Whole Genome Sequencing Reveals Potential New Targets for Improving Nitrogen Uptake and Utilization in *Sorghum bicolor*

Karen Massel[1]*, Bradley C. Campbell[1], Emma S. Mace[2], Shuaishuai Tai[3], Yongfu Tao[4], Belinda G. Worland[1], David R. Jordan[4], Jose R. Botella[1] and Ian D. Godwin[1]

[1] School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD, Australia, [2] Department of Agriculture and Fisheries, Warwick, QLD, Australia, [3] BGI-Shenzhen, Shenzhen, China, [4] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia

Nitrogen (N) fertilizers are a major agricultural input where more than 100 million tons are supplied annually. Cereals are particularly inefficient at soil N uptake, where the unrecovered nitrogen causes serious environmental damage. *Sorghum bicolor* (sorghum) is an important cereal crop, particularly in resource-poor semi-arid regions, and is known to have a high NUE in comparison to other major cereals under limited N conditions. This study provides the first assessment of genetic diversity and signatures of selection across 230 fully sequenced genes putatively involved in the uptake and utilization of N from a diverse panel of sorghum lines. This comprehensive analysis reveals an overall reduction in diversity as a result of domestication and a total of 128 genes displaying signatures of purifying selection, thereby revealing possible gene targets to improve NUE in sorghum and cereals alike. A number of key genes appear to have been involved in selective sweeps, reducing their sequence diversity. The ammonium transporter (AMT) genes generally had low allelic diversity, whereas a substantial number of nitrate/peptide transporter 1 (NRT1/PTR) genes had higher nucleotide diversity in domesticated germplasm. Interestingly, members of the distinct race *Guinea margaritiferum* contained a number of unique alleles, and along with the wild sorghum species, represent a rich resource of new variation for plant improvement of NUE in sorghum.

Keywords: nitrogen uptake and utilization, nitrogen use efficiency (NUE), sorghum (*Sorghum bicolor*), domestication, selection

INTRODUCTION

Quantitatively, Nitrogen (N) is the most important nutrient required for plant growth and development and is often supplied in the form of synthetic N fertilizers. Their intensive commercial use and generation from fossil fuels, means N fertilizers have a high economic volatility and are often the most expensive additive used in achieving high yielding crops. Application can lead to a variety of environmental problems such as leaching into the waterways contributing to soil and water contamination, as well as the formation of nitrous oxide which plays a role in long-lived greenhouse gas accumulation (Vitousek et al., 1997; Lewandowski and Schmidt, 2006).

One-hundred million tons of N fertilizers are used worldwide each year, 50% of which is applied to the three major cereals (Ladha et al., 2016). Cereals are inherently inefficient at N uptake resulting in only ∼40% of total N applied being harvested in the grain (Peoples et al., 1995; Raun and Johnson, 1999; Sylvester-Bradley and Kindred, 2009). By the year 2050 world population is projected to reach nine billion (Tester and Langridge, 2010), pressuring farmers and breeders to increase commercial crop yields to meet the demand. Population growth is most accentuated in developing countries, including Sub-Saharan Africa, where there are limited resources for agriculture and input-intensive crops are already displaying depleting yields (Cardinale, 2012).

Nitrogen Use Efficiency (NUE) is a broad term incorporating various traits that can be targeted for improvement, such as N uptake or utilization. In cereals, NUE ultimately refers to maximizing grain N yield per unit of soil N (Good et al., 2004). It is often categorized into two subclasses: N Recovery Efficiencies (NRE) referring to the crops' ability to recover N from soil and N Internal Efficiencies (NIE) which refers to the crops' ability to efficiently assimilate and remobilize N. Many phenotypic traits can be manipulated, such as root morphology, total biomass and tiller number to affect NUE (Garnett et al., 2009; Van Oosterom et al., 2010a,b; Olson et al., 2013), however the efficiency of cellular pathways involved in N uptake and utilization also play an important role. For example, elite maize varieties often display enhanced utilization of N; but nonetheless, N uptake has remained constant throughout domestication, suggesting conventional breeding programs have reached NRE capacity (Moose and Below, 2009).

Nitrogen is most commonly available in soils in the inorganic forms of ammonium and nitrate, although amino acids and proteins can also be taken up (Paungfoo-Lonhienne et al., 2008; Näsholm et al., 2009). Uptake of ammonium and nitrate is a tightly regulated system facilitated by both constitutive and inducible molecular transporters that mediate their translocation into the root cells (Krapp et al., 2014). AMT2 and NRT1/PTR are low-affinity transporters which typically function in the presence of high levels of either ammonium or nitrate respectively (Tsay et al., 1993; Sohlenkamp et al., 2000; Neuhäuser et al., 2007; Yuan et al., 2007; Wang et al., 2012). Conversely, AMT1 and NRT2, with assistance from the accessory protein NRT3, function as high-affinity transporters which are most efficient under low concentrations (Masclaux-Daubresse et al., 2010; Zhou et al., 2015). NRT1/PTRs are a multigene family including several solute transporters with affinities for glucosinolates, di- and tri-peptides, as well as plant hormones (Léran et al., 2014; von Wittgenstein et al., 2014; Chiba et al., 2015).

Nitrate is converted into ammonium through enzymatic reactions carried out by Nitrate Reductase (NR) and Nitrite Reductase (NiR). Assimilation of ammonium depends on Glutamine Synthetase (GS), the gateway enzyme responsible for converting inorganic N into the organic amino acid glutamine, utilizing ATP. GS has two isoforms with distinct physiological functions. In higher plants, GS2 is typically encoded within a single gene where it functions in primary N assimilation within plastids. Alternatively, GS1 is encoded by a large multigene family

where cytosolic GS1 is predominantly involved in ammonium recycling (Bernard and Habash, 2009). Glutamine is an important signaling molecule providing a negative feedback loop to reduce nitrate uptake through down-regulation of transporter genes as well as driving post-translational modifications (Miller et al., 2007). Glutamine is converted into two glutamate molecules by the enzyme Glutamine Oxoglutarate Aminotransferase (GOGAT) with the reducing power varies depending on isoform: NADH for cytosolic, or ferredoxin (Fd) for plastidic (Suzuki and Knaff, 2005). Within the mitochondria, Glutamate Dehydrogenase (GDH) balances relative levels of ammonium and glutamate (Suzuki and Knaff, 2005; Purnell and Botella, 2007).

Ultimately, glutamate is incorporated into aspartate and asparagine by reactions catalyzed by Aspartate Aminotransferase (AST) and Asparagine Synthetase (AS). Asparagine has the highest N:C ratio, making it a useful vehicle for long-range N transport and storage. These reactions are reversible, where Asparaginase (ASPG) converts asparagine to aspartate and Aspartate Ammonia Lyase (AAL) which converts aspartate to fumarate, each step releasing an ammonium ion. The four major amino acids, glutamine, glutamate, aspartate, and asparagine are typically the most abundant amino acids in plant roots and leaves (Lam et al., 1995; Paynel et al., 2001) where they are transported to sink tissues to fulfill their essential roles in various biosynthetic pathways (Forde and Lea, 2007; Pratelli and Pilot, 2014; Gaufichon et al., 2015).

Attempts to improve cereal NUE using breeding programs and genetic modification have mostly been unsuccessful due to the complexity of regulatory controls and a lack of accurate phenotyping (Garnett et al., 2015). Improvement of uptake efficiencies are difficult to achieve since uptake is dependent upon plant demand and is regulated through downstream assimilates where the mechanisms of regulation are poorly understood in cereals (Malagoli et al., 2005; Garnett et al., 2009, 2013). Overexpression of key enzymes such as NR, NiR, GS, and GOGAT has led to inconsistent results that often do not facilitate an improvement of NUE or phenotypic change (Yamaya et al., 2002; Good et al., 2004; Hirel et al., 2007; Masclaux-Daubresse et al., 2010). However, one example in maize involves two characterized GS1 genes involved in kernel number and size (Martin et al., 2006; Cañas et al., 2010) where constitutive overexpression of ZmGln1;3 increased kernel number (Martin et al., 2006). However, this was not reliably reproduced under field conditions (Thomsen et al., 2014). Constitutive expression of GDH from *Escherichia coli* and *Aspergillus* in tobacco and rice respectively resulted in biomass and seed size increases (Ameziane et al., 2000; Abiko et al., 2010). However, when two GDH genes from *Nicotiana plumbaginifolia* were expressed in tobacco, overall biomass was decreased (Terce-Laforgue et al., 2013). These inconsistent results demonstrate the lack of understanding of the intricate functions and regulation of key enzymes and their isoforms involved in the N assimilatory pathways.

Sorghum is an ideal candidate to study traits relating to NUE due to its higher grain yield per unit of N available from the

soil during N limiting conditions, especially when compared to other major cereal crops (Lemaire et al., 1996; Hirel et al., 2007; Oikeh et al., 2007; Zhu et al., 2010; Vogan and Sage, 2011). Sorghum is also an important staple food for 500 million people in Asia and sub-Saharan Africa (Anglani, 1998). Its ability to tolerate periods of drought, high temperatures, and low inputs has made it a preferred cereal to grow in semi-arid climates. Domestication of sorghum occurred ~8000 years ago in the Eastern regions of Africa (Wendorf et al., 1992) and has given rise to the landraces examined within this study. Molecular evidence suggests a secondary domestication event occurred in Western Africa, where *Guinea margaritiferum* races originated (Mace et al., 2013). These sorghum accessions are morphologically distinct and cultivated in nutrient deficient soils (Mace et al., 2013; Morris et al., 2013). Studying population genetics throughout domestication and cultivation of sorghum will facilitate a clearer understanding of key genes facilitating NUE, as well as identifying genes maintaining allelic variation within a population, possibly as an adaption to variable soil environments.

Numerous traits have been selected for throughout domestication and improvement, the most obvious being grain number and size (Campbell et al., 2016), and lack of seed shattering. Other traits selected for throughout domestication and improvement include rounded seeds, photoperiod insensitivity, disease resistance, and reduction of seed tannins, however, depending on the end-use of the grains these features will vary throughout accessions. Landrace and wild and weedy sorghums have a larger overall biomass, mainly attributable to their height. Although assumed to require high available soil N, it is suggested that elite breeding lines have enhanced NIE as they typically remobilize 70–90% of plant N to the grain (Olson et al., 2013), and can continue to uptake significant levels of N post-anthesis (Sinclair et al., 1997; Kamoshita et al., 1998, 1999; Borrell and Hammer, 2000; Van Oosterom et al., 2010b). Thus, there is an expectation that key transporter genes involved in N storage and remobilization are more likely to be under strong selection within improved accessions, while genes involved in primary N uptake from the soil, especially under limited conditions, are more likely to be under selection in wild and weedy accessions.

This study provides the first broad population genetic analysis of the overall trends in the selection of genes involved in N uptake and utilization pathways in sorghum as viewed through the prism of domestication. Here we have identified key candidate genes for further targeted research with the ultimate aim of breeding cereal crops with improved NUE.

MATERIALS AND METHODS

Whole Genome Sequencing (WGS) and Plant Material

Resequencing data was previously described in Mace et al. (2013) wherein a total of 44 *Sorghum bicolor* genotypes underwent WGS by Illumina paired-end sequencing with ~22x coverage per line (16x–45x). The sequencing had a SNP calling accuracy of 99.85 and 99.72%, validated through targeted sequencing and whole-genome *de novo* assembly of representative lines. The accessions included are genetically representative of 18 Landrace and seven Wild and Weedy sorghums along with two *G. margaritiferums* and two progenitors *S. propinquum* genotypes (**Table S1**). Additionally, this analysis includes 19 inbred lines which are the product of various degrees of selective breeding.
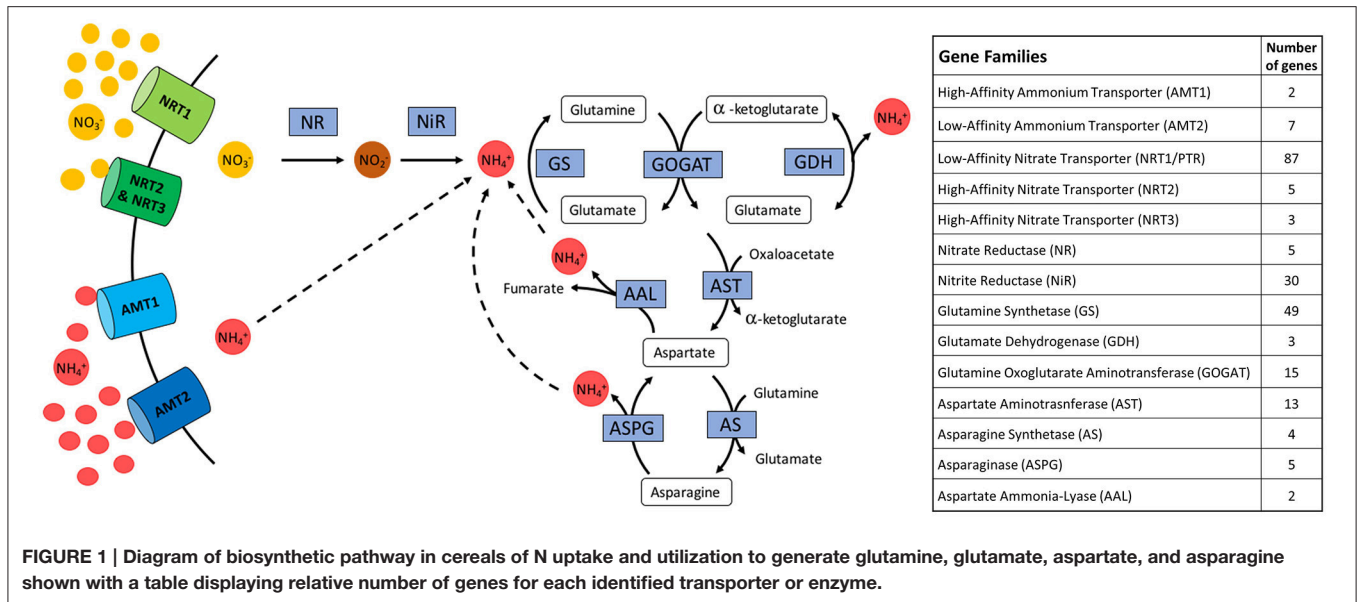
Identification of Genes Involved in N Uptake and Utilization

Nitrogen uptake commences with the recovery of nitrate and ammonium ions from the soil, followed by the rapid assimilation of these inorganic ions into organic forms (**Figure 1**). For the purposes of this study, the end point of the metabolic pathway was set at the generation and cycling of aspartate and asparagine, which facilitates the incorporation of N into four major amino acids: glutamine, glutamate, asparagine, and aspartate. Genes were grouped based on their gene family, where analysis was focused on the five transporter protein families, NRT1/PTR, NRT2, NRT3, AMT1, and AMT2; genes involved in N assimilation by two reductases NR and NiR, and conversions into glutamine and glutamate by GS, GOGAT, and GDH; and also within the aspartate and asparagine cycling involving AST, AS, ASPG, and AAL.

Identification of enzymes involved in N uptake and utilization were retrieved from the SorghumCyc (Tello-Ruiz et al., 2016) and Kyoto Encyclopedia of Genes and Genomes (KEGG) online sorghum pathway database (Kanehisa et al., 2012). Ammonium and nitrate uptake and transporter genes described by von Wittgenstein et al. (2014) and Léran et al. (2014) were added for analysis. All gene candidates were then filtered for expression utilizing the publicly available Morokoshi sorghum transcriptome database (Makita et al., 2015). Only those genes found to be expressed were used for analysis in this study, irrespective of expression levels. Of note, this database also includes expression data from sorghum exposed to N-stress, highlighting upregulation of genes involved in N stress response, however this analysis was limited to roots (Gelli et al., 2014). A total of 230 expressed genes were identified for potential involvement in the N biosynthetic pathway. Orthologous genes under selection identified via WGS studies of maize (Hufford et al., 2012; Jiao et al., 2012) and rice (He et al., 2011; Huang et al., 2012; Xu et al., 2011) were also compared to selected sorghum genes (**Table S2**). Genes within NRT1/PTR family were identified based on their homology with characterized Arabidopsis NTR1/PTR genes, as well as the unified nomenclature which divides these transporters into subfamilies and is denoted with a NPF number (**Table S3**; Léran et al., 2014; von Wittgenstein et al., 2014).

Selection Assessments

Population genetic assessments were performed on all identified genes to establish the effects of domestication and improvement upon genetic diversity, as well as determining if any genes displayed signatures of selection. To probe the direction and severity of selection, coding regions were analyzed using

FIGURE 1 | Diagram of biosynthetic pathway in cereals of N uptake and utilization to generate glutamine, glutamate, aspartate, and asparagine shown with a table displaying relative number of genes for each identified transporter or enzyme.

K_a:K_s ratios which were computed using the software package Calculator 1.2 (MYN Method) following the protocol by Zhang et al. (2006). The K_a:K_s ratio is the frequency of non-synonymous substitutions per non-synonymous site (K_a) to the frequency of synonymous substitutions per synonymous site (K_s). The ratio differentiates between positive selection if results are greater than one, purifying if less than one, and neutral if equal to one.

Analysis was performed at both whole gene and coding sequence (CDS) level for all 230 genes by evaluating nucleotide diversity ($\theta\pi$; Nei, 1987), Watterson's estimator ($\theta\omega$; Watterson, 1975), and neutrality test *Tajima's D* (TajD; Tajima, 1989) using BioPerl modules with an in-house script (Mace et al., 2013). Further, F_{ST}-values were calculated using an alternative BioPerl module with the same genetic components to estimate population differentiation. Signatures of purifying and balancing selection were identified utilizing the population genetic data mirroring the criteria outlined in Mace et al. (2013) and Campbell et al. (2016). Specifically, the thresholds applied to identify signatures of purifying selection were as follows; when $\theta\pi$- and $\theta\omega$-values for a gene were in the lower 5% of the empirical distribution in the descendent population, F_{ST}-values > 95% of the empirical distribution, and with a negative TajD. In contrast, the thresholds applied to identify signatures of balancing selection were as follows; when $\theta\pi$- and $\theta\omega$-values were in the upper 25% of the empirical distribution in the descendent population, TajD in the upper 5% of the empirical distribution, and F_{ST} > 10% but lower than <90% of the population pair wise distribution.

Validation of genes under selection was performed using the mlHKA test (Wright and Charlesworth, 2004). Specifically, genes identified as being under purifying and balancing selection throughout domestication were compared against 34 neutral sorghum genes. The mlHKA test was run under a neutral model, the selection model was performed where numselectedloci = 0 and numselectedloci > 0. Mean log likelihood ratio test statistics were used to determine significance following a χ^2 distribution where df equals the number of different parameters.

Genes were also analyzed for selection at the base pair level using the R package PopGenome (Pfeifer et al., 2014). Pairwise comparisons between ancestor/descendant populations were used to determine selection. To identify specific base pairs under purifying selection, the criteria utilized a reduction in diversity between pairwise population comparisons greater for the specific base pair than the mean gene diversity in the whole genome, an F_{ST} > 0, and a negative TajD-value. To identify balancing selection, the criteria included an overall increase in diversity between pairwise population comparisons, greater for the specific base pair than the mean gene diversity in the whole genome, F_{ST}-values greater than zero, and a positive TajD-value.

Haplotype Network Analysis of Genes under Selection

Haplotype network analysis was performed using the package PEGAS (Population and Evolutionary Genetics Analysis System; Paradis, 2010) in R and integrated using the ape (Paradis et al., 2004) and adegenet packages (Jombart, 2008) for genes under selection. In addition to the sorghum accessions that underwent population genetic assessments, with some exclusions, accessions from the *G. margaritiferum* race and *S. propinquum* were used in haplotype network analyses (**Table S1**), totaling 45 sorghum genotypes. This analysis was performed to validate selection and to establish how domestication events have shaped these haplotypes, with a particular focus on *G. margaritiferum* accessions which have resulted from a secondary domestication event (Mace et al., 2013). Haplotype network analysis were performed at both gene and CDS levels.

RESULTS

N Uptake and Utilization Pathway and Sequence Diversity

A combination of online databases and gene identification based on homology identified a total of 290 putative genes. Of these only those with published expression data were further analyzed (Makita et al., 2015). This resulted in a total of 230 genes potentially involved in the N uptake and utilization pathways (**Figure 1**). This pathway begins with uptake of either ammonium or nitrate through the transporter genes AMT1 (2 genes), AMT2 (7 genes), NRT1/PTR (87 genes; **Table S3**), NRT2 (5 genes), and NRT3 (3 genes). Nitrate requires assimilation into ammonium through the enzymatic roles of NR (5 genes) and NiR (30 genes), ammonium can then be converted into organic amino acid glutamine by GS (49 genes), and further into glutamate by GOGAT (15 genes) where GDH (3 genes) balances levels of glutamate and ammonium. Glutamate can be incorporated into the aspartate-asparagine shuttle which involves AST (13 genes), AS (4 genes), ASPG (5 genes), and AAL (2 genes). A subset of these enzymes require additional cofactors in order to function, such as the reducing agents ferredoxin (Fd) or NADH, as well as chemical energy from ATP.

Of the 230 transcribed genes identified as putatively playing a role in N uptake and utilization, 79.5% (183 genes) displayed an average decrease in nucleotide diversity after domestication (**Table S4**). The examination of average nucleotide diversity within gene families revealed that there were reductions in sequence diversity in all the gene families in the early steps of the N pathway, up to the stage where glutamate is generated. Conversely, three gene families in the aspartate-asparagine cycling, specifically AS, ASPG, and AAL, displayed an increase

in diversity (**Figure 2**). This disparity may be caused by the small number of genes within the families in the later stages of the pathway resulting in outliers having sizeable effects on the averages. Of the five ASPG genes, one gene (*Sobic.006G169500*) had a maximum of 32.7-fold increase of diversity ($\theta\pi_{Landrace}$ = 0.007699, $\theta\pi_{Wild\,\&\,Weedy}$ = 0.0002358) in Landraces compared to the Wild and Weedy genotypes.

Amongst the 230 genes studied, eight genes displayed invariance at the nucleotide level throughout domestication (**Table S5**). Of these, AMT2 (*Sobic.003G400300*), a low affinity ammonium transporter gene, was invariant within all 44 sorghum accessions analyzed. Of the invariant genes, two were also determined to have significantly high F_{ST}-values (F_{ST} > 0.5): a NiR gene (*Sobic.001G422300*) with $F_{ST:Domestication}$ = 0.695 and an AST gene (*Sobic.002G413100*) with $F_{ST:Domestication}$ = 0.593. Both also displayed signatures of purifying selection. The most predominant gene families displaying invariance were NRT1/PTR with three genes and AST with two genes.

Signatures of Selection

The K_a:K_s ratio is a tool used to determine the direction and magnitude of selection. The majority of genes in all families showed a trend toward purifying selection with a K_a:K_s ratio of less than one (**Figure S1**). GS genes had the highest K_a:K_s ratio, while the ammonium transporters (AMT1 and AMT2) were low in all the sorghum genotype groups. Of the genes with a K_a:K_s ratio less than one, 56 out of 146 genes in Landrace, and 77 of the 169 in Wild and Weedy groups were statistically significant ($p < 0.05$; data not shown). Average K_a:K_s ratios throughout the N uptake and utilization pathway were K_a:$K_{s:Landrace}$ = 0.245 and K_a:$K_{s:Wild\,\&\,Weedy}$ = 0.253.
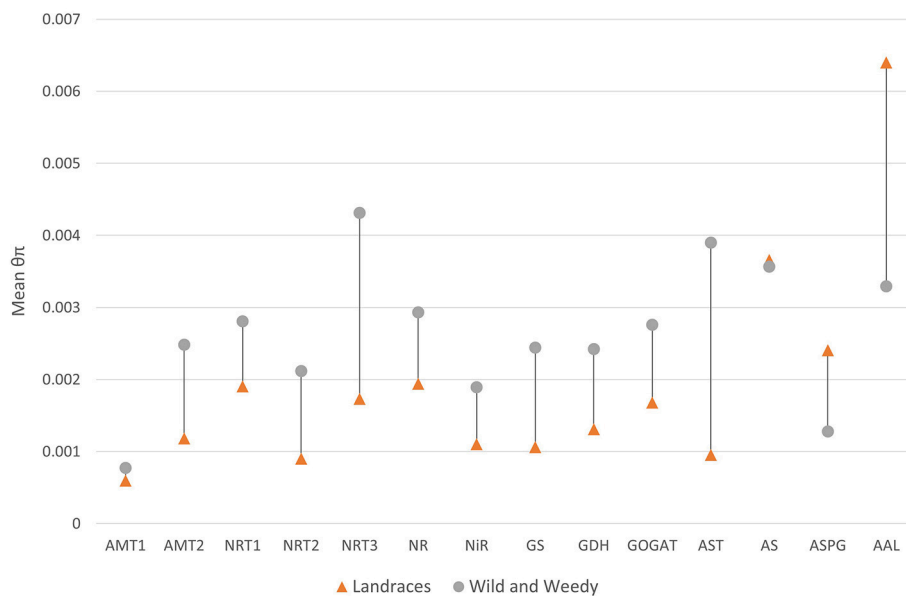
FIGURE 2 | The mean nucleotide diversity ($\theta\pi$) for each gene family in Landrace and Wild and Weedy genotypes. Analysis is performed on the CDS of each gene and averaged within each gene family within the nitrogen uptake and utilization pathway.

Sixteen genes were found to display signatures of positive selection by having K_a:K_s-values greater than one in at least one genotype grouping and/or when comparing these groupings through domestication (**Table 1**). The gene families enriched with the high K_a:K_s ratio were NRT1/PTR (9 genes), GS (5) genes, and NIR (2) genes.

Three genes were identified with signatures of purifying selection during domestication based on their sequence diversity ($\theta\pi$), allele frequencies (TajD), and population differentiation (F_{ST}; **Table 2**). All three of these genes were located in regions of the genome with selective sweeps, where adjacent genes were also determined to have signatures of purifying selection (**Table 1**). The three genes are NiR (*Sobic.001G42230*), AST (*Sobic.002G413100*), and GS (*Sobic.007G078500*), two of which (NiR and AST) are invariant throughout Landrace genotypes. All three genes were also shown to have high F_{ST}-values: the NiR gene (*Sobic.001G422300*) with values of $F_{ST:Domestication} = 0.695$, the AST gene (*Sobic.002G413100*) with $F_{ST:Domestication} = 0.593$, and the GS genes *Sobic.007G078500* with $F_{ST:Domestication} = 0.644$.

Two genes within the N uptake and utilization pathway exhibited evidence of balancing selection throughout domestication (**Table 2**). The ASPG gene (*Sobic.006G169500*) was identified with a signature of balancing selection throughout domestication with a substantial increase in nucleotide diversity of over 32-fold between the Landrace and Wild and Weedy genotypes ($\theta\pi_{Landrace} - 0.007699$, $\theta\pi_{Wild \& Weedy} - 0.000255$). Both genes, the ASPG (*Sobic.006G169500*) and GOGAT (*Sobic.009G225700*) showed contrasting *Tajima's D*-values: the GOGAT with $TajD_{Landrace} =$ 3.008 and $TajD_{Wild \& Weedy} = -1.248$, and the ASPG gene with $TajD_{Landrace} = 2.002$ $TajD_{Wild \& Weedy} = -1.313$.

Genes under selection throughout either domestication, or during subsequent breeding were confirmed using the mlHKA test of the genes under selection (**Table S6**). The N uptake and utilization pathway was found to display significantly fewer genes with signatures of purifying selection during domestication than expected ($p = 0.0281$), and also significantly fewer genes with signatures of balancing selection during improvement than expected ($p = 0.0371$). Chi-squared tests were also performed on the individual population genetics metrics for the genes in the N uptake and utilization pathways in comparison to the whole genome analysis (**Table S7**). Nitrate uptake transporter genes (NRT1/PTR, NRT2, and NRT3) within the Wild and Weedy genotypes had significantly fewer genes within the 5% tail of the empirical distribution of TajD-values in comparison to the whole genome.

Higher resolution analysis of CDS sequence at the base pair level of sorghum accessions identified 135 genes with signatures of purifying selection and 108 genes with signatures of balancing selection. Further examination revealed the majority are candidate domestication genes, where 128 genes presented signatures of purifying selection and 97 presented signatures of balancing selection (**Tables S8, S9**). The number of codons identified with signatures of purifying selection ranged from 1 to 21 in a single gene, with 14 genes containing at least 10 codons under selection. Two genes from the NRT1/PTR family contained the highest number of codons with signatures of selection: *Sobic.001G444000* (NPF8.11) with 21 codons under selection, of which nine were non-synonymous; and

TABLE 1 | List of genes that display pressures of positive selection within either Landrace or Wild and Weedy genotypes, and/or when comparing them throughout domestication.

Gene ID	Function	K_a:K_s ratio			Nucleotide Diversity ($\theta\pi$)	
		Landraces	Wild & Weedy	Domestication	Landraces	Wild & Weedy
Sobic.001G111100	NRT1/PTR	1.574990	1.977870	NA	0.002041	0.003005
Sobic.001G276300	NRT1/PTR	0.336859	0.262895	0.49050	0.006815	0.006517
Sobic.001G276600	NRT1/PTR	0.226060	0.205998	1.39857	0.003121	0.004348
Sobic.003G372000	NRT1/PTR	NA	0.779389	1.91974	0.001054	0.003528
Sobic.005G107500	NRT1/PTR	1.740940	NA	2.43478	0.000925	0.001128
Sobic.006G239000	NRT1/PTR	0.333274	0.221838	1.33638	0.000407	0.001106
Sobic.006G251200	NRT1/PTR	1.120390	NA	1.25673	0.000636	0.000281
Sobic.009G099000	NRT1/PTR	0.000000	1.123060	0.56059	0.000828	0.001677
Sobic.010G133100	NRT1/PTR	0.252617	0.411819	1.73304	0.006464	0.006129
Sobic.004G314200	NiR	1.139690	0.517056	0.90156	0.000470	0.001648
Sobic.008G006900	NiR	NA	0.653504	1.35632	0.000667	0.001840
Sobic.001G002000	GS	3.256560	0.953878	1.05790	0.000615	0.000575
Sobic.001G497300	GS	1.011360	0.673694	NA	0.001000	0.001031
Sobic.002G133400	GS	0	1.712950	NA	0.000690	0.000802
Sobic.006G210900	GS	1.110000	0.211033	0.27145	0.001053	0.003297
Sobic.007G078300	GS	2.487300	3.282810	NA	0.000582	0.002025

K_a:K_s-values are given along with their respective nucleotide diversities, and those highlighted in gray display positive selection (K_a:K_s >1).

TABLE 2 | List of genes and their population genetics data of those found to be displaying pressures of selection at the gene level throughout domestication.

Selection at gene level	Gene ID	Function	Selective Sweep (kb)	Nucleotide diversity ($\theta\pi$)		Tajima's D (TajD)		F_{ST}
				Landraces	Wild & Weedy	Landraces	Wild & Weedy	Domestication
Purifying	*Sobic.001G422300*	NiR	174.4	0	0.000764	NA	−0.345900	0.694667
Purifying	*Sobic.002G413100*	AST	188.0	0	0.013148	NA	1.273849	0.593150
Purifying	*Sobic.007G078500*	GS	79.4	0.000026	0.007312	−0.567739	1.464598	0.644061
Balancing	*Sobic.006G169500*	ASPG	–	0.007699	0.000236	2.001733	−1.313478	0.274193
Balancing	*Sobic.009G225700*	GOGAT	–	0.003671	0.002523	3.008451	−1.247962	0.188190

Sobic.001G443900 (NPF8.7) with 16 codons under selection, two of which were non-synonymous.

The two genes with the highest number of codons under balancing selection were both from the GOGAT family: *Sobic.003G258800* with 23 codons, seven of which were non-synonymous; and *Sobic.009G225700* with 20 codons with signatures of balancing selection, of which 13 were non-synonymous. Both genes were determined to be displaying signatures of balancing selection at the gene level and also had codons with signatures of purifying selection. The first gene was GOGAT gene *Sobic.009G225700*, with seven codons with signatures of purifying selection, of which four were non-synonymous (three nonpolar and one acidic residue). Additionally, the ASPG gene (*Sobic.006G169500*) had one non-synonymous codon (nonpolar residue) with a signature of purifying selection.
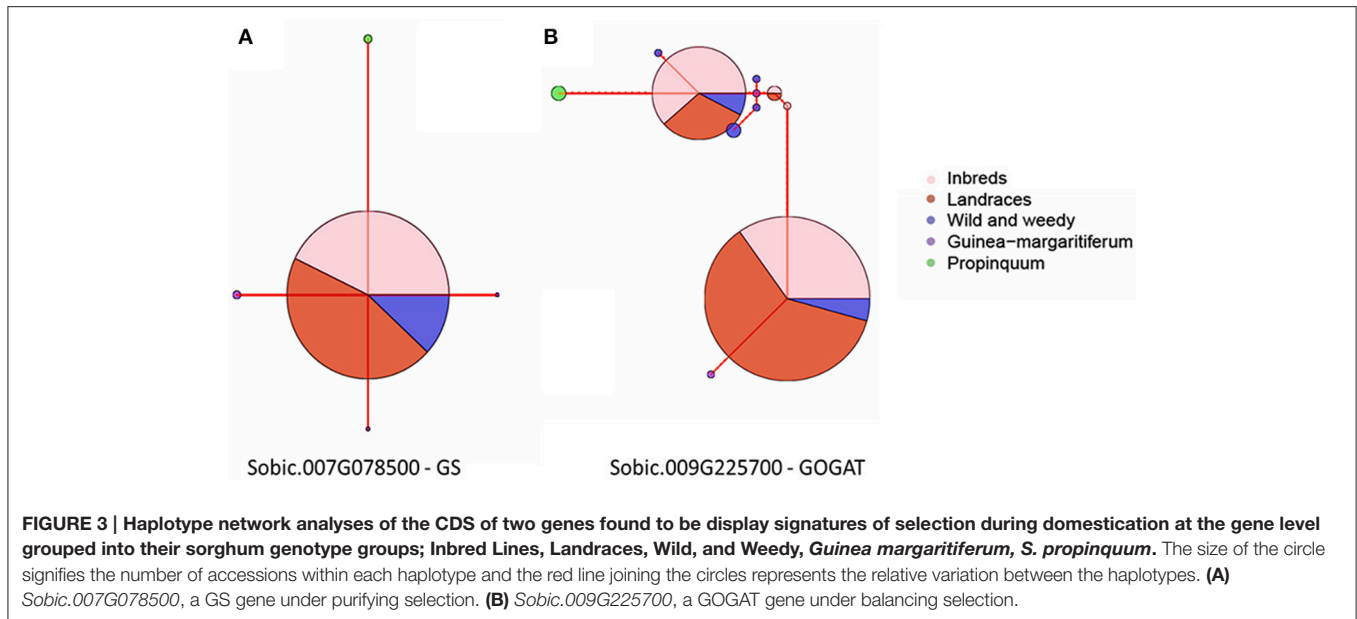
Haplotype Network Analyses

Haplotype network analyses were performed on all five genes with signatures of selection identified at the whole gene level throughout domestication. This analysis was based on their genotype grouping (*S. propinquum*, *G. margaritiferum*, Wild and Weedy, Landrace and Inbred Lines), their geographical location (East Africa, West Africa, South Africa, Asia, New World) and race/species/wild (caudatum, complex, durra, guinea, kafir, *G. margaritiferum*). **Figures 3A,B** show the common trends for two genes under purifying and balancing selection and illustrate the effect of domestication on these traits by either limiting diversity with one major haplotype or maintaining two or more prevalent haplotypes within cultivated genotype groups. The haplotype network analysis for the GS gene (*Sobic.007G078500*) identified with a signature of purifying selection at both whole gene and base-pair levels is presented in **Figure 3A**. **Figure 3B** shows the results of the haplotype network analysis for the GOGAT gene (*Sobic.009G225700*) under balancing selection at both gene and base-pair level throughout domestication. A contrasting example of a gene under balancing selection during domestication, *Sobic.006G169500* in the ASPG gene family is presented in **Figure 4A**, which maintains three high frequency haplotypes, one of which is dominated by the caudatum and complex racial types. **Figure 4B** displays the haplotype network analysis of the AST gene *Sobic.002G413100* which was found to share its haplotype with *G. margaritiferum*.

DISCUSSION

Diversity within Gene Families

This study presents the largest evaluation of sequence diversity in genes within the N uptake and utilization pathway in a diverse panel of sorghum accessions and has identified a significant reduction of diversity in 80% of these genes as a result of domestication. This may be due to selective constraints on specific genes or blocks of genes, or the effects of an evolutionary bottleneck. The overall average reduction in diversity in genes within this pathway after domestication was in line with genome-wide averages (approx. 1.7-fold; Mace et al., 2013) and in the primary starch biosynthesis pathway (Campbell et al., 2016). These results uncover substantial amounts of unexploited genetic diversity within Wild and Weedy lines that could be evaluated by breeding programs. However, this trend was not observed for all gene families; AS, ASPG, and AAL displayed higher sequence diversity within the Landrace genotypes, although this may be attributed to the low number of genes within these families where outliers would exert substantial effects. The AAL gene family (2 genes) displays the highest nucleotide diversity, while the AMT1 family (2 genes) displays the lowest nucleotide diversity, representing differing pressures on gene families throughout the N uptake and utilization pathway.

NRT2, AMT1, and AMT2 gene family members had low nucleotide diversity in the Landrace grouping, suggesting selective constraint on all ammonium transporters and high-affinity nitrate transporters. The high-affinity nitrate transporter gene families (NRT2 and NRT3) co-function as a tetramer containing two subunits of each (Yong et al., 2010) and play a critical role in nitrate uptake as they are upregulated in roots during limited N conditions in sorghum (Gelli et al., 2014). An ancient whole-genome duplication in sorghum (Paterson et al., 2004) resulted in the formation of sub-functional protein complexes, and there are examples of asymmetrical evolutionary rates between such subunits (e.g., ADP-glucose pyrophosphorylase; Georgelis et al., 2008; Corbi et al., 2011). Similarly, the co-functioning gene families, which may have arisen from the duplication event, have contrasting nucleotide diversities with the NRT3 family having a 2-fold higher nucleotide diversity in both Landrace and Wild and Weedy genotypes. Four of the five NRT2 genes display SNPs with signatures of purifying selection, whereas the NRT3 gene family lacks any evidence of selection suggesting greater plasticity in

FIGURE 3 | Haplotype network analyses of the CDS of two genes found to be display signatures of selection during domestication at the gene level grouped into their sorghum genotype groups; Inbred Lines, Landraces, Wild, and Weedy, *Guinea margaritiferum, S. propinquum*. The size of the circle signifies the number of accessions within each haplotype and the red line joining the circles represents the relative variation between the haplotypes. **(A)** *Sobic.007G078500*, a GS gene under purifying selection. **(B)** *Sobic.009G225700*, a GOGAT gene under balancing selection.

NRT3 proteins. As NRT3 proteins induce initiation of lateral roots under low N conditions and modulate their growth under nutritional cues (Orsel et al., 2006), hypothetically, allelic variation within NRT3 genes may be beneficial for optimal transport of nitrate and root growth under variable environmental and soil conditions often found in subsistence farming systems.

Within the Landrace group, 20% of genes exhibited higher levels of nucleotide diversity, which is a preliminary indication of balancing selection, principally found within the NRT1/PTR gene family. Increases in nucleotide diversity can be attributed to a relaxation of selective constraint, or a desired increase to generate allelic variants that may be beneficial under subsistence agriculture N conditions. Surprisingly, the Inbred breeding lines, albeit a small subset of varieties with arguably similar genotypes to Landrace, still demonstrate an increase of diversity in 45% of genes within the N uptake and utilization pathway. These genotypes includes sorghums generated as sweet sorghum types, as well as those developed for a variety of farming practices in Australia, USA, China, India, and countries in west and east Africa. Thus, the variability in desired breeding outcomes may have resulted in this allelic diversity. An example of this is two NRT1/PTR genes, *Sobic.001G444000* & *Sobic.001G444100*, both within the NPF8 family which includes functional peptide transporters (PTR) with no affinity for nitrate (Komarova et al., 2008; Weichert et al., 2012; Léran et al., 2014). Both these genes are invariant within the Landrace category as well as the majority of the additional breeding lines analyzed. The exceptions are within two sorghum genotypes Rio and E-tian, both of which are sweet sorghums. Surprisingly, this variant allele is also similar to those of the two *Guinea-margaritiferums*. This finding provides some evidence to suggest that varying end uses may have placed different selection pressures on sorghum's nitrogen uptake and utilization pathway.

Six other genes are invariant within Landrace accessions. One belongs to the NRT1/PTR family, where *Sobic.003G304000* is within the NPF1 subfamily and functions as a high-affinity nitrate transporter in Arabidopsis (Léran et al., 2014), suggesting this gene is also a critical transporter protein within sorghum. Additionally, an AMT2 gene *Sobic.003G400300,* is invariant among all 44 sorghum lines analyzed. Based on preliminary expression profiles, this gene is likely involved in the mobilization and recycling of ammonium during early grain filling (Makita et al., 2015). As this gene does not have a homologous gene in rice or maize, it may represent a species-specific adaptation within N uptake and utilization in sorghum. This strict selective constraint is likely due to a combination of its crucial function and inflexibility of protein sequence.

Purifying Selection

Signatures of purifying selection were identified in three genes at the whole gene level during domestication: a NiR, an AST and a GS gene. These are likely optimally functioning isoforms and may contribute to phenotypes improving agronomic performance, and therefore are worth further characterization in sorghum. Two of these genes, the NiR and AST were also invariant throughout Landrace genotypes. However, all of these genes were located within a broader selective sweep where surrounding genes also displayed similar pressures of selection and/or invariance. Since the majority of sorghum genes are uncharacterized it is difficult to determine the drivers of such selective pressures.

Additionally, this pathway contains less than expected number of genes displaying purifying selection in comparison to the sorghum genome as a whole. This may be attributable to the lack of methods available to phenotype NUE during crop domestication and thus, genes involved in N uptake and utilization were only indirectly selected for. The N uptake and utilization pathway also requires high levels of flexibility,
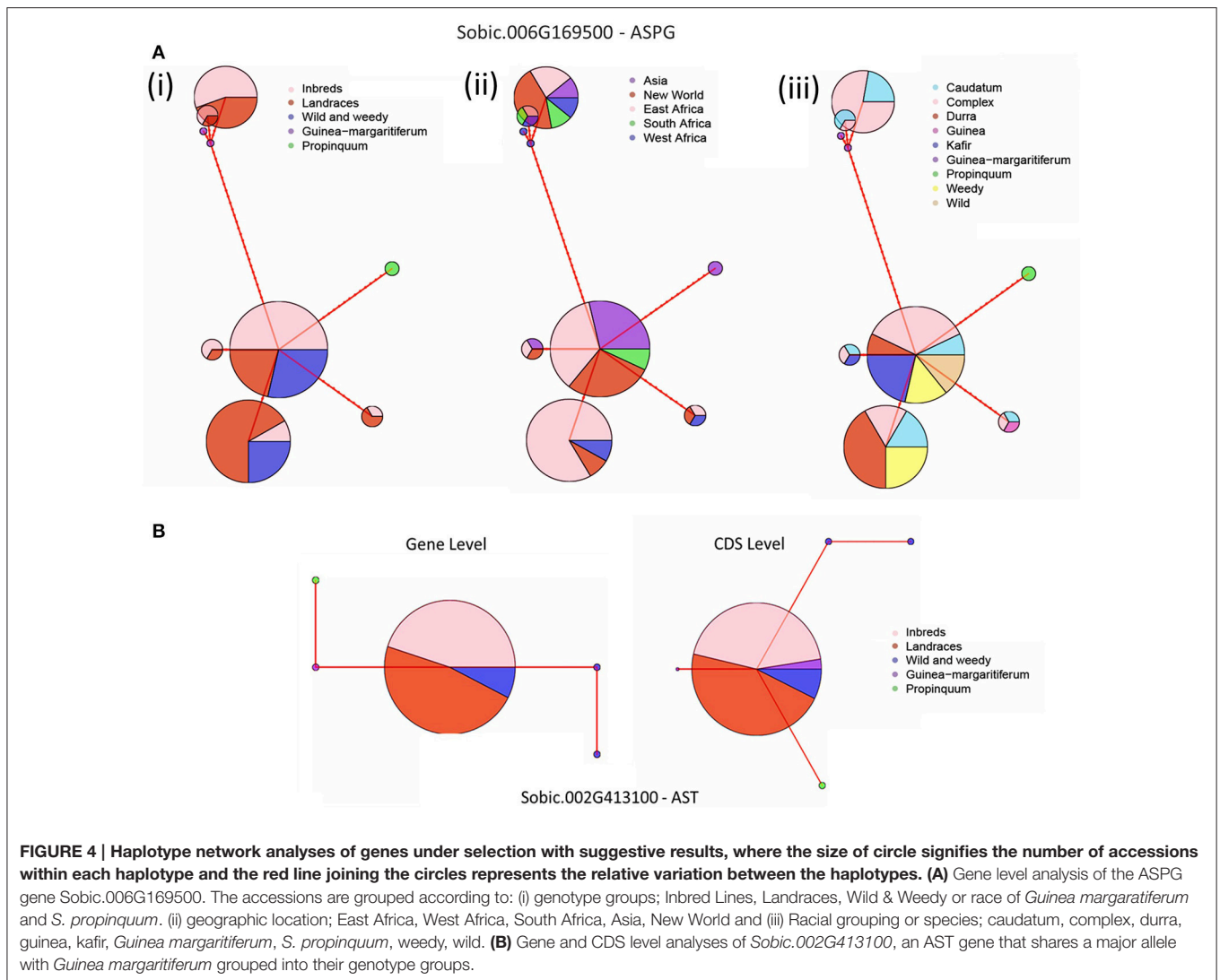
FIGURE 4 | Haplotype network analyses of genes under selection with suggestive results, where the size of circle signifies the number of accessions within each haplotype and the red line joining the circles represents the relative variation between the haplotypes. (A) Gene level analysis of the ASPG gene Sobic.006G169500. The accessions are grouped according to: (i) genotype groups; Inbred Lines, Landraces, Wild & Weedy or race of *Guinea margaratiferum* and *S. propinquum*. (ii) geographic location; East Africa, West Africa, South Africa, Asia, New World and (iii) Racial grouping or species; caudatum, complex, durra, guinea, kafir, *Guinea margaritiferum*, *S. propinquum*, weedy, wild. **(B)** Gene and CDS level analyses of *Sobic.002G413100*, an AST gene that shares a major allele with *Guinea margaritiferum* grouped into their genotype groups.

where Landrace sorghums are placed on soils with a range of fertility, climate conditions, end uses and differing farming practices adopted by an assortment of ethnically and culturally diverse people. Thus, combined genomic analyses of such diverse sorghums may result in the loss of signals of purifying selection, whereas analyzing sorghum subpopulations taken from similar soil types and geographical locations may allow for the discovery of additional genes under purifying selection.

Balancing Selection

Genes identified with signatures of balancing selection maintained several alleles in the population conferring a selective heterozygous advantage and likely function optimally under varied conditions according to specificities in development, tissue, or stressors. Two genes were determined to be under balancing selection at the whole gene level throughout domestication. Although only one GOGAT gene family member was identified with signatures of balancing selection at the gene level, two were identified with at least 20 codons displaying

pressures of balancing selection throughout domestication (*Sobic.009G225700* and *Sobic.003G258800*). GOGAT isoforms can have distinct physiological functions (Vance et al., 1994; Lam et al., 1995), where various genes display cell-specific and organ-specific patterns of expression with differential regulation (Edwards et al., 1990; Thum et al., 2003; Suzuki and Knaff, 2005), and as such, the GOGAT genes may play important roles under variable N conditions.

An ASPG gene (*Sobic.006G169500*) was identified with a signature of balancing selection at both the whole gene level and base-pair level with 11 codons under balancing selection (four non-synonymous—two polar, one nonpolar, one basic) while also having the largest increase in nucleotide diversity throughout domestication. With the above exception, haplotype network analyses of genes displaying pressures of balancing selection generally indicate that heterozygosity is independent of genotype grouping, race, and geographical location. Modern hybrids are generated from crossing two inbred parental lines with genotypes from different racial origins, typically a caudatum-derived line

and either a kafir- or durra-derived line. Caudatums are highly utilized in breeding as they typically have the largest and roundest seeds, resulting in higher yields and a reduction in net loss during post-harvest processing (Doggett, 1988; Dahlberg, 2000). It is possible the dominant use of caudatums in breeding has forced pressures of balancing selection upon certain loci based on racial structure, as the haplotype network analysis demonstrates one of the core alleles is dominated by caudatum and complex races. Such diversity within other genes displaying balancing selection cannot be attributed to racial structure, and is more likely the result of adapting to varied environmental conditions throughout domestication. Further characterization is required to confirm if heterozygous plants confer a selective advantage.

Positive Selection

Positive selection is the accumulation of relatively high levels of amino acid altering mutations, modifying the function of the gene. Signatures of positive selection were determined in 22 genes, most frequently within NRT1/PTR and GS gene families, with a few examples in NiR. In cases such as the NRT1/PTR gene *Sobic.001G111100* (NPF8.5) and the GS gene *Sobic.007G078300*, values demonstrate positive selection in both Landrace and Wild and Weedy sorghums. The former, *Sobic.001G111100* has been described within the NPF8 subfamily involved in the transport of dipeptides (Léran et al., 2014), but may have an altered function in sorghum. Another gene of interest belonging to the GS gene family (*Sobic.001G002000*), came under strong positive selection in Landrace genotypes as well as during domestication. As the GS family likely contains multiple key isoforms, this gene is of particular interest as it may be adapted specifically for subsistence agriculture practices and would be worth exploring variants of this gene in Landrace genotypes.

Two genes *Sobic.009G099000* (NRT1/PTR, NPF2.1) and *Sobic.002G133400* (GS) underwent positive selection exclusively in Wild and Weedy lines. This suggests these genes may be adapting to limited or variable N supply, making them ideal candidates to gain insights into potential modes for improving NUE under low N conditions. The NRT1/PTR gene (*Sobic.009G099000*—NPF2.1) is within the subfamily of nitrate transporters involved in both the efflux of nitrate in mature roots and known as a nitrate excretion transporter (NAXT), critical during early embryo development (Segonzac et al., 2007; Almagro et al., 2008). Alternatively, this NAXT gene in Wild and Weedy may be under exclusive positive selection as Wild and Weedy sorghums are perennials in comparison to cultivated sorghums which are semi-perennials grown as annuals. Thus, changing the root system to operate for longer periods of time, potentially reducing the investment into N assimilation systems. Further characterization of genes displaying such selection could shed insights into sorghum's adaptability to low and varied N conditions within its Wild and Weedy and Landrace groups.

Orthologous NUE Improvement Genes

Nitrogen uptake and utilization genes in sorghum were compared to homologous genes in maize and rice to determine if similar pressures of selection have been identified in the closely related cereals. Orthologs of 22 maize and six rice

genes were found to have signatures of selection throughout their domestication (He et al., 2011; Huang et al., 2012; Hufford et al., 2012; Jiao et al., 2012). However, of these, only three orthologous genes in sorghum (*Sobic.003G295500*—NRT1/PTR, *Sobic.007G125900*—GS, *Sobic.006G169500*—ASPG) were identified with similar selection pressures at the gene level. Additionally, a GS gene (*Sobic.005G213200*) displayed evidence of selection in rice, maize, and sorghum. Comparative analysis studies on pressures of selection and SNP analysis may be informative for other crops since sorghum is a good model for more complex genomes, such as maize and sugarcane.

The gene family member NRT1.1B (NPF6.5) was identified with signatures of selection in both sorghum and rice. In rice, it has been demonstrated that a variant allele contributes to increased nitrate absorption (Hu et al., 2015). The NPF6 subfamily of NRT1/PTR transporters have been found to be involved in dual-affinity nitrate transport (Léran et al., 2014). Additionally, two rice genes described as short panicle 1 (*Os02g48570*—NPF7.11 & *Os11g12740*—NPF4.1) play roles in tillering and panicle elongation (Li et al., 2009; Wang and Li, 2011) and are also found under selection in maize (Hufford et al., 2012; Jiao et al., 2012). The homologs in sorghum are both NRT1/PTR genes (*Sobic.004G260500*—NRT1.5, NPF7.9 & *Sobic.005G090600*—NPF4.1) and each display a synonymous SNP under selection during domestication. They are within differing subfamilies, both of which in Arabidopsis are not always involved in transport of nitrate (Léran et al., 2014), however, *AtNRT1.5* has been linked to xylem loading of nitrate (Lin et al., 2008). Therefore, exploring variations of this gene within diverse sorghums for varying efficiencies of nitrate uptake and transport is of interest.

In maize, the GS1 gene *ZmGln1;3* improves kernel number and size when constitutively overexpressed (Martin et al., 2006) and was identified as being under selection, and shares 96% protein identity with the sorghum gene *Sobic.004G24700*, however, the sorghum ortholog lacks evidence of selection. In rice, overexpression of *GS1;1* (*Os02g50240*) and *GS1;2* (*Os03g0223400*) improved soluble N content within the seeds (Cai et al., 2009; Thomsen et al., 2014). *GS1;1* shares 87% protein identity, and *GS1;2* gene shares 95% protein identity with *Sobic.004G247000*, the same GS1 protein sharing high identity within the maize studies. However, overexpression of this gene in maize transformation studies has led to inconsistent results whereby previous results cannot be reproduced or have not been validated in the field (Martin et al., 2006; Cañas et al., 2010; Thomsen et al., 2014). The limited data available on overexpression, even in one of the most studied enzymes within this pathway, highlights the gaps in our knowledge and the need for further research.

Candidate Genes for NUE Improvement within Sorghum

Re-introduction of diversity within genes can be accomplished by utilizing the diversity from Wild and Weedy accessions, cultivated *G. margaritiferums* or the Asian species *S. propinquum*. Utilizing exotic germplasm could be beneficial for a variety of

improvements, as diversity is often viewed as the key to adaption. This is especially true within the N uptake and utilization pathway as available N sources are often highly variable, even within modern agriculture systems. High affinity transporter genes (NRT2/NRT3, AMT1) often have low nucleotide diversity in cultivated accessions in comparison to Wild and Weedy accessions and are therefore important candidates for the re-introduction of diversity. In rice, *OsNRT2.3* has two splice variants, where *OsNRT2.3B* is expressed in the phloem and regulates nitrate transport activity through sensing pH levels (Cai et al., 2008; Feng et al., 2011). High activity of this transporter enhances the uptake of N, iron, and phosphate through enhancing the pH buffering capacity of the plant, where field trials resulted in a 40% increase in NUE (Fan et al., 2016). The homologous gene in sorghum is *Sobic.003G270800*, an NRT2 gene that displays an increasing nucleotide diversity throughout domestication and improvement, potentially demonstrating selection in modern breeding for increased diversity. For other sorghum NRT2 genes which are not displaying such an increase, it would be worth exploring variants within Wild and Weedy populations. These high-affinity transporter genes are ideal NUE targets and should be introduced into cultivated material to assess their value in modern production systems.

Other candidate genes for NUE include the three genes found under purifying selection at gene level. The corresponding alleles for these genes within the *G. margaritiferum* genotypes were often quite varied and only demonstrated shared haplotypes in two cases. The first is NRT1/PTR gene *Sobic.003G185100* (NPF6.2) which has homologs in maize (*GRMZM2G064091*) and rice (*Os01g37590*). In Arabidopsis, this gene is described as *AtNRT1.4* and is most highly expressed within the petiole, which is linked to storage of nitrate under high N supply (Chiu et al., 2004; De Angeli et al., 2006; Tsay et al., 2007). If this gene retains a similar function in sorghum, it could represent a key gene for nitrate storage or remobilization in leaf sheaths, the grass equivalent to the dicotyledonous petiole. The second gene is an AST gene (*Sobic.002G413100*) that displays pressures of purifying selection during domestication and is invariant throughout Landrace accessions analyzed. Interestingly, this gene only shares the haplotype with *G. margaritiferum* races of sorghum in the CDS, whereas at the gene level its haplotype is not as closely related. This likely occurred through accumulation of synonymous mutations where it has 13 SNPs displaying pressures of purifying selection during domestication, six of which were synonymous. It is also possible this represents differential regulation of this gene between the two domestication events, where the dissimilarity at the gene level is within regulatory or non-CDS regions, and would be worth characterizing. Thus, typically the *G. margaritiferum* genotypes present high levels of genetic diversity which could be utilized for improving NUE.

CONCLUSION

This study focused on the analysis of genes within the N uptake and utilization pathway and has successfully identified a number of key gene targets for further research. Due to limited experimental data regarding sorghum's N assimilation pathway, the vast majority of these genes are uncharacterized and function could only be assumed based on homology. The genes discussed herein may not be an exhaustive list and ultimately may not function as suggested within this study. For example, the NRT1/PTR gene family encompasses a variety of transporters involved in the transport of proteins and other compounds, and not all members will be linked to the uptake and transport of nitrate. Additional bioinformatics analyses concerning diverse sorghums grown under known soil types and fertilizer regimes would be beneficial for breeding programs to produce lines adapted to specific geographic locations and farming practices. Furthermore, performing gene-based association mapping within NAM populations is also worth exploring to link genes and their allelic diversity to NUE phenotypes.

A more complete understanding of the components involved in this pathway is vital toward the progress of improving NUE. Specifically, transcriptome data of sorghum grown under varied environmental conditions, especially in low N levels, could elucidate key stress-induced genes and their respective promoters. Previous overexpression studies often used constitutive promoters, but tissue-specific or inducible promoters have been suggested to yield more consistent results, especially under limited N conditions, and may avoid negative feedback loops that have hindered results of previous studies (Thomsen et al., 2014; Garnett et al., 2015). Due to the tight regulation evident within this pathway, genetic modification is likely required to introduce variants or alter proteins in order to increase the efficiency of enzymes under specific conditions. This would avoid turning off upstream and downstream reactions, while focusing on modifying key SNPs within genes that display pressures of selection. It is with optimism that further characterization of key genes and elucidation of regulatory networks involved in N cycling of plants will facilitate the development of NUE crops.

AUTHOR CONTRIBUTIONS

Conception and design of the work was substantially performed by KM, BC, IG, JB, and DJ. Majority of the sequence data was generously given by EM and DJ and statistically analyzed by KM, BC, ST, YT, and additional analysis into transporter genes was performed by KM and BW. The manuscript underwent revision and was given final approval by all.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.01544

Figure S1 | A comparison of the mean K_A:K_S ratios for each gene family within the N uptake and utilization pathway between Landrace (orange triangle) and Wild & Weedy (gray circle) genotype groups.

Table S1 | Description of sorghum accessions used in the population genetics study, their category, country of origin, general geographic location, and if they were included in the haplotype network analyses. This includes, 18 Landraces, seven Wild and Weedy, 19 Inbred Lines, two *Guinea margaritiferum*, and two *S. propinquum* accessions.

Table S2 | The 230 genes analyzed within the N uptake and utilization pathway displaying orthologous maize and rice genes.

Table S3 | List of NRT1/PTR genes with their respective NRT1/PTR supergroups, homology to Arabidopsis, and NPF family.

Table S4 | The 230 identified genes putatively involved in N uptake and utilization in sorghum with their population genetic analyses ($\theta\pi$, TajD, F_{ST}). These are organized into their gene families and in the order of the biosynthetic pathway. Genes with a reduction in diversity throughout domestication and improvement have a white background in the nucleotide diversity section, while those with an increase in diversity during only domestication are green, throughout only improvement are gray, and throughout both domestication and improvement are blue. Genes with high F_{ST}-values (>0.5) are highlighted in yellow.

Table S5 | List of invariant genes during domestication and their respective population genetic data. Highlighted in green is *Sobic.003G400300* which is invariant throughout all sorghum lines analyzed. High F_{ST}–values (>0.5) are highlighted in yellow.

Table S6 | Statistical Tests (A) mlHKA test results of 3 genes under purifying selection and 2 under balancing selection compared to 34 neutral genes. Seven runs were performed with *p*-values validating selection. **(B)** The 34 neutral genes used as input and their chromosome and bp start and stop site.

Table S7 | χ^2 analysis on the frequency of sorghum population genetic analyses within 5 and 95% tails grouped according to gene families. Significant results are highlighted in yellow. Only significant results were shown.

Table S8 | List of 128 genes that display codons under purifying selection during domestication. Includes base pair position, codon, amino acid, and whether a polymorphism is synonymous or non-synonymous.

Table S9 | List of 91 genes that display codons under balancing selection during domestication. Includes base pair position, codon, amino acid, and whether a polymorphism is synonymous or non-synonymous.

REFERENCES

Abiko, T., Wakayama, M., Kawakami, A., Obara, M., Kisaka, H., Miwa, T., et al. (2010). Changes in nitrogen assimilation, metabolism, and growth in transgenic rice plants expressing a fungal NADP(H)-dependent glutamate dehydrogenase (gdhA). *Planta* 232, 299–311. doi: 10.1007/s00425-010-1172-3

Almagro, A., Lin, S. H., and Tsay, Y. F. (2008). Characterization of the *Arabidopsis* nitrate transporter NRT1.6 reveals a role of nitrate in early embryo development. *Plant Cell* 20, 3289–3299. doi: 10.1105/tpc.107.056788

Ameziane, R., Bernhard, K., and Lightfoot, D. (2000). Expression of the bacterial gdhA gene encoding a NADPH glutamate dehydrogenase in tobacco affects plant growth and development. *Plant Soil* 221, 47–57. doi: 10.1023/A:1004794000267

Anglani, C. (1998). Sorghum for human food – A review. *Plant Foods Hum. Nutr.* 52, 85–95. doi: 10.1023/A:1008065519820

Bernard, S. M., and Habash, D. Z. (2009). The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytol.* 182, 608–620. doi: 10.1111/j.1469-8137.2009.02823.x

Borrell, A. K., and Hammer, G. L. (2000). Nitrogen dynamics and the physiological basis of stay-green in sorghum. *Crop Sci.* 40, 1295–1307. doi: 10.2135/cropsci2000.4051295x

Cai, C., Wang, J. Y., Zhu, Y. G., Shen, Q. R., Li, B., Tong, Y. P., et al. (2008). Gene structure and expression of the high-affinity nitrate transport system in rice roots. *J. Integr. Plant Biol.* 50, 443–451. doi: 10.1111/j.1744-7909.2008.00642.x

Cai, H., Zhou, Y., Xiao, J., Li, X., Zhang, Q., and Lian, X. (2009). Overexpressed glutamine synthetase gene modifies nitrogen metabolism and abiotic stress responses in rice. *Plant Cell Rep.* 28, 527–537. doi: 10.1007/s00299-008-0665-z

Campbell, B. C., Gilding, E. K., Mace, E. S., Tai, S., Tao, Y., Prentis, P. J., et al. (2016). Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol. J.* doi: 10.1111/pbi.12578. [Epub ahead of print].

Cañas, R. A., Quilleré, I., Lea, P. J., and Hirel, B. (2010). Analysis of amino acid metabolism in the ear of maize mutants deficient in two cytosolic glutamine synthetase isoenzymes highlights the importance of asparagine for nitrogen translocation within sink organs. *Plant Biotechnol. J.* 8, 966–978. doi: 10.1111/j.1467-7652.2010.00524.x

Cardinale, B. J. (2012). Biodiversity loss and its impact on humanity. *Nature* 489, 326–326. doi: 10.1038/nature11373

Chiba, Y., Shimizu, T., Miyakawa, S., Kanno, Y., Koshiba, T., Kamiya, Y., et al. (2015). Identification of *Arabidopsis thaliana* NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. *J. Plant Res.* 128, 679–686. doi: 10.1007/s10265-015-0710-2

Chiu, C. C., Lin, C. S., Hsia, A. P., Su, R. C., Lin, H. L., and Tsay, Y. F. (2004). Mutation of a nitrate transporter, AtNRT1: 4, results in a reduced petiole nitrate content and altered leaf development. *Plant Cell Physiol.* 45, 1139–1148. doi: 10.1093/pcp/pch143

Corbi, J., Debieu, M., Rousselet, A., Montalent, P., Le Guilloux, M., Manicacci, D., et al. (2011). Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme. *Theor. Appl. Genet.* 122, 705–722. doi: 10.1007/s00122-010-1480-9

Dahlberg, J. A. (2000). "Classification and characterization of sorghum," in *Sorghum: Origin, History, Technology, and Production*, eds C. W. Smith and R. A. Frederikson (New York, NY: John Wiley & Sons), 99–130.

De Angeli, A., Monachello, D., Ephritikhine, G., Frachisse, J. M., Thomine, S., Gambale, F., et al. (2006). The nitrate/proton antiporter AtCLCa mediates nitrate accumulation in plant vacuoles. *Nature* 442, 939–942. doi: 10.1038/nature05013

Doggett, H. (1988). *Sorghum.* New York, NY: Wiley.

Edwards, J. W., Walker, E. L., and Coruzzi, G. M. (1990). Cell-specific expression in transgenic plants reveals nonoverlapping roles for chloroplast and cytosolic glutamine synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1459–3463. doi: 10.1073/pnas.87.9.3459

Fan, X. R., Feng, H. M., Tan, Y. W., Xu, Y. L., Maio, Q. S., and Xu, G. H. (2016). A putative 6-transmembrane nitrate transporter OsNRT1.1b plays a key role in rice under low nitrogen. *J. Integr. Plant Biol.* 58, 590–599. doi: 10.1111/jipb.12382

Feng, H. M., Yan, M., Fan, X. R., Li, B. Z., Shen, Q. R., Miller, A. J., et al. (2011). Spatial expression and regulation of rice high-affinity nitrate transporters by nitrogen and carbon status. *J. Exp. Bot.* 62, 2319–2332. doi: 10.1093/jxb/erq403

Forde, B. G., and Lea, P. J. (2007). Glutamate in plants: metabolism, regulation, and signalling. *J. Exp. Bot.* 58, 2339–2358. doi: 10.1093/jxb/erm121

Garnett, T., Conn, V., and Kaiser, B. N. (2009). Root based approaches to improving nitrogen use efficiency in plants. *Plant Cell Environ.* 32, 1272–1283. doi: 10.1111/j.1365-3040.2009.02011.x

Garnett, T., Conn, V., Plett, D., Conn, S., Zanghellini, J., Mackenzie, N., et al. (2013). The response of the maize nitrate transport system to nitrogen demand and supply across the lifecycle. *New Phytol.* 198, 82–94. doi: 10.1111/nph.12166

Garnett, T., Plett, D., Heuer, S., and Okamoto, M. (2015). Genetic approaches to enhancing nitrogen-use efficiency (NUE) in cereals: challenges and future directions. *Funct. Plant Biol.* 42, 921–941. doi: 10.1071/FP15025

Gaufichon, L., Rothstein, S. J., and Suzuki, A. (2015). Asparagine metabolic pathways in *Arabidopsis. Plant Cell Physiol.* 57, 675–689. doi: 10.1093/pcp/pcv184

Gelli, M., Duo, Y., Konda, A. R., Zhang, C., Holding, D., and Dweikat, I. (2014). Identification of differentially expressed genes between sorghum genotypes with contrasting nitrogen stress tolerance by genome-wide transcriptional profiling. *BMC Genomics* 15:179. doi: 10.1186/1471-2164-15-179

Georgelis, N., Braun, E. L., and Hannah, L. C. (2008). Duplications and functional divergence of ADP-glucose pyrophosphorylase genes in plants. *BMC Evol. Biol.* 8:232. doi: 10.1186/1471-2148-8-232

Good, A. G., Shrawat, A. K., and Muench, D. G. (2004). Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? *Trends Plant Sci.* 9, 597–605. doi: 10.1016/j.tplants.2004.10.008

He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X. M., et al. (2011). Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* 7:e1002100. doi: 10.1371/journal.pgen.1002100

Hirel, B., Le Gouis, J., Ney, B., and Gallais, A. (2007). The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58, 2369–2387. doi: 10.1093/jxb/erm097

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., et al. (2015). Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* 47, 834–838. doi: 10.1038/ng.3337

Huang, X., Kurata, N., Wei, X. H., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhajarvi, T., Chia, J. M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–118. doi: 10.1038/ng.2309

Jiao, Y., Zhao, H., Ren, L. H., Song, W. B., Zeng, B., Guo, J. J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44, 812–815. doi: 10.1038/ng.2312

Jombart, T. (2008). *adegenets*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Kamoshita, A., Fukai, S., Muchow, R. C., and Cooper, M. (1998). Sorghum hybrid differences in grain yield and nitrogen concentration under low soil nitrogen availability. I. Hybrids with similar phenology. *Aust. J. Agric. Res.* 49, 1267–1276. doi: 10.1071/A98018

Kamoshita, A., Fukai, S., Muchow, R. C., and Cooper, M. (1999). Grain nitrogen concentration differences among three sorghum hybrids with similar grain yield. *Aust. J. Agric. Res.* 50, 137–145. doi: 10.1071/A98029

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114. doi: 10.1093/nar/gkr988

Komarova, N. Y., Thor, K., Gubler, A., Meier, S., Dietrich, D., Weichert, A., et al. (2008). AtPTR1 and AtPTR5 transport dipeptides in planta. *Plant Physiol.* 148, 856–869. doi: 10.1104/pp.108.123844

Krapp, A., David, L. C., Chardin, C., Girin, T., Marmagne, A., Leprince, A., et al. (2014). Nitrate transport and signalling in Arabidopsis. *J. Exp. Bot.* 65, 789–798. doi: 10.1093/jxb/eru001

Ladha, J. K., Tirol-Padre, A., Reddy, C. K., Cassman, K. G., Verma, S., Powlson, D. S., et al. (2016). Global nitrogen budgets in cereals: a 50-year assessment for maize, rice, and wheat production systems. *Sci. Rep.* 6:19355. doi: 10.1038/srep19355

Lam, H. M., Coschigano, K., Schultz, C., Meolooliveira, R., Tjaden, G., Oliveira, I., et al. (1995). Use of *Arabidopsis* mutants and genes to study amide amino-acid biosynthesis. *Plant Cell* 7, 887–898. doi: 10.1105/tpc.7.7.887

Lemaire, G., Charrier, X., and Hebert, Y. (1996). Nitrogen uptake capacities of maize and sorghum crops in different nitrogen and water supply conditions. *Agronomie* 16, 231–246. doi: 10.1051/agro:19960403

Léran, S., Varala, K., Boyer, J. C., Chiurazzi, M., Crawford, N., Daniel-Vedele, F., et al. (2014). A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci.* 19, 5–9. doi: 10.1016/j.tplants.2013.08.008

Lewandowski, I., and Schmidt, U. (2006). Nitrogen, energy and land use efficiencies of miscanthus, reed canary grass and triticale as determined by the boundary line approach. *Agric. Ecosyst. Environ.* 112, 335–346. doi: 10.1016/j.agee.2005.08.003

Li, S., Qian, Q., Fu, Z., Zeng, D. L., Meng, X. B., Kyozuka, J., et al. (2009). Short panicle1 encodes a putative PTR family transporter and determines rice panicle size. *Plant J.* 58, 592–605. doi: 10.1111/j.1365-313X.2009.03799.x

Lin, S. H., Kuo, H. F., Canivenc, G., Lin, C. S., Lepetit, M., Hsu, P. K., et al. (2008). Mutation of the *Arabidopsis* NRT1.5 nitrate transporter causes defective root-to-shoot nitrate transport. *Plant Cell* 20, 2514–2528. doi: 10.1105/tpc.108.060244

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Commun.* 4, 2320. doi: 10.1038/ncomms3320

Makita, Y., Shimada, S., Kawashima, M., Kondou-Kuriyama, T., Toyoda, T., and Matsui, M. (2015). MOROKOSHI: transcriptome database in *Sorghum bicolor. Plant Cell Physiol.* 56:e6. doi: 10.1093/pcp/pcu187

Malagoli, P., Laine, P., Rossato, L., and Ourry, A. (2005). Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest. II. An N-15-labelling-based simulation model of N partitioning between vegetative and reproductive tissues. *Ann. Bot.* 95, 1187–1198. doi: 10.1093/aob/mci131

Martin, A., Lee, J., Kichey, T., Gerentes, D., Zivy, M., Tatout, C., et al. (2006). Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* 18, 3252–3274. doi: 10.1105/tpc.106.042689

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., and Suzuki, A. (2010). Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Ann. Bot.* 105, 1141–1157. doi: 10.1093/aob/mcq028

Miller, A., Fan, X., Shen, Q., and Smith, S. J. (2007). Amino acids and nitrate as signals for the regulation of nitrogen acquisition. *J. Exp. Bot.* 59, 111–119. doi: 10.1093/jxb/erm208

Moose, S., and Below, F. E. (2009). "Biotechnology approaches to improving maize nitrogen use efficiency" in *Molecular Genetic Approaches to Maize Improvement*, eds A. L. Kriz and B. A. Larkins (Heidelberg: Springer), 65–77.

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110

Näsholm, T., Kielland, K., and Ganeteg, U. (2009). Uptake of organic nitrogen by plants. *New Phytol.* 182, 31–48. doi: 10.1111/j.1469-8137.2008.02751.x

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

Neuhäuser, B., Dynowski, M., Mayer, M., and And Ludewig, U. (2007). Regulation of NH_4^+ transport by essential cross talk between AMT monomers

through the carboxyl tails. *Plant Physiol.* 143, 1651–1659. doi: 10.1104/pp.106.094243

Oikeh, S. O., Chude, V. O., Kling, G. J., and Horst, W. J. (2007). Comparative productivity of nitrogen-use efficient and nitrogen-inefficient maize cultivars and traditional grain sorghum in the moist Savanna of West Africa. *Afr. J. Agric. Res.* 2, 112–119.

Olson, S. N., Ritter, K., Medley, J., Wilson, T., Rooney, W. L., and Mullet, J. E. (2013). Energy sorghum hybrids: functional dynamics of high nitrogen use efficiency. *Biomass Bioenergy* 56, 307–316. doi: 10.1016/j.biombioe.2013.04.028

Orsel, M., Chopin, F., Leleu, O., Smith, S. J., Krapp, A., Daniel-Vedele, F., et al. (2006). Characterization of a two-component high-affinity nitrate uptake system in Arabidopsis. Physiology and protein-protein interaction. *Plant Physiol.* 142, 1304–1317. doi: 10.1104/pp.106.085209

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26, 419–420. doi: 10.1093/bioinformatics/btp696

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9903–9908. doi: 10.1073/pnas.0307901101

Paungfoo-Lonhienne, C., Lonhienne, T. G. A., Rentsch, D., Robinson, N., Christie, M., Webb, R. I., et al. (2008). Plants can use protein as a nitrogen source without assistance from other organisms. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4524–4529. doi: 10.1073/pnas.0712078105

Paynel, F., Murray, P. J., and Cliquet, J. B. (2001). Root exudates: a pathway for short-term N transfer from clover and ryegrass. *Plant Soil* 229, 235–243. doi: 10.1023/A:1004877214831

Peoples, M. B., Herridge, D. F., and Ladha, J. K. (1995). Biological nitrogen-fixation: an efficient source of nitrogen for sustainable agricultural production? *Plant Soil* 174, 3–28. doi: 10.1007/BF00032239

Pfeifer, B., Wittelsburger, U., Ramos Onsins, S. E., and Lercher, M. J. (2014). *PopGenome:* an efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31, 1929–1936. doi: 10.1093/molbev/msu136

Pratelli, R., and Pilot, G. (2014). Regulation of amino acid metabolic enzymes and transporters in plants. *J. Exp. Bot.* 65, 5535–5556. doi: 10.1093/jxb/eru320

Purnell, M. P., and Botella, J. R. (2007). Tobacco isoenzyme 1 of NAD(H)-dependent glutamate dehydrogenase catabolizes glutamate *in vivo*. *Plant Physiol.* 143, 530–539. doi: 10.1104/pp.106.091330

Raun, W. R., and Johnson, G. V. (1999). Improving nitrogen use efficiency for cereal production. *Agron. J.* 91, 357–363. doi: 10.2134/agronj1999.00021962009100030001x

Segonzac, C., Boyer, J. C., Ipotesi, E., Szponarski, W., Tillard, P., Touraine, B., et al. (2007). Nitrate efflux at the root plasma membrane: identification of an *Arabidopsis* excretion transporter. *Plant Cell* 19, 3760–3777. doi: 10.1105/tpc.106.048173

Sinclair, A. G., Johnstone, P. D., Smith, L. C., Roberts, A. H. C., O'Connor, M. B., and Morton, J. D. (1997). Relationship between pasture dry matter yield and soil Olsen P from a series of long-term field trials. *N. Z. J Agric. Res.* 40, 559–567.

Sohlenkamp, C., Shelden, M., Howitt, S., and Udvardi, M. (2000). Characterization of *Arabidopsis* AtAMT2, a novel ammonium transporter in plants. *FEBS Lett.* 11, 273–278. doi: 10.1016/S0014-5793(00)01153-4

Suzuki, A., and Knaff, D. B. (2005). Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. *Photosyn. Res.* 83, 191–217. doi: 10.1007/s11120-004-3478-0

Sylvester-Bradley, R., and Kindred, D. R. (2009). Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. *J. Exp. Bot.* 60, 1939–1951. doi: 10.1093/jxb/erp116

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

Tello-Ruiz, M. K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., et al. (2016). Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 44, D1133–D1140. doi: 10.1093/nar/gkv1179

Terce-Laforgue, T., Bedu, M., Dargel-Grafin, C., Dubois, F., Gibon, Y., Restivo, F. M., et al. (2013). Resolving the role of plant glutamate dehydrogenase: II. Physiological characterization of plants overexpressing the two enzyme subunits individually or simultaneously. *Plant Cell Physiol.* 54, 1635–1647. doi: 10.1093/pcp/pct108

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700

Thomsen, H. C., Eriksson, D., Moller, I. S., and Schjoerring, J. K. (2014). Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? *Trends Plant Sci.* 19, 656–663. doi: 10.1016/j.tplants.2014.06.002

Thum, K. E., Shasha, D. E., Lejay, L. V., and Coruzzi, G. M. (2003). Light- and carbon-signaling pathways. Modeling circuits of interactions. *Plant Physiol.* 132, 440–452. doi: 10.1104/pp.103.022780

Tsay, Y. F., Chiu, C. C., Tsai, C. B., Ho, C. H., and Hsu, P. K. (2007). Nitrate transporters and peptide transporters. *FEBS Lett.* 581, 2290–2300. doi: 10.1016/j.febslet.2007.04.047

Tsay, Y. F., Schroeder, J. I., Feldmann, K. A., and Crawford, N. M. (1993). The herbicide sensitivity gene *CHL1* of Arabidopsis encodes a nitrate-inducible nitrate transporter. *Cell* 72, 705–713 doi: 10.1016/0092-8674(93)90399-B

Vance, C. P., Gregerson, R. G., Robinson, D. L., Miller, S. S., and Gantt, J. S. (1994). Primary assimilation of nitrogen in alfalfa nodules – molecular-features of the enzymes involved. *Plant Sci.* 101, 51–64. doi: 10.1016/0168-9452(94)90164-3

Van Oosterom, E. J., Borrell, A. K., Chapman, S. C., Broad, I. J., and Hammer, G. L. (2010a). Functional dynamics of the nitrogen balance of sorghum: I. N demand of vegetative plant parts. *Field Crops Res.* 115, 19–28. doi: 10.1016/j.fcr.2009.09.018

Van Oosterom, E. J., Chapman, S. C., Borrell, A. K., Broad, I. J., and Hammer, G. L. (2010b). Functional dynamics of the nitrogen balance of sorghum. II. Grain filling period. *Field Crop Res.* 115, 29–38. doi: 10.1016/j.fcr.2009.09.019

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., et al. (1997). Human alteration of the global nitrogen cycle: sources and consequences. *Ecol. Appl.* 7, 737–750. doi: 10.1890/1051-0761(1997)007[0737:haotgn]2.0.co;2

Vogan, P. J., and Sage, R. F. (2011). Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of Flaveria Juss. (Asteraceae). *Plant Cell Environ.* 34, 1415–1430. doi: 10.1111/j.1365-3040.2011.02340.x

von Wittgenstein, N. J. J. B., Le, C. H., Hawkins, B. J., and Ehlting, J. (2014). Evolutionary classification of ammonium, nitrate, and peptide transporters in land plants. *BMC Evol. Biol.* 14:11. doi: 10.1186/1471-2148-14-11

Wang, Y. H., and Li, J. Y. (2011). Branching in rice. *Curr. Opin. Plant Biol.* 14, 94–99. doi: 10.1016/j.pbi.2010.11.002

Wang, Y. Y., Hsu, P. K., and Tsay, Y. F. (2012). Uptake, allocation and signaling of nitrate. *Trends Plant Sci.* 17, 458–467. doi: 10.1016/j.tplants.2012.04.006

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7, 256–276. doi: 10.1016/0040-5809(75)90020-9

Weichert, A., Brinkmann, C., Komarova, N. Y., Dietrich, D., Thor, K., Meier, S., et al. (2012). AtPTR4 and AtPTR6 are differentially expressed, tonoplast-localized members of the peptide transporter/nitrate transporter 1 (PTR/NRT1) family. *Planta* 235, 311–323. doi: 10.1007/s00425-011-1508-7

Wendorf, F., Close, A. E., Schild, R., Wasylikowa, K., Housley, R. A., Harlan, J. R., et al. (1992). Saharan exploitation of plants 8,000 years BP. *Nature* 359, 721–724. doi: 10.1038/359721a0

Wright, S. I., and Charlesworth, B. (2004). The HKA test revisited: a maximum-likelihood-ratio test of standard neutral model. *Genetics* 168, 1071–1076. doi: 10.1534/genetics.104.026500

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050

Yamaya, T., Obara, M., Nakajima, H., Sasaki, S., Hayakawa, T., and Sato, T. (2002). Genetic manipulation and quantitative-trait loci mapping for nitrogen recycling in rice. *J. Exp. Bot.* 53, 917–925. doi: 10.1093/jexbot/53.370.917

Yong, Z. H., Kotur, Z., and Glass, A. D. M. (2010). Characterization of an intact two-component high-affinity nitrate transporter from

Arabidopsis roots. *Plant J.* 63, 739–748. doi: 10.1111/j.1365-313X.2010.04278.x

Yuan, L., Loque, D., Kojima, S., Rauch, S., Ishiyama, K., Inoue, E., et al. (2007). The organization of high-affinity ammonium uptake in *Arabidopsis* Roots depends on the spatial arrangement and biochemical properties of AMT1-Type transporters. *Plant Cell* 19, 2636–2652. doi: 10.1105/tpc.107.052134

Zhang, Z., Li, J., and Yu, J. (2006). Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evol. Biol.* 6:44. doi: 10.1186/1471-2148-6-44

Zhou, Y., Bai, L., and Song, C.-P. (2015). Ammonium homeostasis and signaling in plant cells. *Sci. Bull.* 60, 741–747. doi: 10.1007/s11434-015-0759-2

Zhu, X. G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206