nature genetics

Article

https://doi.org/10.1038/s41588-026-02655-2

Multiomics analysis of primary metabolism reveals the genetic basis of nitrogen partitioning modulated by *ZmAVT1A-1* in maize

Received: 20 December 2023

Accepted: 2 June 2026

Published online: 29 June 2026

Check for updates

Min Jin[1,2,3,10], Shijuan Yan[4,5,10], Yuting Wu[1,2], Zhaowei Zhai[1,2], Saleh Alseekh[6,7], Yuanyuan Chen[1,2], Wenjie Huang[4], Niannian Ma[3], Keyu Tao[3], Kangyi Xiao[1], Yongli Zhu[1,2], Yanhui Yu[1,2], Yang Shao[1], Chenglin Jiang[1,2], Xiangguo Liu[8], Jiamin Sun[1,2], Yanzhi Qu[1,2,3], Wenjie Wei[1,2], Wenqiang Li[1,2], Jieting Xu[1], Jingyun Luo[1,2], Xiaqing Wang[9], Lin Zhuo[1], Junpeng Zhan[1,2], Fazhan Qiu[1], Ning Yang[1,2], Yingjie Xiao[1,2], Hai-Jun Liu[1,2], Alisdair R. Fernie[6,7] & Jianbing Yan[1,2] ✉

Maize primary metabolism drives complex agronomic traits, yet its genetic regulation remains difficult to resolve. Here we integrated genomic, transcriptomic and metabolomic data from 1,404 maize progenies derived from 24 diverse founders to dissect the genetic architecture of primary metabolism. We constructed a high-confidence regulatory network that resolved causal genes underlying metabolic quantitative trait loci and successfully identified targets for improving maize nutritional quality. This systems-level framework further prioritized *ZmAVT1A-1*, encoding a putative amino acid transporter, as a key regulator of amino acid accumulation. Natural variation and transgenic analyses showed that *ZmAVT1A-1* modulates nitrogen partitioning between vegetative tissues and kernels, revealing pleiotropic effects on agronomic traits. These findings illustrated the intricate trade-offs inherent in metabolic regulation. Together, our study provides a comprehensive multiomics resource for decoding metabolic networks and underscores the necessity of a systems approach to navigate the pleiotropic nature of crop improvement targets.

Maize (*Zea mays*) is a crucial global staple food, yet the improvement of its biomass, grain production, quality and nutrient content to meet the demands of increasing worldwide demands remains a key challenge[1]. Modern agricultural systems rely on applying excessive inorganic nitrogen fertilizers in the pursuit of high yields, leading to substantial environmental pollution. Consequently, developing new crop varieties that maintain yields with lower nitrogen inputs, that is, crops with high nitrogen use efficiency (NUE), is urgently needed for more sustainable agriculture[2].

A number of genes exhibiting natural variation have been cloned for traits such as plant architecture, flowering time, yield components, grain quality and both biotic or abiotic stress resistance[3]. These traits are normally visible or at least relatively straightforward to measure. However, they are also polygenic and subject to dynamic developmental, physiological and/or environmental interactions. Many genes or loci contribute only a minor direct effect to a given phenotype while interacting with one another. Therefore, it is usually difficult, labor intensive and time consuming to identify and clone the genes underlying such complex phenotypes.

A full list of affiliations appears at the end of the paper. ✉e-mail: yjianbing@mail.hzau.edu.cn

Plant metabolites are categorized into primary metabolites and secondary/specialized types. Primary metabolites play direct and fundamental roles in plant growth and development, acting as cellular building blocks and energy sources. They also serve as the precursors to secondary metabolites with various functions[4,5] and often function directly or indirectly as signaling molecules[6]. Primary metabolism, characterized by high-flux pathways such as photosynthesis, nitrogen assimilation and respiration shape the developmental and physiological status of plant cultivars and can serve as mediators that ultimately affect high-level breeding targets such as quality or yield. Exploring the genetic control of primary metabolite levels can, therefore, inform new breeding approaches.

Large-scale metabolomic studies have flourished in the past two decades[7–10]. These studies, combined with genomic information, have been widely applied in model plants and major crops to help characterize complex quantitative traits[5,9,11–15]. Previous studies suggest that plant metabolite abundances are moderately heritable and largely under the polygenic control of several loci[14]. Generally, secondary metabolites are regulated by a few loci with large effects[16], whereas primary metabolites are controlled by a greater number of loci with smaller effects[12,14]. In addition, the variance in secondary metabolites is typically larger than that in primary metabolites[17]. Nevertheless, our understanding of plant metabolic heredity follows a relatively simple model of inheritance; this contrasts with complex end-phenotype traits such as flowering time and grain yield, which are controlled by many small-effect loci. Metabolites, therefore, can be regarded as intermediate phenotypes and correlated biomarkers to bridge genotype–phenotype gaps for more complex breeding traits[18]. Recent works have revealed the genetic basis of metabolomic responses to abiotic and biotic stresses, such as maize adaptation to drought[19] and tobacco resistance to leafhoppers[20]. However, the direct links between traits remain incomplete and there is still a lack of systematic integration of metabolite quantitative trait loci (mQTLs), phenotypic traits or their associated quantitative trait loci (QTLs), biological pathways and/or compound annotations[21]. Exploring metabolic variation can also aid in gene functional characterization through simple follow-up experiments, and novel metabolites could be identified based on their links to functionally identified genes[16]. Ultimately, characterizing the genetic basis of the plant metabolome could facilitate the reconstruction of biochemical pathways, the discovery of novel genes and the delineation of regulatory networks that govern metabolism. This would, in turn, deepen our understanding of the genetic regulation of complex traits.

Current genetic designs are used to dissect the genetic architecture of plant metabolome primarily based on recombinant inbred lines or natural populations[12]. Such designs are often insufficient to identify loci influencing primary metabolite levels as these are subjected to a complex network of chemical processes regulated at both genetic and post-translational levels[17,22], making genotype-to-metabolite relationships difficult to resolve with limited mapping resolution or sample size. In parallel, the breadth of the plant chemical space—much of which remains structurally uncharacterized[13]—further constrains biological interpretation of associated loci. The integration of multiomics data offers a complementary route to link genetic loci to molecular mechanisms by connecting genomic variation to transcriptomic regulation and metabolite phenotypes[9,10].

Here we used the statistical power of a large-scale synthetic population, the Complete-diallel plus Unbalanced Breeding-derived Inter-Cross (CUBIC) population derived from 24 parents and consisting of 1,404 progenies[23], to dissect the genetic architecture of maize primary metabolism. We performed metabolic profiling of maize leaves and kernels using gas chromatography–mass spectrometry (GC–MS) and integrated these data with high-density genotypic information. Using both single-variant-based genome-wide association study (sGWAS) and haplotype-based genome-wide association study (hGWAS) approaches, we systematically investigated the genetic architecture underlying primary metabolite variation and its connectivity to agronomic phenotypes. Among multiple candidate genes identified from the systems-level analysis, we prioritized *ZmAVT1A-1*, encoding a putative amino acid transporter, as a proof-of-concept for network-guided gene characterization. Functional analyses demonstrated that *ZmAVT1A-1* modulates amino acid accumulation and influences nitrogen partitioning. Moreover, these metabolic shifts were accompanied by pleiotropic changes in agronomic and yield-related traits.

Together, our study establishes a robust multiomics resource for exploring the genetic regulation of primary metabolism in maize. Furthermore, our characterization of *ZmAVT1A-1* underscores a critical complexity in crop genetics: natural variation in metabolic hub genes often drives agronomic traits through intricate pleiotropic trade-offs, which must be carefully navigated during the utilization of such alleles for maize improvement.

Results

Variation in primary metabolite abundance within the CUBIC population

The metabolomes of seedling leaves and mature kernels from 1,404 progeny and 24 parents were quantified using a GC–MS-based metabolomics approach. There were 144 metabolites identified in this study (Supplementary Table 1) and 86 metabolites confirmed by authentic standards were selected for further analysis (Methods). In total, 67 metabolites were identified in maize leaves and 68 in kernels, with 49 common to both tissues (Fig. 1a). These metabolites were categorized into five classes: organic acids, amino acids, sugars, amines and others (Supplementary Table 1). The progeny showed greater metabolic variation than the parents in both leaves and kernels (independent Student's *t*-test, $P = 2.4 \times 10^{-3}$ and 4.8×10^{-4}, respectively; Fig. 1b and Supplementary Table 2), probably due to increased recombination events during population development.

The metabolites could be grouped into six clusters on the basis of their Spearman correlations. Five of these clusters (I, II, III, IV and VI) consisted of highly correlated compounds, primarily driven by tissue-specific patterns and similar biochemical classifications (Fig. 1c). Notably, amino acids in clusters I and VI were predominant in kernel and leaf. Meanwhile, ethanolamine and allantoin showed strong correlations with amino acids in kernels, whereas 1,3-diaminopropane and sucrose were tightly associated with amino acids in leaves (Fig. 1c).

Metabolite GWAS reveals complex genetic control of primary metabolite variation

To identify loci responsible for metabolite variation, we performed a genome-wide association study (GWAS) using 11.8 million single-nucelotide polymorphisms (SNPs)[23] with both sGWAS and hGWAS, resulting in the identification of 'sQTLs' and 'hQTLs', respectively. We identified a total of 1,373 metabolite QTLs (mQTLs) (Fig. 2a), including 286 sQTLs and 395 hQTLs associated with 51 and 67 metabolites in leaves, respectively (Supplementary Table 3) and 207 sQTLs and 485 hQTLs for 52 and 65 metabolites in kernels (Supplementary Table 4). On average, each metabolite was associated with 5.6 sQTLs and 5.9 hQTLs in leaves, with 3.9 sQTLs and 7.2 hQTLs in kernels (Extended Data Fig. 1a).

The percentage of phenotypic variation explained (PVE) by individual QTL varied widely, with averages of 4.4% for sQTLs and 5.9% for hQTL in leaves, and 4.6% for sQTLs and 6.5% for hQTLs in kernels (Supplementary Tables 3 and 4). The total PVE for individual metabolite also ranged significantly, with higher averages in kernels (33.5%) compared to leaves (28.3%) ($P = 0.018$; Extended Data Fig. 1b and Supplementary Table 5). By comparing the number of mQTLs and PVE values to previously published works on biparental populations and natural association populations[24–26], we found that this study uncovered numerous novel mQTLs, many of which have minor effects (Extended Data Fig. 1c,d and Supplementary Tables 3 and 4).

https://doi.org/10.1038/s41588-026-02655-2

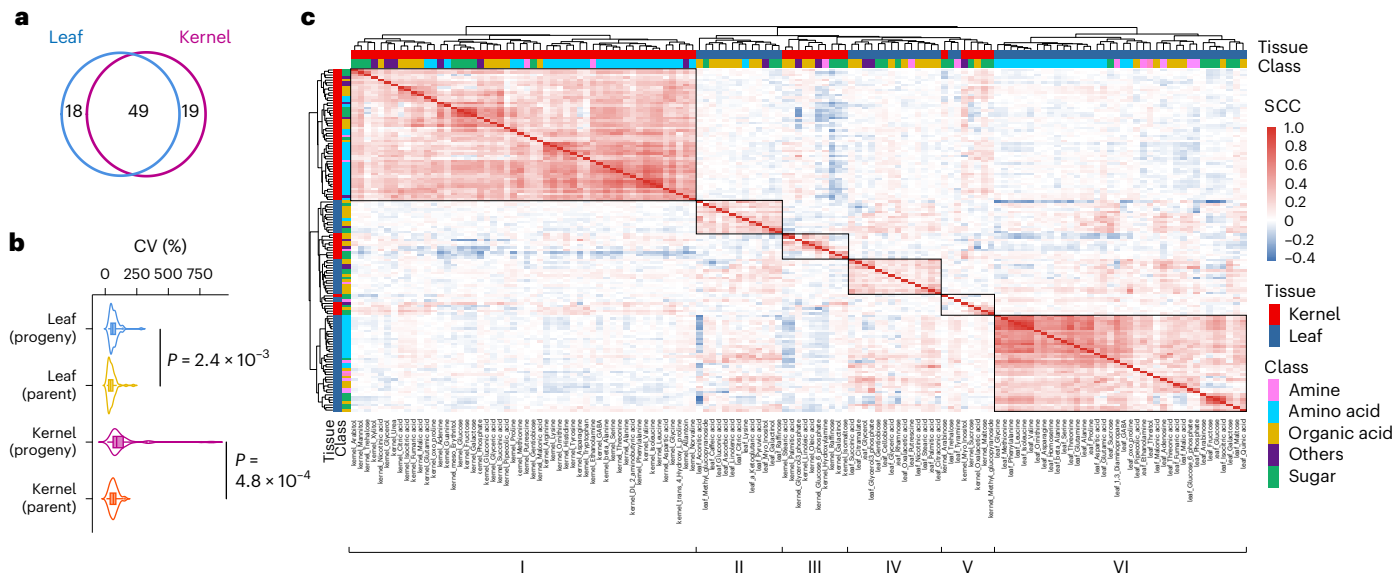


Fig. 1 | Metabolite variation across the CUBIC population. a, A Venn diagram of metabolites detected in maize leaves and kernels from the CUBIC population. **b**, The coefficient of variation (CV) between metabolites in leaves and kernels of parents and progenies. Each data point represents one metabolite and was calculated from population-level metabolite values within each group. The box plots show the median and interquartile range (IQR; 25–75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR. The sample sizes were as follows: leaf metabolites, $n = 67$; kernel metabolites, $n = 68$. *P* values were calculated using two-tailed Student's *t*-tests. **c**, A heat map of Spearman's correlation coefficient (SCC) between metabolites in kernels and leaves.

Analysis of mQTL distribution across the genome revealed 16 hotspots (based on 100 permutation tests) (Fig. 2a), consistent with previous observations that mQTLs clustered nonrandomly across the genome[25]. Enrichment analysis revealed that metabolites associated with these hotspots were enriched in specific Kyoto Encyclopedia of Genes and Genomes (KEGG) modules, including the tricarboxylic acid (TCA) cycle and the reductive carboxylate cycle (metabolites associated with hotspots 3, 6 and 7). KEGG pathways were also enriched, for example, β-alanine metabolism and galactose metabolism (metabolites associated with hotspot 2 and hotspot 13, respectively; Supplementary Table 6). These mQTL hotspots represent promising targets for investigating the regulation of complex metabolic networks.

To compare the genetic bases of metabolites to different tissues, we examined mQTLs for the 49 metabolites detected in both leaves and kernels. Only 6.9% (72/1049) of loci were colocalized (Extended Data Fig. 1e and Supplementary Table 7), indicating substantial distinct genetic control across tissues, although the degree of tissue specificity may be overestimated because the two tissues were profiled in independent grow outs. The shared loci involved 16 metabolites including six sugars (sucrose, raffinose, galactinol, glucose, maltose and methyl-glucopyranoside), four amino acids (GABA, glutamate, proline and threonine), two organic acids (fumarate and isocitric acid), two amines (ethanolamine and putrescine) and two other compounds (glycerol-3-phosphate and *myo*-inositol). Notably, at a raffinose locus on chromosome 8, we highlighted a SWEET transporter as a candidate regulator and eight other colocalized loci contained genes encoding enzymes involved in metabolic biosynthesis (Supplementary Table 7), suggesting both tissue-specific genetic regulation and environmental effects.

Multiomics network prioritizes functional regulators of the maize metabolome

To systematically identify novel components affecting variation in the leaf and kernel metabolome and characterize the interplay among genetic variation, gene expression and metabolite abundance, we integrated transcriptomic data to connect genomic loci and metabolites with candidate causal genes. This genome–transcriptome–metabolome integration generated a high-confidence network containing 114 metabolic traits associated with 409 mQTLs and 1,630 genes (Fig. 2b and Supplementary Table 8), successfully resolving candidates for 30.0% of the identified loci (409/1,373 mQTLs).

The network demonstrates high biological relevance, as 27.9% (454/1,630) of the candidate genes encode enzymes that catalyze reactions within two steps of the target metabolite based on KEGG database annotations (Supplementary Fig. 1 and Supplementary Table 9). This observation underscores the network's capacity to capture core metabolic regulators. To validate the network's predictive power for maize quality traits, we prioritized three candidate genes expressed during kernel filling: *ZmASN2* (*GRMZM2G093175*), encoding asparagine synthetase 2; *ZmMIPS1* (*GRMZM2G155242*), encoding *myo*-inositol-3-phosphate synthase; and *ZmMGL* (*GRMZM2G450498*), encoding methionine γ-lyase (Fig. 3a–c and Supplementary Fig. 2). It is worth noting that as a homolog of *THP9*[27], *ZmASN2* exhibits distinct expression patterns in kernels and roots, which may indicate a specialized role in regulating asparagine accumulation during kernel filling and is consistent with its function in kernel nitrogen metabolism (Supplementary Fig. 2). We generated two independent CRISPR–Cas9 edited events for each gene to confirm their metabolic functions (Fig. 3d–f). The *ZmASN2* knockouts (KOs) exhibited significantly ($P < 0.01$) lower kernel asparagine levels (Fig. 3g). *ZmMIPS1* KO lines showed reduced ($P < 0.01$) *myo*-inositol (Fig. 3h), a precursor to the antinutrient phytate, suggesting potential for enhancing bioavailability[28]. KO lines of *ZmMGL* displayed increased ($P < 0.01$) methionine levels (Fig. 3i), mitigating the deficiency of this essential amino acid in kernels[29].

Collectively, these results demonstrate that our multiomics approach effectively prioritizes functional candidates driving metabolite accumulation and nutritional value. We summarized the metabolic effects observed across candidates in a pathway-level schematic (Fig. 3j), highlighting that *ZmASN2*, *ZmMIPS1* and *ZmMGL* perturbations consistently shift their corresponding primary metabolites (asparagine, *myo*-inositol and methionine, respectively) and thereby connect

https://doi.org/10.1038/s41588-026-02655-2

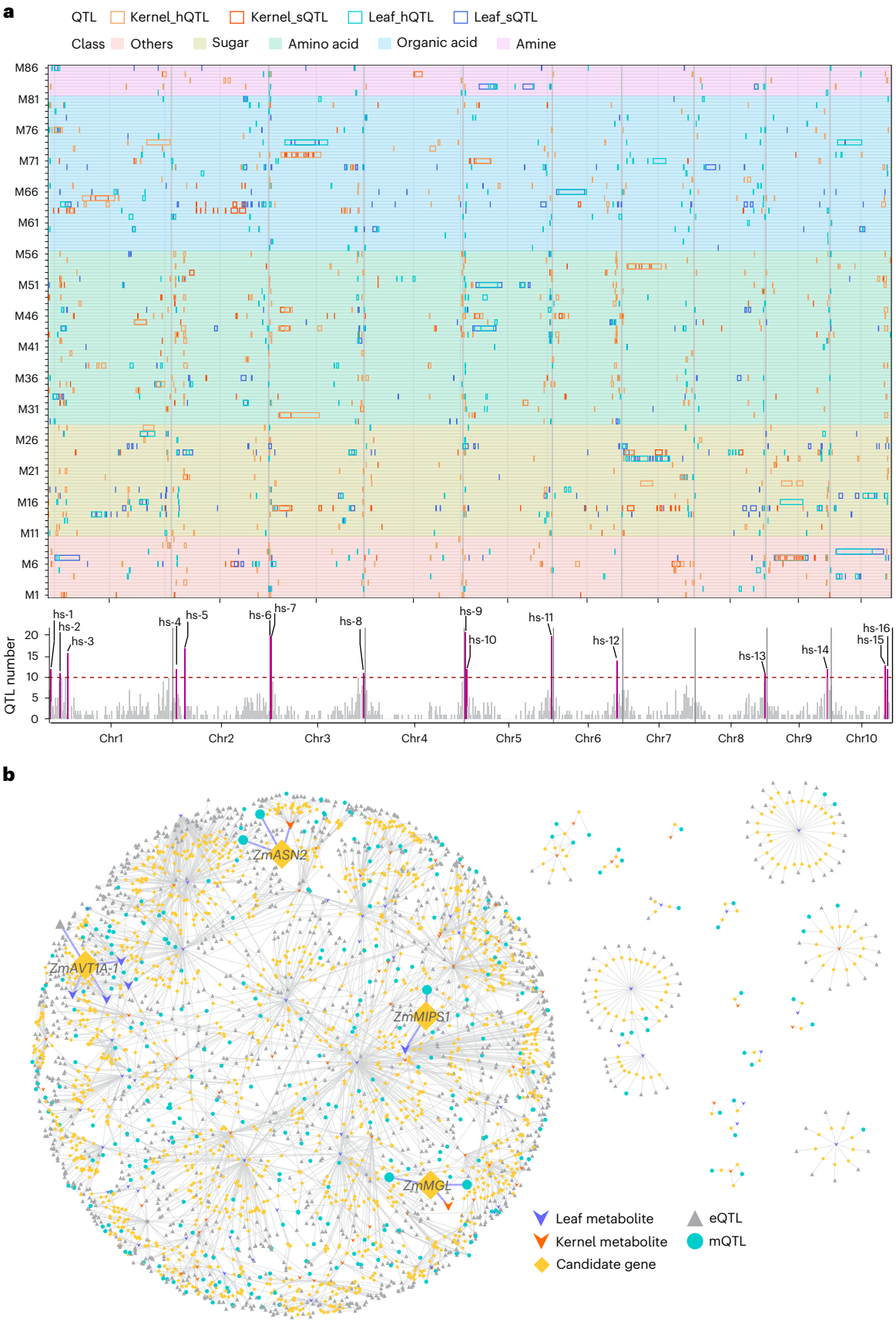


Fig. 2 | Genetic architecture of primary metabolites and candidate genes revealed by multiomics data. a, Top: the distribution of mQTLs detected by sGWAS (sQTLs) and hGWAS (hQTLs). The corresponding metabolite names can be found in Supplementary Table 1. Bottom: the number of QTLs in a 3-Mb sliding window. Pink bars indicate QTL hotspots. **b**, Network of associations among genes, QTLs and metabolites based on transcriptome data.

https://doi.org/10.1038/s41588-026-02655-2

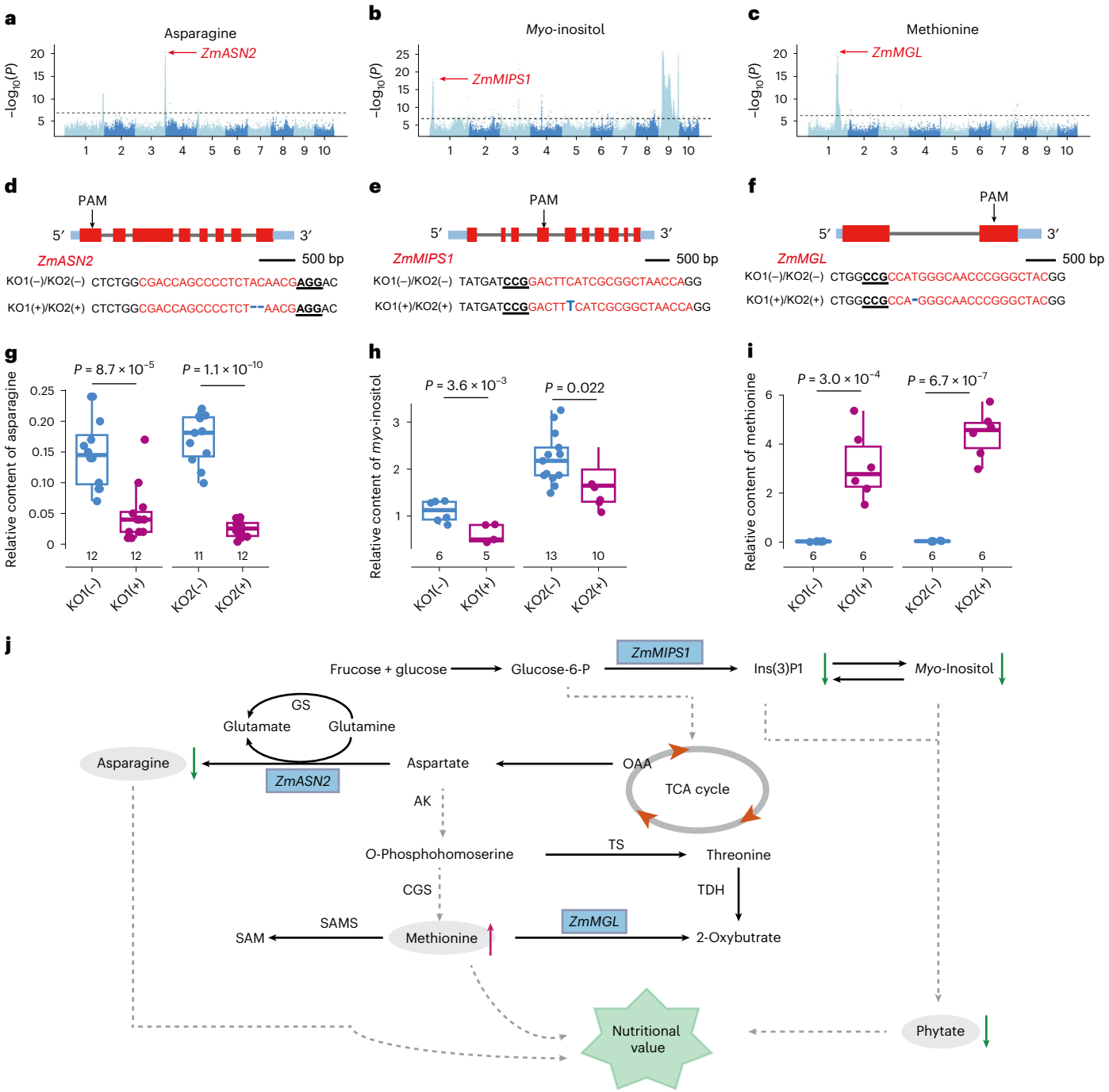


Fig. 3 | Candidate gene validation for central metabolism. a–**c**, Manhattan plots for relative contents of asparagine (**a**), *myo*-inositol (**b**) and methionine (**c**). In **a**–**c**, the dashed line indicates the Bonferroni-adjusted significance threshold ($P = 8.5 \times 10^{-8}$). **d**–**f**, Gene structures and targeted sequences for *ZmASN2* (**d**), *ZmMIPS1* (**e**) and *ZmMGL* (**f**) of CRISPR editing. **g**–**i**, Asparagine (**g**), *myo*-inositol (**h**) and methionine (**i**) content quantified with GC–MS followed by methanol/MTBE biphasic protocol. *P* values were calculated using a two-sided Student's *t*-test. The box plots show the median and IQR (25–75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR. **j**, Proposed links across primary metabolism and nutritional value in maize. AK, aspartate kinase; CGS, cystathionine γ-synthase; GS, glutamine synthetase; SAM, *S*-adenosylmethionine; SAMS, SAM synthase; TDH, threonine dehydratase; TS, threonine synthase. Candidate genes are shown in blue and quality-associated traits are in gray. Upward arrows represent an increase in the KO lines and downward arrows represent a decrease. In **g**–**i**, KO(−) denotes event-matched null segregants lacking the edit and KO(+) denotes plants carrying the KO allele.

core nitrogen and carbon metabolism to nutrition-associated traits. This summary provides a compact framework for interpreting how manipulating key nodes in primary metabolism can influence multiple metabolite classes relevant to grain nutritional value. Building on this network, we next focused on amino acid metabolism and examined its links to complex agronomic traits.

Network analysis identifies a novel gene affecting amino acid accumulation in leaves

A locus on chromosome 5 associated with four leaf amino acids (valine, isoleucine, leucine and phenylalanine) was identified (Fig. 4a,b). Branched-chain amino acids are key indicators of nitrogen status[30], making them particularly relevant to our analysis.

https://doi.org/10.1038/s41588-026-02655-2

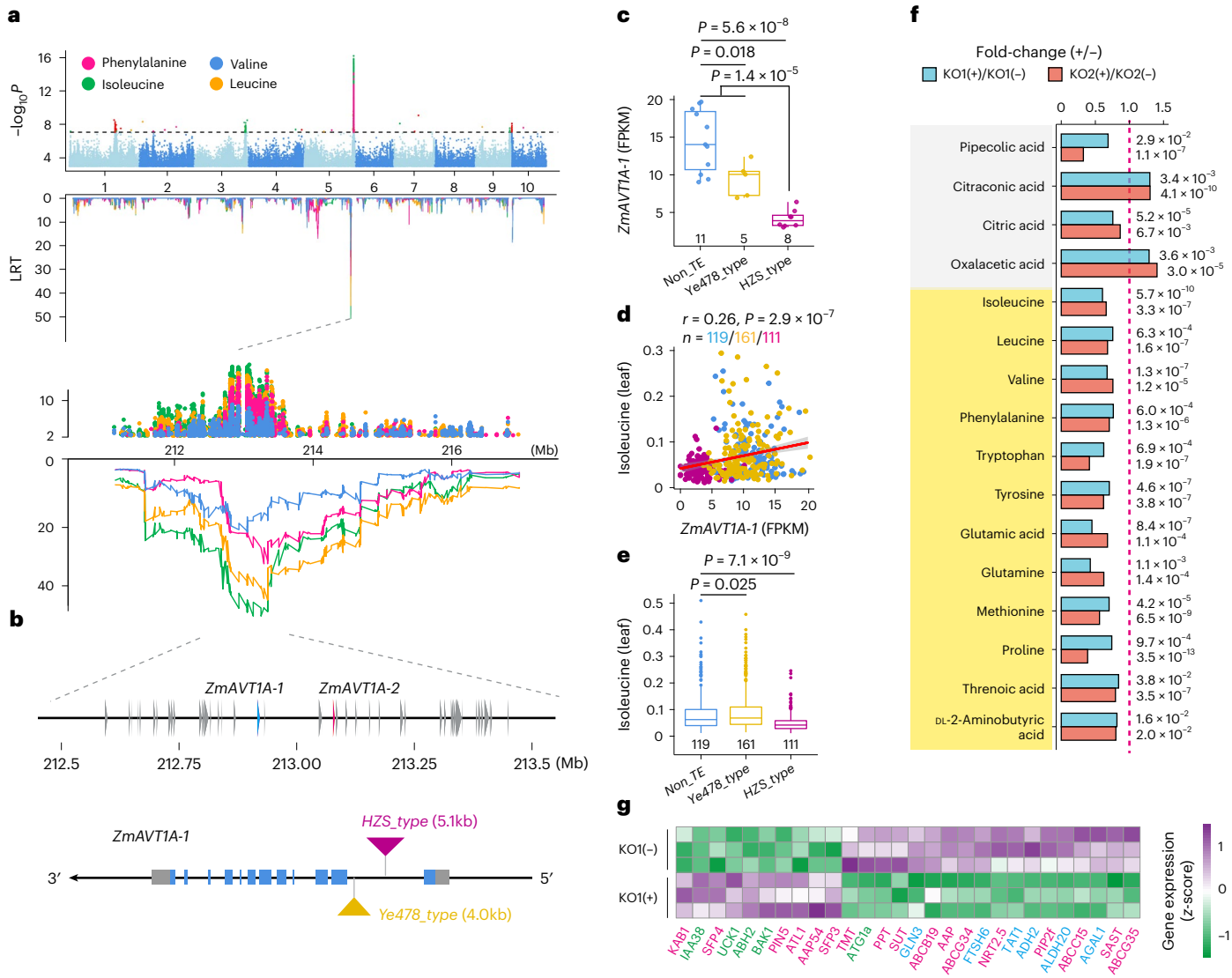


Fig. 4 | *ZmAVT1A-1* is a novel determinant of multiple leaf metabolite levels. **a**, GWAS results for the contents of four amino acids on chromosome 5. The dashed line indicates Bonferroni-adjusted significance threshold ($P = 8.5 \times 10^{-8}$). **b**, Structure of the candidate gene *ZmAVT1A-1* from the region in **a**. **c**, Insertion/deletion variations affect expression of *ZmAVT1A-1* in parent leaves; the numbers above the *x* axis represent the number of individuals. **d**, *ZmAVT1A-1* expression is positively correlated with isoleucine content in leaves. Pearson correlation was used. Blue represents the *Non_TE* haplotype, yellow the *Ye478_type* haplotype and pink the *HZS_type* haplotype. **e**, Insertion/deletion variations affect the isoleucine content of progeny leaves; the numbers above the *x* axis represent the number of individuals. In **c** and **e**, the box plots show the median and IQR (25–75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR. *P* values were calculated using a two-sided Student's *t*-test. **f**, Fold changes (KO(+)/KO(−)) of significantly altered leaf metabolites in two independent *ZmAVT1A-1* KO lines. For each event, KO(−) denotes event-matched null segregant siblings lacking the edit and KO(+) denotes siblings carrying the KO allele; the fold change was calculated as KO(+)/KO(−). The *P* values were calculated using a two-sided Student's *t*-test and are shown above each bar. Yellow and gray shading represents amino acids and organic acids, respectively. The dashed pink line indicates a fold change of 1 (no change). Sample sizes: line 1, $n = 18$ KO(+) and 29 KO(−); line 2, $n = 30$ KO(+) and 28 KO(−). **g**, A heat map of differentially expressed genes related to signaling (green), transport (pink) and enzymatic reactions (blue) in KO(−) and KO(+) plants. The gene expression is shown in transcript per million values scaled by *z*-score.

Within the confidence interval lie two homologous genes, GRMZM2G083788 and GRMZM2G080843, members of the *ATLb* subfamily of amino acid transporters[31,32] (Extended Data Fig. 2a). Their *Arabidopsis* homolog, *AtAVT1A*, localizes to the tonoplast[33], leading us to name the maize genes *ZmAVT1A-1* and *ZmAVT1A-2*. *ZmAVT1A-1* was prioritized owing to its leaf-specific expression, while *ZmAVT1A-2* is primarily expressed during early kernel development[34] (Extended Data Fig. 2b). Interestingly, *ZmAAP31*, located on chromosome 4 and phylogenetically close to *ZmAVT1A-1*, exhibits a similar constitutive expression profile[34] (Extended Data Fig. 2c and Supplementary Table 3), suggesting a potential role in both vegetative and reproductive stages.

Genome analysis revealed two distinct transposable element (TE) insertions in the first intron of *ZmAVT1A-1* (~5.1 kb in HZS, ~4.0 kb in Ye478) but absent in B73[35] (Extended Data Fig. 3a,b). Three haplotypes were detected among the panel[23], with the 5.1 kb *HZS_type* TE (PIF/Harbinger_TIR_transposon) insertions clearly reducing *ZmAVT1A-1* expression compared to the *Non_TE* haplotype and the *Ye478_type* (L1_LINE_retrotransposon) across both parents and progenies. Meanwhile, the *Ye478_type* also reduced its expression with minor effect (Fig. 4c and Extended Data Fig. 3c). *ZmAVT1A-1* expression correlated positively ($r = 0.13–0.26$) with amino acid contents (Fig. 4d and Extended Data Fig. 3d). The 5.1-kb intronic TE with *HZS_type* significantly reduced amino acid levels, explaining 2.7–8.8% of phenotypic

variation for leaf valine, phenylalanine, leucine and isoleucine (Fig. 4e and Extended Data Fig. 3d). Tonoplast localization of *ZmAVT1A-1* was confirmed using a GFP fusion in maize protoplasts, which co-localized with the tonoplast marker ZmTPK[36] (Supplementary Fig. 3a). Heterologous complementation in a yeast amino acid transporter mutant showed its broad substrate selectivity[37] (Supplementary Fig. 3b).

Two independent CRISPR–Cas9 KO events generated in the KN5585 background (*Non_TE* haplotype) (Supplementary Fig. 4a,b) showed reduced levels of 12 amino acids (0.39–0.84-fold-change compared to corresponding null-segregant controls) and four organic acids (pipecolic, citraconic, citric and oxalacetic acid) in leaves (Fig. 4f and Supplementary Table 10), indicating a disruption in nitrogen metabolism. RNA-sequencing of the KOs identified 121 differentially expressed genes (≥1.5-fold change) (Supplementary Table 11), enriched in signaling pathways, such as 'protein serine/threonine kinase activity', 'response to external stimulus', 'membrane' and 'transmembrane transporter activity' (adjusted $P < 0.05$) (Supplementary Fig. 4c). Notably, Zm*NRT2.5* (nitrate transporter 2.5) and Zm*GLN3* (glutamine synthetase 3), involved in nitrogen metabolism, were among the differentially expressed genes (Fig. 4g). Together, these results indicate that *ZmAVT1A-1* contributes to leaf amino acid accumulation and nitrogen metabolism.

ZmAVT1A-1 modulates NUE-related traits through source-to-sink nitrogen partitioning

Amino acid transporters underpin nitrogen uptake, remobilization and utilization—and thus NUE, as shown in *Arabidopsis* studies[38,39]. Given the transcriptomic signature consistent with altered nitrogen-associated processes in *ZmAVT1A-1* KO plants, we next evaluated its effects on nitrogen-related traits using transgenic materials in both pot and field experiments.

We first evaluated the gene's role in nitrogen-related traits with transgenic plants under high nitrogen (HN) and low nitrogen (LN) (Extended Data Fig. 4). In pot assays under LN, KOs exhibited accelerated senescence and increased leaf chlorosis than null-segregant controls, whereas overexpression (OE) plants showed slightly higher chlorophyll content during early kernel filling (Extended Data Fig. 4a–d). Weight of harvest ear were assessed in the corresponding validation experiments. Under LN, ear weight was 49.2% lower in KO plants, whereas OE lines in pots exceeded controls by 21.7% under HN and 47.3% under LN (Extended Data Fig. 4e).

To further investigate the cellular dynamics of this trait, we performed a ^{15}N-urea pulse-chase experiment (pulsed at V10, sampled at 20 days after pollination (DAP)), followed by the quantification of total N and ^{15}N in kernels and six vegetative tissues (Extended Data Fig. 4f). Although the overall biomass was similar (Extended Data Fig. 4g), *ZmAVT1A-1* could lead to tissue-specific shifts: KO accumulated biomass in lower stems particularly under LN, whereas OE gained kernel biomass particularly under HN, and the leaves remained no significant changes (Extended Data Fig. 4h–j). Isotopic tracking confirmed the partitioning effects: KO showed reduced total nitrogen and ^{15}N allocation to kernels, coupled with greater retention in the lower stems (Extended Data Fig. 5k–o); by contrast, OE exhibited enhanced ^{15}N uptake at the seedling stage and superior remobilization to the kernel (Extended Data Fig. 5k–p). Consistently, an independent ^{15}N-labeling assay at the seedling stage indicated higher ^{15}N uptake in OE plants (Extended Data Fig. 5p), together supporting a role for *ZmAVT1A-1* in both nitrogen acquisition and source-to-sink remobilization.

Additional field trials corroborated these physiological mechanisms. We further evaluated *ZmAVT1A-1* using controlled pools with extreme low and normal high nitrogen levels, and fields with fertilizer applied or withheld (Supplementary Fig. 5 and Supplementary Table 12). Temporal profiling of ear-leaf relative chlorophyll content, measured as soil and plant analyzer development (SPAD) values from 20 DAP to harvest, confirmed that KO plants underwent earlier senescence with lower SPAD, with the phenotypic divergence starting at around 45 DAP; however, no significant differences were observed under HN (Fig. 5a and Extended Data Fig. 5a,b). By contrast, OE underwent enhanced stay-green phenotype, which showed delayed senescence and higher SPAD, and thus maintained photosynthetic capacity longer than null-segregant controls under LN (Fig. 5c and Extended Data Fig. 5c,d).

Agronomically, the dosage of *ZmAVT1A-1* also impacted per-plant yield components. The KO(+) suffered per-plant yield penalties of 16.0–21.1% in LN field conditions (Fig. 5c,d), while OE(+), increased per-plant yield by 10.8–20.0% (Fig. 5e,f and Supplementary Fig. 6) under LN and slightly (up to 6.8%) under HN (Fig. 5e,f and Supplementary Fig. 6). Furthermore, while kernel protein concentration by the Dumas method was unchanged in KOs, it increased 3.0-5.6% in OE plants (Fig. 5g), driving a net protein content per plant (kernel weight × protein concentration) increase of up to 16.6% (Fig. 5h). While OE plants showed a slight positive trend in plot-level yield, the gains were not statistically significant. By contrast, KO(+) suffered a significant yield penalty (Supplementary Fig. 7). These data suggest that while this gene is essential for maintaining baseline yield potential, increasing its individual dosage offers only marginal productivity gains.

It is valuable to evaluate the impact of *ZmAVT1A-1* on NUE-related traits using large-scale field-trails. We compared hybrids derived from the three haplotypes (*Non_TE*, *Ye478_type* and *HZS_type*, $n = 11$, 10 and 12 hybrids in two plots, respectively, in Gongzhuling and $n = 10$, 12 and 11 hybrids in two plots, respectively, in Xiangyang) crossed with a common tester (SK, *Non_TE type*) under both HN and LN conditions at two locations (Xiangyang in central China and Gongzhuling in north-east China). The hybrid NUE index was calculated using the formula[40] $(\text{yield}_{HN} - \text{Yield}_{LN})/(N_{Input})$. At both locations, hybrids of *HZS_type* (with significantly lower *ZmAVT1A-1* expression) showed a significantly reduced NUE index compared to *Non_TE* hybrids, while the *Ye478_type* hybrids showed more modest effects (Fig. 5i).

Evolutionary trajectory and pleiotropic effects of *ZmAVT1A-1* haplotypes

Haplotype frequency analysis revealed distinct selection patterns during domestication and improvement. Teosinte populations contained only *Non_TE* and *HZS_type* haplotypes. During the transition from teosinte to landraces, the *Non_TE* haplotype declined from 89% to 59%, whereas the *HZS_type* increased from 11% to 37% (Fig. 6a and Supplementary Table 13). Given the estimated soil nitrogen levels[41] was similar across teosinte origin sites[42] (Supplementary Fig. 8), the 5.1-kb insertion defining the *HZS_type* was probably not a direct adaptation to soil N availability. The third haplotype, *Ye478_type*, emerged in landraces (4%) and increased to 12% in tropical maize and 29% in temperate maize. Concurrently, the *HZS_type* frequency decreased from 36% in tropical maize to 12% in temperate maize, while the ancestral *Non_TE* haplotype remained relatively stable postdomestication (Fig. 6a).

To dissect the phenotypic drivers of these shifts, we analyzed the CUBIC panel at low- (Xinxiang, 35° 30′ N, 113° 58′ E) and high-latitude (Yushu, 43° 42′ N, 123°36′ E) locations. Both derived haplotypes (*HZS_type* and *Ye478_type*) were associated with reduced plant height at both sites (Fig. 6b). However, the *HZS_type* exhibited a strong genotype-by-environment interaction, slightly promoting flowering at low latitude but significantly delayed it at high latitude, whereas *Ye478_type* imposed a mild flowering delay across environments (Fig. 6b,c). Similar trends were observed in the large-scale hybrid populations[43] derived from crosses between 1,428 lines and the other two testers (Zheng58 and Jing724), where *HZS_type* consistently conferred reduced height and delayed flowering, particularly at higher latitudes (Supplementary Fig. 9). Validating these effects, *ZmAVT1A-1* KO plants showed reduced plant height (Fig. 6d) and delayed flowering compared to wild type across two environments (Fig. 6e).

Transcriptome profiling further suggested that these pleiotropic phenotypes are accompanied by changes in signaling pathways. Among

https://doi.org/10.1038/s41588-026-02655-2

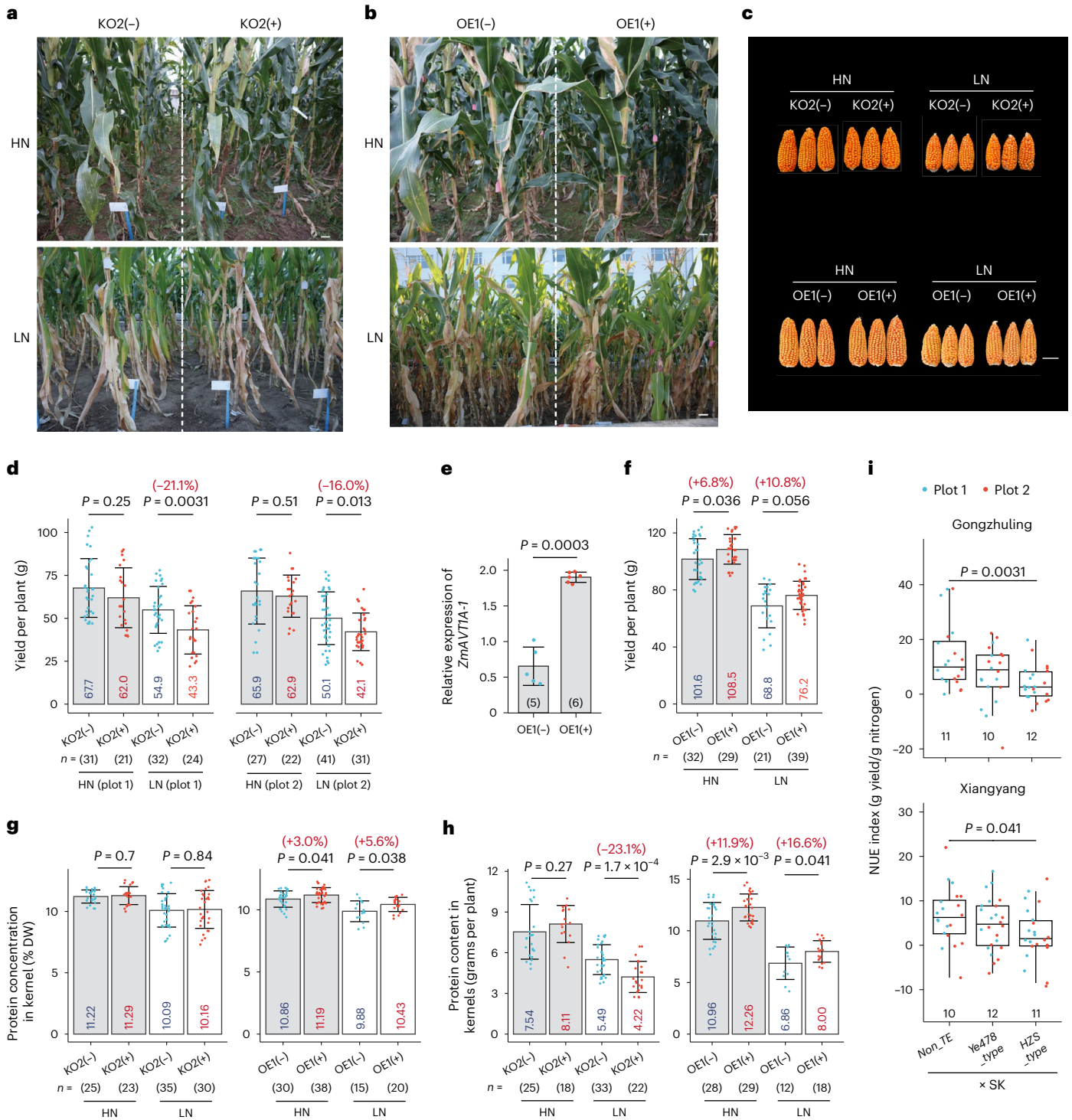


Fig. 5 | Field performance for *ZmAVT1A-1*. **a**, Phenotypes of bottom leaves for KO lines in pools. **b**, Phenotypes of bottom leaves for OE lines in pools. **c**, Representative ears of transgenic plants in LN and HN conditions. Scale bar in **a**–**c**, 5 cm. **d**, The yield per plant for KO lines. **e**, RT–qPCR analysis of *ZmAVT1A-1* expression in the leaves. **f**, The yield per plant for OE lines. The number in each bar denotes the average yield for each genotype. **g**, Kernel protein concentration for transgenic plants. The number in each bar denotes the average yield for each genotype. DW, dry weight. **h**, The total kernel protein content for transgenic plants. The number in each bar denote the average yield for each genotype. **i**, NUE index of hybrids evaluated at Gongzhuling and Xiangyang. The box plots show the median and IQR (25–75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR. In **d**–**h**, data are presented as mean ± s.d. A minus sign denotes null-segregant controls (siblings lacking the edit/transgene) and a plus sign denotes plants carrying the KO allele or OE transgene. The numbers inside bars indicate the mean value for each genotype. Red numbers above paired bars indicate the percent decrease for KO(+) relative to KO(−) or the percent increase for OE(+) relative to OE(−). In **d**–**i**, *P* values were calculated using two-tailed Student's *t*-tests.

differentially expressed genes, the upregulation of *ZmBAK1* (brassinosteroid insensitive 1-associated receptor kinase like 1) and downregulation of *ZmBKI1* (brassinosteroid insensitive 1-kinase interacting protein) are consistent with altered brassinosteroid signaling that could contribute to reduced height[44]. Furthermore, upregulation of the circadian clock regulator *ZmPRR7* (pseudoresponse regulator7)[45]

https://doi.org/10.1038/s41588-026-02655-2

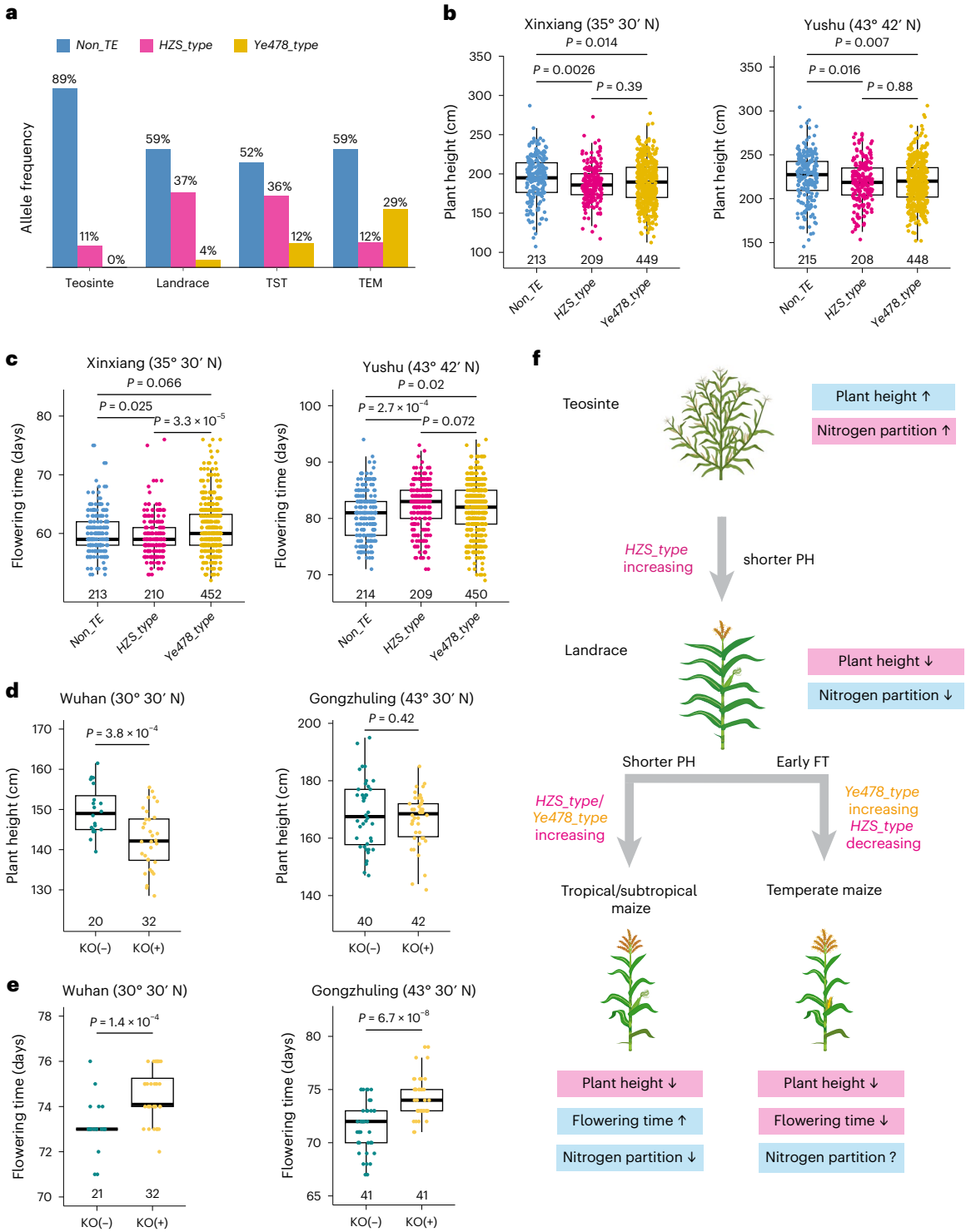


Fig. 6 | Evolution and domestication of three *ZmAVT1A-1* haplotypes. **a**, Haplotype frequency in teosintes, landraces and tropical and subtropical (TST) and temperate (TEM) maize inbred lines. **b**, The effects of three haplotypes (*Non_TE*, *HZS_type* and *Ye478_type*) on plant height in Henan (lower latitude) and Jilin (higher latitude). **c**, The effects of three haplotypes on maize flowering time in Henan and Jilin. **d**, The plant height of *ZmAVT1A-1* KO plants and their null-segregant controls evaluated in Hubei and Jilin. KO(+) denotes KO individuals and KO(−) denotes null-segregant controls. The numbers above the *x* axis indicate sample size. **e**, The flowering time of KO plants and null-segregant controls evaluated in Hubei and Jilin. KO(+) denotes KO individuals and KO(−) denotes null-segregant controls. In **b**–**e**, the box plots show the median and IQR (25–75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR. *P* values were calculated using two-tailed Student's *t*-tests. **f**, Proposed working model summarizing putative trade-offs associated with *ZmAVT1A-1* haplotypes during domestication and improvement; the nitrogen-related effect of *Ye478_type* is unresolved (question mark). Traits in pink boxes indicate putatively favorable directions for breeding, whereas traits in blue boxes indicate potentially unfavorable directions. Landrace and maize illustrations in **f** adapted with permission from Depositphotos (https://www.depositphotos.com/). Teosinte illustration in **f** created in BioRender; Jin, M. https://biorender.com/cp9ck9i (2026).

was consistent with the observed flowering delay. Nitrogen metabolism genes (*ZmNRT2.5* and *ZmGLN3*) were also differentially expressed (Supplementary Figs. 10 and 11).

Together, these results support a model of balancing nitrogen partitioning, flowering time and plant architecture (Fig. 6f). We propose that *HZS_type* was initially favored for its height-reducing effects despite costs to nitrogen partition. Subsequently, *Ye478_type* was favored in temperate breeding programs as it decoupled height reduction from severe flowering delays, optimizing the trade-off for high-latitude adaptation.

Discussion

Plant primary metabolism is a complex, highly coordinated network influenced by multiple factors[46]. While previous metabolomics studies combined with genomic information identified numerous genetic factors affecting metabolite abundance[19,22,24,47], limitations in biparental or natural populations hindered a full understanding. This study used an innovative synthetic CUBIC population, combining advantages of both population types, to investigate the genetic underpinnings of maize metabolite variation in two tissues. Integrating GWAS, transcriptomics and metabolic pathway information revealed that metabolites serve as integrative readouts of physiological status, reflecting trade-offs between growth, development and resource allocation, rather than acting merely as biochemical end products[16,19,20,48,49].

We constructed a multiomics map linking genomic variants to the primary metabolism via transcriptome as intermediates, which allowed us to pinpoint high-confidence candidate genes. Among these, *ZmASN2*, *ZmMGL* and *ZmMIPS1* were functionally verified, confirming the predictive power of our network. Metabolites may further act as signaling molecules that trigger downstream regulatory cascades. As a consequence, our multiomics network may additionally offer a metabolic roadmap for discovering novel genes and decoding the genetic architecture of complex traits. Although rarely acting in isolation to drive complex traits such as yield, the network prioritized hub genes that served as crucial entry points for dissecting metabolic networks.

A prime example was *ZmAVT1A-1*, a predicted amino acid transporter. Through ^{15}N-labeling experiments and multienvironment field trials with transgenic plants and hybrids, we demonstrated that *ZmAVT1A-1* regulated nitrogen partitioning between vegetative tissues and kernels, rather than acting as a simplistic determinant of total yield. The gene exhibits clear pleiotropy, influencing plant height, flowering time and senescence. These effects mirror those of orthologous transporters in *Arabidopsis*[50], wheat[51] and rice[52–54], suggesting that *ZmAVT1A-1* coordinates internal nitrogen distribution with developmental phase transitions, potentially via phytohormone signaling pathways. These parallels suggested that amino acid transporters act as pleiotropic hubs, modulating a diverse array of developmental and agronomic phenotypes. Our findings highlighted that this gene family was a largely untapped resource in crop breeding, warranting a re-evaluation of their role in driving yield and stress resilience.

Notably, alleles with strong loss-of-function (for example, KO or the *HZS_type* haplotype) exerted stronger phenotypic effects than gene-modified OE lines. This observation aligns with recent perspectives emphasizing that robust yield improvements from single-gene manipulations are rare and difficult to validate in the field[55]. Instead of viewing *ZmAVT1A-1* as a standalone solution, our data suggested a working model where its variation constrained agronomic performance through pleiotropic trade-offs. Practically, this provided a framework for breeders to purge unfavorable haplotypes or to use *ZmAVT1A-1* as a component in gene-stacking strategies, combining it with other beneficial alleles to engineer synergistic improvements in complex traits.

Limitations of the study

While our multiomics framework identifies key regulators of nitrogen partitioning, several limitations remain. First, while *ZmAVT1A-1* is predicted to be an amino acid transporter, its precise biochemical activity remains to be validated. Although many transporters exhibit broad specificity, functional constraints often arise from tissue-specific expression or environmental responses[50,56]. Future heterologous expression studies (for example, in *Xenopus* oocytes) are required to definitively determine its substrate range and transport kinetics. Second, the mechanistic link between *ZmAVT1A-1* and nitrogen responsiveness is incompletely resolved. Our isotope labeling captured nitrogen dynamics at discrete stages; more detailed temporal and spatial analyses are necessary to understand how the gene modulates redistribution under fluctuating environmental conditions. Third, the molecular basis of the haplotype effects requires further dissection. The *HZS_type* haplotype (associated with PIF/Harbinger_TIR_transposon) behaves like a KO, whereas the *Ye478_type* haplotype (insertion with L1_LINE_retrotransposon) shows distinct, context-dependent effects. Resolving how these specific sequence variations alter transcript structure or protein function remains a task for future molecular characterization. Finally, we explicitly note that this study does not establish *ZmAVT1A-1* as a direct driver of NUE per se. As noted previously, translating single-gene effects into agronomic NUE gains is nontrivial[55]. Consistent with this, although our field data suggest a limited, environment-dependent benefit under low nitrogen, the plot-level yield differences were small and will require additional replication across environments and genetic backgrounds. Future work must incorporate multienvironment trials in diverse hybrid backgrounds to evaluate whether *ZmAVT1A-1* can contribute to stable yield improvements when stacked with other QTL or genes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02655-2.

References

1. Xiao, Y., Liu, H., Wu, L., Warburton, M. & Yan, J. Genome-wide association studies in maize: praise and stargaze. *Mol. Plant* **10**, 359–374 (2017).
2. Liu, Q. et al. Improving crop nitrogen use efficiency toward sustainable green revolution. *Annu. Rev. Plant Biol.* **73**, 523–551 (2022).
3. Liang, Y., Liu, H. J., Yan, J. & Tian, F. Natural variation in crops: realized understanding, continuing promise. *Annu. Rev. Plant Biol.* **72**, 357–385 (2021).
4. Wurtzel, E. T. & Kutchan, T. M. Plant metabolism, the diverse chemistry set of the future. *Science* **353**, 1232–1236 (2016).
5. Medeiros, D. B., Brotman, Y. & Fernie, A. R. The utility of metabolomics as a tool to inform maize biology. *Plant Commun.* **2**, 100187 (2021).
6. Fàbregas, N. & Fernie, A. R. The interface of central metabolism with hormone signaling in plants. *Curr. Biol.* **31**, R1535–R1548 (2021).
7. Alseekh, S. et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat. Methods* **18**, 747–756 (2021).
8. Zampieri, M., Sekar, K., Zamboni, N. & Sauer, U. Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* **36**, 15–23 (2017).
9. Scossa, F., Alseekh, S. & Fernie, A. R. Integrating multi-omics data for crop improvement. *J. Plant Physiol.* **257**, 153352 (2021).
10. Zhu, G. et al. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261 (2018).
11. Fang, C., Fernie, A. R. & Luo, J. Exploring the diversity of plant metabolism. *Trends Plant Sci.* **24**, 83–98 (2019).

12. Fernie, A. R. & Tohge, T. The genetics of plant metabolism. *Annu. Rev. Genet.* **51**, 287–310 (2017).
13. Li, D. & Gaquerel, E. Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. *Annu. Rev. Plant Biol.* **72**, 867–891 (2021).
14. Luo, J. Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* **24**, 31–38 (2015).
15. Shen, S., Zhan, C., Yang, C., Fernie, A. R. & Luo, J. Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. *Mol. Plant* **16**, 43–63 (2023).
16. Chen, W. et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721 (2014).
17. Alseekh, S. et al. Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* **27**, 485–512 (2015).
18. Keurentjes, J. J. Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* **12**, 223–230 (2009).
19. Zhang, F. et al. Genomic basis underlying the metabolome-mediated drought adaptation of maize. *Genome Biol.* **22**, 260 (2021).
20. Bai, Y. et al. Natural history-guided omics reveals plant defensive chemistry against leafhopper pests. *Science* **375**, eabm2948 (2022).
21. Carreno-Quintero, N., Bouwmeester, H. J. & Keurentjes, J. J. Genetic analysis of metabolome–phenotype interactions: from model to crop species. *Trends Genet.* **29**, 41–50 (2013).
22. Wu, S. et al. Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. *PLoS Genet.* **12**, e1006363 (2016).
23. Liu, H. J. et al. CUBIC: an atlas of genetic architecture promises directed maize improvement. *Genome Biol.* **21**, 20 (2020).
24. Wen, W. et al. Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell* **27**, 1839–1856 (2015).
25. Li, K. et al. Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *Plant J.* **99**, 216–230 (2019).
26. Wen, W. et al. An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J.* **93**, 1116–1128 (2018).
27. Huang, Y. et al. THP9 enhances seed protein content and nitrogen-use efficiency in maize. *Nature* **612**, 292–300 (2022).
28. Shukla, S., VanToai, T. T. & Pratt, R. C. Expression and nucleotide sequence of an INS (3) P1 synthase gene associated with low-phytate kernels in maize (*Zea mays* L.). *J. Agric. Food Chem.* **52**, 4565–4570 (2004).
29. Wu, Y., Wang, W. & Messing, J. Balancing of sulfur storage in maize seed. *BMC Plant Biol.* **12**, 77 (2012).
30. Gent, L. & Forde, B. G. How do plants sense their nitrogen status? *J. Exp. Bot.* **68**, 2531–2539 (2017).
31. Liu, H. et al. Genome-wide analysis of the AAAP gene family in moso bamboo (*Phyllostachys edulis*). *BMC Plant Biol.* **17**, 29 (2017).
32. Zhao, H., Ma, H., Yu, L., Wang, X. & Zhao, J. Genome-wide survey and expression analysis of amino acid transporter gene family in rice (*Oryza sativa* L.). *PLoS ONE* **7**, e49210 (2012).
33. Fujiki, Y. et al. Functional identification of AtAVT3, a family of vacuolar amino acid transporters, in *Arabidopsis*. *FEBS Lett.* **591**, 5–15 (2017).
34. Stelpflug, S. C. et al. An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* **9**, plantgenome2015.04.0025 (2016).
35. Wang, B. et al. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nat. Genet.* **55**, 312–323 (2023).
36. Liu, C. et al. A dual-subcellular localized β-glucosidase confers pathogen and insect resistance without a yield penalty in maize. *Plant Biotechnol. J.* **22**, 1017–1032 (2024).
37. Guo, N. et al. Oryza sativa Lysine-Histidine-type Transporter 1 functions in root uptake and root-to-shoot allocation of amino acids in rice. *Plant J.* **103**, 395–411 (2020).
38. The, S. V., Snyder, R. & Tegeder, M. Targeting nitrogen metabolism and transport processes to improve plant nitrogen use efficiency. *Front. Plant Sci.* **11**, 628366 (2020).
39. Tegeder, M. & Masclaux-Daubresse, C. Source and sink mechanisms of nitrogen transport and use. *New Phytol.* **217**, 35–53 (2018).
40. Congreves, K. A. et al. Nitrogen use efficiency definitions of today and tomorrow. *Front. Plant Sci.* **12**, 637108 (2021).
41. Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).
42. Chen, L. et al. Genome sequencing reveals evidence of adaptive variation in the genus Zea. *Nat. Genet.* **54**, 1736–1745 (2022).
43. Xiao, Y. et al. The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* **22**, 148 (2021).
44. Wang, D. et al. BKI1 regulates plant architecture through coordinated inhibition of the brassinosteroid and ERECTA signaling pathways in *Arabidopsis*. *Mol. Plant* **10**, 297–308 (2017).
45. Zhao, Y. et al. The evening complex promotes maize flowering and adaptation to temperate regions. *Plant Cell* **35**, 369–389 (2023).
46. Galili, G., Amir, R. & Fernie, A. R. The regulation of essential amino acid synthesis and accumulation in plants. *Annu. Rev. Plant Biol.* **67**, 153–178 (2016).
47. Schauer, N. et al. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447–454 (2006).
48. Riedelsheimer, C. et al. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl Acad. Sci. USA* **109**, 8872–8877 (2012).
49. Li, J. et al. Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science* **371**, 255–260 (2021).
50. Dinkeloo, K., Boyd, S. & Pilot, G. Update on amino acid transporter functions and on possible amino acid sensing mechanisms in plants. *Semin. Cell Dev. Biol.* **74**, 105–113 (2018).
51. Chen, H. et al. Amino acid transporter gene TaATLa1 from *Triticum aestivum* L. improves growth under nitrogen sufficiency and is down regulated under nitrogen deficiency. *Planta* **256**, 65 (2022).
52. Peng, B. et al. OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice. *Nat. Commun.* **5**, 4847 (2014).
53. Lu, K. et al. Blocking amino acid transporter OsAAP3 improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol. J.* **16**, 1710–1722 (2018).
54. Ding, L., Huang, W., Li, Z. & Fang, Z. Amino acid transporter OsATL13 coordinately regulates rice yield and quality by transporting phenylalanine and methionine. *Plant Sci.* **353**, 112398 (2025).
55. Khaipho-Burch, M. et al. Genetic modification can improve crop yields—but stop overselling it. *Nature* **621**, 470–473 (2023).
56. Yao, X. et al. Disruption of the amino acid transporter CsAAP2 inhibits auxin-mediated root development in cucumber. *New Phytol.* **239**, 639–659 (2023).

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China. [2]Hubei Hongshan Laboratory, Wuhan, China. [3]Yazhouwan National Laboratory, Sanya, China. [4]Guangdong Key Laboratory for Crop Germplasm Resources Preservation and Utilization, Guangzhou, China. [5]Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, China. [6]Department of Molecular Physiology, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany. [7]Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria. [8]Institute of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun, China. [9]Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Beijing Academy of Agricultural and Forestry Sciences, Beijing, China. [10]These authors contributed equally: Min Jin, Shijuan Yan. ✉e-mail: yjianbing@mail.hzau.edu.cn

Methods

Plant materials and growth conditions

The CUBIC population was derived from 24 genetically diverse maize inbred lines[23] and includes 1,404 progenies. The parents and progeny were planted in Sanya (18° 15′ N 109° 30′ E) in winter 2015 for mature kernel harvest and winter 2016 for leaf sampling. Each line was planted in a randomized design in a single 2.5-m row field plot. In 2015, plants were self-pollinated and three ears per line were randomly selected for mature kernel sampling. In 2016, the tenth leaves from three V9-stage plants per line were pooled and immediately frozen in liquid nitrogen. Gene-edited plants were grown in Sanya in winter 2020 and 2021 and sampled using the same procedures. All samples were stored at −80 °C before metabolite extraction.

Metabolite extraction and metabolome profiling with GC–MS

Primary metabolites were extracted from maize leaves and kernels as described previously[57]. In brief, 50 mg leaf powder was extracted with 1 ml methanol:methyl-*tert*-butyl-ether (MTBE) solution (1:3, v/v). A 300-µl aliquot of the lower polar phase was dried in a SpeedVac concentrator without heating, derivatized with *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) and analyzed by GC–MS (7890A-5975C, Agilent)[58]. Samples were injected at 270 °C in split mode (50:1) with helium carrier gas at 1 ml min^{-1} and separated on a DB-35MS UI capillary column (30 m × 0.25 mm, 0.25 µm). The oven temperature was held at 90 °C for 4 min, increased to 205 °C at 8 °C min^{-1}, held for 2 min, then increased to 310 °C at 15 °C min^{-1} and held for 2 min. The transfer line and ion source temperatures were 300 °C and 230 °C, respectively, and the mass range was m/z 85–700. GC–MS data were processed using Agilent MassHunter Qualitative Analysis B06.00 and Quantitative Analysis B.07.01[59]. Metabolites were identified using the NIST library and an in-house database of authentic standards. In total, 67 leaf and 68 kernel metabolites were confirmed by authentic standards and used for further analysis; other metabolites were annotated using public databases (Supplementary Table 1). Metabolite intensities were normalized to the internal standard (^{13}C-ribitol) to calculate relative contents.

Genotype data

Genotype data for the CUBIC population were obtained by sequencing on the Illumina HiSeq 2500 platform. After read alignment, SNP calling and filtering, 11.8 M SNPs from whole-genome resequencing of the panels were obtained from a previously reported analysis of the CUBIC population[23]. The dentity by descent (IBD) mosaic map of all CUBIC lines was collapsed into 27,005 bins with a hidden Markov model[60].

Genome-wide association analysis and mQTL identification

sGWAS and IBD-based GWAS (hGWAS) were performed as below[23]. For sGWAS, high-quality SNPs with minor allele frequency ≥2% were analyzed using a mixed linear model in TASSEL 3.0[61] to control for population structure. The significance threshold was set as $1/N$, where N represents the number of independent tests. Significant SNPs were grouped into loci if consecutive SNPs were <20 kb apart and each locus contained at least two significant SNPs. Adjacent loci were merged when significant SNPs between them showed linkage disequilibrium ($r^2 \geq 0.2$). Nearby loci were further combined, and the SNP with the lowest P value was defined as the peak SNP. Each significant locus was defined as an sQTL, with its interval determined by the span of significant SNPs.

For hGWAS, 27,005 IBD bins were analyzed using a mixed linear model with restricted maximum likelihood. Bin significance was assessed by likelihood ratio test, and the average likelihood ratio test (LRT) threshold of 7.1 was determined by 500 permutations. Significant bins within ≤1 Mb or ≤5 bins were merged into one hQTL, and the hQTL interval was defined by the physical range of significant bins. In this study, both sQTLs and hQTLs were collectively referred to as mQTLs.

mQTL hotspot identification and enrichment analysis

Potential mQTL hotspots were identified using a 3-Mb sliding-window analysis. The number of mQTLs in each window was counted based on the peak SNP position or the midpoint of the peak bin. To determine the hotspot threshold, 100 permutation tests were performed for each chromosome by randomly assigning mQTLs to their corresponding chromosomes and recalculating the mQTL number in each window. The maximum mQTL count across the ten chromosomes was recorded in each permutation, and the highest value from 100 permutations was used as the hotspot threshold. Windows containing more than ten mQTLs were defined as hotspots. KEGG pathway, module and role enrichment analyses of metabolites within each hotspot were performed using MBROLE 2.0, with FDR <0.01 considered significant[62].

Transcriptomic data and eQTL analysis

At the V9 stage (when leaf tissue growth is fastest), total RNA was extracted from the eleventh leaf from a subset of 391 randomly selected progenies[28]. Illumina sequencing (150 bp, paired end) was performed using HiSeq X Ten protocols and reads were mapped to the B73 AGPv3.25 reference genome. Expression data for all genes were normalized for subsequent analysis using DESeq2 software. Genes expressed in >60% of the lines were retained for eQTL mapping with EMMAX software, using 1.89×10^{-8} as the significance threshold. eQTLs and mQTLs were considered to be colocalized if the peak variant of an mQTL was included among the significant SNPs of an eQTL.

Network construction for the genome, transcriptome and metabolome

mQTL signals were used to connect metabolites with genetic loci, and eQTL signals for genes within the corresponding regions were used to link loci with candidate genes. Genes located in the overlapping regions of their eQTL intervals and colocalized sQTL intervals were considered potential mediators between genetic loci and metabolites, linking genomic and metabolic variation through gene expression.

For transcriptome data were missing in the kernel, genes matching one of three criteria were prioritized: (1) located within 800 kb (the linkage disequilibrium decay distance is 800 kb when $r^2 = 0.1$ in CUBIC)[23] of the peak SNP and within the two steps of the metabolic reaction according to KEGG information, (2) genes whose expression was correlated with the corresponding metabolite content and that were within the two steps of the metabolic reaction according to KEGG information and (3) genes with local-eQTL. Finally, a network containing 409 loci, 1,630 genes and 114 metabolic traits was constructed; the network was displayed using Cytoscape software[63].

Phylogenetic analysis

Protein sequences of *AVTs* from maize were obtained from Gramene (http://www.gramene.org). Sequences of full-length proteins were aligned with Clustal W[64], and a neighbor-joining phylogenetic tree was constructed using MEGA 11 with 500 bootstrap replicates. Maize, *Arabidopsis* (https://www.arabidopsis.org/) and rice (http://rice.plantbiology.msu.edu/) genes were named according to previous publications[31,32].

Expression profile analysis

Expression data were obtained from the published studies. One study was focused on dynamic kernel development with spatiotemporal transcriptome atlas of B73 maize seed with 53 samples from fertilization to maturity for embryo, endosperm and whole seed tissues[39] and the other one is the expanded B73 gene atlas, comprising 79 diverse tissues representing 11 major organ systems and varying developmental stages of the maize plant[34].

Identification of TE insertions in *ZmAVT1A-1*

Based on the published maize genome[35,65], the sequences of *ZmAVT1A-1* were downloaded and aligned by the BLAST program. According to the

insertion site by TEs identified, PCR markers (Supplementary Table 14) were developed to identify the parental variations. PCR amplification was performed using 2× Phanta Max Master Mix (Vazyme), following the manufacturer's protocol. Combining the raw reads of each line with next-generation sequencing[42,66,67], the haplotypes with different insertions were identified based on the coverage pattern.

Design of CRISPR–Cas9 constructs and generation of transgenic plants

To confirm the effects of selected candidate genes on target traits, we edited their sequences using the CRISPR–Cas9 system. Guide RNAs targeting the gene coding sequences were designed with CRISPR-Local[68]. One target for each gene was integrated into the highly efficient BGK053 vector with the U6 promoter, guide RNA sequence and Cas9 coding sequence. The coding sequence of *ZmAVT1A-1* was amplified from B73 cDNA using the primers (TsingKe Biotech) listed in Supplementary Table 14 and cloned into the *pZZ01* vector using the ClonExpress II One Step Cloning Kit (Vazyme) to generate UBI::*ZmAVT1A-1-GFP* verified by Sanger sequencing. The recombinant vectors were transformed into immature embryos of KN5585 via *Agrobacterium* infection by Wimi Biotechnology[69].

Primary edited plants were screened by PCR and sequencing of the target region, and heterozygous edited individuals were self-pollinated. In the subsequent generation(s), plants were genotyped to identify individuals carrying the desired edited allele but lacking the Cas9 transgene (Cas9-free), which were then advanced by selfing for two to three generations before phenotyping. OE was verified by RT–qPCR in leaf tissue collected at the indicated jointing stage, using *ZmActin* as the internal reference. OE events showing increased expression relative to its sibling controls were retained for downstream phenotyping.

To eliminate potential confounding effects from background heterogeneity or somaclonal variation induced by tissue culture, phenotypic comparisons were performed within each independent transformation event using segregating sibling controls (null segregants). Specifically, for each event-derived family, progenies were genotyped and classified as '+' (carrying the edited allele for KO lines, or the OE construct for OE lines) or '−' (null segregants lacking the corresponding edit/transgene). Phenotyping was conducted by comparing KO(+) versus KO(−) or OE(+) versus OE(−) individuals derived from the same event and advanced through the same generations, thereby controlling for event-specific background and regeneration effects. The primers used to detect the genotypes are listed in Supplementary Table 14.

Subcellular localization of *ZmAVT1A-1*

The *ZmAVT1A-1* coding sequence was PCR amplified and cloned into a ubi promoter-driven GFP expression vector using the ClonExpress II One Step Cloning Kit (Vazyme) to generate the ZmAVT1A-1-GFP fusion construct. The plasmid was transformed into protoplasts according to a protocol developed by Jen Sheen (http://genetics.mgh.harvard.edu/sheenweb/). After transformation, protoplasts were incubated in the dark at room temperature for 14 h and GFP fluorescence was observed using a confocal microscope (Zeiss LSM980).

Functional complementation of *ZmAVT1A-1* in yeast

Full-length *ZmAVT1A-1* cDNA was amplified and cloned into yeast expression vector pYES2. The constructs along with the empty vector were transformed into yeast mutant strain 22Δ10α (ref. 37) as negative and *OsAUX5* (ref. 70) transformed as positive control. The heat shock method was used for yeast transformation[71]. Transformants were selected on solid agar with yeast nitrogen base (without amino acids or NH_4^+) media without uracil, supplemented with 5 g l^{-1} $(NH_4)_2SO_4$ and 20 g l^{-1} galactose. Transformed yeast cells were incubated in the shaker at 28 °C and 200 rpm to an OD_{600} of 1.0. Cells were collected by centrifugation, washed with sterilized water and diluted to an OD_{600} of 0.1, 0.01 or 0.001. Drops of 3 µl were aligned on uracil-free yeast nitrogen base media containing 1 mM of individual amino acids or $(NH_4)_2SO_4$ as the sole N source. Plates were incubated at 28 °C for 2 days.

Differentially expressed gene analysis

Leaves at joining were sampled in the field for the KO event of *ZmAVT1A-1*, for RNA-sequencing the Illumina NovaSeq 6000 platform was used for sequencing with paired-end reads generation. Clean data of RNA-sequencing were obtained and compared to the maize reference genome (B73 AGPv3.25) by RSEM v1.3.3[72] combined with Bowtie v.2.4.1 (ref. 73) and the expression levels were calculated with maize annotation information. Then the differentially expressed genes were calculated by EBSeq[74] and the differentially expressed genes were selected with significances of $P < 0.05$ and ≥1.5-fold change for the KO lines.

Pot experiment and ^{15}N labeling

In 2022, the KO(−) and KO(+) lines of *ZmAVT1A-1* were planted in pots in Wuhan (30° 30′ N, 114° 18′ E) under two nitrogen levels. For HN soil, 1.6305 g urea, 1.324 g KCl and 4.609 g $Ca(H_2PO_4)_2$ were added to 3 kg Pindstrup substrate for each plant. For LN soil, no urea and 1.324 g KCl and 4.609 g $Ca(H_2PO_4)_2$ were added to the same soil. At the jointing stage, 5 g urea was added to the HN soil. At the same time, the KO(−) and KO(+) lines of *ZmAVT1A-1* were planted in Gongzhuling (43° 30′ N, 124° 48′ E) under two nitrogen conditions. At the same time, for HN, 7.98 g m^{-2} nitrogen fertilizer was applied in the field and for LN, no nitrogen was applied. Open-pollinated ears were harvested.

In winter 2022, we planted *ZmAVT1A-1* OE(−) and OE(+) lines in pots with two nitrogen levels in Sanya (18° 15′ N, 109° 30′ E). For HN, 3.125 g urea was added to 10 kg soil per pot when seeds were sown and no nitrogen was added for LN. At the V5 stage, 1 g urea was added to the HN soil and 0.25 g (25%) urea to the LN soil. At the V10 stage, 2 g urea was added to the HN soil and 0.5 g (25%) urea to the LN soil. Open-pollinated ears were harvested.

In winter 2022, KO(−), KO(+), OE(−) and OE(+) plants were grown in pots at the same site under HN and LN conditions. Five plants with similar growth status per genotype and nitrogen treatment were selected for ^{15}N labeling at the jointing stage (V10). HN pots received 2 g of ^{15}N-labeled urea (10 atom% ^{15}N excess) and LN pots received 0.5 g of ^{15}N-labeled urea. Labeled plants were harvested at 20 days after self-pollination and separated into kernels and vegetative tissues for total nitrogen and ^{15}N measurement.

In winter 2024, we planted *ZmAVT1A-1* OE segregating lines in pots with two nitrogen levels in Sanya (18° 15′ N, 109° 30′ E). For HN, 3.0 g ^{15}N-labeled urea (at 10% ^{15}N atom excess) was added to 10 kg soil per pot when seeds were sown and 0.3 g ^{15}N-labeled urea (at 10% ^{15}N atom excess) for LN. At the V5 stage, plants were harvested and split into shoot and root before dried to constant weight and grinded for ^{15}N measurement.

Biomass, N content and ^{15}N measurements

All ^{15}N-labeled plants were harvested at 20 DAP and divided into leaves and stems on the basis of their positions relative to the ear. The samples were heated at 105 °C for 30 min and then dried to a constant weight at 70 °C. After dry weight measurement, the samples were ground into powder for N and ^{15}N quantification by isotope ratio mass spectrometry (Elementar isoprime100). The content of ^{15}N in each sample was calculated as: N% × biomass × ^{15}N abundance.

Nitrogen-depletion investigation

Experiment in nitrogen depletion pools built with artificial filled soil was conducted in Gongzhuling (43° 30′ N, 124° 48′ E) from May to October in 2024 under two nitrogen conditions: normal nitrogen fertilizer (7.98 g m^{-1}) and no nitrogen fertilizer, which has been used to identify high NUE maize treated without nitrogen fertilizer in the past 10 years. Three rows for KO, OE lines and their corresponding wild-type sibling controls were planted in with 0.25-m spacing between individuals.

Chlorophyll contents of ear leaves were measured using a SPAD-502 Plus chlorophyll meter during the kernel-filling period from 20 DAP to harvest. During the period, photographs were taken concurrently. Statistical significance was calculated using a two-tailed, independent Student's *t*-test.

Nitrogen-associated trait investigation in field

Field trials were conducted in Gongzhuling from May to October in 2024 under two nitrogen conditions: normal nitrogen fertilizer (7.98 g m^{-2}) and no nitrogen fertilizer were applied in the field (Supplementary Table 12). The field experiment followed a randomized block design with three replicates for KO lines, OE lines and their corresponding wild-type sibling controls. Each plot consisted of four rows of 4 m with 17 plants in a single row, with 0.25-m spacing between plants and 0.6 m between rows. Individuals from each plot were harvested and air dried for measuring grain yield. Statistical significance was calculated using a two-tailed, independent Student's *t*-test.

For selected hybrid crosses, field trials were conducted in Xiangyang City from Hubei Province (32° 6′ N, 112° 12′ E) and Gongzhuling City from Jilin Province, following a randomized block design with two replicates. Each plot contained two rows with 11 plants per row (4-m row length, 0.4-m spacing between plants and 0.6 m between rows in Gongzhuling; 3-m row length, 0.3-m spacing between plants and 0.6 m between rows in Xiangyang). All ears of each plot were harvested to estimate plot yields. The NUE index was calculated using the formula[40] NUE index = $(yield_{HN} - yield_{LN})/(N_{Input})$. Where $yield_{HN}$ and $yield_{LN}$ represent plot yields under HN and LN conditions, respectively, and N_{Input} represents the nitrogen fertilizer applied under HN conditions.

Phenotype analysis for hybrids

Phenotype data (plant height and flowering time) for inbreds were collected from CUBIC populations (n = 2,856) from two locations: Xinxiang City from Henan Province (35° 30′ N, 113° 58′ E), and Yushu City from Jilin Province (43° 42′ N, 123° 36′ E)[23]. Phenotype data (plant height and flowering time) were collected from two F1 hybrid populations (n = 2,856) derived from crossing 1,428 diverse maternal lines with two tester lines Zheng58 and Jing724 (ref. 43). All lines were genotyped with the different haplotypes of *ZmAVT1A-1* and used to explore its effect in hybrids. The phenotypes were collected from two locations: Xinxiang City from Henan Province, and Shenyang City from Liaoning Province (42° 6′ N, 123° 24′ E).

Measurement of total nitrogen content in kernels

Kernels from each ear were dried at 60 °C to a constant weight and ground in a mixer mill (MM 400, Retsch). For each sample, 50–60 mg of powder was wrapped in tin foil and analyzed using a Dumas rapid nitrogen analyzer (rapid N exceed, Elementar). Five asparagine standards were included before each measurement batch and an aspartic acid standard was included every 20 samples as internal controls. Sample weights were entered into rapid N exceed software v.1.1.25 with the following settings: O_2 dosing time, 60 s; O_2 dosing flow, 225 ml min^{-1}; O_2 cutoff threshold, 15%; autozero delay, 30 s; and peak anticip, 90 s. Wrapped samples were loaded into the sample tank according to their assigned serial numbers.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Full methodological details for all statistical analyses, including specific tests applied, sample sizes and *P* values, are provided in the figures or their respective legends. Field experiments were conducted using randomized designs as described in the relevant Methods sections. Statistical significance was assessed using Student's *t*-tests and one-way analysis of variance (ANOVA) followed by least significant difference (LSD) post hoc tests in SPSS (IBM SPSS Statistics 22). All statistical results including correlation analysis and *t*-test and graphs were generated in R (v.4.3.2) (https://cran.r-project.org) and Excel.

Reporting summary

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

Data availability

Raw whole-genome sequencing and RNA sequencing reads of CUBIC population[23] are available in the Genome Sequence Archive[75] with accession numbers CRA000171 (https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA000171) and CRA001241 (https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA001241), respectively. Resequence data of the teosinte, landrace and maize can be found under BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) with accession numbers PRJNA641489 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA641489), PRJNA783885 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA783885) and PRJNA531553 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA531553), respectively. The RNA-sequencing data for KOs can be found in the Genome Sequence Archive[75] under accession number CRA042606 (https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA042606). The raw GC–MS data can be obtained at Metabolomics Workbench (https://dev.metabolomicsworkbench.org:22222/data/DRCCMetadata.php?Mode=Study&StudyID=ST003797&Access=NkqL1083). The B73 reference genomic sequences are collected from the MaizeGDB (https://www.maizegdb.org/). Gene model information is retrieved from MaizeGDB for *ZmAVT1A-1* (GRMZM2G083788), *ZmAVT1A-2* (GRMZM2G080843), *ZmAAP31* (GRMZM2G101125), *ZmASN2* (GRMZM2G093175), *ZmMGL* (GRMZM2G450498), *ZmMIPS1* (GRMZM2G155242), *ZmBKI1* (GRMZM2G035741), *ZmBAK1* (GRMZM2G145720), *ZmPRR7* (GRMZM2G095727), *ZmNRT2.5* (GRMZM2G455124) and *ZmGLN3* (GRMZM2G046601). Processed data supporting the findings of this study are provided in this Article and its Supplementary Information. Gene-editing or OE maize plants are available from the corresponding author. Source data are provided with this paper.

Code availability

All software used in the study is publicly available as described in Methods and Reporting Summary.

References

57. Salem, M. A., Jüppner, J., Bajdzienko, K. & Giavalisco, P. Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. *Plant Methods* **12**, 45 (2016).
58. Wang, H. et al. A subsidiary cell-localized glucose transporter promotes stomatal conductance and photosynthesis. *Plant Cell* **31**, 1328–1343 (2019).
59. Yan, S., Huang, W., Gao, J., Fu, H. & Liu, J. Comparative metabolomic analysis of seed metabolites associated with seed storability in rice (*Oryza sativa* L.) during natural aging. *Plant Physiol. Biochem.* **127**, 590–598 (2018).
60. Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl Acad. Sci. USA* **97**, 12649–12654 (2000).
61. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
62. López-Ibáñez, J., Pazos, F. & Chagoyen, M. MBROLE 2.0—functional enrichment of chemical compounds. *Nucleic Acids Res.* **44**, W201–W204 (2016).
63. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

64. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
65. Li, C. et al. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. *Nat. Plants* **8**, 750–763 (2022).
66. Gui, S. et al. A pan-Zea genome map for enhancing maize improvement. *Genome Biol.* **23**, 178 (2022).
67. Wang, Y. et al. Three types of genes underlying the Gametophyte factor1 locus cause unilateral cross incompatibility in maize. *Nat. Commun.* **13**, 4498 (2022).
68. Sun, J. et al. CRISPR-Local: a local single-guide RNA (sgRNA) design tool for non-reference plant genomes. *Bioinformatics* **35**, 2501–2503 (2019).
69. Liu, H. J. et al. High-throughput CRISPR–Cas9 mutagenesis streamlines trait gene identification in maize. *Plant Cell* **32**, 1397–1413 (2020).
70. Shi, Y. et al. Natural variations of OsAUX5, a target gene of OsWRKY78, control the neutral essential amino acid content in rice grains. *Mol. Plant* **16**, 322–336 (2023).
71. Suga, M. & Hatakeyama, T. A rapid and simple procedure for high-efficiency lithium acetate transformation of cryopreserved *Schizosaccharomyces pombe* cells. *Yeast* **22**, 799–804 (2005).
72. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Leng, N. et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043 (2013).
75. CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30–D44 (2025).

Acknowledgements

We thank H. Liu from the high-throughput computing platform of the National Key Laboratory of Crop Genetic Improvement.

Author contributions

J.Y. designed and supervised this study. X.W., W.L. and S.Y. prepared the materials. J.S., J.L., H.-J.L., N.Y. and Y.X. provided the genetic and transcriptomic data. M.J. and Y.C. performed the association analysis. M.J., Y.Y. and W.W. performed the data analyses. M.J., S.Y., S.A. and A.R.F. evaluated the candidate genes. M.J., S.Y., W.H., Y.Z. and J.X. performed genetic transformation and mutant validation. M.J., N.M., Y.W. and Z.Z. performed the subcellular localization experiments and field trials. M.J., Y.W., Z.Z., Y.S., K.X., K.T., C.J., J.Z., X.L. and Y.Q. performed the nitrogen-related trait experiments for *ZmAVT1A-1*. M.J. and Y.W. performed the ^{15}N-labeling experiment and protein content measurements. M.J. and Y.W. performed the agronomic trait and flowering time investigation. M.J., Y.Q., W.L., F.Q. and L.Z. performed the field experiments. M.J., S.Y., H.-J.L., A.R.F. and J.Y. prepared the paper.

Funding

This research was supported by the National Key Research and Development Program of China (grant number 2021YFF1000500 to F.Q.), the National Natural Science Foundation of China (grant number 32321005 to J.Y.), the Science and Technology Major Program of Hubei Province (grant number 2021ABA011 to J.Y.), the 111 Project Crop genomics and Molecular Breeding (grant number B20051 to J.Y.), the Science and Technology Program of Guangdong Province (grant number 2023A0505090005 to S.Y. and grant number 2021TQ06N115 to S.Y.), Innovation Fund Projects of Guangdong Academy of Agricultural Sciences (grant number 202205 to W.H.). The Research Innovation and Digitalisation for Smart Transformation (grant number BG16RFPR002-1.014-0003-C01 to S.A.) and the HORIZON-WIDERA-2022-TALENTS-01 (grant number 720101087091 to S.A.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

J.X. is an employee of WIMI Biotechnology Company. The other authors declare no competing interests.

Additional information

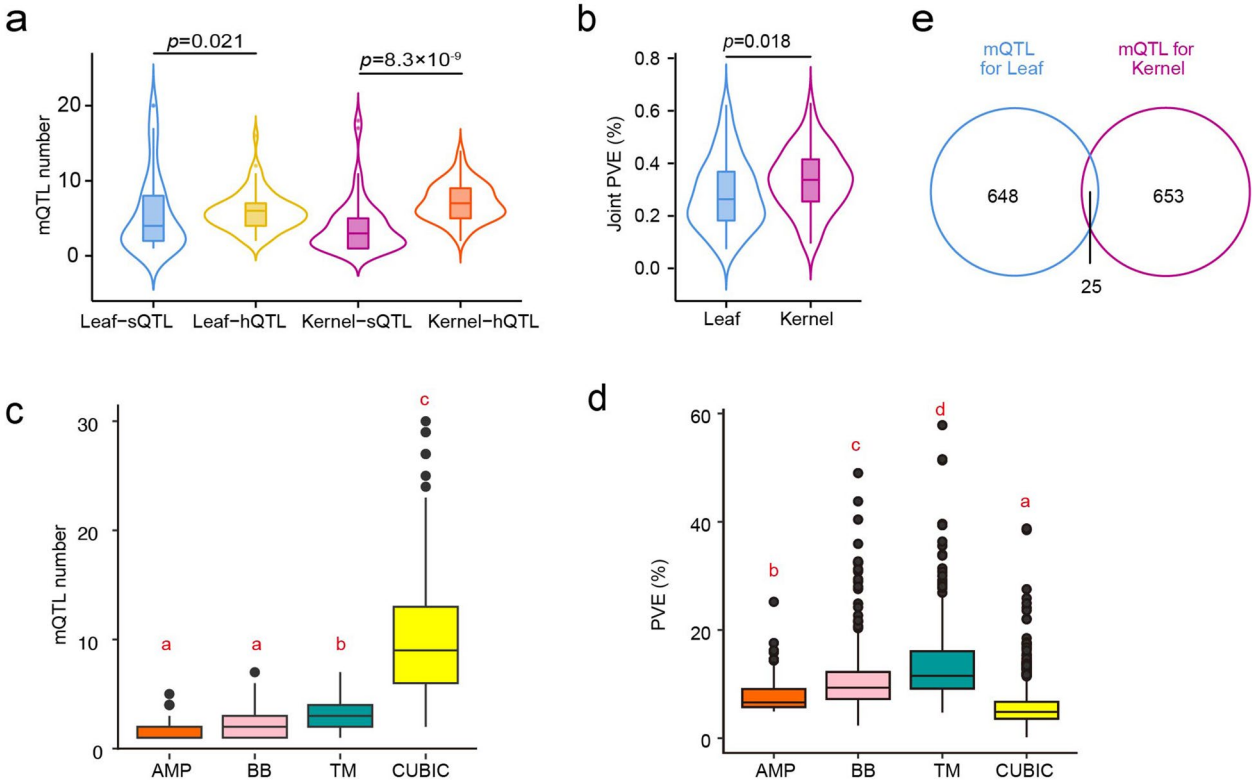
Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02655-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02655-2.

Correspondence and requests for materials should be addressed to Jianbing Yan.

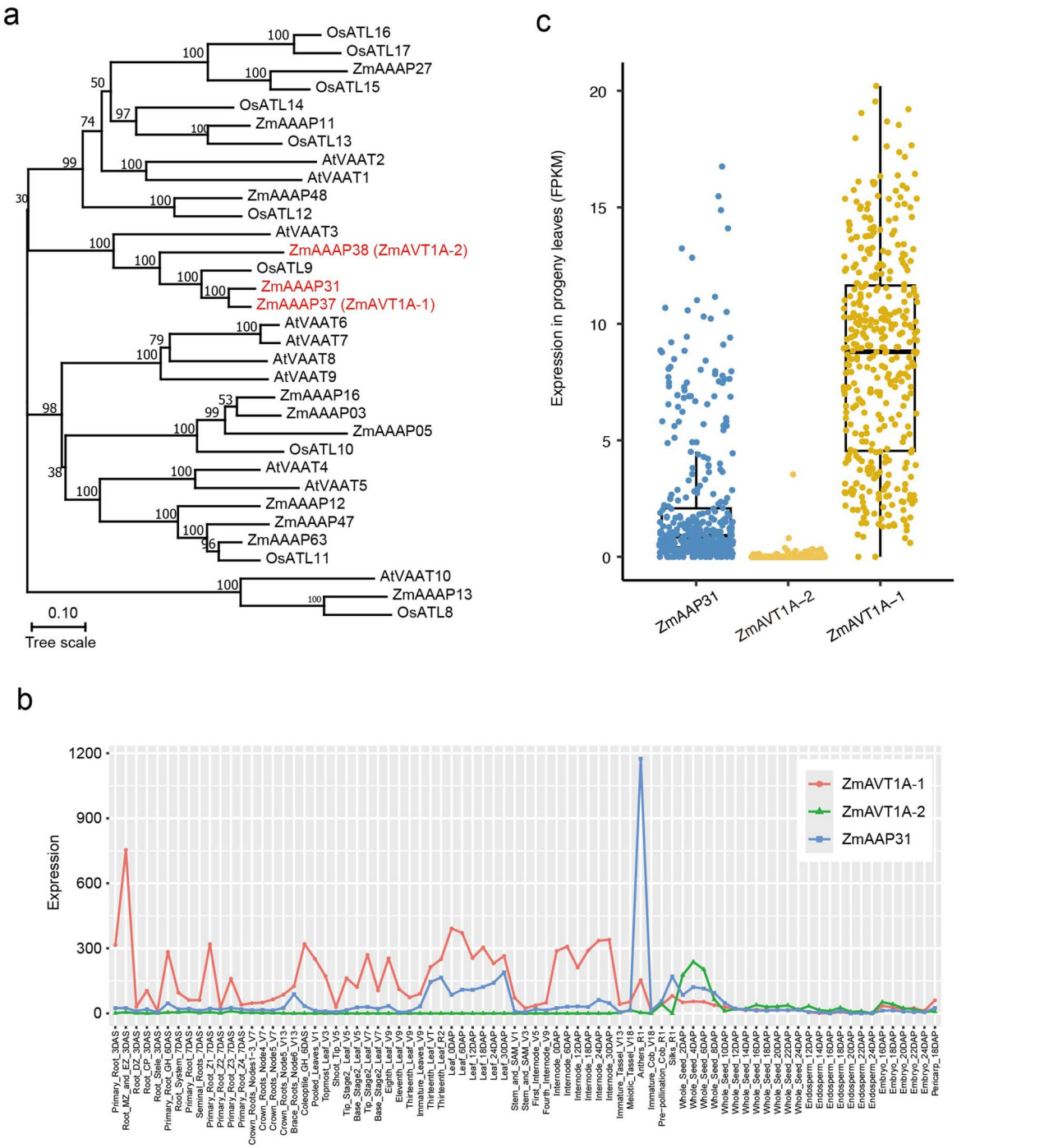
Peer review information *Nature Genetics* thanks Xiangdong Fu, Stephen Moose and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-026-02655-2



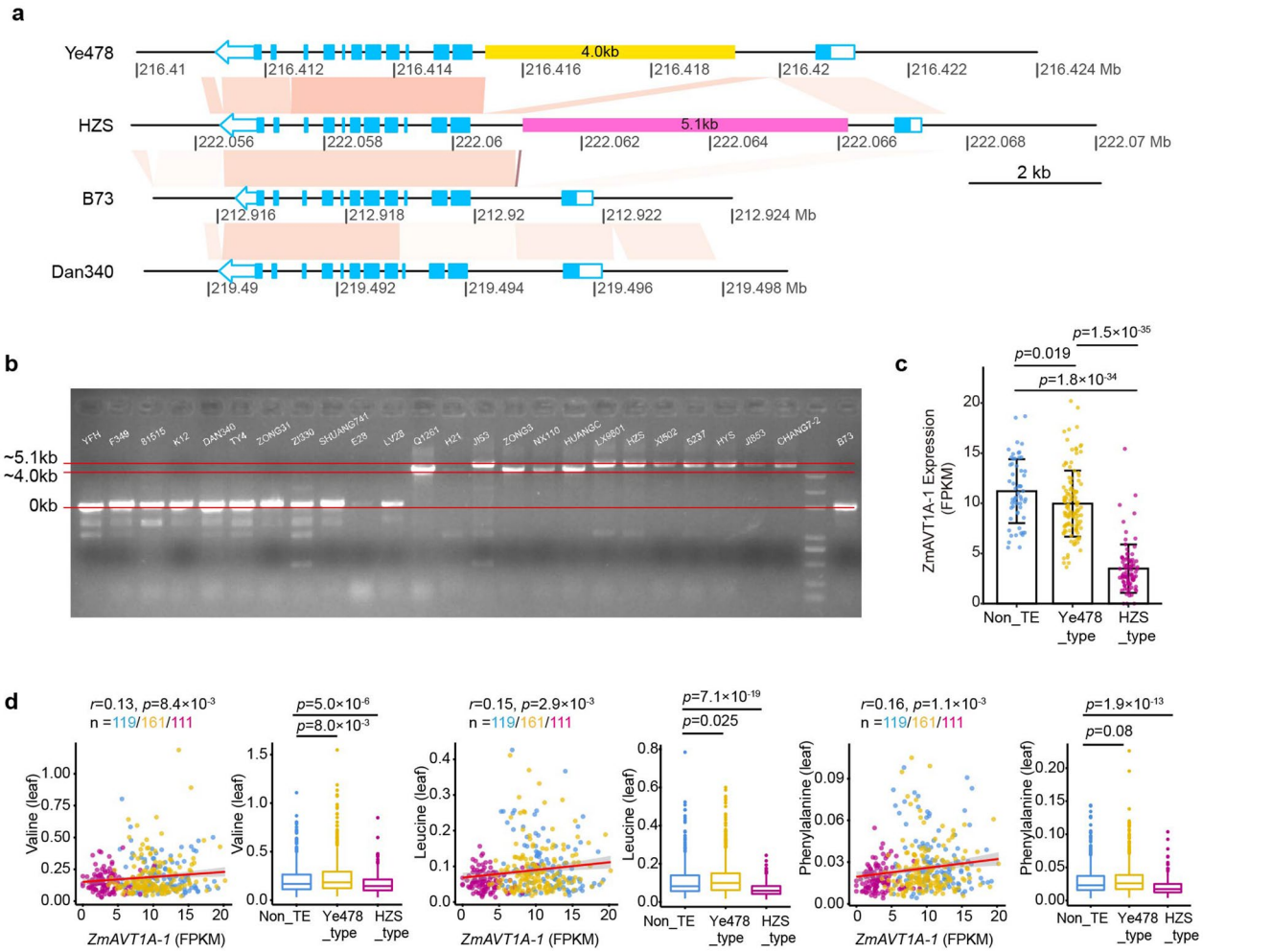
Extended Data Fig. 1 | Genetic architecture of metabolites in maize leaves and kernels across the CUBIC population. a, Numbers of mQTLs for leaves and kernels. Each point represents one metabolite. n = 67 for leaf metabolites, and 68 for kernel metabolites. **b**, Joint percentage variation explained (PVE) of mQTLs for individual metabolites in leaves and kernels. Each point represents one metabolite. n = 67 for leaf metabolites, and 68 for kernel metabolites. **c**, Comparison of mQTL number per metabolite across mapping populations. Each point represents one metabolic trait from the corresponding population[24–26]. n = 100, 152, 350 and 135 metabolic traits in AMP[24], BB[26], TM[25] and CUBIC population, respectively. **d**, Comparison of Phenotypic variance explained by individual mQTLs (PVE) across mapping populations. Each point represents one mQTL from the corresponding study or population[24–26]. n = 153, 312, 350 and 1371 mQTL in AMP[24], BB[26], TM[25] and CUBIC population, respectively. In **a**-**d**, the box plots show the median and interquartile range (25th-75th percentiles), with whiskers extending to the minimum and maximum values within 1.5×IQR. In **a**, **b**, *P* values were calculated using two-tailed Student's *t*-tests. In **c**-**d**, different red letters indicate significant difference ($P < 0.05$) based on one-way ANOVA (LSD test). AMP = association mapping population[24], BB = B73 × By804 RIL population[26], TM = Teosinte × Mo17 backcross population[25]. **e**, Venn diagram of mQTL numbers detected in leaves, kernels, or both.

https://doi.org/10.1038/s41588-026-02655-2



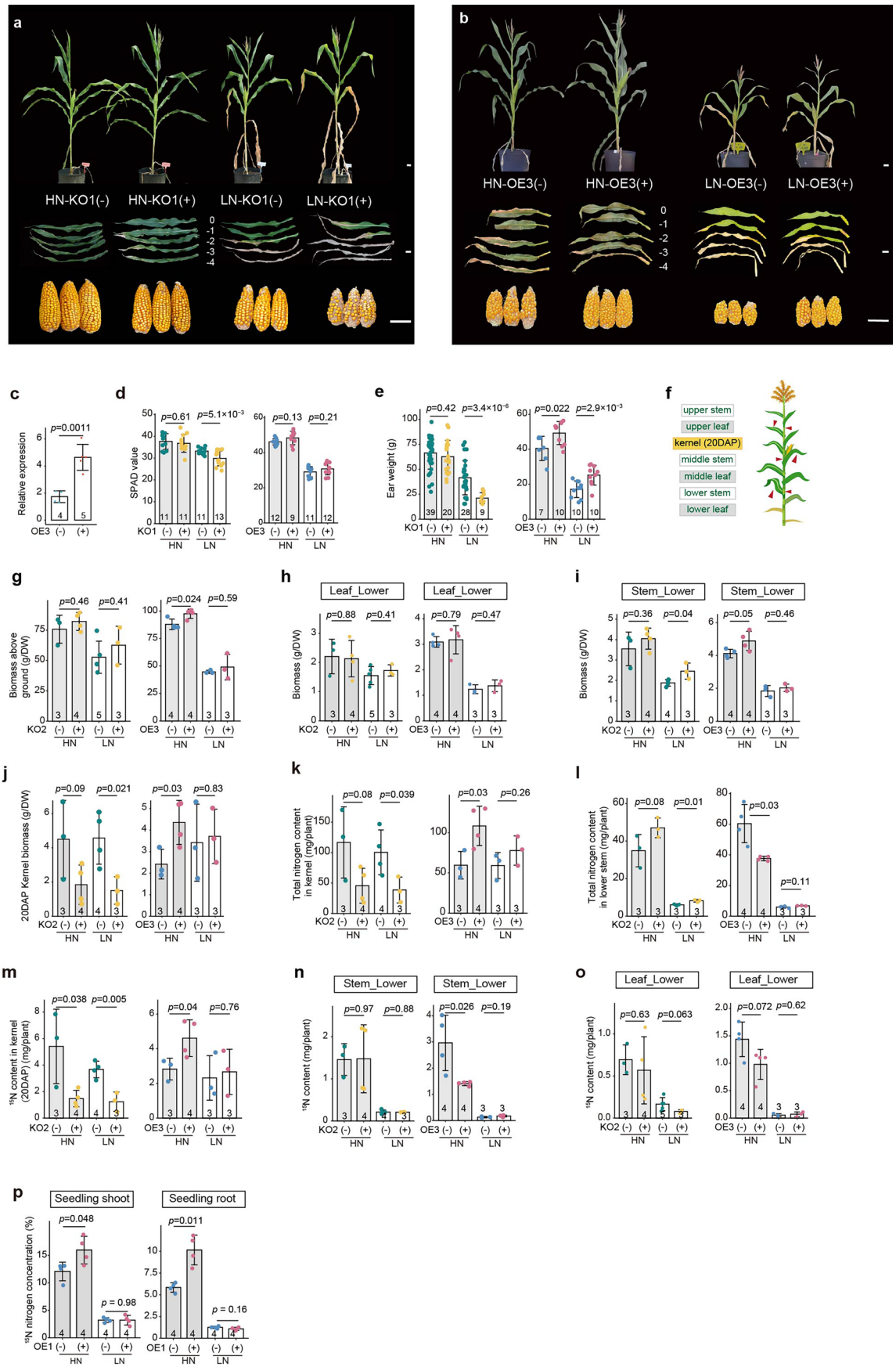
Extended Data Fig. 2 | Phylogenetic relationships among members of the *ATLb* subfamily. a, Phylogenetic relationships among members of the *ATLb* subfamily of amino acid transporters in maize, rice and Arabidopsis. **b**, Expression profiles for *ZmAVT1A-1* and its two homologous genes across 79 diverse tissues representing major organs and developmental stages of B73 maize[34]. **c**, Expression variations in progeny leaves across CUBIC population. Each point represents one CUBIC progeny with leaf transcriptome data (n = 391) and was treated as one biological data point. The box plots show the median and interquartile range (25th-75th percentiles), with whiskers extending to the minimum and maximum values within 1.5×IQR.

https://doi.org/10.1038/s41588-026-02655-2



Extended Data Fig. 3 | Structural variations in *ZmAVT1A-1*. **a**, Alignment of *ZmAVT1A-1* and its flanking regions from Ye478, HZS, B73, and Dan340. **b**, Detection of insertion/deletion variants in 24 CUBIC parents using a DNA marker. **c**, The expression ranges (unscaled) for *ZmAVT1A-1* among three haplotypes in CUBIC. n = 58, 117 and 84 for *Non_TE*, *Ye478_type* and *HZS_type*, respectively. Data are presented as mean±s.d. *P* values were calculated using two-tailed Student's *t*-tests. **d**, Effects of insertion/deletion variation on the contents of valine, leucine, and phenylalanine in progeny leaves and positive correlations between contents of these amino acids and expression of *ZmAVT1A-1*. Blue dots, *Non_TE* type; yellow dots, *Ye478_type*; pink dots, *HZS_type*. *P* values were calculated using a two-sided Student's *t*-test. The box plots show the median and interquartile range (25th-75th percentiles), with whiskers extending to the minimum and maximum values within 1.5× IQR.

https://doi.org/10.1038/s41588-026-02655-2

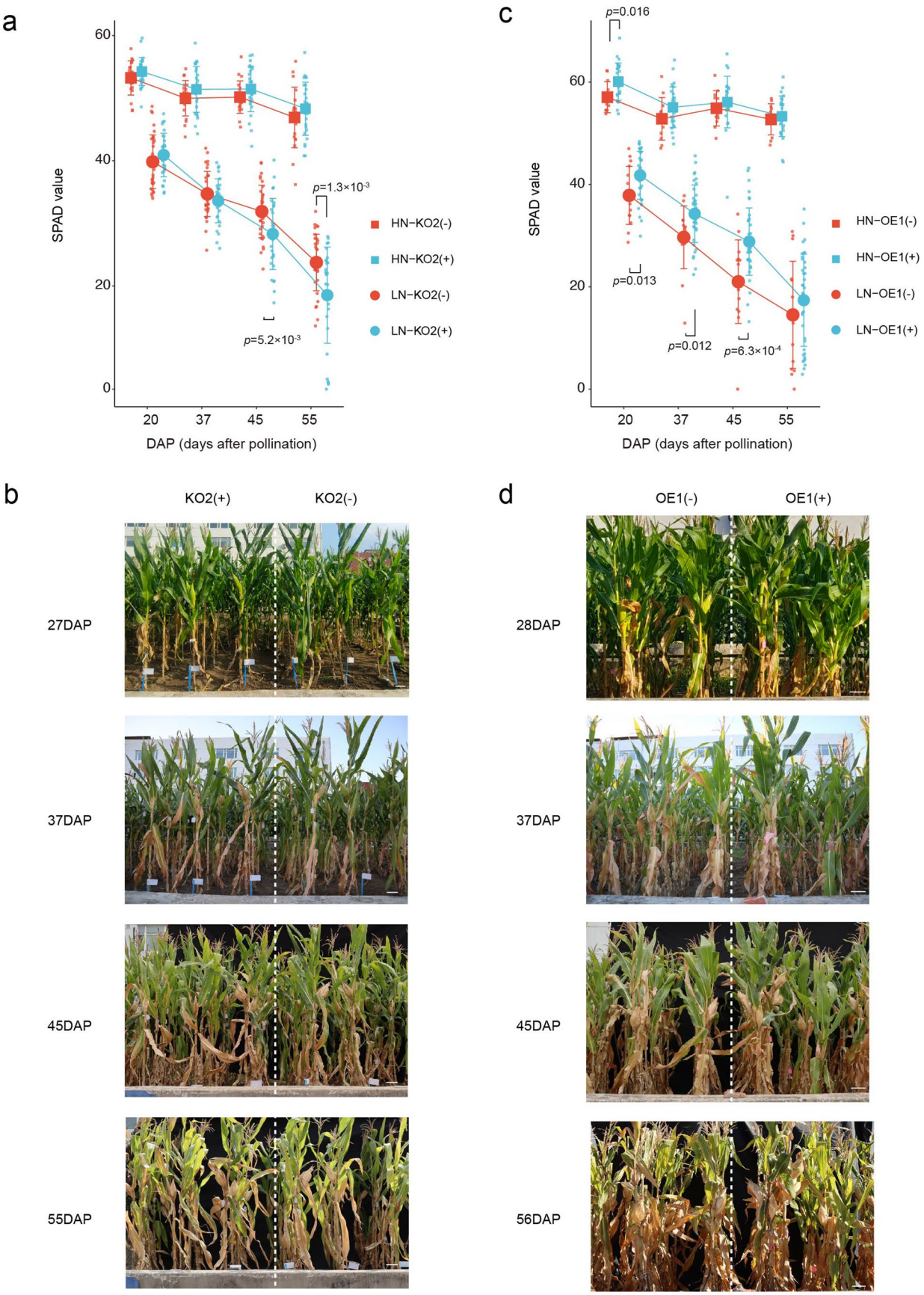


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41588-026-02655-2

Extended Data Fig. 4 | *ZmAVT1A-1* regulates maize nitrogen-associated traits through nitrogen partitioning. a, Leaf performance of *ZmAVT1A-1* knockout plants and corresponding null-segregant controls under high-nitrogen (HN) and low-nitrogen (LN) conditions. 0, -1, -2, -3, and -4 indicate the ear leaf and leaves 1-4 below the ear leaf. **b**, Leaf performance of *ZmAVT1A-1* overexpression plants and corresponding null-segregant under HN and LN conditions. In **a**-**b**, scale bars are 5 cm. **c**, qRT-PCR analysis of *ZmAVT1A-1* expression in the leaves. **d**, Leaf chlorophyll content estimated by SPAD readings for knockout and overexpression plants under HN and LN conditions. **e**, Harvest ear weight of knockout and overexpression plants under HN and LN conditions. Knockout and overexpression ears were assessed in the corresponding validation experiments. Data are mean± s.d. *P* values were calculated using a two-sided Student's *t*-test. **f**, Diagram of plant sampling for ^{15}N detection. Five individuals of similar growth status from each genotype under HN and LN conditions were labeled with ^{15}N at the V10 stage. At 20 days after pollination, at least 3 plants pollinated on the same day were harvested, divided into seven parts as shown, and used for ^{15}N detection. For details, see Methods. **g**-**j**, Dry weight of tissues from ^{15}N-labeled plants grown under HN and LN: g, shoot biomass 20 days after pollination; **h**, lower stems; **i**, lower leaves; **j**, kernel biomass 20 days after pollination; **k**-**l**, Total nitrogen content from different genotypes grown under HN and LN conditions for **k**, kernels and **l**, lower stems. **m**-**o**, Total amount of ^{15}N recovered in tissues: **m**, kernels; **n**, lower stems; **o**, lower leaves. **p**, ^{15}N concentration in the seedling shoot and root for overexpression lines under low nitrogen (LN) and high nitrogen (HN) conditions. In **c**-**d** and **g**-**p**, data is mean ± s.d., and the number of observed individuals is shown at the bottom of each bar. "+" denotes plants carrying the knockout allele or overexpression transgene, and "-" denotes null-segregant controls. *P* values were calculated using a two-sided Student's *t*-test. Maize illustration in **f** adapted with permission from Depositphotos (https://www.depositphotos.com/).

https://doi.org/10.1038/s41588-026-02655-2



Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41588-026-02655-2

Extended Data Fig. 5 | Leaf senescence phenotypes during grain filling. **a**, Relative ear leaf SPAD values over time for *ZmAVT1A-1* knockout plants. n = 28 for KO(+) plants and 20 for KO(-) under HN, n = 30 for KO(+) plants and 34 for KO(-) plants under LN at each time point. **b**, Representative time-course photographs of plants grown in low-nitrogen (LN) pools, comparing knockout versus null-segregant controls. **c**, Relative ear leaf SPAD values over time for *ZmAVT1A-1* overexpression plants. n = 27 for OE(+) plants and 12 for OE(-) plants under HN, n = 38 for OE(+) plants and 15 for OE(-) plants under LN at each time point. **d**, Representative time-course photographs of plants grown in low-nitrogen (LN) pools, comparing overexpression versus null-segregant controls. In **a** and **c**, only *P* value < 0.05 displays, and *P* value > 0.05 don't display. Exact *P* values are provided in Source Data. Each data point represents the mean relative SPAD value for the indicated genotype and nitrogen treatment at the indicated time point, and each biological replicate is one individual plant. Solid circles represent the average SPAD value of each group under LN conditions; the solid squares represent the average SPAD value under HN conditions; and the jittered points indicate individual observations. Data is mean±s.d. *P* values were calculated with two-sided Student's *t*-test between "+" and "-" plants within the same nitrogen treatment at each time point. "+" indicates plants carrying the knockout allele or overexpression transgene, and "-" indicates null-segregant controls. Scale bars are 10 cm.

nature portfolio

Corresponding author(s): Jianbing Yan

Last updated by author(s): May 10, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection

GC-MS (7890A-5975C, Agilent);
Confocal microscope (Zeiss LSM980);
Illumina NovaSeq 6000;
CFX96 Touch Real-Time PCR Detection System (Bio-Rad);
C1000 Touch Thermal Cycler (Bio-Rad);
SPAD-502 Plus chlorophyll meter (Tokyo, Japan);
isotope ratio mass spectrometry (Elementar isoprime 100);
Dumas rapid nitrogen analyser (rapid N exceed, Elementar);

Data analysis	Mass quantification: Agilent MassHunter Qualitative Analysis (vB06.00), Agilent MassHunter Quantitative Analysis (vB.07.01); Mass feature identification: The NIST library; GWAS: TASSEL3 (v3.0.174), EMMAX (beta-07Mar2010)； LD calculation: PLINK (v1.9); Phylogenetic tree construction: MEGA 11; One-way ANOVA: SPSS (IBM SPSS Statistic 22), Data display and statistic analysis: R version 4.3.2 with ggplot2 package. Student's t-test:Microsoft 2019, R version 4.3.2; Network construction: Cytoscape (v3.2.1), RNA-seq and DEG analysis: Bowtie2 (v2.4.1), RSEM(v1.3.3) GO analysis: g:Profiler.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw whole-genome sequencing and RNA sequencing reads of CUBIC population are available in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/) with accession number CRA000171 and CRA001241, respectively. Re-sequence data of the tesosinte, landrace and maize can be found under BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) accession PRJNA641489, PRJNA783885 and PRJNA531553, respectively. The RNA-sequencing data for knock-outs can be found in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/) under accession number CRA042606. The raw GC-MS data can be obtained at Metabolomics Workbench (https://dev.metabolomicsworkbench.org:22222/data/DRCCMetadata.php?Mode=Study&StudyID=ST003797&Access=NkqL1083). The B73 reference genomic sequences are collected from the MaizeGDB (https://www.maizegdb.org/). Gene model information was retrieved from MaizeGDB for ZmAVT1A-1 (GRMZM2G083788), ZmASN2 (GRMZM2G093175), ZmMGL (GRMZM2G450498) and ZmMIPS1 (GRMZM2G155242). Gene model information was retrieved from MaizeGDB for ZmAVT1A-1 (GRMZM2G083788), ZmAVT1A-2 (GRMZM2G080843), ZmAAP31 (GRMZM2G101125), ZmASN2 (GRMZM2G093175), ZmMGL (GRMZM2G450498), ZmMIPS1 (GRMZM2G155242), ZmBKI1 (GRMZM2G035741), ZmBAK1(GRMZM2G145720), ZmPRR7 (GRMZM2G095727), ZmNRT2.5 (GRMZM2G455124) and ZmGLN3 (GRMZM2G046601). Processed data supporting the findings of this study are provided in the Supplementary Information and Source Data files. Source data are provided with this paper. Gene-editing or overexpression maize plants are available from the corresponding author.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human participants or human data used in this study.
Reporting on race, ethnicity, or other socially relevant groupings	No human participants or human data used in this study.
Population characteristics	No human participants or human data used in this study.
Recruitment	No human participants or human data used in this study.
Ethics oversight	No human participants or human data used in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes are indicated in the relevant figures and figure legends. No statistical method was used to predetermine sample size. 1404 progenies planted and metabolic profiled. For GWAS, the sample size was determined by the number of inbred lines available in the mapping population with MAF filtered. For RNA-seq, pot experiment, isotope labelling and metabolic profiling, at least 3 independent biological

	replicates were included for each genotype/condition. For field phenotyping experiments, at least 12 independent biological replicates were collected for each genotype or field treatment. For the NUE index analysis of hybrids, at least 10 different crosses were included for each maternal haplotype at each location.
Data exclusions	Outliers of phenotype data such as ear and plant architecture were excluded for abnormal growth and development.
Replication	All molecular experiments in this study were independently repeated at least three times with similar results. For RT–qPCR, three biologically independent samples were analysed in each experiment. Field trials were performed using a randomized plot design with three biological replications where applicable. For transgenic analyses, independent transformation events and their corresponding segregating sibling controls were used to define the treatment and control groups. The hybrid NUE index analysis and field trials were reproduced in multiple independent experiments, as indicated in the relevant figure legends, with at least two independent transgenic events or two independent randomized plots where applicable.
Randomization	Progenies were planted in a randomized order in the field. For leaf metabolic profiling, leaves from three randomly selected individuals within each line were pooled, with each line comprising 12 individuals. For kernel metabolic profiling, three ears were randomly collected from each line. For transgenic plants, positive transgenic plants and their negative segregating sibling controls were planted, cultivated and analysed under the same environmental conditions for RNA-seq, metabolic profiling and phenotyping. Randomly shuffled genomic regions were used as controls in the QTL hotspot enrichment analysis.
Blinding	The investigators were blinded to the maize genotypes during data collection. All sample names were coded during data production which ensures that the data were not influenced by knowing the identity of samples.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	The seeds for parents and progeny in CUBIC and trans-genic or gene-edited plants are stored in Yan's lab (http://www.maizego.org/).
Novel plant genotypes	Transgenic lines were obtained via Agrobacterium infection into immature embryos of KN5585. The gene-edited lines were obtained with the CRISPR-Cas9 system with U6 promoter.
Authentication	No authentication procedures were done for the seed stock used and the novel genotype generated.