A historically balanced locus under recent directional selection in responding to changed nitrogen conditions during modern maize breeding

Gen Xu[1,2,†], Jing Lyu[1,2,3,†], Toshihiro Obata[2,4], Sanzhen Liu[5], Yufeng Ge[2,6], James C. Schnable[1,2], 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 68588, USA.
[3]Current address: Department of Biology, Indiana University, Bloomington, IN 47405, USA.
[4]Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA.
[5]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA.
[6]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68583, USA.
[†]These authors contributed equally to this work.
[*]Corresponding author: jinliang.yang@unl.edu

ABSTRACT

Understanding the patterns of selection during plant evolution and recent crop improvement processes is the central topic in plant breeding and genetics. As an essential macronutrient for plant growth and development, nitrogen (N) is a key factor in affecting plant adaptation and crop improvement. The widespread adoption of less expensive industrial N fixation has dramatically reshaped plant morphology by favoring compact maize plants to tolerant crowding stress. The associated genetic changes, however, have not been systematically studied. Here, we investigated maize inbred lines developed before and after the 1960s — the time point when inorganic N fertilizer started to be widely used for maize production. We identified a strong selective sweep exhibiting pronounced genomic differentiation between Old-Era (pre-1960s) and New-Era (post-1960s) inbred lines. Further study revealed population genetics statistics in the sweep exhibited patterns consistent with historical balancing selection. This balanced genomic interval is associated with a number of morphological, physiological, and metabolite traits related to vegetative N responses. A cluster of three glutamate receptor-like (GLR) genes is located within the region targeted by selection. Functional characterizations suggested differences in transcriptional activity of the GLR genes between the haplotypes carried by Old-Era and New-Era inbred lines likely play an essential role in mediating distinct N responses. The identification of both targets of selection and changes in the regulation of N responsive genes between maize lines developed in different eras sheds light on the N sensing and regulation pathways and paves the way to developing N resilient crops.

Introduction

Through usually early hybridization events followed by selective breeding, about 150 wild plants have been domesticated into crops to meet human needs (1), including the major cereal crops of maize, rice, and wheat (2). Understanding the selection forces during these domestication and improvement processes has long been the central topic in plant genetics and breeding. Depending on the allele effects relative to fitness, the modes of selection in a diploid species include positive selection to increase the frequencies of advantageous alleles, negative selection to remove the deleterious alleles, or balancing selection to maintain both alleles (3). Unlike advantageous or deleterious alleles, alleles under balancing selection are not universally beneficial or detrimental, whose fitness changes with time, space, or population frequency (4). With the increasing availability of population-level genomic data, a number of studies have been conducted to understand the patterns of advantages or deleterious alleles (5–9). However, studies focusing on balancing selection in plants are limited (10–13), likely due to the balanced alleles being difficult to detect (14), which prevent the accurate evaluation of the roles that balanced alleles played during the crop domestication processes.

Nitrogen (N), as one of the essential macronutrients, its availability changes with time and space and, therefore, plays a critical role in plant adaptation and recent crop improvement. N is a major constituent of proteins, nucleic acids, chlorophyll, coenzymes, phytohormones, and secondary metabolites (15; 16). Plants take up inorganic N mainly in the forms of nitrate (NO_3^-) and ammonium (NH_4^+) from agricultural soils via specific assimilation and mobilization processes (17–20). For most cereal crops, such as maize or sorghum, achieving high yields in an intensive agricultural system requires a large quantity of supplemental N fertilizer. However, N utilized by most plants ranges from 30% to 50 % (21), resulting in N runoff in farm fields to form nitrous oxide (N_2O) — a potent greenhouse gas that has 300 times the warming ability of carbon dioxide (CO_2). In addition to the substantial adverse effects on natural ecosystems and global warming (22; 23), the poor N usage imposes the economic cost on farmers and reduces human life expectancies around the globe (24).

Maize (*Zea mays* ssp. *mays L.*) is a major crop grown around the world and consumes 17% N fertilizers worldwide (21). In the past, the breeding efforts in maize mainly focused on increasing grain yield, resulting in steady yield improvement over the last century (25). Prior to the 1960s (Old-Era) in the U.S. Corn Belt, the selection in maize breeding primarily occurred in nitrogen-limited agricultural systems. Subsequence to the green revolution in the 1960s (New-Era), inorganic nitrogen fertilizers became increasingly available due to the Haber-Bosch process (26) and maize selection and breeding has been mainly conducted in systems where nitrogen was not the limiting constraint on productivity or yield. The shift in the crucial environmental factor of N availability has changed the breeders' preference to select hybrids with high planting density to take advantage of sufficient nitrogen fertilizers (27), resulting in a number of changes in physiological and morphological traits (28; 29). However, previous studies have not systematically examined the mode of selection act on the shifted nitrogen condition and to what extent the selection has reshaped the genomic architecture in affecting N responses.

In this study, we employed a set of Old-Era and New-Era maize inbred lines to evaluate genome wide signatures of selection to changing N conditions. We characterized N-related traits in field conditions under sufficient and N limited conditions in a two-year field trial and validated differences in phenotypic performance in controlled environment studies. Leveraging publicly available genomics dataset as well as newly generated phenomics, metabolomics, and transcriptomics datasets, our integrative analyses revealed a region which was historically under balancing selection because a direct target of positive selection during modern crop breeding. Functional characterization identified differences in transcriptional activity associated with differences in response to N. Our results shed light on the selection patterns of an N-associated locus and provide a potential target for developing N resilient crops in the future.

Results

Genome-wide selection scan identified a differentiated genomic region during recent breeding

We collected five in-field leaf physiological traits, including leaf nitrogen level, leaf chlorophyll content, leaf dry weight, leaf fresh weight, and leaf area from replicated field trials of the maize association panel (MAP) (30) grown under both conventional agronomic practices (high N, or HN) and under nitrogen-limited conditions (low N, or LN) in 2018 and 2019 (See **Materials and Methods**) using hyperspectral reflectance phenotyping (31) (**Table S1**). Among the 231 lines phenotyped, 37 have been previously classified as Old-Era inbred lines (i.e., lines developed before the 1960s) and 33 have been previously classified as New-Era inbred lines (i.e., lines developed post the 1960s), respectively (32) (**Table S2**). Old-Era inbred lines exhibited smaller differences than New-Era inbred lines for the leaf nitrogen level, leaf chlorophyll content, and leaf dry weight between plants grown in low N and high N conditions (**Figure 1A**, see **Figure S1** for other traits). In-field phenotypic data, especially the HN/LN ratios, suggested the New-Era lines were more responsive to take advantage of increased N availability at the vegetative stage, consistent with a previous study (33).

Employing both XP-CLR and F_{ST} approaches (**Materials and Methods**) to scan for signatures of selection between Old-Era and New-Era inbred lines using whole genome resequencing data resulted in the identification of 491 selective sweeps (**Table S3**). The regions identified exhibited significant overlap (75/491=15.3%, permutation test, P-value = 7×10^{-3}) with a set of regions linked to recent sweeps associated with maize improvement (29). The

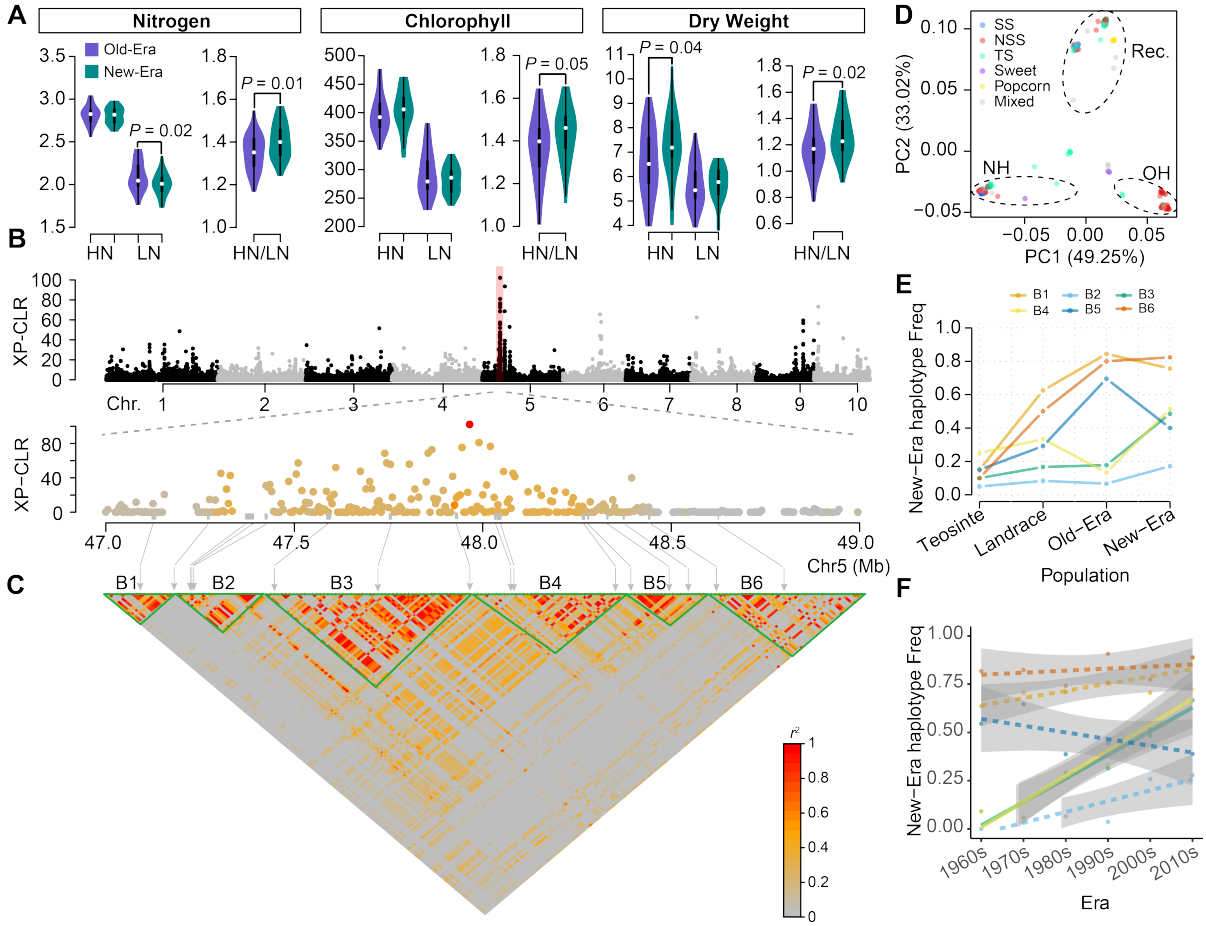

Figure 1. Phenomic and genomic characteristics of maize inbred lines developed before and after the 1960s. (**A**) Comparison of values for three leaf physiological traits scored across 70 maize inbreds grown under high (HN) and low N (LN) levels. (**B**) Results of a genome wide scan for selective sweeps between Old-Era and New-Era maize (see **Figure S2** for results using F_{ST} approach) including a zoomed in view of the XP-CLR scores individual windows located near the peak highlighted in red on chromosome 5. The color in the zoom-in plot reflects the LD level (r^2) of windows with the leading signal (the red dot). (**C**) Linkage disequilibrium (LD) relationship for genetic markers in the highlighted region of chromosome 5. The grey arrows indicate the positions o f annotated genes. Six LD blocks (block1 to block6) are indicated with green triangles and labeled B1-B6. (**D**) PCA analysis conducted using genetic markers of 271 maize inbred lines within LD Block4. Dashed ovals indicate clusters corresponding to three haplotypes within this region, one (NH) abundant in New-Era maize, one abundant in Old-Era maize and another set of lines carrying a recombinant haplotype. Individual points are color-coded by the subpopulations of maize inbreds assigned to in Flint-Garcia et al. 2005 (30). (**E**) The frequency of the New-Era haplotype of each LD block in populations of teosinte, landrace, Old- and New-Era maize inbred lines. (**F**) Changes in the frequency of the New-Era haplotype for each LD block in elite inbred lines developed in China between the 1960s to the 2010s. Solid and dashed lines indicate significant (linear regression analysis, P-value < 0.001) and nonsignificant linear regressions, respectively, with 95% confidence intervals for each regression indicated in grey.

most significant sweep was located on chromosome 5 and detected by both XP-CLR and F_{ST} approaches (**Figure 1B** and **Figure S2**). This sweep colocalized with genetic loci associated with nitrogen uptake efficiency (34), plant height (35), and grain weight per plant (36). We detected six linkage disequilibrium (LD) blocks (**Materials and Methods**) in the region from 1Mb upstream to 1 Mb downstream of the leading signal. These LD blocks (B2, B3, B4) exhibited pronounced genomic differentiation between New-Era and Old-Era inbred lines (**Figure 1B-C**).

For each LD block, we conducted haplotype analysis and assigned the New-Era and Old-Era haplotypes with a membership coefficient of Q >= 0.7 (**Materials and Methods**). As shown in **Figure 1D**, the New-Era, Old-Era, and recombinant haplotypes were each observed in maize inbreds from different subpopulations (30), suggesting population structure is unlikely to explain the observed pattern of LD (See **Figure S3** for other LD blocks). We found New-Era haplotypes of the B2, B3, and B4 LD blocks were also present at intermediate frequencies in teosinte (the maize wild ancestor) and maize landrace populations (**Figure 1E**). As expected, the frequency of the New-Era haplotypes for these three blocks were lower in Old-Era lines and then exhibit a dramatic increase in frequency after the 1960s. Consistent with the pattern observed in the MAP which composes primarily of lines developed in the Americas, in Chinese elite inbred lines (29), the frequencies of New-Era haplotypes for B3 and B4 also rose dramatically from 0.1 to 0.7 over the past 60 years (**Figure 1F**). Taken together, these data suggested the New-Era haplotype exists in the maize ancestral population and has undergone recent positive selection in both China and US elite maize populations.

Balancing selection maintains genetic variation at the N associated locus

In teosinte population, the B4 New-Era haplotype exhibited an intermediate frequency (> 0.2) (**Figure 1E**), higher than that in Old-Era maize inbreds, which drive us to hypothesize that this N associated locus may be under historical balancing selection (4). To address this hypothesis, we calculated the site frequency spectrum (SFS) for each LD block. Using sorghum as the ancestral alleles, we found the derived alleles, especially in the maize population, showing an excess of intermediate frequencies for B4 (**Figure 2A**), a signature consistent with balancing selection that maintains different alleles (in this case, both Old-Era and New-Era alleles) at the selected loci for a long evolutionary period (4). We observed a similar pattern for B2, B3, and B5 (**Figure S4B-D**) in the selective regions that was different from the genome-wide pattern (**Figure S4F**).

Historical balancing selection is predicted to result in high sequence diversity (4). Consistent with this model, nucleotide diversity (π) and Tajima's D results (**Figure S5**) in the region are significantly higher than genome-wide level, especially within the B4 block (**Figure 2B**). Furthermore, using a newly developed composite B statistics (37), we detected balancing selection signals in the chromosome 5 region for both teosinte and maize (**Figure 2C**). These results suggest the N associated locus might be a historically balanced site to maintain for both New-Era and Old-Era alleles.

The selective haplotypes associated with plant morphology, physiology, and metabolite traits

Old-Era and New-Era inbreds were further characterized under controlled environment conditions in the plant phenotyping facility (**Materials and Methods**). A haplotype-based association analysis (**Materials and Methods**) detected significant differences in the N content of the lower leaves of maize lines carrying the New-Era and Old-Era haplotypes of LD B4 block (**Figure 3A**), a pattern similar to the field data under low N condition (**Figure 1A**). In contrast to the field study, leaf chlorophyll content was not significantly different under controlled environment conditions (**Figure S6**). Plants carrying the B4 New-Era haplotypes exhibited significantly larger leaf areas, greater leaf dry weights, and more compact plant architectures (**Figure 3B**). Using data on the abundance of primary metabolites collected from leaf tissue of the same plants (40), we found the abundance of lysine ($C_6H_{14}N_2O_2$), an essential amino acid, was significantly higher in inbreds carrying the B4 New-Era haplotype at the B4 linkage block (**Figure 3B**). In contrast, the abundance of fructose ($C_6H_{12}O_6$) was significantly lower in inbreds carrying the New-Era haplotype at the B4 linkage block compared to inbreds carrying the Old-Era haplotype (**Figure 3B**). These differences in both leaf morphology and physiological characteristics are consistent with the view that modern maize lines were preferentially selected to take advantage of the N-oversupplied condition (41; 42).

In total, 17 genes were annotated in the extended area of the selective sweep (**Figure 1C**). Transcriptome data revealed that 10 of these 17 genes were differentially expressed (DE) between Old-Era and New-Era inbred

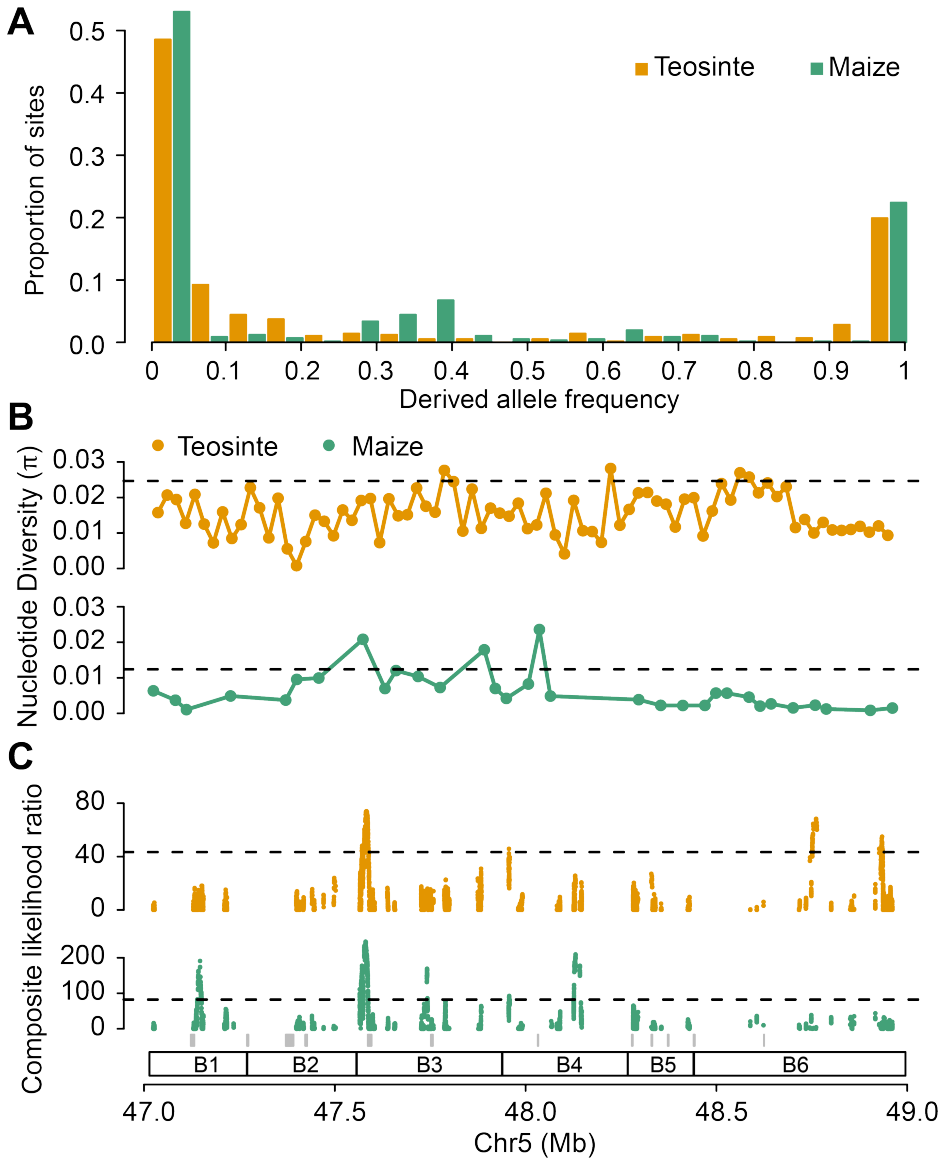

Figure 2. Site frequency spectrum (SFS) and neutrality test statistics at the N responsive locus. (**A**) The SFS of LD Block4 for teosinte and maize populations considering the allele shared with sorghum as the ancestral allele and the non-shared allele as the derived allele. Nucleotide diversity (π) (**B**) and composite likelihood ratio based on $B_{0,MAF}$ statistic (**C**) for the chromosome 5 region. The horizontal dashed lines represent the 5% significance level across the genome. The grey rectangles at the bottom of the panel C indicate the position of annotated gene models. The labels of B1 to B6 indicate the positions of the six LD blocks.

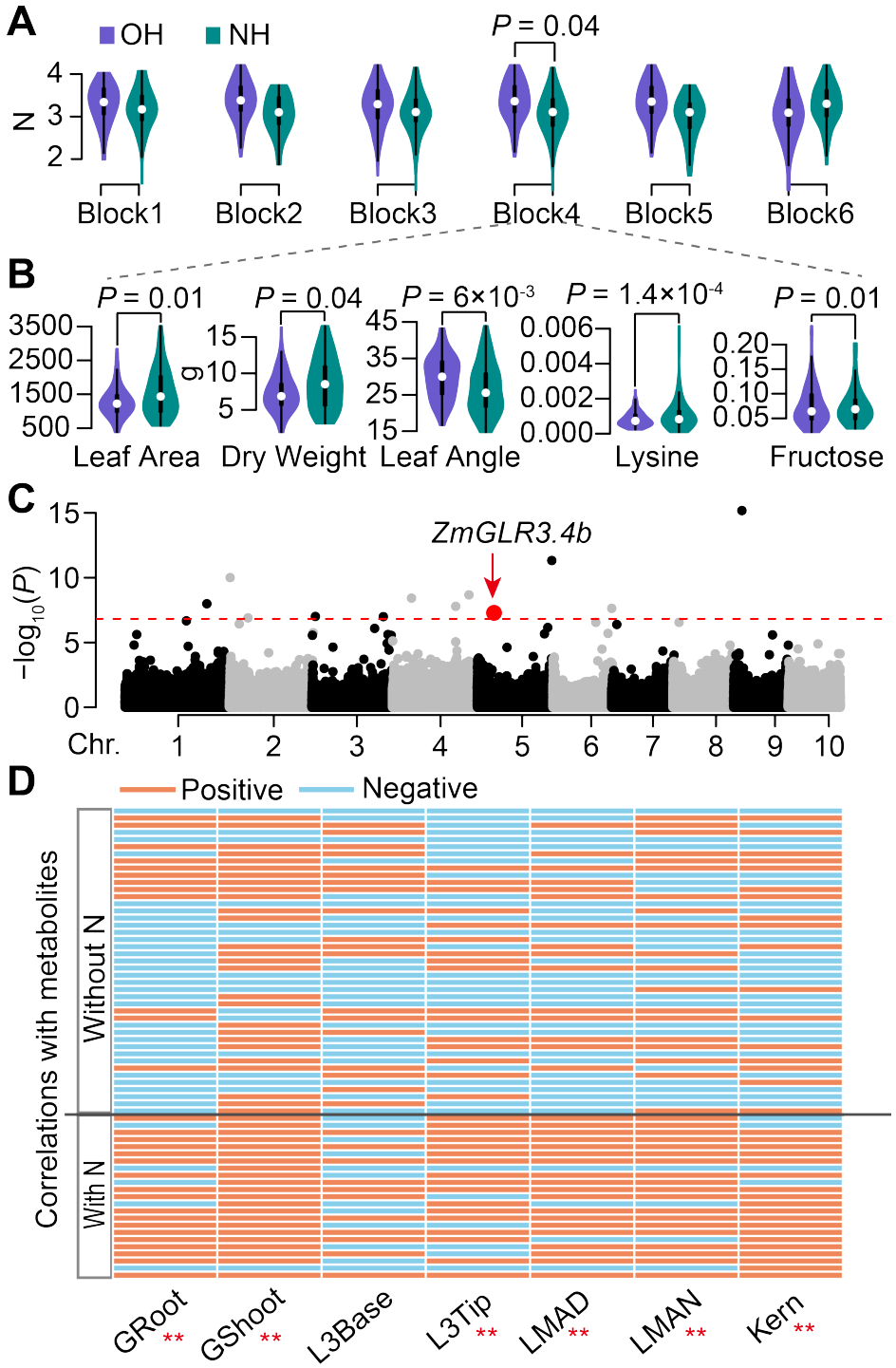

Figure 3. Association results for plant morphology, physiology, and metabolite traits. (**A**) Haplotype-based association analysis for leaf nitrogen level across six different LD blocks. (**B**) Phenotypic performance between Old-Era haplotype (OH) and New-Era haplotypes (OH) at LD block 4 (B4) for leaf area, leaf dry weight, leaf angle, leaf lysine content, and leaf fructose content traits. (**C**) The Manhattan plot for leaf chlorophyll *a* content using the NAM population (38). The red dot indicates the GWAS signal at chromosome 5 overlapped with *ZmGLR3.4b* gene in B4. The red horizontal dashed line denotes the Bonferroni threshold ($P < 1.5\times10^{-7}$). (**D**) Correlation analysis between gene expression of *ZmGLR3.4b* and 66 metabolites. "With N" and "Without N" denote the metabolites containing or not containing N in their chemical formulas. ** denotes Chi-squared test *P*-value < 0.01. Gene expression data were collected from seven tissues (39), including germinating root (GRoot), germinating shoot (GShoot), third leaf base (L3Base), third leaf tip (L3Tip), adult leaf during the day (LMAD), adult leaf during the night (LMAN), and kernel (Kern).

lines (two-sided Student's *t*-test, *P*-value <0.05) in at least one of the seven tissues (39). These DE genes are particularly common in the LD blocks B3 (2/3) and B4 (3/3) (**Figure S7**). Among the six genes within B3 and B4, noticeably, a cluster of three glutamate receptor-like (GLR) genes was identified. GWAS for N-related traits using public data collected from the NAM population (38) identified a significant signal for variation in the abundance of chlorophyll *a* within the second exon of *Zm00001d014456*, one of the three GLR genes located within B4 (**Figure 3C**). Phylogenetic analysis of 18 maize and 20 *Arabidopsis* GLRs indicated that the cluster of three GLRs in the selective sweep was most closely related to the Arabidopsis gene *AtGLR3.4* (**Figure S8**). We refer to them below as *ZmGLR3.4a* (*Zm00001d014451*), *ZmGLR3.4b* (*Zm00001d014456*), and *ZmGLR3.4c* (*Zm00001d014458*).

The mRNA abundance of one of the three GLR genes (*ZmGLR3.4b*) exhibited a statistically significant trend towards positive correlations with the abundance of N-containing metabolites, such as lysine ($C_6H_{14}N_2O_2$), serine ($C_3H_7NO_3$), allantoin ($C_4H_6N_4O_3$), gamma-aminobutyric acid (GABA, $C_4H_9NO_2$), and negative correlations with the abundance of metabolites that do not contain the element N, such as fructose ($C_6H_{12}O_6$), glyceric acid ($C_3H_6O_4$), puruvic acid ($C_3H_4O_3$) (**Figure 3D**). This pattern was observed when using expression data for the *ZmGLR3.4b* gene in six out of seven tissues evaluated. However, any pattern of correlation between gene expression and metabolite abundance was substantially less clear for other genes within the region (**Figure S9**).

Expression of ZmGLR3.4b is affected by altered cis-regulatory modulation

The overall expression levels of *ZmGLR3.4a* was much lower than that of *ZmGLR3.4b* or *ZmGLR3.4c* (**Figure S10**). Both *ZmGLR3.4b* and *ZmGLR3.4c* were predominantly expressed in leaf tissues. In the leaf three tip (L3Tip) and adult leaf collected during the day (LMAD), the expression levels of *ZmGLR3.4b* was significantly higher in New-Era than in Old-Era inbreds. In contrast, the expression of *ZmGLR3.4c* in L3Tip was significantly lower in New-Era inbred lines than in Old-Era inbreds (**Figure S10**). Genome-wide analysis identified significant *cis*-eQTL for both *ZmGLR3.4b* (**Figure 4A**) and *ZmGLR3.4c* (**Figure S11**).

Leveraging the *de novo* assembled maize genomes that include both Old-Era and New-Era haplotypes (43–45), we investigated the structural variation (SV) of the GLR genes. No apparent structural variation was present in the *ZmGLR3.4a* and *ZmGLR3.4c* genes between *de novo* assembled genomes for inbreds carrying the Old-Era and New-Era haplotypes (43–45) (**Figure S12**). However, two transposable element (TE) insertions, present in the first and third introns, distinguished *ZmGLR3.4b* in the genome of inbreds carrying the Old-Era and New-Era haplotypes (**Figure 4B**). The 2,786-bp TE insertion in the third intron in the New-Era haplotype, is likely associated with the spread of both CG (**Figure 4C**) and CHG (**Figure S13**) DNA methylation into the surrounding exons of the *ZmGLR3.4b* gene. Published H3K4me3 and H3K27ac ChIP-seq data (46; 47), as well as the STARR-seq (47) data, suggested that multiple putative promoters or enhancers exist at the selective region (i.e., B3 and B4). HiChIP data further illustrated physical contacts among the three GLR genes in B73, an inbred line carrying the New-Era haplotype (**Figure 4D**). In addition, *ZmGLR3.4b* physically interacted with two putative promoters or enhancers (highlighted areas in **Figure 4D**), as evidenced by the high STARR-seq or ChIP-seq peaks; one of the physical interactions overlaps with a differentially methylated region (DMR) that was previously identified between maize and teosinte (48). Taken together, these results suggest *cis*-regulatory modulations likely alter the transcriptional activities of the GLR genes.

Phenotypic and transcriptional responses under different N conditions

We conducted additional growth chamber experiments using four selected inbred lines (two lines for each Era) based on their haplotypes and field performance. After growing for two weeks with different N treatments, we harvested the aboveground and belowground tissues for phenotyping (**Materials and Methods**). The dry weights of both aboveground and belowground tissue of Old-Era lines were not significantly different between plants grown in high N and low N treatments (**Figure 5A**). A similar phenomenon was observed in the fresh weight (**Figure S14**), consistent with the N resilient effect of the Old-Era haplotype observed in the field. For the New-Era lines, both genotypes exhibited significantly better performances under high N as compared to low N conditions, except for the dry root weight of B73, consistent with previous observations that the New-Era haplotypes were more responsive to N.

The expression of *ZmGLR3.4a* was consistently low (FPKM < 2) in both public and newly generated RNA-seq

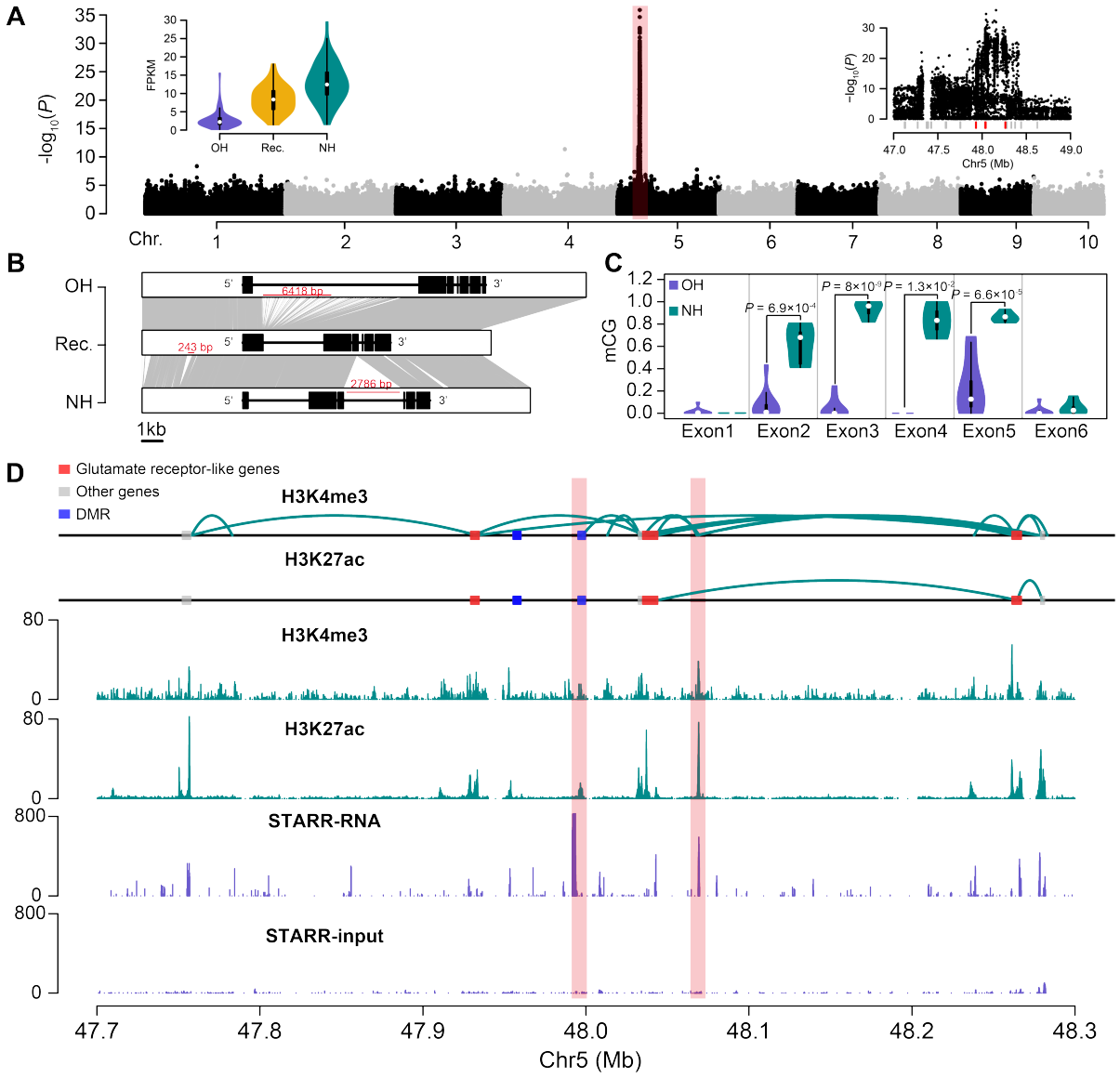

Figure 4. Functional genomic characterization of the GLR genes at the chromosome 5 genomic interval. (**A**) Results of a genome-wide eQTL analysis using the expression of the *ZmGLR3.4b* gene collected from third leaf tip as the trait. The distribution of the expression levels of the *ZmGLR3.4b* gene within the Old-Era (OH), New-Era (NH), and recombinant (Rec.) haplotypes are shown in the top left panel. The top right panel shows a zoom-in view of the region of the genome wide Manhattan plot highlighted in red. The positions of the three GLR genes in the top right panel are indicated by the three red tick marks. (**B**) Comparison of the annotated structure of the *ZmGLR3.4b* gene in *de novo* assembled genomes of maize lines carrying the Old-Era (OH, A188), New-Era (NH, B73), and recombinant (Rec., IL14H) haplotypes. (**C**) Levels of DNA methylation in exons of *ZmGLR3.4b* belonging to Old-Era (OH, $n = 9$) and New-Era haplotypes (NH, $n = 5$). *P* values were determined using a two-sided Student's *t*-test. (**D**) Physical interactions (two upper panels), colocalization with H3K27ac and H3K4me3 (two middle panels), and STARR profiles (two lower panels) around the GLR genes. Curved green lines denote interacting regions with evidence of physical contact. Red and gray boxes indicate the GLR and other gene models, respectively. Blue boxes indicate teosinte-maize differentially methylated regions. The regions highlighted in pink indicate anchors showing enhancer activities.

data from the plants grown in this study (**Table S4**), indicating it is a potentially malfunctional gene (**Figure 5B**). In leaf tissue, the abundance of mRNA transcripts derived from both *ZmGLR3.4b* (**Figure 5C**) and *ZmGLR3.4c* (**Figure 5D**) responded positively to high N treatments; in root, no apparent transcriptional reactions to the N treatments were observed for either gene. New-Era lines exhibited significantly higher expression of *ZmGLR3.4b* in leaf tissue than did Old-Era lines (fold change = 2.2, FDR corrected *P*-value or *q*-value = 4.6×10^{-8}) under both high N (fold change = 2.5, *q*-value = 1.5×10^{-8}) and low N (fold change = 2.1, *q*-value = 1.7×10^{-3}) conditions (**Figure 5C**). In root tissue, the opposite pattern was observed with a moderate decrease in the expression of *ZmGLR3.4b* New-Era inbred lines relative to Old-Era inbred lines (fold change = 2.4 for HN and fold change = 1.5 for LN). We did not detect any significant differences in the expression of *ZmGLR3.4c* between Old-Era and New-Era inbred lines (**Figure 5D**).

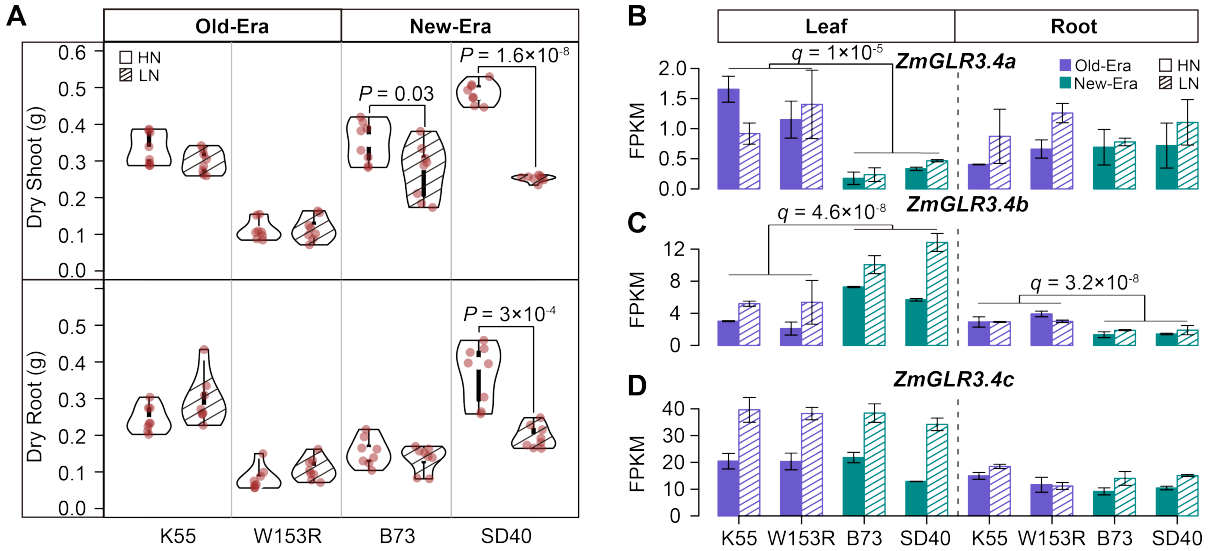

Figure 5. Phenotypic and transcriptional responses of selected Old-Era and New-Era inbred lines to different N treatments. (**A**) The dry weight of two-week-old shoot and root for Old-Era and New-Era inbred lines growing with high N (HN) and low N (LN) treatments . *P* values were determined using a two-sided Student's *t*-test. Gene expression of *ZmGLR3.4a* (**B**), *ZmGLR3.4b* (**C**), and *ZmGLR3.4c* (**D**) of Old-Era and New-Era inbred lines with different N treatments. FDR corrected *P*-value (*q*) was calculated between Old-Era and New-Era inbred lines.

We detected *n* = 2,264 differentially expressed (DE) genes in leave tissue and *n* = 699 DE genes in root tissue, comparing expression between the two N treatments (**Table S5**). These genes are referred to below as N responsive genes. Additionally, we detected *n* =1,600 leaf and *n* = 1,609 root DE genes between Old-Era and New-Era inbreds (**Table S5**). Notably, the Old-Era vs. New-Era DE genes were significantly enriched in the N responsive gene sets (**Figure S15**). KEGG analysis suggested the DE genes were highly enriched in metabolism pathways (**Figure S16**). Old-Era vs. New-Era DE genes were enriched in genes encoding enzymes from multiple amino acids metabolism pathways, including the glutamate metabolism pathway (**Figure S16**). Plant GLRs are largely glutamate non-specific and can be gated by a broad range of amino acids (49). To gain further insight into the roles of the *ZmGLR3.4* genes, we employed the predicted protein-protein interaction (PPI) networks using New-Era and Old-Era DE genes in the leaf tissue. After pulling down the network involving *ZmGLR3.4a* and *ZmGLR3.4b* (*ZmGLR3.4c* was not differentially expressed between Old-Era and New-Era inbreds as shown in **Figure 5D**), we found *ZmGLR3.4* genes positioned in between the N synthesis and transportation pathways and the ion signaling pathways (**Figure 6**). Noticeably, a number of known N assimilation and transportation genes, i.e., *ZmNIR1.1* (*Zm00001d052165*) (50), *ZmNR1.1* (*Zm00001d018206*) (50), *ZmAMT1.1B* (*Zm00001d003025*) (51), were up-regulated in New-Era lines, while the ion transporters, i.e., *ZmHKT1* (*Zm00001d040627*) (52), *cbl4* (*Zm00001d038730*), *ZmHAK25* (*Zm00001d017925*) (53), were down-regulated.

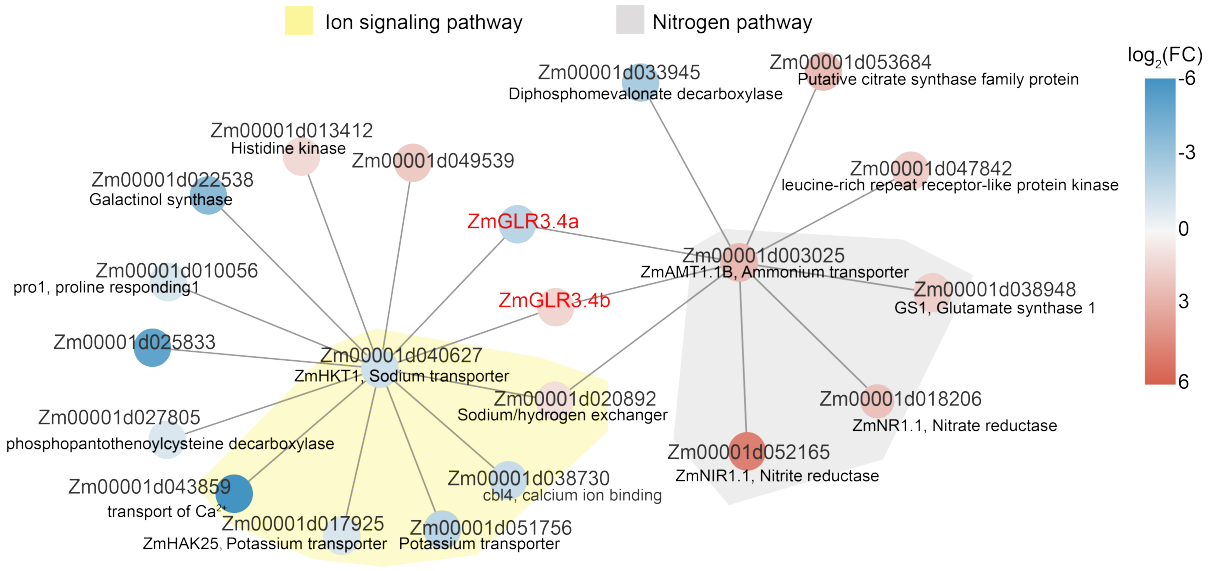

Figure 6. Network analysis of differentially expressed (DE) genes between Old-Era and New-Era inbreds. Protein-protein interaction network predicted from Old-Era vs. New-Era DE genes detected in the leaf tissue. The gradient colors of the dots denote the log2 fold change of NE/OE ratio. The yellow and gray colors shaded areas highlight genes involved in ion signaling and nitrogen pathways, respectively.

Discussion

In this study, we found an N-associated locus that remained at an intermediate frequency from teosinte to landrace but rapidly increased its frequency in maize inbred lines developed after the 1960s when inorganic N fertilizer became increasingly available in maize production. This genetic locus is associated with vegetative N-related traits, as evidenced by field, greenhouse, and growth chamber experiments. Noticeably, the New-Era haplotype is more responsive to high N conditions, and the Old-Era haplotype is more resilient to N stress. Population genomics analyses detected an excess of SNPs with intermediate frequencies, and the sweep showed slightly increased nucleotide diversity. Additionally, a weak balancing selection signal was detected in the region. Because we only considered SNPs for the balancing selection scan, the actual sequence diversity might be underestimated due to large insertion or deletion polymorphisms. By leveraging the recently *de novo* assembled maize genomes that include both Old-Era and New-Era haplotypes, structural variation (SV) analyses found two transposon insertions in the first and third introns of the *ZmGLR4.3b* gene. Both transposable elements are historical insertion events as their sequences have been largely degraded, consistent with the idea that long-term balancing selection might maintain both Old-Era and New-Era haplotypes.

Notably, three glutamate receptor-like (GLR) genes in a tandem array were located within the selected region. The GLR genes, plant homologs of mammalian ionotropic glutamate receptors (iGluR), have been hypothesized to play a crucial role in sensing the amino acid level at the cellular level (17). Our data revealed that the gene expression levels of *ZmGLR4.3b*, the strongest candidate in the GLR gene cluster, are positively correlated with the abundance of primary metabolites containing N and negatively correlated with metabolites lacking N element. These correlations are consistent with the potential role of the GLR genes in regulating the N/C metabolic balance, as suggested by studies in Arabidopsis (54; 55). Unlike in mammals, the plant GLR genes were reported as broad-spectrum amino acid receptors (56; 57). Glutamine and glutamate, as the products of the N assimilation, are precursors of other amino acids (58; 59). The binding of GLR with amino acid likely induces a conformational change and opens the ion channel (60). Consistent with this idea, our PPI network analysis positioned GLR genes between two groups of functional genes, one group for nitrogen biosynthesis or transportation and the other for ion signaling exchanging or transportation. After ions pass through the channel, the ion signaling can potentially feedback in regulating N uptake and utilization (17; 49; 57). Note that the known N uptake and transportation genes in the PPI network were

upregulated in New-Era inbred lines, likely in responding to the ion signaling feedback.

Collectively, our results suggest a functional role of the GLR genes in N responses. Population genomics analyses show signs of historical balancing selection for a selective sweep region harboring a cluster of three GLR homologous genes. In addition, the GWAS study detects genomic variation associated with several N-related traits. Around the GLR genes, we also identify functional variations, i.e., SV and DMRs, and other genomic features important for gene activation or chromatin interaction. These results suggest that further investigation of the GLRs in affecting N-related traits is warranted and that Old-Era alleles may provide a promising alternative allele for tolerating N stress and developing N resilient crops.

Materials and Methods

Field experimental design

The Maize Association Panel (30) including Old-Era and New-Era inbred lines were grown at the Havelock Research Farm of the University of Nebraska-Lincoln using an incomplete split plot block design in 2018 and 2019. Two nitrogen treatments were applied to the association panel, one under low nitrogen condition (no additional N fertilizer) and the other following conventional N application practice at the rate of 135 kg/ha dry urea. Each treatment was replicated twice in each year, with each replicate consisting of 288 plots including both the lines of the Maize Association panel and between 27 and 37 plots planted with check varieties. Each plot was 1.6 m wide and 6.3 m long, comprising of two rows, 38 plants were grown in each row. In-field vegetative N related traits were quantified follow a high-throughput phenotyping protocol described previously (31).

Calculation of the best linear unbiased predictions for the field phenotypic data

To minimize the effects of environmental variation, best linear unbiased predictions (BLUPs) were performed using the R package lme4 (61) to estimate the phenotypic value. The BLUP model was:

$$y_{ij} = \mu + G_i + E_j + \varepsilon_{ij}$$

where y_{ij} is the observed phenotype for the i^{th} genotype of the j^{th} environment, μ is the overall mean, G_i is the random genotypic effect of the ith genotype, E_j is the random effect of the jth environment, and ε_{ij} is a random residual error.

Plant materials and growth conditions in the controlled environments

We conducted nitrogen treatment experiments in the growth chamber with a photoperiod of 16/8 h at 28/24°C (light/dark). Seeds of two Old-Era inbreds (K55 and W153R) and two New-Era inbreds (SD40 and B73) were sterilized in 75% (v/v) ethanol for 15 min, washed with distilled water, and then soaked in distilled water for 8 h. Afterward, two seeds were planted in a plastic pot consisting of an autoclaved mixture (volume based) of 50% medium size (0.5–0.3 mm) commercial grade sand, 30% fine vermiculite, 20% MetroMix 200. Two days before planting, each tray containing 12 pots was irrigated with 3 L of a nutrient solution adjusted to pH 5.8. The high-N treatment nutrient solution contained 6.5 mM KNO_3, 4 mM Ca $(NO_3)_2{\cdot}4H_2O$, 1 mM $NH_4H_2PO_4$, 2 mM $MgSO_4{\cdot}7H_2O$, 46 M H_3BO_3, 9 M $MnCl_2{\cdot}4H_2O$, 7 M $ZnSO_4{\cdot}7H_2O$, 0.8 M $Na_2MoO_4{\cdot}2H_2O$, 0.32 M $CuSO_4{\cdot}5H_2O$, 77 M Fe-EDTA. The low-N treatment nutrient solution contained 0.65 mM KNO_3, 4.95 mM KCl, 0.40 mM Ca $(NO_3)_2{\cdot}4H_2O$, 3.60 mM $CaCl_2{\cdot}2H_2O$, 0.10 mM $NH_4H_2PO_4$, 0.90 mM KH_2PO_4, 2 mM $MgSO_4{\cdot}7H_2O$, 46 M H_3BO_3, 9 M $MnCl_2{\cdot}4H_2O$, 7 M $ZnSO_4{\cdot}7H_2O$, 0.8 M $Na_2MoO_4{\cdot}2H_2O$, 0.32 M $CuSO_4{\cdot}5H_2O$, 77 M Fe-EDTA. On day 5, each pot was thinned to one plant and received 50 mL of high-N or low-N nutrient solution every two days. After two weeks, the shoots and roots were harvested for further experiments.

Sequence Alignment and Phylogenetic Analysis

We downloaded the GLR protein sequences of *Arabidopsis thaliana* from the TAIR database (http://www.arabidopsis.org/), which have been reported in a previous study (62). Then the *Arabidopsis* GLR family protein sequences were aligned to the maize protein database using BLASTP (63) with an e-value of 10^{-5} to get a list of hits against maize protein. The GLR protein sequences of *Arabidopsis* as well as their orthologous genes in

maize were aligned with MUSCLE (64) using MEGA6 software (65). A neighbor-joining (NJ) method was then used for phylogenetic tree construction, with 1,000 bootstrap resampling.

Linkage disequilibrium (LD) analysis and haplotype construction

We estimated LD with the r^2 statistics using plink1.9 (66). Heat maps of pairwise LD between SNPs were plotted using the R package LDheatmap (67). The R package gpart (68) was used to partition the sweep region into blocks. Haplotypes of each block was determined by Admixture (69), individuals with membership coefficients of $Q >= 0.7$ were assigned to a specific haplotype. We defined the haplotypes from New-Era and Old-Era inbreds as New-Era and Old-Era haplotypes, respectively. Haplotype based target association mapping was performed for each block using the first three principal components as the covariates in the model.

Genome-wide association study

To determine the contribution of regulatory variants that influence gene expression of *ZmGLR3.4* genes, we conducted eQTL analysis using a mixed linear model (MLM) implemented in GEMMA (v 0.98.3) (70). The genotype was downloaded from maize HapMap 3 (71) and gene expression data was obtained from Kremling et al., 2018. The kinship matrices and the first three principal components were calculated by GEMMA (70) and Plink 1.9 (66), respectively and then fitted into the GWAS model. GWAS for leaf chlorophyll *a* content in the NAM population (38) was performed using FarmCPU method implemented in the R package rMVP (72). The NAM genotypic data was downloaded from the Panzea website (`http://www.panzea.org`).

Genome-wide scanning to detect selective signals.

We performed genome scanning for selective signals using F_{ST} (73) and the latest version of XP-CLR (74) approaches. In the XP-CLR analysis, we used a 50 kb sliding window and a 5 kb step size. To ensure comparability of the composite likelihood score in each window, we fixed the number assayed in each window to 200 SNPs. In the F_{ST} analysis, we calculated F_{ST} using VCFtools (75) with a 50 kb sliding window and a 5 kb step size. We merged nearby windows falling into the 10% tails into the same window. After window merging, we considered the 0.1% outliers as the selective sweeps.

Detection of balancing selection

According to the previous study (76; 77), the locus under balancing selection has an elevated nucleotide diversity (π), Tajima's D, and site frequency spectrum. Utilizing Sorghum bicolor alleles as the ancestral state, we calculated the site frequency spectrum for each block at the selection region. The nucleotide diversity (π) and Tajima's D were estimated using ANGSD software (78). All bam files of 17 teosinte, 23 landrace and 269 modern maize lines were derived from the Maize HapMap3 panel (71) which were downloaded from CyVerse database (`/iplant/home/shared/panzea/raw_seq_282/bam/`) as described in Panzea database (`www.panzea.org`). In the analysis, we first inferred the unfolded genome-wide site frequency spectrum (SFS) and calculated the thetas for each site. We then used the "thetaStat" program, which was implemented in ANGSD, to summarize the nucleotide diversity and Tajima's D values with 25 kb non-overlapping sliding windows. We also calculated the $B_{0,MAF}$ statistics using BalLeRMix (version 2.3) (37) with default parameters.

RNA sequencing and data analysis

Two Old-Era inbreds (K55 and W153R) and two New-Era inbreds (SD40 and B73) that were planted in the growth chamber were used for RNA-seq. For each genotype, we conducted high nitrogen and low nitrogen treatments. For each treatment, we used two biological replicates for conducting RNA-seq. Two weeks after sowing, shoot and root were harvested. Total RNA was extracted and purified using the RNeasy Plant mini kit (Qiagen), followed by DNase treatment. We sequenced the libraries using Illumina NovaSeq 6000 with 150 bp paired-end reads. Raw reads were trimmed and preprocessed using fastp (79) in default settings. Using the "Rsubread" software package (80), all clean reads were then mapped to the B73 reference sequence (AGPV4) (81; 82) by the "align" function, and transcript counts were extracted by using the "featureCounts" function. We carried out differential gene expression analysis using DESeq2 (83). A gene was identified as differentially expressed if the false discovery rate (FDR) is < 0.05 and

have at least twofold expression change. The expression of each gene was normalized to fragments per kilobase of transcript per million reads (FPKM).

Pathway enrichment and protein-protein interaction analysis

We performed pathway enrichment analysis using KOBAS v2.0 (84) with a significance cutoff of P-value < 0.01. Protein–protein interaction (PPI) networks was established using STRING v11 with a combined score >= 0.4 (85) and visualized using Cytoscape app (86).

Data and code availability

RNA-seq data produced from this study have been deposited in the NCBI GEO database under PRJNA800008. The code used for the analyses can be accessed through GitHub repository (https://github.com/GenXu1/Nitrogen_project).

Acknowledgements

We thank J. Ross-Ibarra for comments on the manuscript. 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, the National Science Foundation under award number OIA-1557417 for Center for Root and Rhizobiome Innovation (CRRI), and the University of Nebraska-Lincoln Start-up fund. This work was conducted using the Holland Computing Center of the University of Nebraska-Lincoln, which receives supports from the Nebraska Research Initiative.

Author contributions

J.Y. designed this work. J.L., T.O., Y.G., J.C.S. generated the data. G.X., J.L., and J.Y. analyzed the data. S.L. provided conceptual advice. J.Y. and G.X. wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Fernie, A. R. & Yan, J. De novo domestication: an alternative route toward new crops for the future. *Mol. plant* **12**, 615–631 (2019).

2. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).

3. Bamshad, M. & Wooding, S. P. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4**, 99–110 (2003).

4. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS genetics* **2**, e64 (2006).

5. Zhou, Z. *et al.* Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. biotechnology* **33**, 408–414 (2015).

6. Zhao, H. *et al.* Analysis of 427 genomes reveals moso bamboo population structure and genetic basis of property traits. *Nat. communications* **12**, 1–12 (2021).

7. Wang, L. *et al.* The interplay of demography and selection during maize domestication and expansion. *Genome biology* **18**, 1–13 (2017).

8. Lozano, R. *et al.* Comparative evolutionary genetics of deleterious load in sorghum and maize. *Nat. Plants* **7**, 17–24 (2021).

9. Yang, J. *et al.* Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLoS genetics* **13**, e1007019 (2017).

10. Qiu, J. *et al.* Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nat. communications* **8**, 1–12 (2017).

11. Liu, C. *et al.* Early selection of bzip73 facilitated adaptation of japonica rice to cold climates. *Nat. Commun.* **9**, 1–12 (2018).

12. Okuyama, Y. *et al.* A multifaceted genomics approach allows the isolation of the rice pia-blast resistance gene consisting of two adjacent nbs-lrr protein genes. *The Plant J.* **66**, 467–479 (2011).

13. Zhai, C. *et al.* The isolation and characterization of pik, a rice blast resistance gene which emerged after rice domestication. *New Phytol.* **189**, 321–334 (2011).

14. Fijarczyk, A. & Babik, W. Detecting balancing selection in genomes: limits and prospects. *Mol. ecology* **24**, 3529–3545 (2015).

15. O'Brien, J. A. *et al.* Nitrate transport, sensing, and responses in plants. *Mol. plant* **9**, 837–856 (2016).

16. Weih, M., Westerbergh, A. & Lundquist, P.-O. Role of nutrient-efficient plants for improving crop yields: bridging plant ecology, physiology, and molecular biology. In *Plant Macronutrient Use Efficiency*, 31–44 (Elsevier, 2017).

17. Krapp, A. Plant nitrogen assimilation and its regulation: a complex puzzle with missing pieces. *Curr. opinion plant biology* **25**, 115–122 (2015).

18. Tegeder, M. & Masclaux-Daubresse, C. Source and sink mechanisms of nitrogen transport and use. *New phytologist* **217**, 35–53 (2018).

19. Vidal, E. A. & Gutierrez, R. A. A systems view of nitrogen nutrient and metabolite responses in arabidopsis. *Curr. opinion plant biology* **11**, 521–529 (2008).

20. Xu, G., Fan, X. & Miller, A. J. Plant nitrogen assimilation and use efficiency. *Annu. review plant biology* **63**, 153–182 (2012).

21. Ahmed, M., Rauf, M., Mukhtar, Z. & Saeed, N. A. Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health. *Environ. Sci. Pollut. Res.* **24**, 26983–26987 (2017).

22. Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**, 889–892 (2008).

23. Wang, Y.-Y., Hsu, P.-K. & Tsay, Y.-F. Uptake, allocation and signaling of nitrate. *Trends plant science* **17**, 458–467 (2012).

24. Gu, B. *et al.* Abating ammonia is more cost-effective than nitrogen oxides for mitigating pm2. 5 air pollution. *Science* **374**, 758–762 (2021).

25. Andorf, C. *et al.* Technological advances in maize breeding: past, present and future. *Theor. Appl. Genet.* **132**, 817–849 (2019).

26. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nat. geoscience* **1**, 636–639 (2008).

27. Duvick, D. N. & Cassman, K. G. Post–green revolution trends in yield potential of temperate maize in the north-central united states. (1999).

28. Gage, J. L., White, M. R., Edwards, J. W., Kaeppler, S. & de Leon, N. Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. *Genetics* **210**, 1125–1138 (2018).

29. Wang, B. *et al.* Genome-wide selection and genetic improvement during modern maize breeding. *Nat. Genet.* **52**, 565–571 (2020).

30. Flint-Garcia, S. A. *et al.* Maize association population: a high-resolution platform for quantitative trait locus dissection. *The plant journal* **44**, 1054–1064 (2005).

31. Ge, Y. *et al.* 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 (2019).

32. van Heerwaarden, J., Hufford, M. B. & Ross-Ibarra, J. Historical genomics of north american maize. *Proc. Natl. Acad. Sci.* **109**, 12420–12425 (2012).

33. York, L. M., Galindo-Castañeda, T., Schussler, J. R. & Lynch, J. P. Evolution of us maize (zea mays l.) root architectural and anatomical phenes over the past 100 years corresponds to increased tolerance of nitrogen stress. *J. Exp. Bot.* **66**, 2347–2358 (2015).

34. He, K. *et al.* Mining of candidate genes for nitrogen use efficiency in maize based on genome-wide association study. *Mol. Breed.* **40**, 1–17 (2020).

35. Peiffer, J. A. *et al.* The genetic architecture of maize height. *Genetics* **196**, 1337–1356 (2014).

36. Xu, C. *et al.* Genome-wide association study dissects yield components associated with low-phosphorus stress tolerance in maize. *Theor. applied genetics* **131**, 1699–1714 (2018).

37. Cheng, X. & DeGiorgio, M. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Mol. biology evolution* **37**, 3267–3291 (2020).

38. Zhang, N. *et al.* Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol.* **168**, 575–583 (2015).

39. Kremling, K. A. *et al.* Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).

40. Yang, Z., Xu, G., Zhang, Q., Obata, T. & Yang, J. Genome-wide mediation analysis: bridging the divide between genotype and phenotype via transcriptomic data in maize. *bioRxiv* (2021).

41. Duvick, D. N. The contribution of breeding to yield advances in maize (zea mays l.). *Adv. agronomy* **86**, 83–145 (2005).

42. Tian, F. *et al.* Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. genetics* **43**, 159–162 (2011).

43. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).

44. Lin, G. *et al.* Chromosome-level genome assembly of a regenerable maize inbred line a188. *Genome biology* **22**, 1–30 (2021).

45. Springer, N. M. *et al.* The maize w22 genome provides a foundation for functional genomics and transposon biology. *Nat. genetics* **50**, 1282–1288 (2018).

46. Li, E. *et al.* Long-range interactions between proximal and distal regulatory regions in maize. *Nat. communications* **10**, 1–14 (2019).

47. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat. plants* **5**, 1237–1249 (2019).

48. Xu, G. *et al.* Evolutionary and functional genomics of dna methylation in maize domestication and improvement. *Nat. communications* **11**, 1–12 (2020).

49. Gent, L. & Forde, B. G. How do plants sense their nitrogen status? *J. Exp. Bot.* **68**, 2531–2539 (2017).

50. Ge, M. *et al.* The nin-like protein 5 (zmnlp5) transcription factor is involved in modulating the nitrogen response in maize. *The Plant J.* **102**, 353–368 (2020).

51. Dechorgnat, J. *et al.* Tissue and nitrogen-linked expression profiles of ammonium and nitrate transporters in maize. *BMC plant biology* **19**, 1–13 (2019).

52. Zhang, M. *et al.* A hak family na+ transporter confers natural variation of salt tolerance in maize. *Nat. Plants* **5**, 1297–1308 (2019).

53. Zhang, F. *et al.* Characterization of the calcineurin b-like (cbl) gene family in maize and functional analysis of zmcbl9 under abscisic acid and abiotic stress treatments. *Plant Sci.* **253**, 118–129 (2016).

54. Kang, J. & Turano, F. J. The putative glutamate receptor 1.1 (atglr1. 1) functions as a regulator of carbon and nitrogen metabolism in arabidopsis thaliana. *Proc. Natl. Acad. Sci.* **100**, 6872–6877 (2003).

55. Kang, J., Mehta, S. & Turano, F. J. 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.* **45**, 1380–1389 (2004).

56. Alfieri, A. *et al.* The structural bases for agonist diversity in an arabidopsis thaliana glutamate receptor-like channel. *Proc. Natl. Acad. Sci.* **117**, 752–760 (2020).

57. Forde, B. G. & Roberts, M. R. Glutamate receptor-like channels in plants: a role as amino acid sensors in plant defence? *F1000prime reports* **6** (2014).

58. Pratelli, R. & Pilot, G. Regulation of amino acid metabolic enzymes and transporters in plants. *J. Exp. Bot.* **65**, 5535–5556 (2014).

59. Tabuchi, M., Abiko, T. & Yamaya, T. Assimilation of ammonium ions and reutilization of nitrogen in rice (oryza sativa l.). *J. experimental botany* **58**, 2319–2327 (2007).

60. Green, M. N. *et al.* Structure of the arabidopsis thaliana glutamate receptor-like channel glr3. 4. *Mol. Cell* **81**, 3216–3226 (2021).

61. Bates, D., Maechler, M., Bolker, B. *et al.* lme4: Linear mixed-effects models using s4 classes. r package version 0.999999-0 (2012).

62. Davenport, R. Glutamate receptors in plants. *Annals botany* **90**, 549–557 (2002).

63. Camacho, C. *et al.* Blast+: architecture and applications. *BMC bioinformatics* **10**, 1–9 (2009).

64. Edgar, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).

65. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. Mega6: molecular evolutionary genetics analysis version 6.0. *Mol. biology evolution* **30**, 2725–2729 (2013).

66. Chang, C. C. *et al.* Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015 (2015).

67. Shin, J.-H., Blay, S., McNeney, B. & Graham, J. Ldheatmap: an r function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. statistical software* **16**, 1–9 (2006).

68. Kim, S. A. *et al.* gpart: human genome partitioning and visualization of high-density snp data by identifying haplotype blocks. *Bioinformatics* **35**, 4419–4421 (2019).

69. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655–1664 (2009).

70. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. genetics* **44**, 821–824 (2012).

71. Bukowski, R. *et al.* Construction of the third-generation zea mays haplotype map. *Gigascience* **7**, gix134 (2018).

72. Yin, L. *et al.* rmvp: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, proteomics & bioinformatics* (2021).

73. Weir, B. S. & Cockerham, C. C. Estimating f-statistics for the analysis of population structure. *evolution* 1358–1370 (1984).

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

75. Danecek, P. *et al.* The variant call format and vcftools. *Bioinformatics* **27**, 2156–2158 (2011).

76. Lenz, T. L., Spirin, V., Jordan, D. M. & Sunyaev, S. R. Excess of deleterious mutations around hla genes reveals evolutionary cost of balancing selection. *Mol. biology evolution* **33**, 2555–2564 (2016).

77. Nunez, J. C. *et al.* Ecological load and balancing selection in circumboreal barnacles. *Mol. biology evolution* **38**, 676–685 (2021).

78. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. Angsd: analysis of next generation sequencing data. *BMC bioinformatics* **15**, 1–13 (2014).

79. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

80. Liao, Y., Smyth, G. K. & Shi, W. The r package rsubread is easier, faster, cheaper and better for alignment and quantification of rna sequencing reads. *Nucleic acids research* **47**, e47–e47 (2019).

81. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).

82. Schnable, P. S. *et al.* The b73 maize genome: complexity, diversity, and dynamics. *science* **326**, 1112–1115 (2009).

83. Love, M., Anders, S. & Huber, W. Differential analysis of count data–the deseq2 package. *Genome Biol* **15**, 10–1186 (2014).

84. Xie, C. *et al.* Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research* **39**, W316–W322 (2011).

85. Szklarczyk, D. *et al.* String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**, D607–D613 (2019).

86. Demchak, B. *et al.* Cytoscape: the network visualization tool for genomespace workflows. *F1000Research* **3** (2014).

Supporting Information

Supporting Tables

Table S1. The best linear unbiased predictors (BLUP) values of leaf nutrients in maize association panel. (https://github.com/GenXu1/Nitrogen_project/tree/main/table/TableS1_field_phenotype.xlsx)

Table S2. Old- and New-Era samples. (https://github.com/GenXu1/Nitrogen_project/tree/main/table/TableS2_Old_New_samples.xlsx)

Table S3. Selective sweeps detected between Old- and New-Era inbreds. (https://github.com/GenXu1/Nitrogen_project/tree/main/table/TableS3_Sweeps.xlsx)

Table S4. Summary for RNA-Seq data. (https://github.com/GenXu1/Nitrogen_project/tree/main/table/TableS4_RNA_seq_data_summary.xlsx)

Table S5. Differentially expressed (DE) genes detected in this study. (https://github.com/GenXu1/Nitrogen_project/tree/main/table/TableS5_differentially_expressed_genes.xlsx)

Supporting Figures

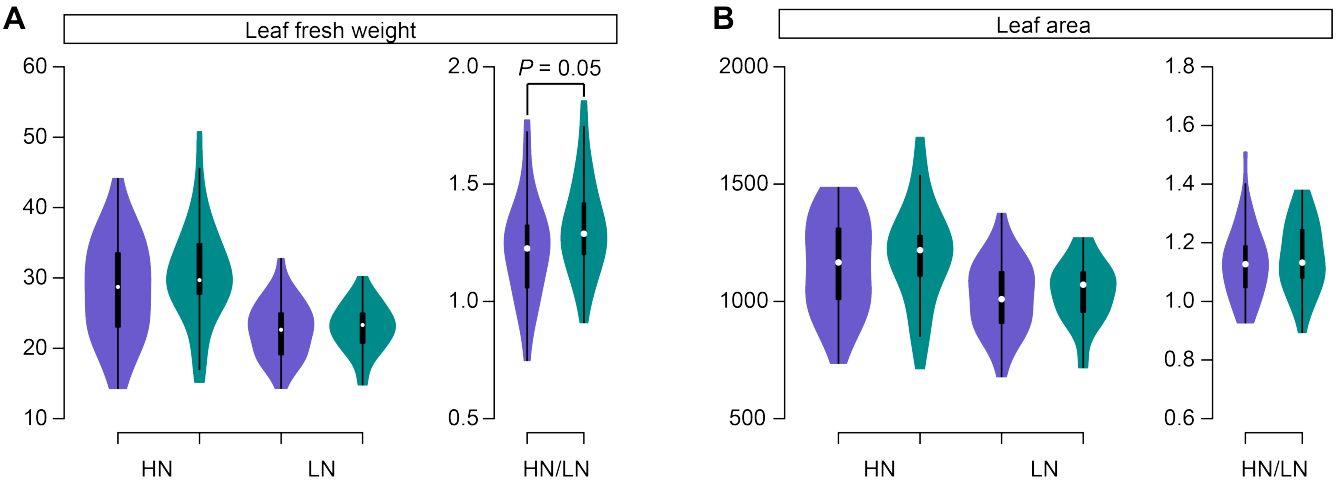

Figure S1. In-field phenotypic performance of Old-Era and New-Era inbred lines. Leaf fresh weight ((**A**)) and leaf area ((**B**)) under high N and low N conditions, and the ratio of high N over low N.

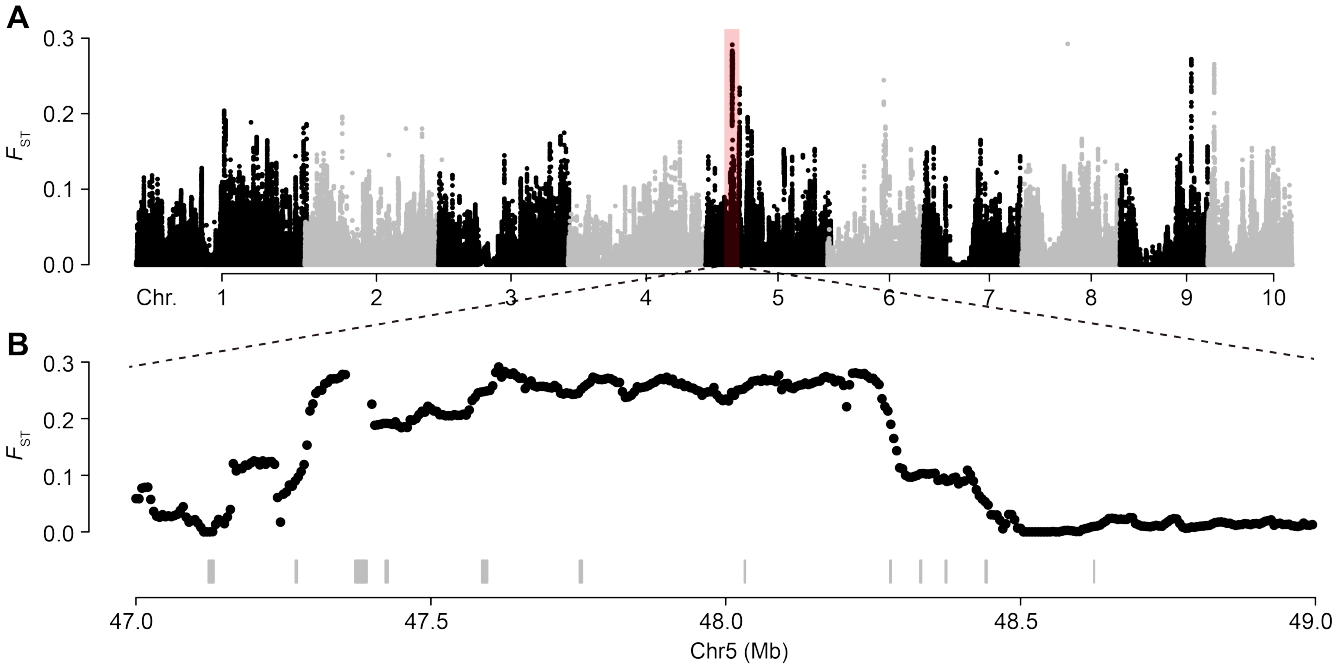

Figure S2. Selective sweeps detected using F_{ST} approach. Genome-wide F_{ST} values **A** and the zoom-in plot of the highlighted region on chromosome 5 **B**.

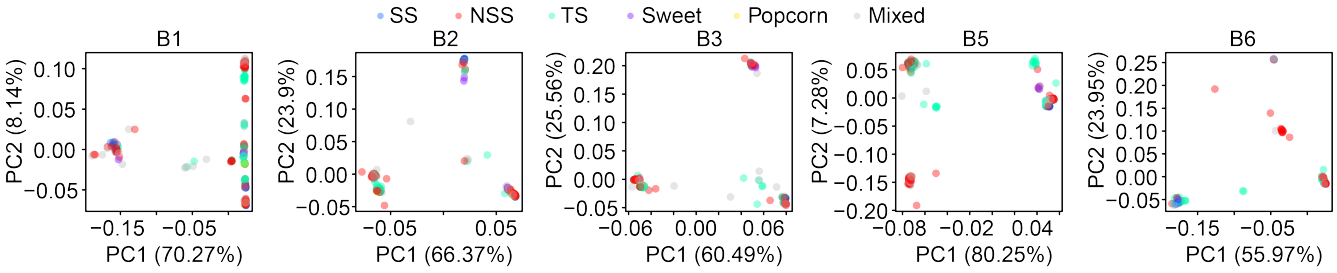

Figure S3. The principal component analysis (PCA) plots for linkage disequilibrium (LD) blocks. Colors represent subpopulations based on Flint-Garcia et al. 2005. B1, B2, B3, B5, and B6 denotes LD block1, block2, block3, block5 and block6, respectively.

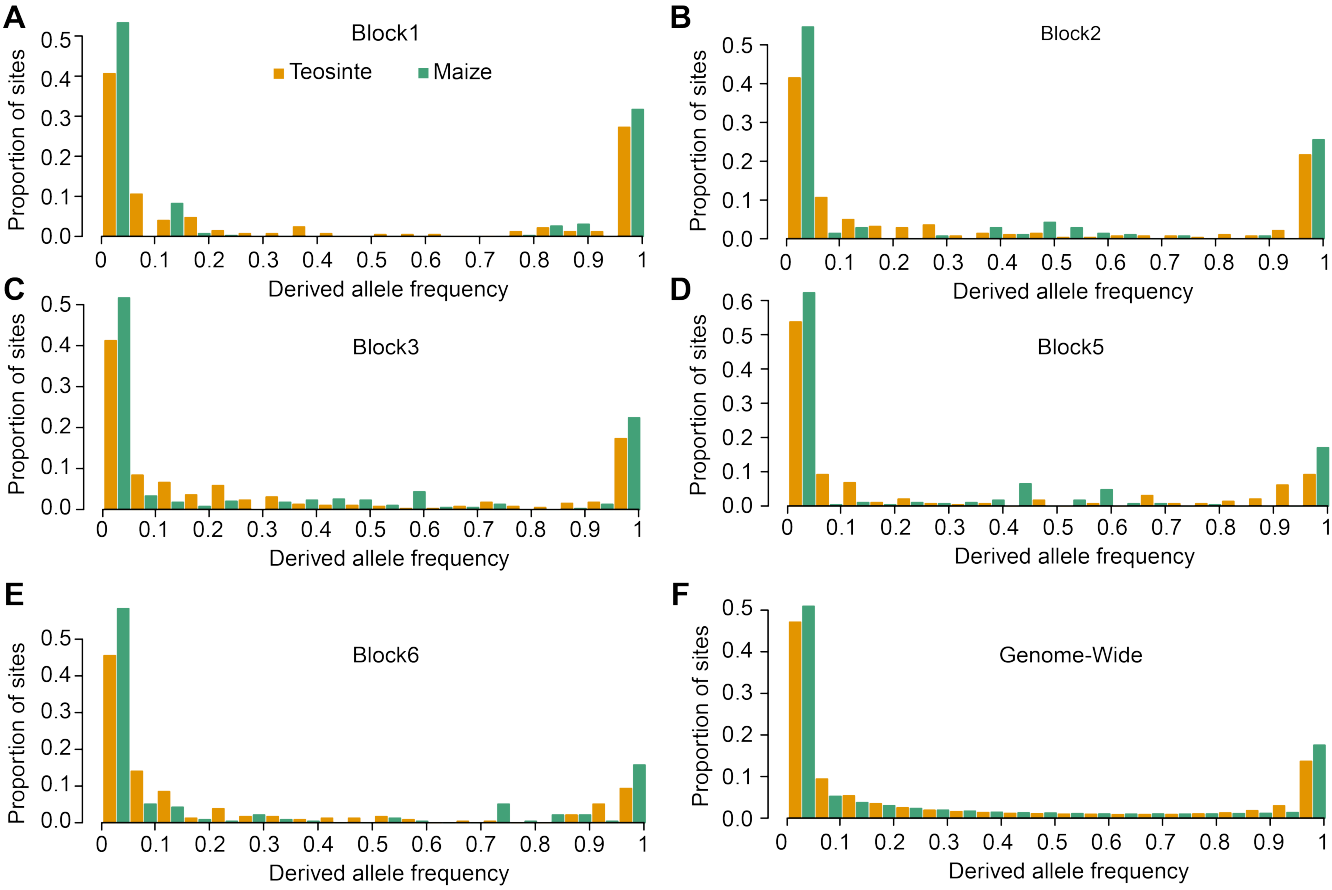

Figure S4. Unfolded site frequency spectrum in teosinte and maize. **(A-E)** The site frequency spectrum of variants within the LD Blocks. **(F)** Genome-wide site frequency spectrum. Sorghum bicolor alleles were used as the ancestral state.

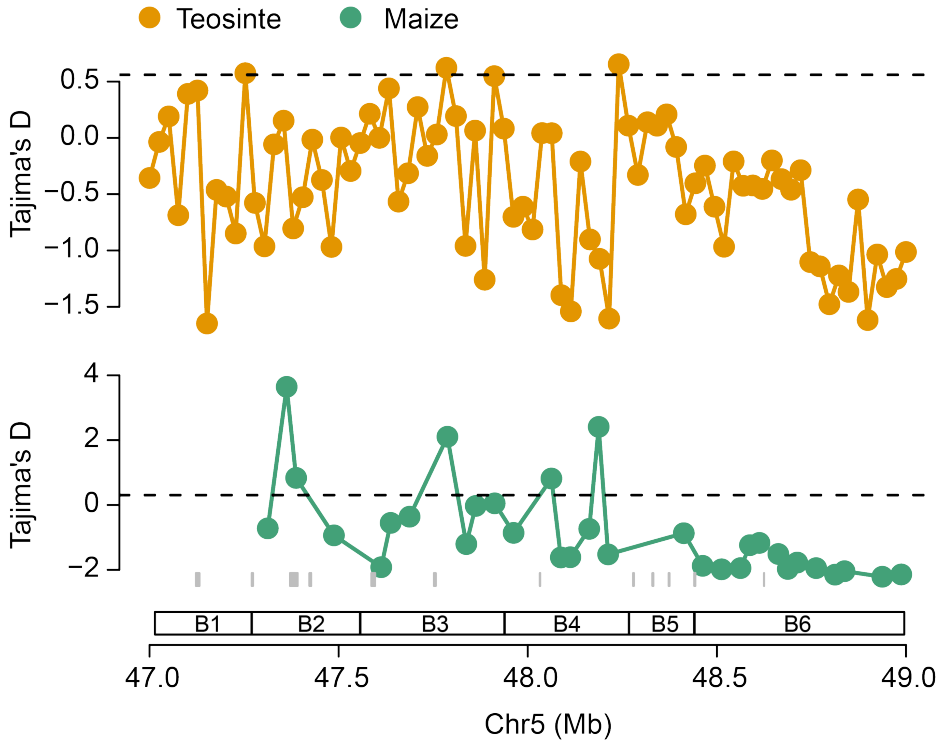

Figure S5. Tajima's D at the N responsive locus on chromosome 5. The horizontal dashed lines represent the 5% significance level across the genome. The underneath grey rectangles represent gene models. B1 to B6 indicate LD Block1 to Block6.

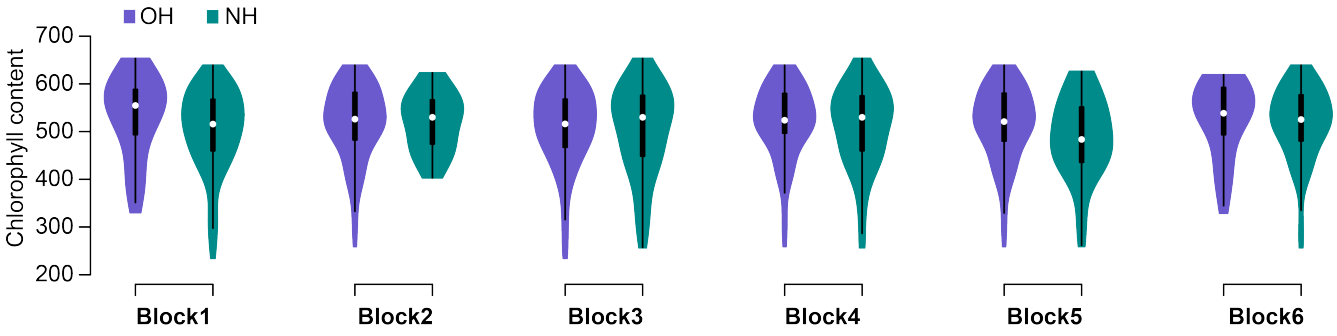

Figure S6. Phenotypic performance between Old-Era haplotype (OH) and New-Era haplotypes (NH) at Block1 to Block6 for leaf chlorophyll content.

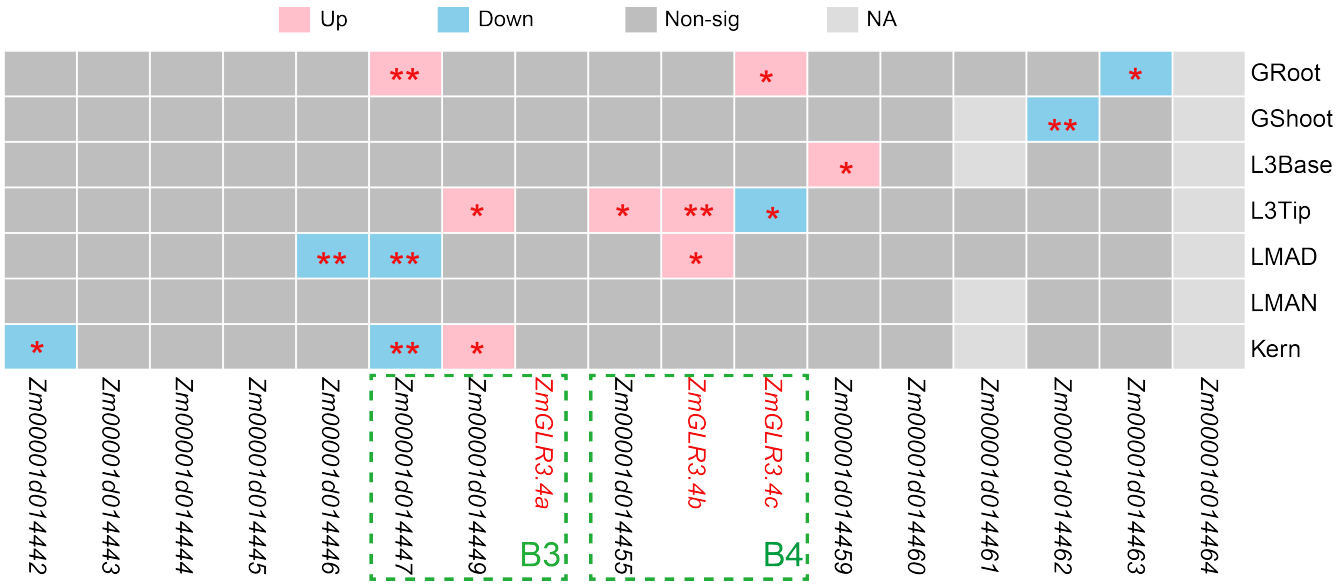

Figure S7. Gene expression differences between Old-Era and New-Era inbred lines for genes underneath the strongest selection signal at Chromosome 5. The pink and sky-blue colors indicate gene expression was upregulated and downregulated significantly in New-Era inbred lines, respectively. The dashed rectangles highlight genes located with LD block3 (B3) and block4 (B4). **P*-value < 0.05 and ***P*-value < 0.01 from two-sided Student's *t*-test.

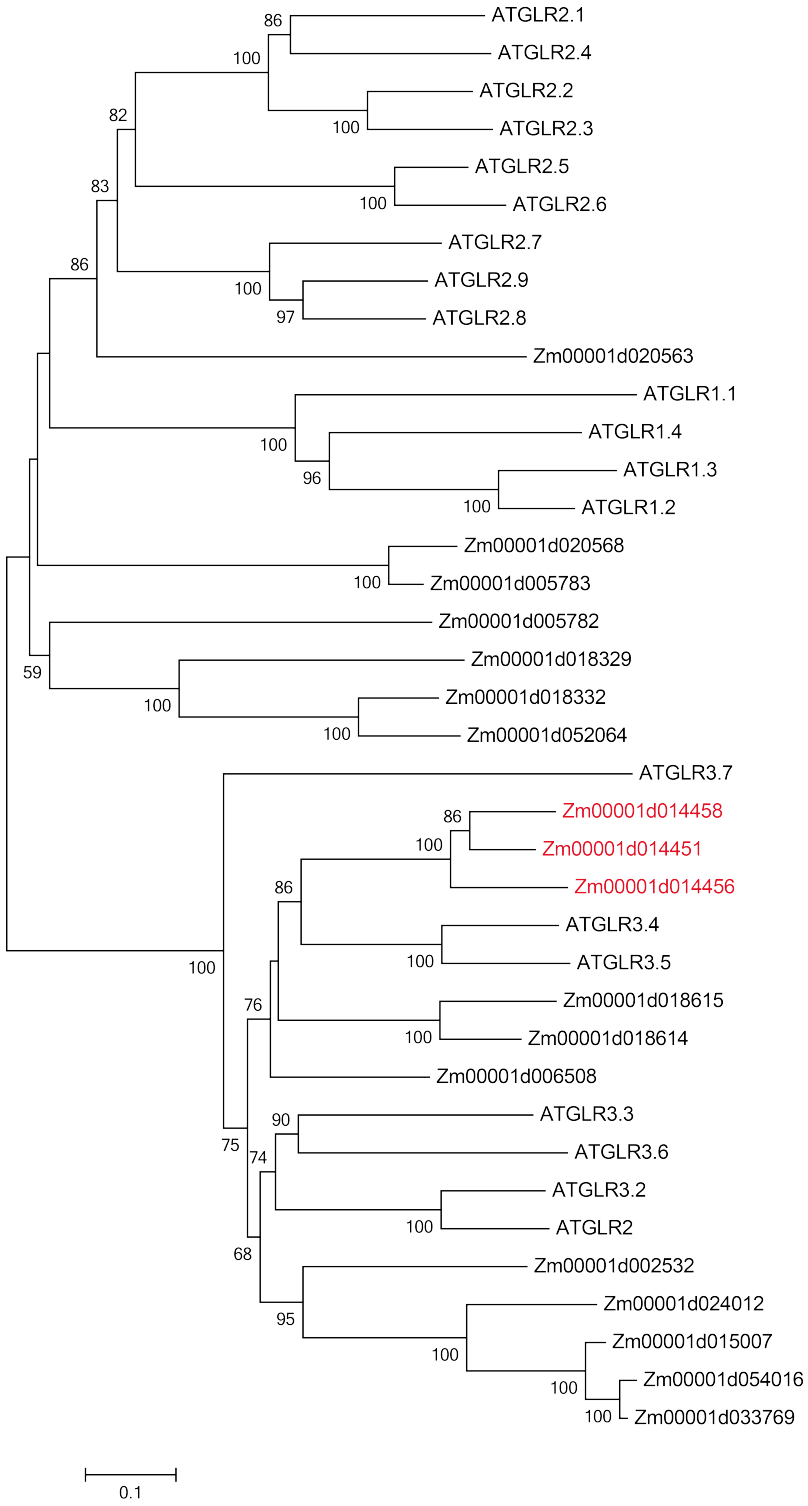

Figure S8. Phylogenetic tree generated from *Arabidopsis* and maize GLRs protein sequences. The gene names with red color indicate GLR genes under the strongest selection signal at chromosome 5.

Figure S9. Correlation analysis between gene expression 66 metabolites. All these genes are located at the most significant sweep at Chromosome 5. "With N" and "Without N" denote the metabolites containing or not containing N in their formulas. ** denotes Chi-squared test *P*-value < 0.01.

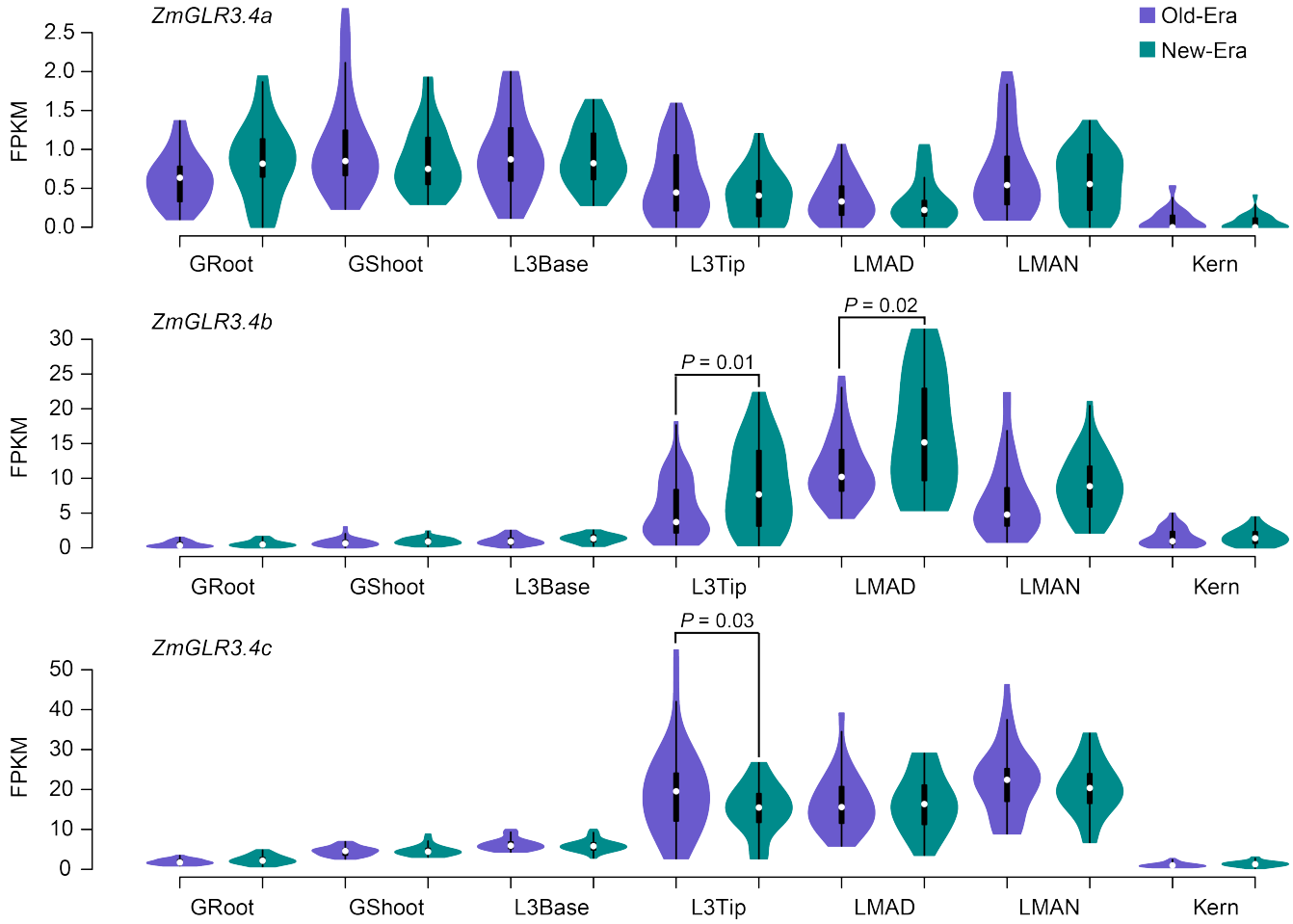

Figure S10. The distribution of gene expression levels of *ZmGLR3.4a*, *ZmGLR3.4b*, and *ZmGLR3.4c* in different tissues of Old- and New-Era inbred lines. *P* values were determined with a two-sided Student's *t*-test. The data was obtained from Kremling et al. 2018.

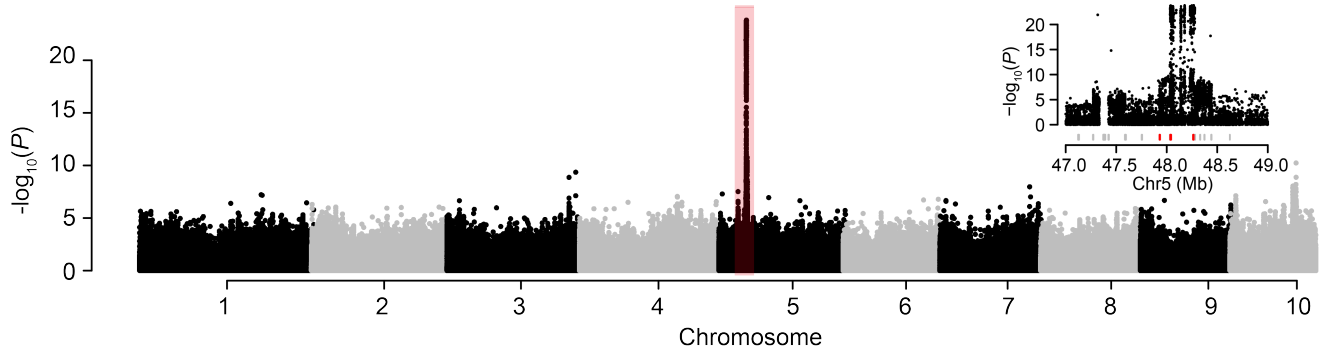

Figure S11. The Manhattan plot of the eQTL results using gene expression of *ZmGLR3.4c* collected from third leaf tip as the trait. The top right panel shows the zoom-in plot of the highlighted region in the Manhattan plot. The red ticks indicate the three GLR genes.

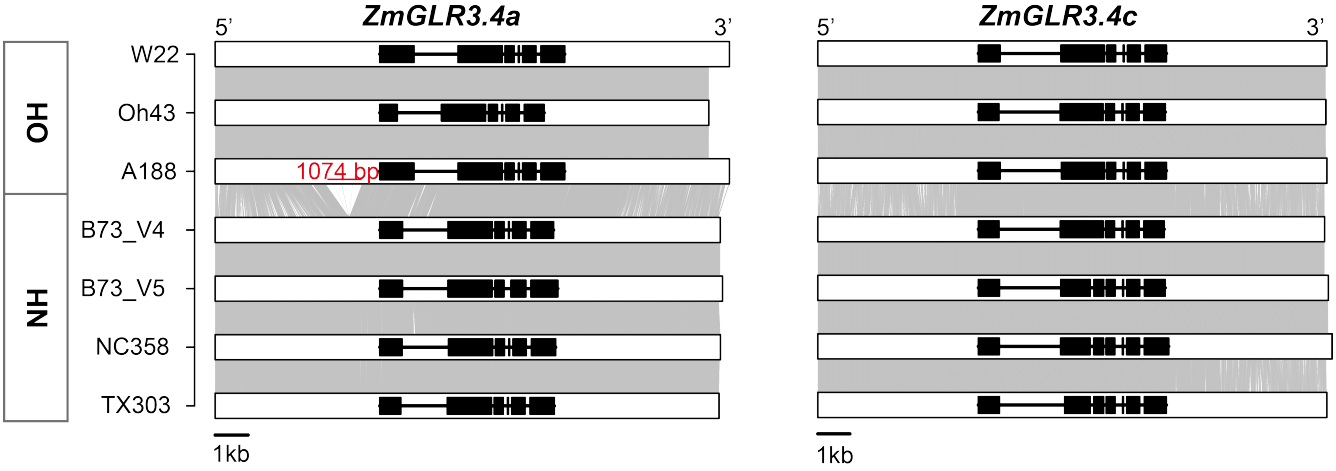

Figure S12. Comparison of *ZmGLR3.4a* and *ZmGLR3.4c* gene structures annotated from *de novo* assembled genomes. OH and NH indicate the inbred lines carrying the Old-Era and New-Era haplotype, respectively.

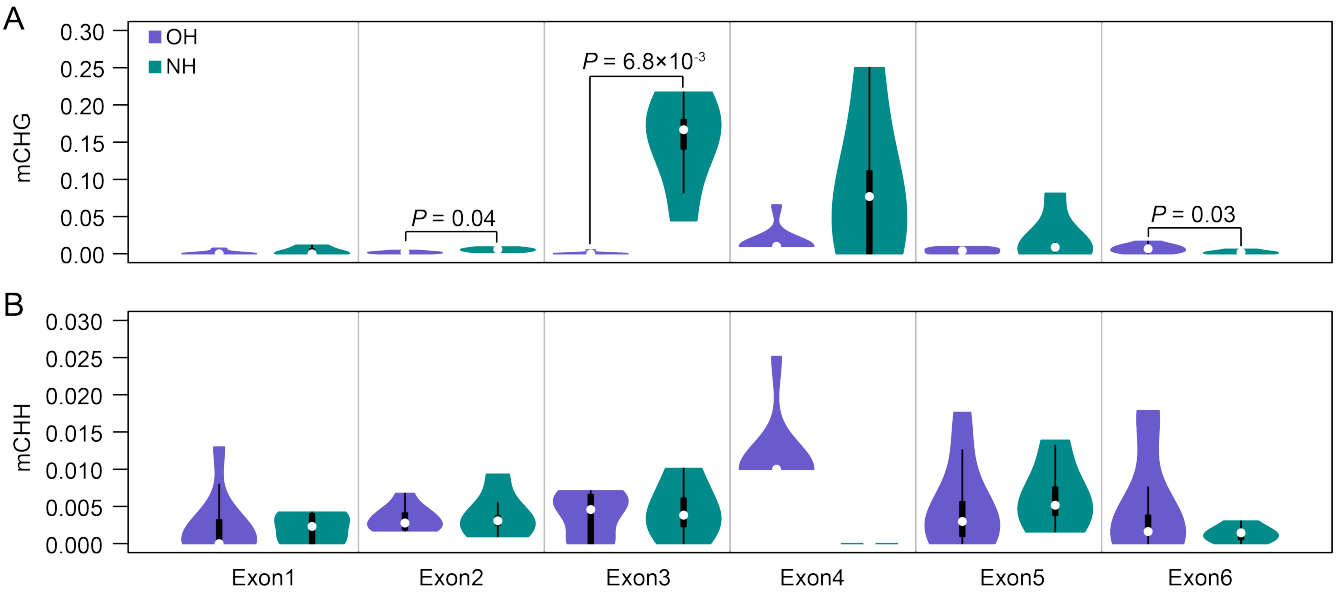

Figure S13. Levels of DNA CG (A) and CHG (B) methylation in exons of *ZmGLR3.4b*. OH and NH indicate the inbred lines carrying the Old-Era and New-Era haplotype, respectively. *P* values were determined with a two-sided Student's *t*-test.

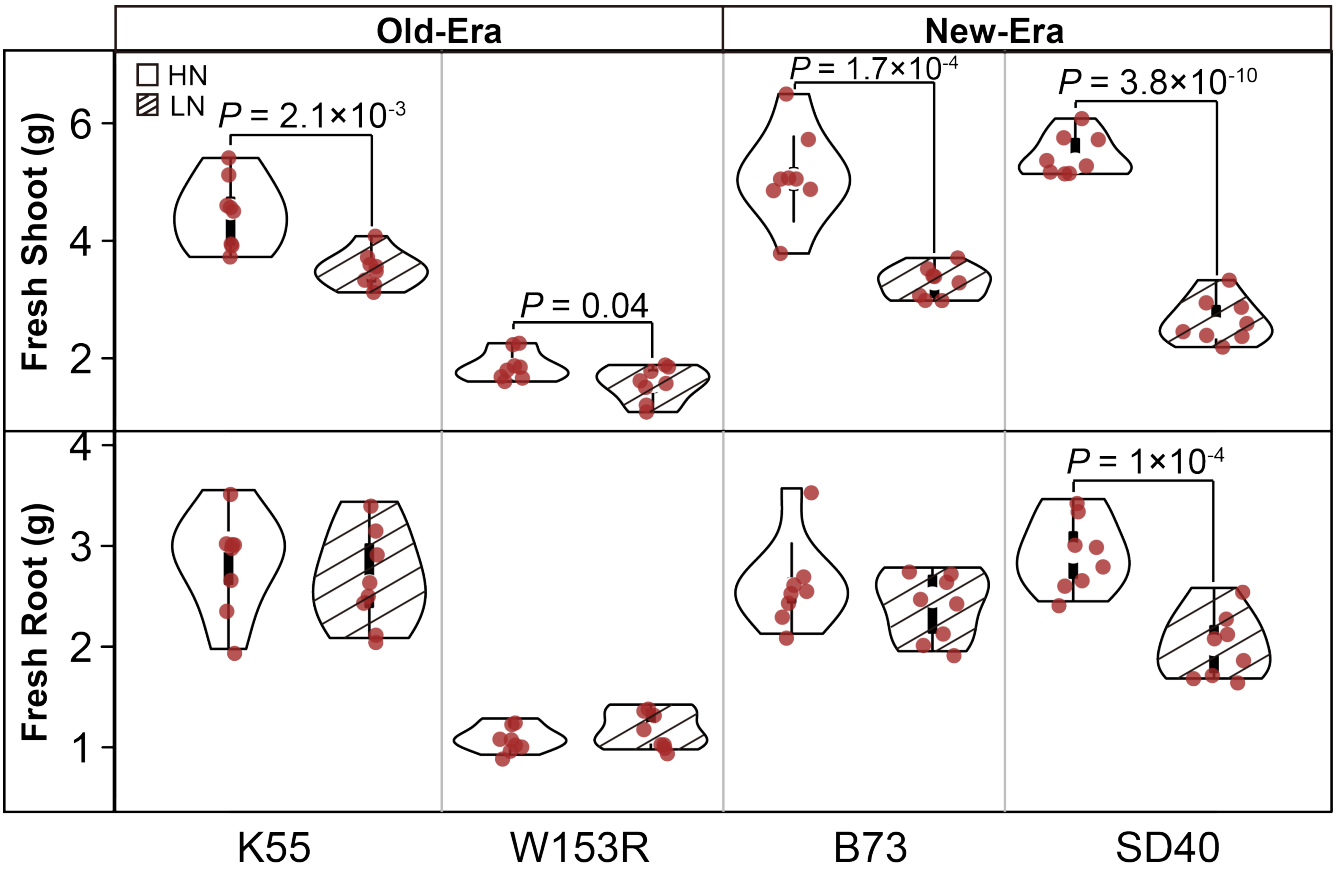

Figure S14. The fresh weight of shoot and root for Old-Era and New-Era inbred lines growing with high N and low N treatments. *P* values were determined with a two-sided Student's *t*-test.

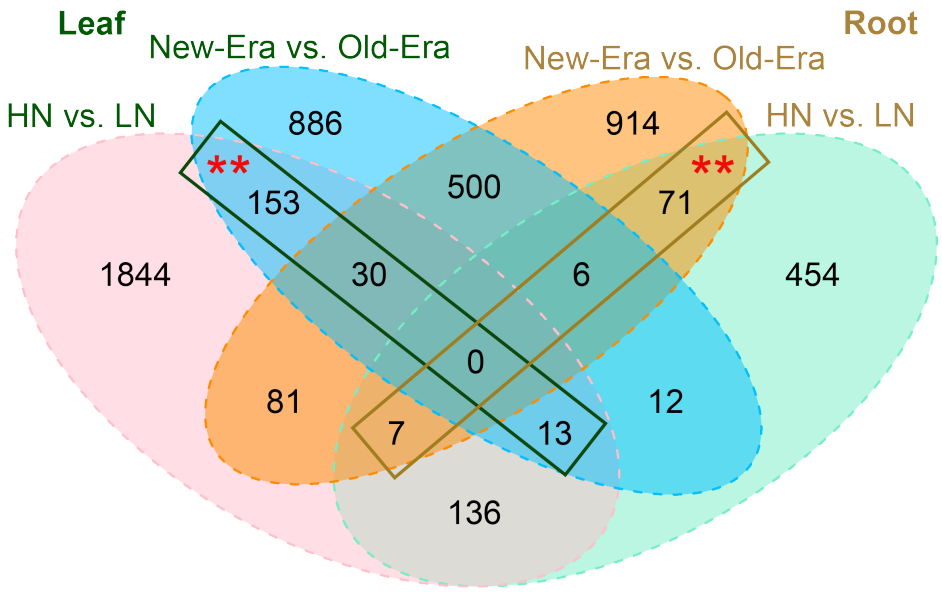

Figure S15. Differentially expressed (DE) genes between Old-Era and New-Era inbreds. Overlaps between the sets of DE genes of high N (HN) vs. low N (LN) and New-Era (NE) vs. Old-Era (OE) in leaf and root tissues. The red asterisks indicate categories with more genes than expected at a permutation based *P*-value cutoff of < 0.01.

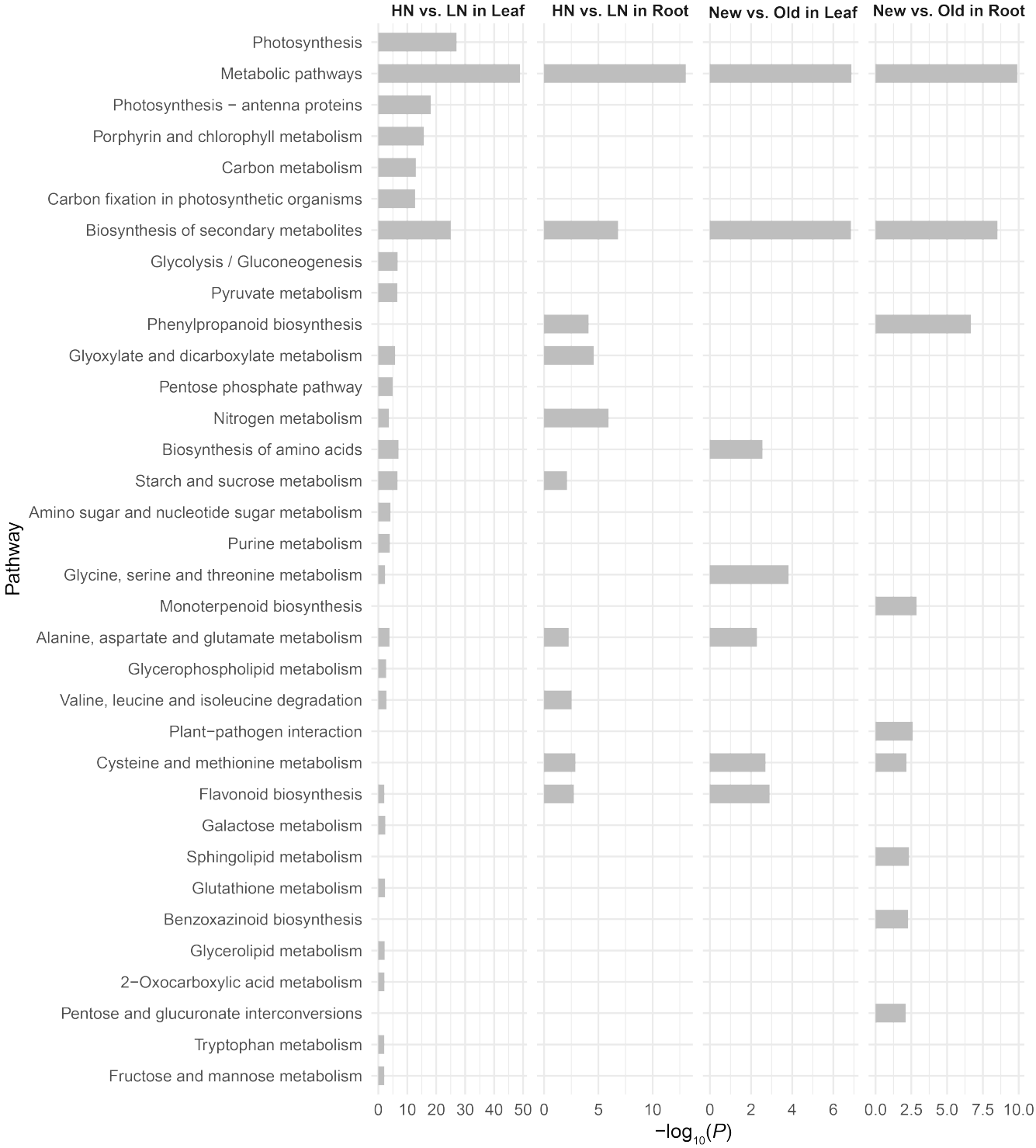

Figure S16. Enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways detected by N and Era responsive differentially expressed genes.