Teosinte high protein 9 enhances the seed protein content and nitrogen utilization efficiency in maize

Yongrui Wu (yrwu@cemps.ac.cn)
National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China
https://orcid.org/0000-0003-3822-0511

Yongcai Huang
National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Haihai Wang
National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Yidong Zhu
National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Xing Huang
National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Shuai Li
Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University

Xingguo Wu
Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University

Yao Zhao
State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University

Zhigui Bao
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

Li Qin
Institute of Molecular Breeding for Maize, Qilu Normal University

Yongbo Jin
Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University

Yahui Cui

Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University

Guangjin Ma

National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Qiao Xiao

National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Qiong Wang

National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Jiechen Wang

National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai 200032, China

Xuerong Yang

State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University

Hongjun Liu

Shandong Agricultural University https://orcid.org/0000-0001-5123-688X

Xiaoduo Lu

Qilu Normal University

Brian Larkins

School of Plant Sciences, University of Arizona

Wenqin Wang

Shanghai Normal University https://orcid.org/0000-0001-6427-6338

Biological Sciences - Article

Keywords: maize, teosinte, high protein, quality, NUE

Posted Date: June 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1701760/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature on November 16th, 2022. See the published version at https://doi.org/10.1038/s41586-022-05441-2.

Teosinte high protein 9 enhances the seed protein content and nitrogen utilization efficiency in maize

Yongcai Huang[1,8], Haihai Wang[1,8], Yidong Zhu[1,2,8], Xing Huang[1,2], Shuai Li[3], Xingguo Wu[3], Yao Zhao[5], Zhigui Bao[7], Li Qin[4], Yongbo Jin[3], Yahui Cui[3], Guangjin Ma[1,2], Qiao Xiao[1,2], Qiong Wang[1], Jiechen Wang[1], Xuerong Yang[5], Hongjun Liu[5], Xiaoduo Lu[4], Brian A. Larkins[6], Wenqin Wang[3*], and Yongrui Wu[1*]

[1]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Chinese Academy of Sciences, Shanghai 200032, China.

[2]University of the Chinese Academy of Sciences, Beijing 100049, China.

[3]Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, 200234, Shanghai, China.

[4]Institute of Molecular Breeding for Maize, Qilu Normal University, Jinan, China.

[5]State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural UniversityTai'an, China.

[6]School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA

[7]Shenzhen 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, Buxin Road 97, Shenzhen, 518120, Guangdong, China

[8]These authors contributed equally to this work.

[*]Correspondence should be addressed to Y.W. (yrwu@cemps.ac.cn) and W.W. (wang2021@shnu.edu.cn)

Running title: *Thp9* confers high-protein and high-NUE traits in maize

Keywords: maize, teosinte, high protein, quality, NUE

Abstract

The maize (*Zea mays* ssp*. mays*) wild ancestor, teosinte, has three times the seed protein content of most modern inbreds and hybrids, but the mechanisms responsible for this trait are unknown. We created a contiguous haplotype DNA sequence of a teosinte, *Zea mays* ssp*. Parviglumis*, with Trio-Binning, and map-based cloned a major high-protein QTL, *teosinte high protein 9* (*Thp9*) on chromosome 9. *Thp9* encodes an asparagine synthetase 4 that is highly expressed in teosinte, but not in the B73 inbred, where a deletion in the 10th intron of *Thp9-B73* causes incorrect splicing of *Thp9-B73* transcripts. Transgenic expression of *Thp9-teosinte* in B73 significantly increased seed protein content. Introgression of *Thp9-teosinte* into modern maize inbreds and hybrids greatly enhanced free amino acid accumulation, especially asparagine, throughout the plant, increasing seed protein content without affecting yield. *Thp9-teosinte* appears to increase nitrogen utilization efficiency, important for promoting a high yield under low nitrogen conditions.

INTRODUCTION

Seeds of plants contain stored metabolites, e.g., carbohydrates, proteins, lipids, and nucleic acids, all of which are important for rapid cell division and growth during the transition from dormancy to photosynthetic autotrophy when environmental conditions are suitable for germination[1,2]. These metabolites also make seeds a valuable source of food for a variety of animals, including humans[3]. Over millennia, plant breeders have genetically altered plant species to create seeds with greater proportions of these metabolites and improve their nutritional value and utility as food and feed[4]. Perhaps one of the most striking examples was the conversion of the wild maize relative, teosinte, to modern maize[5,6].

Native Americans selected mutations that modified a variety of teosinte traits, notably its floral inflorescence and seed size, structure, and yield[7,8]. Because of its importance in their diet, the maize they created had a high protein content, enhanced favor, and utility for food making. However, as corn became a commodity and used for livestock feed, starch content (yield) was of primary importance, and less attention paid to protein content and flavor[9]. Also, the use of supplemental nitrogen (N) fertilizer reduced the importance of seed N content, and as a consequence modern maize hybrids contain only 5%-10% protein, compared with 20%-30% in teosinte[10].

While N fertilizer dramatically improves maize yield, its excessive use often leads

to run off, causing eutrophication of rivers and other bodies of water[11]. Consequently, future maize breeding must design plants with greater N use efficiency (NUE)[12]. In addition, seed protein content and quality will have greater importance, as vegetable protein will become of increased importance in human diets[13].

To identify genes responsible for variation of the protein content in maize and teosinte, we analyzed progeny of their cross and characterized the quantitative trait loci (QTL) affecting this trait. We sequenced a teosinte haplotype genome (*Zea mays* ssp. *parviglumis*, Ames21814) and localized loci linked with high seed protein content. 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 (ASN4) that plays an important role in amino acid accumulation throughout the plant. The *Thp9-teosinte* allele, *Thp9-T*, is highly expressed in teosinte, while the corresponding allele in the maize B73 inbred, *Thp9-B73*, has a 48-bp deletion in the 10th intron that affects intron splicing. Several versions of the *Thp9-B* transcripts contain a premature stop codon, rendering undetectable levels of the ASN4 enzyme. Introgression of *Thp9-T* into maize inbreds and hybrids increased the amino acid and protein content in roots, stems, and leaves, as well as seeds. These plants elicited higher NUE under low N conditions than those with the *Thp9-B* allele and show promise for improving maize germplasm in general.

Result

High protein content in teosinte

During maize domestication and artificial selection, many visible (such as plant and glume architectures) and invisible (seed composition) traits were dramatically modified (Fig. 1a). To investigate variation in seed protein content between teosinte and modern maize inbreds, we collected 20 lines of *Zea mays* ssp. *parviglumis*, 10 lines of *Zea mays* ssp. *mexicana* and 518 modern maize inbreds. In maize seeds, most nitrogen (N) occurs in storage proteins, so the total N content is basically equal to that. In roots, stems, and leaves, total N reflects the sum of N in free amino acids and proteins, but most of it is in protein. We determined the N content in roots, stems, leaves and seeds of B73 by two procedures, namely acid hydrolysis and the Dumas method, the latter using a Dumas rapid N analyzer (Elementar, Germany). There was no significant difference in N content measured by the two methods (Extended Data 1a). Therefore, we could use the high throughput Rapid N analyzer to assay seed protein

content and N in plant tissues. Seed protein content of all teosinte lines was around 30%, while that of maize inbreds (435 finally harvested for measurement) ranged from 7% to 17% (Fig. 1b). These differences suggested loci controlling high seed protein content were not under selection in recent plant breeding programs, and that these loci are genetically variable among inbred lines.

We selected one line of *Zea mays* ssp. *parviglumis* (accession number: Ames21814) as a representative high protein genotype for analysis. When total N content in roots, stems, and leaves of Ames21814 and B73 was measured, we found total N content in all tissues of Ames21814 was higher than in B73 (Extended Data 1b). The composition of free amino acids differed to some extent (Extended Data 1c), in particular, asparagine was notably higher in all tissues of Ames21814 than B73 (Fig. 1c, Extended Data 1c). This is consistent with a previous observation in rice seed showing an increased asparagine level is associated with a high protein content[14].

Maize seed proteins are classified as prolamins (called zeins), albumins, globulins and glutelins, based on their solubility[15]. Zeins are the main endosperm storage proteins, accounting for more than 60% of the total. Based on their structure, zeins are divided into four families: α (19- and 22-kD, designated α19 and α22), β (15-kD), γ (50-, 27- and 16-kD) and δ (18- and 10-kD). α19 is further divided into z1A, z1B and z1D subgroups, and α22 into z1C[16,17]. SDS-PAGE of zein proteins in teosinte and B73 seeds revealed α19 and α22 were more abundant in teosinte (Extended Data 2), indicating α-zeins are the major component contributing to their higher protein content.

We analyzed zein content in 512 inbred lines by SDS-PAGE (Source Extended Data 3a) and found the abundance of zeins in this population varied greatly (Extended Data 3a). We graded α-zein content into three phenotypes base on differential accumulation (i.e., α19 more than, equal to, and less than α22, Source Extended Data 3b), and performed a genome-wide association analysis (GWAS). GWAS revealed a predominant locus controlling variation in α19 and α22 content on the short arm of chromosome 4 (Extended Data 3b), precisely where *z1A* and *z1C* genes are clustered[18]. This suggested gene copy number could lead to differential accumulation of α19 and α22 in the inbred population. By implication, this encouraged us to investigate the copy number of *α-zein* genes in teosinte.

Haplotype-resolved assembly of Ames21814 genome using Trio-binning

Because of advances in long-read DNA sequencing technologies, many maize

inbred genomes have been characterized. For an inbred, the two genome haplotypes are theoretically identical and can be easily collapsed into one type[19]. However, teosinte lines are open-pollinated and their genomes have a high degree of heterozygosity. So, routine genome sequencing and assembly would collapse different homologous haplotypes into a consensus representation of frequent heterozygous alleles. This could cause loss of half of the genetic information in a diploid genome.

To create a high-quality teosinte genome, we sequenced the DNA of a single F_1 plant from the B73 x Ames21814 cross. The genome sequence of B73 is known, and we used the graph-based Trio-binning strategy[20] to untangle the haplotype information (Extended Data 4a and b). The initial Ames21814 haplotype was assembled into 2,424 Mb by using 104-Gb HiFi long reads with a 47-fold coverage, resulting in 543 contigs with an N50 of 62.29 Mb (Supplementary Table 1), longer than the recently sequenced 26 diverse maize genomes used as the maize Nested Association Mapping (NAM) population generated with the contig N50 of 6.26-52.36 Mb[19]. Two haplotigs were combined and scaffolded into 10 pseudo-chromosomes using 375.56 Gb HiC reads (Fig. 1d and Extended Data 4c). We also assembled a B73 haplotype that showed an excellent collinearity with the reference B73_v5, indicating the correctness of the assembled pipeline (Extended Data 4d). The final assembly size of the genome was 2,436 Mb, with 452 scaffolds and an N50 scaffold of 245.32 Mb (Supplementary Table 1), 11.64% more sequences than B73_v5 (2,182 Mb), which could either be from capturing more obstinate sequences by long-read sequencing or from abundant gene pools in teosinte. The teosinte genome was predicted to harbor 58,092 protein-coding genes (Supplementary Table 2), of which 46,473 (81.3%) were supported by RNA-seq data and IsoSeq. BUSCO[21] analysis, indicated that 96.8% of embryophyta genes were complete in the Ames21814 assembly (Supplementary Table 3). A total of 2,109-Mb (equal to 86.57% of the genome) wase identified as repetitive sequences, including the most abundant retrotransposons (67.61%) (Supplementary Table 4). The total repeat content in Ames21814 was much higher than 1860 Mb in B73, due to the contribution of the Knob and CentC repeats in maize (~10.15%) (Fig. 1d and Supplementary Table 4), where previous study showed that Knobs could affect local recombination[22]. The LTR Assembly Index (LAI) score[23] (LAI=33.07) indicated the Ames21814 haplotype could be a useful reference to discover valuable alleles in the wild maize ancestoral genome. Although there was good collinearity between B73_v5 and Ames21814, 197,346 structural variations (>50 bp) were identified (Supplementary Table 5), especially 71 inversions larger than 100 kb. Whether these structural variations correlate with phenotypic variation between teosinte and modern

maize needs experimental validation.

Using the highly contiguous Ames21814 haplotype, we were able to annotate all the *α-zein* genes. The total copy number of *α19* (*z1A1*，*z1A2*，*z1B*，*z1D*) and *α22* (*z1C1*, *z1C2*) genes in Ames21814 was 22 and 12, respectively, compared with 25 and 15 in B73, and 25 and 19 in W22, respectively. Both the total number of genes and the number of genes in the *α-zein* cluster on the short arm of chromosome 4 were smaller in Ames21814 than in B73 and W22 (Extended Data 5a and b), suggesting the high-protein trait in teosinte is not conferred by a larger number of *α-zein* genes. Indeed, the accumulation of non-zein proteins in teosinte seeds was also apparently higher than that in B73 (Extended Data 2), which suggests the high-protein QTLs in teosinte generally, rather than specifically, increase protein content.

Map-based cloning of the high-protein locus in Ames21814

To identify the QTLs associated with the high-protein trait in Ames21814, we created a series of continuous backcrossing populations using Ames21814 as the high-protein donor parent and B73 as the recurrent backcrossing parent. Due to unidirectional incompatibility between teosinte and modern maize, we used Ames21814 to pollinate B73 for the F_1 progeny (Extended Data 4a). We measured protein content with the Rapid N analyzer and found the F_1 seeds (B73 x Ames21814, 11.6±0.8%) had a protein level like B73 (10.8±1.0%), and Ames21814 seeds had a protein content of 28.6±1.0% (Fig. 2a), consistent with the accumulation pattern of zein proteins, where α-zeins (α19 and α22) are a major indicator of the total protein content (Extended Data 6a). However, the F_2 seeds had nearly double the protein content (19.9±1.2%) of F_1 and B73 seeds (Fig. 2a). When F_2 seeds were analyzed individually by SDS-PAGE, there was no apparent variation in zein protein accumulation, and α-zeins in particular were remarkably more abundant compared with B73 seeds (Extended Data 6b and Source Extended Data 6b). These results indicated the high-protein trait is determined by the maternal rather than the filial genotype.

Because the F_1 plants displayed many rudimentary teosinte phenotypes in vegetative and reproductive growth (Extended Data 4a), we used B73 as the ear parent to make the F_1BC_1. Afterwards, we used B73 as the pollen source for backcrossing. In the F_1BC_2 ((B73 x Ames21814) x B73), we observed segregation of zein protein content among different ears in a quantitative rather than a qualitative pattern (Extended Data 6c and Source Extended Data 6c), indicating the high-protein trait is regulated by multiple genetic loci. Like the F_2 seeds (Extended Data 6b), when

individual seeds from a high-protein F_1BC_2 ear were analyzed, each seed uniformly accumulated more α-zeins than B73 (Extended Data 6d). Subsequent backcrossing generated eight ears with the highest protein content (about 15%); they were saved and seeds from each ear were planted for analysis. Similarly, quantitative measurement with the Rapid N analyzer of the F_1BC_3 and F_1BC_4 generations confirmed that protein content varied among different ears (ranging from 10-15%), but it was uniform in individual seeds of the same ear (Extended Data 6e-h).

To identify the genetic loci influencing protein content, we planted the F_1BC_3 seeds; a piece of leaf from each plant was saved for DNA extraction. Zein and non-zein proteins from 500 F_1BC_3 ears were analyzed by SDS-PAGE (Extended Data 7a, Source Extended Data 7a). Based on 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 revealed evidence of several QTLs, with a significant peak in the region between 130 Mb and160 Mb (based on Teo_v1) on chromosome 9 (Fig. 2b and Extended Data 7b) that contained 315 introgressed teosinte genes. Accordingly, this locus was designated *teosinte high protein 9* (*Thp9*).

Using the same approach, we created F_1BC_6 (n=1314) and F_1BC_8 (n=1386) populations. BSA of F_1BC_6 and F_1BC_8 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 0.025 threshold) (Extended Data 7c-f). We performed high-coverage (>20x) resequencing of this region in five high-protein and five low-protein individuals from the F_4BC_6, and found 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 (135.5-143 Mb) should be the candidate interval (Extended Data 8). Nearly isogenic lines, NILTeo and NILB73, with high and low protein levels, respectively, were created based on this interval.

To fine map *Thp9*, we created a F_1BC_9 generation (n = 2000) that narrowed *Thp9* to a 150-kb region containing two genes based on the B73 reference genome (B73_v4) (Fig. 2c). One gene, Zm00001d047732, encodes a protein phosphatase and lacks structural variation between Ames21814 and B73, while the other, *Zm00001d047736*, corresponding to *Teo09G002926* in Ames21814, encodes an asparagine synthetase 4 (ZmASN4*)*. Analysis of the Ames21814 and B73 sequences revealed *Teo09G002926* is an intact *ASN4* gene (hereafter referred to as *Thp9-Teosinte*, *Thp9-T*), while *Zm00001d047736* has a 48-bp deletion in the 10th

intron of *ASN4* (hereafter referred to as *Thp9-B73*, *Thp9-B*; Fig. 2c).

Based on published data, this deletion creates altered splicing of *Thp9-B* transcripts, resulting in 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 9a). RNA-seq revealed *Thp9-T* transcripts (*ASN4-Teo*) accumulate abundantly in roots and leaves of Ames21814, whereas the *ZmASN4-T003* isoform was barely detectable in these tissues of B73 (Extended Data 9b). Further RNA-seq analysis of NILTeo and NILB73 confirmed *Thp9-T* is highly expressed, while *Thp9-B* is barely expressed in root and leaf tissues (Fig.2d, Extended Data 9c). Consistent with the transcript levels, ASN4 protein is abundantly accumulated in NILTeo, but is absent in NILB73 (Fig.2e). The results suggest the 48 bp deletion in the 10th intron of the *ASN4* gene in B73 seriously affects RNA splicing and stability of *ASN4* transcripts, making them and the ASN4 protein difficult to detect. Therefore, *Thp9-B* can be considered a null allele.

We developed a molecular marker for *Thp9-B* and used it to genotype 200 individuals in the F_3BC_7 population (Extended Data 10a). When we measured the free asparagine content in roots of *Thp9-T* and the heterozygote of *Thp9-H* (*T/B*), we found significantly higher aspargine than in *Thp9-B* (Extended Data 10b). The protein content of *Thp9-T* and *Thp9-H* seeds was also significantly higher than in *Thp9-B* seeds (Extended Data 10c), confirming the *Thp9-T* allele is associated with higher protein content. The protein content of NILTeo seeds grown in Harbin (northeast China), Shanghai (east China), and Sanya (south China) was 12 ±0.7%, 13.1±0.4% and 15.4 ±0.9%, respectively, while that of NILB73 was 9.2±0.5%, 9.7±0.4% and 11.2 ±0.9%, respectively. Thus, the *Thp9-T* allele increased protein content by 30.4%, 35.2% and 37.8% at the three different geographic locations (Fig. 2f). We also compared other phenotypes of NILTeo and NILB73 in Sanya. NILTeo showed a 7.6% increase in plant height (Extended Data 11a and b) and 15.1% increase in plant fresh weight (root and above-ground mass) compared with NILB73 (Extended Data 11c).

Since seed storage proteins function as a sink for N storage, we wondered whether the total nitrogen content, most existing as free amino acids and proteins, was elevated in source tissues. We used the Rapid N analyzer to measure N in the stems and corresponding ears of 1,334 F_1BC_8 plants and found they were highly correlated (Extended Data 11d). This correlation was also observed in NILs, where NILTeo had increased total N content in roots, stems, and leaves (Extended Data 11e), as well as

total free amino acid contents in roots and leaves, compared to NILB73 (Extended Data 11f). In addition, the levels of free asparagine in NILTeo roots and leaves were significantly higher than those in NILB73 (Fig. 2g), indicating increased accumulation of asparagine via *Thp9-T* facilitates increased synthesis of proteins in roots, stems, leaves and seeds.

Genetic 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-OE* plants had greatly enhanced levels of *ASN4* transcript and protein in leaves and roots compared with the non-transgenic B73 control (Fig. 3a-c). Two representative *Thp9-OE* lines grown in Sanya were analyzed. The seed protein contents of *Thp9-OE-1* and *Thp9-OE-2* were 15.2 ±1% and 15.8 ±1%, a 25.7% and 30.9% increase, respectively, over the B73 control (12.08±0.88%), (Fig. 3d). These results are consistent with the hypothesis that the mutation in *Thp9* is responsible for the low protein phenotype of B73.

We measured the seed protein content of 405 and 438 maize inbreds grown at Sanya in 2019 and 2020. The protein content of the 2019 crop ranged from 6.5% to 16%, with an average of 11.5%, while that in 2020 varied from 7.7% to 16.8%, with an average of 12.30% (Source Data Fig. 3e). GWAS analysis of seed protein content in these inbreds identified a region with physical coordinates near the *Thp9* locus (Fig. 3e). PCR and sequencing of 215 inbreds revealed the *Asn4* gene in this population had three haplotypes (HAP1-3) based on an InDel polymorphism in the 10th intron: HAP1 inbreds (23.7%) belong to the Ames21814 genotype (*Thp9-T*) with an intact *ZmASN4* coding sequence; HAP2 inbreds (45.1%) harbor a 22-bp deletion in the 10th intron that apparently doesn't affect splicing (the allele designated *Thp9-T'*); HAP3 inbreds (31.2%) have the B73 genotype (*Thp9-B*) with a 48-bp deletion in the intron (Fig.3f). We found the high-protein inbreds had a protein content higher than 13%, and low-protein inbreds less than 10%. HAP1 had the highest percentage of high-protein inbreds (21 out of 51, 41.2%), followed by HAP2 (20 out of 97, 20.6%) and HAP3 (3 out of 67, 4.5%), whereas HAP3 had the highest percentage of low-protein inbreds (17 out of 67, 25.4%), followed by HAP2 (6 out of 97, 6.2%) and HAP1 (0 out of 51, 0.0%). HAP1 inbreds also had the highest average protein content (12.6±1.1%), while HAP2 was medium (11.9±1.3%) and HAP3 the lowest (10.9±1.3%, Fig. 3g). These results support the hypothesis that *Thp9* is a major QTL influencing variation of seed protein content among inbred lines.

Thp9-T increases NUE

Since *Thp9-T* increases the free amino acid composition in plants, which in turn promotes plant development and protein accumulation in seeds, we investigated whether *Thp9-T* can increase NUE. To this end, we set up an experimental site on our farm in Shanghai to test the effects of applying different levels of N fertilizer on plant growth. Several above ground concrete containers with different soil N concentrations were constructed. A plastic film covered the containers to prevent rainwater from affecting the soil N concentration (Extended Data 12a-c). More than 50 NILTeo and 50 NILB73 plants were grown side by side in containers with normal N application (40 g/plant) and without N application. NILTeo plants appeared to grow better than NILB73 plants under normal and low N conditions (Extended Data 12d-f). We measured N level in the soil and found the normal N pool contained 76.7% more N than the low N pool (Extended Data 12g). RT-qPCR showed that *Thp9-T* but not *Thp9-B* was dramatically induced for expression when N was applied, suggesting *Thp9-T* is sensitive to soil N level (Extended Data 12h). Without N application, both NILTeo and NILB73 plants were slender and had a smaller amount of root mass than with normal N, but NILTeo plants with less N were comparable in size to NILB73 plants with normal N (Extended Data 12e and f). Root fresh weight and above-ground biomass of NILTeo and NILB73 plants were greatly reduced by low N, but there was no significant difference between NILTeo under low N and NILB73 under normal N (Extended Data 12i and j). The total N content (mostly free amino acids and proteins) in roots, stems, and leaves, and the protein content in seeds of NILTeo and NILB73 were affected by low N, but their contents in NILTeo under low N were comparable to those of NILB73 under normal N (Extended Data 12k-n).

Subsequently, in 2021 we set up a larger field trial in Sanya, where we applied different amounts of N, i.e., 100% (32 g/plant), 50% (16g/plant), 25% (8g/plant) and 0%. In each trial, 300 seeds of NILTeo and NILB73 were planted together (Fig. 4a). NILTeo showed an apparent growth advantage over NILB73 in terms of plant height (Fig. 4b) and above-ground biomass (Fig. 4c) under the different N conditions. Total N content in roots, stems, and leaves of NILTeo was significantly higher than in NILB73 in all trials (Fig. 4d-f). After reducing N application from 100% to 0%, the protein content in NILTeo seeds decreased from 14.2%, to 13.5%, 12% and 10.7%, while in NILB73 seeds it decreased from 11.4%, to 11.2%, 10.6% and 8.9% (Fig. 4g). The results indicate seed protein content is sensitive to soil N level, and in each treatment NILTeo seeds always maintained a higher level of protein than NILB73 seeds. The protein content in NILTeo seeds harvested at 25% N reached 12%, which was higher than

NILB73 seeds (11.4%) with normal N. These results are consistent with the hypothesis that *Thp9-T* confers higher NUE than *Thp9-B* in NILB73 at normal and low N conditions.

Creation of high-protein maize germplasms with *Thp9-T*

B73 and Mo17 (HAP2 type) are inbreds frequently used to study hybrid vigor. To examine whether *Thp9-T* can increase the protein content of their hybrid and influence other agronomical traits, we made two sets of F_1 seeds, i.e., NILTeo x Mo17 and NILB73 x Mo17, and planted them at Harbin, Northeast China. The protein content of F_2 seeds from the NILTeo by Mo17 cross (9.2±0.6%) was 7.8% higher than that made from the NILB73 by Mo17 (8.6±0.4%) cross, while the 100-kernel weight was nearly identical for the two hybids (35.4 g vs 35.6 g, Fig. 5a-c).

We also introgressed *Thp9-T* into Zheng58 (HAP3 type) and Chang7-2 (HAP2 type), two elite inbred lines that make the Zhengdan958 hybrid, which has the largest growing area in China. The *Thp9-T* modified (designated Zhengdan958-T) and the unmodified Zhengdan958 (designated Zhengdan958-B) hybrids were grown for comparison in Sanya in 2021 (Fig. 5d-e). Zhengdan958-T manifested a significantly increased above-ground plant weight and height, compared with Zhengdan958-B (Fig. 5f and g), and Zhengdan958-T seeds had a protein content of 11.1±1.1%, a 12.8% increase compared to Zhengdan958-B seeds (9.9±0.6%) (Fig. 5h). The total nitrogen content in roots (Fig. 5i), stems (Fig. 5j) and leaves (Fig. 5k) of Zhengdan958-T also significantly increased. However, the 100-kernel weights of Zhengdan958-T and Zhengdan958-B was not significantly different (Fig. 5l). The results suggest *Thp9-T* has potential value to improve the protein content of maize seeds and plants through plant breeding.

Discussion

Genetic complexity of seed protein content and NUE in maize

Genetic variability for seed protein content is well documented in maize. More than 100 years ago, the University of Illinois initiated a breeding program to demonstrate the consequence 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 materials: Illinois High Protein (IHP), Reverse High Protein (RHP), Illinois Low Protein (ILP) and Reverse Low Protein (RLP), which had protein contents ranging from about 30%, 7%, 4% and 15%, respectively[24]. The results

of this experiment suggest the existence of both positive and negative genetic factors influencing protein content in natural maize populations that are likely controlled by multiple genetic loci[25].

Because modern maize was domesticated from teosinte, we reasoned that characterization of the genes responsible for the high-protein trait in teosinte might reveal a more diverse set of QTLs than those found in recent maize inbred populations. The results might also help us understand the reasons why seed protein content decreased during maize domestication[10]. Also, teosinte contains high levels of free amino acids, especially asparagine, in the roots, stems and leaves (Fig. 1c and Extended Data 1c), suggesting it has high N assimilation, which could contribute to 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* ssp. *parviglumis*, Ames21814 (Fig. 1d and Extended Data 4).

Zeins are the principal storage proteins in maize endosperm and are indicative of total seed protein content[26]. There are multiple factors that influence zein accumulation: 1) zein gene number and transcriptional regulation; 2) zein mRNA level and association of these transcripts with ribosomes and aminoacyl tRNAs; and 3) translational and post-translational processing of zeins into protein bodies. Sink strength of the ear is also a factor because it draws sugars and amino acids from plant tissue into the kernel, providing energy and source materials for storage metabolite synthesis. In principle, one or more of these processes could be the most limiting factor for seed protein content.

The results of our experiments show gene copy number variation influences the relative amount of α19 and α22 zeins (Extended Data 3), but it doesn't determine total seed protein content (Fig 2a). The copy number of *α-zeins* in Ames21814 is less than that in B73 and W22, consistent with the hypothesis that high seed protein content in teosinte seeds is not a consequence of *α-zein* gene copy number (Extended Data 5a and b). A previous study showed transformation of 10 gene copies of *Sorghum bicolor* 22-kD *α-kafirins* (homologues to the maize 22-kD *α-zeins*) into maize increased the α22 fraction, but not total seed protein[27]. These observations suggest maize seed protein content is less dependent on zein gene expression than other factors associated with protein synthesis.

Thp9 **encodes asparagine synthetase**

We assembled a high-quality teosinte haplotype genomic sequence that helped

us identify genes responsible for QTLs associated with its high seed protein phenotype[28]. Significant QTLs were found on chromosomes 1, 3, 5 and 9 (Fig. 2b), but we focused on *Thp9* due to it having the greatest impact and being the highest peak revealed by BSA DNA sequencing. *Thp9* encodes an asparagine synthetase 4 (ASN4) gene that is highly expressed in teosinte roots and leaves, but it is not expressed in these tissues in B73 (Fig. 2d), which probably leads to differences in N assimilation.

ASN is as an important enzyme in the N metabolism[29] and it plays a key role in the N response network[30]. Previous research on *Arabidopsis* [31-33], rice [14,34,35], wheat [36,37] and barley [38] showed changes in *ASN* gene expression alter plant growth and N 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 N content in seeds, floral organs, leaves, and plants [31-33,39]. In rice (*Oryza sativa*), studies on *OsASN1* confirmed the importance of asparagine on plant N and grain protein content [14]. The increase in ASN activity leads to enhanced N assimilation, resulting in more asparagine transported to the seed for protein synthesis[29,40].

In B73, there are four *ASN* genes: *ZmASN1-ZmASN4* (*Zm00001d045675*, *Zm00001d044608*, *Zm00001d028750* and *Zm00001d047736*)[41]. *ZmASN1* appears to be expressed in all maize tissues, including the root, stem, leaf, endosperm, and embryo. *ZmASN2* is mainly expressed in the endosperm and embryo based on public RNA-seq data [42]. *ZmASN3* (on chromosome 1) and *ZmASN4* (on chromosome 9) could have resulted from an ancestral gene duplication[43]. These two genes are functional in Ames21814 and could have an additive effect in asparagine synthesis.

ZmASN3 was annotated as an intact gene in the B73 genome, but it is expressed at low levels in leaves, cobs and silks [42]. The four maize ASN genes could have a redundant function for asparagine synthesis, but the absence of *ZmASN4* activity leads to asparagine insufficiency in the plant. When *Thp9-T* was introgressed into B73, the asparagine content in roots and the N content of the entire plant were significantly enhanced (Fig. 2g; Extended Data 11e), supporting the importance of *Thp9-T* in NUE and increasing seed protein content.

Our data suggest *Thp9-B* is a null allele, as the ZmASN4 protein is missing in B73 (Fig. 2e). Public data showed *Thp9-B* gives rise to three mRNA isoforms, of which the functional one was undetectable in our assays and the other two are defective because the 48-bp deletion in the 10th intron leads to mis-splicing, which creates a premature stop codon in *Thp9-B* transcripts. Based on prediction via Pfam searches

(http://pfam.xfam.org/search), the *Thp9-T* allele produces a protein of 588 amino acids, whereas the barely detectable isoform 3 of *Thp9-B* can only be translated into a truncated protein with 480 amino acids.

Amino acids are essential substrates for protein synthesis, and their level in the plant is influenced by soil N availability and the NUE of the plant[44]. During amino acid synthesis, asparagine plays a primary role in N recycling, and it acts as a N donor for multiple aminotransferases[45]. Due to its high nitrogen-to-carbon ratio and inert nature, free asparagine is an important carrier for N storage and long-distance transport in the plant[29]. ASN, which is responsible for transferring amide groups from glutamine to aspartate and forming asparagine, determines N assimilation, remobilization, and allocation in the plant[31,39].

With limited soil N, the amino acid supply for protein synthesis can be increased by greater NUE[40]. NUE is determined by multiple processes, namely N uptake, transport, assimilation, and remobilization, of which N assimilation has been actively studied[46]. Looking to the future there is economic and environmental pressure to maintain high yielding maize while reducing the level of N applied to the soil. Therefore, it's important to identify genetic factors that increase NUE. Several studies have shown an association between QTL and N metabolism-related enzymes[47-52]; however, genes responsible for QTL controlling N assimilation has not been cloned.

Potential use of *Thp9-T* for genetic improvement

Thp9-T manifested a stable phenotypic effect at different geographic locations and under different levels of N, which is essential for its practical application. However, we found 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 significant contributions to the high protein phenotype of teosinte[28] (Fig. 2b); 2) Seed protein content is affected by N distribution 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. Seeds of modern inbred lines are larger than those of teosinte, and there are more seeds per ear and per plant; consequently, a lower concentration of amino acids allocated from the source to the seed could limit protein synthesis. The protein content of hybrids with the introgressed *Thp9-T* allele was lower than in NILTeo (Fig. 2f; Fig. 5c and h), supporting this hypothesis.

Since the most of the increased protein in teosinte and NILTeo is α-zein, which is

essentially devoid of the essential amino acid lysine, the increased protein content has limited nutritional value for monogastric animals, which includes humans. However, *Thp9-T* can be introgressed into Quality Protein Maize (QPM), which contains less zein and more non-zein proteins due to the *o2* mutation, and create high-protein, high-lysine hybrids.

The IHP genotype implies some, if not all, of the ancestral high-protein QTLs were retained in domesticated maize populations. Nevertheless, seed protein content declined during modern maize breeding[53]. Since *Thp9-T* is superior to *Thp9-B* for protein synthesis and has no apparent negative effect on yield, why wasn't *Thp9-T* retained in elite maize germplasm? One possible explanation is that *Thp9* was not under selection pressure due to ample application of N fertilizer. This could have become a vicious cycle, with low NUE of *Thp9-B* requiring more N fertilizer to ensure yield and protein content.

NUE has important environmental and economic implications for global food security, and research to understand NUE is important if we are to maintain high yield and high protein quality with low N input[54]. Several genes/QTLs affecting rice NUE, including *NRT1.1B*, *OsTCP19*, *GRF4* and *NGR5*, have been cloned[55]. Superior alleles of these genes/QTLs offer potential to achieve high stable rice yields with low N application. Root N sensing was found to be affected by multiple external factors, and there is a novel strategy to increase N acquisition efficiency under varying N conditions for crop production[46,56].

Our research demonstrates the potential value of hybrids containing the *Thp9-T* allele. They perform well in a N poor environment, maintaining a normal yield with reduced N application. Additional research on NUE based on the high-quality teosinte genome sequence could lead to other QTLs that improve modern hybrids. The tremendous structural variation between the genomes of *Zea mays* ssp. *parviglumis*, Ames21814 and B73 will also be valuable for investigating genes potentially responsible for phenotypic modifications that occurred during teosinte domestication.

References

1 Rosental, L., Nonogaki, H. & Fait, A. Activation and regulation of primary metabolism during seed germination. *Seed science research* **24**, 1-15 (2014).

2 Han, C., Zhen, S., Zhu, G., Bian, Y. & Yan, Y. Comparative metabolome analysis of wheat embryo and endosperm reveals the dynamic changes of metabolites during seed germination. *Plant Physiol Biochem* **115**, 320-327, doi:10.1016/j.plaphy.2017.04.013 (2017).

3 De Lumen, B. O. Molecular approaches to improving the nutritional and functional properties of plant seeds as food sources: developments and comments. *Journal of Agricultural and Food Chemistry* **38**, 1779-1788 (1990).

4 Palacios‐Rojas, N. *et al.* Mining maize diversity and improving its nutritional aspects within agro‐food systems. *Comprehensive Reviews in Food Science and Food Safety* **19**, 1809-1834 (2020).

5 Doebley, J. The genetics of maize evolution. *Annu Rev Genet* **38**, 37-59, doi:10.1146/annurev.genet.38.072902.092425 (2004).

6 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. *Theoretical and Applied Genetics* **119**, 1129-1142 (2009).

7 Matsuoka, Y. *et al.* A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences* **99**, 6080-6084 (2002).

8 Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309-1321, doi:10.1016/j.cell.2006.12.006 (2006).

9 Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S. & Buckler, E. S. Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences* **99**, 12959-12962 (2002).

10 Flint-Garcia, S. A., Bodnar, A. L., Scott, M. P. J. T. & Genetics, A. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theoretical and Applied Genetics* **119**, 1129-1142 (2009).

11 Wani, S. H. *et al.* Nitrogen use efficiency (NUE): elucidated mechanisms, mapped genes and gene networks in maize (Zea mays L.). *Physiol Mol Biol Plants* **27**, 2875-2891, doi:10.1007/s12298-021-01113-z (2021).

12 Ciampitti, I. A. & Lemaire, G. From use efficiency to effective use of nitrogen: A dilemma for maize breeding improvement. *Sci Total Environ* **826**, 154125, doi:10.1016/j.scitotenv.2022.154125 (2022).

13 Day, L. Proteins from land plants–potential resources for human nutrition and food security. *Trends in Food Science & Technology* **32**, 25-42 (2013).

14 Lee, S. *et al.* OsASN1 Overexpression in Rice Increases Grain Protein Content and Yield under Nitrogen-Limiting Conditions. *Plant Cell Physiol* **61**, 1309-1320, doi:10.1093/pcp/pcaa060 (2020).

15 Wu, Y., and Messing, J. (2017). Understanding and improving protein traits in maize. In Achieving sustainable cultivation of maize Vol 1: From improved varieties to local applications, D. Watson, ed (Cambridge, UK Burleigh Dodds Science Publishing.

16 Esen, A. (1987). A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (Zea mays L.). Journal of Cereal Science 5, 117-128.

17 Thompson, G., and Larkins, B. (1994). Characterization of Zein Genes and Their Regulation in Maize Endosperm. In The Maize Handbook, M. Freeling and V. Walbot, eds (Springer New York), pp. 639-647.

18 Dong, J. *et al.* Analysis of tandem gene copies in maize chromosomal regions reconstructed from long sequence reads. *Proceedings of the National Academy of Sciences* **113**, 7949-7956 (2016).

19 Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655-662, doi:10.1126/science.abg5289 (2021).

20 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175, doi:10.1038/s41592-020-01056-5 (2021).

21 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

22 Carvalho, R. F., Aguiar-Perecin, M. L. R., Clarindo, W. R., Fristche-Neto, R. & Mondin, M. A Heterochromatic Knob Reducing the Flowering Time in Maize. *Front Genet* **12**, 799681, doi:10.3389/fgene.2021.799681 (2021).

23 Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res* **46**, e126, doi:10.1093/nar/gky730 (2018).

24 Moose, S. P., Dudley, J. W. & Rocheford, T. R. Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends in plant science* **9**, 358-364, doi:10.1016/j.tplants.2004.05.005 (2004).

25 Goldman, I., Rocheford, T. & Dudley, J. Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. *Theoretical and Applied Genetics* **87**, 217-224 (1993).

26 Larkins B.A. Wu Y. Song R.Messing J. Maize seed storage proteins. in: Larkins B.A. Maize Kernel Development. CABI, Oxfordshire)2017: 175-189

27 Wu, Y. & Messing, J. RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. *Plant physiology* **153**, 337-347, doi:10.1104/pp.110.154690 (2010).

28 Karn, A., Gillman, J. D. & Flint-Garcia, S. A. Genetic analysis of teosinte alleles for kernel composition traits in maize. *G3: Genes, Genomes, Genetics* **7**, 1157-1164 (2017).

29 Lea, P. J., Sodek, L., Parry, M. A., Shewry, P. R. & Halford, N. G. J. A. o. A. B. Asparagine in plants. *Annals of Applied Biology* **150**, 1-26 (2007).

30 Jiang, L. *et al.* Analysis of Gene Regulatory Networks of Maize in Response to Nitrogen. *Genes (Basel)* **9**, doi:10.3390/genes9030151 (2018).

31 Gaufichon, L. *et al.* Arabidopsis thaliana ASN2 encoding asparagine synthetase is involved in the control of nitrogen assimilation and export during vegetative growth. *Plant, Cell & Environment* **36**, 328-342 (2013).

32 Gaufichon, L. *et al.* ASN1-encoded asparagine synthetase in floral organs contributes to nitrogen filling in Arabidopsis seeds. *Plant J* **91**, 371-393, doi:10.1111/tpj.13567 (2017).

33 Lam, H.-M. *et al.* Overexpression of the ASN1 gene enhances nitrogen status in seeds of Arabidopsis. *Plant physiology* **132**, 926-935 (2003).

34 Luo, L. *et al.* OsASN1 Plays a Critical Role in Asparagine-Dependent Rice Development. *Int J Mol Sci* **20**, doi:10.3390/ijms20010130 (2018).

35 Ohashi, M. *et al.* Asparagine synthetase1, but not asparagine synthetase2, is responsible for the biosynthesis of asparagine following the supply of ammonium to rice roots. *Plant Cell Physiol* **56**, 769-778, doi:10.1093/pcp/pcv005 (2015).

36 Curtis, T. Y., Bo, V., Tucker, A. & Halford, N. G. Construction of a network describing asparagine metabolism in plants and its application to the identification of genes affecting asparagine metabolism in wheat under drought and nutritional stress. *Food Energy Secur* **7**, e00126, doi:10.1002/fes3.126 (2018).

37 Raffan, S. *et al.* Wheat with greatly reduced accumulation of free asparagine in the grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene TaASN2. *Plant Biotechnol J* **19**, 1602-1613, doi:10.1111/pbi.13573 (2021).

38 Avila-Ospina, L., Marmagne, A., Talbotec, J., Krupinska, K. & Masclaux-Daubresse, C. The identification of new cytosolic glutamine synthetase and asparagine synthetase genes in barley (Hordeum vulgare L.), and their expression during leaf senescence. *J Exp Bot* **66**, 2013-2026, doi:10.1093/jxb/erv003 (2015).

39 Moison, M. *et al.* Three cytosolic glutamine synthetase isoforms localized in different-order veins act together for N remobilization and seed filling in Arabidopsis. *Journal of experimental botany* **69**, 4379-4393 (2018).

40 Seebauer, J. R., Moose, S. P., Fabbri, B. J., Crossland, L. D. & Below, F. E. Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling. *Plant Physiol* **136**, 4326-4334, doi:10.1104/pp.104.043778 (2004).

41 Todd, J. *et al.* Identification and characterization of four distinct asparagine synthetase (AsnS) genes in maize (Zea mays L.). **175**, 799-808 (2008).

42 Chen, J. *et al.* Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant physiology* **166**, 252-264, doi:10.1104/pp.114.240689 (2014).

43 Raffan, S. & Halford, N. G. Cereal asparagine synthetase genes. *Ann Appl Biol* **178**, 6-22, doi:10.1111/aab.12632 (2021).

44 The, S. V., Snyder, R. & Tegeder, M. Targeting nitrogen metabolism and transport processes to improve plant nitrogen use efficiency. *Frontiers in Plant Science* **11**, 628366 (2021).

45 Sieciechowicz, K. A., Joy, K. W. & Ireland, R. J. The metabolism of asparagine in plants. *Phytochemistry* **27**, 663-671 (1988).

46 Liu, X., Hu, B. & Chu, C. Nitrogen assimilation in plants: current status and future prospects. *J Genet Genomics*, doi:10.1016/j.jgg.2021.12.006 (2021).

47 Silva, I. T. *et al.* Biochemical and genetic analyses of N metabolism in maize testcross seedlings: 2. Roots. *Theor Appl Genet* **131**, 1191-1205, doi:10.1007/s00122-018-3071-0 (2018).

48 Gallais, A. & Hirel, B. An approach to the genetics of nitrogen use efficiency in maize. *J Exp Bot* **55**, 295-306, doi:10.1093/jxb/erh006 (2004).

49 Hirel, B. *et al.* Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. *Plant Physiol* **125**, 1258-1270, doi:10.1104/pp.125.3.1258 (2001).

50 Zhang, N. *et al.* Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the maize IBM mapping population. *Plant Physiol* **154**, 1753-1765, doi:10.1104/pp.110.165787 (2010).

51 Zhang, N. *et al.* Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol* **168**, 575-583, doi:10.1104/pp.15.00025 (2015).

52 Trucillo Silva, I. *et al.* Biochemical and genetic analyses of N metabolism in maize testcross seedlings: 1. Leaves. *Theor Appl Genet* **130**, 1453-1466, doi:10.1007/s00122-017-2900-x (2017).

53 Duvick, D. N. in *Advances in Agronomy* Vol. 86 83-145 (Academic Press, 2005).

54 Liu, Q. *et al.* Improving Crop Nitrogen Use Efficiency Toward Sustainable Green Revolution. *Annu Rev Plant Biol* **73**, 523-551, doi:10.1146/annurev-arplant-070121-015752 (2022).

55 Hou, M., Yu, M., Li, Z., Ai, Z. & Chen, J. Molecular Regulatory Networks for Improving Nitrogen Use Efficiency in Rice. *Int J Mol Sci* **22**, doi:10.3390/ijms22169040 (2021).

56 Xuan, W., Beeckman, T. & Xu, G. Plant nitrogen nutrition: sensing and signaling. *Curr Opin Plant Biol* **39**, 57-65, doi:10.1016/j.pbi.2017.05.010 (2017).

57 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**, lqaa108, doi:10.1093/nargab/lqaa108 (2021).

58 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

59 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. *The Plant Cell* **27**, 1162-1172 (2015).

60 Liu, H. *et al.* Gene duplication confers enhanced expression of 27-kDa γ-zein for endosperm modification in quality protein maize. *Proceedings of the National Academy of Sciences* **113**, 4964-4969 (2016).

61 Bukowski, R. *et al.* Construction of the third-generation Zea mays haplotype map. *Gigascience* **7**, 1-12, doi:10.1093/gigascience/gix134 (2018).

62 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824, doi:10.1038/ng.2310 (2012).

63 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

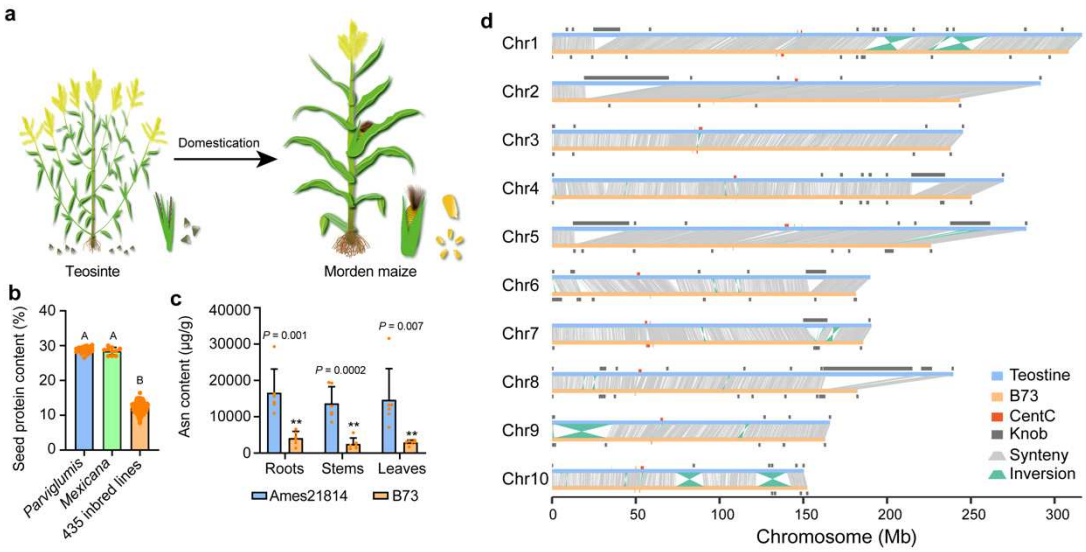
64 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

65 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

66 Mansfeld, B. N. & Grumet, R. QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. *Plant Genome* **11**, doi:10.3835/plantgenome2018.01.0006 (2018).

678 **Figure legends**

679



680

681 **Fig1. Variation in the protein content and genome sequence of teosinte and modern**
682 **maize. a**, Schematic diagram of teosinte and modern maize plant. **b**, Comparison of seed
683 protein content of teosintes (*Zea mays* ssp. *parviglumis* and *mexiana*) and modern inbreds.
684 **c**, Free asparagine content in roots, stems and leaves of Ames21814 and B73. Data are
685 mean ±s.d (n= 6 independent biologically samples). **d**, Synteny between the B73 reference
686 genome (B73_v5, orange) and the Ames21814 haplotype (blue). Grey lines represent
687 synteny blocks, green lines represent inversions larger than 5 kb. Knob density of 10-kb
688 windows is marked with black rectangles above the chromosomes, and CentC density is
689 represented by red rectangles along the chromosomes. In **b**, different letters indicate
690 significant differences ($P < 0.01$, one-way ANOVA and further Tukey's test); In **c**, P values,
691 see Source Data.

692

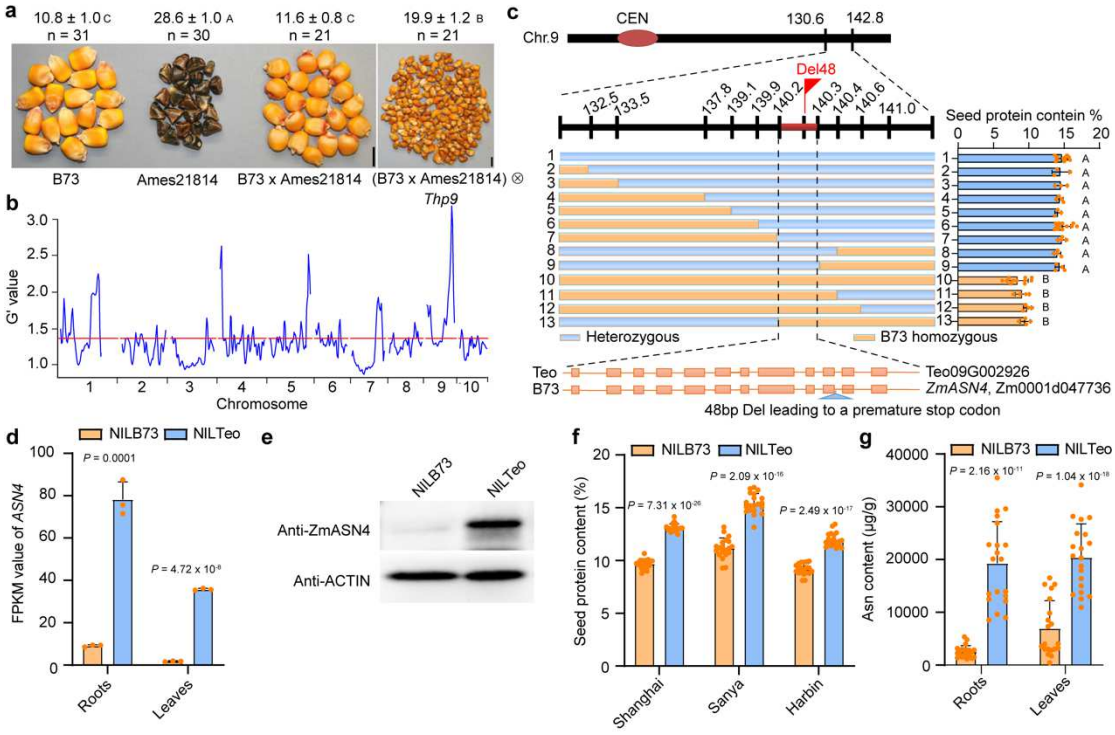


Fig2. Map-based cloning and expression of *Thp9*. **a**, Phenotypes of B73, teosinte (Ames21814), and B73 x teosinte and self-pollinated B73 x teosinte seeds. The protein content and number are indicated above the corresponding seeds. Scale bar, 1 cm. **b**, G'value of QTLs from the BSA of the F_1BC_4 population. **c**, Map-based cloning of *Thp9*. *Thp9* was located in a 150-kb interval between markers 140.2 and 140.3 (based on B73_v4), where *Zm00001d047736*, corresponding to the teosinte gene *Teo09G002926*, is contained. **d**, Analysis of FPKM value of *Thp9* in NILTeo and NILB73. Data are mean ±s.d (n= 3 biologically independent samples). **e**, Immunoblot analysis reflecting the protein level of ASN4 in roots of NILTeo and NILB73. **f**, The protein content in NILTeo and NILB73 seeds harvested in Shanghai, Sanya, and Harbin. Data are mean ±s.d (n= 20 biologically independent samples). **g**, Free asparagine content in NILTeo and NILB73 roots and leaves. Data are mean ±s.d (n= 20 biologically independent samples). In **a**, different letters indicate significant differences ($P < 0.01$, one-way ANOVA and further Tukey's test); In **d, f and g**, *P* values, see Source Data.

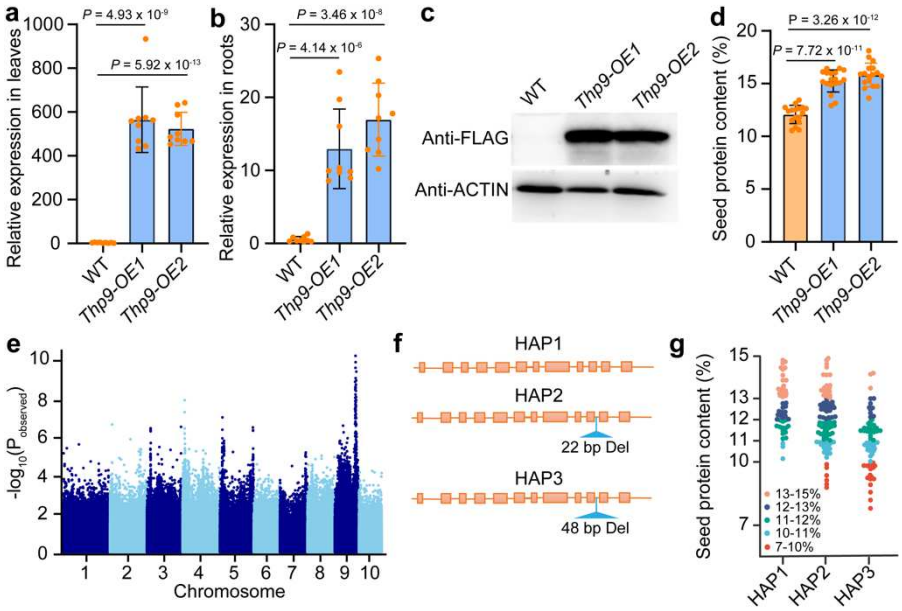


Fig3. Genetic confirmation and natural variation of *Thp9*. a and **b**, RT-qPCR analysis of *Thp9* expression in leaves (**a**) and roots (**b**) of *Thp9-OE1* and *Thp9-OE2*. Expression levels were normalized to that of *ZmActin*. Data are presented as mean values ± SD (n = 9 biologically independent samples). **c**, Immunoblot analysis showing the protein level of ASN4 in leaves of *Thp9-OE1* and *Thp9-OE2*. ACTIN was used as an internal control. **d**, Protein content of *Thp9-OE1* and *Thp9-OE2* seeds. Data are presented as mean values ± SD (n = 17 biologically independent samples). **e**, GWAS analysis of seed protein content in 405 and 438 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 maize population. HAP1, HAP2, and HAP3 identify inbred types containing the three different *Thp9* alleles. **g**, The protein content in HAP1, HAP2 and HAP3 seeds. Data are mean±s.d. (n = more than 50 biologically independent samples). In **a, b, and d**, *P* values, see Source Data.

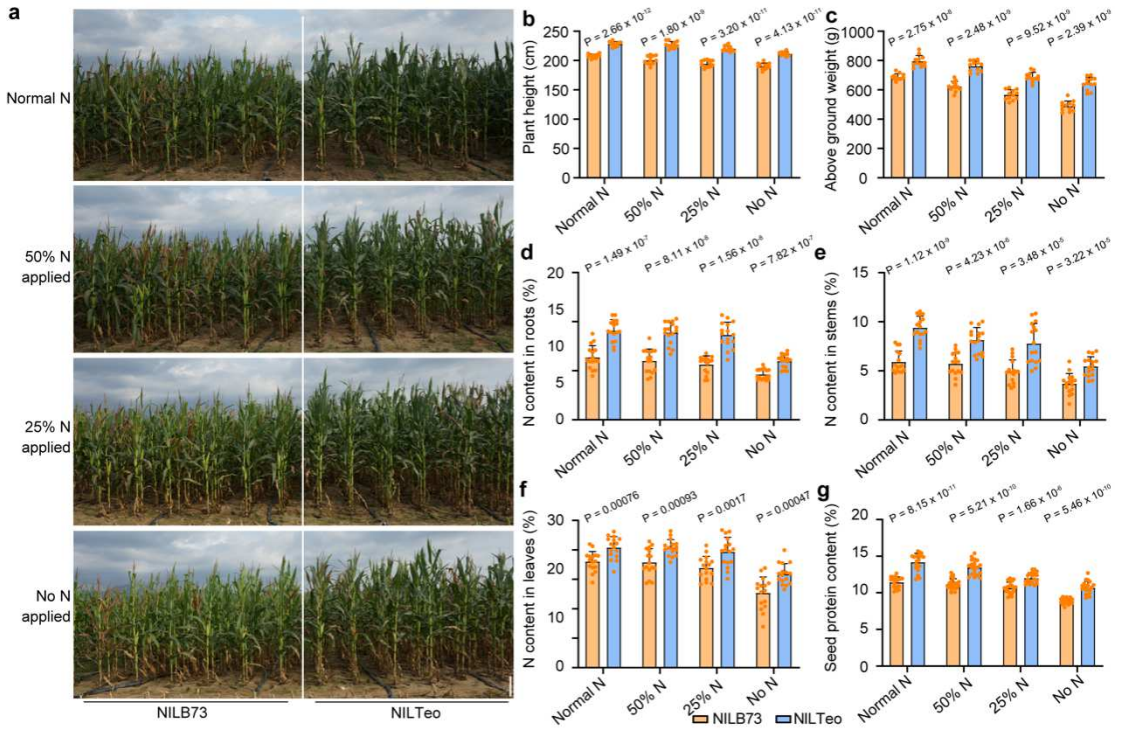


Fig4. NUE of NILTeo and NILB73. a, NILTeo and NILB73 plants grown with decreasing amounts of N, i.e. 100% (32 g/plant), 50%, 25% and 0%. In each trial, NILB73 is on the left and NILTeo on the right. Scal bar, 40 cm. **b**, Measurement of NILB73 and NILTeo plant height when grown under varying N conditions. Data are presented as mean values ± SD (n = 12 biologically independent samples). **c**, Measurement of above-ground biomass of NILB73 and NILTeo grown under different N conditions. Data are presented as mean values ± SD (n = 12 biologically independent samples). **d**, Total N content in NILB73 and NILTeo roots grown under different N conditions. Data are presented as mean values ± SD (n = 16 biologically independent samples). **e**, Total N content in NILB73 and NILTeo stems grown under different N conditions. Data are presented as mean values ± SD (n = 16 biologically independent samples). **F**, Total N content in NILB73 and NILTeo leaves grown under different N conditions. Data are presented as mean values ± SD (n = 16 biologically independent samples). **g,** Protein content of NILB73 and NILTeo seeds grown under different N conditions. Data are presented as mean values ± SD (n = 20 biologically independent samples). In **b-g**, *P* values, see Source Data.

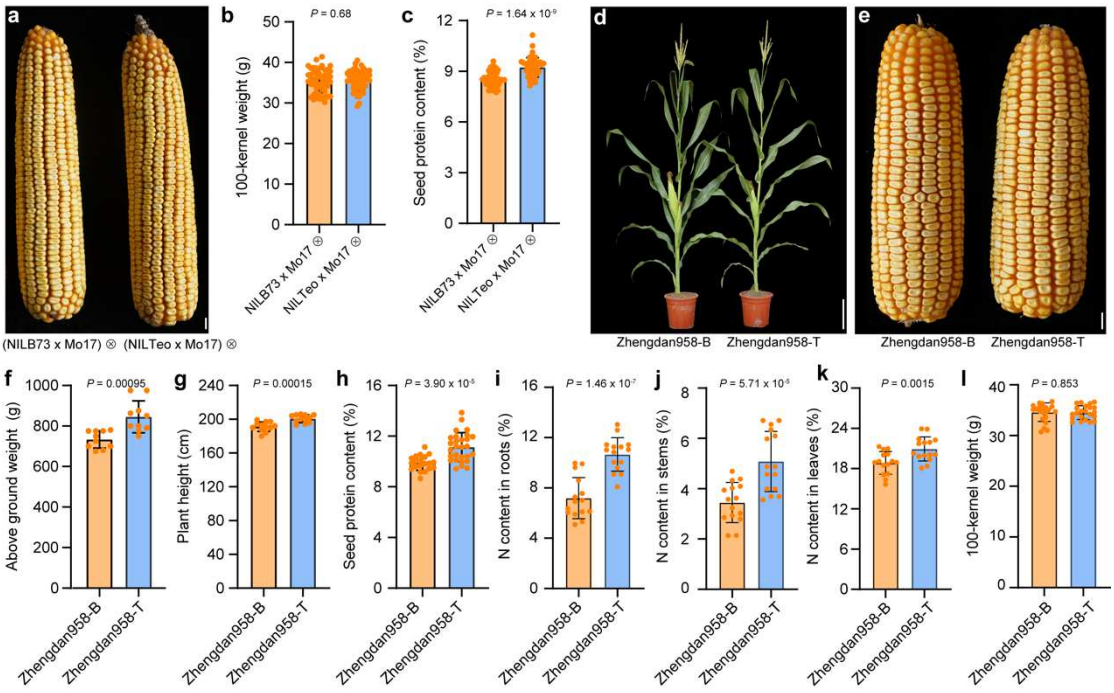
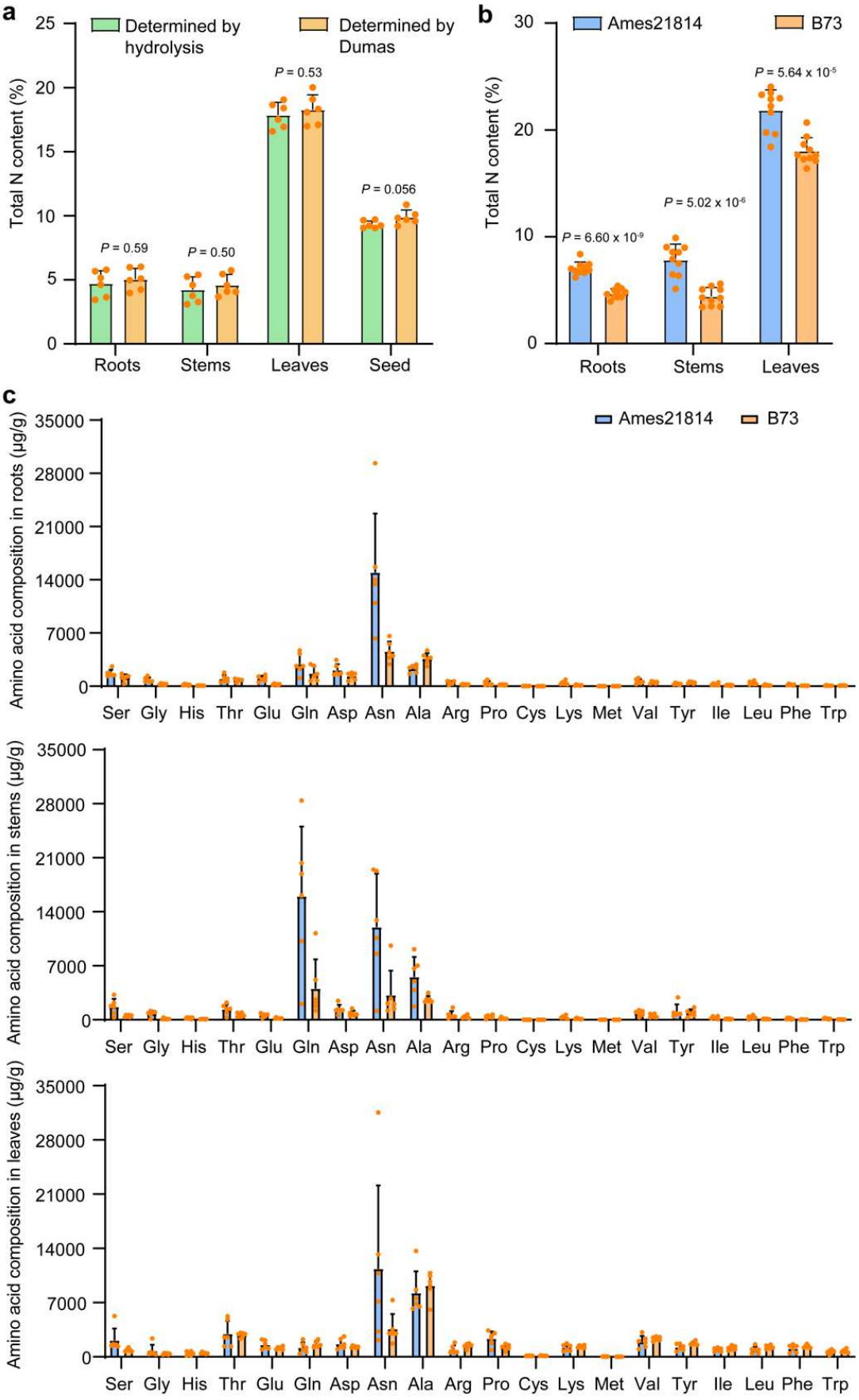
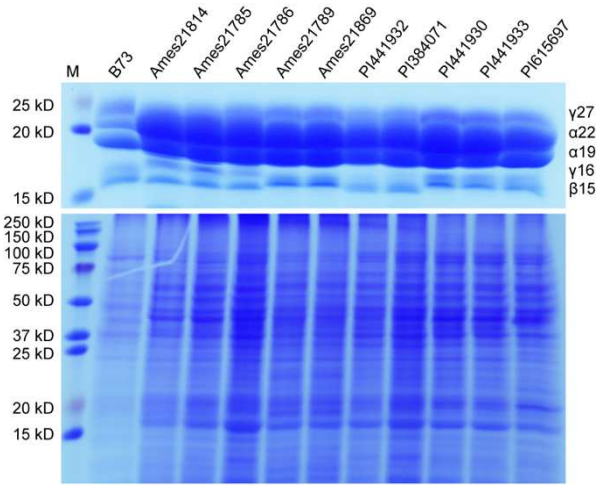


Fig5. Improved N content in hybrids containing *Thp9-T*. **a**, F_2 ears of NILB73 x Mo17 and NILTeo x Mo17. Scale bar, 1 cm. **b**, 100-kernel weight of F_2 seeds of NILB73 x Mo17 and NILTeo x Mo17. **c**, Protein content in F_2 seeds of NILB73 x Mo17 and NILTeo x Mo17. **d**, Plant phenotype of Zhengdan958-B and Zhengdan958-T. Scale bar, 30 cm. **e**, Ear phenotype of Zhengdan958-B and Zhengdan958-T. Scale bar, 1 cm. **f**, Above ground weight of Zhengdan958-B and Zhengdan958-T. Data are presented as mean values ± SD (n = 10 biologically independent samples). **g**, Plant height of Zhengdan958-B and Zhengdan958-T. Data are presented as mean values ± SD (n = 10 biologically independent samples). **h**, Protein content in Zhengdan958-B and Zhengdan958-T seeds. Data are presented as mean values ± SD (n = more than 20 biologically independent samples). **i**, The total N content in Zhengdan958-B and Zhengdan958-T roots. Data are presented as mean values ± SD (n = 15 biologically independent samples). **j**, Total N content in Zhengdan958-B and Zhengdan958-T stems. Data are presented as mean values ± SD (n = 15 biologically independent samples). **k**, Total N content in Zhengdan958-B and Zhengdan958-T leaves. Data are presented as mean values ± SD (n = 15 biologically independent samples). **l**, 100-kernel weight of Zhengdan958-B and Zhengdan958-T seeds. Data are presented as mean values ± SD (n = 22 biologically independent samples). In **b-c and f-l**, *P* values, see Source Data.

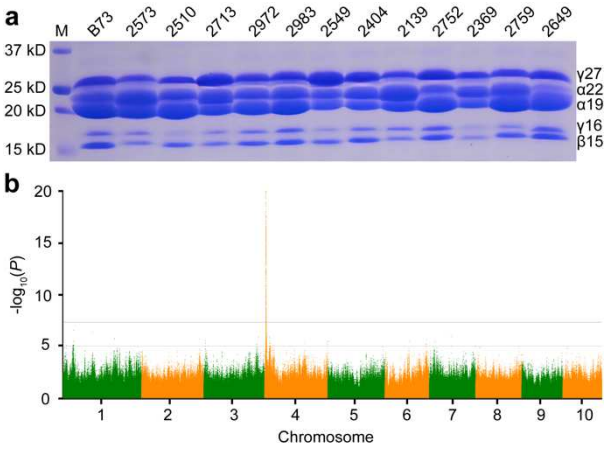
781 **Extended data legends**



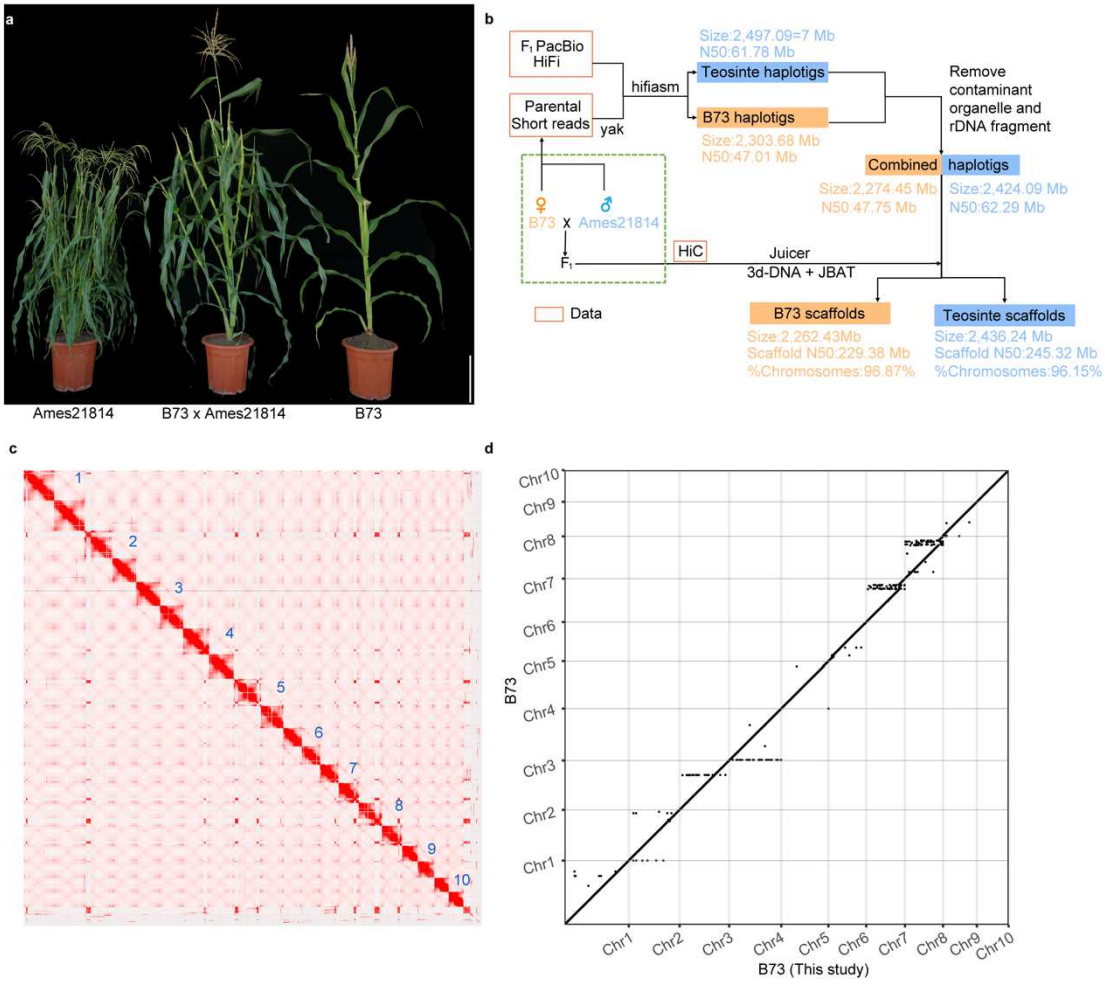
Extended Data Fig.1 |Content of total N and free amino acids in roots, stems and leaves of Ames21814 and B73. a, N content in roots, stems, leaves and seeds of B73 determined by acid hydrolysis and the Dumas method. Data are presented as mean values ± SD (n = 6 biologically independent samples for each method). **b**, Total N content in roots, stems, and leaves of Ames21814 and B73. Data are presented as mean values ± SD (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 Ames21814 and B73. Data are presented as mean values ± SD (n = 6 biologically independent samples). In **a** and **b**, *P* values, see Source Data.



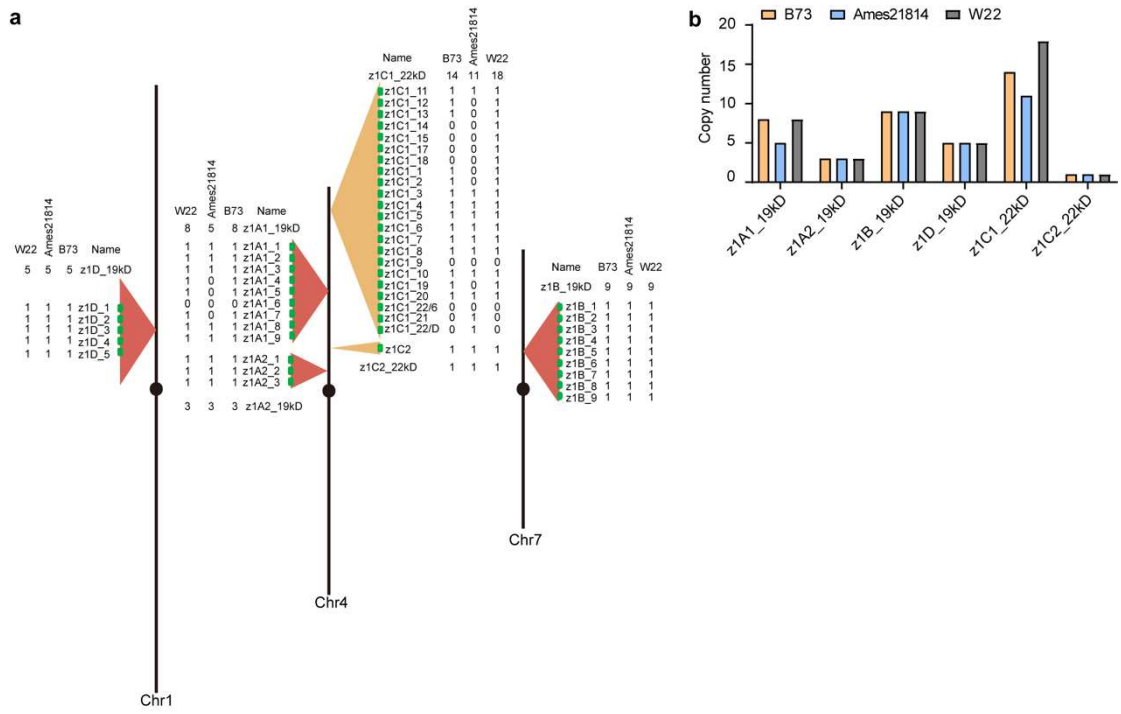
Extended Data Fig.2| SDS-PAGE of zein and non-zein proteins in B73 and 10 teosinte lines. The apparent size in kD of each protein band is indicated on the left. M, protein markers. γ27, 27-kD γ-zein; α22, 22-kD α-zein; α19, 19-kD α-zein; γ16, 16-kD γ-zein; γ15, 15-kD γ-zein; δ10, 10-kD δ-zein.



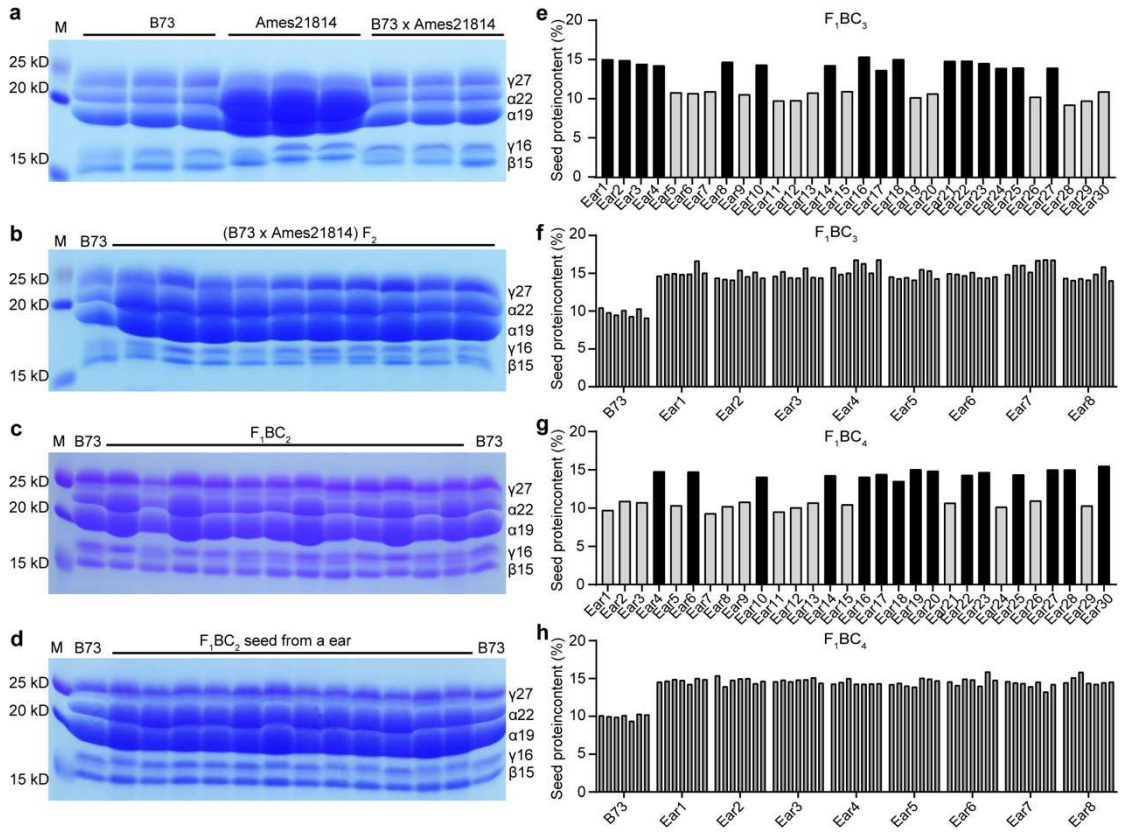
Extended Data Fig.3| GWAS analysis of loci controlling α-zein accumulation in maize populations. a, Representative SDS-PAGE analysis of α-zein accumulation patterns occurring in 512 inbred lines. **b**, GWAS identified a significant peak in the α-zein locus on the short arm of chromosome 4.



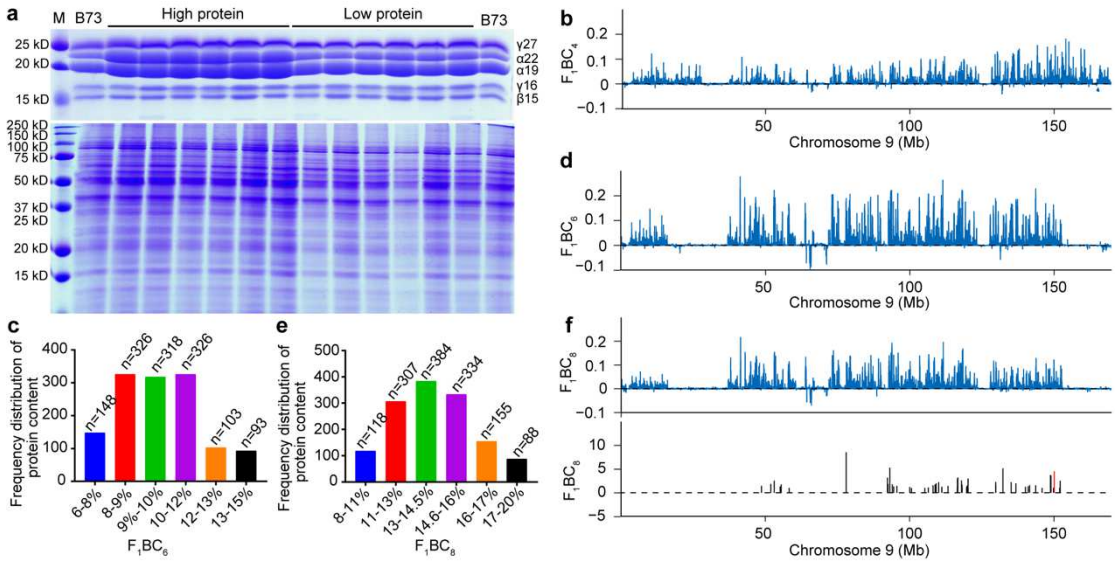
Extended Data Fig.4 |Ames21814 haplotype assembled by Trio-Binning. a, Phenotypes of teosinte *Z. mays* ssp. *Parviglumis* (accession number: Ames21814), B73 x Ames21814 and B73. Scale bar, 35 cm. **b**, Teosinte haplotype genome assembly flow chart. To perform a de novo assembly of the teosinte haplotype, 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 Ames21814. **c**, Whole genome HiC interaction heatmap of 2.5Mb windows. Each blue number indicates the corresponding chromosome. Each cluster represents a chromosome in the haplotype. In a set of chromosomes, the top cluster represents the hap1 (teosinte Ames21814) chromosome, the bottom cluster represents the hap2 (B73) chromosome. **d**, Dotplot of B73 genome assembly (hap2, this study) and B73_v5 genome assembly. Alignment less than 10 kb was filtered out.



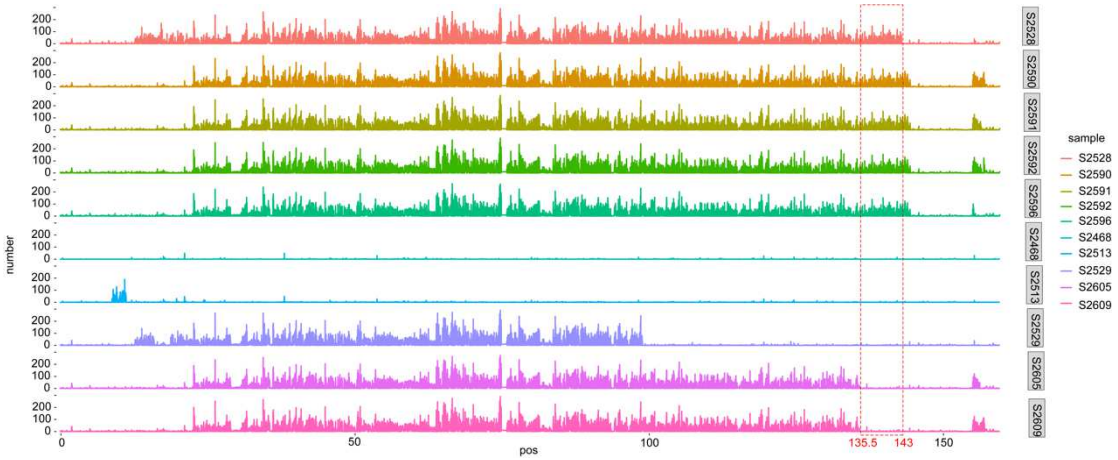
Extended Data Fig.5| Variation in *α-zein* gene copies between Ames21814 and maize inbreds. a, Copy numbers of *α-zein* genes in Ames21814, B73 and W22. Gene numbers of *α19* (*z1A1*, *z1A2*, *z1B* and *z1D*) and *α22* (*z1C1* and *z1C2*) are indicated beside each locus. **b**, Statistical analysis of copy number of *α19* (*z1A1*, *z1A2*, *z1B* and *z1D*) and *α22* (*z1C1* and *z1C2*) in Ames21814, B73 and W22.



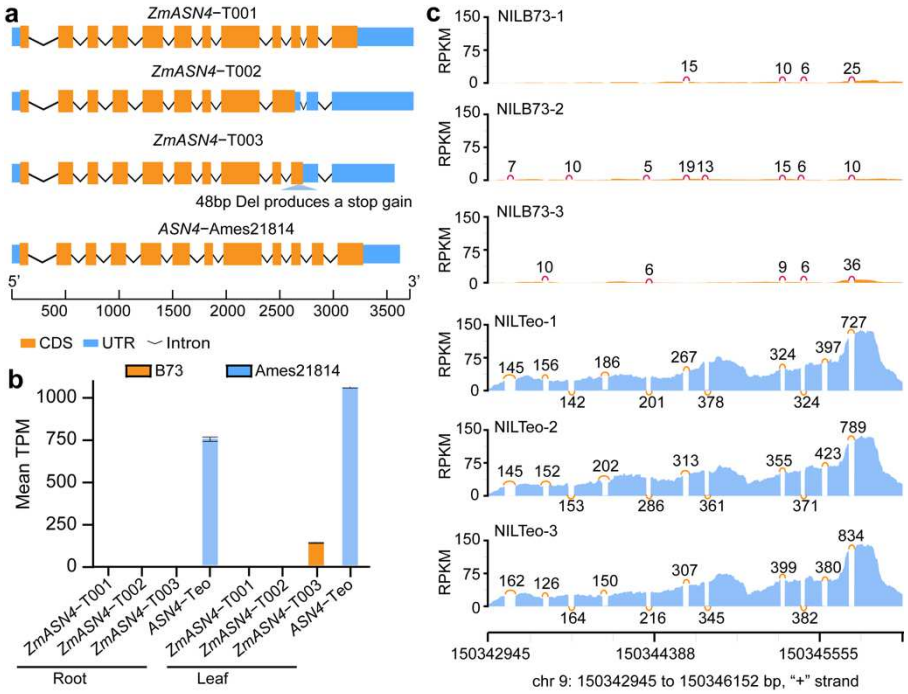
852
853
854 **Extended Data Fig.6 |Measurement of zein accumulation and protein content in**
855 **seeds of backcrossing populations. a**, SDS-PAGE showing zein accumulation in B73,
856 teosinte (Ames2184), and B73 x teosinte seeds. **b**, SDS-PAGE of zein accumulation in 10
857 F_2 seeds of B73 x teosinte, showing no segregation of α-zein accumulation. B73 seed was
858 used as the control. **c**, SDS-PAGE of zein accumulation in seeds from 12 different F_1BC_2
859 eras, showing a quantitative segregation pattern. B73 seed was used as the control. **d**,
860 SDS-PAGE of zein accumulations in 12 F_1BC_2 seeds from a single high-protein ear,
861 showing no segregation of α-zein accumulation. B73 seed was used as the control. **e**,
862 Protein content in seeds from 30 different F_1BC_3 ears. The protein content varied from 10%
863 to 15%. **f**, Protein content in 8 high-protein ears in the F_1BC_3 population. Seven single
864 seeds for each ear were measured. The protein content in each seed was ~15%. B73 was
865 used as the control. **g**, The protein content in seeds from 30 different F_1BC_4 ears. The
866 protein content varied from 10% to 15%. **h**, The protein content in 8 high-protein ears in
867 the F_1BC_4 population. Seven single seeds for each ear were measured. The protein
868 content in each seed was ~15%. B73 was used as the control.
869



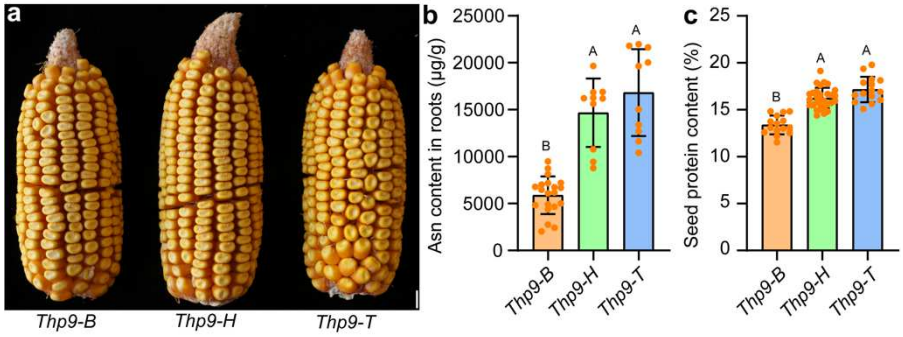
Extended Data Fig.7|Mapping *Thp9* by BSA sequencing of three populations of F_1BC_4, F_1BC_6 and F_1BC_8. a, Phenotyping high- and low-protein ears by SDS-PAGE of zein (upper panel) and non-zein protein (lower panel) accumulation in the F_1BC_4 population. Size of the zein protein band is indicated on the left. M, protein mol wt markers. γ27, 27-kD γ-zein; α22, 22-kD α-zein; α19, 19-kD α-zein; γ16, 16-kD γ-zein; γ15, 15-kD γ-zein; δ10, 10-kD δ-zein. **b**, Gene introgression analysis based on BSA sequencing of the F_1BC_4 population. Introgression of 315 teosinte genes was detected in the region between 130 Mb and160 Mb (based on Teo_v1) on chromosome 9, based on a threshold of 0.025. **c**, Frequency distribution analysis of seed protein content in the F_1BC_6 population. A group of 1,314 ears were phenotyped and classified. **d**, Gene introgression analysis based on BSA sequencing of the F_1BC_6 population. Introgression of 271 teosinte genes was detected in the region of 130 Mb-160 Mb (based on Teo_v1) on chromosome 9, based on a threshold of 0.025. **e**, Frequency distribution analysis of protein content in the F_1BC_8 population. A group of 1,386 ears was phenotyped and classified. **f**, Gene introgression analysis by BSA sequencing of the F_1BC_8 population. Introgression of 190 teosinte genes was detected in the region of 130 Mb-160 Mb on chromosome 9, based on a threshold of 0.025. Bottom is the differential expression of introgressed genes based on RNA-Seq analysis of F_1BC_8 leaves.



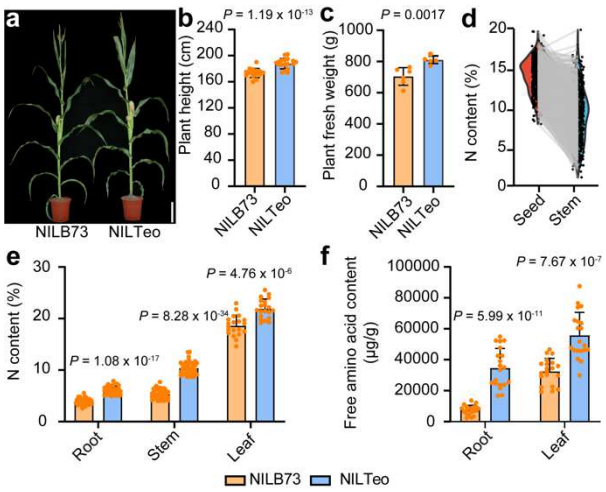
Extended Data Fig.8|Deep resequencing of 5 high-protein and 5 low-protein F_4BC_6 lines. A total of 10 lines were analyzed for gene introgression by resequencing. The high protein lines are S2528, S2590, S2591, S2592 and S2596, and the low-protein lines are S2468, S2513, S2529, S2605 and S2609. The peak 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.



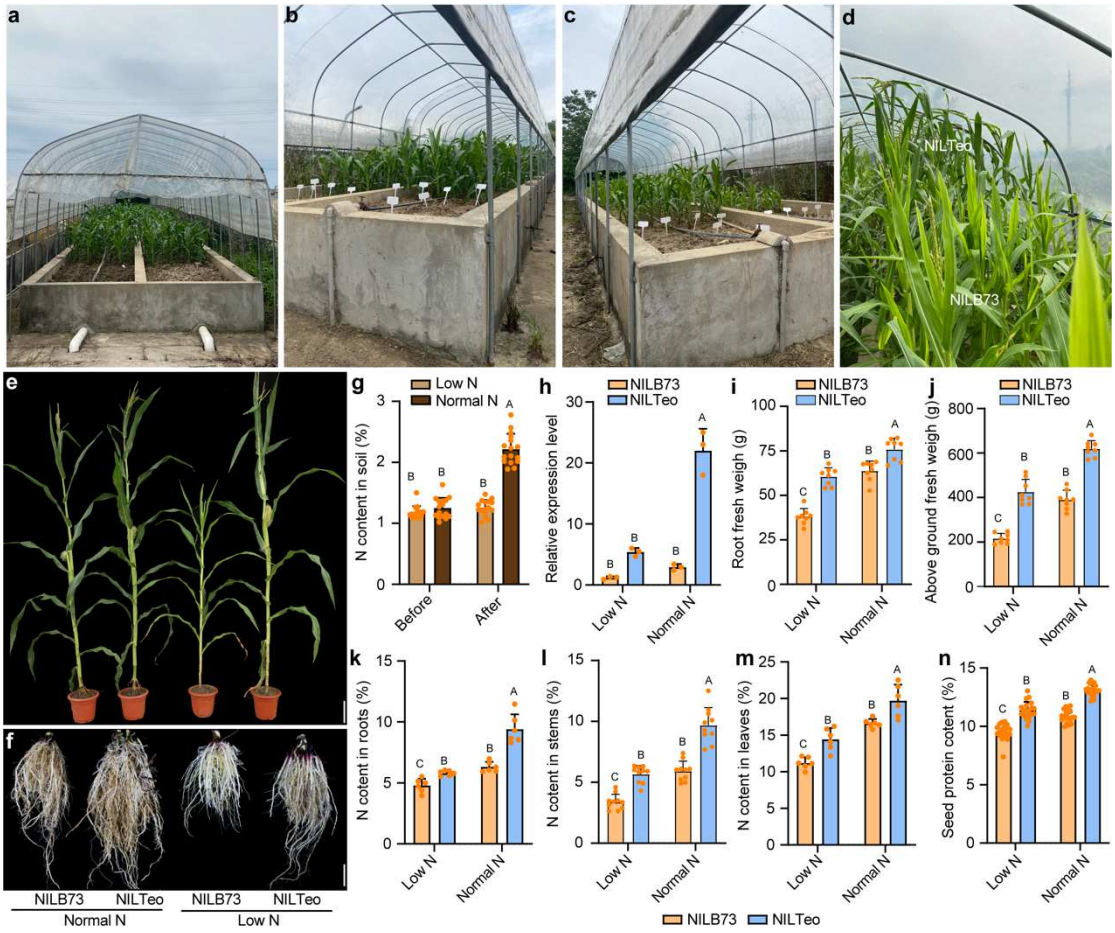
Extended Data Fig.9|*ASN4* transcript analysis. a, Schematic representation of *ASN4* transcripts in B73 and Ames21814. **b**,TPM mean analysis of *ASN4* transcripts in B73 and Ames21814 leaves and roots based on the RNA-seq data. TPM, Transcripts Per Million. **c**, RNA-seq reads of *Thp9* in NILB73 (upper panel) and NILTeo (lower panel) leaves. The number refers to the number of reads across the junction.



Extended Data Fig.10|Linkage of *Thp9-T* with high asparagine and high protein contents in the F_3BC_7 population. a, Phenotypes of three representative F_3BC_7 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 F_3BC_7 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 F_3BC_7 population. Each ear was used as one unit. Data are mean ±s.d. (n = more than 20 biologically independent samples). In **b-c**, different letters indicate significant differences ($P < 0.01$, one-way ANOVA and further Tukey's test).



Extended Data Fig.11|Phenotypic comparison of NILTeo and NILB73. a, Plant phenotypes of NILTeo and NILB73. Scale bar, 30 cm. **b**, Plant height of NILTeo and NILB73. The plants were grown in Sanya in 2021. Data are mean ±s.d (n = 18 biologically independent samples). **c**, Plant fresh weight (root and above-ground mass) of NILTeo and NILB73. Data are mean ±s.d (n = 6 biologically independent samples). **d**, Association analysis of seed protein content with total stem N content in the F_1BC_8 population. The corresponding seed protein content and stem N content of the same plant are connected by a solid grey line. (n = 1334 biologically independent samples). **e**, The total N content in NILTeo and NILB73 roots, stems, and leaves. Data are mean ±s.d. (n = 33, 41 and 20 biologically independent samples, respectively). **f**, 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 and e, f**, *P* values, see Source Data.



Extended Data Fig.12|NILB73 and NILTeo under normal and low N conditions. a-c, Construction of four above ground concrete containers with plastic film covering the containers used for NUE test. **d**, NILB73 and NILTeo grown in ground concrete containers without N fertilizer application. **e**, Plant phenotypes of NILB73 and NILTeo with and without N application. Scale bar, 30 cm. **f**, Root phenotypes of NILB73 and NILTeo with and without N application. Scale bar, 5 cm. **g**, The N content of soil in containers with and without N application. Data are mean ±s.d. (n = 16 biologically independent samples). **h**, RT-qPCR analysis of *Thp9* expression in NILB73 and NILTeo roots with and without N application. Data are mean ±s.d. (n = 3 biologically independent samples). **i**, The root fresh weight of NILB73 and NILTeo with and without N application. Data are mean ±s.d. (n = 8 biologically independent samples). **j**, The above ground biomass of NILB73 and NILTeo with and without N application. Data are mean ±s.d. (n = 8 biologically independent samples). **k.** Total N content in NILB73 and NILTeo roots with and without N application. Data are mean ±s.d. (n = 6 biologically independent samples). **l**, Total N content in NILB73 and NILTeo stems with and without N application. Data are mean ±s.d. (n = 10 biologically independent samples). **m**, Total N content in NILB73 and NILTeo leaves with and without N application. Data are mean ±s.d. (n = 6 biologically independent samples). **n**, Protein content in NILB73 and NILTeo seeds with and without N application. Data are mean ±s.d. (n = 20 biologically independent samples). In **g-n**, letters indicate significant differences ($P < 0.01$, one-way ANOVA and further Tukey's test).

Material and methods

Plant materials

We obtained 30 teosinte lines (20 lines of *Zea mays* ssp. *parviglumis* and 10 lines of *Zea mays* ssp. *Mexicana*) from Joachim Messing's lab at Rutgers University, USA. They were originally obtained from the North Central Regional Plant Introduction Station (NCRPIS), USA. The 518 inbred lines used for GWAS were obtained from Jinsheng Lai's lab at China Agricultural University. A teosinte line (*Zea mays* ssp. *parviglumis*, accession number: Ames21814) was used for genome sequencing, creation of mapping populations, and NILs. Maize genetic materials were grown in the experimental fields in Shanghai (30.5°N, 121.1°E), Harbin (44.0°N, 125.4°E) and Sanya (18.2°N, 109.3°E).

Ames21814 pollen was used to fertilize B73 ears, and the resulting F_1 used for pollen to backcross with B73, yielding the F_1BC_1. We chosed a single F_1BC_1 ear for planting. In the following generations, we used B73 pollen for backcrossing. Zein and non-zein protein accumulation patterns of 108 F_1BC_2 ears were characterized, and the protein accumulation pattern showed quantitative segregation. In a single ear, all seeds contained a uniformly high or low α-zein content . The F_1BC_2 seeds from ears with a high protein content were selected for planting. We created continuous backcrossing populations, yielding F_1BC_3 (n=500), F_1BC_4 (n=500), F_1BC_5 (n=1000), F_1BC_6 (n=1314), F_1BC_7 (n=1200), F_1BC_8 (n=1386) and F_1BC_9 (n=2000). In each generation, we measured the protein content ear by ear.

To obtain homozygous NILTeo and NILB73, 20 F_1BC_6 independent ears with a high protein content (about 15%) were planted as 20 groups. Thirty plants of each group were self-pollinated, yielding 600 F_2BC_6, ears, which formed 30 x 20 subgroups. The protein content of all subgroup ears was measured and 50 F_2BC_6 subgroup ears with a high protein content were planted. Twenty plants of each subgroup were self-pollinated, yielding 1,000 F_3BC_6 ears that were measured for protein content. If individual ears in a subgroup had uniformly high protein content, namely no segregation, they should be homozygous for the high protein locus and identified as NILTeo. In 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 F_4BC_6 NILTeo and five NILB73 individuals were selected for 20x resequencing. The linkage analysis was performed by genotyping 200 F_2BC_7 plants and measuring the protein content of the corresponding F_3BC_7 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 (Jiangsu) (http://www.wimibio.com/). The primer sequences used in this study are shown in Supplemental Table 6.

Resolving the Ames21814 haplotype with trio-binning

High-quality genomic DNA was extracted from fresh leaves of the F_1 crossed by B73

and Ames21814, followed by library construction according to the standard protocol of PacBio (Pacific Biosciences, CA, USA). DNA sequencing on the PacBio sequel II HiFi platform, which produces high-fidelity reads, was done by Shanghai OE Biotech Co., Ltd. In addition, we generated 50X Illumina PE 150 reads for the parental B73 and Ames21814 genomic DNA, respectively. We used yak (https://github.com/lh3/yak/issues/11) to separate parental- and maternal-derived HiFi reads with the guidance of Illumina reads. We applied the hifiasm (0.16.1) trio mode[20], a de novo assembler that could faithfully preserve the contiguity of all haplotypes, to assemble the haplotypes of Ames21814. Contaminants, such as organelle DNA or rDNA fragments, were removed by blastn. We mapped HiC reads to the assembly and scaffolded by 3d-dna with -r 0 -m haploid. False duplications and phase error were manually curated based on yak trioeval within Juicebox. Finally, we used KAT and yak to evaluate the completeness of the genome assembly.

Transposon elements were annotated by EDTA using the pan-genome TE database (https://github.com/HuffordLab/NAM-genomes/tree/master/te-annotation). Protein-coding genes were predicted using the MAKER2[57] with the homolog evidence of RNA-seq and protein databases. RNA was collected from six tissues (root, leaf, stem, seed, cob and tassel) of Ames21814 and aligned against the genome with HISAT2 (v2.10.2). Protein sequences were downloaded from SwissProt (Viridiplantae) (https://www.uniprot.org) and for six plant species (*Arabidopsis thaliana, Oryza sativa, Setaria italica, Sorghum bicolor, Triticum aestivum, Zea mays*) , which were integrated with CD-HIT (v4.6)[58] using the parameter '-c 0.99'. *Ab initio* gene prediction was performed using SNAP (version 2006-07-28), AUGUSTUS (v3.3.3) and GeneMark (v4.3.8). SNAP was trained using the first-round annotation, whereas AUGUSTUS and GeneMark were trained by RNA-seq and protein databases. The gene models with AED values less than 0.5 were retained.

Measurement of protein content with the Rapid N analyzer

To measure total N content in seeds and other tissues (root, stem, and leaf), the samples were first dried to constant weight at 65°C and then powdered using a grinder (60hz, 60s). A total of 50-70 mg of powder was wrapped in tin foil as the test sample. Determination of total N was carried out using the Dumas rapid nitrogen analyzer (rapid N exceed) from Elementar, Germany. Before each round of measurement, it was necessary to weigh about four standard asparagine 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, and the option (O2 dosing time: 60s; O2 dosing flow: 120ml/min; O2 cut off threshold: 15%; Autozero delay: 30s; Peak anticip: 90s) was selected as program settings. 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 was exported in Excel format for analysis.

Measurement of free amino acids

Roots, stems, and leaves of different genetic materials were analyzed to determine the content of free amino acids at the flowering stage. Plant materials were dried at 65°C to constant weight and ground. Thirty mg of powder was treated in 1-ml distilled water at 4°C for 8 h and then homogenized. The powder was hydrolyzed with 6 N 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 5,500 g for 5 min, 50-µl of supernatant was mixed with amino mixed standards (MSLAB50AA) and 50-µl 4% sulfosalicylic acid solution, and the mixture was centrifuged at 17,370 g at 4°C for 4 min. The supernatant was mixed with 50-µl borate buffer (0.1M, pH 8.8) and then derivatized with 20-µl 6-aminoquinoline-N-hydroxyl succinimide carbamate at 55°C for 15 min. After cooling and centrifuging at 4°C, 50-µl of supernatant was analyzed by UPLC (Ultra Performance Liquid Chromatography UPLC, Ultimate 3000)-MS/MS (Tandem mass spectrometry, API 3200 Q TRAP). Chromatographic separations were performed on an MSLab 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}. Conditions for MS-MS detection were as follows: positive-ion mode; ion spray voltage, 5500 V; nebulizer gas pressure, 55 psi; curtain gas pressure, 20 psi; collision gas pressure, medium; turbo gas temperature, 500 °C; entrance potential, 10 V; collision cell exit potential, 2 V. Nitrogen gas was used as the collision gas in a multiple reaction monitoring mode. The data were obtained using Analyst software version 1.5.1 (Applied Biosystems). Amino acid detection and data analysis were performed by Beijing Mass Spectrometry Medical Research Co., Ltd.

Extraction and SDS-PAGE analysis of zein and non-zein proteins

Endosperm was first dried to constant weight at 65°C and then ground to a fine powder in a tissue grinder. A total of 100-mg of flour was extracted with 1-ml of zein extraction buffer (3.75 mM sodium borate, 2% 2-mercaptoethanol [v/v], 0.3% SDS and 70% ethanol). After incubation for 2h or overnight, the mixture was centrifuged at 17,370 g for 10 min. One hundred µL of supernatant was transferred to a new tube and mixed with 10 µl of 10% SDS. The solution was vacuum-dried in a Concentrator Plus (Eppendorf) and the precipitate redissolved in 100 µl of ddH_2O. For non-zein protein extraction, a total of 100-mg of flour was extracted with 1 ml of zein extraction buffer three times. After each centrifugation, the supernatant was discarded. Finally, the precipitate was vacuum-dried and then redissolved in 1 ml of non-zein extraction buffer (12.5 mM sodium tetraborate, 2% 2-mercaptoethanol (v/v) and 5% SDS) for 2 h at room temperature. After centrifugation at 17,370 g for 10 min, the supernatant containing the non-zein proteins was transferred to a new tube. Then, 3 µl of zein and non-zein proteins were analyzed by 15% SDS-PAGE[59] .

GWAS analysis

We planted 518 inbred lines at Harbin in 2014, and 512 were harvested for zein protein analysis by SDS-PAGE. Three phenotypes were distinguished based on differential accumulation of α19 and α22 (i.e. α19 more than, equal to and less than α22). GWAS was performed based on the α-zein accumulation patterns. The GWAS method was described previously[60].

We planted 512 inbred lines at Sanya in 2019 and 2020, and 405 and 438, respectively, were harvested for measurement of seed protein content using the Dumas rapid nitrogen analyzer. For each inbred, three ears were used for biological repetition. For each ear, six seeds were dissected for measurement. Genome-wide association analysis was conducted with 1.63 million high-quality SNPs from maize haplotype map[61] by GEMMA

software[62]. Three principal components (PCs) were fitted, and the centered identical by state kinship matrix was used as random effects in the GWAS model.

Gene location

1）BSA sequencing

Plants of F_1BC_4, F_1BC_6 and F_1BC_8 populations were labelled, and a piece of leaf was sampled for DNA extraction. After measurement of the protein content of the next-generation seeds by SDS-PAGE (F_1BC_4) and the Dumas rapid nitrogen analyzer (F_1BC_6 and F_1BC_8), the corresponding high-protein and low-protein plants were identified. DNA samples of 75, 150 and 50 for each phenotype in F_1BC_4, F_1BC_6 and F_1BC_8 populations, respectively, were pooled and the library construction, sequencing and data analysis were completed by Shanghai OE Biotech Co, Ltd.

2) Data quality evaluation

The raw reads generated by high-throughput sequencing are in fastq format. In order to obtain high-quality reads that could be used for subsequent analysis, further quality filter of raw reads was required. Preprocessed by fastp (Version: 0.19.5) software[63], the quality filter included 4 steps: (1) Removing the linker sequence; (2) Removing reads with N (non AGCT) bases greater than or equal to 5; (3) The sliding window was performed with 4 bases as the window size, and an average base quality value less than 20 was removed; (4) After filtering, reads with length less than 75 bp or the average base quality value less than 15 were removed. Then, BWA (Version: 0.7.12) software[64] was used to align the clean reads to the reference genome to determine the position of the reads. The alignment algorithm was bwa mem, and the parameters were the default parameters. After the alignment, results were formatted and sorted by SAMtools (Version: 1.9) software, and the duplication reads were removed by Picard (Version: 4.1.0.0) software.

3) Variation information detection

Based on alignment of the sample sequencing data with the reference genome, SNP and InDel detection were performed using the Haplotypecaller module of the GATK (Version: 4.1.0.0) software [65].

4）G-value analysis

The G-value analysis was implemented by R package QTLseqr[66]. After manually filtering the snp data, a smoothed version of the standard G statistic called G' were calculated in 8 Mb window size and plotted grouped by chromosome. Green line indicates threshold of G' value corresponding to a q value of 0.01.

5）Introgression genes analysis

6）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 of R (v3.5.1). Coordinate positions are based on the teosinte Ames21814

haplotype. Peaks with delta depth >0.025 indicate introgression of teosinte genes. According to the G-value analysis results, there was a peak in the region between 120 Mb and 150 Mb (based on B73_V4) on chromosome 9, and the corresponding region aligned to the teosinte Ames21814 haplotype between 130 Mb and160 Mb (based on Teo_V1) on chromosome 9. Based on the teosinte Ames21814 haplotype annotation gff file, the number of introgressed teosinte genes in the 130 Mb and160 Mb region of F_1BC_4, F_1BC_6 and F_1BC_8 is 315, 271 and 190, respectively. By gene homologous alignment (blast v2.2.26), *teo09G002926* in Ames21814 genome corresponds to *Zm00001d047736* (*ZmASN4*) in B73 genome.

6) Resequencing Mapping Analysis

7) The introgressed genes on chromosome 9 from Ames21814 were analyzed. The physical coordinates of the extracted regions were based on the B73 reference genome (B73_V4). Based on the genomic sequence differences between Teosinte Ames21814 and B73, the SNPs between teosinte and B73 were used as mapping markers. The number of SNPs was counted every 10 kb as a window (SNP density) in each resequencing sample. Then the R (v3.5.1) package of ggplot2 was used to plot.

7) Fine mapping

More than 2000 F_1BC_9 individuals were planted, numbered, sampled, and self-pollinated. The seed protein content of F_2BC_9 was determined by a Rapid N analyzer. According to the teosinte Ames21814 haplotype sequence and B73 reference genome sequence, we designed molecular marker primers (Supplementary Table 6). Based on 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).

Structural variation analysis of *Thp9*

We analyzed gene structural variation using GSDS 2.0 (Gene Structure Display Server 2.0) based on the Ames21814 haplotype and B73 genome sequences. *ASN* transcripts in the root and leaf of B73 x Ames21814 were analyzed using kallistio (v0.44.0) (https://pachterlab.github.io/kallisto/download.html). The index was established, the default parameters for the paired-end sequencing results were used for alignment, and finally the abundance of the transcripts was obtained.

Genetic confirmation of *Thp9* in B73

The full-length coding sequence of *Thp9-T* was amplified from Ames21814 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 Wimi Biotechnology (Jiangsu).The primer sequences used in this study are shown in Supplemental table 6.

RNA extracted, reverse transcription, and RT-qPCR

Leaf and root tissues of B73, Ames21814 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 extracted with TRIzol reagent (Invitrogen, catalog number 15,596,018). RNA was purified with an RNeasy Mini Kit (Qiagen, catalog number 74,106) after DNaseI digestion (Qiagen, catalog number 79,254) and used for reverse transcription with a SuperScript III First Strand Synthesis Kit (Invitrogen, catalog number 18,080,051). cDNA was diluted to 80 ng/μl for RT-qPCR 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 control was set to 1. All data were generated from three replicate biological samples, and means and SD were calculated. The primer sequences are shown in Supplementary table 6.

Zein copy number analysis

To accurately locate zein gene clusters，BLASTN was used to align the assembled Ames21814 haplotype with the known α-zein clusters and flanking genes of the B73 and W22 inbreds[18] 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 588th amino acid was used to make antibodies by ABclonal (Wuhan, China). To analyze protein accumulation of ASN4 in roots of NILTeo and B73, total proteins were extracted using the non-zein buffer. Twenty μg of total protein was separated by 10% SDS-PAGE and then transferred electrophoretically to a PVDF membrane. The protein was detected with ASN4 antiserum at a dilution of 1:1000 at 4°C overnight, followed by secondary anti-rabbit-HRP at a concentration of 1:5000 (Abmart, catalog number: M21002L). The control protein, ACTIN, was detected with mouse monoclonal ACTIN antibody (Abmart, catalog number M20009L), and a secondary antibody, anti-mouse IgG-HRP (Abmart, catalog number M21001L). The membranes were treated with chemiluminescence substrate reagent (Invitrogen, catalog number: 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 an Anti-FLAG (Sigma, A8592) as primary antibody at a dilution of 1:1000, and anti-mouse IgG-HRP (Abmart, catalog number M21001L) as the secondary antibody at a diluttion of 1:5000. Imaging was with a Tanon-5200 system (Tanon).

NUE test

In 2021, we planted NILTeo and NILB73 at the Songjiang experimental filed in Shanghai, using soil in cement tanks with or without normal N application. For normal N, 20 g of N fertilizer was applied to each plant at the seedling stage (V4) and 20 g at the jointing stage (V12). The N content of the fertilizer is 17%. Gene expression, above-ground biomass, root biomass and seed protein content were investigated.

Larger field trials were performed in Sanya in 2021. Four different N applications were tested: i.e. normal application (16 g/plant applied at each seedling stage (V4) and jointing stage (V12)), 50% (8 g/plant applied at each seedling stage (V4) and jointing stage (V12)),

25% (4 g/plant applied at each seedling stage (V4) and jointing stage (V12)) and 0% (none applied). The N content of the fertilizer is 17%. Each treatment contained 300 plants grown at 0.6 m x 0.25 m for each plant. Plant height, above-ground biomass, total N content of the root, stem and leaf, seed protein content and amino acid content were measured.

Introgression of *Thp9-T* in hybrids

NILTeo and NILB73 were crossed with Mo17 to create hybrids that were grown in Harbin. Using molecular marker selection, *Thp9-T* was introgressed into Zheng58 and Chang7-2 by backcrossing for four generations. The marker was developed based on an InDel polymorphism between *Thp9-T* and *Thp9-B*. After backcrossing, the resulting plant materials were self-pollinated for two generations, creating Zheng58-T and Chang7-2-T. The cross of Zheng58-T and Chang7-2-T produced a modified hybrid, Zhengdan958-T, that carried the *Thp9-T* allele. Zhengdan958-T and Zhengdan958-B, with the *Thp9-B* allele, were grown in Sanya in 2021. Plant height, above-ground biomass, total N content of the root, stem, and leaf and seed protein content were measured.

Acknowledgments

We thank Yunping Xiao, Hongrui Zhan, Bo Lin, Yincong Gu, Ting Zhang and Dong An from Shanghai OE Biotech Co., Ltd for help with BSA sequencing and related analysis. We thank Kunyan Liu for help with the high-protein QTL analysis and gene mapping. We thank Changsheng Li from Anhui Agricultural University for help with the bioinformatics analysis. We thank Junpeng Shi from Sun Yat-Sen University, Weiya Li from University of Wisconsin-Madison for help in analysis of teosinte domestication and sequence variation. We thank Zhenhua Wang, Lin Zhang, Xing Zeng and Yu Zhou from Northeast Agricultural University for help with material planting in Harbin. This research was supported by the Chinese Academy of Sciences (XDB27010201 to Y.W.) and the National Natural Science Foundation of China (31830063 and 31925030 to Y.W.), China Postdoctoral Science Foundation (2020M681412 to YH), Shanghai "Super Postdoctoral" Incentive Program (2020456 to YH).

Reporting summary

Additional information regarding research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The Ames21814 genome sequences have been deposited in NCBI (BioProject: 822523) (https://www.ncbi.nlm.nih.gov/bioproject/822523). The RNA-sequencing data of Ames21814, B73 x Ames21814 and B73 (roots and leaves at flowering stage) have been deposited in NCBI (BioProject: 832948) (https://www.ncbi.nlm.nih.gov/bioproject/832948). Source data are provided with this paper.

Author contributions

Y.W., W.W., Y.H., and H.W. designed research, analyzed the data and supervised the project. Y.W.,Y.H., H.W., Y.Z., X.H. and H.L. created genetic populations and materials. Y.H., H.W., Y.Z., X.H., Y.Z., L.Q., Y.J., Y.C., Q.X., Q.W., J.W., H.L., X.L. performed experiments. W.W., S.L., X.W. Z.B. and Y.H. performed teosinte Ames21814 haplotype sequencing, assembly and annotation. G. M. and X.Y. performed the GWAS analysis. Y.W., X.L., Y.H., H.W. and Y.Z. performed the field NUE tests in Sanya. Y.W., W.W., Y.H., H.W. and B.L. explained the data, drafted and edited the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Table 1. Summary of teosinte genome assembly.

	Hifi Reads	**HiC reads**	**Final assembly**
Library	10-20 kb	PE150	-
Platform	PacBio Hifi	Illumina HiSeq	-
Clean Base (Gb)	104	376	-
Read Number	6,752,166	2,534,951,604	-
Sequencing Depth*	47	174	-
Assembled Size (Mb)	2,424	-	2,436
Contig N50 (Mb)	62.29	-	62.29
Contig Number	543	-	545
Scaffold N50 (Mb)	-	-	245.33
Scaffold Number	-	-	452

*Estimated with the genome size of 2.16 Gb.

Supplementary Table 2. Gene model annotation.

Chromosome	**Size (Mb)**	**Gene Number**
Chr1	316.34	7238
Chr2	291.63	6054
Chr3	245.33	5094
Chr4	269.67	5426
Chr5	283.13	5468
Chr6	189.99	4581
Chr7	190.65	3848
Chr8	239.54	4474
Chr9	165.99	3795
Chr10	150.29	3467
unplaced	93.44	8647
total	2436	58092

Supplementary Table 3. BUSCO analysis.

BUSCO	Number	Percent(%)
Complete BUSCOs (C)	1395	96.80%
Complete and single-copy BUSCOs (S)	1314	91.20%
Complete and duplicated BUSCOs (D)	81	5.60%
Fragmented BUSCOs (F)	12	0.80%
Missing BUSCOs (M)	33	2.40%
Total BUSCO groups searched	1440	100%

Supplementary Table 4. Comparison of repetitive elements between teosinte and B73.

Class		Teo		B73	
		Length (bp)	Percent(%)	Length (bp)	Percent(%)
Class I: Retrotransposon	All	1,647,168,798	67.61%	1,627,831,680	74.60%
	Copia	524,384,025	21.52%	543,774,336	24.92%
	Gypsy	1,009,605,405	41.44%	965,570,400	44.25%
	L1 LINE	5,573,192	0.23%	5,673,408	0.26%
	LINE element	1,362,291	0.06%	1,963,872	0.09%
	RTE LINE	1,783,030	0.07%	1,309,248	0.06%
	Others	104,460,855	4.29%	109,540,416	5.02%
Class II: DNA Transposon	All	198,664,789	8.15%	187,658,880	8.64%
	CACTA	73,199,220	3.00%	64,589,568	3.00%
	Mutator	26,943,771	1.11%	21,602,592	0.99%
	PIF Harbinger	20,492,037	0.84%	23,130,048	1.06%
	Tc1 Mariner	5,219,228	0.21%	12,001,440	0.55%
	hAT	27,227,736	1.11%	25,093,920	1.15%
	Helitron	45,582,797	1.87%	41,241,312	1.89%
Unclassified	All	263,241,107	10.80%	44,296,224	2.39%
	Centromeric repeat	14,165,249	0.58%	4,800,576	0.58%
	Knob	247,237,582	10.15%	36,440,736	1.67%
	Subtelomere	1,024,677	0.04%	1,091,040	0.05%
	rDNA spacer	813,599	0.03%	1,963,872	0.09%
Total		2,109,074,694	86.57%	1,860,659,616	84.34%

1301 Supplementary Table 5. Strutural variation.

Name	Type	SV Counts	>50bp Counts	>100kb counts
Copy gain	CPG	741	741	1
Deletion in non-reference genome	DEL	18,431	18,431	0
Copy loss	CPL	2,070	2,070	2
Duplication	DUP	65,101	65,101	13
Insertion	INS	16,559	16,559	0
Inversion	INV	312	312	71
Inverted Duplication	INVDP	63,678	63,678	11
Inverted Translocation	INVTR	27,351	27,351	0
Tandem repeat	TDM	170	170	0
High diverged region	HDR	58,678	58,678	254
Translocation	TRANS	27,790	27,790	25
Total		280,881	280,881	377

1302

1303

1304 Supplementary Table 6. Primers (5'-3') used in this study.

Primers	Sequence	Purpose
Primers for fine mapping		
130.6-F	CCTGTTTGAATCCATGGTGCTAAA	fine mapping
130.6-R	CATAGATGTCGTTTCTGGCATGAC	fine mapping
132.5-F	AAGAGCAGAACAATGATGGTACCT	fine mapping
132.5-R	TTGATGATGCTGTTGGTAAGTTGG	fine mapping
133.5-F	AGAAAAGGAGTGCAGCTTCAGATA	fine mapping
133.5-R	ACTTTCTTTACCGTCTGAACTCGA	fine mapping
137.8-F	AGTTTCATGCCTCGTTAGTACCAT	fine mapping
137.8-R	GGAGCCCACATGTTATTTGATCTC	fine mapping
139.1-F	TCCTCGATCCAAACAGACTTTTCT	fine mapping
139.1-R	TCCCGGTTTATTTTGTAGACACCT	fine mapping
139.9-F	TGAACTGTGAGGTATGAGGATGTG	fine mapping
139.9-R	TTAAAACTAGTCACGTTGCTGCTC	fine mapping
140.2-F	AGAAATATGTACCACCATGCACA	fine mapping
140.2-R	TCGATTGACTGTAAAAGGCGA	fine mapping
Del47-F	CTCTGTGCCATGCATCCTCC	fine mapping
Del47-R	CGTCAGCGCTGGTTAGC	fine mapping
140.3-F	ACTCAGGGTACGAAGTTTTGAGTC	fine mapping
140.3-R	CAAATCCATAGAGCTTTTTGCGACC	fine mapping
140.4-F	TCAGGGTACGAAGTTTTGAGTC	fine mapping
140.4-R	CCATAGAGCTTTTTGCGACCTG	fine mapping
140.6-F	ATCGATGTCAGGAGGAGTATAGGA	fine mapping
140.6-R	TCCAATGGCTAACAGTTGTAGTGA	fine mapping

141.0-F	AGACTATCAAAGCAAGTGGTCCAT	fine mapping
141.0-R	GGTTGCTACTGTTCAATTCTTGCT	fine mapping
142.8-F	GGAAAGAAATTAAAGCGGACACCA	fine mapping
142.8-R	CTTATCAGCTGTCGAGAGTCTTGT	fine mapping
Primers for near-isogenic lines (NILs)and breeding materials construction		
Asn4is9	CTCTGTGCCATGCATCCTCC	Identification
Asn4is9	CGTCAGCGCTGGTTAGC	Identification
Asn4is9	CAAGCCTGAACTGACGCCT	Identification
Asn4is9	CGTCAGCGCTGGTTAGC	Identification
Asn4ind	CCGTTCCTCGACAAGGAGTT	Identification
Asn4ind	ATCAGAGCTGAAAGTGGGGC	Identification
Asn4ind	GCATGGACCCCGAATGGAAA	Identification
Asn4ind	GCGATCAGAGCTGAAAGTGG	Identification
Primers for Geneclone and vector construct		
A4cds-	GTCTCCTCCTCCCCACAAAA	Cloned gene
A4cds-	TCGTTCGTCCCTATCCTCCA	Cloned gene
ASN4-3300-FlgF	aggtcgactctagaggatccATGgactacaaggaccatgacggtgactacaaggaccatgacattgactacaaggatgacgatgacaagggaggaggatgtggcatcttagccgtg	Over-expression
ASN4-3300-	cgatcggggaaattcgagctcCTAACTTCGACGACGACCATCAACC	Over-expression
Primers for RT-qPCR		
Asn4-rt-	catcgcggtgtaacagtaatgaa	gene
Asn4-rt-	gcttgtactgctctgctttgtca	gene
actin-F	gctacgagatgcctgatggtc	Internal
actin-R	cccccactgaggacaacg	Internal
qA4OE-	GGAGGTGATCTACCACGACG	RT-qPCR
qA4OE-	CAGTCGTACTGGTGCAGGG	RT-qPCR
Primers for search for natural variation of Zm*Asn4*		
A4cds-	GTCTCCTCCTCCCCACAAAA	PCR
A4cds-	TCGTTCGTCCCTATCCTCCA	PCR
A4ID-3F	TTTCAAGCCTGAACTGACGC	Detect
A4ID-3R	ACCTACCAGTCGTTCGTTCG	Detect

1305

1306

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- nrreportingsummary.pdf