Chen et al. Genome Biology (2025) 26:67
https://doi.org/10.1186/s13059-025-03541-6

Genome Biology

RESEARCH

Open Access

Dissecting the genetic architecture of key agronomic traits in lettuce using a MAGIC population

Hongyun Chen[1,2,6†], Jiongjiong Chen[1†], Ruifang Zhai[3†], Dean Lavelle[4], Yue Jia[1], Qiwei Tang[1], Ting Zhu[1], Menglu Wang[1], Zedong Geng[5], Jianzhong Zhu[3], Hui Feng[3], Junru An[1], Jiansheng Liu[1], Weibo Li[1], Shenzhao Deng[2], Wandi Wang[2], Weiyi Zhang[1], Xiaoyan Zhang[1], Guangbao Luo[1], Xin Wang[1], Sunil Kumar Sahu[2,6], Huan Liu[6], Richard Michelmore[4], Wanneng Yang[5*], Tong Wei[2,6,7,8*] and Hanhui Kuang[1*]

†Hongyun Chen, Jiongjiong Chen and Ruifang Zhai contributed equally to this work.

*Correspondence:
ywn@mail.hzau.edu.cn;
weitong@genomics.cn;
kuangfile@mail.hzau.edu.cn

[1] National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops; Hubei Hongshan Laboratory, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China
[2] BGI Research, Wuhan 430074, China
[5] National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China
Full list of author information is available at the end of the article

Abstract

Background: Lettuce is a globally important leafy vegetable that exhibits diverse horticultural types and strong population structure, which complicates genetic analyses. To address this challenge, we develop the first multi-parent, advanced generation inter-cross (MAGIC) population for lettuce using 16 diverse founder lines.

Results: Whole-genome sequencing of the 16 founder lines and 381 inbred progeny reveal minimal population structure, enabling informative genome-wide association studies (GWAS). GWAS of the lettuce MAGIC population identifies numerous loci associated with key agricultural traits, including 51 for flowering time, 11 for leaf color, and 5 for leaf shape. Notably, loss-of-function mutations in the *LsphyB* and *LsphyC* genes, encoding phytochromes B and C, dramatically delay flowering in lettuce, which is in striking contrast to many other plant species. This unexpected finding highlights the unique genetic architecture controlling flowering time in lettuce. The wild-type *LsTCP4* gene plays critical roles in leaf flatness and its expression level is negatively correlated with leaf curvature. Additionally, a novel zinc finger protein (*ZFP*) gene is required for the development of lobed leaves; a point mutation leads to its loss of function and consequently converted lobed leaves to non-lobed leaves, as exhibited by most lettuce cultivars.

Conclusions: The MAGIC population's lack of structure and high mapping resolution enables the efficient dissection of complex traits. The identified loci and candidate genes provide significant genetic resources for improving agronomic performance and leaf quality in lettuce.

Keywords: Lettuce, GWAS, PhyB, PhyC, Flowering time, Lobed leaves

Background

Lettuce (*Lactuca sativa*), a member of the Asteraceae (Compositae) family, is one of the most important vegetable crops worldwide. The total global production of lettuce and chicory was 27.149 million tons in 2022 (FAOSTAT, https://www.fao.org/faostat/en/). As a leafy salad vegetable, the color and shape of the leaf are critical sensory qualities of lettuce. There are several distinct horticultural types of lettuce that vary in shape, color, and degree of heading; these types constitute different genetic populations within cultivated lettuce.

Genetic studies of variations in leaf color have revealed at least seven polymorphic genes associated with the color of lettuce leaves [1–5]. The *RLL1-4* genes encode transcription factors that regulate the anthocyanin biosynthesis pathway in lettuce, resulting in different levels of red color [3]. Mutation of *LsF3'H* and *LsCHI*, two structural genes, reduces anthocyanin accumulation but significantly increases the contents of kaempferol and naringenin in lettuce [6]. A CACTA transposon inserted in the 5' untranslated region of the *LsANS* gene converted red leaves to green leaves [2]. In addition to their red color, lettuce cultivars also vary from pale green to dark green; two genes, *LsGLK* and *LsNRL4*, have been shown to regulate the chlorophyll content and depth of green leaves [1, 5]. However, whether there are additional genes contributing to variations in leaf color in lettuce has yet to be explored.

Leaf shape is established by cell proliferation and expansion along the proximal-distal, medial-lateral, and adaxial-abaxial axes [7]. Developmental dysregulation of leaf dorsoventrality (adaxial-abaxial axis) may result in the development of leafy heads in some vegetable crops. The loss of function of *LsSAW1* led to the upregulation of abaxial genes and the downregulation of adaxial genes to promote the development of leafy heads in lettuce [8]. The upregulation of the *LsKN1* gene by the insertion of a CACTA transposon caused multiple morphological changes in lettuce, including leafy heads, wavy leaves, and/or palmately lobed leaves, depending on the genetic background [9–11]. The downregulation of the expression levels of *LsTCP4*, a CIN-like TCP transcription factor, has been proposed to be associated with wavy leaves in lettuce [12].

Leaf lettuce is harvested as a vegetative rosette. Bolting results in a bitter taste and unmarketable heads; therefore, late bolting and flowering are important breeding goals for lettuce [13]. Flowering time is regulated by multiple pathways, including the photoperiod, ambient temperature, autonomous, vernalization, gibberellin, and age pathways [13, 14]. Numerous genes, including the floral integrator genes *CONSTANS* (*CO*) and photoreceptors *phytochrome A-E* (*phyA-E*), are key players in the photoperiod pathway [14–16]. The effects of the loss of function of *phyB* or *phyC* on flowering time vary dramatically among different plant species. Loss-of-function of *phyB* or *phyC* typically accelerates flowering in most species investigated, including Arabidopsis, rice, poplar, and barley [17–21]. However, the loss of functions of *phyB* or *phyC* delays flowering in wheat under long-day photoperiods [22, 23].

Genome-wide association studies (GWAS) are considered the most effective approach for the systematic identification of genetic variations associated with multiple traits [24]. However, the presence of population structure and rare alleles may compromise its efficacy [25]. The use of multiparent populations for GWAS provides a solution for such constraints. Multi-parent populations that are derived from multiple, diverse parental lines can sample genetic diversity and balance allele frequencies. Multi-parent genetic designs, such as nested association mapping (NAM), multi-parent advanced generation inter-crossing (MAGIC), random-open parent association mapping (ROAM), and complete-diallel design plus unbalanced breeding-like intercrossing (CUBIC), have all been advocated to exploit the advantages of multi-parent populations for GWAS. MAGIC populations have been the most popular design and used to identify polymorphic genes controlling important traits in several crop species, including rice, cotton, and tomato [26–30].

In this study, we developed a MAGIC population to address the strong population structure present among the different horticultural types of lettuce. The MAGIC population was genotyped by sequencing and phenotyped using three imaging technologies: hyperspectral imaging (HSI), RGB imaging, and 3D scanning. GWAS was performed to identify polymorphic genes controlling leaf morphology and flowering time. The identified loci were verified using bi-parental segregating populations. Major genes associated with flowering time and leaf shape were subsequently cloned, and their functions confirmed using CRISPR/Cas9-mediated mutagenesis. Loss of function of the *phyB* and *phyC* genes was shown to delay flowering time in lettuce. Expression level of *LsTCP4* was negatively correlated with leaf curvature. We also identified a novel gene encoding a zinc finger protein required for the development of lobed leaves in lettuce. The large number of loci associated with important traits identified in this study provides foundational genetic resources for future breeding programs for lettuce.

Results

Construction of the first MAGIC population for lettuce

We chose 14 accessions of cultivated lettuce (*L. sativa*) and two accessions of *L. serriola*, the wild progenitor of cultivated lettuce, as founder lines for construction of a MAGIC population (Fig. 1a, b and Additional file 1: Table S1). The 14 cultivated accessions included horticultural types of crisphead, butterhead, romaine, loose leaf, Youmaicai, and stem lettuce. The 16 founder lines were inter-crossed for four generations, and their hybrid progeny were then self-pollinated for at least six generations to generate inbred lines. A total of 381 inbred progeny were chosen for further genotyping and phenotyping. Images of 2-month-old plants of these inbred progeny can be found in Additional file 2: Fig. S1.

Sequencing and minimal population structure of the MAGIC population

We performed whole-genome sequencing of the MAGIC population, including the 16 founder lines and their 381 inbred progeny. A total of 17.3 Tb of sequencing data were obtained, with an average mapping depth of 19.4 × and an average genome coverage of 95.8% relative to the lettuce reference genome of cv. Salinas version 8 (Additional file 1: Table S1 and Additional file 2: Fig. S2). We identified a total of 52,004,146 single-nucleotide polymorphisms (SNPs) and 7,450,810 insertions/deletions (indels) in the MAGIC population, with densities of 21.4 SNPs/kb and 3.1 indels/kb, respectively (Additional file 1: Table S2). The 381 inbred progeny in the MAGIC population captured 98.2% and 97.6% of the total SNPs and indels between the founders, respectively. Approximately

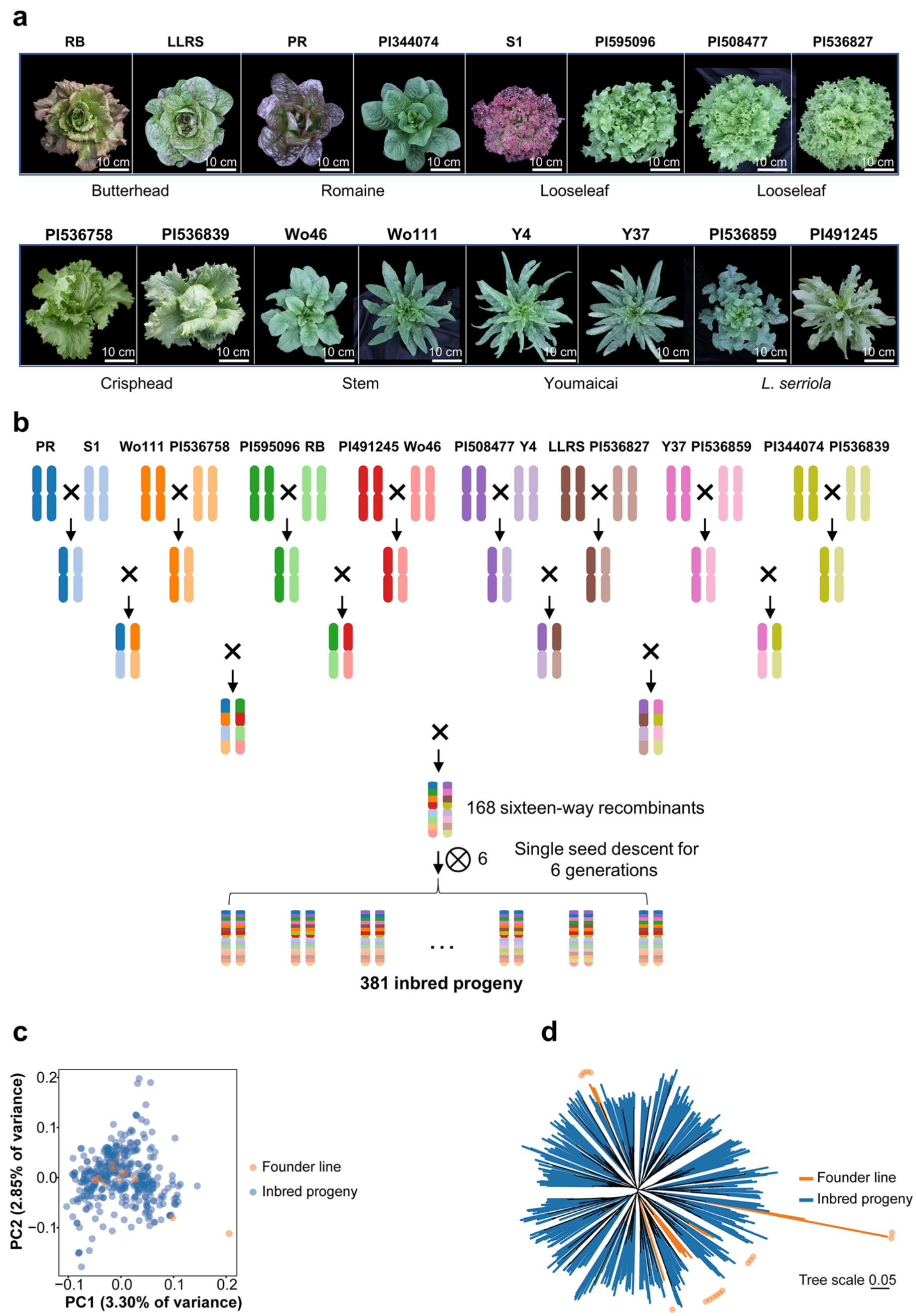

Fig. 1 Construction and population structure of the MAGIC population of lettuce. **a** Morphology of the 16 founder lines. The name and horticultural type of each accession are shown. **b** Pedigree of the MAGIC population. After four consecutive generations of intercross, the hybrids were self-pollinated to generate inbred progeny. **c** Principal component analysis (PCA). **d** Neighbor-joining (NJ) phylogenetic tree for the 16 founder lines and their 381 inbred progeny. Only polymorphic sites were used, and the scale bar refers to the number of substitutions per polymorphic site. The orange dots represent the founder lines

2.78% of the SNPs (1,443,390) and 4.98% of the indels (370,947) were located in genic regions (Additional file 1: Table S3).

Using the SNPs unique to each founder line, we traced the genomic bins of each inbred progeny back to the 16 founder lines (Additional file 2: Fig. S3), which revealed that the inbred progeny were the result of extensive recombination between chromosomal inputs from the 16 founder lines. The average rate of genomic heterozygosity in the inbred progeny was 0.81%, which was slightly lower than the expected heterozygosity rate of 1.5625% after six generations of selfing (Additional file 1: Table S4). After filtering the SNPs (see the "Methods" section), 22,023,376 SNPs were retained for further analysis. Principal component analysis (PCA) of the MAGIC population revealed that the first five PCs explained only 13.56% of the variance (Fig. 1c and Additional file 2: Fig. S4). Consistent with the PCA results, the neighbor joining (NJ) tree of the MAGIC population revealed that most of the 381 inbred progeny were evenly distributed, although the founder lines formed five groups (Fig. 1d). The clustering of the founder lines into a few groups is consistent with the results of previous studies [4, 31]. In summary, the MAGIC population for lettuce that we developed showed no obvious population structure, in striking contrast to the strong population structure found in diversity panels that include multiple types of lettuce [4, 31]. The linkage disequilibrium (LD) decay distance for the MAGIC population was estimated to be 279.3 kb ($r^2 = 0.32$; Additional file 2: Fig. S5).

High-throughput phenotyping of the MAGIC population

The MAGIC population presented rich variations in terms of leaf shape, leaf size, leaf color, and flowering time, reflecting the phenotypic diversity of the 16 founder lines (Additional file 2: Fig. S6). Leaf shape, leaf size, leaf color, and flowering time all exhibited continuous distributions, suggesting that these quantitative traits are controlled by multiple genes or a few genes with large environmental variance. Lobed/non-lobed leaves and seed color, on the other hand, showed non-continuous distributions in the MAGIC population (Additional file 2: Fig. S6).

A high-throughput phenotyping platform was used to measure leaf shape, leaf size, and leaf color in the MAGIC population. The largest leaf of each plant was phenotyped by using HSI, RGB imaging, and 3D scanning (Additional file 2: Fig. S7). A total of 1851 image-based traits (i-traits) were obtained, including 1793 HSI i-traits (Additional file 1: Table S5, S6), 33 RGB i-traits (Additional file 1: Table S7, S8), and 25 3D i-traits (Additional file 1: Table S9, S10). In addition, leaf color, stem color, seed color, lobed leaves, and flowering time were manually scored for the MAGIC population through visual inspection (Additional file 1: Table S11, S12).

More than ten loci are associated with leaf and stem color in lettuce

We first performed a GWAS of leaf color, which was phenotyped using HSI and RGB imaging. For the HSI, the data under visible light represented the trait of leaf color, which is likely associated with the relative concentrations of anthocyanins (465 ~ 560 nm), chlorophylls (400 ~ 500 nm and 600 ~ 700 nm), and carotenoids (400 ~ 530 nm). The HSI data at invisible wavelengths ranging from 780 to 998 nm might reflect variations in the concentrations of some colorless metabolites. A total of nine significant loci for leaf color, *Leaf Color 1-9* (*qLC1-9*), were identified (Additional file 2: Fig. S8), four of which are colocalized with known genes associated with leaf color in lettuce (Additional file 1: Table S13). Among these genes, *qLC6*, *qLC7*, and *qLC8* encode bHLH, MYB, and ANS proteins, respectively, which were previously shown to control red and green leaves in lettuce [2, 3]. The *qLC9* locus overlaps with the *LsGLK* gene, which is the major QTL for chlorophyll concentration in lettuce leaves [5]. The remaining five loci (*qLC1-5*) were identified for the first time in this study. We used the S-TEX parameter to quantify the uniformity of leaf color (Additional file 1: Table S7, S8). GWAS of S-TEX identified one significant locus, which is located on chromosome 6 and overlaps with the locus *qLC1* (Additional file 2: Fig. S8). We predicted candidate genes for the five newly identified loci on the basis of the sequence variations among the 16 founder lines and the predicted functions of genes from the candidate region [32–38] (Additional file 1: Table S13).

Like the leaves, the stems also presented red/green variations in the MAGIC population due to the accumulation or lack of anthocyanins in the epidermal cells of the stems. We performed a GWAS of stem color and identified two significant loci, *Red Stem 1-2* (*qRS1-2*) (Fig. 2a). The stem color loci *qRS1* and *qRS2* overlap with the leaf color loci *qLC6* and *qLC8*, respectively (Additional file 1: Table S13). The *bHLH* gene at the *qRS1*/*qLC6* locus had a frameshift deletion in six of the 16 founder lines, whereas the *ANS* gene at the *qRS2/qLC8* locus had a nonsense mutation leading to a premature stop codon. We knocked out the *bHLH* and *ANS* genes from lettuce with red leaves and red stems; the knockout mutants exhibited green leaves and stems, confirming that the *bHLH* and *ANS* genes are required for both red leaves and red stems in lettuce (Fig. 2b, c). GWAS suggested that *qLC7*, which encodes a MYB transcription factor, did not contribute to the variation in stem color in the MAGIC population, although it controls leaf color in lettuce. Indeed, knockout of the MYB-encoding gene at *qLC7* did not alter stem color (Fig. 2d). Therefore, the red stems and red leaves of lettuce result from overlapping regulatory pathways, but each pathway has its own specific components.

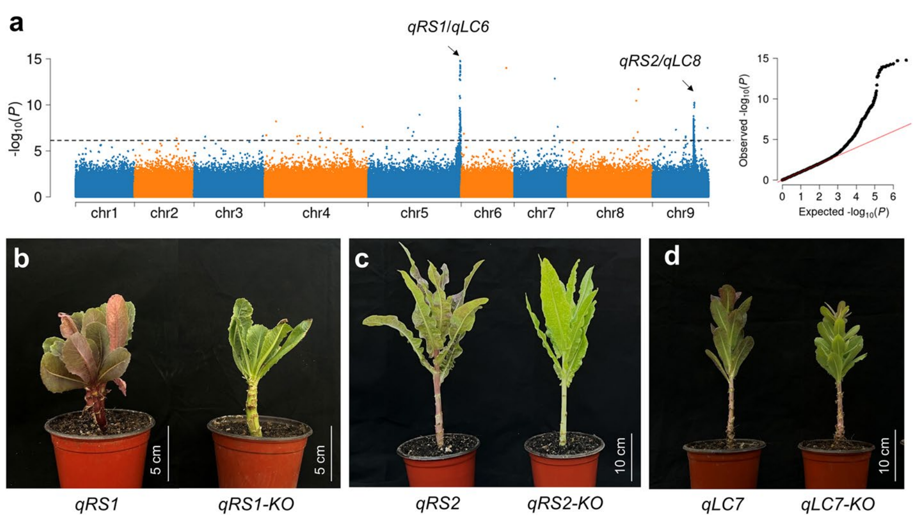

Fig. 2 The genetics of the stem color of lettuce. **a** GWAS of stem color identified three significant loci. **b**, **c** Knockout of the *bHLH* gene at *qRS1* (**b**) or the *ANS* gene at *qRS2* (**c**) converted the red stem (left) to the green stem (right). Scale bar = 5 cm or 10 cm. **d** Knockout of the *MYB* gene at locus *qLC7* did not alter stem color. Scale bar = 10 cm

Numerous loci are associated with flowering time in lettuce

We performed a GWAS of flowering time in the MAGIC population. Three-week-old seedlings were transplanted to the field in Wuhan, China, in three different seasons that differed in temperature and daylength: February 2021, September 2021, and November 2022 (hereafter referred to as the February, September, and November seasons, respectively). The GWAS revealed 44, 14, and 37 loci related to flowering time in the February, September, and November seasons, respectively (Fig. 3a-c). Among the 51 significant loci associated with flowering time, 34 were detected in at least two seasons; 33 were detected for the first time in this study (Additional file 1: Table S14). We chose candidate genes for 16 loci [18, 22, 39–53], on the basis of their sequence variations, predicted protein functions related to flowering time [54], and eQTLs in the candidate regions [4] (Additional file 1: Table S13, S14). For example, *LsAGL24*, which encodes a MADS-box protein and has a missense SNP among the 16 founder lines, is significantly associated with flowering time in all three seasons [49, 50]. *LsAGL4*, another MADS-box protein-encoding gene and from a significant locus on chromosome 2, has a frameshift indel in six parents and is a distant eQTL for the *AP1* homologs *LG2221258* and *LG2221298* [4, 39]. *LsPUB13*, from a significant locus on chromosome 8, encodes a protein with UND, U-box, and ARM domains and acts as a distant eQTL for *AGL16* (*LG8741644*) and *DCL3* (*LG8700429*) homologs [51].

Loss-of-function mutations of LsphyB and LsphyC lead to late flowering in lettuce

We used bi-parental segregating populations to verify some of the flowering loci identified above. We developed a segregating population by crossing the stem lettuce cultivar Wo111 and the *L. serriola* accession CGN10978 and identified an F_5 family that segregated in flowering time (Fig. 3d). BSR analysis revealed that a locus controlling flowering time in the F_5 population was located on chromosome 1, overlapping with the *qFLT1.1* region identified above through GWAS (Additional file 2: Fig. S9a). We then fine mapped *qFLT1.1* using 2,394 F_5 individuals and the causal gene was localized to an interval of 0.34 Mb between 47.32 and 47.66 Mb on chromosome 1 (Additional file 2: Fig. S9b). This interval contains the ortholog of *phyB*, which is associated with flowering time in Arabidopsis [18]. The *LsphyB* allele in the late-flowering genotype had a nonsense mutation leading to a premature stop codon (Fig. 3e). Notably, *LsphyB* in three of the 16 founder lines presented the nonsense mutation. To verify the function of the *LsphyB* gene in flowering time, we knocked out the *LsphyB* gene in cultivated lettuce accessions PI251246 and Wo703, which have the wild-type *LsphyB* allele. Flowering time of the knockout mutants of both accessions was longer relative to the wild-types and flowering time cosegregated with the mutated *LsphyB* gene in the T_1 generations (Fig. 3f and Additional file 2: Fig. S9c-e). Therefore, *LsphyB* is the *qFLT1.1* gene and its loss of function causes late flowering in lettuce; this is in striking contrast to the early flowering of *phyB* mutants in Arabidopsis [18].

Using the same BSR strategy, we identified another F_5 family (resulting from four generations of selfing for the M152 line of the MAGIC population), which was segregating for SNPs at the *qFLT7.1* locus on chromosome 7 (Fig. 3g). This locus co-segregated with flowering time in this F_5 family, which is consistent with the above GWAS results and

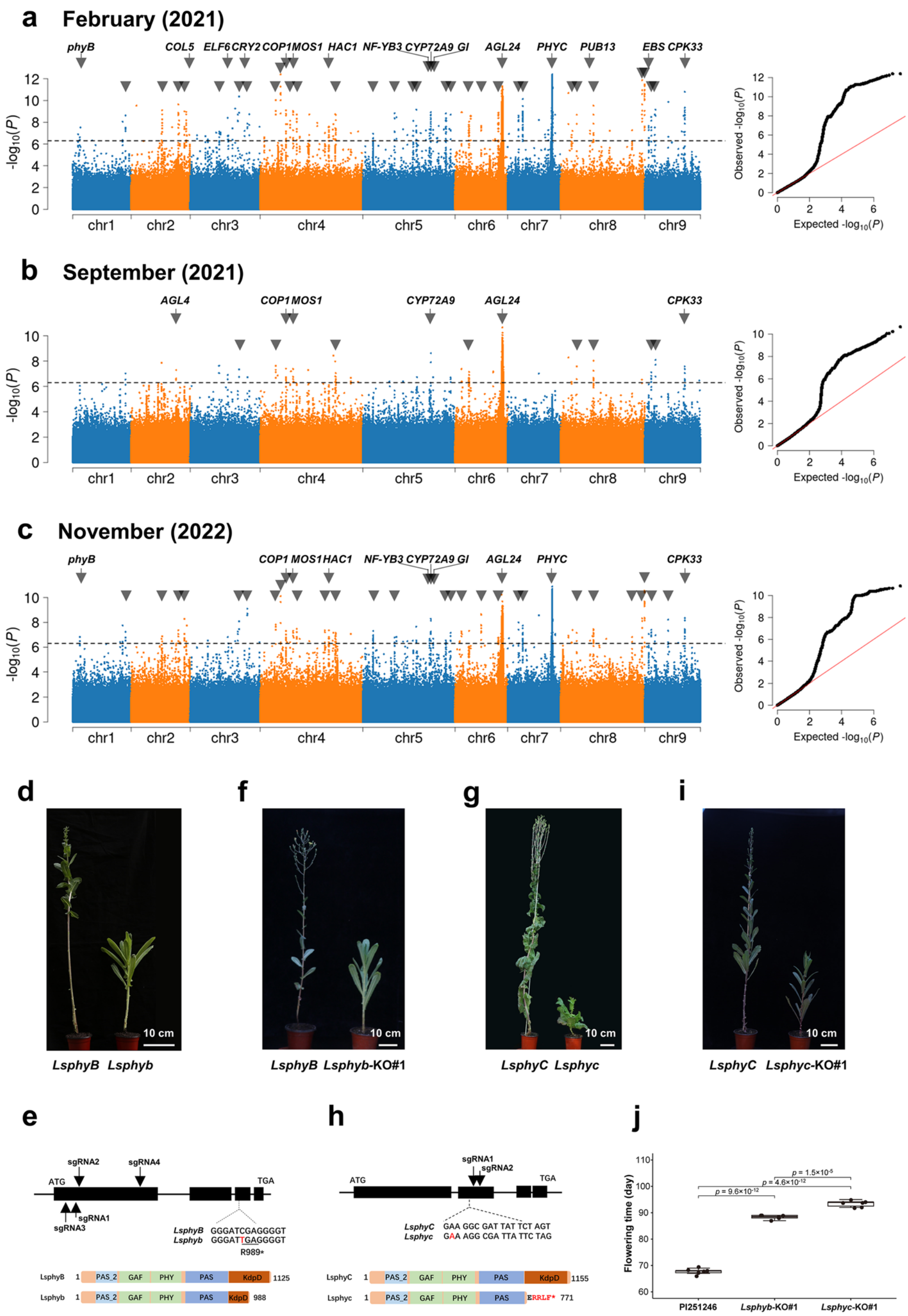

Fig. 3 GWAS of flowering time in the MAGIC population and validation of flowering time-related genes. **a** Planted in February 2021. **b** Planted in September 2021. **c** Planted in November 2022. The black horizontal dashed line in the Manhattan plots represents the suggestive genome-wide significance threshold. The arrow heads show significant loci with predicted candidate genes, whereas the triangles represent significant loci with no known candidate genes. **d** Near-isogenic lines (NILs) of *LsphyB*. Scale bar = 10 cm. **e** Natural loss-of-function mutation in *LsphyB*. The nonsense point mutation (in red) in *Lsphyb* at the 989th codon led to partial loss of the KdpD domain. The positions of the four sgRNAs used to knock out *LsphyB* are shown. **f** Knockout mutants of *LsphyB* flowered considerably later than the wild-type. Scale bar = 10 cm. **g** NILs of *LsphyC*. Scale bar = 10 cm. **h** Natural loss-of-function mutation in *LsphyC*. A nucleotide adenine (in red) was inserted in the non-functional allele *Lsphyc*, leading to the loss of the entire KdpD domain. The positions of the two sgRNAs used to knock out *LsphyC* are shown. **i** Knockout mutants of *LsphyC* in lettuce flowered considerably later than the wild-type. Scale bar = 10 cm. **j** Delayed flowering time in *Lsphyb* and *Lsphyc* knockout mutants. The statistical significance of the differences in flowering time was calculated via Student's *t* test

previous studies [4, 31] (Additional file 2: Fig. S10). *LsphyC* was considered the candidate gene at the *qFLT7.1* locus. We sequenced the two *LsphyC* alleles segregating in this population and found that *LsphyC* from late-flowering individuals had a 1-bp insertion, leading to a premature stop codon (Fig. 3h). To verify the function of *LsphyC*, we then knocked out the wild-type *LsphyC* from accession PI251246. As expected, the knockout mutants flowered later than the wild-type accession (Fig. 3i and Additional file 2: Fig. S10). Therefore, the causal gene at the *qFLT7.1* locus is *LsphyC* and a 1-bp insertion in its coding region leads to loss of function, which results in late flowering. The *Lsphyc* knockout mutants flowered 26 days later than the wild-type (PI251246), in comparison to a delay of 21 days in *Lsphyb* mutants, when grown in a protected field in March 2024 (Fig. 3j).

Several loci are associated with leaf shape and size in lettuce

We measured the largest leaf of each individual from the MAGIC population using three imaging methodologies, thereby measuring multiple i-traits that correspond to leaf shape and size. The GWAS detected one significant locus each for leaf curvature,leaf lobing, leaf length, leaf width, leaf area, and tip shape (Additional file 1: Table S13 and Additional file 2: Fig. S11).

We predicted candidate genes for the significant loci associated with leaf shape and size [12, 37, 55, 56], on the basis of predicted protein function and sequence variation (Additional file 1: Table S13, S15). The candidate gene for the significant locus on chromosome 5 controlling leaf curvature is *LG5495730.* The *LG5495730* gene encodes a TCP homolog, and there is an insertion of a 5411 bp retrotransposon in its 3' UTR in six of the 16 founder lines leading to curved leaves (See next section). The candidate gene for lobed leaves encodes a zinc finger protein (See below; Additional file 1: Table S13). The *LG1103550* gene, the candidate gene controlling leaf length, encodes a homolog of CDC48, a key regulator of cell division [55]. The *LsCDC48* gene has two distinct alleles among the 16 founder lines, and the allele from PI 491245 results in longer leaves than the other alleles do ($p < 0.05$). Similarly, the allele from PI 491245 had the highest expression level among the 16 founder lines (Additional file 2: Fig. S12). The significant locus for leaf width co-localizes with that for leaf area, and the candidate gene is predicted to be *LG6568660*, which encodes a protein kinase and has a missense mutation in two of the 16 founder lines.

The insertion of a retrotransposon disrupts the UTR of LsTCP4 and downregulates its expression resulting in wavy leaves

We had previously constructed a segregating population by crossing the loose-leaf parent S1 and the stem lettuce Y37, which are two founder lines of the MAGIC population [3]. Parent S1 has wavy leaves, whereas parent Y37 has flat leaves. We obtained an F_5 family that segregated into three distinct leaf phenotypes, severely wavy leaves, moderately wavy leaves, and flat leaves, with a Mendelian ratio of 1:2:1 ($\chi^2 = 2.24$, $p > 0.05$). BSR analysis suggested that the degree of leaf flatness in this family is controlled by a locus on chromosome 5, which overlaps with the locus detected via GWAS (Additional file 2: Fig. S13). We then fine-mapped the locus using 3400 individuals from the F_5 family. The causal gene was ultimately mapped to a 1.0-Mb interval between 252.34 and

253.35 Mb on chromosome 5 (Additional file 2: Fig. S13). This interval contains 16 predicted genes, including the *LsTCP4* gene, which encodes a CIN-like TCP transcription factor. In a previous study, *LsTCP4* was predicted to be responsible for wavy leaves, and downregulation of *LsTCP4* was due to an insertion of a Ty3/gypsy retrotransposon in its downstream region [12]. We sequenced the *LsTCP4* gene from the two parents. *LsTCP4* from parent S1 (wavy leaves) has the insertion of the Ty3/gypsy retrotransposon, but parent Y37 (flat leaves) lacks the retrotransposon (Additional file 2: Fig. S13). Furthermore, homozygotes of *Lstcp4* from the F_5 segregating population had low expression, while homozygotes of *LsTCP4* had high expression, and heterozygotes had a medium level of expression, which was consistent with their corresponding phenotypes (Additional file 2: Fig. S13).

To verify the function of *LsTCP4* in determining wavy leaves, we knocked out the *LsTCP4* gene in both the *LsTCP4* genotype and the *Lstcp4* genotype (which had the insertion of the retrotransposon). Three knockout mutants were obtained for the *LsTCP4* genotype, and all three mutants had extremely wavy leaves in comparison to the flat leaves of the wild-type *LsTCP4* plants (Fig. 4a). Three knockout mutants were also obtained for the *Lstcp4* genotype (with the insertion of the retrotransposon). The leaves of all three homozygous knockout mutant lines became more wavy than those of the progenitor *Lstcp4* genotype, suggesting that the naturally mutated allele retained partial functions (Additional file 2: Fig. S14). In contrast, the leaves of *LsTCP4* overexpression lines changed from wavy to flat (Fig. 4b). We also used CRISPR/Cas9-mediated mutagenesis to delete 1255-bp sequences in the downstream region of *LsTCP4*; the deletion mutant had lower expression of *LsTCP4* and converted flat leaves to wavy leaves (Fig. 4c, d). We conclude that the insertion of the retrotransposon disrupted the downstream sequence of *LsTCP4*, lowered its expression, and consequently generated wavy leaves in lettuce.

A point mutation in a novel ZFP gene leads to its loss of function and non-lobed leaves in lettuce

Lettuce has lobed or non-lobed leaves, and lobed phenotypes vary dramatically (Fig. 5a). GWAS of lobed leaves revealed a highly significant locus on chromosome 3 (Fig. 5b). The most significant SNP was located in the coding region of the gene *LG3316063*, which encodes a zinc finger protein of unknown function. The prominent point mutation at + 200 caused an amino acid change from glycine in genotypes with lobed leaves to valine in genotypes with non-lobed leaves. To validate the GWAS results, we constructed two bi-parental segregating populations to study the genetics underlying the variation of lobed/non-lobed leaves. BSR analysis revealed that lobed leaves, regardless of their complexity, are controlled by the same locus on chromosome 3 (Fig. 5c) and the *LG3316063* gene cosegregated with lobed/non-lobed leaves in the two populations. We then generated CRISPR/Cas9-mediated knock-outs of the *LG3316063* gene in accessions CGN04971 and PI595096, which have lobed leaves (Additional file 1: Table S16). The homozygous knockout of the *LG3316063* gene changed the lobed leaves to non-lobed leaves (Fig. 5d). We therefore conclude that the novel *ZFP* gene is required for the development of all types of lobed leaves in lettuce.

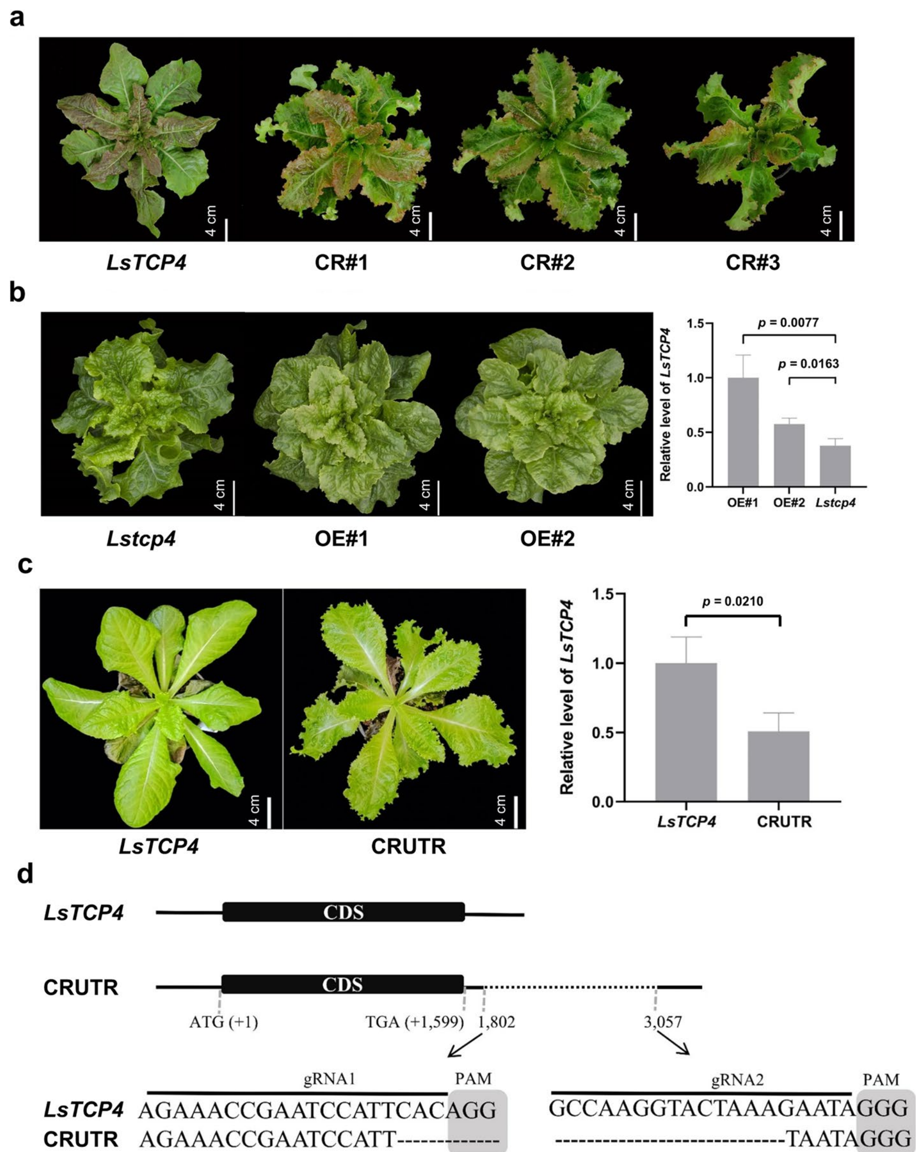

Fig. 4 Functional verification of the *LsTCP4* gene in determining the curvature of leaf margins in lettuce. **a** Knockout of the *LsTCP4* gene led to wavy leaves. A genotype with wild-type *LsTCP4* (no insertion of retrotransposon, left) and its three *Lstcp4* knockout mutants (CR#1 - CR#3). Scale bar = 4 cm. **b** Overexpression of the *LsTCP4* gene converted wavy leaves (*Lstcp4*) to flat leaves (OE#1 and OE#2). Scale bar *r* = 4 cm. The right panel shows the *LsTCP4* expression in the overexpression lines. The data represent the means ± SDs ($n = 3$). The statistical significance of the differences in gene expression was calculated via Student's *t* test, and the *p* values are shown above the lines connecting two genotypes. **c** Deletion of 1255 bp in the downstream region of the *LsTCP4* gene promoted wavy leaves. Scale bar = 4 cm. RT-qPCR analysis of *LsTCP4* expression in the *LsTCP4* genotype and its mutant CRUTR line (right panel). The statistically significant differences were calculated using Student's *t* test, and the *p* values are shown above the lines connecting two genotypes. **d** Sequences around the target sites in the CRUTR mutant. The dashed lines indicate deletions. The horizontal lines refer to the sgRNA sequences. PAM sequences are shaded in gray

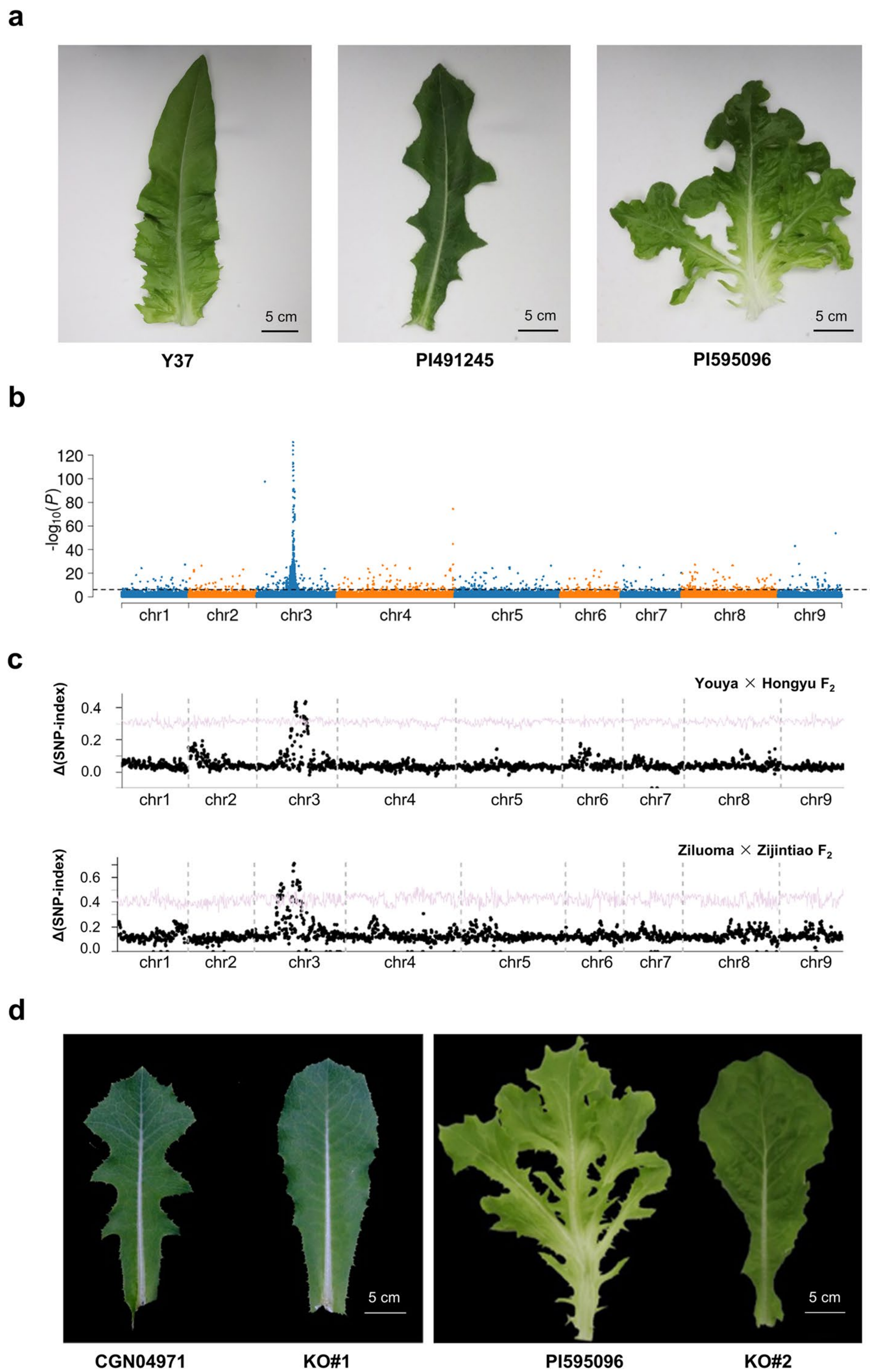

Fig. 5 A gene encoding a zinc finger protein controls the variation of lobed/non-lobed leaves in lettuce. **a** The non-lobed (left), pinnately lobed (middle), and palmately lobed (right) leaves of lettuce. **b** GWAS of lobed leaves. The black horizontal dashed line corresponds to the threshold of statistical significance. **c** BSR of lobed leaves from two F_2 segregating populations (Youya × Hongyu and Ziluoma × Zijintiao). **d** Knockout of the *LsZFP* gene in different lobed types changed lobed leaves to non-lobed leaves

Discussion

The efficiency of genetic studies using the MAGIC population of lettuce

Lettuce is a selfing species that has diverged into several phenotypically distinct horticultural types. Previous studies have shown that lettuce accessions of the same type tend to be closely related. This strong population structure hinders GWAS of important traits using existing cultivars of lettuce [4]. We therefore constructed a MAGIC population for lettuce, which successfully disrupted the population structure. GWAS using the MAGIC population identified significant loci associated with flowering time, leaf color, and leaf morphology, showing that this newly developed population enhanced the efficiency of genetic dissection of complex traits in lettuce.

An important agronomic trait in lettuce is heading, the formation of compact structures due to the unequal cell expansion/division in adaxial and abaxial leaf surfaces. Unfortunately, we could not perform GWAS of the heading phenotype due to the lack of inbred lines in the MAGIC population that formed heads, although two founder lines had the heading phenotype. We had previously identified two loci associated with heading in lettuce [8, 11]. Thirty-four of the 381 inbred lines had heading-promoting alleles of these two genes, *LsKN1*∇ and *Lssaw1*, but none of them developed heads, suggesting that additional loci are required to develop heads in lettuce. Heading or non-heading seems to be a qualitative trait, but it is synergistically controlled by a large number of loci [57, 58]. Loss of an individual heading-promoting allele may convert a heading phenotype to a non-heading one [8, 11]. The lack of heading phenotypes in our MAGIC population is consistent with the hypothesis that a large number of genes are required to develop the heading phenotype in lettuce. We suggest that a higher proportion (for example, half) of the founder lines should form heads when constructing a MAGIC population if heading is the main target trait for GWAS.

High-throughput phenotyping in GWAS

One of the main challenges in GWAS is phenotyping. In this study, we phenotyped the color of lettuce leaves using HSI. The hyperspectral indices of the corresponding peaks may represent the concentrations of anthocyanin or chlorophyll. The traits associated with anthocyanin or chlorophyll can be also phenotyped by other methods including visual inspection. However, hyperspectral imaging-based traits at invisible wavelengths may reflect the concentrations of colorless metabolites [59]; in this case, the use of HSI is highly advantageous over other approaches to detect the controlling loci, although it remains challenging to identify the metabolites and their controlling genes.

The leaf morphology of the 16 founder lines and their 381 inbred progeny was highly variable (Fig. 1). The overlapping of rosette leaves and heterophylly of lettuce make automatic phenotyping a challenge. We measured conventional leaf traits, such as length, width, aspect ratio, area, tip shape, and curvature, but GWAS of these traits revealed only six significant loci (Additional file 1: Table S13). Furthermore, those traits do not explain all the leaf shapes found in the population. To identify additional genes controlling leaf shape, it is critical to develop a phenotyping approach to obtain morphological data that are meaningful in terms of developmental biology.

Phytochromes B and C accelerate flowering in lettuce

Lettuce is a cool season crop that is sensitive to ambient temperature. High temperatures promote the early flowering of lettuce. As a leaf or stem vegetable, early flowering is an unfavorable trait for lettuce. The current global warming trend has led to the frequent occurrence of abnormal weather and sporadic, unpredictable high temperatures, which poses a serious challenge for lettuce cultivation. The use of late-flowering cultivars, particularly late-flowering cultivars at high temperatures, is a practical approach to address this challenge. In this study, we identified a large number of loci associated with flowering time, and we verified two major QTLs, *LsphyB* and *LsphyC*. We showed that natural *LsphyB* and *LsphyC* mutants, as well as their knockout mutants, flowered later than the wild-type genotypes. The loss-of-function mutation of *LsphyB* occurred only in some stem lettuce cultivars, whereas the nonfunctional *LsphyC* allele is present in some leafy lettuce cultivars. It remains to be investigated whether *LsphyB*/*LsphyC* double mutants will be useful in lettuce breeding programs.

The effects of phyB and phyC on flowering time vary dramatically among different plant species. In Arabidopsis, phyB facilitates the degradation of COSTANTS (CO) to downregulate FLOWERING LOCUS T (FT) and consequently delays flowering [18, 60, 61]. As expected, flowering is accelerated in *phyB* mutants in Arabidopsis [18]. Interestingly, the overexpression of *phyB* also delayed flowering in Arabidopsis, since the overexpression of *phyB* attenuated COP1 activity to promote the accumulation of CO protein [18]. In wheat, CO homologs have a limited effect on the photoperiodic response [62]. The earlier heading of wheat under long days (LD) than under short days (SD) is regulated mainly by PHOTOPERIOD1 (PPD1), a pseudoresponse regulator (PRR) with a pseudoreceiver domain and a CCT domain [62]. *PPD1* in wheat is positively regulated by phyB and phyC, and wheat *phyB* and *phyC* mutants flowered later than their wild-type counterparts did [63]. ELF3 interacts with and operates downstream of phyB as a direct transcriptional repressor of PPD1 [64]. It will be interesting to investigate whether the ELF3-PPD1 module is responsible for the role of phyB in promoting flowering in lettuce.

Dozens of loci for flowering time were significant in this GWAS, and many of them were detected in more than one season. A large number of loci for flowering time in a species is not unprecedented [65, 66]. We did not use segregating populations to verify all the significant loci controlling flowering time, and it remains possible that some were false positives. We did verify two major QTLs controlling flowering time in lettuce, namely *LsphyB* and *LsphyC*. The mutated *LsphyC* has been exploited in some cultivars of leafy lettuce, whereas the mutated *LsphyB* is present only in cultivars of stem lettuce. It remains to be investigated whether the mutated *LsphyB* and *LsphyC* genes can be used alternatively or simultaneously in different horticultural types of lettuce. Furthermore, the polymorphic loci controlling leaf shape and color identified in this study will facilitate the development of cultivars rich in anthocyanins with various leaf shapes.

Conclusions

This MAGIC population enabled the efficient identification of beneficial genes in lettuce, which provides the foundation for breeding by design in lettuce. This population, which has been demonstrated to be highly efficient for the genetic analysis of agronomic traits, can be used in the future to dissect the genetics of other important traits, such as nutritional quality and stress tolerance in lettuce.

Methods

Plant materials

Sixteen accessions of lettuce were used as founder lines to construct a MAGIC population. Some of them (named with the prefix of PI) were ordered from the USDA GRIN website (http://www.ars-grin.gov/); others were bought from the local market and maintained in our laboratory. The MAGIC population was constructed by intercrossing for four generations, following the design described by Scott et al. [67]. The hybrids, after four generations of intercrossing, were self-pollinated for six generations to generate 381 inbred progeny using the single seed descent method, as illustrated in Fig. 1b. The detailed information for the founder lines and inbred progeny is shown in Additional file 1: Table S1. The MAGIC population was planted in a protected field in Wuhan, China, in February 2021, November 2021, and November 2022. Each line was represented by 3 to 5 plants.

To genetically map the gene controlling lobed leaves, the cultivar Youya with non-lobed leaves was crossed with the cultivar Hongyu with lobed leaves to develop a segregating population. Similarly, the cultivar Zijintiao with lobed leaves was crossed with the cultivar Ziluoma with non-lobed leaves to construct the second segregating population for lobed leaves.

Phenomics

For hyperspectral imaging (HSI) of a single plant leaf, hyperspectral images containing 224 different spectral bands (401-998 nm) were generated. Following image segmentation and trait calculation, a total of 1793 hyperspectral phenotypic traits were computed, encompassing traits related to total reflectance, average reflectance, and logarithm [59]. Each accession had 2 to 4 biological replicates, and the mean values of the phenotypic traits were calculated (Additional file 1: Table S4, S5).

For RGB imaging of a single plant leaf, top-view RGB images were obtained. Following the automatic calculation of the maximum width, image segmentation, and trait extraction, a total of 33 RGB phenotypic traits were obtained [68]. Each accession had 2 to 3 biological replicates, and the mean values of the phenotypic traits were calculated (Additional file 1: Table S7, S8).

For each 3D scan of a single plant leaf, we utilized a GScan™ 3D scanner (ZG Technology Ltd.) to reconstruct 3D point clouds. Then, through customized scripts, we computed 25 leaf 3D phenotypic traits (Additional file 1: Table S9, S10), including maximum leaf length, maximum leaf width, aspect ratio, surface area, perimeter, curliness, leaf tip angle, average leaf length, average leaf width, leaf width variability, leaf serration, and leaf flatness.

In addition, we manually scored leaf color, stem color, seed color, lobed leaves, and flowering time for the MAGIC population through visual inspection (Additional file 1: Table S11, S12). A total of nine traits were obtained. The phenotypic data were tested for normal distribution before GWAS were performed.

DNA sequencing and data analysis

Genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method from the leaf tissue of 381 inbred progeny and 16 founder lines to construct standardized whole-genome libraries, which were PE100 sequenced on the DNBseq™ high-throughput sequencing platform at BGI-Shenzhen. After sequencing, the raw data from each sample were filtered via Trimmomatic software (version 0.38) [69] to remove adapters and low-quality reads. The resulting data were aligned to the lettuce reference genome (*L. sativa* cv. Salinas v8) [70] using the Sentieon DNAseq workflow (version: sentieon-genomics-201911) [71, 72]. The aligned data were then sorted, and duplicates were removed before variant calling using the Sentieon Haplotyper module, producing variant call format (GVCF) data for each sample. The GVCF data for all samples were merged to create an initial variant dataset using the Sentieon GVCFtyper workflow and genotyped to obtain the initial variant calls for the population samples. The initial variant dataset was filtered for quality using GATK software (version 4.0.10.0) [72] with the recommended hard-filtering criteria, and biallelic variant sites were retained [72, 73]. Among them, only SNPs/indels with a missing rate ≤ 10% and a minimum allele frequency ≥ 5% were retained in the population SNP dataset.

SNPs and indels were annotated using SnpEff (version 5.0e) [74]. SNPs were categorized based on their positions on the chromosomes, including intergenic regions and genic regions. To explore the genetic population structure of the MAGIC population, a principal component analysis (PCA) was performed using GCTA (version 1.93.3 beta) [75]. Phylogenetic relationships were inferred by constructing a neighbor-joining tree (NJtree) by using VCF2Dis (version 1.47, available at https://github.com/BGI-shenzhen/VCF2Dis) in conjunction with the PHYLIP package (version 3.69.650) [76], and the NJtree was visualized using iTOL (version 6.9.1) [77]. Linkage disequilibrium (LD) was assessed within a 1000-kb window via PopLDdecay (version 3.41) [78], and the LD decay was calculated as the distance at which the squared Pearson correlation coefficient (r^2) decreased to half of the maximum.

GWAS

The GWAS was performed using GEMMA (version 0.98.3) [79]. The SNPs with a minor allele frequency ≥ 0.05 and a missing data rate ≤ 10% were retained. The retained SNPs were used for GWAS. To improve the power of the GWAS and avoid confounding factors, the GWAS was conducted using a linear mixed model with a correction of the first ten PCA components and a kinship matrix using GEMMA. The suggestive genome-wide significance thresholds were determined using a uniform threshold of 1/Me, where Me is the effective number of independent SNPs calculated using the Genetic Type 1 Error Calculator (version 1.0) [80]. Variant annotations for the SNP sites were performed using SnpEff (version 5.0e) [74], and Manhattan and QQ plots were generated using the qqman package in R [81].

Neighboring significant regions with less than one LD decay distance were combined. Candidate genes from each significant region were chosen based on their sequence variations (such as SV, nonsense, and missense mutations) and predicted functions. The database of flowering time genes in plants and the eQTLs of flowering time genes in lettuce were also used for the identification of candidate genes [4, 54]. Haplotype analysis of the candidate genes was performed using the geneHapR package in R [82].

Bulked Segregant Analysis (BSA) and RNA-seq

BSA coupled with RNA sequencing (BSR) was used to identify regions controlling flowering time and leaf shape. Twenty individuals with one extreme phenotype and 20 individuals with another extreme phenotype from a segregating population were selected to construct two extreme pools. RNA-seq was performed with the Illumina HiSeq 2500 platform, and approximately 5 Gb of clean data for each pool were obtained. We used hisat2 (version 0.1.6) [83] software to map the clean data to the lettuce reference genome (*L. sativa* cv. Salinas V8) [70] and SAMtools (version 0.1.19) [84] software to identify SNPs between the two pools. The Δ(SNP-index) was calculated by subtracting the SNP index of one pool from that of the other pool. Each peak in the Δ(SNP-index) plot may suggest that a gene contributes to the variation of the trait of interest. All primers of the molecular markers used for genetic mapping are listed in Additional file 1: Table S16.

Vector construction

For knockout assays, we used CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/) to design specific sgRNAs [85]. We amplified PCR fragments from the pCBC-DT1T2 vector or fused multiple sgRNA scaffolds and different tRNAs [86]. Then, we recombined the purified PCR fragments with the *Bsa*I linearized vector pKSE401. After DNA sequence analysis, we transformed the construct into GV3101 through the freeze–thaw method. The primers used in this study are shown in Additional file 1: Table S16.

For the overexpression assays, the coding sequences of the *LsTCP4* gene were amplified and recombined with the *Nde*I linearized pRI101 vector, which uses the CaMV 35S promoter to drive the expression of the inserted sequence. We also transformed the overexpression construct into GV3101 through the freeze–thaw method. The primers used in this study are shown in Additional file 1: Table S16.

Plant transformation

The transformation experiment in lettuce was performed following the procedure described by Curtis [87]. First, we used 50% bleach on sterile seeds for approximately 15 min. After sterilization, we washed the seeds and germinated them on 1/2 strength MS media. We excised the cotyledons of 5-day-old seedlings and immersed them in a suspension of the *Agrobacterium* culture for 10-15 min. We co-cultivated the infected cotyledons in the dark on 1/2 MS media for approximately 2 days. We subsequently transferred the samples to selection media consisting of 1 × MS supplemented with 300 mg/L timentin, 60 mg/L kanamycin, 0.1 mg/L 1-naphthylacetic acid (NAA), and 0.1 mg/L 6-benzylaminopurine (6-BA). After 1 month, we excised the emerging young buds and transferred them to 1/2 MS media supplemented with 300 mg/L Timentin. Finally, we planted the transgenic plants in the soil in the growth room before transfer to the greenhouse.

Gene expression analysis

We extracted total RNA from young leaves using TransZol reagent (TRANSGEN, Beijing, China) following the manufacturer's instructions. The procedures for qRT-PCRs followed the protocols in our previous study [88]. We used the HiScript® II 1st Strand cDNA Synthesis Kit (Vazyme Biotech, Nanjing, China) to obtain cDNA. ChamQ SYBR qPCR Master Mix (Vazyme Biotech, Nanjing, China) was used for qRT-PCR. The reaction system contained 1.0 µl of cDNA, 0.4 µl of primers, 8.2 µl of ddH_2O, and 10.0 µl of 2 × ChamQ SYBR qPCR Master Mix. We conducted qRT-PCR on the QuantStudioTM 6 Real-Time PCR System (Thermo, USA). Three biological replicates and three technical replicates were performed for each reaction. We normalized the expression levels of genes with *ubiquitin* as an internal control. We used the $2^{-\Delta\Delta CT}$ method to calculate the gene expression level [89]. The primers used are listed in Additional file 1: Table S16.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03541-6.

Additional file 1: Table S1: Sequencing and mapping statistics of the MAGIC population. Table S2: Statistics of called SNPs and indels in the MAGIC population. Table S3: Summary of SNP and indel locations in the genome for the MAGIC population. Table S4: The rate of overall genomic heterozygosity for the 381 inbred progeny of the MAGIC population. Table S5: Data types of HSI i-traits. Providing detailed description for each HSI i-trait. Table S6: HSI i-traits of the MAGIC population. Providing the raw data for each HSI i-trait. Table S7: Data types of RGB i-traits. Providing detailed description for each RGB i-trait. Table S8: RGB i-traits of the MAGIC population. Providing the raw data for each RGB i-trait. Table S9: Data types of 3D i-traits. Providing detailed description for each 3D i-trait. Table S10: 3D i-traits of the MAGIC population. Providing the raw data for each 3D i-trait. Table S11: Data types of leaf color, stem color, seed color, lobed leaves, leaf tip shape, and flowering time of the MAGIC population. Describing how these traits were measured and recorded. Table S12: Leaf color, stem color, seed color, lobed leaves, leaf tip shape and flowering time of the MAGIC population. The recorded raw data for each trait. Table S13: Significant loci identified by the GWAS of the lettuce MAGIC population. Table S14: Locus and candidate genes that were significantly associated with the flowering time of the MAGIC population. Table S15: Locus and candidate genes that were significantly associated with the leaf morphology of the MAGIC population. Table S16: Primers used for genetic mapping and verification of gene functions.

Additional file 2: Fig. S1: Images of 2-month-old plants of 381 inbred progeny. Fig. S2: Average mapping depths and average genome coverage rates for the MAGIC population. Fig. S3: Schematic diagram of the genomic bins of six representative inbred progeny. Fig. S4: PCA of the MAGIC population. Fig. S5: Linkage disequilibrium (LD) decay distance for the MAGIC population. Fig. S6: Variations in leaf color, leaf shape, seed color, leaf size, and flowering time in the MAGIC population. Fig. S7: Illustration of HSI, RGB imaging, and 3D scanning. Fig. S8: GWAS of leaf, stem, and seed color in the MAGIC population. Fig. S9: Map-based cloning of the *phyB* gene. Fig. S10: Genetic mapping of the *Lsphyc* gene. Fig. S11: GWAS of leaf curvature, leaf area, lobed leaves, leaf width, leaf length, and tip shape in the MAGIC population. Fig. S12: The haplotype, phenotype, and expression of the leaf length candidate gene *CDC48* (*LG1103550*). Fig. S13: Map-based cloning and identification of *LsTCP4*. Fig. S14: Partial functions of *Lstcp4*. Knockout mutants in *Lstcp4* genotype made leaves further more wavy, suggesting the *Lstcp4* allele retains partial function.

Acknowledgements

We highly appreciate the members of the Lettuce Group at HZAU who participated in the construction of the MAGIC population and its phenotyping. We thank the Germplasm Resources Information Network, US Department of Agriculture, and Centre for Genetic Resources, the Netherlands (CGN), for providing some of the lettuce materials used in this study.

Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

HK conceived the project. HC performed sequencing and bioinformatics with the help of GL, XW, SD, WW, TW, and HL. DL, MW, RM, and JC worked on the cloning of the gene controlling lobed leaves. RZ, ZG, JZ, HF, HC, SD, and WY performed phenomics. TZ, WZ, JA, WL, and JC cloned the PhyB and PhyC genes. YJ and QT cloned the TCP gene controlling leaf flatness. JL and JC worked on the genetics of leaf and stem color. XZ worked on the expression of genes controlling leaf shape. The manuscript was prepared by JC, HC, and HK with the help of RM, WY, SKS and TW. All the authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China awards no. 31830079, 32372719 and U21A20205, the Major Project of Hubei Hongshan Laboratory # 2022hszd020, the Open Project Program of the State Key Laboratory of Agricultural Genomics [No. 2011DQ782025], the Key Projects of the Natural Science Foundation of Hubei Province (2021CFA059), and the GDMPA Science and Technology Innovation Project (2023ZDZ01). This work was also supported by the China National GeneBank (CNGB).

Data availability
All sequencing data generated in this project have been deposited in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under the accession number PRJNA1226568 [90]. This includes the DNA sequences of the 397 individuals from the lettuce MAGIC population, as well as the sequencing data obtained from BSR-seq experiments for leaf curvature, lobed leaves, and flowering time. Source data for this study are available from the corresponding author Hanhui Kuang (kuangfile@mail.hzau.edu.cn) upon request. The genotyping matrix of the lettuce MAGIC population dataset is available at FigShare (https://doi.org/10.6084/m9.figshare.28553273.v1) [91]. No other scripts or software other than those mentioned in the Methods section were used.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Hongyun Chen, Shenzhao Deng, Wandi Wang, Sunil Kumar Sahu, Huan Liu, and Tong Wei are affiliated with BGI Research. The authors declare that they have no competing interests.

Author details
[1]National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops; Hubei Hongshan Laboratory, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China. [2]BGI Research, Wuhan 430074, China. [3]College of Informatics, Huazhong Agricultural University, Wuhan 430070, China. [4]Genome Center and Department of Plant Sciences, University of California, Davis, CA 95616, USA. [5]National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China. [6]State Key Laboratory of Genome and Multi-Omics Technologies, BGI Research, Shenzhen 518083, China. [7]Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen 518083, China. [8]Guangdong Provincial Key Laboratory of Core Collection of Crop Genetic Resources Research and Application, BGI Research, Shenzhen 518083, China.

Received: 4 October 2024 Accepted: 11 March 2025
Published online: 23 March 2025

References
1. An GH, Qi YT, Zhang WY, Gao HR, Qian JL, Larkin RM, Chen JJ, Kuang HH. *LsNRL4* enhances photosynthesis and decreases leaf angles in lettuce. Plant Biotechnol J. 2022;20:1956–67.
2. Gurdon C, Kozik A, Tao R, Poulev A, Armas I, Michelmore RW, Raskin I. Isolating an active and inactive CACTA transposon from lettuce color mutants and characterizing their family. Plant Physiol. 2021;186:929–44.
3. Su WQ, Tao R, Liu WY, Yu CC, Yue Z, He SP, Lavelle D, Zhang WY, Zhang L, An GH, et al. Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. Plant Biotechnol J. 2020;18:479–90.
4. Zhang L, Su WQ, Tao R, Zhang WY, Chen JJ, Wu PY, Yan CH, Jia Y, Larkin RM, Lavelle D, et al. RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. Nat Commun. 2017;8:2264.
5. Zhang L, Qian JL, Han YT, Jia Y, Kuang HH, Chen JJ. Alternative splicing triggered by the insertion of a CACTA transposon attenuates *LsGLK* and leads to the development of pale-green leaves in lettuce. Plant J. 2022;109:182–95.
6. Gurdon C, Poulev A, Armas I, Satorov S, Tsai M, Raskin I. Genetic and phytochemical characterization of lettuce flavonoid biosynthesis mutants. Sci Rep. 2019;9:3305.
7. Nakata M, Okada K. The leaf adaxial-abaxial boundary and lamina growth. Plants. 2013;2:174–202.
8. An GH, Yu CC, Yan CH, Wang ML, Zhang WY, Jia Y, Shi CM, Larkin RM, Chen JJ, Lavelle D, et al. Loss-of-function of *SAWTOOTH 1* affects leaf dorsiventrality genes to promote leafy heads in lettuce. Plant Cell. 2022;34:4329–47.
9. Jia Y, Yu P, Shao W, An GH, Chen JJ, Yu CC, Kuang HH. Up-regulation of *LsKN1* promotes cytokinin and suppresses gibberellin biosynthesis to generate wavy leaves in lettuce. J Exp Bot. 2022;73:6615–29.
10. Wang ML, Lavelle D, Yu CC, Zhang WY, Chen JJ, Wang X, Michelmore RW, Kuang HH. The upregulated *LsKN1* gene transforms pinnately to palmately lobed leaves through auxin, gibberellin, and leaf dorsiventrality pathways in lettuce. Plant Biotechnol J. 2022;20:1756–69.
11. Yu CC, Yan CH, Liu YL, Liu YL, Jia Y, Lavelle D, An GH, Zhang WY, Zhang L, Han RK, et al. Upregulation of a *KN1* homolog by transposon insertion promotes leafy head development in lettuce. Proc Natl Acad Sci U S A. 2020;117:33668–78.

12. Seki K, Komatsu K, Tanaka K, Hiraga M, Kajiya-Kanegae H, Matsumura H, Uno Y. A CIN-like TCP transcription factor (LsTCP4) having retrotransposon insertion associates with a shift from salinas type to empire type in crisphead lettuce (Lactuca sativa L.). Hortic Res. 2020;7:15.
13. Han RK, Truco MJ, Lavelle DO, Michelmore RW. A composite analysis of flowering time regulation in lettuce. Front Plant Sci. 2021;12:632708.
14. Kim DH. Current understanding of flowering pathways in plants: focusing on the vernalization pathway in Arabidopsis and several vegetable crop plants. Hortic Environ Biote. 2020;61:209–27.
15. Osnato M, Cota I, Nebhnani P, Cereijo U, Pelaz S. Photoperiod control of plant growth: flowering time genes beyond flowering. Front Plant Sci. 2022;12:805635.
16. Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science. 2000;288:1613–6.
17. Ding JH, Zhang B, Li Y, André D, Nilsson O. Phytochrome B and PHYTOCHROME INTERACTING FACTOR8 modulate seasonal growth in trees. New Phytol. 2021;232:2339–52.
18. Hajdu A, Adám É, Sheerin DJ, Dobos O, Bernula P, Hiltbrunner A, Kozma-Bognár L, Nagy F. High-level expression and phosphorylation of phytochrome B modulates flowering time in Arabidopsis. Plant J. 2015;83:794–805.
19. Monte E, Alonso JM, Ecker JR, Zhang YL, Li X, Young J, Austin-Phillips S, Quail PH. Isolation and characterization of *phyC* mutants in Arabidopsis reveals complex crosstalk between phytochrome signaling pathways. Plant Cell. 2003;15:1962–80.
20. Nishida H, Ishihara D, Ishii M, Kaneko T, Kawahigashi H, Akashi Y, Saisho D, Tanaka K, Handa H, Takeda K, Kato K. *Phytochrome C* is a key factor controlling long-day flowering in barley. Plant Physiol. 2013;163:804–14.
21. Takano M, Inagaki N, Xie XZ, Yuzurihara N, Hihara F, Ishizuka T, Yano M, Nishimura M, Miyao A, Hirochika H, Shinomura T. Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. Plant Cell. 2005;17:3311–25.
22. Chen A, Li CX, Hu W, Lau MY, Lin HQ, Rockwell NC, Martin SS, Jernstedt JA, Lagarias JC, Dubcovsky J. Phytochrome C plays a major role in the acceleration of wheat flowering under long-day photoperiod. Proc Natl Acad Sci U S A. 2014;111:10037–44.
23. Pearce S, Kippes N, Chen A, Debernardi JM, Dubcovsky J. RNA-seq studies using wheat *PHYTOCHROME B* and *PHYTOCHROME C* mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. BMC Plant Biol. 2016;16:141.
24. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20:467–84.
25. Liang YM, Liu HJ, Yan JB, Tian F. Natural variation in crops: realized understanding, continuing promise. Annu Rev Plant Biol. 2021;72:357–85.
26. Chen HZ, Zhai LY, Chen K, Shen CC, Zhu SB, Qu PP, Tang J, Liu JP, He HH, Xu JL. Genetic background- and environment-independent QTL and candidate gene identification of appearance quality in three MAGIC populations of rice. Front Plant Sci. 2022;13:1074106.
27. Islam MS, Thyssen GN, Jenkins JN, Zeng LH, Delhom CD, McCarty JC, Deng DD, Hinchliffe DJ, Jones DC, Fang DD. A MAGIC population-based genome-wide association study reveals functional association of *GhRBB1_A07* gene with superior fiber quality in cotton. BMC Genomics. 2016;17:903.
28. Wang MJ, Qi ZY, Thyssen GN, Naoumkina M, Jenkins JN, McCarty JC, Xiao YJ, Li JY, Zhang XL, Fang DD. Genomic interrogation of a MAGIC population highlights genetic factors controlling fiber quality traits in cotton. Commun Biol. 2022;5:60.
29. Yuan GD, Sun KF, Yu WL, Jiang ZP, Jiang CH, Liu D, Wen LY, Si H, Wu FY, Meng H, et al. Development of a MAGIC population and high-resolution quantitative trait mapping for nicotine content in tobacco. Front Plant Sci. 2023;13:1086950.
30. Burgos E, De Luca MB, Diouf I, de Haro LA, Albert E, Sauvage C, Tao ZJ, Bermudez L, Asís R, Nesi AN, et al. Validated MAGIC and GWAS population mapping reveals the link between vitamin E content and natural variation in chorismate metabolism in tomato. Plant J. 2021;105:907–23.
31. Wei T, van Treuren R, Liu XJ, Zhang ZW, Chen JJ, Liu Y, Dong SS, Sun PN, Yang T, Lan TM, et al. Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. Nat Genet. 2021;53:752–60.
32. Upadhyay N, Kar D, Mahajan BD, Nanda S, Rahiman R, Panchakshari N, Bhagavatula L, Datta S. The multitasking abilities of MATE transporters in plants. J Exp Bot. 2019;70:4643–56.
33. Xu ZW, Escamilla-Treviño LL, Zeng LH, Lalgondar M, Bevan DR, Winkel BSJ, Mohamed A, Cheng CL, Shih MC, Poulton JE, Esen A. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. Plant Mol Biol. 2004;55:343–67.
34. Oda-Yamamizo C, Mitsuda N, Sakamoto S, Ogawa D, Ohme-Takagi M, Ohmiya A. The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves. Sci Rep. 2016;6:23609.
35. Howe CJ, Schlarb-Ridley BG, Wastl J, Purton S, Bendall DS. The novel cytochrome c_6 of chloroplasts: a case of evolutionary *bricolage*? J Exp Bot. 2006;57:13–22.
36. Li DP, Zhu HF, Liu KF, Liu X, Leggewie G, Udvardi M, Wang DW. Purple acid phosphatases of *Arabidopsis thaliana*: Comparative analysis and differential regulation by phosphate deprivation. J Biol Chem. 2002;277:27772–81.
37. Xie S, Luo GB, An GH, Wang BC, Kuang HH, Wang X. Lskipk Lsatpase double mutants are necessary and sufficient for the compact plant architecture of butterhead lettuce. Hortic Res. 2024;11:uhad280.
38. Zhang CG, Liu HH, Zong YX, Tu ZH, Li HG. Isolation, expression, and functional analysis of the geranylgeranyl pyrophosphate synthase (GGPPS) gene from *Liriodendron tulipifera*. Plant Physiol Biochem. 2021;166:700–11.
39. Savidge B, Rounsley SD, Yanofsky MF. Temporal relationship between the transcription of two Arabidopsis MADS box genes and the floral organ identity genes. Plant Cell. 1995;7:721–33.
40. Hassidim M, Harir Y, Yakir E, Kron I, Green RM. Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. Planta. 2009;230:481–91.

41. Hirose F, Shinomura T, Tanabata T, Shimada H, Takano M. Involvement of rice cryptochromes in de-etiolation responses and flowering. Plant Cell Physiol. 2006;47:915–25.
42. Noh B, Lee SH, Kim HJ, Yi G, Shin EA, Lee M, Jung KJ, Doyle MR, Amasino RM, Noh YS. Divergent roles of a pair of homologous Jumonji/Zinc-Finger-Class transcription factor proteins in the regulation of Arabidopsis flowering time. Plant Cell. 2004;16:2601–13.
43. Tanaka N, Itoh H, Sentoku N, Kojima M, Sakakibara H, Izawa T, Itoh JI, Nagato Y. The *COP1* ortholog *PPS* regulates the juvenile-adult and vegetative-reproductive phase changes in rice. Plant Cell. 2011;23:2143–54.
44. Lu S, Zhang N, Xu YZ, Chen H, Huang J, Zou BH. Functional conservation and divergence of *MOS1* that controls flowering time and seed size in rice and *Arabidopsis*. Int J Mol Sci. 2022;23:13448.
45. Deng WW, Liu CY, Pei YX, Deng X, Niu LF, Cao XF. Involvement of the histone acetyltransferase AtHAC1 in the regulation of flowering time via repression of *FLOWERING LOCUS C* in Arabidopsis. Plant Physiol. 2007;143:1660–8.
46. Li CX, Lin HQ, Debernardi JM, Zhang CZ, Dubcovsky J. *GIGANTEA* accelerates wheat heading time through gene interactions converging on *FLOWERING LOCUS T1*. Plant J. 2024;118:519–33.
47. Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ. The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in Arabidopsis. Planta. 2008;228:709–23.
48. He J, Chen QW, Xin PY, Yuan J, Ma YH, Wang XM, Xu MM, Chu JF, Peters RJ, Wang GD. CYP72A enzymes catalyse 13-hydrolyzation of gibberellins. Nat Plants. 2019;5:1057–65.
49. Gregis V, Sessa A, Colombo L, Kater MM. *AGL24*, *SHORT VEGETATIVE PHASE*, and *APETALA1* redundantly control *AGAMOUS* during early stages of flower development in *Arabidopsis*. Plant Cell. 2006;18:1373–82.
50. Liu C, Zhou J, Bracha-Drori K, Yalovsky S, Ito T, Yu H. Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. Development. 2007;134:1901–10.
51. Li W, Ahn IP, Ning YS, Park CH, Zeng LR, Whitehill JGA, Lu HB, Zhao QZ, Ding B, Xie Q, et al. The U-Box/ARM E3 ligase PUB13 regulates cell death, defense, and flowering time in Arabidopsis. Plant Physiol. 2012;159:239–50.
52. López-González L, Mouriz A, Narro-Diego L, Bustos R, Martínez-Zapater JM, Jarillo JA, Piñeiro M. Chromatin-dependent repression of the *Arabidopsis* floral integrator genes involves plant specific PHD-containing proteins. Plant Cell. 2014;26:3922–38.
53. Kawamoto N, Sasabe M, Endo M, Machida Y, Araki T. Calcium-dependent protein kinases responsible for the phosphorylation of a bZIP transcription factor FD crucial for the florigen complex formation. Sci Rep. 2015;5: 8341.
54. Liu D, Li J, Wang S, Huang T, Tao F, Lin Y, Lin W, Zhao X, Huang Y, Jia Y. PlantCFG: A comprehensive database with web tools for analyzing candidate flowering genes in multiple plants. Plant Commun. 2024;5: 100733.
55. Park S, Rancour DM, Bednarek SY. In planta analysis of the cell cycle-dependent localization of AtCDC48A and its critical roles in cell division, expansion, and differentiation. Plant Physiol. 2008;148:246–58.
56. Guo P, Yoshimura A, Ishikawa N, Yamaguchi T, Guo Y, Tsukaya H. Comparative analysis of the RTFL peptide family on the control of plant organogenesis. J Plant Res. 2015;128:497–510.
57. Bassett MJ. The role of leaf shape in the inheritance of heading in lettuce (Lactuca sativa L.). J Am Soc Hortic Sci. 1975;100:104–5.
58. Robinson RW, McCreight JD, Ryder EJ. The genes of lettuce and closely related species. In: Janick J, editor. Plant Breed Rev. Boston: Springer, US; 1983. p. 267–93.
59. Wu X, Feng H, Wu D, Yan SJ, Zhang P, Wang WB, Zhang J, Ye JL, Dai GX, Fan Y, et al. Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. Genome Biol. 2021;22:185.
60. Endo M, Nakamura S, Araki T, Mochizuki N, Nagatani A. Phytochrome B in the mesophyll delays flowering by suppressing *FLOWERING LOCUS T* expression in Arabidopsis vascular bundles. Plant Cell. 2005;17:1941–52.
61. Lazaro A, Mouriz A, Piñeiro M, Jarillo JA. Red light-mediated degradation of CONSTANS by the E3 ubiquitin ligase HOS1 regulates photoperiodic flowering in Arabidopsis. Plant Cell. 2015;27:2437–54.
62. Shaw LM, Li CX, Woods DP, Alvarez MA, Lin HQ, Lau MY, Chen A, Dubcovsky J. Epistatic interactions between *PHOTOPERIOD1*, *CONSTANS1* and *CONSTANS2* modulate the photoperiodic response in wheat. PLoS Genet. 2020;16: e1008812.
63. Pearce S, Shaw LM, Lin HQ, Cotter JD, Li CX, Dubcovsky J. Night-break experiments shed light on the photoperiod1-mediated flowering. Plant Physiol. 2017;174:1139–50.
64. Alvarez MA, Li CX, Lin HQ, Joe A, Padilla M, Woods DP, Dubcovsky J. *EARLY FLOWERING 3* interactions with *PHYTOCHROME B* and *PHOTOPERIOD1* are critical for the photoperiodic regulation of wheat heading time. PLoS Genet. 2023;19:e1010655.
65. Arrones A, Vilanova S, Plazas M, Mangino G, Pascual L, Díez MJ, Prohens J, Gramazio P. The dawn of the age of multi-parent MAGIC populations in plant breeding: novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. Biology. 2020;9: 229.
66. Fourquet L, Barber T, Campos-Mantello C, Howell P, Orman-Ligeza B, Percival-Alwyn L, Rose GA, Sheehan H, Wright TIC, Longin F, et al. An eight-founder wheat MAGIC population allows fine-mapping of flowering time loci and provides novel insights into the genetic control of flowering time. Theor Appl Genet. 2024;137:277.
67. Scott MF, Ladejobi O, Amer S, Bentley AR, Biernaskie J, Boden SA, Clark M, Dell'Acqua M, Dixon LE, Filippi CV, et al. Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity. 2020;125:396–416.
68. Li HT, Feng H, Guo CC, Yang SJ, Huang W, Xiong X, Liu JX, Chen GX, Liu Q, Xiong LZ, et al. High-throughput phenotyping accelerates the dissection of the dynamic genetic architecture of plant growth and yield improvement in rapeseed. Plant Biotechnol J. 2020;18:2345–53.
69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
70. Reyes-Chin-Wo S, Wang ZW, Yang XH, Kozik A, Arikit S, Song C, Xia LF, Froenicke L, Lavelle DO, Truco MJ, et al. Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. Nat Commun. 2017;8:14953.

71. Kendig KI, Baheti S, Bockol MA, Drucker TM, Hart SN, Heldenbrand JR, Hernaez M, Hudson ME, Kalmbach MT, Klee EW, et al. Sentieon DNASeq variant calling workflow demonstrates strong computational performance and accuracy. Front Genet. 2019;10: 736.
72. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
73. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
74. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. Fly. 2012;6:80–92.
75. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
76. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Seattle: Department of Genome Sciences, University of Washington; 2005.
77. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:78–82.
78. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35:1786–8.
79. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.
80. Li MX, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2012;131:747–56.
81. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J Open Source Softw. 2014;3:731.
82. Zhang RL, Jia GQ, Diao XM. geneHapR: an R package for gene haplotypic statistics and visualization. BMC Bioinf. 2023;24:199.
83. Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
84. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
85. Liu H, Ding YD, Zhou YQ, Jin WQ, Xie KB, Chen LL. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. Mol Plant. 2017;10:530–2.
86. Wang YQ, Li XF, Liu ML, Zhou YJ, Li F. Guide RNA scaffold variants enabled easy cloning of large gRNA cluster for multiplexed gene editing. Plant Biotechnol J. 2024;22:460–71.
87. Curtis IS. Lettuce (*Lactuca sativa* L.). Walker JM. Methods Mol Biol. 2006;343:449–58.
88. Tang QW, Kuang HH, Yu CC, An GH, Tao R, Zhang WY, Jia Y. Non-vernalization requirement in Chinese kale caused by loss of *BoFLC* and low expressions of its paralogs. Theor Appl Genet. 2022;135:473–83.
89. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25:402–8.
90. Chen HY, Chen JJ, Wei T, Kuang HH. DNA sequences of the 397 individuals from the lettuce MAGIC population and the sequencing data obtained from BSR-seq experiments. PRJNA1226568. Sequence read archive. 2025. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1226568.
91. Chen HY, Chen JJ, Wei T, Kuang HH. The genotyping matrix of the lettuce MAGIC population dataset. Dataset figshare. 2025. https://doi.org/10.6084/m9.figshare.28553273.v1.

Publisher's Note