RESEARCH

Open Access

DNA methylation variations underlie lettuce domestication and divergence

Shuai Cao[1], Nunchanoke Sawettalake[1], Ping Li[1], Sheng Fan[1] and Lisha Shen[1,2]*

*Correspondence:
lisha@tll.org.sg

[1] Temasek Life Sciences
Laboratory, 1 Research Link,
National University of Singapore,
Singapore 117604, Singapore
[2] Department of Biological
Sciences, Faculty of Science,
National University of Singapore,
Singapore 117543, Singapore

Abstract

Background: Lettuce (*Lactuca sativa L.*) is an economically important vegetable crop worldwide. Lettuce is believed to be domesticated from a single wild ancestor *Lactuca serriola* and subsequently diverged into two major morphologically distinct vegetable types: leafy lettuce and stem lettuce. However, the role of epigenetic variation in lettuce domestication and divergence remains largely unknown.

Results: To understand the genetic and epigenetic basis underlying lettuce domestication and divergence, we generate single-base resolution DNA methylomes from 52 *Lactuca* accessions, including major lettuce cultivars and wild relatives. We find a significant increase of DNA methylation during lettuce domestication and uncover abundant epigenetic variations associated with lettuce domestication and divergence. Interestingly, DNA methylation variations specifically associated with leafy and stem lettuce are related to regulation and metabolic processes, respectively, while those associated with both types are enriched in stress responses. Moreover, we reveal that domestication-induced DNA methylation changes could influence expression levels of nearby and distal genes possibly through affecting chromatin accessibility and chromatin loop.

Conclusion: Our study provides population epigenomic insights into crop domestication and divergence and valuable resources for further domestication for diversity and epigenetic breeding to boost crop improvement.

Keywords: Lettuce, DNA methylation, Domestication, Divergence, Epigenetic variation

Background

Lettuce (*Lactuca sativa* L.), an important representative of the *Asteraceae* family, is one of the most commonly cultivated and consumed vegetable crops worldwide and serves as a prominent natural source of phytonutrients for humans [1]. As a morphologically and genetically diverse vegetable crop, modern lettuce cultivars could be classified into several distinct horticultural types based on their morphological features, including butterhead, crisp, cos (also known as romaine), looseleaf, Latin, stem (also called stalk), and

oilseed lettuce [2]. Among these, the former five types are collectively known as the leafy type as their leaves are commonly consumed, while the latter two are referred to as the non-leafy type as stem or seeds are harvested for consumption or oil production in stem or oilseed lettuce, respectively. Leafy and stem lettuce, the two main types of cultivated lettuce, are globally favored vegetables, particularly in the United States, where they rank among the top vegetable crops in terms of production and consumption (https://www.fao.org/). Despite great morphological variations, different horticultural types of lettuce share some common domestication traits, such as absence of leaf prickles, non-shattering seed pods, and late bolting [3, 4]. Different types of cultivated lettuce are believed to originate from a single domestication event from the wild progenitor prickly lettuce (*L. serriola*) near the Caucasus in middle east of Asia at approximately 4000 BC [5–7]. The initial domestication and subsequent diversification of lettuce give rise to distinct plant morphologies, such as leaf and stem shapes and head formation, which also involves genomic introgressions from its wild relative species, like *L. serriola* and *L. virosa* [6, 8, 9]. Three wild relatives of lettuce, *L. serriola*, *L. virosa*, and *L. saligna*, are compatible with cultivated lettuce to different degrees, serving as important sources of novel traits in lettuce breeding [2, 9, 10]. Although abundant genetic variations during lettuce domestication have been revealed by a large-scale whole genome resequencing study [6], whether epigenetics contribute to shape the domestication traits remains unexplored.

As a conserved and pervasive epigenetic mark in most eukaryotes, DNA methylation at the C-5 position of cytosine acts as a key regulator of gene expression and modulates numerous biological processes [11–13]. In plants, DNA methylation occurs in CG, CHG, and CHH (H = A, T, or C) contexts which are maintained by different mechanisms [13]. In the model plant *Arabidopsis thaliana*, methylation of CG and CHG is maintained by METHYLTRANSFERASE 1 (MET1) [14] and CHROMOMETHYLASE 3 (CMT3) [15], respectively, while de novo CHH methylation is established by DOMAINS REARRANGED METHYLASE 2 (DRM2) through RNA-directed DNA methylation or CMT2 [13, 16]. While CHH methylation patterns exhibit developmental or tissue-specific variations in plants, CG and CHG methylations are quite stable across different vegetative tissues [17–23].

DNA methylation changes give rise to meiotically stable epialleles, which could be transmitted to offspring through natural evolution and artificial selection [24]. CG methylation inherited from parents exhibits stable transgenerational inheritance over several generations in plant genomes [25, 26]. Epialleles via DNA methylation in gene promoters have been shown to confer symmetric flower development in yellow toadflax (*Linaria vulgaris*) [27], fruit non-ripening phenotype, and change in vitamin E content in tomato [28, 29]. Despite its critical function, the role of DNA methylation in crop evolution is just beginning to be explored. Population epigenetic studies on maize [30, 31], soybean [32], and rice [33] have revealed epigenetic evolution in the long-term crop domestication processes. However, epigenome resources for vegetable crops are very limited, especially at the population level.

In this study, we report single-base resolution DNA methylomes from a natural population of 52 *Lactuca* accessions including major lettuce cultivars and wild relatives. We have defined the comprehensive DNA methylation landscape of the *Lactuca* genus and discovered DNA methylation variations that are associated with the domestication and

divergence of different horticultural types of lettuce. By an integrated analysis of DNA methylomes, transcriptomes, and chromatin accessibility and interaction profiles, we reveal the effect of DNA methylation variations on influencing gene expression. Our data provide valuable epigenomic resources for vegetable improvement through epigenetic engineering.

Results

DNA methylation variations in the *Lactuca* genus

To uncover the DNA methylation footprints during lettuce domestication, we generated single-base resolution DNA methylomes of 52 *Lactuca* accessions reflecting most of morphological diversity in our collection, including 28 accessions of lettuce cultivars *L. sativa* (10 crisp, 6 butterhead, 6 cos, and 6 stem lettuce) and 24 accessions of wild lettuce relatives (15 *L. serriola*, 5 *L. virosa*, 2 *L. saligna*, and 2 *L. indica*) (Fig. 1a; Additional File 1: Table S1). We performed whole-genome bisulfite sequencing (WGBS) using leaves from seedlings at 30 days after planting (DAP). Meanwhile, a total of 600 Gb whole-genome sequencing (WGS) data of these accessions were generated for genomic variant calling. To exclude the effect of genomic variations on the DNA methylation analysis, we used single nucleotide polymorphisms (SNPs) of each accession to construct their corresponding pseudo-reference genome for subsequent methylation analysis [30, 34]. Nearly 600 Gb of WGBS data of all accessions were mapped to pseudo-reference genomes. For each accession, 72 million WGBS paired-end reads (> 10 Gb) on average were generated for methylation cytosine calling. We found that the percentages of unmethylated cytosines at non-CG sites (> 17.2% for CHG and > 80.2% for CHH) were much higher than that at CG sites (< 6.89%) (Additional file 2: Fig. S1a). This observation, together with the feature of meiotically stable inheritance of CG methylation across the plant life-cycle [12, 23], promoted us to focus on analyzing the variation and evolution of CG methylation during lettuce domestication, with concurrent analysis of CHG and CHH methylation when needed.

Methylation levels at all CG sites were used to generate a neighbor-joining phylogenetic tree (Fig. 1b). This phylogenetic tree recapitulated the known evolutionary relationships of lettuce species [6], including an ancestral role of *L. serriola* for lettuce cultivars and a distant relationship between the wild relatives (*L. virosa* and *L. indica*) and cultivars (Fig. 1b). Remarkably, *L. saligna* were classified to the stem lettuce (Fig. 1b), which is different from the classifications based on SNPs (Additional file 2: Fig. S1b). Besides, phylogenetic trees based on DNA methylation levels of all CHG or CHH sites are similar to that based on CG sites (Additional file 2: Fig. S2). Moreover, the phylogenetic relationships of different lettuce groups were also supported by principal component analysis (PCA) using CG methylation levels (Fig. 1c), in which lettuce cultivars were more closely related to *L. serriola* than *L. virosa*, *L. indica*, or *L. saligna*. However, accessions of each cultivar type exhibited distribution diversities that were similar to wild lettuce accessions (Fig. 1c). In addition, the diversity (π) estimated based on CG methylation variation was higher than that based on SNPs in both lettuce cultivars *L. sativa* and the wild lettuce *L. serriola* (Fig. 1d). *L. sativa* displayed slightly higher diversity of CG methylation variation as compare with *L. serriola*, in contrast to the observed significantly lower SNP diversity
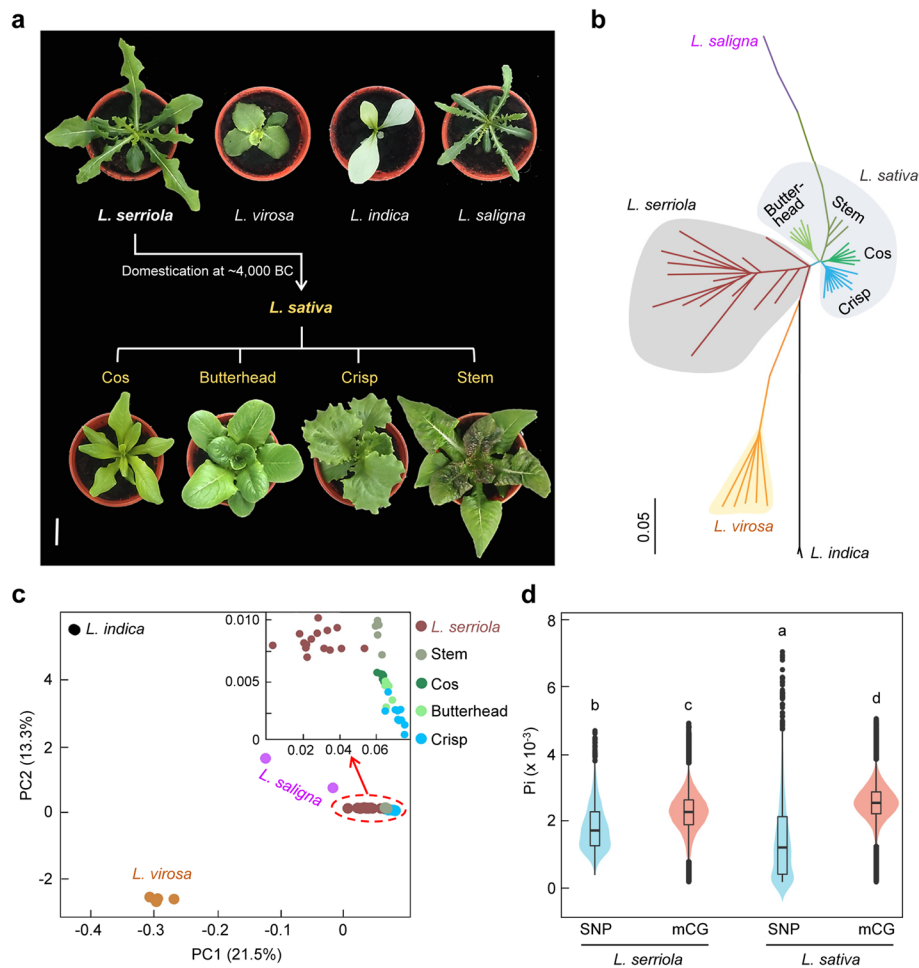
Fig. 1 Increased CG methylation diversity during lettuce domestication. **a** Seedling morphology of the wild relative species (*L. serriola*, *L. virosa*, *L. indica*, and *L. saligna*) and cultivated lettuce including cos, butterhead, crisp, and stem lettuce at 2 weeks after planting. Different types of cultivated lettuce are believed to originate from a single domestication event from *L. serriola* at approximately 4000 BC [5, 6]. Scale bar, 10 cm. **b** A neighbor-joining phylogenetic tree of 52 *Lactuca* accessions based on the methylation levels of all CG loci. **c** Principal component analysis (PCA) of all *Lactuca* accessions based on CG methylation levels. **d** Diversity of CG methylation (mCG) and SNP variations of the wild lettuce (*L. serriola*) and cultivated lettuce (*L. sativa*). Different letters above the violins indicate significant differences ($P < 0.01$, two-sided Wilcoxon signed-rank test) in a pairwise comparison

in *L. sativa* than *L. serriola* (Fig. 1d) [6], suggesting that lettuce domestication is associated with rapid methylation evolution compared with genomic variation.

A global increase of DNA methylation during lettuce domestication

Modern lettuce cultivars are believed to be domesticated from the wild lettuce *L. serriola* [3, 4, 6, 7]; we thus compared methylation levels in cultivars with *L. serriola*. We found that global CG methylation levels were significantly increased in cultivated lettuce, including cos, butter, crisp, and stem lettuce, compared with *L. serriola* (Fig. 2a). Similarly, CHG and CHH methylation levels were also increased during lettuce domestication (Additional file 2: Fig. S3). CG methylation changes were enriched in euchromatic regions with protein-coding genes, especially far away from pericentromeric regions

Cao *et al. Genome Biology*   (2024) 25:158

Page 5 of 22
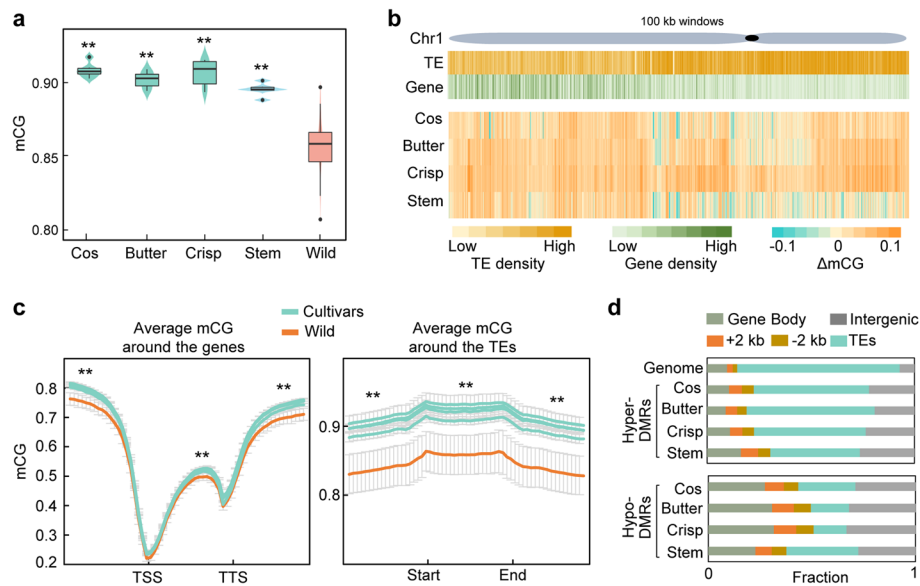
Fig. 2 Elevated CG methylation levels during lettuce domestication. **a** Increased CG methylation levels in cultivated lettuces compared to the wild lettuce *L. serriola* (wild). Asterisks indicate significant differences of CG methylation levels (mCG) in cos, butterhead (butter), crisp, and stem lettuce as compared with the wild lettuce *L. serriola* (**P < 0.01, two-tailed paired Student's *t*-test). **b** Heatmap of methylation level changes along chromosome 1 (Chr1) during lettuce domestication. Methylation level changes (ΔmCG) in butterhead vs. *L. serriola* (butter), cos vs. *L. serriola* (cos), crisp vs. *L. serriola* (crisp), and stem vs. *L. serriola* (stem) were calculated for each 100-kb window. Distribution of genes and TEs along Chr 1 is shown above. **c** Average CG methylation levels around genes (left) and TEs (right). TSS indicates transcription start site, TTS indicates transcription termination site. Asterisks indicate significant differences (**P < 0.01, Wilcoxon signed-rank test) of CG methylation levels in cultivars as compared with the wild lettuce *L. serriola* (wild). Cultivars include cos, butterhead, crisp and stem lettuce. See Additional file 2: Fig. S1, c and d, for the comparison of each cultivar with the wild lettuce *L. serriola*. **d** Distribution of DMRs in different genomic regions divided into gene body, +2 kb flanking region (2 kb upstream of TSS), − 2 kb flanking region (2 kb downstream of TTS), TEs, and intergenic regions excluding TEs. The average distributions of different genomic regions across the whole genome are shown as references

(Fig. 2b). To better understand the distribution of CG methylation changes in different regions of protein-coding genes and TEs, we calculated average methylation levels for every 100-bp interval of each gene and TE, encompassing 2-kb upstream and downstream flank regions. This analysis revealed that CG methylation levels increased greatly in the 5′ and 3′ regions and slightly in gene bodies, while they were much higher across the whole TE regions in all four horticultural types of lettuce cultivars as compared with *L. serriola* (Fig. 2c; Additional file 2: Fig. S1c,d).

To further explore the role of CG DNA methylation during lettuce domestication, we determined the differentially methylated regions (DMRs) between lettuce cultivars and the wild lettuce *L. serriola*. We identified 25,536 hyper- and 4656 hypo-DMRs in the cos group (totally 6.1 Mb, 0.24% of genome), 25,257 hyper- and 5661 hypo-DMRs in the butterhead group (totally 6.0 Mb, 0.24% of genome), 37,138 hyper- and 10,330 hypo-DMRs in the crisp group (totally 8.6 Mb, 0.34% of genome), and 7610 hyper- and 6068 hypo-DMRs in the stem group (totally 2.4 Mb, 0.09% of genome), relative to the wild lettuce *L. serriola* (Additional file 1: Table S3; Additional file 2: Fig. S4a). On average, the count of hyper-DMRs was approximately 3.5-fold higher than that of hypo-DMRs, in line with the observed global increase of methylation in lettuce cultivars (Fig. 2a). We randomly

Cao *et al. Genome Biology*   (2024) 25:158

Page 6 of 22

selected 36 domestication-related DMRs and most of them (34 out 36) were further confirmed by quantitative real-time PCR (qPCR; Additional file 1: Table S4), confirming the reliability of the identified DMRs. We further analyzed the distribution of DMRs across genomic features and found that hyper-DMRs were more prevalent in intergenic regions and 5′ and 3′ flanking regions of coding sequences compared to their average distributions across the whole genome (Fig. 2d). In contrast, most of the hypo-DMRs were enriched in coding sequences, 5′ and 3′ flanking regions, and intergenic regions (Fig. 2d).

Conservation and divergence of methylation changes in cultivated lettuce

We next examined the correlation of CG methylation changes of all DMRs including both hyper- and hypo-DMRs among different types of lettuce cultivars. Notably, correlation coefficients of 0.85–0.94 were observed among pairwise comparisons of the three leafy types: cos, butterhead, and crisp lettuce (Fig. 3a). Compared to the high correlation coefficients among leaf types, lower correlation coefficients of 0.73–0.78 were found between stem lettuce and leafy types (Fig. 3a). This result indicates high degree of conservation of methylation changes among different cultivar types, which is not surprising as they are suggested to be originated from a single domestication event [5, 6]. However,
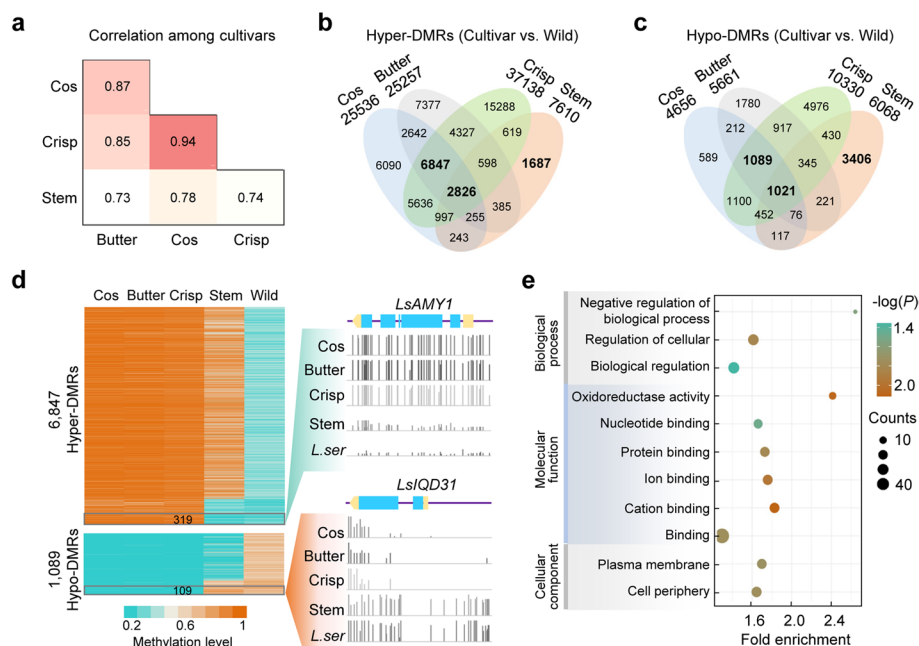
Fig. 3 High correlation of DNA methylation changes among leafy lettuce cultivars. **a** Correction of CG methylation changes among lettuce cultivars, relative to the wild lettuce *L. serriola*. **b**, **c** Shared and unique hyper- (**b**) and hypo- (**c**) DMRs in different types of lettuce cultivars, relative to the wild lettuce *L. serriola*. **d** Heatmap showing methylation levels of type-specific hyper- and hypo-DMRs found in leafy lettuce. Leafy lettuces include cos, butterhead (butter), and crisp lettuce. Grey blocks indicate DMRs showing significant differential methylation levels between leafy lettuce and stem lettuce, representing high-fidelity leafy-specific DMRs. The left panel shows examples for a high-fidelity leafy-specific hyper-DMR across *LsAMY1* (*Lsat_1_ v5_gn_6_112060*) and a high-fidelity leafy-specific hypo-DMR across *LsIQD31* (*Lsat_1_v5_gn_3_102080*). Shown above the methylation profiles are the gene structures of *LsAMY1* and *LsIQD31*, in which blue and yellow boxes indicate exons and untranslated regions, respectively, and purple lines indicate introns and other genomic regions. **e** Gene Ontology (GO) enrichment of genes associated with these high-fidelity leafy-specific DMRs

Cao *et al. Genome Biology*   (2024) 25:158

Page 7 of 22

significant portions of DMRs were specific for each lettuce type (Fig. 3b,c). It is noteworthy that stem and leafy (cos, butterhead, and crisp) types possessed 5093 and 7936 type-specific DMRs relative to wild lettuce, respectively (Fig. 3b,c), supporting the hypothesis that stem and leafy type lettuces are evolved independently after the initial domestication event [5].

Nevertheless, further analysis on these type-specific DMRs found in leafy lettuce revealed that stem lettuce also displayed consistent but insignificant changes in methylation levels on most of these DMRs relative to the wild lettuce (Fig. 3d). In contrast, 319 (4.7%) hyper-DMRs and 109 (10%) hypo-DMRs of these type-specific DMRs found in leafy lettuce exhibited significant methylation changes as compared to stem lettuce and were exclusively detected in leafy lettuce, which were called high-fidelity leafy-specific DMRs (Fig. 3d). For examples, *LsAMY1* (*Lsat_1_v5_gn_6_112060*), homolog to *AtAMY1* involved in starch mobilization and response to biotic and abiotic stress [35–37], showed high methylation levels across its gene region only in leafy lettuce (Fig. 3d). On the contrary, *LsIQD31* (*Lsat_1_v5_gn_3_102080*), homolog to *AtIQD31* involved in plant development [38], displayed low methylation levels particularly in leafy lettuce (Fig. 3d). These high-fidelity leafy-specific DMRs tended to occur in flanking sequences of genes and intergenic regions (Additional file 2: Fig. S4b). Genes associated with these DMRs were enriched in regulation processes including negative regulation of biological process, biological regulation, and regulation of cellular process (Fig. 3e; Additional file 1: Tables S5 and S6).

Independent evolution of stem lettuce

We then further analyzed the type-specific DMRs found in stem lettuce (Fig. 3b,c) and found that most of them showed similar methylation levels in leafy lettuce as compared to the wild lettuce *L. serriola* (Fig. 4a). Among them, 219 (13.0%) hyper-DMRs and 1217 (35.7%) hypo-DMRs exhibited significant methylation changes compared to leafy lettuce (Fig. 4a), which are referred hereafter as high-fidelity stem-specific DMRs. This result implies that evolution of stem lettuce, characterized by many distinct DMRs, is eccentric and independent of the leafy lettuce. Interestingly, the wild relative *L. saligna* was clustered into a subgroup of stem lettuce in the phylogenetic tree (Fig. 1b) despite their phenotypic divergence (Fig. 1a) and huge nucleotide diversity evident from the phylogenetic tree based on SNPs (Additional file 2: Fig. S1b). We thus further analyzed the methylation levels of high-fidelity stem-specific DMRs in *L. saligna*. We found that *L. saligna* showed low methylation levels on the hypo-DMRs compared to the average methylation level across the whole genome but no significant change on the hyper-DMRs (Fig. 4b). In contrast, two other wild relatives *L. virosa* and *L. indica* showed opposite changes with low methylation levels on the hyper-DMRs and high methylation levels on the hypo-DMRs (Additional file 2: Fig. S5a,b). These results imply the possible existence of a unique epigenetic evolution mechanism in stem lettuce in relation to *L. saligna*.

To uncover the origin of these high-fidelity stem-specific DMRs, we utilized global SNPs to identify windows where these DMRs exhibited the same genotype as stem lettuce in comparison with only one species of wild relatives but different from other wild relatives and leafy lettuce. In the global 9,998,549 windows for DMR identification, 26% of these windows exhibited a unique genotype that was shared by stem lettuce and *L.*
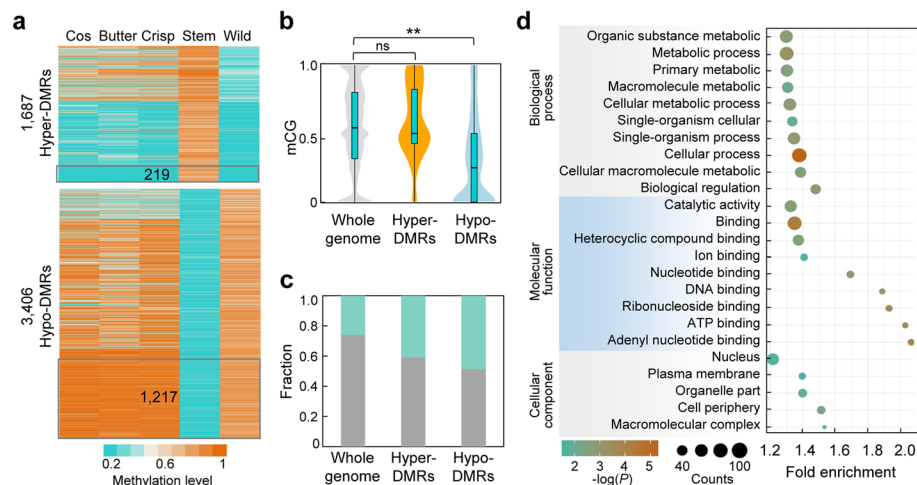
Fig. 4 Independently altered DNA methylation in stem lettuce. **a** Heatmap showing methylation levels of the stem lettuce specific hyper- and hypo-DMRs, relative to the wild lettuce *L. serriola*. DMRs in the grey blocks show significant differences between leafy type and stem lettuce, representing high-fidelity stem-specific DMRs. **b** Methylation levels of the high-fidelity stem-specific DMRs in the wild lettuce *L. saligna*. Asterisks and ns indicate significant difference (**$P < 0.01$, Wilcoxon signed-rank test) and no statistical difference ($P \geq 0.01$, Wilcoxon signed-rank test), respectively. **c** Enrichment of regions with the same SNPs in *L. saligna* and stem lettuce observed in high-fidelity stem-specific DMRs. Blue bars indicate fractions of regions with the same SNPs in stem lettuce and *L. saligna*, but different from other *Lactuca* species, including cos, butter, and crisp lettuce, *L. serriola*, *L. virosa*, and *L. indica*. **d** GO enrichment of genes associated with the high-fidelity stem-specific DMRs

saligna, which was more than 8-fold higher than other wild relatives (3.1% for *L. serriola*, 0.13% for *L. virosa*, and 0.08% for *L. indica*) (Additional file 2: Fig. S5c). Moreover, 41% and 48% of high-fidelity stem-specific hyper-DMRs and hypo-DMRs, respectively, overlapped with the identified windows where stem lettuce shares same genotypes as *L. saligna* (Fig. 4c). This result indicates that *L. saligna* could be the potential origin for the distinctive methylation variations in stem lettuce, thereby distinguishing it from leafy lettuce.

In addition, the high-fidelity stem-specific DMRs tended to occur in flanking sequences of genes as well as intergenic regions (Additional file 2: Fig. S5d). Genes associated with these DMRs were enriched into metabolic processes including organic substance metabolic, primary metabolic, and macromolecule metabolic (Fig. 4d; Additional file 1: Tables S7 and S8), implying a distinct selection on the metabolic processes during diversification of stem lettuce.

Domestication-induced DMRs are associated with stress response

In addition to type-specific DMRs, it was noteworthy that 2826 (7.61–37.1%) of hyper-DMRs and 1021 (9.88–21.9%) of hypo-DMRs were shared between leafy and stem lettuce (Figs. 3b, c and 5a). These shared domestication-induced DMRs accounted for 0.027% (0.698 Mb) of lettuce genomes which was significantly higher than that expected by chance ($P = 0$, hypergeometric test) and were called shared DMRs. These DMRs tended to occur in flanking sequences of genes and intergenic regions (Additional file 2: Fig. S6a). Genes associated with shared hyper-DMRs were enriched in cellular process, cell communication process, transport process, phosphate-containing compound
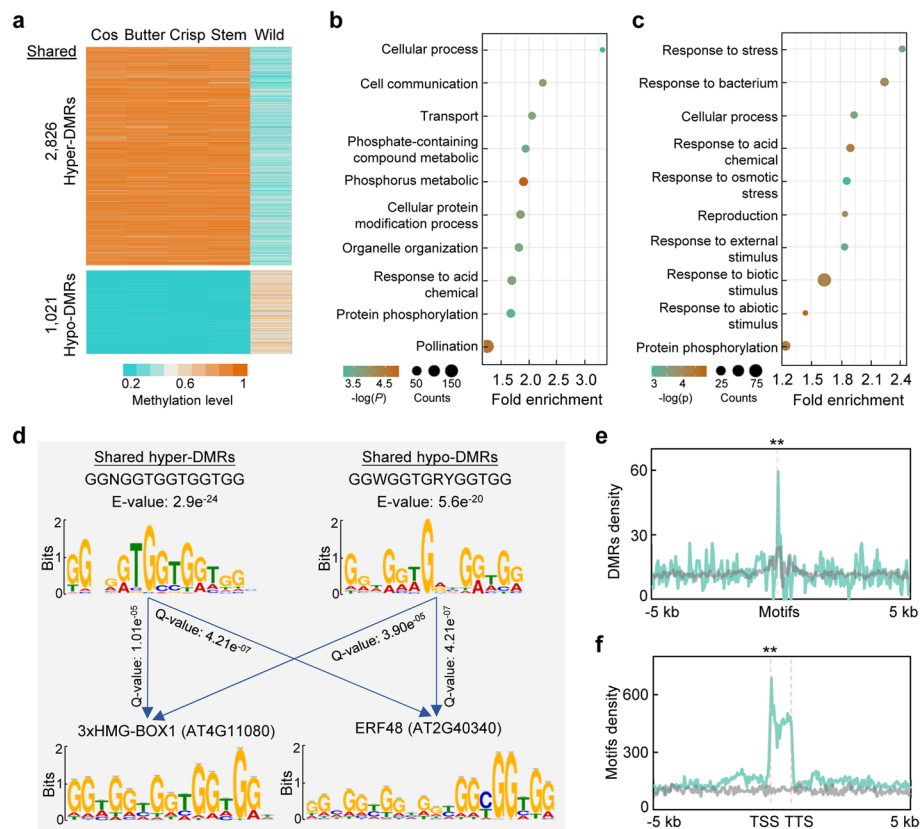
Fig. 5 Shared domestication-induced DMRs of lettuce cultivars. **a** Heatmap showing methylation levels of shared hyper- and hypo-DMRs of lettuce cultivars. **b**, **c** GO enrichment of genes associated with shared hyper- (**b**) and hypo-DMRs (**c**). The plots show 10 top-scoring biological processes. **d** DNA motifs enriched in the shared domestication-induced DMRs (upper panels) are similar to the motifs of binding sites of 3XHMG-BOX1 and ERF48 in published dataset (lower panels). *E*-value (*e*) indicates an estimate of the expected number of motifs by the MEME, while *Q*-value (*q*) indicates the probability that a random motif has an optimal alignment as the target motif. **e** Density of shared DMRs within 10 kb of ERF48 binding motif. The grey line indicates random genomic regions. Asterisk indicates significant difference (**$P < 0.01$, Wilcoxon signed-rank test). **f** Density of ERF48 binding motif within 10 kb of genes. The grey line indicates random genomic regions. Asterisk indicates significant difference (**$P < 0.01$, Wilcoxon signed-rank test)

metabolic process, response to acid chemical process, etc. (Fig. 5b; Additional file 1: Tables S9 and S10). In contrast, genes associated with domestication-induced shared hypo-DMRs were mainly enriched in biotic and abiotic stress response pathways including response to bacterium, response to acid chemical, and response to osmotic stress (Fig. 5c; Additional file 1: Tables S9 and S11), which is consistent with commonly speculated growth-defense trade-off during the process of crop domestication. In addition, we observed consistent changes of non-CG (CHG and CHH) methylation levels with CG methylation in these shared-DMRs between cultivated and wild lettuce (Additional file 2: Fig. S7), implying conserved changes of DNA methylation on different cytosine contexts during lettuce domestication. Notably, shared hypo-DMRs with very low CG methylation levels also contained low levels of CHG and CHH methylation in cultivated lettuce (Additional file 2: Fig. S7a).

As shared DMRs frequently occurred around genes (Additional file 2: Fig. S6a), we further performed MEME motif analysis [39] and found that shared hyper-DMRs possessed

Cao *et al. Genome Biology* (2024) 25:158

Page 10 of 22

a top-scoring motif of GGNGGTGGTGGTGG ($E = 2.9 \times 10^{-24}$), while shared hypo-DMRs had a top-scoring motif of GGWGGTGRYGGTGG ($E = 5.6 \times 10^{-20}$) (Fig. 5d). These two motifs resembled the binding sites of 3xHMG-BOX1 and AtERF48 [40]. 3xHMG-BOX1 belongs to a plant-specific family of DNA-binding proteins that interact with mitotic and meiotic chromosomes [41], while AtERF48, member of APETALA2/ ethylene response factor (ERF) transcription factor family, is responsible for modulating stress responses [42–44]. Remarkably, shared DMRs showed enrichment in the binding motif of ERF48 (Fig. 5e), and this motif tended to enrich in the gene body, particularly at the TSS (Fig. 5f). We then further identified the target genes of LsERF48 (Lsat_1_v5_gn_8_13141) through DNA affinity purification sequencing (DAP-seq) in lettuce (Additional file 1: Table S12). A GGWGGTGGTGG motif was found to be enriched in the binding sites of LsERF48 (Additional file 2: Fig. S8a), which resembled the motifs found in AtERF48 binding sites and shared DMRs (Fig. 5d). Consistent with the motif distribution observed in shared DMRs (Fig. 5e), LsERF48 binding sites were also more prevalent in the gene body (Additional file 2: Fig. S8b). Interestingly, LsERF48 target genes were enriched in stress response (Additional file 2: Fig. S8c), implying its possible role in modulating stress response in lettuce. Remarkably, 38 out of 671 genes associated with shared DMRs were also identified as LsERF48 targets (Additional file 1: Table S12). In addition, *35S:LsERF48-GFP*, but not the *35S:GFP* control, reduces reactive oxygen species (ROS) accumulation in response to salt stress in *N. benthamiana* (Additional file 2: Fig. S9), supporting the role of LsERF48 in modulating stress response. Together, these results imply domestication-induced DNA methylation variations in lettuce may play a potential role in regulating stress response.

Domestication-induced DNA methylation variations influence gene expression through *cis*-or trans-acting effects

As the majority of DNA methylation signatures are related to chromatin accessibility in plants [23, 45], we proceeded to examine whether domestication-induced DNA methylation variations affect chromatin accessibility in lettuce. To this end, we analyzed the ATAC-seq data obtained from stem lettuce [46] to investigate the chromatin accessibility patterns on the shared DMRs. We classified these shared DMRs based on their distance to genes: DMRs near genes that were located in 2 kb flanking sequences (1124/3847, 29%) and DMRs far from genes (2723/3847, 71%). Interestingly, we found that chromatin accessibility was significantly increased at shared hypo-DMRs near genes, but insignificantly decreased at shared hyper-DMRs near genes (Fig. 6a). Moreover, shared hypo-DMRs far away from genes showed higher levels of chromatin accessibility, whereas shared hyper-DMRs far away from genes showed lower levels of chromatin accessibility (Fig. 6a). To further explore whether DNA methylation changes influences chromatin accessibility during lettuce domestication, we assessed chromatin accessibility in the wild lettuce *L. serriola* through ATAC-seq. Notably, we observed high levels of chromatin accessibility in wild lettuce in shared hyper-DMRs near genes and far away from genes (Additional file 2: Fig. S10), where wild lettuce possessed low DNA methylation levels (Fig. 5a). This observation, together with the relatively low chromatin accessibility in shared hyper-DMRs in stem lettuce that possessed high DNA methylations in these regions (Fig. 6a), indicate that domestication is associated with an increase in DNA

Cao *et al. Genome Biology* (2024) 25:158

Page 11 of 22
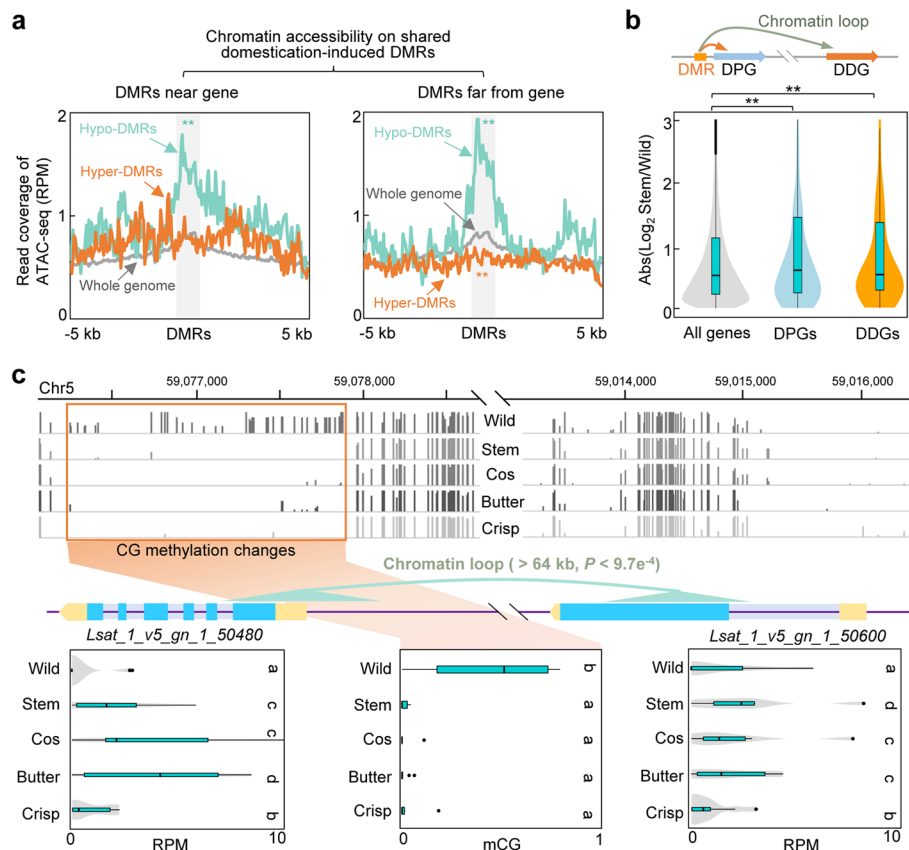
Fig. 6 Domestication-induced DNA methylation changes contribute to gene expression changes through *cis-* or *trans*-acting effects. **a** Metaplots showing chromatin accessibility in the shared domestication-induced DMRs of lettuce cultivars, located near the genes (left panel) and far from genes (right panel). Asterisks indicate significant differences (**$P < 0.01$, Wilcoxon signed-rank test) of the chromatin accessibility in DMRs compared with the whole genome. **b** Increased gene expression changes (absolute values) between stem lettuce and the wild lettuce *L. serriola* in DMR-associated proximal genes (DPGs) and distal genes (DDGs) compared with all genes. Asterisks indicate significant differences (**$P < 0.01$, Wilcoxon signed-rank test). **c** An example showing methylation changes (Chr1: 59,076,400–59,077,800; upper panel and middle box plot in the lower panel) related to the proximal gene (DPG: *Lsat_1_v5_gn_1_50480*; right box plot in the lower panel) and the distal gene (DDG: *Lsat_1_v5_gn_1_50600*; left box plot in the lower panel) that are in one chromatin loop with more than 64 kb physical distance. The middle panel shows the gene structures of these genes, in which blue and yellow boxes indicate exons and untranslated regions, respectively, and purple lines indicate introns and other genomic regions. Different letters on the boxes indicate significant differences ($P < 0.01$, Wilcoxon signed-rank test) in a pairwise comparison

methylation alongside a decrease in chromatin accessibility in stem lettuce compared to wild lettuce.

To further examine whether DNA methylation changes influence gene expression, we identified 673 DMR-associated proximal genes (DPGs) that had shared DMRs located within 2 kb flanking sequences and detected 149 DMR-associated distal genes (DDGs) that interact with shared DMRs via long-range chromatin loops through reanalyzing the published Hi-C datasets of stem lettuce (Fig. 6b; Additional file 1: Table S13) [46, 47]. We also conducted RNA-seq analysis of all our accessions of wild and cultivated lettuce (Additional file 1: Table S2). DNA methylation changes significantly induced expression changes of both DPGs and DDGs in stem lettuce (Fig. 6b; Additional file 2: Fig. S11). For example, a shared hypo-DMR in Chr 1 (59,076,400–59,077,800) was associated

with increased expression of its neighboring gene *Lsat_1_v5_gn_1_50480*, a homolog of *AT2G36540* with unknown function, as well as a distal gene *Lsat_1_v5_gn_1_50600*, a homolog of *AT5G04550* related to abiotic stress responses [48], that is located at ~ 64 kb downstream of this DMR (*P* value < 9.7×10^{-4} for the interaction loop) (Fig. 6c). The expression changes of *Lsat_1_v5_gn_1_50480* and *Lsat_1_v5_gn_1_50600* in cultivated lettuce compared with wild lettuce were further verified by qPCR (Additional file 2: Fig. S12). Inhibition of DNA methylation by 5-azacytidine (5-AzaC) in the wild lettuce (*L. serriola*) seedlings resulted in a reduction of DNA methylation levels in Chr1: 59,076,400–59,077,800 but elevated expression levels of *Lsat_1_v5_gn_1_50480* and *Lsat_1_v5_gn_1_50600* (Additional file 2: Fig. S13). Similar to stem lettuce, DNA methylation variations were also associated with significant expression changes of DPGs and DDGs in cos lettuce (Additional file 2: Fig. S6b). In addition, we also found that DNA methylation changes were associated with increased expression of DPGs in butterhead and crisp lettuce (Additional file 2: Fig. S6b-d). These results indicate that variations in DNA methylation may affect chromatin accessibility and hence alter expression levels of the associated proximal and distal genes.

Discussion

As an essential and pervasive epigenetic mark in most eukaryotes, DNA methylation controls multiple biological and physiological processes [13, 49]. Several recent population epigenetic studies suggest roles of epigenetics evolution in the long-term domestication processes of several crops including maize [30, 31], soybean [32], cotton [50], and rice [33]. In this study, we have defined the DNA methylation landscape in a natural population of 52 *Lactuca* accessions including major lettuce cultivars and wild relatives and uncovered extensive DNA methylation variation during lettuce domestication and divergence.

  Lettuce belongs to the core group of crops that are cultivated worldwide and represents one of the earliest domesticated vegetable crops (~ 6000 years ago) [2, 6]. Lettuce cultivars are believed to be derived from *L. serriola* in a single domestication event [5]. It has been estimated that stem and leafy type lettuce (butterhead, crisp and cos) were diverged from the ancestral cultivated lettuce at ~1922 and ~500 years ago [5]. Our phylogenetic analysis based on DNA methylation of the *Lactuca* genus further supports the ancestral role of *L. serriola* for lettuce cultivars but also indicates an interesting relationship between the wild relative *L. saligna* and stem lettuce. Transfer of genetic material between crops and their wild relatives through spontaneous hybridization and subsequent backcrossing, known as introgression, widely occurred and has been suggested to promote crop domestication and diversification [51–54]. Our analysis has revealed that 41–48% of the high-fidelity stem-specific DMRs are related to the introgression of *L. saligna*, greatly higher compared to other wild relatives *L. serriola*, *L. virosa*, and *L. indica* (Fig. 4c; Additional file: Fig. S5c). Among the noticeably increased fraction of *L. saligna*-related DMRs, the hypo-methylations of *L. saligna* could potentially exhibit transgenerational inherence during stem lettuce evolution and improvement, while most of hyper-methylations could be unstable (Fig. 4b). We envisage that these unstable hyper-methylations might undergo functional demethylation and/or be influenced by the dilution of factors for maintaining methylation, as previously reported [34].

These methylation variations are likely a response to artificial domestication. Our result implies that although wild relatives are commonly used to introduce advanced traits such as high production, rich nutrition, and resistances to abiotic and/or biotic stimulus [6, 55], epigenetics may play an important role in determining the fate of functionalization during genetic introgression.

It has been hypothesized that stem and leafy (butterhead, crisp, and cos) types of lettuce are evolved independently after the initial domestication in the domestication center [5]. We have uncovered abundant DNA methylation variations that are specifically associated with leafy and stem types, in addition to those commonly associated with both types. Interestingly, these variations are found to be enriched in different GO categories, likely contributing to morphological domestication and diversification of lettuce. Moreover, these variations, together with the dataset showing expression levels of the genes related to the stem lettuce-specific DMRs in different types of cultivated lettuce and the wild lettuce *L. serriola* (Additional file 1: Table S14), serve as valuable resources for further functional genomic research on lettuce divergence.

Although abundant genetic variations during lettuce domestication have been profiled by a large-scale whole genome resequencing study [6], there are few examples of genes that have been shown to be directly linked to lettuce domestication. Characterized lettuce genes that are possibly involved in lettuce domestication include lettuce *APETALA2* regulating seed shape [56], *LsKN1* controlling leafy head development [57], and *LsNRL4* regulating photosynthesis and leaf angles [58]. Unfortunately, these genes do not contain domestication-induced DMRs. Nevertheless, we have identified hundreds of genes that are associated with these domestication-induced DMRs, which will serve as valuable resources to facilitate future functional genomic studies in investigating domestication genes in lettuce. Furthermore, it has been suggested that the domestication process is driven significantly by the adaptation of stress tolerance to human-provided environments and conditions, leading to considerable changes in morphology, behavior, and physiology, and notably, reduced stress resistance is a common trait observed in many crop domestications [59, 60]. In this study, the enrichment of stress response genes has been found in the genes associated with the shared hypo-DMRs (Fig. 5c), and interestingly, the shared DMRs show enrichment in the binding motif of ERF48 that is possibly involved in modulating stress response (Fig. 5e; Additional file 2: Fig. S9). Stress conditions has been shown to induce DNA methylation in plants [61–63], and cultivated crops grown in human-manipulated environments are often accompanied by tradeoffs in stress resistance to promote growth and yield [64, 65]. It is possible that DNA methylation related to stress response may passively decrease during lettuce domestication. Differential responses to pathogens have been observed in lettuce cultivars and wild relatives [9]; it would be intriguing to further explore the involvement of the identified epialleles in responses to various stress conditions in lettuce.

Long-range interactions between genomic elements are critical for many cellular processes through regulation of gene expression [47, 66, 67]. For example, in maize, a distal enhancer *KERNEL ROW NUMBER4* (*KRN4*) located at more than 30 kb downstream from the *UNBRANCHED3* (*UB3*) gene could spatially interact with the *UB3* promoter to affect *UB3* expression and hence regulate inflorescence development [68]. Likewise, our analysis has revealed that DNA methylation variations influence not only

gene expression changes in their proximal genes but also their distal genes involved in long-range chromatin loops (Fig. 6b) and that expression of shared hypo-DMR-associated genes is significantly increased, but not vice versa for hyper-DMR-associated genes (Additional file 2: Fig. S14). However, we have observed some discrepancies among different lettuce cultivars. DNA methylation variations influence expression levels of both proximal and distal genes in stem and cos lettuce but mainly affect proximal genes in butterhead and crisp lettuce (Additional file 2: Fig. S6c,d). This could be due to the differences in the chromatin loop profiles in different lettuce cultivars. It has been suggested that chromatin structures are dynamically altered during soybean polyploidization, diploidization, and domestication [69]. Vast chromatin loops between genes and genes and those between gene and distal regulator elements are generated or eliminated during soybean domestication [69]. Thus, further mapping the chromatin interaction profiles in the *Lactuca* genus will be necessary to better understand roles of epigenetic evolution in influencing gene expression and morphological changes during lettuce domestication.

Conclusions

In summary, our study reveals epigenome variations among major cultivated lettuce and wild relatives, providing important insights into epigenetic evolution in gene regulation during lettuce domestication and divergence. These identified epialleles serve as valuable resources for facilitating future functional genomic research into lettuce domestication and divergence as well as for epigenetic breeding to improve lettuce production and stress resistance for sustainable agriculture.

Methods

Plant materials and growth conditions

Lettuce seeds were obtained from the Center for Genetic Resources, the Netherlands (http://www.wageningenur.nl/) (Additional file 1: Table S1). Seeds were surface sterilized with 10% sodium hypochlorite and grown on soil in a growth chamber with 16 h light/8 h dark at 24℃ (day)/22 °C (night). The third pair of leaves at 30 days after planting (30 DAP) was collected at ZT4 (Zeitgeber time) for isolating genomic DNA and total RNA for library construction.

Library construction for MethylC-seq and DNA-seq

Genomic DNA was isolated from the leaf samples using the cetyltrimethylammonium bromide (CTAB) method [70]. After removing RNA with RNase A (NEB), genomic DNA (about 3 µg) was fragmented into 300–500 bp long, end-repaired, and 3′-end adenylated followed by ligation of the methylated adapter (AITbiotech) according to the protocol of NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB). For MethylC-seq, around 1 µg of adapter-ligated DNA fragments was treated with bisulfite using the Zymo EZ DNA Methylation-Gold™ kit (Zymo Research), followed by a 10-cycle PCR amplification with Q5U® HiFi Hot Start DNA Polymerase (NEB). For DNA-seq libraries, the rest (around 1 µg) of the above adapter-ligated DNA fragments was amplified by a 6-cycle PCR amplification with Q5® HiFi Hot Start DNA Polymerase (NEB). After purification with VAHTSTM DNA Clean Beads (Vazyme), both MethylC-seq and DNA-seq

libraries were sequenced on a NovaSeq platform (Illumina), generating 150 bp paired-end reads.

SNP calling

After filtering the raw reads of DNA-seq libraries with fastp [71], clean reads were mapped to the lettuce reference genome (Salinas_v8, https://phytozome-next.jgi.doe.gov/info/Lsativa_V8) [72] using BWA (v0.7.15) with default parameters [73]. Uniquely mapped reads were extracted, and potential PCR duplicates were removed using Picard-tools (version 2.0.1). The remaining reads were used for variant detection using Genome Analysis Toolkit (GATK, version 3.5.0) [74]. SNPs and Indels were separated using the GATK "SelectVariants" function. To reduce the variant discovery rate, the SNP calls were filtered according to the following threshold: QD (quality by depth) < 10.0 || MQ (mapping quality) < 20.0 || FS (fisher strand) > 30.0 || SOR (symmetric odds ratio) > 3.0 || MQRankSum (rank sum test for mapping quality) < -2.5 || ReadPosRankSum (rank sum test for read position bias) < -3.5 || DP (read depth) <5 || DP > 100.0. The remaining SNPs were used for further analyses.

Single cytosine methylation calling and population parameter (Pi, π)

After filtering the raw reads of MethylC-seq libraries with fastp [71], the clean reads of each accession were mapped to the corresponding pseudo reference genome [30, 34], in which the lettuce reference genome (Salinas_v8) sequences were replaced by the corresponding nucleotides of each accession at the SNPs loci [75, 76], using Bismark (v0.15.0) with options (--score_min L,0,-0.2 -X 1000). Bisulfite conversion of all libraries was evaluated using the chloroplast genome as a control. The average conversion rate of 98.5% indicates a high level of reproducibility. To reduce clonal bias, the reads mapped to the same sites were collapsed into a single consensus read for calling methylation level on each cytosine site covered by at least three uniquely mapped reads.

To distinguish unmethylated cytosines in the CG contexts, we analyzed methylation level using the mixtools [77] under "normalmixEM" and set the thresholds as follows: unmethylation (≤ 0.1), heterozygous methylation (> 0.1 and < 0.9), methylation (≥ 0.9). To calculate the diversity of DNA methylation in lettuce population, we used methylation haplotype (meplotype) by converting the methylation status of each cytosine to nucleobase based on the threshold of the methylation level: methylation level ≤ 0.1 marked as "TT" representing unmethylated cytosine, methylation level within 0.1–0.9 marked as "CT" representing partially methylated cytosine, and methylation level ≥ 0.9 marked as "CC" representing methylated cytosine [78]. The meplotype and SNPs after removing variants with > 40% missing calls and minor allele frequency (MAF) < 0.05 were used to calculate population parameter (π) using VCFtools (0.1.13) [79] in 100-kb windows with 10-kb steps across the genome.

Identification of differentially methylated regions (DMRs)

To identify DMRs among the lettuce population including the wild lettuce *L. serriola* population and each horticultural types of cultivated lettuce population, accessions from each population were considered as repeats and cytosine (CG) sites detected in at least 3 accessions from each population which were considered as effective cytosine sites. The

Cao *et al. Genome Biology*   (2024) 25:158

Page 16 of 22

genomic windows containing eight consecutive effective cytosine sites (< 500 bp) were used to identify DMR candidate regions using DMRfinder (v0.3) [80], in conjunction with R language (V4.3.1) for statistical tests and adjusting *P*-values. Within these candidate regions, CG DMRs between two populations were determined by applying cut-off values for average methylation level differences (> 0.4) and a corrected false discovery rate (FDR < 0.05). The FDR was calculated by adjusting the P-values (obtained from ANOVA tests) of pairwise comparisons using the Benjamini-Hochberg method.

Analysis of DMRs using qPCR

The DMRs were validated using qPCR with the digested methylated genomic DNA as previously reported [81]. Briefly, genomic DNA (1 μg) was digested for 16 h using MspJI (NEB), with mock treatments performed by substituting glycerol for MspJI. qPCR was performed using Taq Pro Universal SYBR Green Master Mix (Vazyme) on the CFX Opus 384 Real-Time PCR System (Bio-Rad). Relative DNA methylation levels were assessed by calculating the difference between mock Ct and digested Ct values, and a larger differential Ct value represents a higher methylation level. The primers for the randomly selected 36 DMRs are listed in Additional file 1: Table S15.

Library construction and analysis of DAP-seq

DAP-seq libraries were constructed as previously reported [82]. To generated adapter-ligated DNA fragments, the genomic DNA (~3 μg) was fragmented into 300–500 bp long, end-repaired, and 3′-end adenylated followed by ligation of the adapter (AITbiotech) according to the protocol of NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB). The full-length coding sequence of *LsERF48* was amplified with P1F and P1R (Additional file 1: Table S14) and cloned into the pCAMBIA1300-GFP vector. The resulting pCAMBIA1300-*LsERF48*-GFP plasmid was added to the TNT® SP6 Coupled Wheat Germ Extract System (Promega) to synthesize the LsERF48-GFP protein. The LsERF48-GFP protein was then immobilized on the GFP-Trap® Agarose Beads (ChromoTek) and subsequently incubated with 100 ng of adapter-ligated DNA fragments. After extensive washing, the eluted DNA was amplified through 20 cycles of PCR using Q5 HiFi HotStart DNA Polymerase (NEB). Following purification, the DAP-seq libraries were sequenced on the NovaSeq platform (Illumina), generating 150 bp paired-end reads.

After filtering the raw reads of DAP-seq with fastp [71], clean reads were mapped to the lettuce reference genome (Salinas_v8) using BWA (v0.7.15) [73]. Uniquely mapped reads were extracted, followed by removing potential PCR duplicates using Picard-tools (version 2.0.1). The remaining paired reads were then used for peak calling using MACS2 (2.2.6). Total peaks were used to perform DNA motifs analysis using MEME (5.1.0).

ATAC-seq library construction

Construction of ATAC-seq libraries was performed using the ATAC-seq kits (Active-motif) according to the manufacturer's instructions. DNA was recovered by MinElute PCR Purification Kit (Qiagen) and amplified through 10 cycles of PCR. Libraries were sequenced on the NovaSeq platform (Illumina) to generate 150 bp paired-end reads.

Cao *et al. Genome Biology*   (2024) 25:158

Page 17 of 22

Analysis of ATAC-seq and Hi-C seq datasets

The ATAC-seq data and published ATAC-seq datasets downloaded from https://www.lettucegdb.com/data/ATAC-Seq/ (Additional file 1: Table S2) were filtered with fastp [71]. Clean reads were mapped to the lettuce reference genome (Salinas_v8) using BWA (v0.7.15) [73]. Uniquely mapped reads were extracted, followed by removing potential PCR duplicates using Picard-tools (version 2.0.1). Using the remaining paired reads, read counts in 50-bp bins across the whole genome were normalized as reads per million mapped reads (RPM). To reveal changes in chromatin accessibility between DMRs and random genomic regions, the genome was divided into windows to match the average length of DMRs, and subsequently the same numbers of windows were randomly selected for comparison with DMRs.

Hi-C data were downloaded from BIG Data Center (https://bigd.big.ac.cn/gsa/index.jsp) under the accession no. PRJCA007442 (Additional file 1: Table S2) [46]. After filtering with fastp [71], clean reads were mapped to the lettcue reference genome (Salinas_v8) using HiC-Pro (2.11.1) [83]. The aligned reads of both fragment mates were then paired in a single paired-end BAM file generated by HiC-Pro (2.11.1) [83]. Subsequently, dangling-end reads, same-fragment reads, self-circled reads, self-ligation reads, and other invalid Hi-C reads were discarded, and potential PCR duplicates were removed. Retained valid paired-end reads were used to generate raw Hi-C matrix by HiC-Pro (2.11.1) [83]. These matrices were normalized by the iterative correction and eigenvector decomposition (ICE) method of HiC-Pro (2.11.1) [83]. The normalized Hi-C matrices were used to identify intra-chromosomal interaction loops (q-value <0.05) by the "FitHiC" package (2.0.7) with default parameters [84].

RNA-seq library construction and analysis

Total RNA was extracted from the same batch of plant tissues used for MethylC-seq and DNA-seq analysis. RNA-seq libraries were constructed as previously reported [82]. After DNase I treatment, aliquots of total RNA were individually fragmented. First-strand cDNA synthesis was performed using oligo(dT)-index primer, followed by second strand cDNA synthesis, end-repair, and adapter ligation. Subsequently, cDNA was amplified by 15 cycles of PCR reaction using Q5 HiFi HotStart DNA Polymerase (NEB). After purification, the 3′ RNA-seq libraries were sequenced on NovaSeq platform (Illumina) to generate 150 bp paired-end reads.

After filtering with fastp (v0.23.2) [71], clean reads were mapped to the lettuce reference genome (Salinas_v8) using Hisat2 (v2.2.1) [85]. After removing potential PCR duplicates using Picard-tools (version 2.0.1), uniquely mapped reads on each gene were normalized as reads per million mapped reads (RPM) to represent the expression level for each gene using featureCounts (v2.0.4) [86] with default parameters.

5-Azacytidine (5-AzaC) treatment

Seeds of the wild lettuce (*L. serriola*) were placed on sterile filter papers soaked in water with or without 5-AzaC (20 mg/L) and incubated in a climate-controlled chamber with a 16 h light/8 h dark cycle at 24 °C (day)/22°C (night). After 3 days after sowing (DAS), the old filter papers were replaced with new ones that soaked in water with or without

5-AzaC (20 mg/L). On 6 DAS, these wild lettuce seeds were transferred to soil and grown in a growth chamber under 16 h light/8 h dark at 24°C (day)/22 °C (night). The third pair of leaves at 30 DAS was collected at ZT4 (Zeitgeber time) for isolating genomic DNA and RNA. The mock- and 5-AzaC-treated genomic DNAs were used for constructing MethylC-seq libraries as described above, followed by sequencing on the NovaSeq platform (Illumina) to generate 150 bp paired-end reads. MethylC-seq data were analyzed as described above, and the CG methylation levels of Chr1: 59,076,400–59,077,800 were shown in Additional file 2: Fig. S13a.

Functional validation of LsERF48

The full-length coding sequence of *LsERF48* was amplified with P1F and P1R (Additional file 1: Table S15) and cloned into the pCAMBIA1300-35S:GFP vector to generate *35S:LsERF48-GFP*. *Agrobacterium* culture ($OD_{600} = 0.9$) harboring *35S:LsERF48-GFP* or the empty *35S:GFP* vector were infiltrated into leaves of *N. benthamiana* at 30 days after planting. After 2 days post-inoculation (dpi), *N. benthamiana* with inoculated leaves were irrigated with water containing 150 mM NaCl (salt treatment) or 0 mM NaCl (mock treatment). After 4 dpi, ROS accumulation was analyzed using NBT (Nitro tetrazolium blue chloride) and DAB (diaminobenzidine) staining [87]. For DAB staining, salt- and mock-treated infiltrated *N. benthamiana* leaves were stained with DAB-HCl (1 mg/mL) solution overnight in darkness. Subsequently, the stained leaves were washed five times with sterile water, followed by fixation in 70% (v/v) ethanol for 10 h at room temperature. Decolorization was then performed in 96% ethanol at 40 °C to remove chlorophyll. For NBT staining, detached leaves were immersed in 100 mL of staining solution containing 0.05% (w/v) NBT, 10 mM sodium azide, 50 mM potassium phosphate, pH 6.4 for 6 h. After stopping the reaction with 95% ethanol, the samples were decolorized in 96% ethanol at 40 °C. The photos were taken under a light microscope. The IMAGEJ software (http://rsbweb.nih.gov/ij) was used to measure the relative intensities of DAB staining and NBT staining.

qPCR analysis

Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) and reverse transcribed with the HiScript III RT SuperMix (Vazyme) according to the manufacturer's instructions. qPCR was performed using Taq Pro Universal SYBR Green Master Mix (Vazyme) on the CFX Opus 384 Real-Time PCR System (Bio-Rad). Relative gene expression levels were determined as previously described [88], with *LsActin* serving as the internal control. Primers used for qPCR analysis are listed in Additional file 1: Table S15.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03310-x.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3. Review history.

Declarations

Ethic approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Shatilov MV, Razin AF, Ivanova MI. Analysis of the world lettuce market. IOP Conference Series. 2019;395:012053.
2. Lebeda A, Ryder EJ, Grube R, Dolezˇalova´ I, Krˇ´ıstkova E. Lettuce (Asteraceae; Lactuca spp.). Genetic Resources, Chromosome Engineering, and Crop Improvement Vol 3 (ed Singh, R J). 2006:377-472.
3. de Vries IM. Origin and domestication of Lactuca sativa L. Genet Resour Crop Ev. 1997;44:165–74.
4. Lindqvist K. On the origin of cultivated lettuce. Hereditas. 1960;46:319–50.
5. Zhang L, Su W, Tao R, Zhang W, Chen J, Wu P, Yan C, 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.
6. Wei T, van Treuren R, Liu X, Zhang Z, Chen J, Liu Y, Dong S, Sun P, Yang T, Lan T, et al. Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. Nat Genet. 2021;53:752–60.
7. Park S, Kumar P, Shi A, Mou B. Population genetics and genome-wide association studies provide insights into the influence of selective breeding on genetic variation in lettuce. Plant Genome. 2021;14:e20086.
8. Fertet A, Graindorge S, Koechler S, de Boer GJ, Guilloteau-Fonteny E, Gualberto JM. Sequence of the mitochondrial genome of Lactuca virosa suggests an unexpected role in Lactuca sativa's evolution. Front Plant Sci. 2021;12:697136.
9. Lebeda A, Křístková E, Kitner M, Mieslerová B, Jemelková M, Pink DAC. Wild Lactuca species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. Eur J Plant Pathol. 2014;138:597–640.
10. Lebeda A, Křístková E, Kitner M, Majeský Ľ, Doležalová I, Khoury CK, Widrlechner MP, Hu J, Carver D, Achicanoy HA, Sosa CC. Research gaps and challenges in the conservation and use of North American wild Lettuce germplasm. Crop Sci. 2019;59:2337–56.
11. Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. Science. 2010;330:622–7.
12. Song Q, Chen ZJ. Epigenetic and developmental regulation in plant polyploids. Curr Opin Plant Biol. 2015;24:101–9.
13. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol. 2018;19:489–506.
14. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics. 2003;163:1109–22.
15. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE. Requirement of CHROMO-METHYLASE3 for maintenance of CpXpG methylation. Science. 2001;292:2077–80.
16. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205.
17. An YC, Goettel W, Han Q, Bartels A, Liu Z, Xiao W. Dynamic changes of genome-wide DNA methylation during soybean seed development. Sci Rep. 2017;7:12263.
18. Lin JY, Le BH, Chen M, Henry KF, Hur J, Hsieh TF, Chen PY, Pelletier JM, Pellegrini M, Fischer RL, et al. Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proc Natl Acad Sci USA. 2017;114:E9730–9.

19. Narsai R, Gouil Q, Secco D, Srivastava A, Karpievitch YV, Liew LC, Lister R, Lewsey MG, Whelan J. Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. Genome Biol. 2017;18:172.

20. Kawakatsu T, Nery JR, Castanon R, Ecker JR. Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol. 2017;18:171.

21. Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, Colot V. DNA methylation dynamics during early plant life. Genome Biol. 2017;18:179.

22. Song Q, Guan X, Chen ZJ. Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid cotton. PLoS Genet. 2015;11:e1005724.

23. Crisp PA, Marand AP, Noshay JM, Zhou P, Lu Z, Schmitz RJ, Springer NM. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. Proc Natl Acad Sci U S A. 2020;117:23991–4000.

24. Richards EJ. Inherited epigenetic variation–revisiting soft inheritance. Nat Rev Genet. 2006;7:395–401.

25. Hauser MT, Aufsatz W, Jonak C, Luschnig C. Transgenerational epigenetic inheritance in plants. Biochim Biophys Acta. 2011;1809:459–68.

26. Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P, et al. Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet. 2009;5:e1000530.

27. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. Nature. 1999;401:157–61.

28. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet. 2006;38:948–52.

29. Quadrana L, Almeida J, Asis R, Duffy T, Dominguez PG, Bermudez L, Conti G, Correa da Silva JV, Peralta IE, Colot V, et al. Natural occurring epialleles determine vitamin E accumulation in tomato fruits. Nat Commun. 2014;5:3027.

30. Xu G, Lyu J, Li Q, Liu H, Wang D, Zhang M, Springer NM, Ross-Ibarra J, Yang J. Evolutionary and functional genomics of DNA methylation in maize domestication and improvement. Nat Commun. 2020;11:5539.

31. Xu J, Chen G, Hermanson PJ, Xu Q, Sun C, Chen W, Kan Q, Li M, Crisp PA, Yan J, et al. Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. Genome Biol. 2019;20:243.

32. Shen Y, Zhang J, Liu Y, Liu S, Liu Z, Duan Z, Wang Z, Zhu B, Guo YL, Tian Z. DNA methylation footprints during soybean domestication and improvement. Genome Biol. 2018;19:128.

33. Cao S, Chen K, Lu K, Chen S, Zhang X, Shen C, Zhu S, Niu Y, Fan L, Chen ZJ, et al. Asymmetric variation in DNA methylation during domestication and de-domestication of rice. Plant Cell. 2023;35:3429–43.

34. Cao S, Wang L, Han T, Ye W, Liu Y, Sun Y, Moose SP, Song Q, Chen ZJ. Small RNAs mediate transgenerational inheritance of genome-wide trans-acting epialleles in maize. Genome Biol. 2022;23:53.

35. Lloyd JR, Kossmann J, Ritte G. Leaf starch degradation comes out of the shadows. Trends Plant Sci. 2005;10:130–7.

36. Kamranfar I, Xue GP, Tohge T, Sedaghatmehr M, Fernie AR, Balazadeh S, Mueller-Roeber B. Transcription factor RD26 is a key regulator of metabolic reprogramming during dark-induced senescence. New Phytol. 2018;218:1543–57.

37. Doyle EA, Lane AM, Sides JM, Mudgett MB, Monroe JD. An alpha-amylase (At4g25000) in Arabidopsis leaves is secreted and induced by biotic and abiotic stress. Plant Cell Environ. 2007;30:388–98.

38. Burstenbinder K, Mitra D, Quegwer J. Functions of IQD proteins as hubs in cellular calcium and auxin signaling: a toolbox for shape formation and tissue-specification in plants? Plant Signal Behav. 2017;12:e1331198.

39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202-208.

40. O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell. 2016;166:1598.

41. Pedersen DS, Coppens F, Ma L, Antosch M, Marktl B, Merkle T, Beemster GT, Houben A, Grasser KD. The plant-specific family of DNA-binding proteins containing three HMG-box domains interacts with mitotic and meiotic chromosomes. New Phytol. 2011;192:577–89.

42. Muller M, Munne-Bosch S. Ethylene response factors: a key regulatory hub in hormone and stress signaling. Plant Physiol. 2015;169:32–41.

43. Chen H, Hwang JE, Lim CJ, Kim DY, Lee SY, Lim CO. Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. Biochem Biophys Res Commun. 2010;401:238–44.

44. Lee SJ, Kang JY, Park HJ, Kim MD, Bae MS, Choi HI, Kim SY. DREB2C interacts with ABF2, a bZIP protein regulating abscisic acid-responsive gene expression, and its overexpression affects abscisic acid sensitivity. Plant Physiol. 2010;153:716–27.

45. Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat Plants. 2019;5:1250–9.

46. Guo Z, Li B, Du J, Shen F, Zhao Y, Deng Y, Kuang Z, Tao Y, Wan M, Lu X, et al. LettuceGDB: the community database for lettuce genetics and omics. Plant Commun. 2023;4:100425.

47. Liu C, Cheng YJ, Wang JW, Weigel D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat Plants. 2017;3:742–8.

48. Luhua S, Hegie A, Suzuki N, Shulaev E, Luo X, Cenariu D, Ma V, Kao S, Lim J, Gunay MB, et al. Linking genes of unknown function with abiotic stress responses by high-throughput phenotype screening. Physiol Plant. 2013;148:322–33.

49. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41.

50. Song Q, Zhang T, Stelly DM, Chen ZJ. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol. 2017;18:99.

51. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20:229–37.

52. Wang X, Chen L, Ma J. Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. Genome Biol. 2019;20:22.

53. Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. Genome Res. 2017;27:1029–38.
54. Hufford MB, Lubinksy P, Pyhajarvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J. The genomic signature of crop-wild introgression in maize. PLoS Genet. 2013;9:e1003477.
55. Xiong W, Berke L, Michelmore R, van Workum DM, Becker FFM, Schijlen E, Bakker LV, Peters S, van Treuren R, Jeuken M, et al. The genome of Lactuca saligna, a wild relative of lettuce, provides insight into non-host resistance to the downy mildew Bremia lactucae. Plant J. 2023;115:108–26.
56. Luo C, Wang S, Ning K, Chen Z, Wang Y, Yang J, Qi M, Wang Q. The APETALA2 transcription factor LsAP2 regulates seed shape in lettuce. J Exp Bot. 2021;72:2463–76.
57. Yu C, Yan C, Liu Y, Liu Y, Jia Y, Lavelle D, An G, Zhang W, Zhang L, Han R, et al. Upregulation of a KN1 homolog by transposon insertion promotes leafy head development in lettuce. Proc Natl Acad Sci USA. 2020;117:33668–78.
58. An G, Qi Y, Zhang W, Gao H, Qian J, Larkin RM, Chen J, Kuang H. LsNRL4 enhances photosynthesis and decreases leaf angles in lettuce. Plant Biotechnol J. 2022;20:1956–67.
59. Chen YH, Gols R, Benrey B. Crop domestication and its impact on naturally selected trophic interactions. Annu Rev Entomol. 2015;60:35–58.
60. Razzaq A, Saleem F, Wani SH, Abdelmohsen SAM, Alyousef HA, Abdelbacki AMM, Alkallas FH, Tamam N, Elansary HO. De-novo domestication for improving salt tolerance in crops. Front Plant Sci. 2021;12:681367.
61. Secco D, Wang C, Shou HX, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife. 2015;4:e09343.
62. Wang L, Cao S, Wang P, Lu K, Song Q, Zhao FJ, Chen ZJ. DNA hypomethylation in tetraploid rice potentiates stress-responsive gene expression for salt tolerance. Proc Natl Acad Sci U S A. 2021;118:e2023981118.
63. Zheng XG, Chen L, Xia H, Wei HB, Lou QJ, Li MS, Li TM, Luo LJ. Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition. Sci Rep. 2017;7:39843.
64. Wang Z, Hong Y, Zhu G, Li Y, Niu Q, Yao J, Hua K, Bai J, Zhu Y, Shi H, et al. Loss of salt tolerance during tomato domestication conferred by variation in a Na(+)/K(+) transporter. EMBO J. 2020;39:e103256.
65. Li LF, Li YL, Jia Y, Caicedo AL, Olsen KM. Signatures of adaptation in the weedy rice genome. Nat Genet. 2017;49:811–4.
66. Zhao L, Wang S, Cao Z, Ouyang W, Zhang Q, Xie L, Zheng R, Guo M, Ma M, Hu Z, et al. Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. Nat Commun. 2019;10:3640.
67. Peng Y, Xiong D, Zhao L, Ouyang W, Wang S, Sun J, Zhang Q, Guan P, Xie L, Li W, et al. Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. Nat Commun. 2019;10:2632.
68. Du Y, Liu L, Peng Y, Li M, Li Y, Liu D, Li X, Zhang Z. UNBRANCHED3 expression and inflorescence development is mediated by UNBRANCHED2 and the distal enhancer, KRN4, in maize. PLoS Genet. 2020;16:e1008764.
69. Wang L, Jia G, Jiang X, Cao S, Chen ZJ, Song Q. Altered chromatin architecture and gene expression during poly-ploidization and domestication of soybean. Plant Cell. 2021;33:1430–46.
70. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 2006;1:2320–5.
71. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
72. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikit S, Song C, Xia L, 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.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
74. 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.
75. Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. Nat Methods. 2012;9:145–51.
76. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.
77. Benaglia T, Chauveau D, Hunter DR, Young DS. mixtools: an R package for analyzing finite mixture models. J Stat Softw. 2009;32:1–29.
78. Xu J, Zhao L, Liu D, Hu S, Song X, Li J, Lv H, Duan L, Zhang M, Jiang Q, et al. EWAS: epigenome-wide association study software 2.0. Bioinformatics. 2018;34:2657–8.
79. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
80. Gaspar JM, Hart RP. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. BMC Bioinformatics. 2017;18:528.
81. Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh CT, Jia Y, Gendler K, Freeling M, et al. Heritable epigenetic variation among maize inbreds. PLoS Genet. 2011;7:e1002372.
82. Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR. Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc. 2017;12:1659–72.
83. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16:259.
84. Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res. 2014;24:999–1011.
85. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
86. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

87. Jambunathan N. Determination and detection of reactive oxygen species (ROS), lipid peroxidation, and electrolyte leakage in plants. Methods Mol Biol. 2010;639:292–8.

88 Chen Y, Song S, Gan Y, Jiang L, Yu H, Shen L. SHAGGY-like kinase 12 regulates flowering through mediating CON-STANS stability in Arabidopsis. Sci Adv. 2020;6:eaaw0413.

89. Cao S, Sawettalake N, Li P, Fan S, Shen L. DNA methylation variations underlie lettuce domestication and divergence. BioProject accession: PRJCA017183. Genome Sequence Archive. 2024. https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA017183.

Publisher's Note

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