RESEARCH

Open Access



Epigenome profiling reveals distinctive regulatory features and *cis*-regulatory elements in pepper

Hongji Yang¹, Guorong Yu¹, Zhidong Lv¹, Tonghui Li^{1,2}, Xi Wang³, Ying Fu¹, Zhangsheng Zhu¹, Guangjun Guo⁵, Hang He¹, Ming Wang⁶, Guochen Qin¹, Feng Liu^{3,4*}, Zhenhui Zhong^{7*} and Yan Xue^{1*}

*Correspondence: jwszjx@hunau.edu.cn; zhenhuizhong@scu.edu.cn; yan. xue@pku-iaas.edu.cn

¹ State Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China ³ Engineering Research Center for Germplasm Innovation and New Varieties Breeding of Horticultural Crops, Key

Laboratory for Vegetable Biology of Hunan Province, College of Horticulture, Hunan Agricultural University, Hunan, China

⁷ Ministry of Education Key Laboratory for Bio-Resource and Eco-Environment, College of Life Sciences, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China

Full list of author information is available at the end of the article

Abstract

Background: Pepper (*Capsicum annuum*) is one of the earliest and most widely cultivated vegetable crops worldwide. While the large and complex genome of pepper severely hampered the understanding of its functional genome, it also indicates a rich yet unexplored reservoir of regulatory elements (REs). In fact, variations in the REs represent a major driving force in evolution and domestication in both plants and animals. However, identification of the REs remains difficult especially for plants with complex genomes.

Results: Here, we present a comprehensive epigenomic landscape of *Capsicum annuum*, Zhangshugang (ST-8), including chromatin accessibility, histone modifications, DNA methylation, and transcriptome. We also develop comparative crosslinked immunoprecipitation mass spectrometry to reveal the proteome associated with certain chromatin features. Through integrated analysis of these epigenetic features, we profile promoters and enhancers involved in development, heat stress and cucumber mosaic virus challenges. We generate stress responsive expression networks composed of potential transcription activators and their target genes. Through population genetics analysis, we demonstrate that some regulatory elements show lower nucleotide diversity compare to other genomic regions during evolution.

Conclusions: We demonstrate that variations in the REs may contribute to more diversified and agronomically desired phenotypes. Our study provides a foundation not only for studying gene regulation, but also for targeted genetic and epigenetic manipulation for pepper improvement.

Keywords: Epigenome, Pepper (*Capsicum spp.*), *Cis*-regulatory element, Promoter, Enhancer, Proteome, Development, Heat stress, Disease, Transcriptional regulation



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Pepper (*Capsicum* spp.) is the most widely grown spice in the world. It has important uses in pharmaceuticals, as natural coloring and an active ingredient in most defense repellents [1]. In 2021, the production of fresh and dry pepper reached 36.29 and 4.84 million tons respectively (FAOSTAT; http://faostat.fao.org/), harvested from over 3.68 Mha of planting area over the world.

Heat stress presents a devastating abiotic stress to pepper production. Increased flower abscission and failure in fruit setting take place at temperatures above 38 °C [1]. Immune responses are also affected. R gene mediated effector triggered immunity (ETI) is often suppressed at elevated temperatures [2]. For example, the first isolated plant virus resistance gene *N* encodes a TIR-NB-LRR protein, whose activity is attenuated at temperatures above 28 °C [3, 4]. On the other hand, RNA interference (RNAi), the major anti-viral immune response, is more active at high temperatures, potentially compensating for the compromised ETI [5].

Cucumber mosaic virus (CMV) presents a major biotic threat to pepper production. It causes severe systemic symptoms including dwarfism and fruit lesions, leading to drastic yield loss [6]. The broad range of hosts and insect vectors renders CMV control difficult. Intensive efforts have been devoted to search for resistant germplasm. Nevertheless, the levels of CMV resistance in the currently available commercial lines are still insufficient, prompting for a deeper understanding of the molecular mechanisms underlying CMV responses in pepper.

The reference genome of peppers has been recently assembled [7]. Surprisingly, it is significantly larger than that of its closest relative tomato [1, 8–10]. Approximately 80% of its over 3 Gb genome consists of transposable elements (TEs), which often carry their own *cis*-regulatory elements (CREs) to be expressed [1, 8]. Evidences show that TEs are important sources of functional CREs, such as enhancers and promoters [11, 12]. Therefore, the abundant TEs indicate a potentially rich reservoir of REs in pepper. Enhancers are characterized by their ability to drive transcription independently of the orientation, location or distance to the target genes, whereas promoters are positioned immediately upstream of their targets. Both enhancers and promoters are assembly centers for the transcription machinery. They carry a diverse array of transcription factor (TF) binding modules to ensure precise spatial–temporal control of gene expression [11, 13].

It is now evident that CRE diversification is a main driving force of evolution and domestication [14]. A classic example is the maize domestication gene *teosinte branched* 1 (*tb1*), where a transposon insertion in its regulatory region enhanced *tb1* expression. This genetic modification partially accounts for the increased apical dominance observed in modern maize compared to its wild ancestor, teosinte [15]. Similarly, the striking difference in fruit size between modern tomatoes and their wild progenitors can be partly attributed to *fw2.2* [16]. The heterochronic expression pattern and elevated expression levels of the domesticated *fw2.2* allele influenced cell division, thereby modulating fruit size [17].

Compare to the variations in the coding region, which often disrupt the function of a protein, variations in the CREs are less likely pleiotropic and may generate diverse phenotypes by modulating the spatial-temporal expression of a gene. It has been demonstrated in maize and tomato that Clustered Regularly Interspaced Short Palindromic

Repeats (CRISPR)/Cas9-mediated mutagenesis of the promoters regulating key developmental genes can give rise to a wide spectrum of phenotypic variations, enabling the identification of agronomically desired traits without the sacrifice from pleiotropic effects [18, 19].

Despite the potential, identification of CREs and their target genes is non-trivial, especially in organisms with complex genomes. It was first demonstrated in human that accurate prediction of CREs can be achieved with the help of their distinctive chromatin features [20]. This was recently adopted in plants including Arabidopsis, rice, maize and wheat [21–25]. In general, while chromatin features at promoters are relatively conserved, enhancer-associated chromatin features vary in plants. Specifically, plant promoters are marked by open chromatin, H3K4me3 and H3K27ac, while only open chromatin seems to be the conserved feature of plant enhancers.

In this study, we systematically profiled the CREs using *C. annuum* Zhangshugang (ST-8). To achieve a more accurate prediction, we integrated epigenetic features including chromatin accessibility, histone modifications, DNA methylation and transcription. As a result, we found that the promoters in pepper are marked by open chromatin, H3K4me3 and H3K27ac, while the enhancers are primarily characterized by open chromatin with highly variable H3K27ac and H3K4me3 levels. Additionally, we identified tissue specific and stress responsive CREs, reflecting their unique functions in development and stress responses. Functions of some of these CREs were validated experimentally. Through population genetics, we demonstrated that some regulatory elements accumulate fewer SNPs compare to the genome background.

In summary, our study presents a valuable foundation not only for the mechanistic study of gene regulation, but also for the genetic and epigenetic modulation for pepper improvement.

Results

Epigenome profiling in pepper leaves

To comprehensively profile the genome wide chromatin features of pepper, we performed Assay for Transposase-Accessible Chromatin with Sequencing (ATAC-seq) to reveal chromatin accessibility; Whole Genome Bisulfite Sequencing (WGBS) to examine DNA methylation at single nucleotide resolution, Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq) to detect distribution of histone marks, including H3K27ac, H3K4me3, H3K4me1, and adopted datasets of repressive histone marks including H3K27me3 and H3K9me2 from leaf tissue [9].

As expected, DNA methylation, especially in the CG and CHG context, and the constitutive heterochromatin mark H3K9me2 are highly concentrated in the pericentromeric regions populated with TEs. The facultative heterochromatin mark H3K27me3, on the other hand, is enriched in gene dense euchromatic arms. Similar patterns were observed with H3K4me1, H3K4me3, and H3K27ac (Fig. 1a). Interestingly, compared to CG and CHG methylation, methylation in CHH context remains relatively constant, regardless of TE or protein coding gene (PCG) density (Fig. 1a, Additional file 1: Fig. S1a–c). Next, we examined the distribution of these chromatin features relative to the functional elements. Compare to the genome average, H3K9me2 and DNA methylation are mainly distributed in the distal intergenic regions, while



Fig. 1 Epigenome profiling of pepper. **a** Circos plot illustrating epigenetic features across the chromosomes. **b** Distribution of the epigenetic features in the functional genome. **c** Chromatin states classified according to the epigenetic features using a multivariate HMM. **d** Distribution of each chromatin state in the functional genome. **e** Expression levels of each chromatin state. **f** Distance of each chromatin state to their proximal TSS. **g** Percentage of TEs in each chromatin state. **h**–**j** CG (**h**), CHG (**i**), and CHH (**j**) methylation levels in each chromatin state. The H3K27me3 and H3K9me2 ChIP-seq data are adopted from Liao *et al.* (2023)

H3K27me3 is enriched in promoters and genic regions (Fig. 1b). Active chromatin marks, H3K27ac and H3K4me3, are enriched immediately after the transcription start sites (TSS) (Additional file 1: Fig. S2a–b), whereas chromatin accessibility is hightest at TSS (Additional file 1: Fig. S2c). As expected, H3K4me1 is enriched over genic regions (Fig. 1b, Additional file 1: Fig. S2 d). Overall, the intensities of the active chromatin features are positively corelated with the transcription activities (Additional file 1: Fig. S2a–d), while the repressive chromatin marks, H3K9me2 and H3K27me3 are negatively corelated (Additional file 1: Fig. S2e, f). Gene body CG methylation was observed regardless of the expression (Additional file 1: Fig. S2 g) while non-CG methylations exist only over silent genes (Additional file 1: Fig. S2 h, i).

mCHH islands in pepper

Interestingly, we noticed a positive correlation between transcription and the CHH methylation immediately before the TSS (Additional file 1: Fig. S2i), suggesting the presence of mCHH islands in pepper. mCHH islands have been described in several plants, including rice and maize [26–28]. However, their biological relevance remains largely unknown. To functionally characterize them, we defined mCHH islands as CHH methylated patches less than 100 bp in length, located within 3 kb upstream of a TSS, and identified genes with (+ mCHH island) and without (-mCHH island) mCHH islands (Additional file 1: Fig. S3a). We found that the + mCHH island group possesses higher CHH and CHG methylation further upstream of the islands (Additional file 1: Fig. S3a, b), followed by a drastic depletion of DNA methylation in all sequence context. This is not observed in the -mCHH island group. Consistently, the depletion of DNA methylation is accompanied by a sharp increase in chromatin accessibility, H3K27ac and H3K4me3 levels (Additional file 1: Fig. S3c–e). Consistent with previous reports, genes with mCHH islands are more actively expressed (Additional file 1: Fig. S3f).

To answer whether mCHH islands influence transcription, we performed WGBS and RNA-seq in three other tissues, root, stem and floral bud. We identified mCHH islands that are present in leaf but absent in root, stem or bud (Additional file 1: Fig. S3 g–i). With these differential mCHH islands, we compared the expression of their downstream PCGs, as well as their upstream TEs. As a result, we did not observe significant changes in PCG expression (Additional file 1: Fig. S3j–l, left panels). Expression of TEs increases slightly in the absence of mCHH islands (Additional file 1: Fig. S3j–l, right panels), consistent with the role of DNA methylation in TE silencing. However, most TEs still remain silent, possibly due to the loss of functional sequences required for transcription during evolution.

Therefore, we conclude that in pepper, mCHH islands may not directly regulate PCG expression at least under regular growth conditions. Alternatively, redundant mechanisms may exist to safeguard the correct transcription activities.

Epigenetic characteristics of the CREs in pepper

To reveal the CREs within the genome, we applied a multivariate Hidden Markov Model (HMM), which takes into consideration the combinatorial epigenetic features and categorized the genome into 15 chromatin states, reflecting different biological functions [29].

A conserved feature of active CREs is open chromatin, to allow the access of TFs [11, 13, 30]. Therefore, we focused on Chromatin State 1, 4, 5, 9, and 10 for further analysis (Fig. 1c).

State 1 displays open chromatin, enrichment of H3K27ac and H3K4me3. Forty-four percent of the State 1 chromatin are located 3 kb within a TSS and their downstream genes are actively expressed (Fig. 1d, e). They possess low TE content and minimal DNA methylation in all sequence contexts (Fig. 1f–j). Therefore, these 44% of State 1 chromatin represent active promoters.

Twenty-nine percent of State 1 chromatin is annotated as distal intergenic (> 3 kb from a TSS). However, through visual inspection, we observed RNA expression 3 kb

within some of these regions (Additional file 1: Fig. S4a), suggesting that they are in fact, promoters of unannotated genes or non-coding RNAs, hereafter referred to as unannotated expressed regions (UERs). To reveal what percentage of State 1 distal intergenic components are promoters of UERs, we first defined UERs as regions with H3K4me1, a faithful mark of expressed genes, and active transcription but are not annotated (Additional file 1: Fig. S4b, c). We found that more than half of State 1 distal intergenic components are located 3 kb within a UER and therefore are actually promoters. In conclusion, State 1 chromatin are mostly promoters, characterized by open chromatin, enrichment of H3K4me3 and H3K27ac.

State 4 and State 10 are largely distal intergenic (Fig. 1d). Their proximal genes tend to be actively expressed (Fig. 1e). They carry relatively low levels of TEs and DNA methylation (Fig. 1g-j). These features indicate that State 4 and State 10 are potential enhancers. Interestingly, while State 4 is enriched of H3K27ac, State 10 is largely depleted (Fig. 1a), suggesting a conserved feature of enhancers is chromatin accessibility, while H3K27ac levels may vary.

State 5 and State 9 possess H3K4me1 in addition to open chromatin. While State 5 also contains high levels of H3K27ac, State 9 is relatively depleted, suggesting that State 5 and State 9 represent 5' and 3' end, respectively, of actively transcribed genes. Consistently, a large proportion of State 5 and 9 chromatin is genic and actively expressed (Fig. 1e-j).

H3K4me1 potentially marks poised enhancers in pepper

Through visual inspection, we observed H3K4me1 enrichment in intergenic regions, hereafter named as intergenic H3K4me1 (Additional file 1: Fig. S4 d). These intergenic H3K4me1 are clearly depleted of genic features, such as open chromatin, active histone modifications or RNA expression (Additional file 1: Fig. S4e, f). High H3K4me1/H3K4me3 ratio is a hallmark of animal enhancers [31]. However, similar observation has not been reported in plants. To test whether these regions are potential enhancers, we identified their most proximal genes regardless of the orientation and examined their expressions. We included equal number of random genes as control and intergenic H3K27ac proximal genes as potential active enhancer control. As a result, the intergenic H3K4me1 target genes show significantly lower expression compare to the other two groups (Additional file 1: Fig. S4 g).

In animals, lineage specification promoters are marked by H3K4me1 and H3K27me3 [32, 33]. These two marks function cooperatively to keep the genes at a poised state. Surprisingly, we also observed enrichment of H3K27me3 over these intergenic H3K4me1 regions, while H3K9me2 is depleted (Additional file 1: Fig. S4 h). H3K27me3 are known to regulate development or stress responsive genes. Consistently, we found that the intergenic H3K4me1 targeted genes are enriched in processes including reproduction and ion homeostasis (Additional file 1: Fig. S4i). Finally, these intergenic H3K4me1 regions are depleted of DNA methylation in all sequence contexts (Additional file 1: Fig. S4j). Altogether, we propose that these intergenic H3K4me1 are bivalently modified with H3K27me3 and may potentially be poised enhancers in pepper. They are likely reflected in the 15 chromatin states as State 7.

Identification and characterization of the active CREs in pepper

Next, we systemically profiled the active regulatory elements including promoters and enhancers. According to the previous analysis, regions 3 kb upstream or immediately downstream of a TSS, with open chromatin, enrichment of H3K27ac and H3K4me3 were identified as promoters. Enhancers were defined as intergenic regions (> 3 kb from a TSS) with high chromatin accessibility. Luciferase reporter assays were performed to validate 20 promoters and 5 enhancers through transient expression in *Nicotiana benthamiana*. Among these, 11 promoters robustly activated luciferase expression. Two enhancers were capable of further enhancing transcription on top of their corresponding promoters (Fig. 2a; Additional file 1: Fig. S5a, b).



Fig. 2 Characteristics of the CREs identified in pepper. **a** Luciferase reporter assays validating the activation function of a candidate promoter (left) and its predicted enhancer (right). **b** Top 5 enriched transcription factors, their binding motifs and enrichment in the promoters (left) and enhancers (right). **c**–**e** Intensities of chromatin accessibility (**c**), H3K27ac (**d**) and H3K4me3 (**e**) over the promoters and enhancers. **f**–**h** CG (**f**), CHG (**g**) and CHH (**h**) methylation levels over the promoters and enhancers. **i**–**j** Enrichment of H3K4me3 (**i**) and H3K27ac (**j**) over the promoters and enhancers. **k**–**l** Transcription activities ranked by H3K27ac (left panels), chromatin accessibility (middle panels) and H3K4me3 (right panels) intensities at promoters (**k**) and enhancers (**l**). **m**–**n** Percentage of the enhancers interacting with the promoters revealed by HiC ranked by H3K4me3 (**m**) and H3K427ac (**n**) levels. The HiC data are adopted from Liao *et al.* (2023)

Both enhancers and promoters function as transcription organizing centers and binding platforms for TFs. Through motif search, we found that promoters and enhancer are enriched with different classes of TF binding motifs (Fig. 2b). This is consistent with the observations in wheat [22].

Compare to the promoters, active enhancers display higher chromatin accessibility but lower levels of H3K27ac and H3K4me3 (Fig. 2c–e). Consistent with their intergenic localization, enhancers possess higher levels of DNA methylation in all sequence context (Fig. 2f–h).

Consistent with our previous observations that neither H3K27ac nor H3K4me3 is a faithful mark of enhancer, we found a significant portion of the enhancers depleted of these two marks (Fig. 2i, j). This prompted us to ask why the enhancers vary in the levels of these active histone modifications. Interestingly, we found that while the intensities of chromatin accessibility, H3K27ac and H3K4 me3 are all positively correlated with transcription activities at promoters (Fig. 2k), only chromatin accessibility and H3K27ac moderately reflect transcription activities at the enhancers (Fig. 2l). To validate, we performed the same analysis in other tissues and consistently found that unlike promoters, enhancer activities are only moderately reflected by the active chromatin marks (Additional file 1: Fig. S6a–f).

Enhancers may interact with promoters to further activate transcription. We then asked whether the presence of H3K4me3 and H3K27ac is associated with such interactions. Accordingly, we separated the enhancers into 4 groups according to their H3K4me3 or H3K27ac levels, and detected promoter-interacting-enhancers in each group using the published HiC data [9]. As a result, we observed more interactions within the groups with higher H3K4me3 or H3K27ac (Fig. 2m–n). However, we were not able to differentiate whether the active mark at enhancers is a cause (H3K4me3 and H3K27ac promote enhancer-promoter interaction) or consequence (H3K4me3 and H3K27ac are from the promoters that happen to be in contact) of interaction with promoters.

In summary, we characterized promoters and enhancers in the leaf tissue. We found that while the promoter activities are strongly correlated with chromatin accessibility, H3K27ac and H3K4me3, enhancer activities are only moderately associated with these chromatin features in pepper. It should be noted that distal CREs may interact across megabases of DNA, and even across chromosomes [34–36]. Some enhancer may have multiple targets, while others display target specificity [37, 38]. Some genes may also have multiple enhancers that are simultaneously active [39]. We here assumed that the target of an enhancer is its closest gene regardless of the orientation. Despite the simplicity, it is remarkably accurate among the validated enhancers [39, 40]. In addition, reporter assays showed that enhancers inserted closest to the promoters drove the strongest activation [41]. However, it is possible that some enhancers predicted in this study are actually distal REs of a different kind, and may target different genes. Experimental validations are needed to confirm the identity of a CRE and its target gene.

Identification of active promoter associated proteome

In order to reveal proteins associated with different chromatin environment, we performed XLIPMS against modified histones, including H3K4me1 and H3K4me3



Fig. 3 Identification of proteins associated with specific chromatin environment using XLIPMS. **a** Illustration of the XLIPMS workflow. **b** Cluster of proteins identified in different chromatin environment. **c**-**e** GO term analysis of H3K36me3 unique (**c**), H3K4me3 unique (**d**), and H3K4me1 unique (**e**) associated proteome. **f** Examples of transcription factors and chromatin modifying proteins specifically associated with H3K4me1 or H3K4me3

(Fig. 3a). We included H3K36me3 as a control to reveal genic proteins in close proximity with H3K4me3 marked promoters and overlapping with genic H3K4me1. Finally, we clustered the identified proteins according to their association with these chromatin marks (Fig. 3b; Additional file 2: Table S1).

To assess the reliability of the dataset, we first examined the distribution of components involved in different steps of transcription. We detected more mRNA processing proteins, especially pre-mRNA splicing components in H3K4me1 and H3K36me3 associated proteome, in comparison to H3K4me3 (Table 1) [42]. Mediators and DNA directed RNA polymerases are enriched in H3K4me3 marked promoters as well as H3K4me1 marked expressed genes, with a bias to H3K4me1 (Table 2). Interestingly, we also detected a large number of splicing factors in H3K4me3 (Table 1), indicating that splicing is tightly coupled to transcription immediately after it occurs and the resolution of our method is not high enough to demarcate regions in such close proximity (Fig. 3a). It should be noted that we observed multiple genes with the same annotation (Tables 1 and 2), reflecting a high redundancy of genes or a substantial gap in the functional gene annotation in pepper. Finally, we performed GO term analysis of the proteins associated with different chromatin environment. Consistently, we observed "RNA splicing," "RNA processing" among the top enriched functions with H3K36me3 and "regulation of gene expression" with H3K4me3, validating the quality of our results (Fig. 3c-d). Interestingly, "rRNA processing" and "ncRNA metabolic processes" are enriched in H3K4me1 specific clusters, reflecting a potential role of the H3K4me1 in the regulation of noncoding RNAs (Fig. 3e).

In order to search for the TFs associated with active promoters, we examined proteins associated specifically with H3K4me3. Among the annotated TFs, we identified zinc-finger proteins, global transcription factor group E (GTE) and trihelix TFs (Fig. 3f), which are likely broadly expressed and thus easily detected. Additionally, we detected H3K4 methyltransferases ATX3 and ATX5, specifically in the H3K4me3 associated proteins, further validating the reliability of our dataset (Fig. 3f).

Finally, we took advantage of our comparative proteomic data to identify proteins potentially associated with intergenic H3K4me1. Specifically, we searched through the H3K4me1 unique cluster. Interestingly, we detected the H3K4 methyltransferases ATX2, the H3K27 methyltransferase EZ2, the H3K27me3 eraser REF6 and H3K4me3-H3K27me3 dual reader SHL and EBS (Fig. 3f). The presence of H3K27me3 reader, writer and eraser demonstrated the presence and dynamics of H3K27me3. As these regions are depleted of H3K4me3 (Additional file 1: Fig. S3 d), and the gene annotation is based on homology to well characterized proteins, it is likely that the two dual reader homologues in pepper may recognize H3K4me1 instead of H3K4me3. Alternatively, the dual readers might be associated with genic H3K4me1 which cannot be separated from the proximal H3K4me3/H3K27me3 bivalent promoters using our method.

In summary, through comparative XLIPMS, we demonstrated the possibility to profile proteins associated with specific chromatin environment. Using this method, we identified potential TFs and chromatin modifiers that are specifically associated with H3K4me3 marked active promoters. Further experiments are required to validate the presence, specificity and functions of these proteins.

Identification of the tissue specific CREs

Next, we moved onto identifying the tissue specific regulatory elements. We first profiled the CREs in leaf, bud, stem and root using the criteria described before. Based on this, we defined tissue specific promoters if they possess higher levels of H3K4me3,

	מסודט מווווסנמנכת וווויוזאי אוסככסזוווק כסווואסווכוונס ווו מוכ אבו			
	mRNA processing	H3K4me1	H3K4me3	H3K36me3
Caz01g39830	Pre-mRNA-splicing factor CWC22-like protein			
Caz06g01190	Pre-mRNA-splicing factor SLU7			
Caz01g11900	Splicing factor YJU2			
Caz05g15750	splicing factor-like protein 1			
Caz10g15430	splicing factor U2af small subunit B-like			
Caz05g14310	Pre-mRNA-splicing factor ISY1-like protein			
Caz09g11110	pre-mRNA-processing factor 39			
Caz03g20310	Pre-mRNA cleavage factor Im 25 kDa subunit			
Caz05g02630	Pre-mRNA-splicing factor 38			
Caz02g16800	splicing factor 3B subunit 2			
Caz03g04190	pre-mRNA-splicing factor SPF27 homolog			
Caz03g36240	Pre-mRNA-processing factor 19			
Caz03g36360	mRNA cap-binding protein			
Caz05g04260	Splicing factor 3B subunit 3			
Caz05g22780	mRNA-decapping enzyme subunit 2			
Caz06g05800	mRNA-decapping enzyme-like protein			
Caz07g07020	PREDICTED: pre-mRNA-processing-splicing factor			
Caz10g22370	Pre-mRNA-splicing factor ATP-dependent RNA helicase			

Table 1 Detection of annotated mRNA processing components in the XLIPMS

.

	RNA Polymerase/Mediator	H3K4me1	H3K4me3	H3K36me3
Caz09g11180	Mediator of rna polymerase ii transcription subunit 33a			
Caz01g04110	DNA-directed RNA polymerase			
Caz01g23050	DNA-directed RNA polymerase subunit beta			
Caz01g30230	RNA polymerase sigma factor			
Caz02g11870	DNA-directed RNA polymerase V subunit 7			
Caz11g00970	DNA-directed RNA polymerases II, IV and V subunit 11			
Caz07g01640	DNA-directed RNA polymerase subunit			
Caz10g20920	DNA-directed RNA polymerase subunit beta			
Caz08g07790	Mediator of RNA polymerase II transcription subunit 12			
Caz08g10580	Mediator of RNA polymerase II transcription subunit 14			
Caz04g22430	DNA-directed RNA polymerases II and V subunit 8A-like			
Caz05g02100	DNA-directed RNA polymerases IV and V subunit 4			
Caz08g11670	DNA-directed RNA polymerase V subunit 5A-like			
Caz01g33300	DNA-directed RNA polymerases I and III subunit RPAC1			
Caz02g16560	Mediator of rna polymerase ii transcription subunit 16			
Caz01g19300	DNA-directed RNA polymerase subunit beta			
Caz01g20570	DNA-directed RNA polymerase subunit beta			
Caz02g16310	DNA-directed RNA polymerase subunit beta			
Caz02g21640	DNA-directed RNA polymerase subunit			
Caz03g02070	RNA polymerase II-associated, Paf1			
Caz03g12800	DNA-directed RNA polymerase subunit beta			
Caz03g15750	pre-mRNA-splicing factor SYF1			
Caz03g39650	DNA-directed RNA polymerase subunit beta			
Caz05g00370	DNA-directed RNA polymerase V subunit 5A			
Caz06g25970	DNA-directed RNA polymerase II subunit RPB7			
Caz08g09580	DNA-directed RNA polymerase subunit			
Caz11g20520	DNA-directed RNA polymerase II subunit, putative			
Caz12g23780	DNA-directed RNA polymerases II and IV subunit 5A			

Table 2 Detection of annotated RNA polymerases and mediators in the XLIPM	MS
---	----

H3K27ac, chromatin accessibility and their target gene expression is at least four-fold higher than any other tissue studied. Tissue specific enhancers were defined similarly without the consideration of H3K4me3 or H3K27ac (Fig. 4a, b; Additional file 3: Table S2). As enhancers may interact with their target promoters to further activate transcription, we adopted the published HiC data to detect any interaction between these tissue specific enhancers and their target promoters. As a result, 36% and 46% of leaf and bud specific enhancers display physical interaction with their target gene promoters (Additional file 3: Table S2).

GO term analysis were performed with leaf specific CRE target genes. The top enriched biological processes are photosynthesis and energy generation for both promoters and enhancers (Fig. 4c), suggesting that the CREs identified are reliable. Then we asked whether the repression of bud, stem and root specific CREs in leaf tissue is a consequence of H3K27me3 deposition. We adopted the published leaf-H3K27me3 ChIP seq data and examined its enrichment over the tissue specific CREs [9]. We found that in leaf, H3K27me3 levels remain lowest over leaf specific promoters compare to the bud, root and stem specific promoters (Fig. 4d, e). However, no significant differences were observed at the enhancers (Fig. 4d, e). Therefore,



Fig. 4 Identification and characterization of tissue specific CREs. a Screenshots showing examples of tissue specific promoters (left panels) and enhancers (right panels). b Heatmaps showing the enrichment of H3K27ac, chromatin accessibility, H3K4me3 and target gene expression of tissue specific promoters (upper panels) and enhancers (lower panels). c Functional enrichment of leaf specific promoter target genes (left panel) and leaf specific enhancer target genes (right panel). d Heatmaps showing the enrichment of leaf H3K27me3 over tissue specific promoters (left panels) and enhancers (lower panels). c Functional enrichment of leaf specific promoter target genes (left panel) and leaf specific enhancer target genes (right panel). d Heatmaps showing the enrichment of leaf H3K27me3 over tissue specific promoters (upper panel) and enhancers (lower panel) and enhancers (lower panel) and enhancers (lower panel) and enhancers (lower panel)

H3K27me3 is a highly conserved repressive mark for developmentally regulated genes, especially over promoters, while other repressive mechanisms may exist at enhancers in pepper.

Finally, we identified CREs relatively equally active in all four tissues and defined them as non-specific CREs (Additional file 1: Fig. S7a). Their target genes are enriched in housekeeping functions including RNA processing, cellular localization and catabolic process (Additional file 1: Fig. S7b).

In summary, we identified tissue specific and non-specific CREs. We showed that tissue specificity is achieved through the deposition of H3K27me3 over promoters. Although deposition of H3K27me3 is also observed over enhancers, other mechanisms may exist to repress the transcription activation potential of the enhancers.

Identification of the heat stress responsive CREs

Heat stress greatly impacts plant growth and development. The optimal growth temperature for pepper fruit maturation is between 21 and 29 °C. Beyond 32 °C, growth and reproduction are affected, leading to significant yield loss [43]. Epigenetic changes are intricately associated with stress resistance. Studies in Arabidopsis showed that key DNA methyltransferases including MET1, CMT3, and DRM2 are upregulated in response to heat, while natural occurring loss of function alleles of CMT2 underlies adaptation to hot areas [44, 45].

Therefore, we set out to understand how pepper responds to high temperature stresses epigenetically. Seedlings were stressed for two days at 40 °C and heat responsive CREs were identified using the criteria described before (Fig. 5a–c; Additional file 4: Table S3). To validate with transient luciferase activity assays, we first tested survival of agrobacteria under heat stresses. We found that while the expression of luciferase driven by 35S promoter remains relatively constant at 30 °C, it diminished significantly after 3 h at 40 °C (Fig S8a), consistent with the growth condition of agrobacteria. Therefore, we performed validation with heat treatment at 30 °C for 10 h. As a result, 2 out of the 5 promoters were responsive (Fig S8b). It should be noted that the CREs were identified in pepper after 48 h at 40 °C. Therefore, some CREs might be responsive at higher temperatures for a longer exposure.

Through GO term analysis, we found that heat induced promoters regulate genes involved in processes such as protein folding and heterochromatin formation. Among these, we identified components in the RNA directed DNA methylation (RdDM) pathway, including SUVH5 and IDN2 (Fig. 5d, Additional file 4: Table S3). RdDM is a plant specific pathway, the main function of which is de novo DNA methylation and silencing of newly integrated transposons [46]. Upregulation of RdDM components in response to heat has also been reported in Arabidopsis and this confers protection against the stress [47]. Therefore, the involvement of RdDM in heat responses may be conserved in plants. Other induced promoter target genes include LHP1 and ELF7, indicating a finetuned H3K27me3 dynamics.

A well-known damage caused by heat stress is altered membrane fluidity and thus ion homeostasis. Correspondingly, we observed enrichment in "ion transport", including potassium channel like, potassium transporter, cyclic nucleotide-gated ion channel (Fig. 5d, Additional file 4: Table S3). Photosynthesis is one of the most heat sensitive physiological processes [48]. The integrity of thylakoid membranes and therefore the activity of especially Photosystem II are disrupted under heat stress. Consistently, in our study, the most repressed biological function is photosynthesis. Besides, we detected repression in "regulation of development," "glucan metabolism," and "responses to light," indicating a tradeoff between growth and stress response (Fig. 5d).

We then intended to identify the TFs associated with the heat responsive CREs. We performed TF footprint analysis with the ATAC-seq data. Motifs occupied by TFs display a regional closed conformation and therefore are less accessible to the transposases [49]. Notably, the WRKY TFs, well known to be involved in stress responses, are highly enriched in heat repressed promoters, suggesting a key role in repressing transcription under heat stress (Fig. 5e).



Fig. 5 Identification and characterization of heat responsive CREs. **a** Screenshots showing examples of promoters (left panels) and enhancers (right panels) that are induced (upper panels) or repressed (lower panels) in response to heat stress. **b** Heatmaps showing the enrichment of H3K27ac, chromatin accessibility, H3K4me3 and target gene expression of heat induced (upper panel) and heat repressed (lower panel) promoters. **c** Heatmaps showing the enrichment of chromatin accessibility and target gene expression of heat induced (upper panel) and heat repressed (lower panel) enhancers. **d** GO functional enrichment of heat responsive CREs. **f** Expression patterns of the potential heat responsive transcription activators. **g** Coexpression network of potential heat responsive transcription activators and their target genes

Next, we moved on to search for heat responsive transcription activators. We assume that TFs showing stronger interaction with the induced CREs are potential transcription activators and are likely co-expressed with their target genes. Therefore, we focused on the TFs meeting the following criteria: (1) they show increased interaction with heat induced CREs; (2) their target genes are specifically induced by heat; (3) their own expressions are induced by heat.

Accordingly, we first screened for heat activated CREs whose target genes are specifically activated in response to heat stress. Indicated by TF footprint, we then obtained a list of TFs showing stronger interaction with these CREs (Fig. 5e). Pepper homologues were subsequently identified and those specifically activated by heat stress were considered as potential heat responsive transcription activators (Fig. 5f). Finally, combining the potential transcription activators and their target genes, we constructed a heat activated regulatory network (Fig. 5g), consisting of ERF, NAC, bHLH, and Myb family transcription factors.

It should be noted that the abovementioned transcription network is based on several assumptions. First, the candidate TFs in pepper display the same motif preference as their homologous TF models indicated in the TF footprint. Second, the candidate TFs in pepper indeed interact with the heat responsive CREs identified in our study. Third, such interactions increase in response to heat stress. Forth, increased TF-CRE interaction recruit cofactors that modify the chromatin environment and activates transcription. Further experiments including ChIP are required to validate the transcription network proposed in our study.

In summary, we identified potential heat responsive CREs and TFs. Through expression analysis, we proposed potential heat activated transcription network consisting of TFs and their target genes in pepper.

Identification of the CMV responsive CREs

CMV is one of the most devastating diseases in pepper worldwide. It can cause up to 80% yield loss during severe local epidemics [6]. The broad range of insect vectors and hosts render disease control difficult. Therefore, intensive efforts have been spent on the identification of resistant varieties and to understand the defense mechanisms.

In this study, we systemically profiled CMV responsive CREs in ST-8, a CMV sensitive variety and PBC688, a CMV resistant variety [50] (Fig. 6a, b; Additional file 5: Table S4). Among the CMV responsive CRE target genes, 231 are specifically induced and 428 are specifically repressed in PBC688, suggesting their potential involvement in resistance (Fig. 6c). GO term analysis revealed that the PBC688 specific genes, either induced or repressed, are enriched in system development and reproduction (Fig. 6d), indicating that growth and reproduction is finetuned in PBC688. Photosynthesis and circadian rhythm are specifically found in PBC688 induced genes, suggesting that PBC688 is more robust in maintaining energy metabolism under CMV challenge (Fig. 6d).

To identify the transcription activators potentially conferring resistance in PBC688, we performed co-expression analysis with PBC688 specific differential CREs using the criteria described before. As a result, we detected several TFs belonging to the AP2/ERF, DREB, NAC and WRKY transcription factor families. The co-expression network was constructed accordingly.

In summary, by comparing the CMV sensitive ST-8 and resistant PBC688, we identified CMV responsive CREs and TFs that may potentially function as transcription activators. Through co-expression analysis, we proposed potential CMV activated transcription network consisting of TFs and their target genes.

Candidate genomic variations conferring CMV resistance in PBC688

The CMV resistance loci of PBC688 have been previously determined to two genomic regions: a 330 kb region on chromosome 2 and a second 2.05 Mb region on chromosome 11 [50]. Identification of the resistance genes has been hampered partly by the lack of complete PBC688 sequence information. Here we took advantage of our H3K4me1



Fig. 6 Identification and characterization of CMV responsive CREs. **a** Screenshots showing examples of promoters and enhancers responsive to CMV. **b** Heatmaps showing the enrichment of H3K27ac, chromatin accessibility, H3K4me3 and target gene expression of CMV responsive CREs. **c** Venn diagram showing the overlap of DEGs between ST-8 and PBC688. **d** GO term enrichment of the CMV responsive DEGs. **e** Potential transcription activators specifically upregulated by CMV. **f** Transcription network of the potential activators and their target genes

and RNA seq data, covering the genic regions of the expressed genes; and H3K4me3, H3K27ac, ATAC-seq data, covering the regulatory regions, to reveal the genomic variations within the functional elements between ST-8 and PBC688 (Additional file 6: Table S5 and Additional file 7: Table S6).

Previously, one gene (Caz02g20380 in ST-8; CA02g19570 in the previous study using CM334 as the reference genome) encoding an N-like TIR-NBS-LRR protein was proposed to be the candidate resistance gene because its expression is elevated in PBC688. In our study, we consistently found that Caz02g20380 expression is higher in PBC688, which is further elevated after CMV challenge (Additional file 1: Fig. S9a, b), while in ST-8, Caz02g20380 expression is slightly repressed in response to CMV. Interestingly, we did not detect any non-synonymous nucleotide variation within the coding region (Additional file 1: Fig. S9c, left panel). Instead, we identified a patch of SNPs within its

promoter (Additional file 1: Fig. S9c, right panel). Whether these SNPs influence the expression of this R gene and thus lead to CMV resistance require further experimental validation.

The second region explaining 11% of the resistance was defined on chromosome 11. Within this region, we identified 3 genes carrying non-synonymous SNPs (Additional file 7: Table S6). One of them (Caz11g16520) encodes an RNA dependent RNA polymerase (RdRP). Expression of this RdRP is induced in ST-8 after CMV challenge but repressed in PBC688 (Additional file 1: Fig. S9 d, e). Two conserved amino acids in the RdRP domain were found to be different in PBC688 (Additional file 1: Fig. S9f, left panel; Additional file 1: Fig. S9 g). Several SNPs were also detected in the promoter region (Additional file 1: Fig. S9f, right panel). A complete list of sequence variations between ST-8 and PBC688 detected within the two defined region is provided (Additional file 6: Table S5 and Additional file 7: Table S6).

In summary, we demonstrated the possibility to identify genomic variations within the function genome through epigenome profiling. This is particularly informative when the variations occur within regulatory regions. Furthermore, it will be an excellent alternative to whole genome resequencing for crops with large and complex genome.

Heat and CMV stress crosstalk

Current climate prediction indicates a gradual increase in the frequency and amplitude of heat episodes. Heat influences pathogenicity and host defense responses. While some studies showed that high temperature compromises plant immunity, opposite observations have also been reported [3, 51, 52].

Here, we compared the CMV and heat responsive CREs and identified a large number of genes influenced by both heat and CMV (Additional file 1: Fig. S10a, b). Through GO term analysis, we found that genes up regulated by both heat and CMV include pathogenesis related protein PR-4 like, potentially involved in defense response to fungus (Additional file 1: Fig. S10c, d). Therefore, heat stress may stimulate fungal resistance in pepper. As expected, both heat and CMV stress leads to the repression of genes involved in several metabolic pathways, consistent with the observation that growth and development are compromised in stress responses.

Diversity of the CREs

Key functional elements in the genome tend to accumulate fewer mutations. Accordingly, we examined the nucleotide diversity of the CREs at a population level. We adopted the genome sequences of 500 accessions, covering 7 species under 3 clades [10]. Clade Pubescens is likely the most ancient one, from which evolved Clade Baccatum and Clade Annuum. The 5 cultivated species are *C. annuum* var. *annuum*, domesticated from the wild *C. annuum* var. *glabriusculum*; *C. baccatum* var. *pendulum*, domesticated from the wild *C. baccatum* var. *baccatum*; *C. chinense*, *C. frutescens*, and *C. pubescens* [10].

We first examined the nucleotide diversity of the non-specific CREs, involved in the housekeeping functions (Additional file 1: Fig. S6). We found that the overall genetic diversity in the ancient Clade Pubescens remains the lowest, which increased in the wild Baccatum (*C. Baccatum* var. *baccatum*) and Annuum (*C. annuum* var. *glabriusculum*)

(Fig. 7a). Comparison between the wild and domesticated Baccatum species (*C. baccatum* var. *baccatum* and *C. baccatum* var. *pendulum*, respectively), revealed that the overall genetic diversity is much higher in the wild species, consistent with the conclusion



Fig. 7 Nucleotide diversity of the CREs. **a** Nucleotide diversity (indicated by the *y*-axis) of the non-specific promoters (upper panels) and enhancers (lower panels) in the three clades of peppers. **b** Nucleotide diversity of heat repressed promoters (upper panels) and enhancers (lower panels). **c** Nucleotide diversity of heat induced promoters (upper panels) and enhancers (lower panels). **d** Nucleotide diversity of CMV repressed promoters (upper panels) and enhancers (lower panels). **d** Nucleotide diversity of CMV repressed promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** Nucleotide diversity of CMV induced promoters (upper panels) and enhancers (lower panels). **e** 17; *C*. baccatum var baccatum; n = 9; *C*. baccatum var pendulum; n = 109; *C*. annuum var glabriusculum; n = 22; *C*. annuum var annuum; n = 90; *C*. frutescens: n = 99; *C*. chinense: n = 115

that domestication results in a severe tradeoff between desired agronomic traits and genetic diversity. Importantly, we observed a sharp decrease in the nucleotide diversity in the non-specific CREs, which extends to at least 5 kb beyond. The decrease in nucleotide diversity is more significant in the wild population compared to the domesticated population. It is also more significant at the promoters compared to the enhancers. Similar observations were found between the other wild-domesticated pair in Clade Annuum.

Domestication of the other two species *C. frutescens* and *C. chinense* are less intense [10]. Consistently, the overall nucleotide diversity in these two populations is much higher than the intensively domesticated *C. baccatum* var *pendulum* and *C. annuum* var *annuum*. As expected, we observed a strong depletion of nucleotide diversity over the CREs in *C. frutescens* and *C. chinense*. In conclusion, the non-specific CREs, largely involved in the housekeeping functions, accumulate fewer SNPs throughout evolution than the genetic background (Fig. 7a).

We next examine the stress responsive CREs. Overall, heat responsive CREs display similar patterns to the non-specific CREs (Fig. 7b). However, we did not find significant diversity change in CMV responsive enhancers (Fig. 7c), potentially indicating that CMV is not a devastating threat for pepper in their natural habitat. This might also partly explain the lack of species with high resistance to CMV.

In summary, through population genetics, we concluded that CREs may be the selection targets both during evolution and domestication.

Discussion

Being one of the most important vegetable crops worldwide, pepper has wide use in food and pharmaceutical industry. In this study, we systematically profiled and analyzed the epigenome of *C. annuum* Zhangshugang (ST-8).

We detected the presence of mCHH islands, further supporting that this is likely a common feature in plants with large genomes. Previous studies showed that mCHH islands are a product of RdDM. They may function as an insulator to prevent the spread of PCG transcription to the proximal TEs, causing the activation of TEs [26]. In our study, we consistently found that the island proximal genes display higher expression compare to the ones without. Loss of island methylation leads to the reactivation of a number of TEs, but does not influence the expression of PCGs. Therefore, we suggest that the mCHH islands are likely a product of active transcription from PCGs, mediated by the non-canonical RdDM pathway, to maintain TE silencing. During evolution, some TEs may have lost features necessary for expression and therefore, remain silent even at the absence of mCHH islands.

By analyzing the H3K4me1 distribution, we observed a large number of PCGs or ncR-NAs either failed to be annotated or annotated incorrectly. This could be due to the dynamic nature of gene expression. Compare to RNA expression, epigenetic modification is much more stable. For example, H3K4me1 and H3K36me3 both faithfully and stably mark the genic region of expressed genes. Therefore, integration of epigenetic features might ensure a more accurate annotation of functional genome. Surprisingly, we also detected intergenic H3K4me1. Because these intergenic H3K4me1 co-localize with H3K27me3 and their proximal genes are expressed at low levels, we proposed that they might function as poised enhancers. H3K4me1 is a hallmark of enhancers, regardless of the activity, and may also mark poised promoters in animals [32, 33]. However, similar observations have not been reported in plants. Whether these intergenic H3K4me1 are indeed poised enhancers and whether this is a unique feature in pepper require careful experimental analysis.

Conclusions

CREs carry a diverse array of transcription factor (TF) binding modules to ensure precise spatial-temporal control of gene expression [11, 13]. It is now evident that variations in CREs represent a main driving force of evolution, domestication and breeding. They often lead to moderate changes in gene expression pattern, timing or level, and thus are less pleiotropic compare to altering the gene coding sequences. Given sufficient knowledge of the regulatory modules and the protein components executing the regulation, it should be possible to design CREs and generate desired phenotypes. Therefore, a comprehensive identification and characterization of CREs is the foundation not only for the mechanistic understanding of transcription regulation, but also for targeted genetic or epigenetic modification for crop improvement.

Methods

Plant materials and growth conditions

C. annuum "Zhangshugang" (ST-8) was kindly provided by Dr. Feng Liu. PBC688 was kindly provided by Dr. Guangjun Guo. Seeds were germinated at 28 °C in the dark and transferred to ½ Hoagland solution (Phygene Biotechnology, China) in a greenhouse condition (16 h light/8 h dark). Hoagland solution was changed every 3 days. Leaf, stem and root tissues were collected at the 6-leaf stage. Floral buds were collected 60 days after germination. Heat treatment was performed with ST-8 at the 6-leaf stage in a growth chamber at 40 °C for 2 days. Leaf samples were collected at the end of the treatment. CMV treatment was performed with the sensitive line ST-8 and the resistant line *C. frutescens* "PBC688." CMV_{FNY} was provided by Dr. Qian Zhou at Hunan Agricultural University. Leaf sap from infected tobacco leaves were infiltrated into two true leaves at the 3-leaf stage. Symptoms were observed 1 month after inoculation and systemic leaves were collected for analysis.

ChIP seq and data processing

Two grams of tissues pooled from at least three plants were used for each replicate and two biological replicates were performed for each ChIP. All samples were crosslinked in vitro with Nuclei Isolation Buffer (50 mM HEPES, 1 M sucrose, 5 mM KCl, 5 mM MgCl₂, 0.6% Triton X-100, 0.4 mM PMSF, 5 mM benzamidine, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)) supplemented with 1% formaldehyde for 12 min with rotation. Glycine was immediately added to stop the crosslinking. Lysate was filtered through Miracloth and centrifuged for 20 min at 2880 g at 4 °C. The pellet was resuspended in 1 ml of extraction buffer 2 (0.25 M sucrose, 10 mM Tris–HCl pH 8, 10 mM MgCl₂, 1% Triton X-100, 5 mM BME, 0.1 mM PMSF, 5 mM benzamidine, and 1 × protease inhibitor cocktail tablet) and centrifuged at 12,000 g for 10 min at 4 °C. The pellet was resuspended in 500 µL extraction buffer 3 (1.7 M sucrose, 10 mM Tris–HCl pH

8, 2 mM MgCl₂, 0.15% Triton X-100, 5 mM BME, 0.1 mM PMSF, 5 mM benzamidine, $1 \times$ protease inhibitor cocktail tablet) and layered on top of 500 µL extraction buffer 3 (1.7 M sucrose, 10 mM Tris-HCl pH 8, 2 mM MgCl₂, 0.15% Triton X-100, 5 mM BME, 0.1 mM PMSF, 5 mM benzamidine, $1 \times$ protease inhibitor cocktail tablet) and centrifuged at 12,000 g for 1 h at 4 °C. The pellet was lysed with 400 μ L nuclei lysis buffer (50 mM Tris pH 8, 10 mM EDTA, 1% SDS, 0.1 mM PMSF, 5 mM benzamidine, 1 × protease inhibitor cocktail tablet). A total of 1.7 ml of chIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH 8, 167 mM NaCl, 0.1 mM PMSF, 5 mM benzamidine, 1 × protease inhibitor cocktail tablet) was added to the lysed nuclei. Chromatin was sheared with Bioruptor Plus (Diagenode) for 30 min with 30 s on and 30 s off and incubated with the corresponding antibodies overnight at 4 °C (anti H3 K4 me1 antibody (Abcam), anti H3 K4 me3 antibody (Abcam), anti H3 K27ac antibody (Abcam)). Chromatin-bound proteins were immunoprecipitated with Protein A and Protein G magnetic Dynabeads (Invitrogen) for 2 h at 4 °C. Dynabeads were washed twice with low salt buffer (150 mM NaCl, 0.2% SDS, 0.5% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8), once with high salt buffer (200 mM NaCl, 0.2% SDS, 0.5% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8), once with LiCl buffer (250 mM LiCl, 1% Igepal, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris pH 8), and once with TE buffer (10 mM Tris pH 8, 1 mM EDTA). Elution was performed with 250 ul elution buffer (1% SDS, 10 mM EDTA, 0.1 M NaHCO₃) by incubating at 65 °C with shaking twice. Four hundred microliters of eluted complexes were reverse crosslinked by incubation at 65 °C overnight with the addition of 20 μ l of 5 M NaCl followed by protease K treatment (20 µg in 10 mM EDTA and 40 mM Tris pH 8) at 45 °C for 1 h. DNA fragments were precipitated with EtOH overnight at -20 °C. Libraries were prepared with Ovation Ultra Low System V2 kits following the manufacturer's instructions.

All libraries were sequenced at a length of 150 bps pair-end with the NovaSeq 6000 platform (Illumina). Raw reads were aligned to the ST-8 reference genome with Bowtie2 (v2.5.3) [53]. Peaks were called with MACS2 (v2.2.7.1) [54] (q value < 0.01).

Promoters were defined as regions containing H3K27ac, H3K4me3 and chromatin accessibility, which are located 3 kb within a TSS, as defined by ChIPseeker [55, 56]. In addition, their predicted target genes need to be expressed (TPM >0 in all replicates). Enhancers were defined similarly without the consideration of histone modifications. Tissue specific promoters were defined if the active chromatin features are most enriched in one tissue (FC > 1) and their target gene expression is at least fourfold higher than any other tissue included in this study.

RNA seq and data processing

RNA seq was performed with three biological replicates. Each biological replicate contains pooled samples from three plants. Total RNA was extracted using FreeZol Reagent according to the manufacturer's instruction (Vazyme). Libraries were made with VAHTS Unive V8 RNA-seq Library Prep Kit for Ilumina according to the manufacturer's instruction (Vazyme). Sequencing was performed with the Illumina NovaSeq 6000 system.

Reads were mapped to ST-8 reference genome using HISAT2 (v2.2.1) [57]. Expression abundance was calculated with Rsubread package (v2.16.1) and Trinity (v2.14.0). The

expression data presented in this study are normalized using TPM (transcripts per million). DEGs were called with DEseq2 (FDR < 0.01, $|\log 2 \text{ FC}| > 1$).

Whole Genome Bisulfite Sequencing and data processing

WGBS was performed with two biological replicates. Each biological replicate includes pooled leaf samples from at least three plants.

WGBS libraries were prepared by Novogene (Novogene Co., Ltd.). Briefly, DNA was extracted using the CTAB method. Lamda DNA was used as control. Samples were fragmented using Covaris S220 (Covaris) to 200–400 bps and treated with sodium bisulfite. Libraries were prepared using Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the manufacturer's instructions. Sequencing was performed using Illumina 6000 platform (Illumina).

Reads were mapped to ST-8 reference genome with Bismark (v 0.24.2) [58], which is not aware of SNPs or mismatches. Methylation level was calculated with BatMeth2 [59]. DMR was called with the R package methylKit (v1.24.0) [60] and defined as (meth.diff) > 25% with q value < 0.05. mCHH islands were defined by regions within 3 kb upstream of a TSS, with CHH methylation over 25%. Transposons were annotated with Extensive de-novo TE Annotator (EDTA v2.0.1) [61].

ATAC seq and data processing

ATAC seq was performed with two biological replicates. Each biological replicate includes pooled leaf samples from at least three plants.

ATAC seq libraries were prepared with the help of Biomarker Technologies (Biomarker). Briefly, 50,000 nuclei were treated with Tn5 transposases for 30 min at 37 °C. DNA was purified with the MinElute PCR purification kit (Qiagen) according to the manufacturer's instructions. Libraries were prepared using TruePrep DNA library prep kit V2 for Illumina (Vazyme). Sequencing was performed with the Illumina 6000 platform (Illumina).

ATAC seq data was mapped to ST-8 reference genome with bowtie2 (v2.5.3). Picard (v3.0.0) was used for deduplication. Peaks were called with MACS2 (q value < 0.01).

Crosslinked IPMS

Crosslinked IPMS was performed with two technical replicates. Thirty grams of leaf tissues pooled from at least 10 plants were equally separated into two replicates. For each replicate, 15 g of leaf tissues and resuspended in Nuclei Isolation Buffer (50 mM HEPES, 1 M sucrose, 5 mM KCl, 5 mM MgCl2, 0.6% Triton X-100, 0.4 mM PMSF, 5 mM benzamidine, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)) to reach a final volume of 40 ml and supplemented with 1% formaldehyde for 12 min with rotation. Glycine was added immediately to stop the crosslinking. Clumps were broken by Dounce homogenizer and lysate was filtered through Miracloth and centrifuged at 1500 g for 10 min at 4 °C. Nuclei pellet was resuspended and washed with NRBT buffer (20 mM Tris–HCl pH 7.5, 2.5 mM MgCl2, 25% glycerol, 0.2% Triton X-100) twice and resuspended in 6 ml of RIPA buffer (1 × PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS). Resuspended nuclei were split into 3×2 ml aliquots for sonication for 30 min (30 s on/30 s off) with Bioruptor Plus (Diagenode). Sheared lysate was centrifuged at 8000 g for 15 min at 4 °C and combined supernatant was incubated with the corresponding anti-histone antibodies (same as ChIP) overnight at 4 °C with rotation. ½ IgA and IgG beads were added and incubated at 4 °C with rotation for another 2 h and washed with low salt buffer twice, high salt buffer once, LiCl buffer once similar to the ChIP protocol. Elution was performed with 30 μ l SDS buffer boiled at 90 °C for 5 min. Eluted samples were separated on SDS PAGE gel to remove the heavy and light chains of the antibody.

LC-MS/MS

Peptides were extracted from gel by In-gel digestion. The extracted peptides were dried and then resuspended in 0.1% (v/v) formic acid and then analyzed on the Orbit-rap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to Vanquish neo and FAIMS Pro Interface. Peptides were separated on a 20 cm analytical column packed in-house with the reverse phase material ReproSil-Pur C18–AQ, 1.9 μ m resin (Dr. Maisch, GmbH) with 90 min gradient. Full MS scans were acquired with resolution of 60,000 at m/z 200 in the orbitrap. The HCD fragment ion spectra were acquired in the orbitrap with resolution of 15,000 at m/z 200. The following conditions were used: spray voltage of 2.2 kV, scan range of 350–1800 m/z, ion transfer tube temperature of 320 °C, RF lens of 40%, intensity threshold of 5.0e4. dynamic exclusion durations of 20 s, HCD collision energy of 30%. FAIMS mode was set on standard resolution with -45 V and -65 V voltages.

Database searching

The MS data were analyzed using Proteome Discoverer software (Thermo Fisher Scientific), version 2.4.0.305. Proteins were identified by searching MS and MS/MS data of peptides against the CM334 proteome using the Sequest HT search engine. The parameters used for data analysis included trypsin as the protease with a maximum of two missed cleavages allowed. Carbamidomethylation of cysteine was specified as fixed modification. Oxidation of methionine was specified as variable modifications. The minimum peptide length was specified to be 6 amino acids. The mass error was set to 10 ppm for precursor ions and 0.05 Da for fragment ions.

XLIPMS data analysis

All samples were compared with the no antibody control (CT). Proteins with at least twofold more unique peptides than the CTs were kept for the subsequent analysis. Proteins located in the mitochondria, chloroplast or vesicles were excluded from the analysis. The protein–protein-interaction network was constructed by Gephi (v0.9.2) and layout by the Yifan Hu's algorithm with default settings.

HiC analysis

The HiC data was adopted from Liao et al. performed in pepper leaf tissues [9]. Reads were processed using HiC-Pro (v3.1.0) using the parameter "MIN_MAPQ = 10" and mapped to the reference genome (Zhangshugang) to generate all valid read pairs and interaction matrices [10, 62]. To detect the loop interactions, the hicpro2juice box script from HiC-Pro toolbox was first applied to transfer the.allvalidpairs files to.hic files. Then hicConvertFormat from HiCExplorer (version 3.7.3) was applied to transfer the.hic files

to.cool files and finally hicDetectLoops from HiCExplorer was used to detect the loop interactions with the parameter "–maxLoopDistance 20,000,000".

Nucleotide diversity analysis

VCFtools [63] was used to calculate the nucleotide diversity. Bin size is 500 bp without stepping. The pepper population vcf files were downloaded from http://ted.bti.cornell. edu/pepper. Genomic sequences from one clade were merged into one vcf file. A total of 499 varieties were included in the analysis [10].

No other scripts and software were used other than those mentioned in the "Methods" section.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03595-6.

Additional file 1: Supplementary Figures S1-S10. Fig. S1 Distribution of DNA methylation in the functional genome. Fig. S2 Epigenetic profiles of genes ranked by the expression levels in leaf. Fig. S3 Identification and characterization of mCHH islands. Fig. S4 Characterization of H3K4me1 distal intergenic peaks. Fig. S5 Validation of the promoter and enhancer activities with luciferase reporter assay. Fig. S6 Target gene expression ranked by H3K27ac, chromatin accessibility and H3K4me3 over promoters and enhancers. Fig. S7 Identification of non-specific CREs. Fig. S8 Validation of stress responsive CREs. Fig. S9 Identification of the genomic variations between PBC688 and ST-8 in the CMV resistant regions. Fig. S10 CMV-heat stress crosstalk.

Additional file 2: Table S1 Proteome clusters.

Additional file 3: Table S2 Tissue specific CREs.

Additional file 4: Table S3 Heat responsive CREs.

Additional file 5: Table S4 CMV responsive CREs.

Additional file 6: Table S5 PBC688 genetic variations-InDels.

Additional file 7: Table S6 PBC688 genetic variations-SNPs.

Acknowledgements

We thank Prof. Qian Zhou (Hunan Agricultural University) for providing the CMV_{ENY} strain.

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

Y.X., Z.Z. and F.L. designed this study. H.Y., G.Y., Z.L., T.L., X.W. and M.W. performed the ChIP seq, ATAC seq, RNA seq, stress treatment and validation of the CREs by luciferase reporter assay. H.Y., T.L. performed bioinformatic analysis. Y.F. and G.Q. performed IPMS. Z.Z. and H.H. guided the bioinformatic analysis. Y.X. wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the Taishan Scholars Program of Shandong Province.

Data availability

All high-throughput sequencing data generated in this study are accessible at Genome Sequence Archive (GSA) in the National Genomics Data Center (NGDC) with the access number: CRA018207 [64]. The proteomics data is available at NGDC with the access number: OMIX008621 [65]. H3K27me3, H3K9me2 ChIP-seq and HiC data are accessible with the accession number CNP0001129 at China National GeneBank Database [9]. VCF data adopted in population genetic studies and the raw genome sequencing reads were from the study by Liu et al. [10]. These data are accessible from the Pepper Genomics Database (http://ted.bti.cornell.edu/ftp//pepper) [10] and with the accession number PRJNA800056 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA800056) in NCBI [10].

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹State Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China. ²School of Plant Sciences and Food Security, Tel-Aviv University, Tel-Aviv, Israel. ³Engineering Research Center for Germplasm Innovation and New Varieties Breeding of Horticultural Crops, Key Laboratory for Vegetable Biology of Hunan Province, College of Horticulture, Hunan Agricultural University, Hunan, China. ⁴Yuelushan Laboratory, Hunan, China. ⁵Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China. ⁶State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China.

Received: 5 October 2024 Accepted: 28 April 2025 Published online: 08 May 2025

References

- 1. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet. 2014;46:270–8.
- Zhu Y, Qian W, Hua J. Temperature modulates plant defense responses through NB-LRR proteins. PLoS Pathog. 2010;6: e1000844.
- Kiraly L, Hafez YM, Fodor J, Kiraly Z. Suppression of tobacco mosaic virus-induced hypersensitive-type necrotization in tobacco at high temperature is associated with downregulation of NADPH oxidase and superoxide and stimulation of dehydroascorbate reductase. J Gen Virol. 2008;89:799–808.
- Whitham S, McCormick S, Baker B. The N gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc Natl Acad Sci U S A. 1996;93:8776–81.
- Kim Y, Kim YJ, Paek KH. Temperature-specific vsiRNA confers RNAi-mediated viral resistance at elevated temperature in Capsicum annuum. J Exp Bot. 2021;72:1432–48.
- Li N, Yu C, Yin Y, Gao S, Wang F, Jiao C, Yao M. Pepper crop improvement against Cucumber Mosaic Virus (CMV): a review. Front Plant Sci. 2020;11: 598798.
- Chen W, Wang X, Sun J, Wang X, Zhu Z, Ayhan DH, Yi S, Yan M, Zhang L, Meng T, et al. Two telomere-to-telomere gapless genomes reveal insights into Capsicum evolution and capsaicinoid biosynthesis. Nat Commun. 2024;15:4295.
- Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. 2017;18:210.
- 9. Liao Y, Wang J, Zhu Z, Liu Y, Chen J, Zhou Y, Liu F, Lei J, Gaut BS, Cao B, et al. The 3D architecture of the pepper genome and its relationship to function and evolution. Nat Commun. 2022;13:3479.
- Liu F, Zhao J, Sun H, Xiong C, Sun X, Wang X, Wang Z, Jarret R, Wang J, Tang B, et al. Genomes of cultivated and wild Capsicum species provide insights into pepper domestication and population differentiation. Nat Commun. 2023;14:5487.
- 11. Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. Genome Biol. 2021;22:108.
- 12. Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos Trans R Soc Lond B Biol Sci. 2020;375:20190347.
- Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020;21:71–87.
- Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134:25–36.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet. 2011;43:1160–3.
- 16. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB. Tanksley SD: fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science. 2000;289:85–8.
- Cong B, Liu J, Tanksley SD. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci U S A. 2002;99:13606–11.
- Rodriguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell. 2017;171(470–480):e478.
- Liu L, Gallagher J, Arevalo ED, Chen R, Skopelitis T, Wu Q, Bartlett M, Jackson D. Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize CLE genes. Nat Plants. 2021;7:287–94.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39:311–8.
- Wang M, Li Z, Zhang Y, Zhang Y, Xie Y, Ye L, Zhuang Y, Lin K, Zhao F, Guo J, et al. An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. Plant Cell. 2021;33:865–81.
- Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, et al. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. Genome Biol. 2019;20:139.

- 23. Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li C, Li Y, Buckler ES. Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc Natl Acad Sci U S A. 2015;112:3823–8.
- 24. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J. High-resolution mapping of open chromatin in the rice genome. Genome Res. 2012;22:151–62.
- Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, Kaufmann K. Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nat Commun. 2019;10:1705.
- Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CD, Hirsch CN, Dawe RK, Madzima TF, McGinnis KM, et al. RNAdirected DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc Natl Acad Sci U S A. 2015;112:14728–33.
- 27. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010;328:916–9.
- 28. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 2013;23:628–37.
- 29. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215–6.
- Oka R, Zicola J, Weber B, Anderson SN, Hodgman C, Gent JI, Wesselink JJ, Springer NM, Hoefsloot HCJ, Turck F, Stam M. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. 2017;18:137.
- Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian MII3/MII4. Genes Dev. 2012;26:2604–20.
- 32. Yu Y, Li X, Jiao R, Lu Y, Jiang X, Li X. H3K27me3-H3K4me1 transition at bivalent promoters instructs lineage specification in development. Cell Biosci. 2023;13:66.
- 33. Wang Z, Ren B. Role of H3K4 monomethylation in gene regulation. Curr Opin Genet Dev. 2024;84: 102153.
- Monahan K, Horta A, Lomvardas S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature. 2019;565:448–53.
- 35. Lin X, Liu Y, Liu S, Zhu X, Wu L, Zhu Y, Zhao D, Xu X, Chemparathy A, Wang H, et al. Nested epistasis enhancer networks for robust genome regulation. Science. 2022;377:1077–85.
- Long HK, Osterwalder M, Welsh IC, Hansen K, Davies JOJ, Liu YE, Koska M, Adams AT, Aho R, Arora N, et al. Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. Cell Stem Cell. 2020;27(765– 783): e714.
- Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A. 2002;99:7548–53.
- Symmons O, Uslu W, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014;24:390–400.
- 39. Kim S, Wysocka J. Deciphering the multi-scale, quantitative cis-regulatory code. Mol Cell. 2023;83:373–92.
- 40. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, Jackson D, Leith A, Schreiber J, Noble WS, et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell. 2019;176:1516.
- 41. Zuin J, Roth G, Zhan Y, Cramard J, Redolfi J, Piskadlo E, Mach P, Kryzhanovska M, Tihanyi G, Kohler H, et al. Nonlinear control of transcription through enhancer-promoter interactions. Nature. 2022;604:571–7.
- 42. Shenasa H, Bentley DL. Pre-mRNA splicing and its cotranscriptional connections. Trends Genet. 2023;39:672–85.
- Rajametov SN, Yang EY, Cho MC, Chae SY, Jeong HB, Chae WB. Heat-tolerant hot pepper exhibits constant photosynthesis via increased transpiration rate, high proline content and fast recovery in heat stress condition. Sci Rep. 2021;11:14328.
- 44. Shen X, De Jonge J, Forsberg SK, Pettersson ME, Sheng Z, Hennig L, Carlborg O. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLoS Genet. 2014;10: e1004842.
- Naydenov M, Baev V, Apostolova E, Gospodinova N, Sablok G, Gozmanova M, Yahubyan G. High-temperature effect on genes engaged in DNA methylation and affected by DNA methylation in Arabidopsis. Plant Physiol Biochem. 2015;87:102–8.
- 46. Erdmann RM, Picard CL. RNA-directed DNA Methylation. PLoS Genet. 2020;16: e1009034.
- 47. Popova OV, Dinh HQ, Aufsatz W, Jonak C. The RdDM pathway is required for basal heat tolerance in Arabidopsis. Mol Plant. 2013;6:396–410.
- Zhao J, Lu Z, Wang L, Jin B. Plant Responses to Heat Stress: Physiology, Transcription, Noncoding RNAs, and Epigenetics. Int J Mol Sci. 2020;22:117.
- 49. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21 29 21-21 29 29.
- 50. Guo G, Wang S, Liu J, Pan B, Diao W, Ge W, Gao C, Snyder JC. Rapid identification of QTLs underlying resistance to Cucumber mosaic virus in pepper (Capsicum frutescens). Theor Appl Genet. 2017;130:41–52.
- Carter AH, Chen XM, Garland-Campbell K, Kidwell KK. Identifying QTL for high-temperature adult-plant resistance to stripe rust (Puccinia striiformis f. sp. tritici) in the spring wheat (Triticum aestivum L.) cultivar 'Louise. Theor Appl Genet. 2009;119:1119–28.
- 52. Cheng C, Gao X, Feng B, Sheen J, Shan L, He P. Plant immune response to pathogens differs with changing temperatures. Nat Commun. 2013;4:2530.
- 53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
- 54. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9: R137.
- 55. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3.
- Wang Q, Li M, Wu T, Zhan L, Li L, Chen M, Xie W, Xie Z, Hu E, Xu S, Yu G. Exploring Epigenomic Datasets by ChIPseeker. Curr Protoc. 2022;2: e585.

- 57. 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.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.
- 59. Zhou Q, Lim JQ, Sung WK, Li G. An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping. BMC Bioinformatics. 2019;20:47.
- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13: R87.
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275.
- 62. 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.
- 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.
- 64. Yang H, Yu G, Lv Z, Li T, Wang X, Fu Y, Zhu Z, Guo G, He H, Wang M, et al: Epigenome profiling reveals distinctive regulatory features and cis-regulatory elements in pepper. CRA018207. 2025. https://ngdc.cncb.ac.cn/gsa/browse/CRA018207. Released Apr 2025.
- 65. Yang H, Yu G, Lv Z, Li T, Wang X, Fu Y, Zhu Z, Guo G, He H, Wang M, et al: Epigenome profiling reveals distinctive regulatory features and cis-regulatory elements in pepper. OMIX008621. 2025. https://ngdc.cncb.ac.cn/omix/relea se/OMIX008621. Released Apr 2025.

Publisher's Note

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