($n = 9$)		$prr37^{Broomcorn}$ $prr37^{Milo}$ *PRR37*	0.78 0.11 0.11	*ghd7-1* *Ghd7*	0.14 0.86

Allele freq. refers to detection of specific *PRR37* or *Ghd7* alleles in a panel of supplemental (other) accessions listed in Online Supplementary material 3. Allele designations '*PRR37*' and '*Ghd7*' indicate that no photoperiod-insensitive mutations, as reported by Murphy et al. (2011) and herein, were detected during sequence analysis. 'USA selection' refers to early generation direct selections from sorghum foundation introductions, and 'USA bred' are cultivars with more complex pedigrees. Racial classification refers collectively to a race (e.g., caudatum) plus its racial admixtures (e.g., caudatum-bicolor). Alleles separated with a '/' (e.g., $prr37^{Feterita}$/*PRR37*) indicate that this accession is segregating for those alleles

variants may have occurred repeatedly in the long-day ecosystems across the continent. There are several types of historic sorghums uniquely developed in China: the kaoliangs (He and Bonjean 2010; House et al. 2000; Kimber 2001) whose grain is often distilled into a strong liquor known as Shaojiu, and the broomcorns (Berenji et al. 2011; Kimber 2001) with their long, straight seed branches that are used to make brooms (see images, Online Supplementary material 6). In resequencing the *PRR37* gene space in historic kaoliang and broomcorn cultivars, the unique origins of these photoperiod-insensitive Asian sorghums was evident. Historic Chinese kaoliang cultivars Brown and Manchu Brown encoded a *PRR37* variant ($prr37^{Durra}$) with the same functional mutation candidate p.S443* as historic African durra cultivar Gahan (Fig. 3 and Online Supplementary material 1). By contrast, historic broomcorn cultivars Black Spanish and Acme contained a unique photoperiod-insensitive *PRR37* allele ($prr37^{Broomcorn}$) that harbors a novel nonsense functional mutation candidate (p.Q494*) not observed in kaoliangs (Online Supplementary material 3). Due to the limited number of historical broomcorn accessions available in public germplasm collections, we sequenced a limited panel ($n = 9$) of broomcorn cultivars collected from several diverse geographic locales (Table 1 and Online Supplementary material 3). Within this panel, we observed a relatively high frequency of the $prr37^{Broomcorn}$ allele (allele freq. 0.77), and thus the data do support that the $prr37^{Broomcorn}$ allele was selected during the development of sorghum for broom production in the long-day ecosystems of China dating to about 800 years BP.

Forage sorghums encode a novel *PRR37* allele

Sorghum has a multitude of uses including agronomic types specifically bred throughout history for use as a forage or pasture grass. The "sorghum grasses" (Maunder 2001; Quinby 1974; Vinall and Getty 1921), which include sudangrass domesticated in southern Egypt (local name, garawi), became widely cultivated only after being introduced into the long-day ecosystems of the USA (circa 1909 CE) and other temperate regions worldwide (Vinall and Getty 1921). In the historic forage cultivars Greenleaf and Piper sudangrass, we observed a novel *PRR37* allelic variant that harbors a missense mutation in the pseudo-receiver domain in which a conserved hydrophobic residue is mutated to a charged amino acid (p.I126K). This $prr37^{Sudangrass}$ allele was also found in several temperate-zone grown sudangrass cultivars and several closely related native pasture sorghums (Tunis grass, Kamerun grass) that were collected in Australia and the former Soviet Union (Table 1 and Online Supplementary material 3). The discovery of the $prr37^{Sudangrass}$ allele in native pastures ranging from Australia to the former Soviet Union is likely the result of early twentieth century forage sorghum introductions in these temperate regions that eventually escaped cultivation and became naturalized (Vinall and Getty 1921). The $prr37^{Sudangrass}$ allele was present in about half of the grass sorghums we examined (Table 1 and Online Supplementary material 3; allele frequency 0.55, $n = 9$). While bred primarily for use as forage, the selection of forage sorghums with reduced photoperiod sensitivity, which flower late in the growing season, would permit seed propagation of these annual forage sorghums while also permitting dual use for forage and grain (e.g., silage) in temperate climates worldwide.

Loss-of function *PRR37* alleles in equatorial Africa

Given that sorghum is a short-day species with a center of origin near what is present-day Sudan, we expected a low frequency of photoperiod-insensitive *PRR37* allelic variants in cultivars and landraces indigenous to the lower latitudes of Africa (e.g., Sudan, Ethiopia, Gambia), which was in fact the case (Table 1 and Online Supplementary material 3). However, three historic Sudanese cultivars (Standard Feterita, Geshaish Feterita, and Gahan) each encoded a photoperiod-insensitive *PRR37* allelic variant. The $prr37^{Durra}$ allele (functional mutation p.S443*) encoded by cultivar Gahan was not detected in additional African accessions (Table 1 and Online Resource Data 3, frequency 0.00), whereas the $prr37^{Feterita}$ allele (c.529insA) was present in a limited number of supplemental African accessions (3 of 151 race caudatum or durra cultivars). We also discovered this $prr37^{Feterita}$ allele in historic USA cultivars Spur Feterita and Bonita (Table 1); cultivars whose pedigree traces directly to Standard Feterita introduced from Africa to the USA in 1906 CE (Karper and Quinby 1947; Vinall et al. 1936). The number of historic equatorial-region cultivars encoding a recessive photoperiod-insensitive *PRR37* allele was limited and thus may be a consequence of genetic drift. However, the fixation of a photoperiod-insensitive allele in the Sudanese cultivars could also be the result of natural or human-mediated selection (see "Discussion").

Twentieth century plant breeding and global dissemination of a once rare *PRR37* allele

Systematic breeding programs and the intercontinental exchange of germplasm have markedly altered the genetic diversity that exists within domesticated plant and animal species. In the case of sorghum, one temperate-zone breeding program, the TAES-USDA Sorghum Conversion program (Quinby 1974; Rosenow and Dahlberg 2001; Stephens et al. 1967), had a profound impact worldwide on genetic diversity of sorghum with a substantial portion of temperate-zone breeding materials being directly traced to

this program. From about 1965–1990 CE, the TAES-USDA Conversion Program introgressed a recessive *prr37* allelic variant (along with dwarfing genes and possibly other maturity genes) from one elite inbred, Tx406, into about 1000 day-responsive African sorghum accessions to generate short-stature, photoperiod-insensitive cultivars suitable for temperate-zone agriculture (Quinby 1974; Rosenow and Dahlberg 2001; Stephens et al. 1967). In resequencing the *PRR37* gene in cultivar Tx406, we were able to trace its photoperiod-insensitive *PRR37* allelic variant to the historic USA cultivar Early White Milo, which is the founder cultivar of the $prr37^{Milo}$ day-neutral allele (Klein et al. 2008) (Table 1). Since we did not observe the $prr37^{Milo}$ in any of the 253 supplemental accessions (allele frequency 0.00), we propose that the $prr37^{Milo}$ allele was discovered in the late nineteenth century United States, and the allele was subsequently dispersed throughout the USA in the historic cultivar Early White Milo and subsequently worldwide as the photoperiod-adaptive *PRR37* allele in converted sorghum cultivars.

Ghd7 and the human-mediated movement of sorghum to temperate regions

The recent identification of *Ghd7* as an important determinant of flowering time in sorghum (Murphy et al. 2014) afforded the opportunity to address whether natural variants in this gene were selected prior to the expansion of sorghum cultivation from the equatorial region of Africa to temperate climates worldwide. The recent report of Murphy et al. (2014) definitively showed that one of two photoperiod-insensitive *ghd7* alleles was present in each of 17 historic temperate-zone sorghum cultivars examined. The question remained, however, whether the recessive *ghd7* alleles were present at high frequency in equatorial African sorghum cultivars and landraces prior to the dispersal of sorghum cultivation to long-day ecosystems. To address this question, we examined the frequency of the two recessive alleles reported by Murphy et al. (2014), *ghd7-1* and *ghd7-2*, in our panel of 253 supplemental accessions: a germplasm panel that is composed largely of tropical accessions from Sudan, Ethiopia, and western Africa (Table 1 and Online Supplementary material 3). When examining cultivars of race durra (and durra admixtures) from Ethiopia and elsewhere, a large number of accessions encoded the *ghd7-1* allele ($n = 84$, allele frequency = 0.84), and an even greater proportion of durra cultivars encoded either *ghd7-1* or *ghd7-2* photoperiod-insensitive allele (*ghd7-1* and *ghd7-2,* frequency = 0.88). Race caudatum cultivars from the Sudan, Ethiopia and other lower latitudes of Africa showed a relatively high frequency of recessive *ghd7-2* allele ($n = 67$, allele frequency = 0.72) while a limited subset of caudatums encoded the recessive *ghd7-1* allele ($n = 4$, allele frequency = 0.06). By comparison, only one race guinea accession from western Africa encoded either of the recessive *ghd7* alleles ($n = 63$, allele frequency *ghd7-2* = 0.02, *ghd7-1* = 0.00). Consistent with published results (Murphy et al. 2014), nearly all race kafir cultivars examined from South Africa and the USA encoded the photoperiod-insensitive *ghd7-1* allele ($n = 14$, allele frequency = 0.93). Thus, our supplemental panel of tropical sorghum cultivars originating in Africa exhibited a high frequency of recessive *ghd7* alleles, which is in stark contrast to the very low frequency of photoperiod-insensitive *prr37* alleles in these same cultivars (see Table 1).

We extended our examination of the *Ghd7* and *PRR37* alleles to one additional historic cultivar, Standard Yellow Milo (Table 1). Standard Milo (either White or Yellow) is the late-flowering tropical cultivar from which a nineteenth century USA farmer selected a rogue early-flowering plant, and that rogue plant represents the genetic source of the $prr37^{milo}$ photoperiod-insensitive allele present in founder cultivar Early White Milo and many modern temperate cultivars. When the alleles of Standard Yellow Milo at the *PRR37* and *Ghd7* loci were examined, this late-flowering sorghum harbored a wild-type (photoperiod-sensitive) *PRR37* allele, which is agreement with historical accounts (Maunder 2001; Vinall et al. 1936). By contrast, Standard Yellow Milo was recessive at the *Ghd7* locus (*ghd7-1* allele) indicating that the selection of a rogue plant by USA farmers in the late nineteenth century did not involve the selection of a novel photoperiod-insensitive *ghd7* allele, but did involve the selection of the photoperiod-insensitive $prr37^{Milo}$ allele.

Discussion

Placed in the context of historical accounts, we present a model (Fig. 4) of the origin and dispersal of photoperiod-adaptive *PRR37* alleles during the spread of sorghum cultivation to temperate-zone ecosystems across the continents of Africa, Asia, and the Americas. Originating near the Sudanese-Egyptian border (Harlan and Stemler 1976; Kimber 2001; Mann et al. 1983), sorghum is a short-day plant, and mutations in sorghum floral-initiation genes were critical for grain production as early humans migrated or traded with peoples in long-day ecosystems. By examining the allele genealogy of the key adaptive gene *PRR37*, apparent relationships between sorghum ecotypes from distant geographical locations were revealed, and these relationships were consistent with historic records of the movements of early humans. Some of these relationships between sorghums from distant geographic origins were not readily apparent in our whole-genome sequence-based phylogeny (Online Supplementary material 4, 6), which may

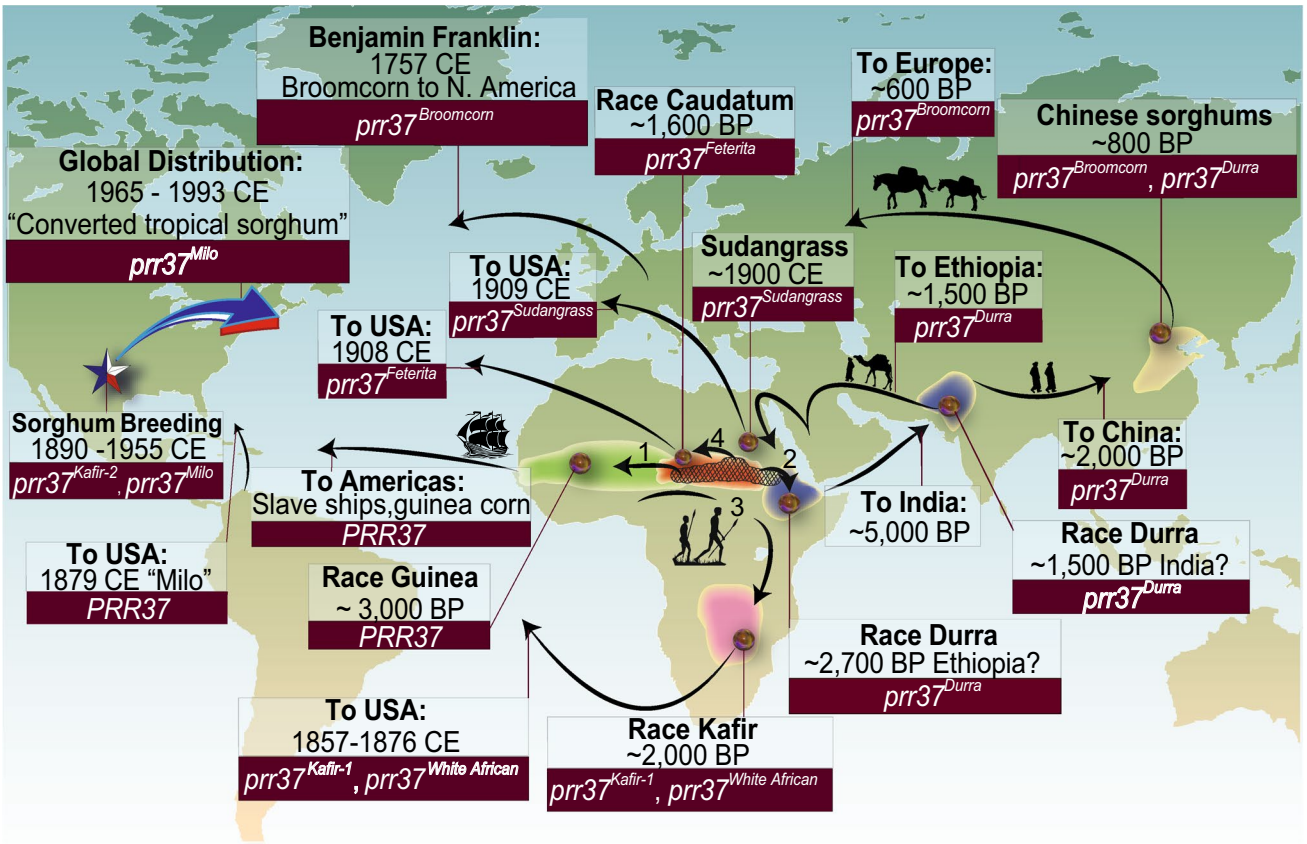

Fig. 4 Origin and global dispersal of adaptive *PRR37* alleles. Movement of specific day-insensitive *PRR37* alleles (*maroon boxes*) is hypothesized based on the sequence of the *PRR37* gene and historical records of movements of humans and the expansion of agriculture. *Arrows* indicate major trajectories of sorghum dispersal and the accepted (or proposed) year associated with the movement. *Hatched area* in present-day Sudan and Ethiopia show the accepted center of origin for sorghum, and *shaded geographic* regions in Africa and Asia mark proposed areas of sorghum racial/cultivar domestication. *Yellow halos* around domestication centers depict wild and weedy sorghums sympatric with cultivated sorghums. *Arrows* within Africa (numbered 1–4) indicate migration of early humans and the dispersal of primitive sorghum from its center of origin as follows: *1* to the savanna of Lake Chad (about 3000 years BP); *2* to the highlands of Ethiopia (about 5000 years BP); *3* Bantu speakers migration from the Guinea Coast to eastern and southern Africa (>2000 years BP); *4* to the lowlands of Ethiopia and Sudan of Chari-Nile speakers (about 1600 years BP). Two proposed centers of origin of race durra and allele $prr37^{Durra}$ are shown: domesticated in Ethiopia (about 2700 years BP) from primitive sorghums; domesticated in India (about 1500 years BP) from primitive sorghums transported from Africa followed by reintroduction of durras to Ethiopia

relate, in part, to genetic recombination and the divergence between the genomes in the geographically isolated regions to which common ancestral sorghums were transported and bred for specific environmental pressures and cultural needs. The resequencing of the gene space of key adaptive genes permits the interpretation of sequence variation within a limited genetic region where hybridization and genetic recombination are less likely to obscure the foundational events of selection of adaptive alleles, and the subsequent dispersal of the adaptive trait by human migration and trade (Gilding et al. 2013; Johanson et al. 2000; Jones et al. 2008; Komatsuda et al. 2007; Lin et al. 2012; Lister et al. 2009; Olsen and Purugganan 2002; Wu et al. 2013). This is particularly true in highly inbred crops such as sorghum, where effective recombination in the gene space is limited, and thus, recent hybridizations are less likely to obscure more historic genetic events. However, discordance between genome-based and allele-based phylogeny are also useful indicators of gene introgression followed by natural or human selection. Thus, care must be exercised in the interpretation of shared allele haplotypes as evidence of a common ancestral gene rather than trait introgression into a divergent genetic background.

Within the continent of Africa and across Asia, *PRR37* allele genealogy revealed relationships between specific ecotypes which serve as evidence of the historical movement of early humans and the selection and dispersal of sorghum that they transported for sustenance or trade. A deep connection existed between the *PRR37* allelic sequence of western and southern African sorghums, which we believe

relates to the transport of primitive sorghum ecotypes by Bantu-speaking people during their slow migration from the Guinea Coast to the long-day ecosystems of southern Africa (Harlan and Stemler 1976; Oliver 1966; Posnansky 1968; Russell et al. 2014). Our results suggest that a common ancestral *PRR37* gene sequence was shared by a progenitor(s) of southern Africa kafir grain sorghums and a subpopulation (guinea margaritiferums) of western Africa ecotypes, and within this ancestral gene sequence, the day-neutral mutation p.K184N was selected and fixed by Bantu people as biologically significant increases in day length in southern Africa mandated photoperiod insensitivity for grain production. This mutation would have been critical to food production for the migrating Bantu as the wild-type (short-day) *PRR37* allele would have delayed or possibly prevented floral initiation in the southern latitudes of Africa. Similarly, as sorghum was transported out of northeastern Africa to northwestern India and subsequently to China along trade or missionary routes (He and Bonjean 2010; Kimber 2001), grain production in the long-day ecosystems would have mandated natural selection and fixation of adaptive photoperiod-insensitive alleles. Our results provide support (albeit limited) for a shared *PRR37* allelic variant by kaoliang cultivars of China and a specific cultivar of race durra: a race of sorghum that was domesticated in Ethiopia or India (Harlan and Stemler 1976). It remains to be determined whether this shared *PRR37* haplotype reflects hybridization followed by fine introgression of the photoperiod-insensitive allele by human agencies or reflects a common ancestry of Chinese kaoliang cultivars and select durra cultivars.

In addition to revealing historical relationships between sorghums, *PRR37* allele genealogy also revealed examples of human-mediated selection for different adaptive alleles within common geographic locations. Originating in northeastern China, historic broomcorn and kaoliang cultivars did not share a common *PRR37* gene haplotype, which supports the assertion of Kimber (2001) that different human agencies in thirteenth-century China began to independently develop broomcorns and kaoliangs from primitive sorghums imported from India. Thus, agriculturists in the long-day ecosystems of northeastern China had developed novel types of sorghums that included kaoliang for grain fermentation and broomcorn with panicles uniquely adapted for broom construction, and in doing so had selected different adaptive *PRR37* allelic variants to initiate and drive the reproductive phase transition. Selection of unique *PRR37* alleles, each conferring a photoperiod-insensitive phenotype, was also evident in historic South African sorghums domesticated for grain and for stem sugar content. The genealogical difference between $prr37^{Kafir-1}$ and $prr37^{White\ African}$ is consistent with selection in the southern latitudes of different adaptive alleles of the same gene, and each confers a photoperiod-adaptive phenotype in ecotypes bred for markedly different human needs.

The discovery of photoperiod-insensitive *PRR37* allelic variants originating in the lower latitudes of the Sudan and Lower Egypt poses questions concerning the genetic forces that may be in play. We discovered three recessive *PRR37* alleles, $prr37^{Feterita}$, $prr37^{Durra}$, and $prr37^{Sudangrass}$, which were present in historic cultivars originating in short-day ecosystems of northeastern Africa. The presence of *PRR37* allelic variants in accessions from what is present-day Sudan and Ethiopia may have been selected as stress-avoidance mechanisms during the "hungry season" (Doggett 1988) where tropical rains begin too late and terminate prematurely prior to the shortening of day lengths that normally signals the end of the rainy season. In these drought-prone years, the early flowering phenotype associated with day-neutral alleles including $prr37^{Feterita}$ and $prr37^{Durra}$ would permit early reproductive transition and seed ripening prior to the premature end of rainfall (Doggett 1988). Thus, selection for early flowering tropical ecotypes as a stress avoidance strategy could account for *PRR37* allelic variants in some equatorial cultivars, although the *PRR37* allelic variants could represent conditionally neutral alleles that drifted to high frequency in a specific genetic background resulting in their fixation in a limited number of cultivars from Africa's lower latitudes.

The introduction of sorghum to the Americas has a unique history that can be initially traced to the slave trade that brought humans and their foods from the Guinea Coast of Africa (Martin 1936; Maunder 2001). However, this initial episode of sorghum cultivation in the United States eventually disappeared, possibly owing to the tropical adaption and late-flowering nature of these first introductions. The first successful introductions of sorghum to the United States were cultivars selected and fixed for adaptive *PRR37* allelic variants in the long-day ecosystems of China and South Africa. Chinese broomcorn ($prr37^{Broomcorn}$ adaptive allele) was introduced in 1757 CE by Benjamin Franklin (Berenji et al. 2011; Martin and Leonard 1949) and is considered the first successful introduction of sorghum to North America. In the mid-nineteenth century, South African sweet sorghums ($prr37^{White\ African}$) were introduced to the USA followed by South African grain cultivars ($prr37^{Kafir-1}$) at the United States Centennial Exposition of 1876 CE (Martin 1936; Maunder 2001; Vinall et al. 1936). All of these historic introductions were of cultivars that were bred for early flowering in long-day ecosystems, and through their introductions to the Americas, the *PRR37* allelic variants they harbored were dispersed to ecosystems far from their geographic origins.

Throughout the first half of the twentieth century, worldwide distribution of elite sorghum cultivars was quite limited owing to the infancy of public breeding programs and

the sorghum seed industry. Thus, sorghums bred in the USA prior to 1965 CE, and the photoperiod-insensitive alleles they encode (e.g., $prr37^{Milo}$, $prr37^{Kafir-2}$), were not dispersed outside the United States. Beginning in 1965 CE, however, this restricted exchange of temperate-adapted sorghums was reversed with the initiation of the TAES-USDA Sorghum Conversion Program (Quinby 1974; Rosenow and Dahlberg 2001; Stephens et al. 1967). The purpose of the TAES-USDA Sorghum Conversion Program was to convert tropical sorghums to short-stature, early-flowering cultivars and to freely distribute 'converted' sorghums to seed companies and public breeding programs worldwide. Prior to this innovated breeding approach, the vast amount of genetic diversity that resided in tropical sorghum accessions was largely inaccessible to temperate-zone sorghum improvement programs. Approximately a thousand diverse tropical accessions, which were too tall and flowered too late for temperate-zone production farmers, were converted by introgressing alleles for dwarfism and day neutrality, and nearly 70 % of these "converted" cultivars harbored the $prr37^{Milo}$ day-neutral haplotype (Thurber et al. 2013). Therefore, while the 'conversion' of tropical ecotypes markedly increased the diversity of temperate-zone sorghum germplasm, introgression of dwarfism and day neutrality from a single donor inbred cultivar also created a genetic bottleneck at the genomic region spanning the *PRR37* locus (Klein et al. 2008; Thurber et al. 2013). Thus, from its discovery in the nineteenth century as a rogue early-flowering plant in the Southern Great Plains, the once rare $prr37^{Milo}$ allele confers day neutrality in a significant percentage of sorghum cultivars and hybrids cultivated in the USA, Australia, South America, Europe, Asia, and Africa.

Sorghum geneticists dating back to the 1960s CE had concluded that the presence of a recessive photoperiod-insensitive allele at maturity locus ma_1 (also known as *PRR37*) was critical for the development of day-neutral sorghum that would produce grain in the long-day environments of higher latitudes. With the recent identification of the *Ghd7* gene as an important determinant of flowering time in sorghum (Murphy et al. 2014), the question has been raised concerning the relative importance of this flowering time gene compared with *PRR37* in the movement of sorghum to temperate climates from its equatorial origin, and to the role of *Ghd7* in the conversion of tropical sorghums to temperate adaptation. While Murphy et al. (2014) clearly showed that both of these genes are critical determinants of flowering time in sorghum, it would appear that early-flowering *PRR37* phenotypes would primarily be under selection during the adaptation to temperate climates of grain sorghums of races durra and caudatum, whereas both *Ghd7* and *PRR37* allelic variants would likely be under selection during the dispersal of guinea accessions to long-day ecosystems. Race kafir cultivars from temperate climates in Africa and elsewhere harbored photoperiod-insensitive *Ghd7* and *PRR37* alleles indicating that both alleles may have been selected during the transport of race kafir progenitors from the Guinea Coast to the southern latitudes of Africa. The results of the present study are in agreement with Thurber et al. (2013) who conducted a retrospective genomic analysis of sorghum adaptation to temperate-zone agriculture and thereby identified specific genomic regions that were critical to the conversion of tropical (photoperiod-sensitive) sorghums to temperate adaptation. Thus, while both *Ghd7* and *PRR37* are critical determinants of flowering time in temperate-zone sorghums, it appears that the need for the introgression of photoperiod-insensitive *Ghd7* and *PRR37* allelic variants to convert tropical sorghums to temperate adaptation is dependent on the race and/or the region of Africa from which the sorghum accession originated.

Over the past several decades, molecular phylogenetic analyses have shown that parallel phenotypic adaptations within and across species have often evolved via parallel genetic changes. The well-studied cases of genetic mechanisms underlying adaptive floral phenotypes (Smith and Rausher 2011; Smith et al. 2012) and pigmentation in animals (Lindgren et al. 2014; Manceau et al. 2010) has shown that evolution is sometimes repeatable and often the same genes are mutated to confer the adaptive phenotype. Despite the complexity of the signaling pathways regulating photoperiodic floral initiation in plants, convergent phenotypic adaptation to day length in the cereals is mirrored by repeated recruitment of *PRR37* gene homologs in the expansion of cultivation to different environments with biologically significant differences in day length. High-yielding hexaploid wheat cultivars of the "green revolution" were insensitive to photoperiod owing to the semi-dominant *Ppd-D1* allele, a *PRR37* gene family member (Beales et al. 2007). Similarly, the expansion of barley across the continent of Europe shows clear latitudinal clines in *PRR37* (*Ppd-H1*) alleles that reflect early patterns of agricultural dispersal to climates with biologically significant changes in photoperiod (Lister et al. 2009; Turner et al. 2005). In rice, natural occurring mutations in *OsPRR37*, in epistatic interaction with *Ghd7*, have contributed to the expansion of rice cultivation to the higher latitudes of Asia (Gao et al. 2014; Koo et al. 2013; Shrestha et al. 2014). With the exception of wheat, the photoperiod-nonresponsive *PRR37* phenotype was conditioned by the premature termination of the *PRR37* protein or by missense mutations in the CCT motif or the pseudo-receiver domain (Koo et al. 2013; Lister et al. 2009; Shrestha et al. 2014; Turner et al. 2005). This similarity in selection of loss-of-function variants in *PRR37* gene homologs across cereals is suggestive of natural and human selection of circadian clock-associated

gene variants that permit diversification of flowering time without marked deleterious pleiotropic effects on fitness (Manceau et al. 2010; Smith and Rausher 2011). The fact that present era plant breeders still exploit these loss-of-function *PRR37* variants in temperate climates indicates that this array of mutations achieve an adaptive phenotype with a high net selection coefficient in multiple genetic backgrounds. While there is a mechanistic convergence at the gene level, there was divergence at the mutational level as unique functional mutations were discovered in *PRR37* allelic variants both within and across cereal species. Thus, while the same gene can be targeted for adaptive change, the precise mutations and their effect on protein function can differ leading to adaptive phenotypes. While it is well documented that natural variation in other genes was also critical to the expansion of cultivation of cereals across broad latitudes, *PRR37* gene homologs play a prominent role in the intricate genetic regulation associated with plant adaptation to changes in day length during the latitudinal expansion of cereal cultivation.

Author contribution statement R.R.K., F.R.M., and P.E.K. conceived and designed the study. R.R.K. and P.E.K. coordinated the different contributions and supervised the study. P.E.K. conducted data curation and bioinformatic analyses associated with whole-genome sequencing. F.R.M. conducted sorghum racial reclassification and cultivar identity analyses of accessions utilized in the study. F.R.M. provided unpublished historical information related to sorghum breeding in the USA and the TAES-USDA Sorghum Conversion Program. D.V.D. assisted in sequencing *PRR37* gene sequences and in graphic illustration of the manuscript. P.J.B. provided unpublished data of the *PRR37* haplotype frequency in sorghum cultivars arising from the TAES-USDA Sorghum Conversion Program and provided purified DNA for supplemental sorghum accessions for estimating *PRR37* allele frequencies. A.M.B. conducted population genetic structure analyses and generated phylogenetic trees of historical sorghum cultivars utilizing genome-wide SNP data. R.R.K. conducted all allele genealogy analyses and wrote the paper.

Acknowledgments We thank the members of the United States Department of Agriculture—Agriculture Research Service's Plant Genetic Resources Conservation Unit, Griffin, GA (USA), for helpful consultation and for provision of seed samples of historical sorghum cultivars, African landraces, and wild and weedy accessions. We thank J. McCollum for expert technical assistance, the entire staff of MMR Genetics LLC for expert field assistance, and Stephanie Sullivan for her graphical design services. Funding was provided by the United States Department of Agriculture—Agriculture Research Service (to R.R.K.), and by MMR Genetics LLC (to F.R.M.).

Conflict of interest The authors declare no conflict of interest.

References

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Beales J, Turner A, Griffiths S, Snape J, Laurie D (2007) A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet 115:721–733

Berenji J, Dahlberg J, Sikora V, Latkovi D (2011) Origin, history, morphology, production, improvement, and utilization of broomcorn [*Sorghum bicolor* (L.) Moench] in Serbia. Econ Bot 65:190–208

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Broadhead DM, Coleman OH (1973) Registration of Dale sweet sorghum. Crop Sci 13:776

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1660

Doggett H (1988) Sorghum, 2nd edn. Wiley-Blackwell, New York

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT*-like gene expression independently of *Hd1*. Gene Dev 18:926–936

Farré EM, Liu T (2013) The PRR family of transcriptional regulators reflects the complexity and evolution of plant circadian clocks. Curr Opin Plant Biol 16:621–629

Gao H, Jin M, Zheng X-M, Chen J, Yuan D, Xin Y, Wang M, Huang D, Zhang Z, Zhou K, Sheng P, Ma J, Ma W, Deng H, Jiang L, Liu S, Wang H, Wu C, Yuan L, Wan J (2014) Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. Proc Natl Acad Sci. doi:10.1073/pnas.1418204111

Gilding EK, Frère CH, Cruickshank A, Rada AK, Prentis PJ, Mudge AM, Mace ES, Jordan DR, Godwin ID (2013) Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. Nature Commun 4:1483. doi:10.1038/ncomms2450

Harlan JR, Stemler ABL (1976) The races of sorghum in Africa. In: Harlan JR, de Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton Press, The Hague, pp 465–478

He Z, Bonjean APA (2010) Cereals in China. CIMMYT, Mexico

Hitchcock AS (1921) Manual of farm grasses. United States Department of Agriculture, Washington, DC

House LR, Gomez M, Murty DS, Sun Y, Verma BN (2000) Development of some agricultural industries in several African and Asian countries. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. John Wiley and Sons, New York, pp 131–190

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290:344–347

Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, Jones G, Jones MK, Brown TA, Powell W (2008) Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol 25:2211–2219

Karper RE, Quinby JR (1947) New varieties of sorghum. Texas Agricultural Experiment Station, College Station

Karper RE, Quinby JR, Kramer NW (1951) New varieties of sorghum. Texas Agricultural Experiment Station, College Station

Kimber CT (2001) Origin of domesticated sorghum and its early diffusion to India and China. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology and production. Wiley and Sons Inc, New York, pp 3–98

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci 48:S12–S26

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci 104:1424–1429

Koo B-H, Yoo S-C, Park J-W, Kwon C-T, Lee B-D, An G, Zhang Z, Li J, Li Z, Paek N-C (2013) Natural variation in *OsPRR37* regulates heading date and contributes to rice cultivation at a wide range of latitudes. Mol Plant 6:1877–1888

Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F, Yu J (2012) Parallel domestication of the *Shattering1* genes in cereals. Nat Genet 44:720–724

Lindgren J, Sjovall P, Carney RM, Uvdal P, Gren JA, Dyke G, Schultz BP, Shawkey MD, Barnes KR, Polcyn MJ (2014) Skin pigmentation provides evidence of convergent melanism in extinct marine reptiles. Nature 506:484–488

Lister DL, Thaw S, Bower MA, Jones H, Charles MP, Jones G, Smith LMJ, Howe CJ, Brown TA, Jones MK (2009) Latitudinal variation in a photoperiod response gene in European barley: insight into the dynamics of agricultural spread from 'historic' specimens. J Archaeol Sci 36:1092–1098

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. Philos Trans R Soc B Biol Sci 365:2439–2450

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Texas Agricultural Experimental Station College Station, pp 107–153

Martin AH (1936) Yearbook of Agriculture. United States Department of Agriculture, Washington, DC, pp 523–623

Martin JA, Leonard WH (1949) Principles of field crop production, 1st edn. The McMillan Company, New York

Maunder AB (2001) History of cultivar development in the United States: from "memoirs of A.B. Maunder-sorghum breeder". In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology and production. Wiley and Sons Inc, New York, pp 191–223

Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 *(PRR37)* controls photoperiodic flowering in sorghum. Proc Natl Acad Sci 108:16469–16474

Murphy RL, Morishige DT, Brady JA, Rooney WL, Yang S, Klein PE, Mullet JE (2014) *Ghd7* (Ma_6) represses sorghum flowering in long days: *Ghd7* alleles enhance biomass accumulation and grain production. Plant Genome. doi:10.3835/plantgenome2013.3811.0040

Oliver R (1966) The problem of the Bantu expansion. J Afr Hist 7:361–376

Olsen KM, Purugganan MD (2002) Molecular evidence on the origin and evolution of glutinous rice. Genetics 162:941–950

Posnansky M (1968) Bantu genesis: archaeological reflections. J Afr His 9:1–11

Quinby JR (1967) The Maturity Genes of Sorghum. In: Norman AG (ed) Adv Agron. Academic Press, New York, pp 267–305

Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioprod Bior 1:147–157

Rosenow DT, Dahlberg JA (2001) Collection, conversion, and utilization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology and production. Wiley and Sons Inc, New York, pp 309–328

Russell T, Silva F, Steele J (2014) Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. PLoS One 9:e87854

Shrestha R, Gómez-Ariza J, Brambilla V, Fornara F (2014) Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals. Ann Bot. doi:10.1093/aob/mcu032

Smith SD, Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. Mol Biol Evol 28:2799–2810

Smith SD, Wang S, Rausher MD (2012) Functional Evolution of an anthocyanin pathway enzyme during a flower color transition. Mol Biol Evol 30:602–612

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Thurber C, Ma J, Higgins R, Brown P (2013) Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol 14:R16

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. Science 310:1031–1034

Vinall HN, Getty RE (1921) Sudan grass and related plants. US Department of Agriculture, Washington, DC

Vinall HN, Stephens JC, Martin JH (1936) Identification, history, and distribution of common sorghum varieties. United States Department of Agriculture, Washington, DC

Wendorf F, Close AE, Schild R, Wasylikowa K, Housley RA, Harlan JR, Krolik H (1992) Saharan exploitation of plants 8000 years BP. Nature 359:721–724

Wu W, Zheng X-M, Lu G, Zhong Z, Gao H, Chen L, Wu C, Wang H-J, Wang Q, Zhou K, Wang J-L, Wu F, Zhang X, Guo X, Cheng Z, Lei C, Lin Q, Jiang L, Wang H, Ge S, Wan J (2013) Association of functional nucleotide polymorphisms at *DTH2* with the northward expansion of rice cultivation in Asia. Proc Natl Acad Sci 110:2775–2780