Journal of Advanced Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original Article

GWAS identifies genetic loci underlying nitrogen responsiveness in the climate resilient C_4 model *Setaria italica* (L.)

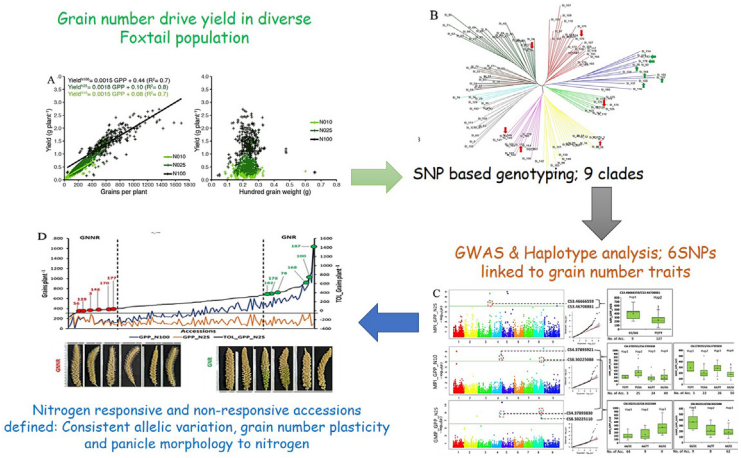
Tirthankar Bandyopadhyay [a,1], Stéphanie M Swarbreck [b,c,1], Vandana Jaiswal [d], Jyoti Maurya [a], Rajeev Gupta [e,f], Alison R. Bentley [g,b], Howard Griffiths [c], Manoj Prasad [a,*]

[a] National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110067, India
[b] The John Bingham Laboratory, NIAB, 93 Lawrence Weaver Rd, Cambridge CB3 0LE, United Kingdom
[c] Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom
[d] CSIR-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh 176061, India
[e] Cereal Crops Research Unit, US Department of Agriculture (USDA) Agricultural Research Service (ARS), Fargo, ND, United States
[f] International Crop Research Institute for the Semi -arid Tropics, Patancheru, Hyderabad, Telangana 502324, India
[g] International Maize and Wheat Improvement Center, Texcoco, México

HIGHLIGHTS

- Evaluating nitrogen(N) responsiveness in crops has many commercial/environmental advantages.
- Current lack of knowledge on its physio-genetic basis is a major bottleneck.
- We demonstrated that N dependent yield increase is driven by grain number (GN) in *S.italica*.
- GN has strong genetic basis –22 unique SNPs; six exhibiting haplotypes in natural population.
- Based on this, we define N responsive and non-responsive accessions with distinct panicle types.
- Few genes lying between SNPs with haplotypes show distinct transcript levels in two genotypes.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 September 2021
Revised 21 January 2022
Accepted 23 January 2022
Available online xxxx

Keywords:
C_4 model crop
Climate resilience
Food security

ABSTRACT

Introduction: N responsiveness is the capacity to perceive and induce morpho-physiological adaptation to external and internal Nitrogen (N). Crop productivity is propelled by N fertilizer and requires the breeding/selection of cultivars with intrinsically high N responsiveness. This trait has many advantages in being more meaningful in commercial/environmental context, facilitating in-season N management and not being inversely correlated with N availability over processes regulating NUE. Current lack of its understanding at the physio-genetic basis is an impediment to select for cultivars with a predictably high N response.

Objectives: To dissect physio-genetic basis of N responsiveness in 142 diverse population of foxtail millet, *Setaria italica* (L.) by employing contrasting N fertilizer nutrition regimes.

Peer review under responsibility of Cairo University.

* Corresponding author at: National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110067, India.
E-mail address: manoj_prasad@nipgr.ac.in (M. Prasad).

[1] Equal contribution.

https://doi.org/10.1016/j.jare.2022.01.010

Nitrogen responsiveness
GWAS
Gene expression

Methods: We phenotyped *S. italica* accessions for major yield related traits under low (N10, N25) and optimal (N100) growth conditions and genotyped them to subsequently perform a genome-wide association study to identify genetic loci associated with nitrogen responsiveness trait. Groups of accessions showing contrasting trait performance and allelic forms of specific linked genetic loci (showing haplotypes) were further accessed for N dependent transcript abundances of their proximal genes.

Results: Our study show that N dependent yield rise in *S. italica* is driven by grain number whose responsiveness to N availability is genetically underlined. We identify 22 unique SNP loci strongly associated with this trait out of which six exhibit haplotypes and consistent allelic variation between lines with contrasting N dependent grain number response and panicle architectures. Furthermore, differential transcript abundances of specific genes proximally linked to these SNPs in same lines is indicative of their N dependence in a genotype specific manner.

Conclusion: The study demonstrates the value/ potential of N responsiveness as a selection trait and identifies key genetic components underlying the trait in *S. italica*. This has major implications for improving crop N sustainability and food security.

Introduction

The co-development of new agronomic practises including the application of nitrogen (N) fertiliser together with the selection of improved crop varieties lead to significant yield enhancement for a few selected crop species during the Green Revolution. However, not all crops benefited to the same level and some of the less improved species tends to be highly relevant to food security in arid and semi-arid regions of the world. One of them is *Setaria italica,* a C_4 cereal crop which is one of the world's ancient and second most cultivated millet globally [1,2]. It is self-pollinated lowland species with demonstrated high biotic and abiotic stress resilience [3]. Being nutritionally rich [4,5], it performs as a major crop in the arid and semi-arid areas of Asia, China as well as sub-Saharan Africa and it is distinctively enriched with slowly digestible and resistant starch making it a healthy low-glycemic index cereal [4]. Taken together, its exceptional adaptability and nutritional attributes have made *S. italica* a promising climate-resilient crop [5] and investigation into the strategies millets employ to regulate productivity in this context is particularly relevant for achieving sustainable future food security. The crop however remains under-investigated in terms of the traits underpinning improvements in breeding.

Agricultural sustainability relies on optimal and resourceful application of fertilizers with nitrogen (N) as a major contributor. At the biochemical and physiological levels, complex interactions between assimilation of N in the form of nitrate (NO_3^-) and carbon dioxide (CO_2) contribute to crop productivity, mainly by coupling N driven leaf growth with photosynthesis (accumulation of higher amounts of light reaction components and CO_2 assimilates) [6]. Insufficient N accessibility is a major constraint to crop productivity worldwide [7]. Despite being expensive, its use in cropping systems in some parts of the developing world is considerably subsidised, often leading to its over application [8,9]. This is associated with undesirable environmental costs including eutrophication of aquatic ecosystems [10], threatening aquatic life and polluting the environment [11]. Furthermore, higher greenhouse gas emission from N fertiliser plants and as N_2O release from fertiliser use are other contributory factors in this regard [12].

Optimization of N provisioning strongly influences yield related agronomic traits [13], N assimilation rates and photosynthetic capacity [14] as well as biomass and many other physiological attributes in cereals [15]. However, in order to optimize N application in crop production, it is essential to appreciate how cereal plants respond to higher N accessibility and the underlying regulation of the process. Any insight in this regard should offer new prospects to select for or help breed new lines that will be more capable of converting applied N to harvestable product with minimal economic and environmental costs [9]. Understanding N responsiveness, defined as the plants ability to induce morpho-physiological adaptation according to external N availability, is key to developing efficient genotypes. Selection of lines with improved ability to utilize available N holds potential genetic, agronomic, environmental and commercial advantages over conventional methods of measuring nitrogen use in crops e.g. nitrogen use efficiency (NUE) [9]. In wheat (*Triticum aestivum* L.) evidence exits that show selection over time has resulted in varieties having better N response compared to landraces characterized by enhanced N responsiveness at early N uptake conditions thereby pushing enhanced performance in field conditions at moderate N levels [16]. Genetic dissection of the trait can therefore highlight hitherto unidentified genomic regions of interest [17-19] with the potential to bridge the gap in our understanding of its regulation at the physiological and genetic levels. This in effect will allow us to understand new questions in crop N biology, for example, how external and internal N availability are perceived by plants and what are the downstream phenotypic responses. Additionally, it offers potential to understand how N is transduced and how plants monitor their N homeostasis at the interface of plant development and primary metabolism.

The present study aimed to reveal the genetic basis for N response in a diverse population of 142 *S. italica* accessions, which are part of core collection of accessions previously studied for agronomic traits under uncontrolled nutrition conditions [20]. We used contrasting N treatments to dissect N responsiveness at the whole plant and genetic marker levels. We found that in *S. italica* yield is mainly driven by grain number per plant instead of grain size. Using genome wide association study (GWAS), we defined major single-nucleotide polymorphisms (SNPs) related to yield traits (e.g., grain number per plant) and derived indices to measure different aspects of N responsiveness. Furthermore, we defined six (6) grain number responsive (GNR) and non-responsive (GNNR) genotypes which exhibit different panicle architectures, contrasting grain number response to low N (ΔN100-N25) and display consistent allelic variation of six SNPs (CS3.46666559, CS3.46708881, CS4.37893830, CS4.37893921, CS8.30225088, CS8.30225110) strongly associated with grain number responsive trait. Transcript abundance profiling of 17 genes proximally linked to these SNPs in the developing panicles of the two genotypes showed that three (3) among them, *Seita.3G363700* (encoding a diacyl glycerol kinase), *Seita.8G160400* (containing a DnaJ chaperon and two DUF domains) and *Seita.8G160500* (encoding T-complex protein 1; TCP-1/cpn60 chaperone family) are differentially regulated while being consistent within each group. This demonstrates that allelic variation of these grains per plant (GPP) linked SNPs and expression of some of their proximal genes are linked in a genotype specific manner.

Materials and methods

Plant material and growth conditions

A collection of 142 diverse *S. italica* accessions (Table S1) were chosen from a *S. italica* core collection (Lata et al., 2011; Lata et al., 2013) and the All India Coordinated Small Millets Improvement Project (AICSMIP, 2014). Accessions represent lines originating from China, India, Bangladesh, Turkey, Kenya, Russia and USA exhibiting relative consistency of germination and viability of seeds.

In order to ensure maximum proximity to seasonal field conditions, plants were grown in pots (19.5 cm height × 20 cm diameter) outdoors under a 70% transparent cover. Three biological replicate pots per accession were settled in a randomized block design. Pots were filled with 3 kg of nitrogen free soilrite mix: vermiculite (2:1 w/w) and saturated with 1.6 L of demineralized water. Mancozeb 75% WP broad spectrum fungicide (2 g/L) was first used to pre-treat the seeds, then dried and sown. 300 ml of demineralized water was used to irrigate the pots 7 days after sowing (DAS). Germinated plants were examined at 14 DAS and seedlings were thinned to keep one plant per pot.

At 14 DAS, pots were fertigated with 0.5 L of Hoagland nutrient solution (Table S10) formulated in demineralized water with three different N levels: N100 (2 mM Ca $(NO_3)_2$)-control/optimal N strength, N25 (25% of the full nutrition, i.e., 0.5 mM) and N10 (10% of full nutrition, i.e. 0.2 mM). All plants were fertigated once every week for 17 times (between 16 *h*00-17 h30) in a manner that allows complete absorption of the nutrient solution by the growth medium without any leaching from the pot throughout the experiment. The three N levels were determined following a test of 5 N levels (N100, N50, N25, N10 and N0) in 9 accessions (Table S11). This showed that N10 was more appropriate than N0 as the lowest viable N level treatment and that N25 allowed greater distinction among accessions as a low N level (yield per plant performance at N50 and N25 were comparable). At maturity, panicles were collected, threshed, seed grains collected, sun dried and stored for the study.

Trait assessment and derivation of indices

Sixteen (16) agronomically significant and yield related traits and 5 derived index traits were assessed at three N levels (2 mM- N100, 0.5 mM- N25 and 0.2 mM- N10) using a full cycle potted experiment of the 142 *S. italica* accessions (Supplementary Table S1; Supplementary Table S2) chosen from a previously reported core collection [20]. A total of sixteen agronomic traits were measured, and data was collected for three replications per accession. A one-way ANOVA analysis was performed to evaluate the relative contribution of the genotype, N dose and their interaction towards the trait performance (Table S13). Five index traits, namely: stability index (SI), tolerance index (TOL), mean productivity index (MPI), geometric mean productivity (GMP) and stress susceptibility index (SSI) were used to further evaluate the differences in trait performances due to any two N conditions (Table S4). Broad sense heritabilites ($h^2 = \sigma_g^2/\sigma_p^2$) of major traits and their indices were calculated (Table S13), where σ_g^2 and σ_p^2 are variances due to genotype and phenotype, respectively.

Grain N and C content analysis

Approximately, 5 mg of the powdered grains were used for CHN analysis (CHNS (O) Analyzer, Italy, FLASH EA 1112 series, Thermo finnigan) using the method elaborated by Dumas [21]. N and C contents from each genotype was obtained as percentages of the sample weight studied with three biological replications per sample.

SNP genotyping and sequencing

Leaves from 4 weeks old plants were used to isolate DNA using the Cetrimonium bromide (Rogers and Bendich, 1985). Post RNAase treatment (Fermentas, USA), the isolated DNA was checked for integrity and then quantified through 1.2% agarose gel electrophoresis and NanoDrop 1000 (Thermo Scientific, USA), respectively. Double digest restriction associated DNA (ddRAD) and Illumina HiSeq4000 platforms were used to genotype and sequence the samples, respectively (Peterson et al., 2012) (Agri-Genome Labs Pvt Ltd, Hyderabad, India). Raw FastQ reads were demultiplexed with only one mismatch to obtain reads for each sample and RAD tags were used to filter the data. 5′ and 3′ ends of the reads were trimmed along with the removal of Illumina adapters (Cutadapt v 2.3), while Bowtie2 (version 2–2.2.9) was used to align trimmed sequences to the reference genome catalogued in the phytozome 12 database version 2.2 at default parameters (https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Sitalica).

Furthermore, bcftools were used for filtering reads based on their depth and quality while sequence alignment map (SAM) tools (version 1.6) were used for variant calling analysis.

Assessing population structure and linkage disequilibrium

STRUCTURE version 2.2 software [22] was used to perform model based population structure analysis wherein Burn-in and MCMC were set as 50,000 and 100,000 respectively. We employed admixture model with five iterations for each run and assumed 2–10 sub-populations, with the real number of determined sub-populations by employing the delta K method [23] through an online tool STRUCTURE HARVESTOR [24]. A genotype was assigned to a specific sub-population when it had ≥ 80% probability of affiliation while those with < 80% of the value were considered " admixtures". Previous information on chromosome- and genome-wide LD [25] were also used in the analysis. The genetic relatedness of the individuals in the panel was ascertained by clustering the filtered SNPs using the phylogenetic tree construction tool implemented in TASSEL v5 (neighbour joining clustering method) and visualizing the same using the Archaeopteryx tool [26] implemented therein under all default settings.

Genome wide association analysis

For genome wide association study (GWAS), a minor allele frequency (MAF) of > 5% and missing data of < 30% were fixed as the basic cut-off values from a total of 29,045 SNPs by implementing the filter feature within Tassel 5 software [27]. We employed fixed and random model circulating probability unification (FarmCPU) package [28-30] for genome wide association which has been regularly used for many crop/cereal studies in the recent years [31-34]. The tool effectively eliminates issues arising due to kinship, population structure, multiple testing therefore making it one of the best models for association mapping currently available [28-30]. Kinship matrix is inbuild in FarmCPU and three PCA were employed (K + PCA model) for GWAS analysis. SNPs with a $p < 0.001$ were deemed significant SNP-trait associations (STAs) followed by p-value adjustment via Bonferroni correction (threshold set at 0.01). Quantile-quantile (Q-Q) plots were used to show how the expected and observed p-values are distributed and fit into the population structure model.

A set of 16 major traits were analysed at three N levels with 10 derived index traits from each major trait (5 indices/ main trait of N10-N100 and N25-N100), totalling 208 traits (Table S2). Broad sense heritability for all traits were found to be > 0.8 (Table S13).

Identification of functional genes proximal to trait specific STAs

S.italica genome 2.2 (available from Phytozome v12, https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias = Org_Sitalica) was used to identify genes proximal to SNPs related to significant STAs (for all traits) within the intervals of 0–1 kb, 1–5 kb, 5–10 kb, 10–20 kb, 20–50 kb and 50–100 kb distances from the SNP position in either direction. A distance of 20 Kb along the chromosome was considered a standard window to look for genes positioned proximal to trait associated SNPs for downstream analysis.

Estimation of transcript abundance of genes linked to grain number responsive SNPs

SNPs found to be significantly associated with GPP traits in the study located within the LD decay distance of 177 kb as previously reported in the crop [35] were considered as prime landmarks for identification and assessment of genes linked to grain number responsiveness in the *S. italica* genome. Based on the above, we identified three pairs of SNPs (CS3.46666559:CS3.46708881,CS4.37893830:CS4.37893921, S8.30225088:CS8.30225110) and profiled the expression of genes located 25 Kb upstream and downstream to them within *Setaria italica* genome (available from Phytozome v12, genome version 2.2, https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Sitalica) for qRT-PCR based assay (Table S7). A similar approach for identification of putative genes related to nutritional traits in *S. italica* is reported [35]. For this purpose, nine GNR and GNNR accessions were grown at low N (N25) and optimal N (N100) conditions as previously described and panicles were harvested at the early stage of panicle development when the spikelet organization of the inflorescence is decided (grain number), just before the onset of anthesis. Collected samples were immediately frozen in liquid nitrogen and stored at −80 °C. Total RNA was isolated using Spectrum Plant Total RNA kit (SIGMA), visualized in 2% agarose native gel, quantified using NanoDrop™ 1000 Spectrophotometer followed by reverse transcription using Verso cDNA Synthesis Kit (Thermo Fischer) as per the recommended guidelines. qRT-PCR assay was performed by using the Power SYBR Green chemistry (Thermo Fischer, USA) and employing the QuantStudio Real-Time PCR (qPCR) for assessing the relative transcript abundance of the target genes between samples and N conditions with three biological and two technical replications. *S. italica* actin gene (*ACT2*) was used as suitable endogenous control previously established [36] for the crop. Exon spanning primers for the target genes (Table S9) were designed using the NCBI Primer-BLAST online tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/)

Protein family and domain analysis

Amino acid sequences of proteins were obtained from the gene view tool using the gene ID compatible with *S. italica* annotation available in Phytozome v 12. The selected sequences were searched within the Pfam database (http://pfam.xfam.org/) and results from only the significant Pfam searches (sequence alignment and hidden markov model-based analysis) under default setting were included in further analysis.

Statistical analysis

Nitrogen dependent trait performances were measured and visualized using the 'dplyr' R package [37] while variances were evaluated using the analysis of variance (ANOVA) function "aov ()" analysed using R (R studio version 1.2.5001) [38]. Linear model regression analysis was accomplished using the ggplot2 R package [39] and ggpubr package (v 0.3.0) [40] with dependencies while data analysis and plotting for multi-trait Pearson's correlation was performed using the "ggcorr" function within the "GGAlly" package (v2.0) [41]. Plots showing contrasting trait dependent and N specific responses in GNR/GNNR were plotted using the "ggline()" function under "ggpubr" R package. Normal distribution of traits were ascertained by the Shapiro-Wilk test using "shapiro.test()" present natively in R. Scatterplot ellipses were plotted using the "ggplot2" R package using the stat_ellipse() function.

Results

Grain number largely drives N dependent yield performance in S. Italica

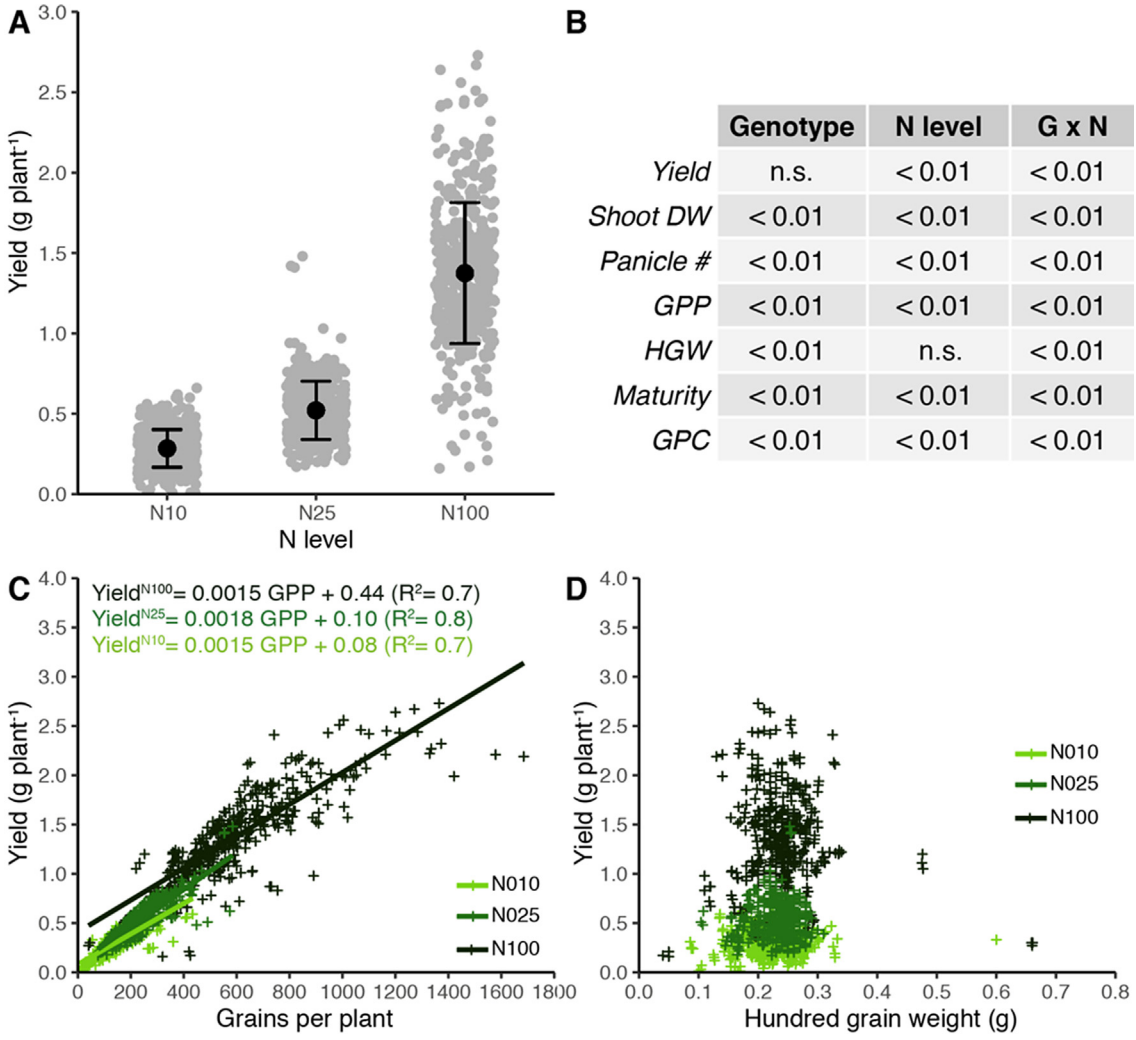
While hundred grain weight (HGW) did not significantly vary between N levels, most of other traits showed a substantial N response (Fig. 1B). Grains per plant (GPP) and yield showed a positive response to increased N accessibility for the majority of accessions (Fig. 1). There was a larger range of yield performance at high N (N100: 0.2 to 2.727 g), in comparison to low N dose settings (N10: 0.035–0.597 g; N25: 0.162 to 0.985 g), implying that the resultant yield plasticity to N increased availability exists in the population (Fig. 1 A), despite having comparable variance at all N levels (N10: 0.41; N25: 0.34; N100: 0.31). All other traits (shoot dry weight, panicle number, grain protein content, maturity time included) showed a noteworthy genotype by N level interaction (Fig. 1B).

Analysis of N dependent yield performance showed that the trait was positively and strongly associated with GPP across all three N levels (R^2 = 0.9, p < 0.01; Fig. 1C, Table S3) unlike HGW (R^2 = 0.01, p < 0.01, Fig. 1D; Figure S1). This indicates that the observed variations in yield are strongly affected by GPP and much less so by the weight of individual grains, regardless of N levels. Notably the GPP range is much higher at high N (40–1700 grains per plant compared to 5–584 grains per plant at N10 and N25). Furthermore, we observe that the increment in GPP is reliant mostly on grain number increase per panicle (Figure S2A) and less on the panicle number (Figure S2B). In *S. italica* multiple panicles originate from the same stem (secondary panicles) which mean that more panicles do not translate into more tillers.

Harvest index (HI), grain per panicle (GPPn; Figure S2C, D) and to a smaller degree shoot dry weight (SDW; Figure S2E) were also positively connected with yield, suggesting that partitioning of N to the panicle may contribute to increased yield. Interestingly we observed a negative (R^2 = 0.11; p < 0.01) association between yield per panicle and panicle number signifying a trade-off between overall yield capacity of a panicle and panicle number (Figure S2F). However, the absence of any correlation between overall plant yield and panicle number in this context suggests some degree of compensation for the negative correlation stated above.

Defining indices for N responsiveness

We employed five (5) derived indices of the traits to specifically appreciate the genetics of N response in *S. italica*. We define these indices focusing on yield as the major trait (Fig. 2A). Yield at N100 is weakly correlated to yield at N25 (R^2 = 0.152) and N10



	Genotype	N level	G x N
Yield	n.s.	< 0.01	< 0.01
Shoot DW	< 0.01	< 0.01	< 0.01
Panicle #	< 0.01	< 0.01	< 0.01
GPP	< 0.01	< 0.01	< 0.01
HGW	< 0.01	n.s.	< 0.01
Maturity	< 0.01	< 0.01	< 0.01
GPC	< 0.01	< 0.01	< 0.01

Fig. 1. High nitrogen availability leads to overall yield increased due to increased grain number per plant in *S. italica* grown. (A) Yield per plant was measured for three biological replicates per genotype and N level, filled circles represent individual plants. Data are mean yield ± SD. Significant differences were observed amongst mean yield at each N level (Tukey , $p < 0.01$). (B) Table showing the results of the ANOVA for specific traits (additional data shown in Supplementary Table S5), data as p value for each factor (genotype and N level) and their interaction. GPC: grain protein content. (C) Yield correlates positively with the number of grains per plants, irrespective of N level. (D) Yield does not correlate with the hundred grain weight. For C and D, each cross represents an individual plant.

(R^2 = 0.056), supporting the idea that it is important to measure responsiveness under different N conditions (Fig. 2B). The tolerance index (TOL) simply indicates in real terms (i.e., g per plant) the yield gain under high N conditions compared to low N conditions and appears to be the best representative index for N responsiveness. Therefore, a higher TOL value indicates greater yield increase after addition of N (from N25 or N10 to N100) whilst low TOL indicates a small value. The mean productivity index (MPI) and the geometric mean productivity (GMP) provide a measure of the mean yield over the range of N levels tested. While MPI is highly correlated to yield at N100 (R^2 = 0.94) and less so to yield at N10 (R^2 = 0.22), GMP is correlated with both indicating that it is less affected by extreme values and perhaps a better representation of an overall yield under contrasting conditions. The stability index (SI) is a ratio that offers a direct comparison between yield under high and low N. In this case, a very low SI (<1) indicates higher yield under high N conditions compared to low N conditions. SI tends to be negatively correlated with other indices (Fig. 2B). The stress susceptibility index (SSI) represents a similar index to SI that is normalised to the overall yield mean of the population under both high and low N. We also calculated N use efficiency (NUE) as the ratio of grain produced per unit of N provided. NUE under each N level is highly correlated to yield for each N level (Fig. 2B). As the NUE calculation simply corresponds to division of yield by a constant for a specific N treatment, this measure was not used in the further analyses.

SNPs are unevenly spread across chromosomes and form distinct phylogenetic clades in S italica

A total of 142 *S. italica* accessions were genotyped with a set of 29,045 high quality SNPs. Heterozygosity of both SNP markers and individuals were within acceptable limits (<25%) (Figure S3A; B). SNP markers showed uneven distribution in nine chromosomes with an overall average of 125.73 SNPs/Mb in *S. italica*. Chromosome 8 and 9 had the highest (235.6 SNPs per Mb) and lowest (87.45 SNPs per Mb) densities, respectively. Overall, chromosome 1 was found to have their maximum evenly distributed densities (Figure S3C). Among chromosomes, the mean polymorphism information content (PIC) ranged from 0.125 to 0.20, the least and maximum values lying in Chr 9 and 8, respectively (Table S3). Population structure analysis showed that about half of accessions were admixed (75 out of 142), and residual 67 accessions being

A

Trait	Description	Adapted From
Tolerance index (TOL)	Y_{HN}-Y_{LN}	Rosielle and Hamblin (1981)
Mean Productivity Index (MPI)	$(Y_{LN}+Y_{HN})/2$	Hossain et al. (1990)
Geometric mean productivity (GMP)	$\sqrt{(Y_{LN}*Y_{HN})}$	Ramirez and Kelly (1998)
Stability index (SI)	Y_{LN}/Y_{HN}	Bouslama and Schapaugh (1984)
Stress Susceptibility Index (SSI)	$\{1-(Y_{LN}/Y_{HN})\}/\{1-(\text{mean } Y_{LN}/\text{ mean } Y_{HN})\}$	Rameeh (2015)
Nitrogen use efficiency (NUE grain)	Grain dry weight/ N supplied per plant	Moll et al.(1982)

Y_{LN}=Mean value in stress(N10/N25); Y_{HN}= Mean value in control(N100); Y_{LN} mean & Y_{HN} mean= average value of the trait in all genotypes at stress(N10/N25) and control(N100) conditions, respectively.

B

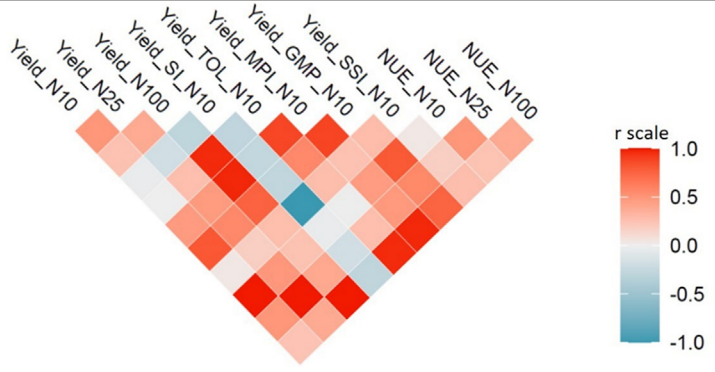


Fig. 2. Index traits of a major trait measure different aspects of its N responsiveness. (A) Tabulation to show the details of all trait indices measured for each of the 16 major traits analysed for 142 *S. italica* accessions (B) Correlation plot to show coefficient of correlation (r) between and within Yield traits at three N levels and its index derivates on one hand and NUE major trait on the other. For ease of understanding and visualization, only N10(Y_{LN}) was considered for plotting index traits of Yield. Since NUE shows strong correlation with Yield main traits at all three N levels, index derivatives for the trait were not plotted. Values are a mean of three replications. Mean data of phenotypic performance of all traits and their indices are available in Supplementary Table S6.

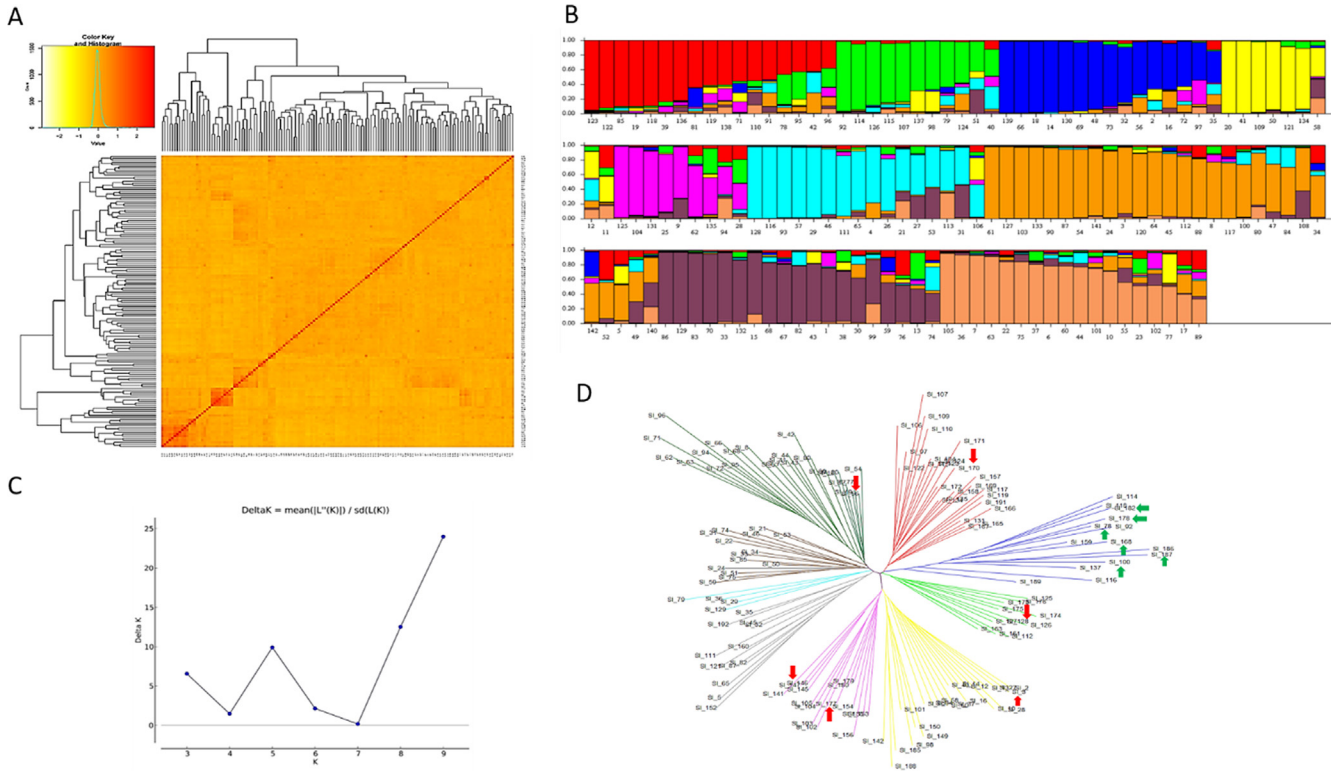


Fig. 3. Population structure in GWAS panel for 142 accessions of *S. italica* natural population and 29,045 SNP data (A) Kinship matrix plot to indicate relationship between accessions. Names in the X-axis and right Y-axis indicate individual accessions and the left Y-axis indicate dendrogram to indicate phylogenetic relationship based on the SNP data (B) Admixture bar plot showing nine subpopulations and assignment of membership. X axis correspond to individual accessions showing their distribution across nine (9) sub-populations within the panel while the Y axis represents Q value depicting affiliation probabilities for assignment within a sub-population (80% or more). (C) Optimization of sub-populations numbers (K value) ranging from from K = 1–9 to fix best possible clustering for foxtail millet accessions (D) Radial cladogram (phylogenetic tree) of the accessions using Neighbour joining clustering method implemented through Archaeopteryx tool (Han and Zmasek, 2009) showing 9 sub-populations in distinct clusters.

randomly spread over 9 sub-populations clustering under nine discrete phylogenetic clades (Fig. 3D).

GWAS identifies new genetic loci linked to N response plasticity

We identified 68 SNP (marker) trait associations (STAs) for the traits measured and their indices from 16 major traits (Table S5; Fig. 4). These STAs comprised of 59 unique SNPs significantly associated with ten major traits (P value set at 5 e^{-07}, Bonferroni correction = 0.01) [42] and related indices: D50F (days to 50% flowering), GPP, grain C/N ratio, grain C, leaf chlorophyll content, panicle number, HGW, days to panicle emergence, days to maturity and shoot length (Table S5; Figure S4). These SNPs were spread throughout the genome, with chromosomes 8 and 9 containing the most (24) and least number (2) of significant SNPs, respectively (Fig. 4). We found that all 68 STAs are highly trait specific (i.e., having no overlap with other major traits) although some SNPs could be associated with more than one trait index within a given major trait (Table S5). Intriguingly, we found more unique STAs associated with index traits (55) than with major traits (13) suggesting that more genetic loci are linked to traits that measure differences in N response due to N availability (N responsiveness) compared to those that don't (Table S5, Fig. 4).

Additionally, we examined the incidence of genes adjoining the SNP loci based on the genes annotated in the *S. italica* genome. Within 50 Kb of such SNPs, we identified a total of 272 genes based on their closeness to nearest genes (protein coding) in six distance ranges of 0–1 Kb, 1–5 kb, 5–10 Kb, 10–20 Kb and 20–50 Kb (Supplementary Figure S5; Table S6). Additionally in this respect, chromosome 8 was found to have the highest gene density, followed by chromosome 5.

Specific SNPs linked to nitrogen responsiveness of grain number exhibit haplotypes

Even though we did not find any significant SNP association with yield trait (or its indices), significant STAs could be identified for yield related traits such as GPP, panicle number and HGW (Table S5). Overall, GPP traits showed the greatest number of detected significant associations (a total of 26 associations from 59 SNPs out of which 17 STAs associated with GPP index traits) suggesting that N responsiveness of the trait is significantly regulated at the genetic level (Figure S4). Furthermore, among the 22 unique GPP linked SNPs, we identified three (3) SNP pairs (CS3.46666559:CS3.46708881, CS4.37893830:CS4.37893921, S8.30225088:CS8.30225110) which are linked to GPP index trait, lie within the linkage disequilibrium(LD) decay distance estimated previously [35] and show haplotypes for their corresponding linked traits, suggesting that allelic variation in these SNPs has significant implications for variability for linked N responsive traits MPI_GPP_N25, MPI_GPP_N10 and GMP_GPP_N25 (Fig. 5).

Subsequent analysis to identify their proximal genes (upstream and downstream 25 Kb) revealed the presence of 17 unique genes (Table S7) out of which four genes (*Seita.3G363300*, *Seita.3G364000, Seita.4G260600, Seita.4G260700*) are unannotated as per Phytozome v2.2. The remaining genes broadly fall in the category of acid phos-

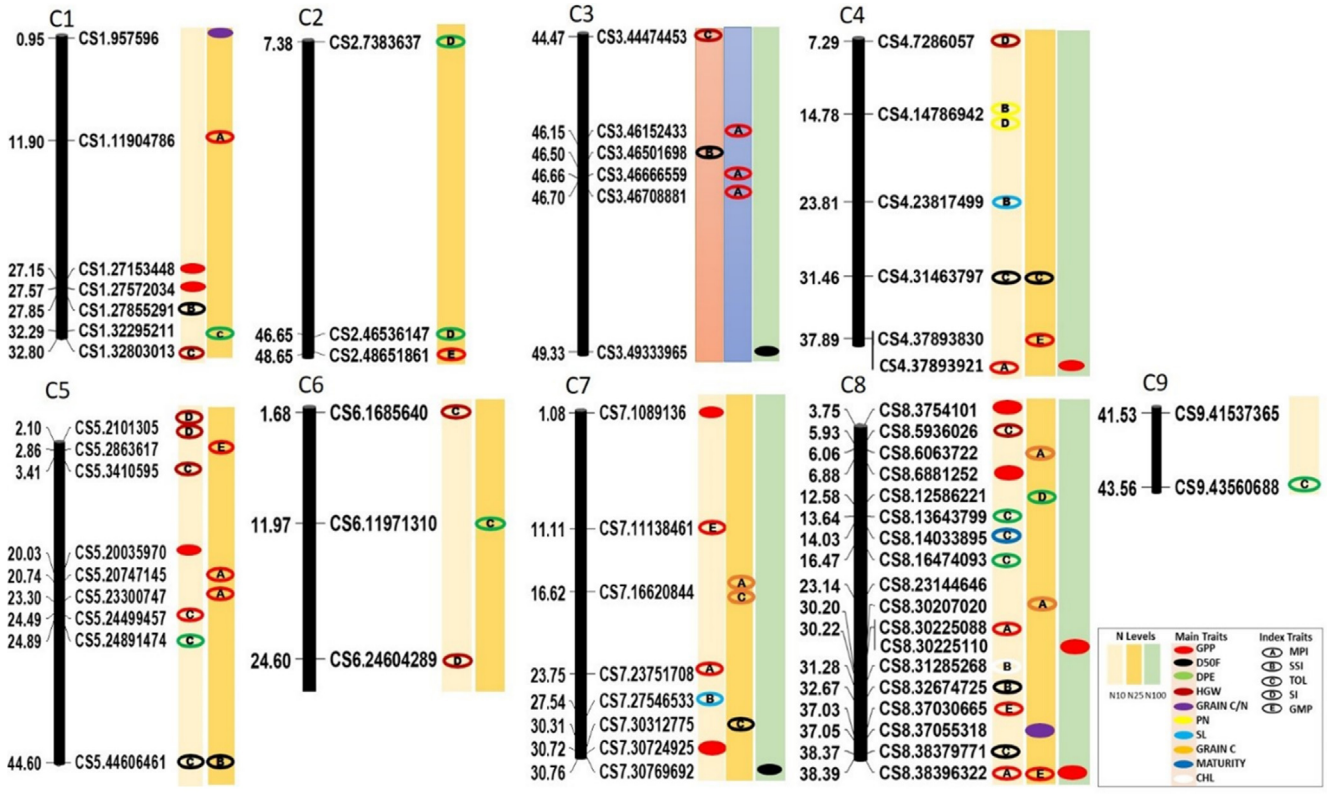


Fig. 4. SNPs linked with N dependent index and major traits are spread across nine chromosomes in *S. italica*. Numbers adjoining to the SNP location specify SNP position in megabases (MB) and discrete oval shapes as single SNP. Oval shape outline and fill pattern represent a major trait, N level type and trait indices, respectively as indicated in the figure. Phytozome version 12 (genome V2.2) was used to map SNPs. MPI: Mean productivity index; SSI: Stress susceptibility index; TOL: Tolerance index; SI: Yield stability index; GMP: Geometric mean productivity.

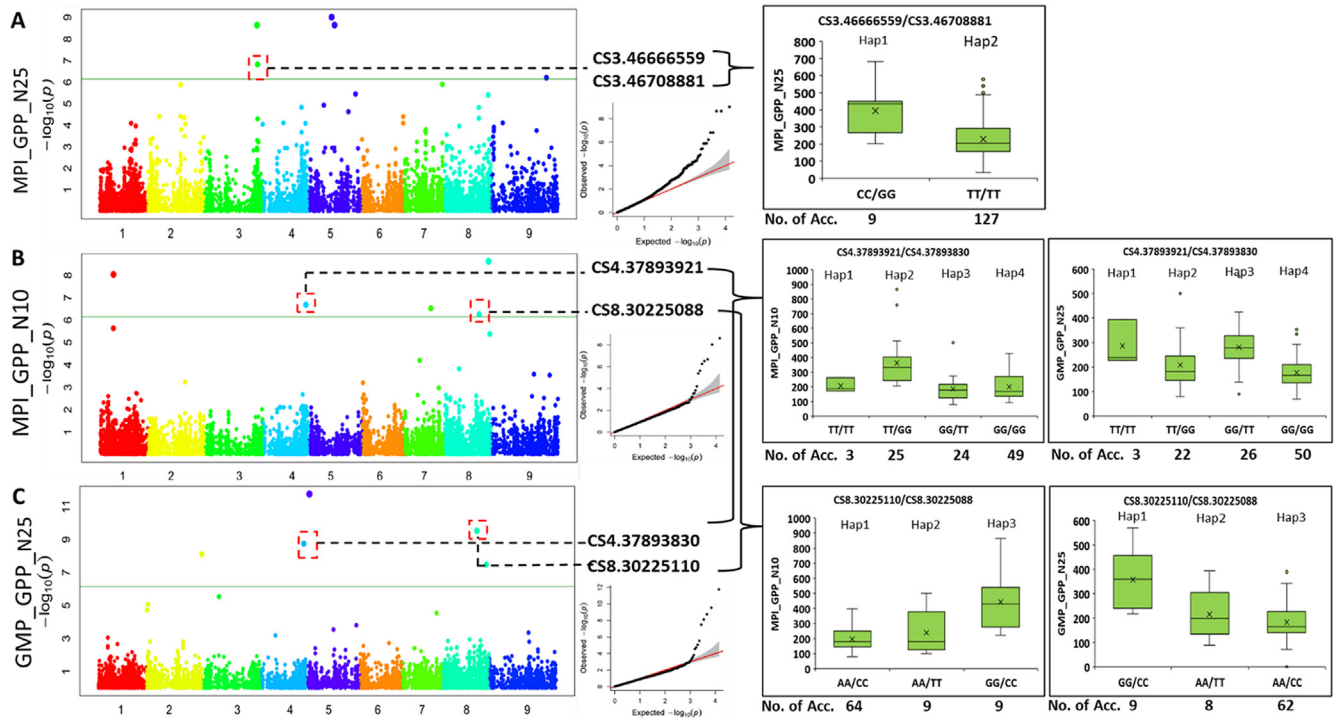


Fig. 5. SNPs in LD and linked to grain number index traits show haplotypes. Polymorphism of the SNP and phenotypic performance variability for main GPP traits are correlated. In each of the panels A-C, left section represents the Manhattan plot. their respective - Quantile-quantile (QQ) in the middle for a given GPP trait while the box-plots on the right show haplotype for strongly trait linked SNPs in LD and their corresponding phenotypic variation within the population. The numbers along the × -axis in the Manhattan plot represent chromosome number while the coloured dots indicate SNPs specific to a given chromosome. The y-axis values show negative logarithm of the association P-value with the threshold p-value of significance indicated by a horizontal line. Box plot defines the mean data from three replications wherein the bold line within each box show the median value, "x" as the mean while the region between the median and edges of the box in both directions represents values that are up to 25% more or less than the median value. The whisker lines outside the edges in both sides comprise the remaining 25% of extreme values and the dots represent outliers.

phatases (*Seita.3G363500*, *Seita.3G363600*), kinases and kinase activators (*Seita.3G363700; Seita.3G363800*), nucleic acid binding and chromatin remodelling (*Seita.3G363900, Seita.8G160300, Seita.8G160400*), cytoskeletal organization (*Seita.3G364100*), hormone biosynthesis and secondary metabolism (*Seita.4G260400*), protein folding (*Seita.8G160500*), ligand-binding and ion channel activity (Seita.4G260500), glucosidase activity (*Seita.8G160600*).

N dependent grain number responsive and non-responsive accessions in foxtail millet

Based on our observation that N dependent yield performance is largely driven by grain number in *S. italica*, we explored whether specific accessions exist in the population which exhibit contrasting grain number responsiveness (GPP_TOL N100-N25) and at the same time are consistent with the allelic variation of SNPs linked to grain number responsive traits (GMP, YI, MPI, SI and TOL). Grains per plant (GPP) at N25 was used to calculate grain number responsiveness since it is appropriately placed to induce N deficiency whilst still allowing ample N for successful grain filling (than at N10) and therefore yield in majority of accessions. Our analysis showed that accessions SI 100, 168, 178, 187, 78,182 and SI 128, 146, 170, 177, 3, 56 show very high and low values for the trait, respectively and exhibit consistent difference in panicle architecture, especially with regard to awn distribution and their lengths (Fig. 6A, B). Furthermore, we observed that these two groups of accessions largely maintain the same allelic variation for six GPP linked SNPs (Table S8) that lie within LD decay distance of 177 kb (CS3.46666559, CS3.46708881, CS4.37893830, CS4.37893921, CS8.30225088, CS8.30225110), previously established for the crop [35]. Such grain number responsive (GNR) and grain number non-responsive (GNNR) accessions were analysed to further examine the basis for N responsiveness in *S. italica*.

Apart from the differences in their capacities to utilize additional N to produce grains, GNR and GNNR also exhibit characteristically different shoot dry weights, yields, harvest indices, panicle dry weight and longer flowering times at least under high N (N100) (Fig. 6C-G) determined chiefly by the ability of GNR accessions to yield more grains. To further dissect their differences in N dependent yield plasticities, we measured four derived indices related to yield, grain number (GPP), hundred grain weight (HGW) and harvest index (HI) based on their respective trait performances at low and high N levels (N10-N100; N25-N100) (Figure S6). Two low N levels (N10 and N25) were considered for the analysis to enable a better understanding into how such trait plasticities play out at very low (N10) and low (N25) N levels against a common control (N100). We observed that except for HGW, indices for all the remaining traits (Yield, GPP and HI) differ significantly between GNR and GNNR genotypes while maintaining the same pattern of behaviour when considering very low to high N (N10-N100) and low to high N (N25-N100) comparisons (Figure S6). Similar to the overall population, yield patterns in the two types shows strong positive correlation with grain number while none were observed for HGW (Figure S7A). Comparative analysis of all these phenotypic trait classes suggest that most vary significantly as a function of genotype and N level (Figure S7B).

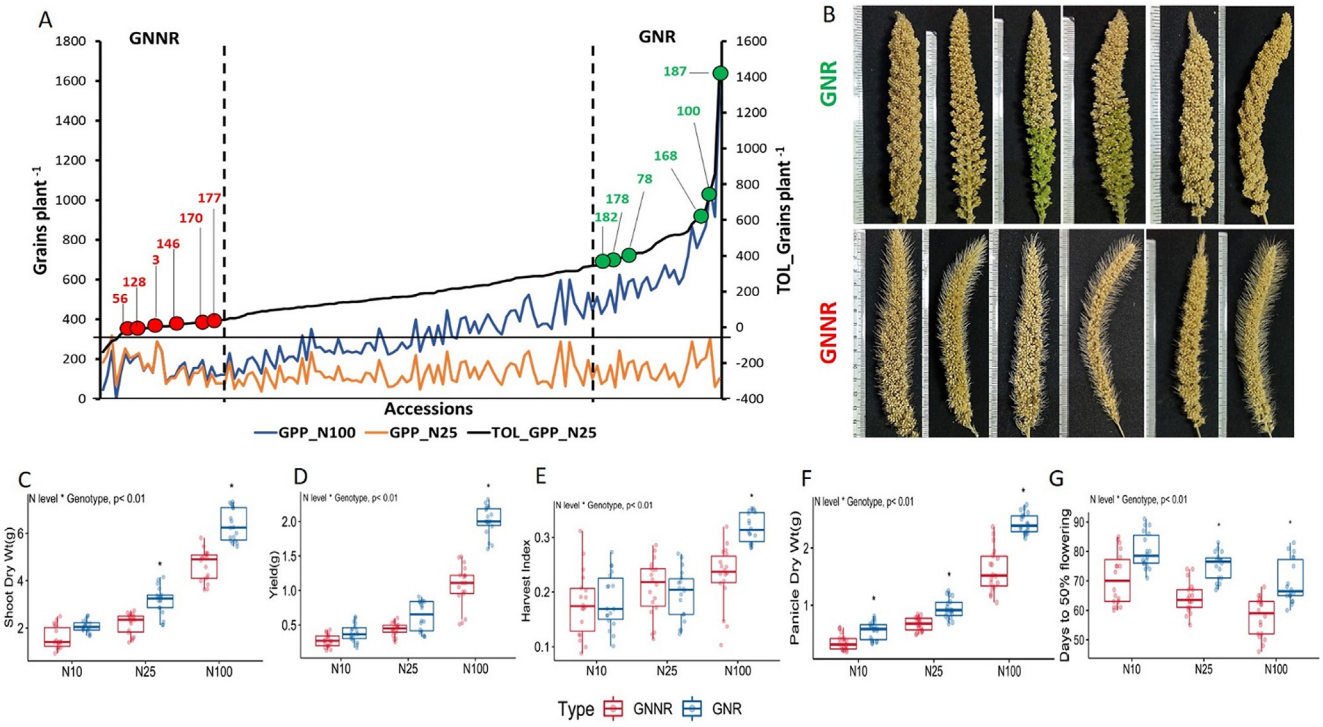


Fig. 6. Grain number responsive (GNR) and non-responsive (GNNR) *S. italica* accessions have opposing grain number tolerances to contrasting N availability. (A) For all the accessions analyzed (on the x-axis), data are plotted from the mean grains per plant (GPP) on left y-axis at N25 (orange lines) and at N100 (blue lines). The grain number tolerance at N25 (GPP at N100-N25) is represented by black line and scaled on the right y-axis. Dotted lines indicate GNNR and GNR limits. Each of the six red and green filled circles indicate specific GNNR and GNR accessions, respectively which largely share the same allelic form of significantly GPP linked SNPs namely CS3.46666559, CS3.46708881,CS4.37893830,CS4.37893921,CS8.30225088, CS8.30225110. Panel B show panicle architectures of these accessions at N100. Panels C, D, E, F and G show data for shoot dry weight (SDW), yield, harvest index (HI), panicle dry weight and days to 50% flowering (D50F) for each of these genotypes at three N levels, respectively. Data shown as the mean +/- SE of three replicates from six GNR or GNNR accessions. Differences due to N level, genotypes and their interaction were analysed using two-way ANOVA followed by Tukey Test with differences indicated by asterix (*). GNNR accessions: SI 128, 146, 170, 177, 3, 56; GNR accessions: SI 100, 168, 178, 182, 187, 78 (Table S6). Error bars show standard error.

These observations suggest that the two groups of genotypes have discrete patterns of phenotypic responses to N provisioning that are consistent within each group and provides evidence that N responsiveness between these two is significantly different across multiple derived interpretations of yield traits. Focussing on these subsets of accessions for further analysing the mechanism of the N dependent yield responses may therefore provide new insights that may still be applicable to the population under study.

Genes linked to GPP associated SNPs have different transcript abundances in GNR and GNNR at low and high N levels

To examine if expression patterns of genes linked to GPP traits are differentially expressed in genotypes with high and low plasticities (GNR and GNNR), qPCR assays of genes proximal to GPP associated SNPs (in LD) were performed (Table S9), using the approach previously implemented in the crop [35]. Three accessions with similar panicle emergence times were chosen from each of the GNR (SI 100, SI 178 and SI 168) and GNNR (SI 128, SI 56 and SI 56) groups to access transcript abundances of 17 genes at high N (N100) against low (N25) N condition to measure their N responsiveness. We observed that three genes (as per Phytozome v12) namely *Seita.3G363700*- Diacyl glycerol kinase; *Seita.8G160400*- a DnaJ domain containing protein; *Seita.8G160500- T*-complex protein 1 (CCT8) out of 17 genes showed largely consistent and distinct expression patterns within and between the two groups, respectively (Fig. 7A). Sequence analysis of their encoded proteins (significant PFAM match, [43]) indicates that Seita.3G363700 has all the domains necessary for diacylgerol kinase activity (with accessory, binding and catalytic domains) while both Seita.8G160400 and Seita.8G160500 are chaperone family proteins containing Hsp40 (DnaJ domain) and Hsp60 (TCP-1/cpn60) proteins, respectively (Figure S8). Seita.8G160400 also contains two DUF (domain of unknown functions), and exploring any connection between them and the DnaJ domain with regard to protein activity will be greatly insightful vis-à-vis N responsiveness. While greater availability of N causes a relative increase in their transcript accumulation in GNR, the opposite is true for GNNR thereby indicating commonality in their regulation leading to potential N responsive processes in a genotype specific manner. Furthermore, we also observed an overall difference of type of correlation between grain number tolerance (TOL_GPP) and expression of all the 17 genes in GNNR/GNR. While we find an overall positive correlation in the case of TOL_GPP/GNR, the same is not true for TOL_GPP/GNNR (Figure S7B), suggesting that these genes largely associate with N responsiveness in a genotype specific manner

Discussion

Enhancement in yield performance has been limited in *S. italica*, especially in comparison to staple cereal crops like wheat, rice or maize. However, the crop can potentially play a larger role in many agro-ecosystems worldwide, including sub-Saharan Africa and India. An important feature that has pushed rise in yield output the major crops is the simultaneous use of synthetic N fertilisers and selection of newer varieties. Intensive agriculture has largely driven selection of varieties that performed better at optimal N conditions[9] and currently information on how crop plants respond to increasing N availability, though crucial is limited. Fill-

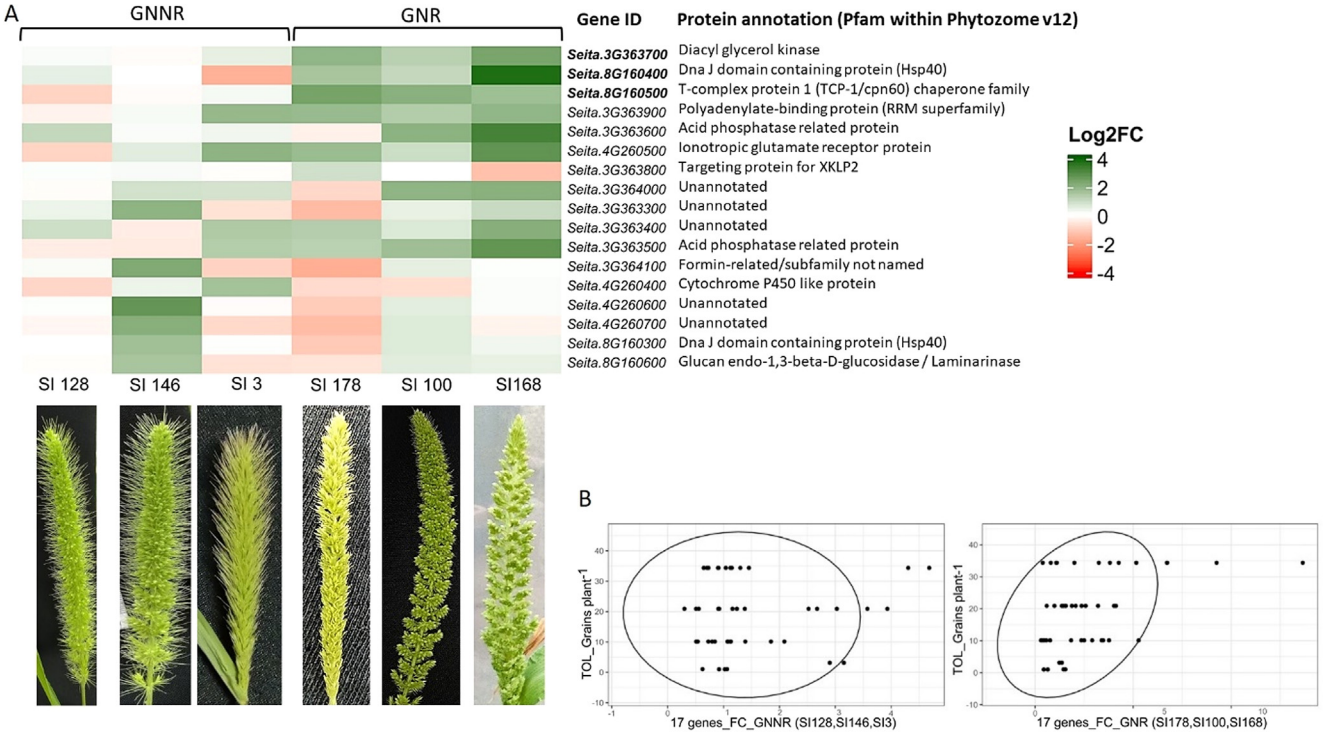


Fig. 7. Transcript abundance of genes linked to SNPs associated with grain number index trait in GNNR and GNR panicles in response to increased N. (A)Transcript abundance of the genes was assessed at high N (N100) against low N (N10) from fully expanded panicle tissues with established floral architecture to evaluate their N responsiveness. All Log2FC values ≥ 1 have a p value of ≤ 0.05 between replicate measurements of gene expression in test (N100) and reference (N10) tissues. Values represent mean of three biological and two technical replications for each gene analyzed. Representative panicle images shown below each accession are at 4th day post emergence (before anthesis) grown at N100 (control). (B) Scatterplot with ellipse to show the overall differences in the correlation between TOL_grain number per plant and absolute fold change (N100 v N10) in transcript abundances of GNNR and GNR in panicle tissues presented in (A).

ing this gap can potentially help selection of varieties that profit from N input in order to yield more and limit N loss to the environment. In this paper, we dissected the response of *S. italica* plants to increased N availability and identified potential genetic markers for high N responsiveness, thus demonstrating a newer approach for variety selection in crops.

Grain number per plant largely regulates nitrogen directed yield increase in S. Italica

Typically, yield is determined by the number of grains produced and their weight per plant. The influence of grain number trait in effecting yield trait in cereal crops is well recognized [44-46]. In C_4 crop like maize, yield is primarily and positively dependent upon the kernel number and number of ears per plant [47,48] although the overall N dependent yield gain is determined by both kernel number and kernel size in the crop [49]. In millets like Sorghum, N dependent yield is largely driven by panicle number, grain number per plant and grain weight [50-52] while it is the panicle number per unit area which largely determines the yield performance (up to 65%) in pearl millet under nitrogen and water stress conditions [53,54]. This indicates that understanding the plasticity of N-dependent response of total number of grains produced (which is dependent on the grain number per panicle and the panicle number), has additional value for C_4 crop species beyond *S. italica*.

Our study in *S. italica* shows that N-dependent increase in grain number per plant has a strong impact on the yield increase in *S. italica*, and that this is mainly effected by an increase in grains per panicle rather instead of rise in the number of panicles per plant. N has a strong effect on branching response in many species [55-57] which would relate to higher panicle number in *S. italica*. A further effect of N supplementation, particularly at earlier developmental stage, is the increased number of flower per panicle [58] which is facilitated by cytokinin amounts in the developing panicle in rice [59]. In our study, we observed that the N-driven yield rise in the GNR type is fixed at (early) panicle developmental stage. Therefore, comparison of earlier developmental stage signalling of GNNR and GNR types at the initial stages of panicle growth may provide major insights on how N determines grain number variation in *S. italica*.

It is noteworthy that we did not observe any agronomic tradeoff to increased yield in terms of grain weight (Fig S6), which is greatly valued by breeders. This will allow identification of the economic N optima threshold for the crop thereby reducing N application rates without affecting crop productivity (grain yield) and positively impacting sustainable agriculture. This will however require the identification and utilization of genotypes that are more capable of translating acquired N into gainful and consistent yield performances with much lesser increment of N fertilizer input as exhibited by GNR genotypes.

N responsiveness is a heritable multigenic trait in S. italica

In the present study, we investigated the effect of higher N availability on yield response in *S. italica* which showed that N responsiveness is a valuable trait with strong genetic basis [9]. We observed that there is no strong correlation between the yield measured at N10 and the yield at N100, or between yield measured at N25 and N100 (Fig. 2). This indicates that for *S. italica*, the N responsiveness or increase in yield under high N conditions cannot be inferred from yield measurements conducted under low N conditions. Therefore, measurements under low and high N conditions are crucial. Likewise, GPP measured at N10 and N100, or N25 and

N100 do not share any correlation. Therefore, measuring yield under only low N does not offer information on the yield potential at high N level and *vice-versa*, yield measurement under high N does not provide information on the yield performance achievable under low N conditions. We have evaluated a series of indices here to estimate the yield gain achieved in the presence of N, with the TOL index being a good representative of N responsiveness *per se*.

N responsiveness is a trait which is heritable and can be mapped genetically, and therefore amenable to breeding programme. The complexity of the trait however is a major challenge as many of the STAs found associated with the trait did not overlap with STAs for major traits, signifying that the genetic basis for high N responsiveness differs from those determining major traits performance including GPP. Furthermore, we did not find any of the STAs close to known genes associated with primary metabolism. In *Arabidopsis,* plasticity of branching due to increased N in greatly responsive lines also less branches under low N and very high shoot branching under high N doses [56]. This is in contrast to our results in this study, where the extent of N responsiveness remains unpredictable when plants were grown under low N.

GWAS analysis highlighted the presence of three pairs of GPP index trait linked SNPs (CS3.46666559:CS3.46708881,CS4.37893830:CS4.37893921,CS8.30225088:CS8.3022511), within close proximity in the chromosome and the existence of correlation of their haplotypes (Fig. 4) with variation in trait performance in the population suggest that NR in the crop is genetically regulated and the underlying components of which are heritable and potential targets of crop improvement strategies. Further investigation of genetic components (SNPs and their proximal genes) pertaining to those linked to N dependent grain number responsive traits (GPP index traits) will be particularly useful to help identify their roles in regulating the trait and the mechanism of regulation thereof. The presence of a significant portion of trait linked SNPs within 3.5 kb upstream to their proximal genes (6 out of 12 GPP index linked genes) suggests that they are likely to have significant influence on their target genes leading to genotype dependent grain number NR (Table S9). Though many protein candidates are known to play a role in N sensing, there is still ample discussion about the molecular machinery underlying N sensing in crop plants [60]. In this regard, we observe that the genes *Seita.3G363700* (diacylglycerol kinase) and *Seita.4G260500* (Ionotropic glutamate receptor) positioned downstream to the GPP linked SNPs CS3.46666559 and CS4.37893830, respectively, have been previously implicated in either N sensing, N/C partitioning [61,62] or lipid metabolism [63] influencing yield response.

Genes related to grain number responsiveness are transcriptionally regulated in a N and genotype dependent manner

Gene expression studies to ascertain the transcriptional regulation of genes proximal to SNP (in LD and showing haplotypes in the population) linked to GPP index traits suggest that few of them are regulated differently in N responsive and non-responsive genotypes. Three genes *Seita.3G363700* (encoding a diacyl glycerol kinase- DAG), *Seita.8G160400* (an uncharacterized chaperone (Hsp40) protein containing a DnaJ domain) and *Seita.8G160500* (encoding T-complex protein 1 belonging toTCP-1/cpn60 chaperonin family) are noteworthy since they showed strong consistent upregulation (from 3 to 13 folds) in GNRs while remaining largely uninduced in their GNNR counterparts in response to N. Diacylglycerol kinases has significant role in lipid metabolism which is altered under high N conditions with low C [64] and perhaps differential activity of the gene in GNNR leads to altered partitioning of C under low N vs high N than in GNR. Higher expression of 'NUMBER OF GRAINS 1′ (NOG1) gene encoding enoyl co-A hydratase/isomerase (ECH)- a vital enzyme in fatty acid b-oxidation pathway was reported to enhance grains per plant [63]. Notably, lipids work as C source for fungi associated with plants in arbuscular mycorrhizal symbiosis [65], only under a low plant N status. Furthermore, DAGs are crucial for generation of phosphatidic acid in plants, a key signal transducer of lipid metabolism/signalling [66] and have been implicated in N sensing in Arabidopsis [67] with contingent effects on organ growth and development. The observed N dependent differential expression of its encoding gene in the developing panicles of the two genotypes in this study is likely to impact the growth and development of these tissues, potentially influencing the observed variation in grain number performance. Exploring how N regulates their behaviour will potentially provide novel insights on hitherto unexplored role of N on genetic regulation of yield responsiveness in cereals.

Plant cytokinin levels are known to be directly associated with N availability [68], thereby potentially modulating assimilation of N and C metabolisms [68,69]. Previous studies in tomato [70] showed that frameshift insertion-deletions (InDels) in two DnaJ encoding genes underlie the expression of a cytokinin oxidase/dehydrogenase gene responsible for cytokinin transport to leaves under higher N availability thereby suggesting their N responsive behaviour. In a previous study, DnaJ proteins have been shown to play important roles in photosystem II maintenance and hence the extent of carbon assimilation through photosynthesis [71] Furthermore, DnaJ/Hsp40 proteins have been implicated to act as transcriptional activators of many genes by binding with many transcription factors [72]. This indicates that differential transcript abundance of *Seita.8G160400* in two genotype groups identified in our study may mediate/regulate N dependent cytokinin metabolism differently leading to their observed differences in N dependent yield response in the crop. Further studies are however needed to substantiate this observation.

T-complex protein 1 subunit theta (CCT8) *Seita.8G160500* is a molecular chaperone which facilitates protein folding and is implicated in stem cell maintenance by transporting transcription factors and other proteins through plasmodesmata [73,74]. The distinct transcriptional responses of the gene (to elevated N provisioning) in the panicle between the two groups suggest that perhaps they target genes/components regulating stem cell maintenance differently potentially leading to differential abolishment of floral stem cell maintenance in the growing inflorescence and hence their different architectures. Furthermore, an overall higher correlation between the transcript abundance of all the 17 genes (N100 vs N10) and TOL-GPP in GNR indicate that they are largely N responsive. However, comprehensive molecular and physiological studies are required to fully explore how enhanced N availability and its perception relates to its transcript abundance and its consequences to inflorescence organization.

Conclusion

Identifying the minimal N amount for optimal yield is key to limit the undesirable ecological impacts of fertilizer dependent cereal cropping. Here we demonstrate that N responsiveness is an important trait to consider in achieving this aim. The present study provides the first exhaustive analysis in *S. italica* of the responsiveness of multiple agronomic traits to applied N and identifies a set of genetic loci strongly linked to N dependent grain number response. Of the putatively associated genes, some showed strongly differential expression in a N, genotype and temporal specific manner in the developing spikelet. The insights gained and resources generated in this will help identify promising N responsive accessions for use by breeders in devising sustainable crop improvement strategies. This study provides key avenues for comprehensive dissection of N responsiveness in the climate resilient

C_4 crop *S. italica* with a potential for translation in additional cereal crop species relevant to sustainable food security.

Compliance with ethics requirements

This research work does not contain any studies with human or animal subjects.

Data availability

Genotyping-by-sequencing data used in the study have been submitted in the NCBI-SRA (Sequence Read Archive) database with the submission ID SUB10121686 and Bioproject ID PRJNA751255.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was funded by Department of Biotechnology (DBT), Govt. of India (Grant No.: BT/IN/UK-VNC/42/RG/2014-15) and by J.C. Bose National Fellowship Grant of Department of Science and Technology (File No.: JCB/2018/000001) to M.P. A.R. B., R.G., and H.G. are supported by GCRF/BBSRC TIGR^{2}ESS programme (BB/P027970/1). T.B. acknowledges the financial support under the CSIR-SRA fellowship scheme [13(9152-A)/2021-Pool] while VJ acknowledges the DST-Inspire fellowship. The authors thank Anand Dangi, Technical Assistant, NIPGR, New Delhi for providing support towards execution of phenotyping experiments and data collection. Thanks are due to Dr. Stephanie Smith (Sainsbury Laboratory Cambridge University, Cambridge, UK) and Dr. Tina Barsby (NIAB, Cambridge, UK) for their observations and inputs. The authors are also thankful to DBT-eLibrary Consortium (DeLCON) for providing access to e-resources.

Appendix A. Supplementary material.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2022.01.010.

References:

[1] Yang X, Wan Z, Perry L, Lu H, Wang Q, Zhao C, et al. Early millet use in northern China. PNAS 2012;109(10):3726–30. doi: https://doi.org/10.1073/pnas.1115430109.
[2] Nadeem F, Ahmad Z, Wang R, Han J, Shen Qi, Chang F, et al. Foxtail millet [setaria italica (L.) beauv.] grown under low nitrogen shows a smaller root system, enhanced biomass accumulation, and nitrate transporter expression, Frontiers. Plant Sci 2018;9. doi: https://doi.org/10.3389/fpls.2018.00205.
[3] Lata C, Gupta S, Prasad M. Foxtail millet: A model crop for genetic and genomic studies in bioenergy grasses. Crit Rev Biotechnol 2013;33(3):328–43. doi: https://doi.org/10.3109/07388551.2012.716809.
[4] Muthamilarasan M, Prasad M. Advances in Setaria genomics for genetic improvement of cereals and bioenergy grasses, TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik. 2015;128:1–14. doi: https://doi.org/10.1007/s00122-014-2399-3.
[5] Bandyopadhyay T, Muthamilarasan M, Prasad M. Millets for next generation climate-smart agriculture, Frontiers in Plant. Science 2017;8:1–6. doi: https://doi.org/10.3389/fpls.2017.01266.
[6] Lawlor DW. Carbon and nitrogen assimilation in relation to yield: mechanisms are the key to understanding production systems. J Exp Bot 2002;53:773–87. doi: https://doi.org/10.1093/jexbot/53.370.773.
[7] Stewart WM, Dibb DW, Johnston AE, Smyth TJ. The contribution of commercial fertilizer nutrients to food production. Agron J 2005;97(1):1–6. doi: https://doi.org/10.2134/agronj2005.0001.
[8] Smith S, Bandyopadhyay T. Opening the dialogue: Research networks between high- and low-income countries further understanding of global agro-climatic challenges. Plants, People, Planet. 2019;1(2):98–101. doi: https://doi.org/10.1002/ppp3.17.
[9] Swarbreck SM, Wang M, Wang Y, Kindred D, Sylvester-Bradley R, Shi W, et al. A Roadmap for Lowering Crop Nitrogen Requirement. Trends Plant Sci 2019;24(10):892–904. doi: https://doi.org/10.1016/j.tplants.2019.06.006.
[10] Herbert RA. Nitrogen cycling in coastal marine ecosystems. FEMS Microbiol Rev 1999;23:563–90. doi: https://doi.org/10.1016/S0168-6445(99)00022-4.
[11] Ascott MJ, Gooddy DC, Wang L, Stuart ME, Lewis MA, Ward RS, et al. Global patterns of nitrate storage in the vadose zone, Nature. Communications. 2017;8:1–6. doi: https://doi.org/10.1038/s41467-017-01321-w.
[12] Bouwman AF, Boumans LJM, Batjes NH. Emissions of N2O and NO from fertilized fields: Summary of available measurement data. Global Biogeochemical Cycles. 2002;16:6–13. doi: https://doi.org/10.1029/2001GB001811.
[13] Saleem Kubar M, Feng M, Sayed S, Hussain Shar A, Ali Rind N, Ullah H, et al. Agronomical traits associated with yield and yield components of winter wheat as affected by nitrogen managements. Saudi Journal of Biological Sciences. 2021;28(9):4852–8. doi: https://doi.org/10.1016/j.sjbs.2021.07.027.
[14] Kocheva K, Kartseva T, Nenova V, Georgiev G, Brestič M, Misheva S. Nitrogen assimilation and photosynthetic capacity of wheat genotypes under optimal and deficient nitrogen supply. Physiol Mol Biol Plants 2020;26(11):2139–49. doi: https://doi.org/10.1007/s12298-020-00901-3.
[15] Sarwar N, Atique-ur-Rehman, Farooq O, Wasaya A, Hussain M, El-Shehawi AM, et al. Integrated nitrogen management improves productivity and economic returns of wheat-maize cropping system. Journal of King Saud University - Science. 2021;33(5):101475. doi: https://doi.org/10.1016/j.jksus.2021.101475.
[16] Melino VJ, Fiene G, Enju A, Cai J, Buchner P, Heuer S. Genetic diversity for root plasticity and nitrogen uptake in wheat seedlings. Funct Plant Biol 2015;42:942–56. doi: https://doi.org/10.1071/FP15041.
[17] Laperche A, Devienne-Barret F, Maury O, Le Gouis J, Ney B. A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of winter wheat adaptation to nitrogen deficiency. Theor Appl Genet 2006;113(6):1131–46. doi: https://doi.org/10.1007/s00122-006-0373-4.
[18] Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, et al. Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 2007;115(3):399–415. doi: https://doi.org/10.1007/s00122-007-0575-4.
[19] Laperche A, Le Gouis J, Hanocq E, Brancourt-Hulmel M. Modelling nitrogen stress with probe genotypes to assess genetic parameters and genetic determinism of winter wheat tolerance to nitrogen constraint. Euphytica 2008;161(1-2):259–71.
[20] Gupta S, Kumari K, Muthamilarasan M, Parida SK, Prasad M. Population structure and association mapping of yield contributing agronomic traits in foxtail millet. Plant Cell Reports. 2014;33:881–93. doi: https://doi.org/10.1007/s00299-014-1564-0.
[21] Shea F, Watts CE. Dumas method for organic nitrogen. Ind Eng Chem Anal Ed 1939;11(6):333–4. doi: https://doi.org/10.1021/ac50134a013.
[22] Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000;155. 945 LP – 959 http://www.genetics.org/content/155/2/945.abstract.
[23] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 2005;14(8):2611–20. doi: https://doi.org/10.1111/j.1365-294X.2005.02553.x.
[24] Earl DA, VonHoldt BM. Structure Harvester: A website and program for visualizing structure output and implementing the Evanno method, Conservation Genetics. Resources. 2012;4:359–61. doi: https://doi.org/10.1007/s12686-011-9548-7.
[25] Jaiswal V, Gupta S, Gahlaut V, Muthamilarasan M, Bandyopadhyay T, Ramchiary N, et al. Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (Setaria italica L.) Using ddRAD Sequencing. Sci Rep 2019;9(1). doi: https://doi.org/10.1038/s41598-019-41602-6.
[26] Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinf 2009;10:356. doi: https://doi.org/10.1186/1471-2105-10-356.
[27] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 2007;23(19):2633–5. doi: https://doi.org/10.1093/bioinformatics/btm308.
[28] Liu X, Huang M, Fan B, Buckler ES, Zhang Z, Listgarten J. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 2016;12(2):e1005767. doi: https://doi.org/10.1371/journal.pgen.1005767.
[29] Kaler AS, Gillman JD, Beissinger T, Purcell LC. Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Front Plant Sci 2020;10:1794. , https://www.frontiersin.org/article/10.3389/fpls.2019.01794.
[30] Malik PL, Janss L, Nielsen LK, Borum F, Jørgensen H, Eriksen B, et al. Breeding for dual-purpose wheat varieties using marker–trait associations for biomass yield and quality traits. Theor Appl Genet 2019;132(12):3375–98. doi: https://doi.org/10.1007/s00122-019-03431-z.
[31] Liang Z, Qiu Y, Schnable JC. Genome–Phenome Wide Association in Maize and Arabidopsis Identifies a Common Molecular and Evolutionary Signature. Molecular Plant. 2020;13:907–22. doi: https://doi.org/10.1016/j.molp.2020.03.003.
[32] Muhammad A, Li J, Hu W, Yu J, Khan SU, Khan MHU, et al. Uncovering genomic regions controlling plant architectural traits in hexaploid wheat using

different GWAS models. Sci Rep 2021;11(1). doi: https://doi.org/10.1038/s41598-021-86127-z.
[33] Kumar D, Sharma S, Sharma R, Pundir S, Singh VK, Chaturvedi D, et al. Genome-wide association study in hexaploid wheat identifies novel genomic regions associated with resistance to root lesion nematode (Pratylenchus thornei). Sci Rep 2021;11(1). doi: https://doi.org/10.1038/s41598-021-80996-0.
[34] Gyawali A, Shrestha V, Guill KE, Flint-Garcia S, Beissinger TM. Single-plant GWAS coupled with bulk segregant analysis allows rapid identification and corroboration of plant-height candidate SNPs. BMC Plant Biol 2019;19:412. doi: https://doi.org/10.1186/s12870-019-2000-y.
[35] Jaiswal V, Bandyopadhyay T, Gahlaut V, Gupta S, Dhaka A, Ramchiary N, et al. Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (Setaria italica L.). J Cereal Sci 2019;85:48–55. doi: https://doi.org/10.1016/j.jcs.2018.11.006.
[36] Kumar K, Muthamilarasan M, Prasad M. Reference genes for quantitative real-time PCR analysis in the model plant foxtail millet (Setaria italica L.) subjected to abiotic stress conditions. Plant Cell, Tissue Organ Cult 2013;115(1):13–22. doi: https://doi.org/10.1007/s11240-013-0335-x.
[37] Wickham H, Francois R. The dplyr package. R Core Team. 2016.
[38] Rstudio T. RStudio: Integrated Development for R., Rstudio Team, PBC, Boston, MA URL Http://Www.Rstudio.Com/. 2020. doi: 10.1145/3132847.3132886..
[39] Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat 2011;3(2):180–5. doi: https://doi.org/10.1002/wics.147.
[40] Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots, R Package Version 0.4.0. (2020)..
[41] Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E. et al., Ggally: Extension to ggplot2, R Package Version 0.5.0. (2020)..
[42] Dunn OJ. Multiple Comparisons among Means. J Am Stat Assoc 1961;56 (293):52–64. doi: https://doi.org/10.1080/01621459.1961.10482090.
[43] Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: The protein families database in 2021. Nucleic Acids Research. 2021;49:D412–9. doi: https://doi.org/10.1093/nar/gkaa913.
[44] Lynch JP, Doyle D, McAuley S, McHardy F, Danneels Q, Black LC, et al. The impact of variation in grain number and individual grain weight on winter wheat yield in the high yield potential environment of Ireland. Eur J Agron 2017;87:40–9. doi: https://doi.org/10.1016/j.eja.2017.05.001.
[45] Würschum T, Leiser WL, Langer SM, Tucker MR, Longin CFH. Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. Theor Appl Genet 2018;131 (10):2071–84. doi: https://doi.org/10.1007/s00122-018-3133-3.
[46] Voss-Fels KP, Stahl A, Wittkop B, Lichthardt C, Nagler S, Rose T, et al. Breeding improves wheat productivity under contrasting agrochemical input levels. Nat Plants 2019;5(7):706–14. doi: https://doi.org/10.1038/s41477-019-0445-5.
[47] Vazin F, Hassanzadeh M, Madani A, Nassiri-Mahallati M, Nasri M. Modeling light interception and distribution in mixed canopy of common cocklebur (Xanthium stramarium) in competition with corn. Planta Daninha. 2010;28 (3):455–62.
[48] Ngoune Tandzi L, Mutengwa CS. Estimation of Maize (Zea mays L.) Yield Per Harvest Area. Appropriate Methods, Agronomy. 2020;10(1):29. doi: https://doi.org/10.3390/agronomy10010029.
[49] Chen K, Camberato JJ, Vyn TJ. Maize Grain Yield and Kernel Component Relationships to Morphophysiological Traits in Commercial Hybrids Separated by Four Decades. Crop Sci 2017;57(3):1641–57. doi: https://doi.org/10.2135/cropsci2016.06.0540.
[50] Buah SSJ, Kombiok JM, Abatania LN. Grain sorghum response to NPK fertilizer in the Guinea Savanna of Ghana. Journal of Crop Improvement. 2012;26 (1):101–15.
[51] Mousavi SGR, Seghatoleslami MJ, Arefi R, Buah SSJ, Mwinkaara S. Response of sorghum to nitrogen fertilizer and plant density in the Guinea savanna zone. Journal of Agronomy. 2012;8:124–30.
[52] Buah SSJ, Mwinkaara S. Response of sorghum to nitrogen fertilizer and plant density in the Guinea savanna zone. Journal of Agronomy. 2009;8(4):124–30.
[53] Ausiku AP, Annandale JG, Steyn JM, Sanewe AJ. Improving Pearl Millet (Pennisetum glaucum) Productivity through Adaptive Management of Water and Nitrogen. Water. 2020;12(2):422. doi: https://doi.org/10.3390/w12020422.
[54] Nagaz K, Masmoudi M, Mechila N. Yield and water use-efficiency of pearl millet (Pennisetum glaucum (L.) R. Br.) under deficit irrigation with saline water in arid conditions of Southern Tunisia, Research. Journal of Agronomy. 2009;3:9–17.
[55] Goron TL, Bhosekar VK, Shearer CR, Watts S, Raizada MN. Whole plant acclimation responses by finger millet to low nitrogen stress. Front Plant Sci 2015;6:652. , https://www.frontiersin.org/article/10.3389/fpls.2015.00652.
[56] de Jong M, Tavares H, Pasam RK, Butler R, Ward S, George G, et al. Natural variation in Arabidopsis shoot branching plasticity in response to nitrate supply affects fitness. Natural variation in Arabidopsis shoot branching plasticity in response to nitrate supply affects fitness 2019;15(9):e1008366. doi: https://doi.org/10.1371/journal.pgen.1008366.
[57] Wu K, Wang S, Song W, Zhang J, Wang Y, Liu Q, et al. Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. Science 2020;367(6478).
[58] Yoshida H, Horie T, Shiraiwa T. A model explaining genotypic and environmental variation of rice spikelet number per unit area measured by cross-locational experiments in Asia. Field Crops Research. 2006;97(2-3):337–43. doi: https://doi.org/10.1016/j.fcr.2005.11.004.
[59] Ding C, You J, Chen L, Wang S, Ding Y. Nitrogen fertilizer increases spikelet number per panicle by enhancing cytokinin synthesis in rice. Plant Cell Rep 2014;33(2):363–71. doi: https://doi.org/10.1007/s00299-013-1536-9.
[60] Gent L, Forde BG. How do plants sense their nitrogen status?. J Exp Bot 2017;68:2531–9. doi: https://doi.org/10.1093/jxb/erx013.
[61] Kang J, Turano FJ. The putative glutamate receptor 1.1 (AtGLR1.1) functions as a regulator of carbon and nitrogen metabolism in Arabidopsis thaliana. Proceedings of the National Academy of Sciences. 2003;100:5. doi: https://doi.org/10.1073/pnas.1030961100. 6872 LP – 6877.
[62] Kang J, Mehta S, Turano FJ. The putative glutamate receptor 1.1 (AtGLR1.1) in Arabidopsis thaliana regulates abscisic acid biosynthesis and signaling to control development and water loss. Plant Cell Physiol 2004;45:1380–9. doi: https://doi.org/10.1093/pcp/pch159.
[63] Huo X, Wu S, Zhu Z, Liu F, Fu Y, Cai H, et al. NOG1 increases grain production in rice. Nat Commun 2017;8(1). doi: https://doi.org/10.1038/s41467-017-01501-8.
[64] Martin T, Oswald O, Graham IA. Arabidopsis seedling growth, storage lipid mobilization, and photosynthetic gene expression are regulated by carbon: nitrogen availability. Plant Physiol 2002;128:472–81. doi: https://doi.org/10.1104/pp.010475.
[65] Luginbuehl LH, Menard GN, Kurup S, Van Erp H, Radhakrishnan GV, Breakspear A, et al. Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. Science 2017;356(6343):1175–8.
[66] Eichmann TO, Lass A. DAG tales: the multiple faces of diacylglycerol—stereochemistry, metabolism, and signaling, Cellular and Molecular. Life Sci 2015;72(20):3931–52. doi: https://doi.org/10.1007/s00018-015-1982-3.
[67] Hong Y, Devaiah SP, Bahn SC, Thamasandra BN, Li M, Welti R, et al. Phospholipase Dε and phosphatidic acid enhance Arabidopsis nitrogen signaling and growth. Plant J 2009;58(3):376–87. doi: https://doi.org/10.1111/j.1365-313X.2009.03788.x.
[68] Sakakibara H, Suzuki M, Takei K, Deji A, Taniguchi M, Sugiyama T. A response-regulator homologue possibly involved in nitrogen signal transduction mediated by cytokinin in maize. The Plant Journal : For Cell and Molecular Biology. 1998;14(3):337–44.
[69] Sugiyama T, Sakakibara H. Regulation of carbon and nitrogen assimilation through gene expression. In: Photosynthetic Nitrogen Assimilation and Associated Carbon and Respiratory Metabolism. Springer; 2002. p. 227–38.
[70] Asins MJ, Albacete A, Martinez-Andujar C, Pérez-Alfocea F, Dodd IC, Carbonell EA, et al. Genetic analysis of rootstock-mediated nitrogen (N) uptake and root-to-shoot signalling at contrasting N availabilities in tomato. Plant Sci 2017;263:94–106. doi: https://doi.org/10.1016/j.plantsci.2017.06.012.
[71] Kong F, Deng Y, Zhou B, Wang G, Wang Y, Meng Q. A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. J Exp Bot 2014;65:143–58. doi: https://doi.org/10.1093/jxb/ert357.
[72] Ham B-K, Park JM, Lee S-B, Kim MJ, Lee I-J, Kim K-J, Kwon CS, Paek K-H. Tobacco Tsip1, a DnaJ-type Zn finger protein, is recruited to and potentiates Tsi1-mediated transcriptional activation. The Plant Cell. 2006;18:2005–20. doi: https://doi.org/10.1105/tpc.106.043158.
[73] Xu XM, Wang J, Xuan Z, Goldshmidt A, Borrill PGM, Hariharan N, et al. Chaperonins Facilitate KNOTTED1 Cell-to-Cell Trafficking and Stem Cell Function. Science 2011;333(6046):1141–4.
[74] Fichtenbauer D, Xu XM, Jackson D, Kragler F. The chaperonin CCT8 facilitates spread of tobamovirus infection. Plant Signaling Behav 2012;7(3):318–21.