OXFORD GENETICS

Estimating the genetic parameters of yield-related traits under different nitrogen conditions in maize

Semra Palali Delen[1,2,†], Gen Xu[1,2,†,*], Jenifer Velazquez-Perfecto[3] and Jinliang Yang[1,2,*]

[1]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
[2]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
[3]Department of Agricultural Leadership, Education and Communication, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
[†]These authors contributed equally to this work.

[*]Corresponding author: Email: Gen Xu (gxu6@unl.edu) and Jinliang Yang (jinliang.yang@unl.edu)

Abstract

Understanding the genetic basis responding to nitrogen (N) fertilization in crop production is a long-standing research topic in plant breeding and genetics. Albeit years of continuous efforts, the genetic architecture parameters, such as heritability, polygenicity, and mode of selection, underlying the N responses in maize remain largely unclear. In this study, about $n = 230$ maize inbred lines were phenotyped under high N (HN) and low N (LN) conditions for two consecutive years to obtain six yield-related traits. Heritability analyses suggested that traits highly responsive to N treatments were less heritable. Using publicly available SNP genotypes, the genome-wide association study (GWAS) was conducted to identify $n = 237$ and $n = 130$ trait-associated loci (TALs) under HN and LN conditions, $n = 164$ for N-responsive (NR) traits, and $n = 31$ for genotype by N interaction (G $\times$ N). Furthermore, genome-wide complex trait Bayesian (GCTB) analysis, a method complementary to GWAS, was performed to estimate genetic parameters, including genetic polygenicity and the mode of selection (S). GCTB results suggested that the NR value of a yield component trait was highly polygenic and that four NR traits exhibited negative correlations between SNP effects and their minor allele frequencies (or the S value < 0) — a pattern consistent with negative selection to purge deleterious alleles. This study reveals the complex genetic architecture underlying N responses for yield-related traits and provides candidate genetic loci for N resilient maize improvement.

Keywords: nitrogen; GWAS; TAL;GCTB;polygenicity; heritability;selection; maize

Introduction

Nitrogen (N), as a fundamental macronutrient, is a major constituent of proteins, nucleic acid, and metabolites and is critical for the high yielding of crops (Pilbeam 2018). Since the 1960s, subsequent to the Green Revolution, due to the Haber-Bosch process, inorganic N fertilizers became increasingly available for crop production, especially in maize, where about 20% of the N fertilizers was applied for maize production (Heffer 2013; Ludemann *et al.* 2022). However, inefficient N usage causes ammonia emission to the environment, accounting for a considerable proportion of fine particulate matter pollution (i.e., PM2.5) and reducing human population life span (Gu *et al.* 2021). Meanwhile, N runoff imposes substantial adverse effects on natural ecosystems, such as reduced water quality and impaired soil health. Therefore, understanding the plant response to N in crop production is crucial for human health, food security, and environmental sustainability and is a long-standing research topic in plant breeding and genetics.

To identify N-responsive genetic loci, many QTL studies were performed, resulting in a number of trait-associated QTLs under different N conditions (Agrama *et al.* 1999; Liu *et al.* 2012) or QTLs for different N-related traits, i.e., grain N yield, N remobilization, and post-silking N uptake (Coque *et al.* 2008; Li *et al.* 2015). Recently, as the technical advances, genetic studies for N-related traits shifted from QTL mapping to GWAS (Morosini *et al.* 2017; He *et al.* 2020), leading to high-resolution mapping results. For example, a recent GWAS using 411 maize inbred lines under optimum and low N conditions detected about 80 significant SNPs and 136 putative candidate genes (Ertiro *et al.* 2020). These N-related QTLs and trait-associated SNPs provide opportunities to investigate the fate of the deleterious alleles — the alleles can potentially affect fitness under different N conditions. During the recent maize improvement process, an excess of the mutational load was enriched in even elite maize inbred lines (Yang *et al.* 2017). However, it is largely unclear how many alleles contribute to NR traits and what is the mode of selection on these alleles, including potentially deleterious alleles, in affecting N responses.

In the current study, by employing two complementary approaches — GWAS and GCTB (Genome-wide Complex Trait Bayesian analysis), we analyzed yield-related traits collected under low N (LN) and high N (HN) conditions (i.e., trait *per se*) as well as the transformed N-responsive (NR) traits. We found higher heritability for most traits *per se* under HN than LN and identified 1,198 trait-associated SNPs in total that locate in 545 genomic regions. Inferring from genome-wide non-zero effects SNPs, including not only significant GWAS SNPs but also SNPs with minor effects, GCTB results suggested the yield-related NR traits were highly polygenic and that NR traits were more likely

under negative selection (Charlesworth *et al.* 1995). The complex genetic architecture revealed from this study, especially for the NR traits, provides guidelines for further genome-enabled selection modeling and N resilient maize development.

Materials and methods

Plant materials and field experimental design

In this experiment, a subset ($n = 226$ genotypes) of the maize diversity panel (Flint-Garcia *et al.* 2005) was planted in a rain-fed experimental field followed commercial maize. For the N treated plots, urea (dry fertilizer) as a source of N was applied at the rate of 120 lbs/acre before planting. The field experiment was conducted using an incompletely randomized block design in two consecutive field seasons (2018-2019). For each replication of a treatment, the field was split into four blocks by plant height and maturity (i.e., tall/early, tall/late, short/early, short/late). Each block was further subdivided into three sub-blocks. Within each sub-blocks, two hybrid varieties B73×Mo17 and B37×Mo17 were planted randomly as check plants (see also (Rodene *et al.* 2022; Meier *et al.* 2022)). Each genotype was planted in a two-row, 20 ft × 5 ft sub-plot (6.096 m × 1.524 m). The between-row spacing was 30 inches (76.2 cm) and the within-row plant spacing was 6 inches (15.24 cm). In each row, 38 seeds were planted, resulting in a planting density of around 30,000 seeds per acre.

Phenotypic data collection

From each two-row plot, three mature ears were harvested from the representative plants. These harvested ears were dried in the oven at 37C for three days to decrease the moisture content. Harvested ears were hand-shelled to prevent kernel loss. After shelling, the kernels and cobs were kept separately with proper barcoded labels. From the cobs, cob diameter (CD), cob length (CL), and cob weight (CW) were manually measured. The total kernel weight (TKW) of each ear was measured from the collected kernels. And then, 20 representative kernels were selected to measure 20 kernel weight (20KW). Finally, the kernel count (KC) was computed using TKW divided by average kernel weight.

Best linear unbiased prediction (BLUP) and N-responsive trait calculation

To obtain the best linear unbiased prediction (BLUP) values of each genotype, we fitted a linear mixed model by treating the year as the fixed effect, and genotype, replication, block, sub-block, and genotype by year interaction as random effects. For each N treatment, the BLUP values were calculated separately using an R package "lme4" (Bates *et al.* 2015).

In the model,

$$y_{ijkrl} = \mu + t_l + g_i + g_i * t_l + b_{jrl} + s_{jkrl} + q_{rl} + \varepsilon$$

where y_{ijkrl} is the phenotypic value of the i^{th} genotype evaluated in the k^{th} sub-block of the j^{th} block of r^{th} replicate nested within the l^{th} year; μ is the overall mean; t_l is the fixed effect of the l^{th} year; g_i is the random effect of the i^{th} genotype; $g_i * t_l$ is the random effect of the i^{th} genotype with the l^{th} year interaction; b_{jrl} is the random effect of the j^{th} block of the r^{th} replicate within the l^{th} year; s_{jkrl} is the random effect of the k^{th} sub-block of the j^{th} block of the r^{th} replicate within the l^{th} year; q_{rl} is the random effect of the r^{th} replicate nested within the l^{th} year; ε is the random residual error.

The N-responsive (NR) traits were calculated from the BLUP values using the equation (Xu *et al.* 2012),

$$NR = \frac{T_{HN} - T_{LN}}{T_{LN}}$$

where T_{HN} and T_{LN} are the BLUP values for a given trait measured from HN and LN field conditions.

Broad sense heritability calculation

The broad-sense heritability (H^2) of yield-related traits was calculated using the equation as:

$$H^2 = \frac{V_G}{V_G + \frac{V_{G\times E}}{i} + \frac{V_E}{i\times j}}$$

where V_G is the genotypic variance; V_E is the environmental variance of different years; and $V_{G\times E}$ is the variance of genotype by year interaction; $i = 2$ is the number of years and $j = 2$ is the number of replications per year.

Genome-wide Association Study (GWAS) and genotype by N interaction (G × N) analysis

The SNP genotype of the maize diversity panel was downloaded from maize HapMap3 (Bukowski *et al.* 2017) with AGPv4 coordinates. After filtering out SNPs with minor allele frequency (MAF) < 0.05 and missing rate < 0.3 among the 226 lines phenotyped in this study, approximately 21 million SNPs were retained.

In GWAS, we employed the QK model that considers both population structure (Q) and kinship relatedness (K) to control for multiple levels of confounding effects (Yu *et al.* 2006; Parisseaux and Bernardo 2004). In the model,

$$\mathbf{y} = \mathbf{Qv} + \mathbf{w_i}m_i + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of BLUP value for a given trait (or the NR trait); $\mathbf{Q}$ is the design matrix of the population structure (i.e., the principle components); $\mathbf{v}$ is the vector of the fixed subpopulation effect; $\mathbf{w_i}$ is a vector of the i^{th} SNP genotype; m_i is the fixed SNP effect to be estimated by an iterative procedure; $\mathbf{Z}$ is the incidence matrix for random effects; $\mathbf{u}$ is the vector of breeding values to be predicted (random effect); $\mathbf{e}$ is the vector of the random residual error.

In the analysis, the $\mathbf{Q}$ matrix was the first three principal components calculated from genome-wide SNPs using PLINK 1.9 software (Chang *et al.* 2015). And the $\mathbf{Z}$ matrix was computed using GEMMA (v 0.98.3) software with option "4" (Zhou and Stephens 2012). The above model was then implemented to estimate significant SNP effects for each trait using GEMMA (Zhou and Stephens 2012). The threshold for the significant association SNPs was set to 1.2×10^{-6} ($1/n$, n = 769,690 is the number of independent SNPs with MAF $\geq$ 5%) according to the method developed previously (Rodene *et al.* 2022).

For the G × N analysis, in addition to considering the population structure and kinship, a genotype by N interaction term was fitted into an unified linear mixed model as the random effect (Yoshida *et al.* 2022). In practice, the first three principal components were used for controlling the population structure, and n = 769,690 independent SNPs were tested for G × N using an R package gwasQxE (Yamamoto and Matsunaga 2021).

From the association analysis results, significant genomic loci were determined by considering a 100 kb window upstream and downstream of the significant SNPs. Overlapping regions were merged, and these regions were defined as trait-associated loci (TAL). We conducted GO term analysis on genes under GWAS peaks using AgriGO2.0 with default parameters (Tian *et al.* 2017). We used the significance cutoff at *P*-value < 0.01.

Genome-wide Complex Trait Bayesian (GCTB) analysis

Genome-wide Complex Trait Bayesian (GCTB-BayesS) approach, which is based on Bayesian multiple regression mixed linear models (Zeng *et al.* 2018), was performed to estimate genetic architecture parameters of yield-related traits. In the BayesS model,

$$\mathbf{y} = \mathbf{1u} + \mathbf{Xb} + \mathbf{e}$$

where **y** is a vector of BLUP value for a given trait (or the NR trait); **u** is the fixed effect; **X** is the SNP genotype matrix; **b** is the vector of random SNP effects; and **e** is the residual error.

The SNP effect b_j for the jth locus was assumed to have a hierarchical mixture prior with a point mass at zero and a univariate normal distribution conditional on σ^2:

$$b_j \mid \pi, S, \sigma^2 \begin{cases} \sim N\left(0, [(2p_j(1-p_j)]^S\sigma^2\right) & probability\ (1-\pi) \\ 0 & probability\ \pi \end{cases}$$

where p_j is the allele frequency at SNP locus j and priors for π, S, and σ^2 are the same as explained in (Zeng *et al.* 2018). In the BayesS model, the S parameter, as an estimation of the selection coefficient, computes the joint distribution between the variance of SNP effects and MAFs; and the $1-\pi$ parameter estimates the number of non-zero SNPs as a proxy of polygenicity.

In practice, the following options were selected for Markov Chain Monte Carlo (MCMC): "–chain-length = 1,010,000, –burn-in = 10,000". In the analysis, n = 834,975 independent SNPs (MAF $\geq$ 1%) were used (Yang *et al.* 2022), which was determined by using PLINK 1.9 (Chang *et al.* 2015) with the "indep-pairwise" option (window size 100 kb, step size 100, $r^2 \geq 0.1$).

The SNP-based heritability (h^2_{SNP}) was estimated using the posterior sampled values of SNP effects after burn-in of MCMC. Briefly, the estimated genetic value for individual i was estimated with sampled non-zero effect SNPs (b_j):

$$g_i = \sum_j X_{ij} b_j$$

where j is the non-zero effect SNPs in the current MCMC cycle and X_{ij} is the SNP genotype of individual i at this SNP j. Then, the estimated genetic variance is

$$\sigma^2_g = \frac{\sum_i g_i^2}{n} - (\frac{\sum_i g_i}{n})^2$$

where n is the number of individuals. Conditional on the sampled value of residual variance (σ^2_e) in the current cycle of MCMC, the SNP-based heritability is estimated as

$$h^2_{SNP} = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}$$

Finally, the mean values of posterior h^2_{SNP} values after burn-in was calculated as the SNP-based heritability estimation.

Results

Phenotypic evaluation of diverse maize lines under different N conditions

A subset of the maize diversity panel (n = 226 lines) was planted in a replicated field trial under high N (HN) and low N (LN) conditions according to an incomplete block design in 2018 and 2019 (see **Materials and Methods**). Using leaf N content and chlorophyll content collected from adult plants in the field (Ge *et al.* 2019), results showed the significantly high N content and chlorophyll content for plants collected in the HN field as compared to LN field (**Figure S1**). From harvested mature ears, six yield-related traits were manually measured, including three cob-related traits (cob diameter, CD; cob length, CL; and cob weight, CW) and three kernel-related traits (20 kernel weight, 20KW; kernel count, KC; and total kernel weight, TKW). For these traits, the best linear unbiased prediction (BLUP) values were calculated separately for each N condition. Besides traits *per se*, we also calculated N-responsive (NR) traits from the BLUP values following a previous method (Liu *et al.* 2021) (see **Materials and Methods** and **Table S1**). Most of these yield-related traits are highly correlated traits in both HN (**Figure S2**) and LN (**Figure S3**) conditions. As expected, these traits exhibited significantly larger BLUP values in HN than LN conditions, except for CD and 20KW (**Figure 1A**). TKW, a trait most closely related to yield (Yang *et al.* 2017), showed the most striking differences from 39.9 g in LN to 46.9 in HN (Paired t-test, *P*-value = 2.2×10^{-74}) and 97.8% of the lines exhibited positive N responses (**Figure 1B**). Similarly, the BLUP value for KC was also significantly improved from 174 in LN to 196 in HN (Paired t-test, *P*-value = 1.8×10^{-30}). And 55.7% and 88.7% of the inbred lines positively responded to the elevated N levels for 20KW and KC, respectively (**Figure 1B**).

Strong N-responsive traits are less heritable

The broad-sense heritability (H^2) of these yield-related traits was estimated separately for each N condition (see **Materials and Methods**). Generally, we found these traits showed higher levels of heritability in HN than LN fields (**Figure 2A**), except for 20KW (H^2 = 0.73 in HN and H^2 = 0.8 in LN) and TKW (H^2 = 0.21 in HN and H^2 = 0.21 in LN), suggesting the environmental effects were less dominant or the data were more repeatable under HN conditions. For cob-related traits, CD (H^2 = 0.81 in HN and H^2 = 0.73 in LN) was more heritable than CL (H^2 = 0.68 in HN and H^2 = 0.51 in LN) and CW (H^2 = 0.53 in HN and H^2 = 0.37 in LN); while for kernel-related traits, 20KW exhibited the highest heritability compared to KC (H^2 = 0.35 in HN and H^2 = 0.16 in LN) and TKW. Additionally, we found the cob-related traits, on average, were more heritable than kernel-related traits, regardless of the N conditions (**Figure 2A**), suggesting kernel-related traits are more sensitive to local environments under N-limiting conditions. Interestingly, the levels of heritability negatively correlated with proportions of inbreds with NR values > 0 (**Figure 1B**), or ratios of inbreds positively responding to N treatments, under both N conditions (**Figure 2B**). These results are consistent with the view that more fitness-related traits, i.e., traits strongly responsive to N treatments, are less heritable (Falconer and Mackay 1996).

Comparing GWAS signals under different N conditions

We then conducted GWAS for the six yield-related traits *per se* as well as the transformed NR traits by fitting a linear mixed

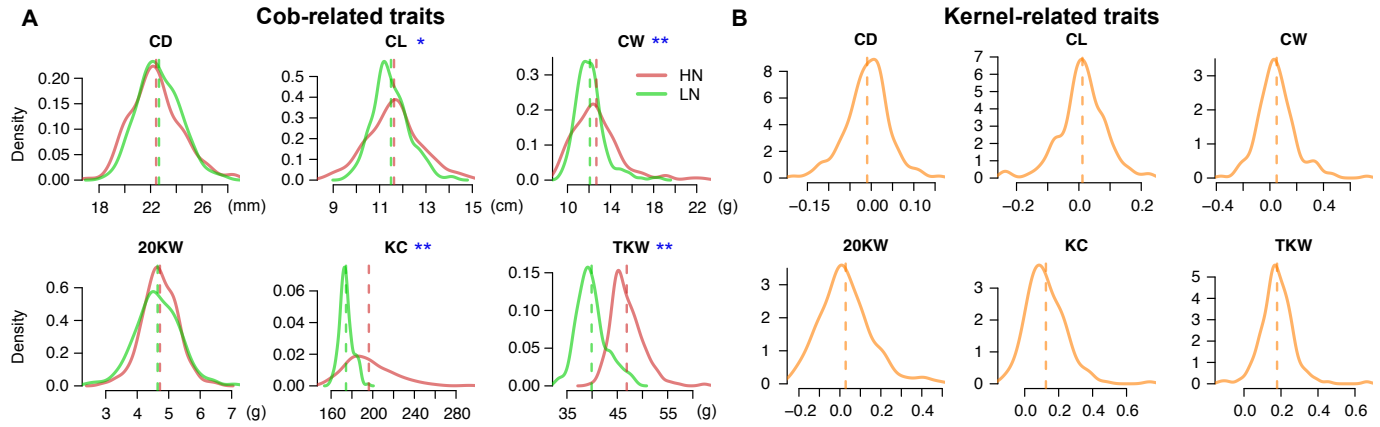

Figure 1 Cob- and kernel-related traits under different N conditions. (**A**) Density plots of the phenotypic traits in low N (LN) and high N (HN) fields. The red and green vertical dashed lines indicate the mean values of each trait. The blue asterisks indicate the traits show significant differences between HN and LN conditions (Paired t-test, P-value < 0.01). (**B**) The distributions of N-responsive (NR) traits. Orange dashed lines denote the mean values.

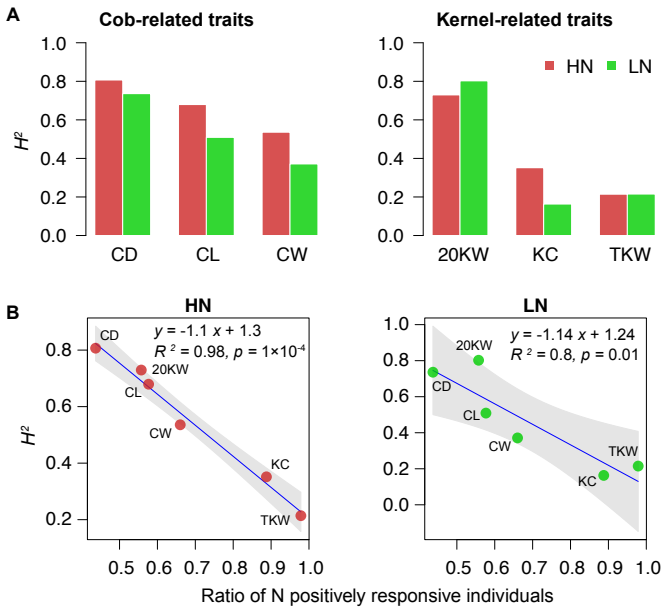

Figure 2 Heritability estimation and correlation analysis with N-response. (**A**) Heritabilities of yield-related traits under different N conditions. (**B**) Correlation analysis between heritability and N-responsive value. Solid blue line indicates the linear regression and the grey shaded area denotes the 95% confidence interval.

model using 21 million SNPs (see **Materials and Methods**). In the GWAS model, the first three principal components were fitted as the fixed effects and the genetic relatedness computed from genome-wide SNPs as the random effects. To control for the false discovery rate (FDR), the modified Bonferroni-adjusted threshold was determined as 1.2×10^{-6} based on $n = 769,690$ independent SNPs (Yang *et al.* 2021b; Li *et al.* 2012). As a result (**Figure 3**, see **Figure S4** and **S5** for the quantile-quantile (Q-Q) plots), a total of 515 unique genomic regions were identified as the trait-associated loci (TALs, see **Table S2** and **S3** for GWAS results).

For HN traits, 237 TALs were identified (**Figure S6**), $n = 25$ of which were detected for at least two yield-related traits (**Figure 3B**) — more than expected by chance (permutation test, P-value = 0.001). Such a large number of overlapping signals were not found for traits under LN conditions ($n = 3$ shared TALs, **Figure 3C** and **Figure S7**). Many shared TALs for HN traits were between highly correlated KC and TKW traits, likely because KC was not a directly measured trait but calculated from TKW and 20KW. Comparatively, very few overlapped TALs were identified for the same trait under different N conditions (**Figure S8**).

Of the 164 TALs for NR traits, $n = 6$ of them were detected for at least two yield-related traits (**Figure 3A** and **Figure 3D**), which is more than expected by chance (permutation test, P-value = 0.02). In addition, we collected 47 N pathway genes (**Table S4**) and found three of these genes (permutation test for gene set enrichment, P-value = 0.012), including *npf2* (*Zm00001d029932*) and *npf3* (*Zm00001d029933*) — two nitrate transporters in a tandem array, and *gln5* (*Zm00001d051804*) — a glutamine synthetase, were underlying the GWAS peaks. Further, the gene ontology (GO) term enrichment test suggested that several GO terms involved in "cell wall organization" were significantly enriched in genes underlying the GWAS peaks for kernel-related NR traits (**Table S5**).

To explore the genetic control for the genotype by N interaction (G × N), we performed G × N analysis using a linear mixed model (see **Materials and Methods**). As a result, 31 TALs for G × N were identified, including 15 for cob-related traits and 16 for kernel-related traits (**Figure S9** and **Table S3**). Among these TALs, two of them were shared between KC and TKW and one promising candidate gene *Zm00001d006592* was underlying this peak. This candidate gene is annotated as chlorophyll a-b binding protein, which is associated with photosystem in plants (Andersson *et al.* 2003).

Estimating genetic architecture parameters for traits *per se* and N-responsive traits

In addition to GWAS, we fitted the *per se* and NR traits to a Bayesian-based model implemented in GCTB (Zeng *et al.* 2018) (see **Materials and Methods**). This method allows the simulta-

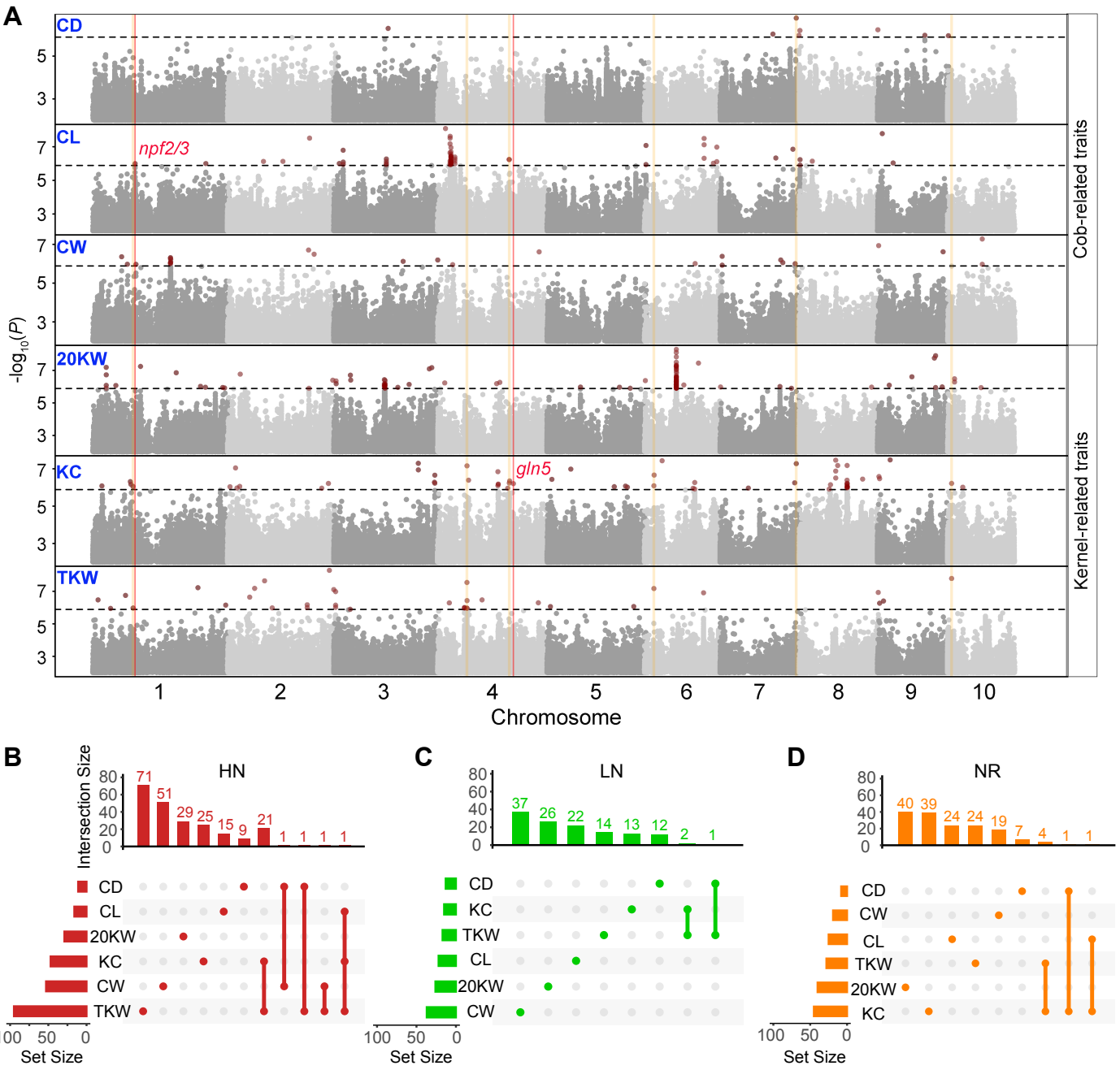

Figure 3 GWAS results for cob- and kernel-related traits under different nitrogen conditions. (**A**) Stacking Manhattan plot for N-responsive (NR) traits. The black horizontal dashed line indicates the GWAS threshold. Each red dot above the threshold represents the SNP significantly associated with a trait. The vertical orange lines indicate the overlapped trait-associated loci (TALs). The vertical red lines indicate the physical position of the N pathway genes underlying the TALs. (**B-D**) Overlapping results of TALs for HN (**B**), LN (**C**), and NR traits (**D**). Numbers on top of the barplots indicate the number of unique (only dots) and shared (dots and lines) TALs.

neous estimation of the genetic architecture parameters, such as polygenicity (π, the percentage of non-zero effect SNPs), variance of BLUP values due to SNPs (h^2_{SNP}, SNP-based heritability), and the mode of selection (S, a proxy using the relationship between variance of SNP effect and MAF). Using relatively independent SNPs with MAF > 1%, we estimated genetic parameters to compare the genetic architecture for the cob and kernel-related traits as well as for *per se* and NR traits (see **Materials and Methods**).

For trait *per se* under HN and LN conditions, we observed no significant differences in the polygenicity (**Figure 4A**), with an average of n = 964 SNPs (or $1-\pi$ =0.1%) exhibiting non-zero effects; but the polygenicity among traits showed a large variation, ranging from n = 66 SNPs for CW in HN to n = 1,987 SNPs for CD in LN. In particular, for the NR trait of TKW, the number of non-zero effect SNPs elevated to n = 52,721 (or $1-\pi$ =6.3%), suggesting TKW — a key yield component trait — was under complex genetic control in responding to changed N conditions.

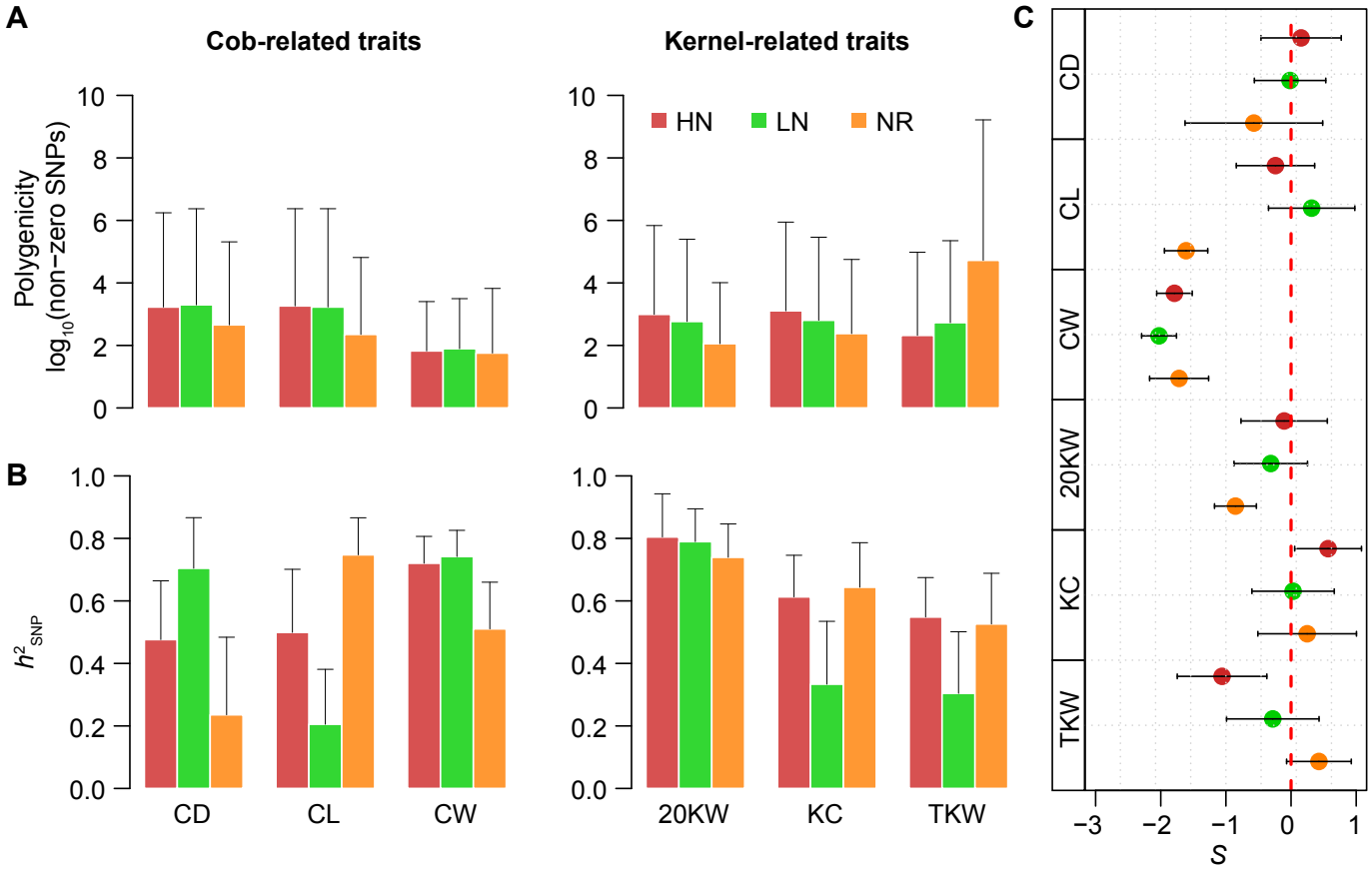

Figure 4 Genetic architecture parameters estimated for HN, LN and NR traits. (**A**) Genetic polygenicity measured by number of non-zero effect SNPs. (**B**) Phenotypic variance explained by SNPs under an additive model. (**C**) The relationships between SNP effects and MAFs of the non-zero effect SNPs.

The h^2_{SNP} values for HN and LN traits were largely in line with the broad sense heritabilities estimated previously with the field data (**Figure 4B**). However, some abnormal values were detected, such as a small h^2_{SNP} value for CL trait in LN, which was likely due to dominance or epistasis mode of inheritance playing an important role as our model considered only additive effect. Or simply because of imperfect model convergence. We also estimated the h^2_{SNP} for NR traits and found, in general, kernel-related NR traits (h^2_{SNP} ranging from 0.51 ± 0.15 to 0.64 ± 0.1) were more heritable than cob-related NR traits (h^2_{SNP} ranging from 0.23 ± 0.24 to 0.73 ± 0.1). Finally, these estimated SNP effects and their allele frequencies in the population allowed us to infer the mode of selection. As pointed out by Zeng et. al., (Zeng *et al.* 2018), $S = 0$ indicates selective neutral, while $S > 0$ and $S < 0$ suggest the positive and negative selection. Our results revealed that S values of four NR traits (CD, CL, CW, and 20KW), two HN traits (CW and TKW), and one LN trait (CW) were significantly smaller than zero (**Figure 4C**), indicating negative selection may be taken into effect to maintain the large effect deleterious SNPs in low frequencies, especially in responding to changed N conditions.

Discussion

In this study, we characterized six yield-related traits under high and low N conditions for two consecutive years and analyzed the data with two complementary approaches — GWAS (to estimate significant effect SNPs) and GCTB (to detect non-zero effect SNPs and infer other genetic parameters). We identified 1,198 GWAS signals located within $n = 545$ genomic regions or TALs. Many of these TALs were repeatedly detected for different traits under the same N conditions, but very few TALs were shared for the same trait under different N conditions, likely because G × N interaction plays an important role in controlling phenotypic variation. Indeed, a G × N analysis identified about 30 TALs. One of the candidate genes, Zm00001d006592 — a chlorophyll a-b binding protein, is located within a repeatedly detected locus for G × N, suggesting certain maize inbreds conduct photosynthesis more efficiently than others under different N conditions. Yet, the actual causal effect of this candidate gene needs to be validated. Genes underlying the GWAS peaks are enriched for cell wall organization, which is consistent with the view that N assimilation is related to cell wall synthesis and remodeling (Landi and Esposito 2017).

In addition to the GWAS SNPs, non-zero effect SNPs estimated from GCTB provided a proxy for evaluating genetic polygenicity. Results suggested that most of these yield-related traits are highly polygenic. In particular, we found the N-responsive trait of TKW was controlled by the highest number of non-zero SNPs (i.e., more than $n = 4 \times 10^4$ SNPs across the genome), consistent with the view that genetic basis for N responses in crop yield is highly complex (Ribeiro *et al.* 2018). Heritability estimation from the field data suggested that traits highly responsive to N treatment tend to be less heritable, further confirming the

genetic complexity of N responses for yield-related traits. The lower heritability in LN conditions suggested that local variation in soil N supply likely contributes greater to phenotypic variance when N is a limiting factor. The higher polygenicity (i.e., more genetic factors contributing to the phenotypic variance) in the LN condition suggested that many small effect genes might contribute to N resilience, and such effect likely be non-detectable under N luxury conditions. Because of the low selection pressure on these N resilient genes over the past several decades of maize improvement under sufficient N conditions, genetic diversity for N resilience might be decreased, and advantage alleles might have been lost. Therefore, to develop N-resilient crops in the face of climate change, the maize lines developed before the inorganic N fertilizer (i.e., before the 1960s) should be re-investigated to search for N-resilient alleles.

In the GCTB result, we detected most of the NR traits exhibiting negative S values, suggesting large effect SNPs for NR traits tend to be rare in the population. It is likely because these rare SNPs were deleterious and, therefore, were maintained in low frequencies to increase the plant fitness in responding to changed N conditions. N, as one of the significant macronutrients for crop development, its composition in the soil varies spatially and temporally (Yang *et al.* 2021a; Zhu *et al.* 2013). Therefore, it is not surprising to expect plant breeding over the past 60 years since the Green Revolution or natural selection on a longer time scale has affected the patterns of deleterious alleles in responding to the N availability. However, the limitations of the current statistical methods (i.e., high false discovery rates and imperfect model convergence) prevent us from pinning down the individual deleterious locus accurately. Note that these genetic parameters were rough estimations derived from the posterior distributions. The point estimations were associated with large standard errors. To get more accurate results, a larger population, especially using a population composed of maize hybrids, or more sophisticated statistical approaches are warranted.

Data and code availability

The raw phenotypic data, supplemental materials, and code used for the analyses can be accessed through the GitHub repository (https://github.com/jyanglab/Genetics-parameters-for-N-related-traits).

Acknowledgements

We thank two anonymous reviewers for their careful review and valuable comments. The computational work was conducted using the Holland Computing Center of the University of Nebraska-Lincoln, which receives support from the Nebraska Research Initiative.

Funding

This work is supported by the Agriculture and Food Research Initiative Grant number 2019-67013-29167 from the USDA National Institute of Food and Agriculture and the University of Nebraska-Lincoln Start-up fund.

Conflicts of interest

The authors declare no competing interests.

Literature cited

Agrama H, Zakaria A, Said F, Tuinstra M. 1999. Identification of quantitative trait loci for nitrogen use efficiency in maize. Molecular Breeding. 5:187–195.

Andersson U, Heddad M, Adamska I. 2003. Light stress-induced one-helix protein of the chlorophyll a/b-binding family associated with photosystem i. Plant physiology. 132:811–820.

Bates D, Maechler M, Bolker B, Walker S. 2015. Mixed-effects models using lme4. J Stat Softw. 67:1–48.

Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C *et al*. 2017. Construction of the third-generation Zea mays haplotype map. GigaScience. 7. gix134.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4. s13742-015-0047-8.

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

Coque M, Martin A, Veyrieras J, Hirel B, Gallais A. 2008. Genetic variation for n-remobilization and postsilking n-uptake in a set of maize recombinant inbred lines. 3. qtl detection and coincidences. Theoretical and Applied Genetics. 117:729–747.

Ertiro BT, Labuschagne M, Olsen M, Das B, Prasanna BM, Gowda M. 2020. Genetic dissection of nitrogen use efficiency in tropical maize through genome-wide association and genomic prediction. Frontiers in Plant Science. 11.

Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Harlow: Longman. fourth edition.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal. 44:1054–1064.

Ge Y, Atefi A, Zhang H, Miao C, Ramamurthy RK, Sigmon B, Yang J, Schnable JC. 2019. High-throughput analysis of leaf physiological and chemical traits with vis–nir–swir spectroscopy: a case study with a maize diversity panel. Plant methods. 15:1–12.

Gu B, Zhang L, Van Dingenen R, Vieno M, Van Grinsven HJ, Zhang X, Zhang S, Chen Y, Wang S, Ren C *et al*. 2021. Abating ammonia is more cost-effective than nitrogen oxides for mitigating pm2. 5 air pollution. Science. 374:758–762.

He K, Xu S, Zhang X, Li Y, Chang L, Wang Y, Shi Y, Cui T, Dong Y, Lan T *et al*. 2020. Mining of candidate genes for nitrogen use efficiency in maize based on genome-wide association study. Molecular Breeding. 40:1–17.

Heffer P. 2013. Assessment of fertilizer use by crop at the global level 2010-2010/11 international fertilizer industry association (ifa), paris, france.

Landi S, Esposito S. 2017. Nitrate uptake affects cell wall synthesis and modeling. Frontiers in Plant Science. 8:1376.

Li MX, Yeung JM, Cherny SS, Sham PC. 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human genetics. 131:747–756.

Li P, Chen F, Cai H, Liu J, Pan Q, Liu Z, Gu R, Mi G, Zhang F, Yuan L. 2015. A genetic relationship between nitrogen use efficiency and seedling root traits in maize as revealed by QTL analysis. Journal of Experimental Botany. 66:3175–3188.

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y. 2012. Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and qtl data. Plant Molecular

Biology Reporter. 30:297–308.
Liu Y, Wang H, Jiang Z, Wang W, Xu R, Wang Q, Zhang Z, Li A, Liang Y, Ou S *et al*. 2021. Genomic basis of geographical adaptation to soil nitrogen in rice. Nature. 590:600–605.
Ludemann C, Gruere A, Heffer P, Dobermann A. 2022. Global data on fertilizer use by crop and by country.
Meier MA, Xu G, Lopez-Guerrero MG, Li G, Smith C, Sigmon B, Herr JR, Alfano JR, Ge Y, Schnable JC *et al*. 2022. Association analyses of host genetics, root-colonizing microbes, and plant phenotypes under different nitrogen conditions in maize. eLife. 11:e75790.
Morosini JS, Mendonca LdF, Lyra DH, Galli G, Vidotti MS, Fritsche-Neto R. 2017. Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. Plant and Soil. 421:453–463.
Parisseaux B, Bernardo R. 2004. In silico mapping of quantitative trait loci in maize. Theoretical and Applied Genetics. 109:508–514.
Pilbeam DJ. 2018. The utilization of nitrogen by plants: a whole plant perspective. Ann. Plant Rev. Online. pp. 305–351.
Ribeiro P, Badu-Apraku B, Gracen V, Danquah E, Garcia-Oliveira A, Asante M, Afriyie-Debrah C, Gedil M. 2018. Identification of quantitative trait loci for grain yield and other traits in tropical maize under high and low soil-nitrogen environments. Crop Science. 58:321–331.
Rodene E, Xu G, Palali Delen S, Zhao X, Smith C, Ge Y, Schnable J, Yang J. 2022. A uav-based high-throughput phenotyping approach to assess time-series nitrogen responses and identify trait-associated genetic components in maize. The Plant Phenome Journal. 5:e20030.
Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agrigo v2. 0: a go analysis toolkit for the agricultural community, 2017 update. Nucleic acids research. 45:W122–W129.
Xu G, Fan X, Miller AJ. 2012. Plant nitrogen assimilation and use efficiency. Annual review of plant biology. 63:153–182.
Yamamoto E, Matsunaga H. 2021. Exploring efficient linear mixed models to detect quantitative trait locus-by-environment interactions. G3. 11:jkab119.
Yang H, Song X, Zhao Y, Wang W, Cheng Z, Zhang Q, Cheng D. 2021a. Temporal and spatial variations of soil c, n contents and c:n stoichiometry in the major grain-producing region of the north china plain. PLOS ONE. 16:1–14.
Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, Mumm RH, Ross-Ibarra J. 2017. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS genetics. 13:e1007019.
Yang Z, Xu G, Zhang Q, Obata T, Yang J. 2021b. Genome-wide mediation analysis: bridging the divide between genotype and phenotype via transcriptomic data in maize. bioRxiv. .
Yang Z, Xu G, Zhang Q, Obata T, Yang J. 2022. Genome-wide mediation analysis: an empirical study to connect phenotype with genotype via intermediate transcriptomic data in maize. Genetics. 221:iyac057.
Yoshida H, Hirano K, Yano K, Wang F, Mori M, Kawamura M, Koketsu E, Hattori M, Ordonio RL, Huang P *et al*. 2022. Genome-wide association study identifies a gene responsible for temperature-dependent rice germination. Nature Communications. 13:1–13.
Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics. 38:203–208.
Zeng J, De Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, Yap CX, Xue A, Sidorenko J, McRae AF *et al*. 2018. Signatures of negative selection in the genetic architecture of human complex traits. Nature genetics. 50:746–753.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. Nature genetics. 44:821–824.
Zhu H, Chen X, Zhang Y. 2013. Temporal and spatial variability of nitrogen in rice–wheat rotation in field scale. Environmental earth sciences. 68:585–590.