**THP9 enhances seed protein content and nitrogen-use efficiency in maize**

Yongcai Huang1,8, Haihai Wang1,8, Yidong Zhu1,2,8, Xing Huang1,2, Shuai Li2, Xingguo Wu3, Yao Zhao3, Zhigui Bao3, Li Qin1, Yongbo Jin1, Yahui Cui2, Guangjin Ma1,2, Qiao Xiao1,2, Qiong Wang3, Jiechen Wang1, Xuerong Yang4, Hongjun Liu4, Xiaoduo Lu6, Brian A. Larkins7, Wenqin Wang3,8 & Yongrui Wu3

Teosinte, the wild ancestor of maize (*Zea mays* subsp. *mays*), has three times the seed protein content of most modern inbreds and hybrids, but the mechanisms that are responsible for this trait are unknown1,2. Here we use trio binning to create a contiguous haplotype DNA sequence of a teosinte (*Zea mays* subsp. *parviglumis*) and, through map-based cloning, identify a major high-protein quantitative trait locus, *TEOSINTE HIGH PROTEIN 9* (*THP9*), on chromosome 9. *THP9* encodes an asparagine synthetase 4 enzyme that is highly expressed in teosinte, but not in the B73 inbred, in which a deletion in the tenth intron of *THP9*-B73 causes incorrect splicing of *THP9*-B73 transcripts. Transgenic expression of *THP9*-teosinte in B73 significantly increased the seed protein content. Introgression of *THP9*-teosinte into modern maize inbreds and hybrids greatly enhanced the accumulation of free amino acids, especially asparagine, throughout the plant, and increased seed protein content without affecting yield. *THP9*-teosinte seems to increase nitrogen-use efficiency, which is important for promoting a high yield under low-nitrogen conditions.

The seeds of plants contain stored metabolites—for example, carbohydrates, proteins, lipids, and nucleic acids—that are crucial for rapid cell division and growth during the transition from dormancy to photosynthetic autotrophy when environmental conditions are suitable for germination1. These metabolites also make seeds a valuable source of food for a variety of animals, as well as humans3. Over millennia, plant breeders have genetically altered plant species to create seeds with greater proportions of these metabolites, to improve their nutritional value and utility as food and feed4. Perhaps one of the most notable examples of this process was the conversion of the wild ancestor of maize, teosinte, to modern maize5.

Native Americans selected mutations that modified a variety of traits of teosinte, including the size and structure of its floral inflorescences and seeds, and its yield6,7. Because of its importance in their diet, the maize that was domesticated by Native Americans had a high protein content, enhanced flavour and utility for making food. However, as corn became a commodity and was used as feed for livestock, starch content (yield) became a primary concern, and less attention was paid to protein content and flavour2,8,9. In addition, the use of nitrogen fertilizer reduced (yield) became a primary concern, and less attention was paid to protein content and flavour2,8,9. In addition, the use of nitrogen fertilizer reduced.

Although nitrogen fertilizer markedly improves the yield of maize, its excessive use often leads to run-off, which causes the eutrophication of rivers and other bodies of water10. Consequently, future maize breeding must design plants with a higher nitrogen-use efficiency (NUE)11. In addition, seed protein content and quality will be more important in the future, as vegetable protein is likely to have a larger role in human diets12.

To identify genes that are responsible for the differences in protein content between maize and teosinte, we analysed the progeny of their cross and characterized the quantitative trait loci (QTLs) that affect this trait. We sequenced a teosinte haplotype genome (*Zea mays* subsp. *parviglumis*, Ames 21814), and localized gene loci that were associated with a high content of protein in the seeds. Using the teosinte haplotype and nearly isogenic line (NIL) populations created from it, we were able to clone a teosinte high-protein locus, *THP9*, which contains an asparagine synthetase 4 gene (*ASN4*) that has a central role in the accumulation of amino acids throughout the plant. The *THP9*-teosinte allele, *THP9*-T, is highly expressed in teosinte, whereas the corresponding allele in the maize B73 inbred, *THP9*-B73 (*THP9*-B) has a 48-bp deletion in the tenth intron that affects intron splicing. Several versions of the *THP9*-B transcripts contain a premature stop codon, which results in undetectable levels of the *ASN4* enzyme. Introggression of *THP9*-T into maize inbreds and hybrids increased the amino acid and protein content in the roots, stems and leaves, as well as the seeds. These plants exhibited a higher NUE under low-nitrogen conditions than did those with the *THP9*-B allele, and show promise for improving maize germplasm in general.

---

1National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology Chinese Academy of Sciences, Shanghai, China. 2University of the Chinese Academy of Sciences, Beijing, China. 3Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai, China. 4State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Tai’an, China. 5Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. 6Institute of Molecular Breeding for Maize, Qilu Normal University, Jinan, China. 7School of Plant Sciences, University of Arizona, Tucson, AZ, USA. These authors contributed equally: Yongcai Huang, Haihai Wang, Yidong Zhu. 8E-mail: wang2021@shnu.edu.cn; yrwu@cemps.ac.cn
High protein content of teosinte

During the domestication and artificial selection of maize, many visible (such as plant and glume architectures) and invisible (seed composition) traits were substantially modified (Fig. 1a). To investigate the variation in seed protein content between teosinte and modern maize inbreds, we collected 20 lines of *Zea mays* subsp. *parviglumis*, 10 lines of *Zea mays* subsp. *mexicana* and 518 modern maize inbreds. In maize seeds, most nitrogen occurs in storage proteins, so the total nitrogen content is essentially equal to the value measured. In the roots, stems and leaves, the total nitrogen reflects the sum of the nitrogen in free amino acids and proteins, but most of it is found in proteins. We determined the nitrogen content of the roots, stems, leaves and seeds of B73 by two procedures: acid hydrolysis and the Dumas method, the latter using a Dumas rapid nitrogen analyser (see Methods). There was no significant difference in the nitrogen content measured by the two methods (Extended Data Fig. 1a). Therefore, we could use the high-throughput rapid nitrogen analyser to assay seed protein content and nitrogen content in plant tissues. The seed protein content of all teosinte lines was around 30%, whereas that of maize inbreds (405 finally harvested for measurement in 2019) ranged from 6.5% to 16.5%, with an average of 11.5% (Fig. 1b). These differences suggest that the loci that control seed protein content are genetically variable in teosinte and modern inbreds.

We selected one line of *Zea mays* subsp. *parviglumis* (accession Ames 21814) as a representative high-protein genotype for analysis. We measured the total nitrogen content in the roots, stems and leaves of Ames 21814 and B73, and found that the nitrogen content was higher in all tissues of Ames 21814 than in B73 (Extended Data Fig. 1b). The composition of free amino acids differed to some extent (Extended Data Fig. 1c); in particular, the levels of asparagine were notably higher in all tissues of Ames 21814 than in B73 (Fig. 1c and Extended Data Fig. 1c). This is consistent with a previous observation in rice seeds, in which increased levels of asparagine were found to be associated with a high protein content.

Maize-seed proteins are classified according to their solubility as prolamins (called zeins), albumins, globulins and glutelins. Zeins are the main endosperm storage proteins, and account for more than 60% of the total. On the basis of their structure, zeins are divided into four families: α (19- and 22-kDa; designated α19 and α22), β (15-kDa), γ (50-, 27- and 16-kDa) and δ (18- and 10-kDa). The α19 family is further divided into z1A, z1B and z1D subgroups, and all members of the α22 family are z1C (refs. 3, 16). SDS–PAGE revealed that the accumulation of both zein and non-zein proteins was apparently higher in teosinte than in B73 seeds, but the most increased fractions seemed to be α19 and α22 (Extended Data Fig. 1d). This encouraged us to investigate the copy number of α-zein genes in teosinte.

Assembly of the Ames 21814 haplotype

To create a high-quality teosinte haplotype assembly for comparing α-zein loci in Ames 21814 and other inbreds, and for the mapping of high-protein QTLs, we sequenced (Supplementary Table 1) the DNA of a single F1 plant from the B73 × Ames 21814 cross using trio binning. The genome sequence of B73 is known, and we used a graph-based trio-binning strategy to decipher the haplotype information (Extended Data Fig. 2a,b). The initial Ames 21814 haplotype was assembled into 2,424 Mb by using 104-Gb HiFi long reads with a coverage of 47-fold. This resulted in 543 contigs with an N50 of 62.29 Mb (Supplementary Table 2)—a length greater than that attained in the previous sequencing of the inbreds that were the founders of the maize Nested Association Mapping population (contig N50 values of 6.26–52.36 Mb). Paternal (Ames 21814) and maternal (B73) haplotigs
Mendelian-like segregation in F2 seeds indicates that the high-protein B73 seeds (Extended Data Fig. 4f and Supplementary Fig. 1). The lack of α-zeins in particular were notably more abundant as compared with (Fig. 2a). When F2 seeds were analysed individually by SDS–PAGE, there was a major indicator of the total protein content (Extended Data Fig. 4b). Accumulation pattern of zein proteins, in which α-zeins (α19 and α22) are measured protein content with the rapid nitrogen analyser and found in incompatibility between teosinte and modern maize, we used Ames backcrossing parent (Extended Data Fig. 4a). Owing to unidirectional 21814, we created a series of continuous backcrossing populations using T0 to identify the QTLs associated with the high-protein trait in Ames Cloning of the high-protein locus toosinte generally, rather than specifically, increase protein content. The Ames 21814 haplotype were 22 and 12, respectively, compared with 25 and (z1A1, z1A2, z1B and z1D) and α22 (z1C1 and z1C2) genes in the Ames maize will need to be determined experimentally. Although there was good collinearity between B73 v5 and the Ames 21814 haplotype, 280,881 structural variations (greater than 50 bp) were identified (Fig. 1d and Supplementary Table 6), including 71 inversions larger than 100 kb. When 14 inversions (larger than 1 Mb) between the two haplotypes were inspected, 13 were supported by the Hi-C contact map (Extended Data Fig. 2f,g). Whether these structural variations correlate with phenotypic variation between teosinte and modern maize will need to be determined experimentally.

Using the highly contiguous Ames 21814 haplotype, we were able to annotate all of the α-zein genes. The total copy numbers of α19 (z1A1, z1A2, z1B and z1D) and α22 (z1C1 and z1C2) genes in the Ames 21814 haplotype were 22 and 12, respectively, compared with 25 and 15 in B73, and 25 and 19 in W22, respectively (Extended Data Fig. 3a,b). This suggests that the high-protein trait in teosinte is not conferred by a larger number of α-zein genes, and that the high-protein QTLs in teosinte generally, rather than specifically, increase protein content.

Cloning of the high-protein locus
To identify the QTLs associated with the high-protein trait in Ames 2184, we created a series of continuous backcrossing populations using Ames 2184 as the high-protein donor parent and B73 as the recurrent backcrossing parent (Extended Data Fig. 4a). Owing to unidirectional incompatibility between teosinte and modern maize, we used Ames 2184 to pollinate B73 for the F1 progeny (Extended Data Fig. 2a). We measured protein content with the rapid nitrogen analyser and found that the F1 seeds (B73 × Ames 2184; 11.6 ± 0.8% (s.d.)) had a protein level similar to that of B73 (10.8 ± 1.0%), whereas the Ames 2184 seeds had a protein content of 28.6 ± 1.0% (Fig. 2a)—consistent with the accumulation pattern of zein proteins, in which α-zeins (α19 and α22) are a major indicator of the total protein content (Extended Data Fig. 4b). However, the total nitrogen content in the roots and leaves of F1 plants was higher than that in B73 (Extended Data Fig. 4c–e), and the F1 seeds had nearly double the protein content (19.9 ± 2.2%) of F1 and B73 seeds (Fig. 2a). When F1 seeds were analysed individually by SDS–PAGE, there was no apparent variation in the accumulation of zein proteins, and α-zeins in particular were notably more abundant as compared with B73 seeds (Extended Data Fig. 4f and Supplementary Fig. 1). The lack of Mendelian-like segregation in F1 seeds indicates that the high-protein trait is determined by the maternal rather than the filial genotype.

Because the F1 (B73 × Ames 2184) plants exhibited many rudimentary teosinte phenotypes in vegetative and reproductive growth (Extended Data Fig. 2a), we used B73 as the ear parent to make the first backcrossing generation (BC1; B73 × F1). Afterwards, we used B73 as the pollen source for backcrossing (Extended Data Fig. 4a). In the BC1 (B73 × B73) population, we observed a segregation of zein protein content among different ears in a quantitative rather than a qualitative pattern (Extended Data Fig. 4g and Supplementary Fig. 2), which indicates that the high-protein trait is regulated by multiple genetic loci. Like the F1 seeds (Extended Data Fig. 4f), when individual seeds from a high-protein BC1 ear were analysed, each seed uniformly accumulated more α-zeins than did B73 (Extended Data Fig. 4h). Subsequent backcrossing generated eight ears with the highest protein content (about 15%); these were retained and seeds from each ear were planted for analysis. Similarly, quantitative measurement of the BC1 and BC2 generations with the rapid nitrogen analyser confirmed that the protein content varied among different ears (ranging from 10 to 15%), but was uniform in individual seeds of the same ear (Extended Data Fig. 4i–l).

To identify the genetic loci that influence protein content, we planted the BC1 seeds, and saved a piece of leaf from each plant for DNA extraction. Zein and non-zein proteins from 500 BC1 ears were analysed by SDS–PAGE (Extended Data Fig. 3a and Supplementary Fig. 3). On the basis of their phenotypes, we pooled leaf DNA samples of the low- and high-protein individuals (n = 75 for each) for bulked segregant analysis (BSA) DNA sequencing. The results highlighted several QTLs on chromosomes 1, 3, 4, 5, 7 and 9, with a significant peak in the region between 130 Mb and 160 Mb (based on Teo v1) on chromosome 9 (Fig. 2b and Extended Data Fig. 5b) that contained 315 introgressed teosinte genes. Accordingly, this locus was designated TEOSINTE HIGH PROTEIN 9 (THP9).

Using the same approach, we created the BC2 (n = 1,314) and BC3 (n = 1,344) generations. BSA of BC2 and BC3 confirmed the existence of THP9. However, continuous backcrossing did not appear to result in more frequent recombination at this locus, as the two latter BSAs still contained 271 and 190 teosinte genes in this region (based on a 0.025 threshold) (Extended Data Fig. 5c–f). We performed high-coverage (higher than 20×) resequencing of this region in five high-protein and five low-protein individuals from the BC3 population, and found that the introgressed teosinte locus in the five high-protein lines recombined in the form of large DNA fragments between 22.7 and 144.4 Mb (based on B73 v4); the smallest common region (135.5–134 Mb) should be the candidate interval (Extended Data Fig. 5g). Nearly isogenic lines, NILTeo and NILLB73, with high and low levels of protein, respectively, were created on the basis of this interval.

To fine-map THP9, we created a BC4 generation (n = 2,000) that narrowed THP9 to a 150-kb region containing three genes (Zm00001d047732, Zm00001d047736 and Zm00001d047737), on the basis of the B73 reference genome (B73 v4) (Fig. 2c). Zm00001d047732, which encodes a protein phosphatase, lacks notable structural variation between Ames 2184 and B73, except for several single-nucleotide polymorphisms (SNPs) in the gene coding sequence (Extended Data Fig. 6a). The fold changes of Zm00001d047732 expression in the roots and leaves of NILTeo and NILLB73 were all slight (Extended Data Fig. 6b). Zm00001d047737, which encodes an uncharacterized protein, was not expressed in the roots and leaves of NILTeo and NILLB73 (according to our RNA-seq data), nor was it expressed in other tissues (according to public RNA-seq data). Zm00001d047736, which corresponds to Teo09G002926 in Ames 2184, encodes an asparagine synthetase 4 gene (ZmASN4). An analysis of the Ames 2184 and B73 sequences revealed that Teo09G002926 is an intact ASN4 gene (hereafter referred to as THP9–Teosinte, THP9–T), whereas Zm00001d047736 has a 48-bp deletion in the tenth intron of ASN4 (hereafter referred to as THP9–B, THP9–B; Fig. 2c).

On the basis of published data, we determined that the intron deletion in ASN4 creates altered splicing of THP9–B transcripts, which results in the formation of three different isoforms of the mRNA. The ZmASN4–T001 isoform is similar to the ASN4–Teo transcript, whereas
ZmASN4-T002 and ZmASN4-T003 are defective, as both contain a premature stop codon (Extended Data Fig. 6c). RNA-seq revealed that THP9-T transcripts (ASN4-Teo) accumulate abundantly in the roots and leaves of Ames 21814, whereas the ZmASN4-T003 isoform was barely detectable in these tissues of B73 (Extended Data Fig. 6d). Further RNA-seq analysis of NILTeo and NILB73 confirmed that THP9-T is highly expressed, whereas THP9-B is barely expressed, in root and leaf tissues (Fig. 2d and Extended Data Fig. 6e).

We amplified the ASN promoter sequences (around 1.9 kb) from Ames 21814 and B73 and tested their activities by dual-luciferase assay. The results showed that there was no significant difference in activity between the two ASN4 promoters \((P > 0.05)\), which suggests that the differential expression of THP9-T and THP9-B is unlikely to be caused by promoter variation (Extended Data Fig. 6f). Consistent with the transcript levels, ASN4 protein accumulates abundantly in NILTeo, but is absent in NILB73 (Fig. 2e and Supplementary Fig. 4). The results suggest that the 48-bp deletion in the tenth intron of the ASN4 gene in B73 considerably affects the RNA splicing and stability of ASN4 transcripts, making them and the ASN4 protein difficult to detect. THP9-B can therefore be considered a null allele.

We developed a molecular marker for THP9-B and used it to genotype 200 individuals in the BC7 population (Extended Data Fig. 7a). When we measured the free asparagine content in the roots of NILTeo and NILB73. The immunoblot analysis experiment was repeated independently at least twice with similar results. f, Protein content of NILTeo and NILB73 seeds that were harvested in Shanghai, Sanya and Harbin. Data are mean ± s.d. \((n = 20 \text{ biologically independent samples})\). g, Free asparagine content in NILTeo and NILB73 roots and leaves. Data are mean ± s.d. \((n = 20 \text{ biologically independent samples})\). In a–d, different letters indicate significant differences \((P < 0.01, \text{one-way ANOVA and further Tukey's test; see Source Data})\). In d,f,g, a two-tailed Student’s test was used to determine \(P\) values (see Source Data). CEN, centromere.
content in the roots, stems and leaves (Extended Data Fig. 7h), as well as an increased total free amino acid content in the roots and leaves, as compared to NILB73 (Extended Data Fig. 7i). In addition, the levels of free asparagine in NILTeo roots and leaves were significantly higher than those in NILB73 (Fig. 2g), indicating that the increased accumulation of asparagine through THP9-T facilitates increased synthesis of proteins in the roots, stems, leaves and seeds.

**Validation and natural variation of THP9**

To investigate whether THP9-T can influence the low-protein phenotype of B73, we expressed this allele in transgenic plants using the ubiquitin promoter. The THP9-overexpressing plants had greatly enhanced levels of ASN4 transcript and ASN4 protein in the leaves and roots, as compared with the non-transgenic B73 control (Fig. 3a–c and Supplementary Fig. 5). Two representative THP9-overexpressing lines (THP9-OE1 and THP9-OE2) grown in Sanya were analysed. The seed protein contents of THP9-OE1 and THP9-OE2 were 15.2 ± 1% and 15.8 ± 1.1%—increases of 25.7% and 30.9%, respectively, over the B73 control (12.1 ± 0.9%) (Fig. 3d). These results are consistent with the hypothesis that the mutation in THP9 is responsible for the low-protein phenotype of B73.

As well as measuring 405 maize inbreds grown in Sanya in 2019 (see Fig. 1b and its corresponding Source Data), we measured the seed protein content of 437 inbreds that were grown in Sanya in 2020 (see Fig. 3e and its corresponding Source Data). The protein content of the 2020 crop ranged from 7.8% to 16.9%, with an average of 12.3%. Inbreds for which data were available on the seed protein content from both 2019 and 2020 were used for a genome-wide association study (GWAS) analysis, which identified a region with physical coordinates significant peak flanking 300 kb included in the THP9 locus on chromosome 9.

**THP9-T increases the NUE**

Because THP9-T increases the free amino acid content in plants, which in turn promotes plant development and the accumulation of protein in seeds, we investigated whether THP9-T could increase the NUE. To this end, we set up an experimental site on our farm in Shanghai to test the

**Fig. 3 | Genetic confirmation and natural variation of THP9.** a, b, qRT–PCR analysis of THP9 expression in the leaves (a) and roots (b) of THP9-OE1 and THP9-OE2. Expression levels were normalized to that of ZmActin. Data are mean ± s.d. (n = 9 biologically independent samples). WT, wild type. c, Immunoblot analysis showing the protein level of ASN4 in the leaves of THP9-OE1 and THP9-OE2. ACTIN was used as an internal control. The immunoblot analysis experiment was repeated independently at least twice with similar results. d, Protein content of THP9-OE1 and THP9-OE2 seeds. Data are mean ± s.d. (n = 17 biologically independent samples). e, GWAS analysis of seed protein content in 405 and 437 inbred lines grown in 2019 and 2020, identifying a significant peak flanking 300 kb included in the THP9 locus on chromosome 9. f, Schematic diagram illustrating three major THP9 alleles in the population of 420 maize inbreds. HAP1, HAP2, and HAP3 identify inbred types containing the three different THP9 alleles. g, Protein content of HAP1, HAP2 and HAP3 seeds. Data are mean ± s.d. (n = 108, 195 and 117 biologically independent samples, respectively). h, Expression analysis of ASN4 in HAP1–HAP3. The transcript levels of ASN4 in the roots of 14 HAP1, 26 HAP2 and 19 HAP3 inbreds were analysed by qRT–PCR. Data are mean ± s.d. (n = 14, 26 and 19 biologically independent samples, respectively). In a, b, d, h, a two-tailed Student’s t-test was used to determine P values (see Source Data).
effects of applying different levels of nitrogen fertilizer on plant growth. Several aboveground concrete containers with different concentrations of soil nitrogen were constructed. A plastic film covered the containers to prevent rainwater from affecting the soil nitrogen concentration (Extended Data Fig. 5a–d). More than 50 NILTeo and 50 NILB73 plants were grown side by side in containers either with a normal application of nitrogen (40 g per plant; normal-nitrogen condition) or without the application of nitrogen (low-nitrogen condition). We measured the levels of nitrogen in the soil and found that the pool of plants that were treated with normal levels of nitrogen contained 76.7% more nitrogen than did the low-nitrogen pool of plants (Extended Data Fig. 5e). NILTeo plants appeared to grow better than NILB73 plants in normal- and low-nitrogen conditions (Extended Data Fig. 5f–h). qRT–PCR showed that the expression of \( \text{THP9-T} \), but not \( \text{THP9-B} \), was strongly induced when nitrogen was applied, suggesting that \( \text{THP9-T} \) is sensitive to the level of soil nitrogen (Extended Data Fig. 5i). Without the application of nitrogen, both NILTeo and NILB73 plants were slender and had a smaller amount of root mass than was observed with the normal-nitrogen treatment, but NILTeo plants that were treated with less nitrogen were comparable in size to NILB73 plants that were treated with normal levels of nitrogen (Extended Data Fig. 5d,f). The root fresh weight and the aboveground biomass of NILTeo and NILB73 plants were greatly reduced by the low-nitrogen condition, but there was no significant difference between NILTeo in low-nitrogen and NILB73 in normal-nitrogen conditions (Extended Data Fig. 5j,k). The total nitrogen content (mostly free amino acids and proteins) in the roots, stems, and leaves, and the protein content in the seeds of NILTeo and NILB73, were affected by low levels of nitrogen, but these values (protein and nitrogen content) in NILTeo under the low-condition condition were comparable to those of NILB73 in the normal-nitrogen condition (Extended Data Fig. 5l–o). We also examined the NUE in \( \text{THP9-OE2} \) transgenic plants. Similar to NILTeo, \( \text{THP9-OE2} \) showed an improved NUE in low-nitrogen conditions (Extended Data Fig. 5a–j).

Subsequently, in 2021 we set up a larger field trial in Sanya, in which we applied different amounts of nitrogen: 100% (32 g per plant), 50% (16 g per plant), 25% (8 g per plant), and 0% in each trial; 300 seeds of NILTeo and NILB73 were planted together (Fig. 4a). NILTeo seemed to have a growth advantage over NILB73 in terms of plant height (Fig. 4b) and aboveground biomass (Fig. 4c) under the different nitrogen conditions. The total nitrogen content in the roots, stems, and leaves of NILTeo was significantly higher than that in NILB73 in all trials (Fig. 4d–f). After reducing the application of nitrogen in the three treatments from 100% to 0%, the protein content in NILTeo seeds decreased from 14.2% to 13.5%, 12.0%, and 10.7%, whereas in NILB73 seeds it decreased from 11.4% to 11.2%, 10.7%, and 8.9% (Fig. 4g). The results indicate that seed protein content is sensitive to the level of soil nitrogen, and in each treatment, NILTeo seeds always maintained a higher level of protein than NILB73 seeds. The protein content in NILTeo seeds that were harvested at 25% nitrogen reached 12.0%, which was higher than that of NILB73 seeds (11.4%) that were treated with normal levels of nitrogen. These results are consistent with the hypothesis that \( \text{THP9-T} \) confers a higher NUE than \( \text{THP9-B} \) in NILB73 in normal- and low-nitrogen conditions.

**Use of \( \text{THP9-T} \) for maize breeding**

B73 and Mo17 (HAP2 type) are inbreds that are frequently used to study the vigour of hybrids. To examine whether \( \text{THP9-T} \) could increase the protein content of hybrids and influence other agronomical traits, we generated two sets of \( F_1 \) seeds—NILTeo × Mo17 and NILB73 × Mo17—and planted them at the Harbin site. The protein content of \( F_1 \) seeds from the NILTeo × Mo17 cross (9.2 ± 0.6%) was 7.8% higher than that of seeds from the NILB73 × Mo17 cross (8.6 ± 0.4%), whereas the 100-kernel weight was nearly identical for the two hybrids (35.4 g versus 35.6 g; Fig. 5a–c).

We also introgressed \( \text{THP9-T} \) into Zheng 58 (HAP3 type) and Chang 7-2 (HAP2 type)—two elite inbred lines that make the Zhengdan 958 hybrid, which is the most widely grown variety of maize in China. The \( \text{THP9-T} \)-modified (designated Zhengdan 958-T) and the unmodified Zhengdan 958 (designated Zhengdan 958-B) hybrids were grown for comparison in Sanya in 2021 (Fig. 5d,e). Zhengdan 958-T plants showed a significantly increased aboveground weight and height, as compared with Zhengdan 958-B (Fig. 5f,g), and Zhengdan 958-T seeds had a protein content of 11.1 ± 1.1%, a 12.7% increase compared to Zhengdan 958-B seeds (9.9 ± 0.6%) (Fig. 5h). The total nitrogen content in the roots (Fig. 5i), stems (Fig. 5j) and leaves (Fig. 5k) of Zhengdan 958-T also increased significantly. However, the 100-kernel weights of Zhengdan 958-T and Zhengdan 958-B were not significantly different (Fig. 5i).

We measured the seed protein content of NILTeo × Mo17 and Zhengdan 958-T hybrids that were harvested in Shanghai in 2022, which also showed a significant increase compared with the corresponding control (Extended Data Fig. 10). The results suggest that \( \text{THP9-T} \) has the potential to improve the protein content of maize seeds and plants through plant breeding.

**Discussion**

**Variable seed protein content in maize**

Genetic variability in terms of seed protein content is well documented in maize. More than 100 years ago, the University of Illinois initiated a breeding program to examine the consequences of artificial selection on seed composition. High-protein and low-protein phenotypes were among the traits selected. Midway through the decades-long process, plant breeders reversed the selection, and converted the high-protein germplasm to a low-protein phenotype, and vice versa with the low-protein selection. The outcome was four genetic strains of maize: Illinois high protein (HPI), reverse high protein (RHP), Illinois low protein (ILP) and reverse low protein (RLP), which had protein contents of about 30%, 7%, 4% and 15%, respectively.

The results of this experiment suggest the existence of both positive and negative genetic factors influencing protein content in natural maize populations, which are likely to be controlled by multiple genetic loci.

Because modern maize was domesticated from teosinte, we reasoned that characterizing the genes responsible for the high-protein trait in teosinte might reveal a more diverse set of QTLs than those found in recent inbred maize populations. The results might also help us to understand the reasons for the decrease in seed protein content during the domestication of maize. In addition, teosinte contains high levels of free amino acids—especially asparagine—in the roots, stems, and leaves (Fig. 1c and Extended Data Fig. 1c); this suggests that it has high nitrogen assimilation, which could contribute to the seed protein content and NUE. The challenge was to create a complete teosinte genome sequence, and this led us to characterize the nucleotide sequence of a high-quality teosinte haplotype, *Zea mays* subsp. *parviglumis*, Ames 21814 (Fig. 1d and Extended Data Fig. 2).

**THP9 encodes an asparagine synthetase**

We assembled a high-quality teosinte haplotype genomic sequence that helped us to identify the genes responsible for QTLs associated with its high seed protein phenotype. Significant QTLs were found on chromosomes 1, 3, 4, 5, 7 and 9 (Fig. 2b), but we focused on \( \text{THP9} \), because it had the greatest effect and was the highest peak revealed by BSA DNA sequencing. \( \text{THP9} \) encodes an asparagine synthetase 4 (ASN4) gene. The intact \( \text{THP9-T} \)-allele is highly expressed in teosinte roots and leaves, but the \( \text{THP9-B} \)-allele is not functional in these tissues in B73, owing to mis-splicing of its transcripts (Fig. 2d); this probably leads to differences in nitrogen assimilation.

ASN is an important enzyme in the metabolism of nitrogen and it has a key role in the nitrogen response network. Previous research in *Arabidopsis* rice, *wheat* and barley showed that changes in the expression of ASN genes alter plant growth and nitrogen content, and that the level of ASN expression is affected by the environment. In *Arabidopsis thaliana*, studies of AtASN1, AtASN2 and AtGS
confirmed the effect of asparagine on the nitrogen content in seeds, floral organs, leaves and plants\textsuperscript{28–30}. In rice (\textit{Oryza sativa}), studies of \textit{OsASN1} confirmed the importance of asparagine for plant nitrogen and for grain protein content\textsuperscript{13}. The increase in ASN activity leads to enhanced assimilation of nitrogen, which results in more asparagine being transported to the seed for protein synthesis\textsuperscript{26,36}.

In B73, there are four \textit{ASN} genes: \textit{ZmASN1–ZmASN4} (\textit{Zm00001d045675, Zm00001d044608, Zm00001d028750} and \textit{Zm00001d047736}) (ref. \textsuperscript{39}). \textit{ZmASN1} appears to be expressed in all maize tissues, including the root, stem, leaf, endosperm and embryo. \textit{ZmASN2} is mainly expressed in the endosperm and embryo, according to public RNA-seq data\textsuperscript{23}. \textit{ZmASN3} (on chromosome 1) and \textit{ZmASN4} (on chromosome 9) could have resulted from an ancestral gene duplication\textsuperscript{38}. These two genes are functional in Ames 21814 and could have an additive effect in asparagine synthesis.

\textit{ZmASN3} was annotated as an intact gene in the B73 genome, but it is expressed at low levels in leaves, cobs and silks\textsuperscript{23}. The four maize \textit{ASN} genes could have a redundant function for asparagine synthesis, but the absence of \textit{ZmASN4} and probably any of the other three \textit{ZmASN} genes leads to asparagine insufficiency in the plant. When \textit{THP9-T} was introgressed into B73, the asparagine content in roots and the nitrogen content of the entire plant were significantly enhanced (Fig. 2g and Extended Data Fig. 7h), providing evidence of the importance of \textit{THP9-T} for NUE and seed protein content. We anticipate that overexpression of other \textit{ZmASN} genes might also increase the seed protein content and NUE in maize.

Our data suggest that \textit{THP9-B} is a null allele, as the \textit{ZmASN4} protein is missing in B73 (Fig. 2e). Public data showed that \textit{THP9-B} gives rise to three mRNA isoforms. Of these, the functional one was undetectable in our assays, and the other two are defective because the 48-bp deletion in the tenth intron leads to mis-splicing, which creates a premature stop codon in \textit{THP9-B} transcripts. On the basis of prediction using Pfam searches (https://www.ebi.ac.uk/interpro/), the \textit{THP9-T} allele produces a protein of 588 amino acids, whereas the barely detectable isoform 3 of \textit{THP9-B} can only be translated into a truncated protein with 480 amino acids.
Amino acids are essential substrates for protein synthesis, and their levels in the plant are influenced by soil nitrogen availability and the NUE of the plant. During amino acid synthesis, asparagine has a primary role in nitrogen recycling, and it acts as a nitrogen donor for multiple aminotransferases. Owing to its high nitrogen-to-carbon ratio and inert nature, free asparagine is a major carrier for nitrogen storage and long-distance transport in the plant. ASN, which is responsible for transferring amide groups from glutamine to aspartate and forming asparagine, determines the assimilation, remobilization and allocation of nitrogen in the plant.

With limited soil nitrogen, the amino acid supply for protein synthesis can be increased by a greater NUE. NUE is determined by multiple processes, namely, nitrogen uptake, transport, assimilation and remobilization, of which nitrogen assimilation has been actively studied. Looking to the future, there is economic and environmental pressure to maintain high-yielding maize while reducing the level of nitrogen applied to the soil. Therefore, it is crucial to identify genetic factors that increase the NUE. Several studies have shown an association between QTLs and enzymes related to nitrogen metabolism; however, the genes responsible for QTLs that control nitrogen assimilation have not been cloned.

Potential use of \textit{THP9-T} for breeding

\textit{THP9-T} had a stable phenotypic effect at different geographical locations and when treated with different levels of nitrogen, which is essential for its practical application. However, we found that the protein content of NILTeo seeds was only half that of teosinte. We can offer two possible explanations: (1) the high protein content of teosinte is regulated by multiple QTLs, and the remaining uncharacterized QTLs on other chromosomes make substantial contributions to the high-protein phenotype of teosinte; (2) the seed protein content is affected by the distribution of nitrogen from source to sink. Teosinte seeds are small and their yield per plant is low. Therefore, the concentration of amino acids allocated to a single seed could be greater for protein synthesis. The seeds of modern inbred lines are larger than those of teosinte, and there are more seeds per ear and per plant; in consequence, a lower concentration of amino acids is needed to produce the high-protein phenotype of teosinte.

The protein content of hybrids with the introgressed \textit{THP9-T} allele was lower than that in NILTeo (Figs. 2f and 5c,h), supporting this hypothesis. As most of the increased protein in teosinte and NILTeo is α-zeins, which is essentially devoid of the essential amino acid lysine, the increased protein content has limited nutritional value for monogastric animals. However, \textit{THP9-T} could be introgressed into quality protein maize (QPM)—which contains less zein and more non-zein proteins, owing to the \textit{o2} mutation—to create high-protein, high-lysine hybrids. The improved QPM will be valuable for future food security, in particular in countries where maize is consumed as a major source of protein.

The HAP1–HAP3 genotypes imply that some— if not all—of the ancestral high-protein QTLs were retained in domesticated populations of maize. Nevertheless, seed protein content declined during modern...
maize breeding. Because THP9-T is superior to THP9-B and THP9-T for protein synthesis and has no apparent negative effect on yield, why THP9-T was not retained in elite maize germplasm is a key question. One possible explanation is that THP9 was not under selection pressure, owing to the ample application of nitrogen fertilizer. This could have become a vicious cycle, with the low NUE of THP9-B requiring more nitrogen fertilizer to improve the yield and protein content.

NUE has key environmental and economic implications for global food security, and research to understand NUE is necessary if we are to maintain high yields and high protein quality with a low input of nitrogen. Several genes and QTLs that affect the NUE in rice, including NRT1.1B, OsTCP19, GRF4 and NGR5, have been cloned. Superior alleles of these genes or QTLs offer the potential to achieve high, stable rice yields with low levels of nitrogen application. Root nitrogen sensing has been found to be affected by several external factors, and strategies are being developed to increase the nitrogen acquisition efficiency under varying nitrogen conditions for crop production.

Our research shows the possible value of hybrids that contain the THP9-T allele, although larger field trials in multiple geographical locations will be needed to fully establish the potential of THP9-T for improving the seed protein content and NUE in maize breeding. These hybrids perform well in a nitrogen-poor environment, and maintain a normal yield when reduced levels of nitrogen are applied. Additional research on NUE, based on the high-quality teosinte genome sequence, could lead to other QTLs that improve modern hybrids. The marked structural variation between the genomes of Zea mays subsp. parviglumis, Ames 21814 and B73 will also be beneficial for investigating the genes that may be responsible for the phenotypic modifications that occurred during the domestication of teosinte.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05441-2.

1. Karn, A., Gillman, J. D. & Flint-Garcia, S. A. Genetic analysis of teosinte alleles for kernel composition traits in maize. G3 7, 1517–1546 (2017).
2. Flint-Garcia, S. A., Bodnar, A. L. & Scott, M. P. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor. Appl. Genet. 119, 1129–1142 (2009).
3. Han, C., Zhen, S., Zhu, G., Bian, Y. & Yan, Y. Comparative metabolome analysis of wheat endosperm reveals the dynamic changes of metabolites during seed germination. Plant Physiol. Biochem. 155, 320–327 (2017).
4. De Lumen, B. O. Molecular approaches to improving the nutritional and functional properties of plant seeds: food sources: developments and comments. J. Agric. Food Chem. 38, 1779–1788 (1990).
5. Palacios-Rojas, N. et al. Mining maize diversity and improving its nutritional aspects within agro-food systems. Comp. Rev. Food Sci. Food Saf. 19, 1809–1834 (2020).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022
seeds from ears with a high protein content were selected for planting. All seeds contained a uniformly high or low content of α-zeins. The BC2 used B73 pollen for backcrossing. The zein and non-zein protein accumulation patterns of 108 BC ears were characterized, and the protein accumulation pattern showed quantitative segregation. In a single ear, all seeds contained a uniformly high or low content of α-zeins. The BC2 seeds from ears with a high protein content were selected for planting. We created continuous backcrossing populations, yielding BC2, (n = 500), BC3, (n = 500), BC4, (n = 1,000), BC5, (n = 1,344), BC6, (n = 1,200), BC7, (n = 1,344) and BC8, (n = 2,000). In each generation, we measured the protein content ear by ear.

To obtain homozygous NILTeo and NILB73, 70 BC7 independent ears with a high protein content (about 15%) were planted as 20 groups. Thirty plants of each group were self-pollinated, yielding 600 BC1F2 ears, which formed 30 × 20 subgroups. The protein content of all subgroup ears was measured and 50 BC1F2 subgroup ears with a high protein content were planted. Twenty plants of each subgroup were self-pollinated, yielding 1,000 BC2F3 ears that were measured for protein content. If individual ears in a subgroup had a uniformly high protein content—namely, no segregation—they should be homozygous for the high-protein locus and were identified as NILTeo. By contrast, if all the ears in a subgroup uniformly had a protein content similar to B73 (about 10%), the ears were designated NILB73. NILTeo and NILB73 were propagated by self-pollination. Five BC3F1 NILTeo and five NILB73 individuals were selected for 20× resequencing. The linkage analysis was performed by genotyping 200 BC2F3 plants and measuring the protein content of the corresponding BC2F3 ears.

The overexpression vector of THP9-T fused with Flag (ubiPro:THP9-T) was constructed and then transformed into B73 via Agrobacterium-mediated transformation by Wimi Biotechnology (http://www.wimibio.com/). The primer sequences used in this study are shown in Supplementary Table 7.

Resolving the Ames 21814 haplotype with trio binning

Because of advances in long-read DNA-sequencing technologies, many high-quality maize inbred haplotypes have been assembled successfully. Inbreeding simplified the assembly consensus process, as most of the regions are homozygous. However, teosinte lines make it hard to untangle the two haplotypes owing to the high heterozygosity caused by open pollination.

High-quality genomic DNA was extracted from fresh leaves of the F1, crossed with B73 and Ames 21814, followed by library construction according to the standard protocol of PacBio (Pacific BioSciences). DNA sequencing on the PacBio Sequel II Hifi platform, which produces high-fidelity reads with CCS (v.4.2.0; https://github.com/PacificBioSciences/ccs), was done by Shanghai OE Biotech. In addition, we generated 50× Illumina PE 150 reads for the parental B73 and Ames 21814 genomic DNA, respectively. We used yak (https://github.com/hh3/yak) to generate the 31-mer database with the parental short reads. We applied the hifiasm (v.0.16.1) trio mode, a de novo assembler that could faithfully preserve the contiguity of all haplotypes, to assemble the haplotypes of Ames 21814. Contaminants, such as organelle DNA or rDNA fragments, were removed by BLASTN. We mapped Hi-C reads to the assembly with the Juicer pipeline (v.1.5.7) (ref. 34) and scaffolded by 3D-DNA (version: 180419) with “r = 0 -m haploid”. False duplications and phase error were manually curated on the basis of yac trinoval within Juicebox (v.1.11.08) (ref. 35). Finally, we used yak to evaluate the base accuracy of the genome assembly.

Transposon elements were annotated by EDTA (v.1.9.5) (ref. 36) using the pan-genome TE database (https://github.com/HuffordLab/NAM-genomes/tree/master/te-annotation). Protein-coding genes were predicted using MAKER2 (v.2.3.0) with the homologue evidence of RNA-seq and protein databases. RNA was collected from six tissues (root, leaf, stem, seed, cob and tassel) of Ames 21814 and aligned against the genome with HISAT2 (v.2.10.2) (ref. 36). Protein sequences were downloaded from UniProt (Viridiplantae) (https://www.uniprot.org) for six plant species (Arabidopsis thaliana, Oryza sativa, Setaria italica, Sorghum bicolor, Triticum aestivum and Zea mays), which were integrated with CD-HIT (v.4.6) (ref. 36) using the parameter ‘-c 0.99’. Ab initio gene prediction was performed using SNAP (v.2006-07-28), AUGUSTUS (v.3.3.3) and GeneMark (v.4.3.8). SNAP was trained using the first-round MAKER2 annotation, whereas AUGUSTUS and GeneMark were trained by RNA-seq and protein databases with BRAKER2 (ref. 37). The gene models with annotation edit distance values of less than 0.5 were retained.

Measurement of protein content with the rapid nitrogen analyser

To measure the total nitrogen content in seeds and other tissues (root, stem and leaf), the samples were first dried to a constant weight at 65 °C and then powdered using a grinder (Tissuelyser-48, Shanghai Jingxin Industrial Development; 60 Hz, 60 s). A total of 50–70 mg of powder was wrapped in tin foil as the test sample. The total nitrogen content was determined using the Dumas rapid nitrogen analyser (‘rapid N exceed’) from Elementar. Before each round of measurement, it was necessary to weigh about four standard asparagus samples for internal controls. After debugging the machine, the weight of each sample was entered in the weight column of the rapid N exceed software (v.1.1.25), and the following options were selected as program settings: O2 dosing time, 60 s; O2 dosing flow, 120 ml min−1; O2 cut-off threshold, 15%; Autozero delay, 30 s; and peak anticip. 90 s. At the same time, the packaged samples were placed into the sample tank according to the corresponding serial number. Fifty-five samples were measured in one round. The data were exported in Excel format for analysis.

Measurement of free amino acids

The roots, stems and leaves of different plant genetic materials were analysed to determine the content of free amino acids at the flowering stage. Plant materials were dried at 65 °C to a constant weight and ground. Thirty milligrams of powder was treated in 1 ml distilled water at 4 °C for 8 h and then homogenized. The powder was hydrolysed with 6 M hydrochloric acid at 110 °C for 24 h; after filtration, 100 µl liquid was added with 100 µl 5 M NaOH and 800 µl distilled water. After centrifugation at 5,500g for 5 min, 50 µl of supernatant was mixed with amino mixed standards (MSLAB50A) and 50 µl 4% sulfosalicylic acid solution, and the mixture was centrifuged at 17,370g at 4 °C for 4 min. The supernatant was mixed with 50 µl borate buffer (0.1 M, pH 8.8) and then derivatized with 20 µl 6-aminoquinoline-N-hydroxysuccinimide carbbamate at 55 °C for 15 min. After cooling and centrifuging at 4 °C, 50 µl supernatant was analysed by ultra-performance liquid chromatography (UPLC: Ultimate 3000)–tandem mass spectrometry (MS/MS; API 3200 Q TRAP). Chromatographic separations were performed on an MS Lab HP-C18 column (150 × 4.6 mm, 5 µm). The mobile phase consisted of water (A) and acetonitrile (B). The solvent was delivered to the column at a flow rate of 0.8 ml min−1. The conditions for MS/MS detection were as follows: positive-ion mode; ion spray voltage, 5,500 V; nebulizer gas pressure, 55 psi; curtain gas pressure, 20 psi; collision gas...
BWA-MEM (v.0.7.12) (ref. 62) was used to align the clean reads to the reference genome. After the alignment, results were formatted and sorted by SAMTools (v.1.9), and the duplication reads were removed by Picard (v.4.1.0.0).

(3) Variant information detection. On the basis of the alignment of the sample sequencing data with the reference genome, SNP and indel detection were performed using the HaplotypeCaller module of the GATK (v.4.1.0.0) software63.

(4) G-value analysis. The G-value analysis was implemented in QTLseqr64. After manually filtering the SNPs, G’, a smoothed value of the standard G statistic, was calculated in an 8-Mb window size. The red line indicates the threshold of the G’ value, corresponding to a p value of 0.01. (5) Analysis of gene introgression. The coverage depth of high bulk (high protein) and low bulk (low protein) on each window was calculated with a 25-kb window, and then normalized (divided by the respective average sequencing depth). The normalized low bulk depth was subtracted from the normalized high bulk depth to obtain the delta depth. The figure was drawn with ggplot2 (R v.3.5.1). Coordinate positions are based on the teosinte Ames 21814 haplotype. Peaks with a delta depth > 0.025 indicate the introgression of teosinte genes. According to the G-value analysis results, there was a peak in the region between 120 Mb and 130 Mb (based on B73_V4) on chromosome 9, and the corresponding region aligned to the teosinte Ames 21814 haplotype between 130 Mb and 160 Mb (based on Teo_V1) on chromosome 9. On the basis of the teosinte Ames 21814 haplotype annotation gff file, the number of introgressed teosinte genes in the 130-Mb and 160-Mb region of BC4, BC6, and BC8, is 315, 271, and 190, respectively. By gene homologous alignment, teo09G002926 in the Ames 21814 genome corresponds to Zm00001d047736 (ZmASN4) in the B73 genome.

(6) Resequencing mapping analysis. The introgressed genes on chromosome 9 from Ames 21814 were analysed. The physical coordinates of the extracted regions were based on the B73 reference genome. The resequencing data were mapped to the B73-V4 reference genome sequence, we designed molecular marker primers (Supplementary Table 7). On the basis of the genotypes of molecular markers and corresponding seed protein contents, THP9 was narrowed down to an interval between two markers, 140.2 and 140.3, on chromosome 9, based on the B73 reference genome (B73_V4). All figures were plotted with ggplot2 (v.3.3.5).

Analysis of the structural variation of THP9
We analysed gene structural variation using GSDS 2.0 (Gene Structure Display Server 2.0), on the basis of the Ames 21814 haplotype and B73 genome sequences. ASN transcripts in the root and leaf of B73 × Ames 21814 were analysed using kallisto (v.0.44.0) for determining the allele-specific expression.

Dual-luciferase reporter assay
We performed a dual-luciferase reporter assay to detect the promoter activities. The ASN4 promoter sequences (around 1.9 kb) were amplified from B73 and Ames 21814, and cloned upstream of the LUC gene in the reporter vector pGreenII 0800. The constructs were transformed into the B73 leaf protoplasts. After incubating for 16 h, the transformed protoplasts were used for total protein extraction and then analysed on a luminometer (Promega 20/20) using a commercial LUC analysis kit according to the manufacturer’s instructions (Promega, E1960). Three biological replicates were performed for each experiment. The primers are listed in Supplementary Table 7.
Genetic confirmation of THP9 in B73
The full-length coding sequence of THP9-T was amplified from Ames 21814 root cDNA and fused with a Flag tag at the N terminus. This DNA fragment was inserted downstream of the ubiquitin promoter. The construct was transformed into B73 by Agrobacterium-mediated transformation. This was done by WinMi Biotechnology. The primer sequences used in this study are shown in Supplementary Table 7.

RNA extraction, reverse transcription and qRT–PCR
The leaf and root tissues of B73, Ames 21814 and the NILs were frozen in liquid nitrogen and stored in −80 °C. These materials were ground into fine powder, and a total of 100 mg was extracted with TRIzol reagent (Invitrogen, 15,596,018). RNA was purified with an RNaseasy Mini Kit (Qiagen, 79,254) and used for reverse transcription with a SuperScript III First Strand Synthesis Kit (Invitrogen, 18,080,051). cDNA was diluted to 80 ng μl⁻¹ for qRT–PCR with SYBR Green (TAKARA) on a CFX Connect Real-Time System (Bio Rad). The maize Actin gene was used as an internal control and the relative gene-expression level was calculated by the comparative CT method (ΔΔCT method). The expression level in the control was set to 1. All data were generated from three replicate biological samples, and means and s.d. were calculated. The primer sequences are shown in Supplementary Table 7.

Zein copy number analysis
To accurately locate zein gene clusters, BLASTN was used to align the assembled Ames 21814 haplotype with the known α-zein clusters and flanking genes of the B73 and W22 inbreds for copy number analysis. To further clarify the copy number, stringent parameters of BLASTN were chosen as follows: -evalue 1e-10.

Antibody preparation and immunoblot analysis
A partial ASN4 protein fragment from the 460th to the 588th amino acid was used to make antibodies (Abclonal, Wuhan). To analyse the protein accumulation of ASN4 in the roots of NIL Teo and B73, total proteins were extracted using the non-zein buffer. Twenty micrograms of total protein was separated by 10% SDS–PAGE and then transferred electrophoretically to a PVDF membrane. The protein was detected with ASN4 antisera at a dilution of 1:1,000 at 4 °C overnight, followed by secondary anti-rabbit-HRP at a concentration of 1:5,000 (Abmart, M21002L). The control protein, ACTIN, was detected with a mouse monoclonal actin antibody (Abmart, M20009L) at a dilution of 1:1,000, and a secondary antibody, anti-mouse IgG-HRP (Abmart, M21001L) at a dilution of 1:5,000. The membranes were treated with chemiluminescence substrate reagent (Invitrogen, WP20005), and then immunoreactive bands were detected using the Tanon-5200 system. To examine ASN4 in THP9-OE1 and THP9-OE2 plants, total protein was extracted from the leaf. Immunoblotting used anti-Flag (Sigma, A8592) as the primary antibody at a dilution of 1:1,000, and anti-mouse IgG-HRP (Abmart, M21001L) as the secondary antibody at a dilution of 1:5,000. Imaging was done with a Tanon-5200 system (Tanon).

NUE testing
In 2021, we planted NILB73 and NIL Teo at the Songjiang experimental field in Shanghai, using soil in cement tanks with or without normal nitrogen application. For normal nitrogen, 20 g of nitrogen fertilizer was applied to each plant at the seedling stage (V4) and 20 g at the jointing stage (V12). The nitrogen content of the fertilizer is 17%. Gene expression, plant height, aboveground biomass, root biomass, root, stem and leaf nitrogen content and seed protein content were investigated. In addition, to perform a test of the NUE of transgenic plants, we planted wild-type and THP9-OE2 plants at the Songjiang experimental field in Shanghai in 2022, under the same nitrogen fertilizer application conditions as in 2021. Larger field trials were performed in Sanya in 2021. Four different nitrogen applications were tested: normal application (16 g per plant applied at each seedling stage (V4) and jointing stage (V12)); 50% (8 g per plant applied at each seedling stage (V4) and jointing stage (V12)); 25% (4 g per plant applied at each seedling stage (V4) and jointing stage (V12)); and 0% (no nitrogen applied). The nitrogen content of the fertilizer is 17%. Each treatment contained 300 plants grown at 0.6 m × 0.25 m for each plant. Plant height, aboveground biomass, total nitrogen content of the root, stem and leaf, seed protein content and amino acid content were measured.

Introgression of THP9-T in hybrids
NIL Teo and NILB73 were crossed with Mo17 to create hybrids that were grown in Harbin. Using molecular marker selection, THP9-T was introgressed into Zheng S8 and Chang 7-2 by backcrossing for four generations. The marker was developed on the basis of an indel polymorphism between THP9-T and THP9-B. After backcrossing, the resulting plant materials were self-pollinated for two generations, creating Zheng S8-T and Chang 7-2-T. The cross of Zheng S8-T and Chang 7-2-T produced a modified hybrid, Zhengdan 958-T, that carried the THP9-T allele. Zhengdan 958-T and Zhengdan 958-B, with the THP9-B allele, were grown in Sanya in 2021. Plant height, aboveground biomass, total nitrogen content of the root, stem, and leaf and seed protein content were measured.

Statistical analysis
GraphPad Prism v.8.0.2 and Microsoft Excel 2016 were used for the statistical analyses (one-way ANOVA, Tukey’s test and two-sided Student’s t-test).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All sequencing data for creation of the Ames 21814 haplotype have been deposited at the National Genomics Data Center (NGDC; https://ngdc.ncbi.ac.cn/) under the BioProject number PRJCA011706, in which the PacBio HiFi sequencing data are under the accession number SAMC873435, the Hi-C data are under the accession number SAMC873392, the PacBio isoform sequencing data are under the accession numbers SAMC873393 and the Illumina WGS data are under the accession numbers SAMC873436 and SAMC873437. The final assembled genome sequence data reported in this paper have been deposited under accession number GWBBKHM000000000 that is publicly accessible at https://ngdc.ncbi.ac.cn/gwh. The RNA-sequencing data of Ames 21814, B73 × Ames 21814 and B73 (roots and leaves at flowering stage) are under the accession numbers SAMS874334–SAMS874357. Source data are provided with this paper.

51. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98 (2016).
52. Durand, N. C. et al. Juicexbox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101 (2016).
53. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20, 275 (2019).
54. Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. 3, 1–48 (2021).
55. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment unifies structural and O2 heterodimerizing proteins. Cell Syst. 3, 1162–1172 (2016).
56. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659 (2006).
57. Zhang, Z., Yang, J. & Wu, Y. Transcriptional regulation of zein gene expression in maize through the additive and synergistic action of opaque2, prolamine-box binding factor, and O2 heterodimerizing proteins. Plant Cell 27, 1162–1172 (2015).
58. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48 (2015).
59. Bukowski, R. et al. Construction of the third-generation *Zea mays* haplotype map. 
Gigascience **7**, gix134 (2018).
60. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**, 821-824 (2012).
61. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. 
Bioinformatics **34**, i884-i890 (2018).
62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. 
Bioinformatics **25**, 1754-1760 (2009).
63. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**, 1297-1303 (2010).
64. Mansfeld, B. N. & Grumet, R. QTLseqr: an R package for bulk segregant analysis with next-generation sequencing. Plant Genome **11**, 180006 (2018).
65. Dong, J. et al. Analysis of tandem gene copies in maize chromosomal regions reconstructed from long sequence reads. Proc. Natl Acad. Sci. USA **113**, 7949-7956 (2016).

Acknowledgements We thank Y. Xiao, H. Zhan, B. Lin, Y. Gu, T. Song and D. An for help with BSA sequencing and related analysis; K. Liu for help with the high-protein QTL analysis and gene mapping; C. Li from Anhui Agricultural University for help with the bioinformatics analysis; J. Shi and W. Li for help with the analysis of teosinte domestication and sequence variation; and Z. Wang, L. Zhang, X. Zeng and Z. Zhou for help with material planting in Harbin. This research was supported by the Chinese Academy of Sciences (XDB27010201 to Y.W.), the National Natural Science Foundation of China (31830063 and 31925030 to Y.W., 32072008 to W.W.), the China Postdoctoral Science Foundation (2020M681412 to Y.H.) and the Shanghai ‘Super Postdoctoral’ Incentive Program (2020456 to Y.H.).

Author contributions Y.W., W.W., Y.H. and H.W. designed the research, analysed the data and supervised the project. Y.W., Y.H., W.W., Y. Zhu, X.H. and H.L. created genetic populations and materials. Y.H., H.W., Y. Zhu, X.H., Y. Zhao, L.Q., Y.J., Y.C., Q.X., Q.W., J.W., H.L. and X.L. performed experiments. W.W., S.L., X.W., Z.B. and Y.H. performed teosinte Ames 21814 haplotype sequencing, assembly and annotation. Q.M. and X.Y. performed the GWAS analysis. Y.W., X.L., Y.H., H.W. and Y. Zhu performed the field NUE tests in Sanya. Y.W., W.W., Y.H., H.W. and B.A.L. explained the data, drafted and edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05441-2.
Correspondence and requests for materials should be addressed to Wenqin Wang or Yongrui Wu.
Peer review information Nature thanks Shujun Ou, Jianming Yu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Content of total nitrogen and free amino acids in roots, stems and leaves of Ames 21814 and B73 and SDS–PAGE of zein and non-zein proteins in B73 and 10 teosinte lines. a. Nitrogen content in roots, stems, leaves and seeds of B73 determined by acid hydrolysis and the Dumas method. Data are presented as mean ± s.d. (n = 6 biologically independent samples for each method).

b. Total nitrogen content in roots, stems, and leaves of Ames 21814 and B73. Data are presented as mean ± s.d. (n = 10 biologically independent samples).

c. Content of individual free amino acids in roots (the top panel), stems (the middle panel), and leaves (the bottom panel) of Ames 21814 and B73. Data are presented as mean ± s.d. (n = 6 biologically independent samples). d. SDS–PAGE of zein and non-zein proteins in B73 and 10 teosinte lines. The apparent size in kDa of each protein band is indicated on the left. M, protein markers. γ27, 27-kDa γ-zein; α22, 22-kDa α-zein; α19, 19-kDa α-zein; γ16, 16-kDa γ-zein; γ15, 15-kDa γ-zein; δ10, 10-kDa δ-zein.

3 times the SDS–PAGE analysis experiment were repeated independently with similar results. In a and b, a two-tailed Student’s t-test was used to determine P values, see Source Data.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Ames 21814 haplotype assembled by trio binning.

a, Phenotypes of teosinte Zea mays subsp. parviglumis (accession number: Ames 21814), B73 x Ames 21814 and B73. Scale bar, 35 cm. b, Teosinte haplotype genome assembly flow chart. To perform a de novo assembly of the teosinte haplotype, we sequenced and assembled its haplotype by integrating three technologies: HiFi long reads with the PacBio Sequel platform, paired-end sequencing with the Illumina HiSeq platform, high-throughput chromatin conformation capture (Hi-C). We completed assembly of the teosinte haplotype based on the trio-binning strategy because of the characteristics of high heterozygosity of Ames 21814.

b, Whole genome Hi-C interaction heat map of 2.5 Mb windows. Each blue number indicates the corresponding chromosome. c, Whole genome Hi-C interaction heat map. Each cluster represents a chromosome in the haplotype. In a set of chromosomes, the bottom cluster represents the hap1 (teosinte Ames 21814) chromosome, the top cluster represents the hap2 (B73) chromosome. d, Each chromosomes Hi-C contact map, Ames 21814 (Teo). e, Dot plot of B73 genome assembly (hap2, this study) and B73_v5 genome assembly. Alignment less than 20 kb was filtered out. f, Dot plot of Ames 21814 haplotype (hap1) and B73 haplotype (hap2). Alignment less than 20 kb was filtered out. g, Haplotype-specific inversions supported by the Hi-C contact map. Fourteen inversions larger than 1 Mb were selected for Hi-C zoom-in inspection by excluding those caused by tandem repeat arrays (CentC or knob). Thirteen inversions were correctly verified, while the Chr1:235 Mb inversion is a misscaffold contig.
Extended Data Fig. 3 | Variation in α-zein gene copies between Ames 21814 and maize inbreds. a, Copy numbers of α-zein genes in Ames 21814, B73 and W22. Gene numbers of α19 (z1A1, z1A2, z1B and z1D) and α22 (z1C1 and z1C2) are indicated beside each locus. b, Statistical analysis of copy number of α19 (z1A1, z1A2, z1B and z1D) and α22 (z1C1 and z1C2) in Ames 21814, B73 and W22.
Extended Data Fig. 4 | Diagram showing how the genetic populations were created and measurement of protein content in backcrossing populations.

a, The population size for each generation is indicated. b, SDS–PAGE showing zein accumulation in B73, teosinte (Ames 2184), and B73 x Ames 21814 seeds. c, Nitrogen content in roots of B73 and F₁ plants of B73 x Ames 21814. Data are presented as mean ± s.d. (n = 6 biologically independent samples for each method). d, Nitrogen content in stems of B73 and F₁ plants of B73 x Ames 21814. Data are presented as mean ± s.d. (n = 6 biologically independent samples for each method). e, Nitrogen content in leaves of B73 and F₁ plants of B73 x Ames 21814. Data are presented as mean ± s.d. (n = 6 biologically independent samples for each method). f, SDS–PAGE of zein accumulation in 10 F₂ seeds of B73 x Ames 21814. B73 seed was used as the control. g, SDS–PAGE of zein accumulation in seeds from 12 different BC₁ eras. B73 seed was used as the control. h, SDS–PAGE of zein accumulations in seeds from 12 different BC₂ eras. B73 seed was used as the control. i, Protein content in seeds from 30 different BC₂ ears. j, Protein content in 8 high-protein ears in the BC₂ population. Seven single seeds for each ear were measured. B73 was used as the control. k, The protein content in seeds from 30 different BC₃ ears. l, The protein content in 8 high-protein ears in the BC₃ population. Seven single seeds for each ear were measured. B73 was used as the control. In b, f, h, more than three times the SDS–PAGE analysis experiment were repeated independently with similar results. In c-e, a two-tailed Student’s t-test was used to determine P-values, see Source Data.
Extended Data Fig. 5 | Mapping THP9 by BSA sequencing of three populations of BC4, BC6, and BC8 and deep resequencing of five high-protein and five low-protein BC6F1 lines. a, Phenotyping high- and low-protein ears by SDS–PAGE of zein (upper panel) and non-zein protein (lower panel) accumulation in the BC4 population. M, protein mol wt markers. γ27, 27-kDa γ-zein; α22, 22-kDa α-zein; α19, 19-kDa α-zein; γ16, 16-kDa γ-zein; γ15, 15-kDa γ-zein; δ10, 10-kDa δ-zein. b, Gene introgression analysis based on BSA sequencing of the BC4 population. c, Frequency distribution analysis of seed protein content in the BC6 population. A group of 1,314 ears were phenotyped and classified. d, Gene introgression analysis based on BSA sequencing of the BC6 population. e, Frequency distribution analysis of protein content in the BC6 population. A group of 1,344 ears was phenotyped and classified. f, Gene introgression analysis by BSA sequencing of the BC8 population. Bottom is the differential expression of introgressed genes based on RNA-Se analysis of BC8 leaves. g, A total of 10 lines were analysed for gene introgression by resequencing. The high-protein lines are S2528, S2590, S2591, S2592 and S2596, and the low-protein lines are S2468, S2513, S2529, S2605 and S2609. The peak of introgressed teosinte DNA fragments in these lines was based on the B73 genome (B73_v4) coordinates are 13 Mb-143 Mb in S2528, 22.7 Mb-144.4 Mb in S2590, S2591, S2592 and S2596, 13 Mb-99 Mb in S2529, 22.7 Mb-135.5 Mb in S2605 and S2609. The smallest common region of the candidate interval is located between 135.5 and 143 Mb as indicated by the dotted box.
Extended Data Fig. 6 | Three genes annotated in the 150-kb fine-mapping region and ASN4 transcript analysis. a, Gene structure of Zm00001d047732 and SNPs in the coding sequences in B73 and Ames 21814. b, Expression fold change of the three genes in roots and leaves of NILB73 and NILTeo based on our RNA-Seq data. c, Schematic representation of ASN4 transcripts in B73 and Ames 21814. d, TPM mean analysis of ASN4 transcripts in B73 and Ames 21814 leaves and roots based on the RNA-seq data. TPM, Transcripts Per Million. Data are mean ± s.d. (n = 3 biologically independent samples). e, RNA-seq reads of THP9 in NILB73 (upper panel) and NILTeo (lower panel) leaves. The number refers to the number of reads across the junction. f, Expression of the ASN4 promoter sequences amplified from Ames 21814 and B73 compared by dual-luciferase (LUC) assay. Data are presented as mean ± s.d. *P* values, a two-tailed Student’s *t*-test was used to determine *P* values, see Source Data.
Extended Data Fig. 7 | Linkage of THP9-T with high asparagine and high protein contents in the BC F1 population and phenotypic comparison of NILTeo and NILB73. a, Phenotypes of three representative BC F1 ears with maternal genotypes that are homozygous THP9-B, heterozygous THP9-H (T/B) and homozygous THP9-T, respectively.
b, Free asparagine content in THP9-B, THP9-H and THP9-T roots in the BC F1 population. Data are mean ± s.d. (n = 10 biologically independent samples).
c, Protein content in seeds from self-pollinated THP9-B, THP9-H and THP9-T plants in the BC F1 population. Each ear was used as one unit. Data are mean ± s.d. (n = more than 20 biologically independent samples).
d, Plant phenotypes of NILTeo and NILB73. Scale bar, 30cm.
e, Plant height of NILTeo and NILB73. The plants were grown in Sanya in 2021. Data are mean ± s.d. (n = 18 biologically independent samples).
f, Plant fresh weight (root and aboveground mass) of NILTeo and NILB73. Data are mean ± s.d. (n = 6 biologically independent samples).
g, Association analysis of seed protein content with total stem nitrogen content in the BC F1 population. The corresponding seed protein content and stem nitrogen content of the same plant are connected by a solid grey line (n = 1334 biologically independent samples).
h, The total nitrogen content in NILTeo and NILB73 roots, stems, and leaves. Data are mean ± s.d. (n = 33, 41 and 20 biologically independent samples, respectively). i, Total free amino acid content in NILTeo and NILB73 roots and leaves. Data are mean ± s.d. (n = 20 biologically independent samples, respectively).
In b–c, different letters indicate significant differences (P < 0.01, one-way ANOVA and further Tukey’s test, see Source Data). In e, f and h, i, a two-tailed Student’s t-test was used to determine P values, see Source Data.
Extended Data Fig. 8 | NILB73 and NILTeo under normal and low-nitrogen conditions. a-c, Construction of four aboveground concrete containers with plastic film covering the containers. d, NILB73 and NILTeo grown in container without nitrogen fertilizer application. e, The nitrogen content of soil in containers with and without nitrogen application. Data are mean ± s.d. (n = 16 biologically independent samples). f, Plant phenotypes of NILB73 and NILTeo with and without nitrogen application. Scale bar, 20 cm. g, Root phenotypes of NILB73 and NILTeo with and without nitrogen application. Scale bar, 5 cm. h, Plant height of NILTeo and NILB73. Data are mean ± s.d. (n = 8 biologically independent samples). i, qRT–PCR analysis of THP9 expression in NILB73 and NILTeo roots with and without nitrogen application. Data are mean ± s.d. (n = 3 biologically independent samples). j, Root fresh weight of NILB73 and NILTeo with and without nitrogen application. Data are mean ± s.d. (n = 8 biologically independent samples). k, Aboveground biomass of NILB73 and NILTeo with and without nitrogen application. Data are mean ± s.d. (n = 14 biologically independent samples). l, Total nitrogen content in NILB73 and NILTeo roots with and without nitrogen application. Data are mean ± s.d. (n = 6 biologically independent samples). m, Total nitrogen content in NILB73 and NILTeo stems with and without nitrogen application. Data are mean ± s.d. (n = 10 biologically independent samples). n, Total nitrogen content in NILB73, and NILTeo leaves with and without nitrogen application. Data are mean ± s.d. (n = 6 biologically independent samples). o, Protein content in NILB73 and NILTeo seeds with and without nitrogen application. Data are mean ± s.d. (n = 20 biologically independent samples). In e and h–o, letters indicate significant differences (P < 0.01, one-way ANOVA and further Tukey’s test).
Extended Data Fig. 9 | Wild-type B73 and THP9-OE2 under normal and low-nitrogen conditions. a, Plant phenotypes of WT (B73), NILTeo and THP9-OE2 with and without nitrogen application. Scale bar, 20 cm. b, Root phenotypes of WT, NILTeo and THP9-OE2 with and without nitrogen application. Scale bar, 5 cm. c, The nitrogen content of soil in containers with and without nitrogen application. Data are mean ± s.d. (n = 8 biologically independent samples). d, Plant height of WT and THP9-OE2. Data are mean ± s.d. (n = 8 biologically independent samples). e, The root fresh weight of WT and THP9-OE2 with and without nitrogen application. Data are mean ± s.d. (n = 14 biologically independent samples). f, The aboveground biomass of WT and THP9-OE2 with and without nitrogen application. Data are mean ± s.d. (n = 14 biologically independent samples). g, Total nitrogen content in WT and THP9-OE2 roots with and without nitrogen application. Data are mean ± s.d. (n = 12 biologically independent samples). h, Total nitrogen content in WT and THP9-OE2 stems with and without nitrogen application. Data are mean ± s.d. (n = 12 biologically independent samples). i, Total nitrogen content in WT and THP9-OE2 leaves with and without nitrogen application. Data are mean ± s.d. (n = 12 biologically independent samples). j, Protein content in WT and THP9-OE2 seeds with and without nitrogen application. Data are mean ± s.d. (n = 20 biologically independent samples). In c-g, letters indicate significant differences (P < 0.01, one-way ANOVA and further Tukey’s test).
Extended Data Fig. 10 | Seed protein content of NILTeo × Mo17 and Zhengdan 958-T hybrids. a, Seed protein content in Zhengdan 958-B and Zhengdan 958-T seeds. Data are presented as mean ± s.d. (n = 20 biologically independent samples). Protein content in F₂ seeds of NILB73 x Mo17 and NILTeo x Mo17. Data are presented as mean ± s.d. (n = 30 biologically independent samples). In a, b, a two-tailed Student’s t-test was used to determine P values, see Source Data.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and/or variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- PacBio Sequel II (CCS reads), Illumina HiSeq PE150 (Hi-C reads), Teosinte Ames21814 haplotype assembly and data collection; Dumas rapid nitrogen analyzer (N content); UPLC [Ultra Performance Liquid Chromatography UPLC, Ultimate 3000]-MS/MS (Tandem mass spectrometry, API 3200 Q TRAP), MSLab HP-C18, [free amino acid measurement]; Illumina novaseq6000 (BSA sequencing and resequencing mapping); Illumina novaseq6000 [RNA-seq]; Bio-Rad CFX Manager v3.1[RT-qPCR data]; Tanon-5200 (immunoblot image).

Data analysis

- GraphPad Prism v8.0.2 (one-way ANOVA and Tukey’s test), Microsoft excel 2016 (two-side Student’s t-test); QC [fastqc/0.19.5 FastQC/0.11.8], hifiasm/0.16.1, minimap2/2.20. samtools/1.9, gflatools/0.14, juicer-tools/1.22.01, juicer/1.5.7, 3D-DNA/20180419, JuicerBox/1.11.08, bwa/0.7.17, yak/0.1+r2-dirty, EDTA/v1.9.4 bedtools/2.27.1, braker/2.2.13, hisat2/2.10.2, StringTie/1.3.3, gffread/0.9.12, gmap/v2018-07-04, diamond/0.9.1, GeneMark-ES/3.4.8, maker/2.31.10, augustus/3.3.3, BUSCO/3.1.0 [Teosinte Ames21814 haplotype assembly and annotation]; Picard [v4.1.0.0], GATK [v4.1.0.0], R [v3.5.1.1] (BSA sequencing and resequencing mapping); htsseq-count, DESeq2 (RNA-seq analysis); Analyst software (v1.5.1) (free amino acid measurement); SnapGene (sequence analysis); Kallisto (v0.44.0), syn/1.5.4 (transcript quantification and structural variation); BLASTN: 2.9.0+ (zinc copy number analysis); readxl/1.4.0, ggplot2/3.3.5, dplyr/0.8.0, ggsci/2.9 (correlation analysis of seed protein content and stem N content); GEMMA (v0.98.1) (GWAS); rapid N exceed software [v1.1.25] (measurement of protein content with the Rapid N analyzer).

For manuscripts utilizing custom algorithms or software that are not central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Portfolio guidelines for submitting code & software for further information.
Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Ames21814 genome sequences have been deposited in NCBI (PRJNA822523). The genome assembled data have been deposited in NCBI (PRJNA861737). The RNA-sequencing data of Ames21814, B73 x Ames21814 and B73 (roots and leaves at flowering stage) have been deposited in NCBI (PRJNA832948). Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/s41590-021-03081-7.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to predetermine sample size. Sample sizes were estimated based on preliminary experiments and previously published results. We made the effort to achieve a large sample size in field trial which proved to be sufficient to reproducibly observe statistically significant differences, sample size is stated in each panel. For the number of near-isogenic lines in population, from BC4 to BC8, more than 2000 were planted in each generation. For NUE test in the field, each treatment contained 300 plants, and 4 nitrogen application levels, 3 replicates were performed for a total of 7200 plants.

Data exclusions

No data was excluded from the analyses.

Replication

All experiments in this study were repeated independently at least three times. This information is shown in figure legends.

Randomization

All samples were arranged randomly into experimental groups. Plants of equal initial sizes were randomly assigned to the treatment and control groups for NUE test.

Blinding

For molecular biology experiments, bias could not be introduced since samples were treated identically and collected randomly. No analyses required being blind to groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | Antibodies |
| ☐ | Eukaryotic cell lines |
| ☒ | Palaeontology and archaeology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |
| ☒ | Dual use research of concern |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | ChiP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

### Antibodies

**Antibodies used**

- Anti-ASN4 (made by Abclonal, 1:1000);
- Anti-FLAG (Sigma, A8592, 1:1000);
- Anti-ACTIN (Abmart, catalog number M20009L, 1:1000);
- anti-rabbit-HRP (Abmart, catalog number: M21002L, 1:5000);
- anti-mouse IgG-HRP (Abmart, catalog number: M21001L, 1:5000) for immunoblot analysis.

**Validation**

ASN4 protein fragment from the 460th to 588th amino acid was used to make antibodies by Abclonal (Wuhan, China), Nil.B73 and
NILOs were used to validate anti-ASNA5 antibodies. Mouse monoclonal ACTIN antibody (Abmart, catalog number M200009), a secondary antibody, repeatedly used in our previous research. The monoclonal antibody anti-FLAG (Sigma, A8592) binds to fusion proteins containing a FLAG peptide sequence at the N-terminus of Thp9. The specific band was shown in the WB using leaf protein of Thp9-OE in Fig 3c, but without in WT.