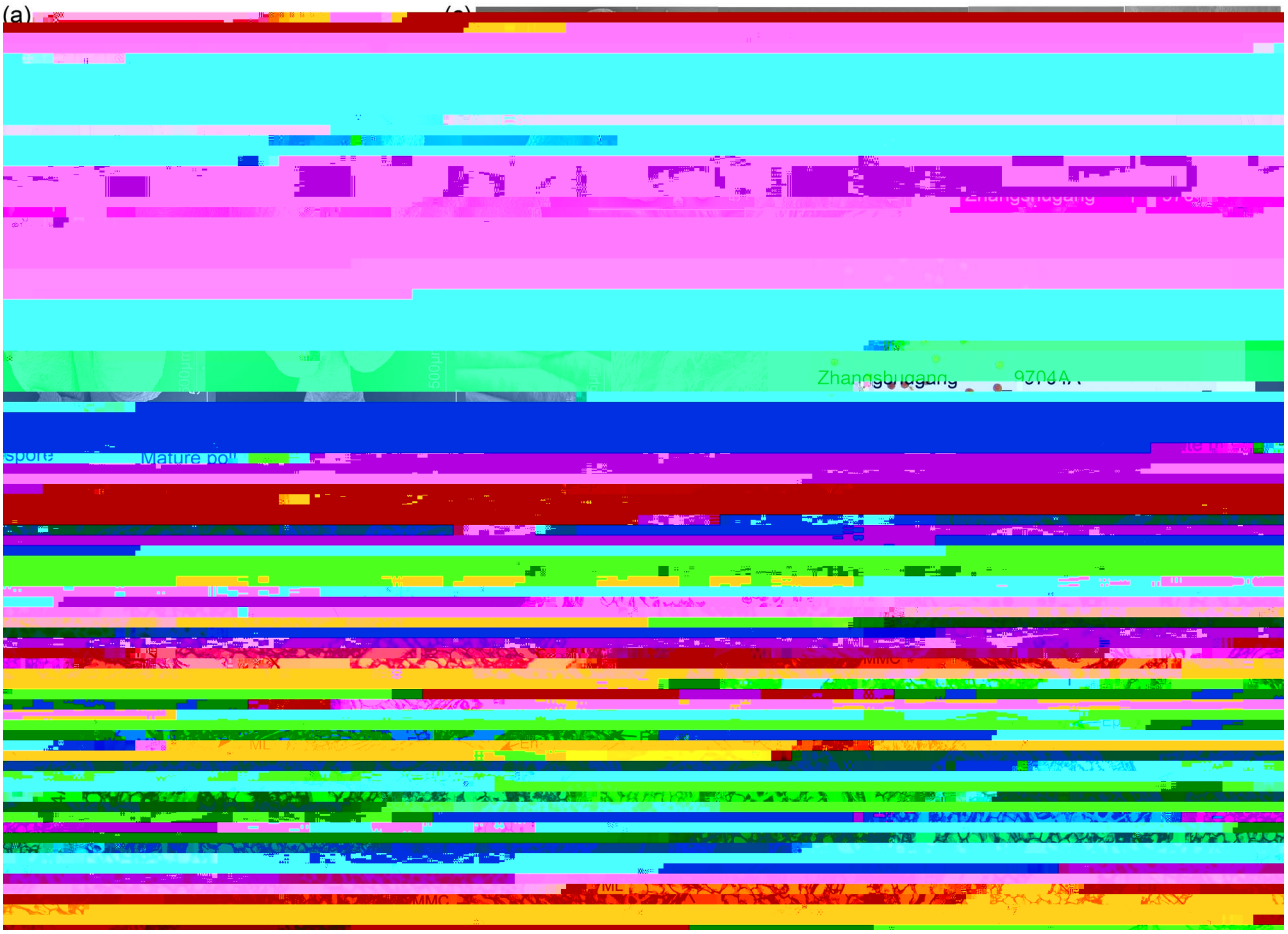

Figure 1. Phenotypes of Zhangshugang and 9704A. a Mature anther phenotypes of Zhangshugang and 9704A. Scale bars, 1 mm. b Pollen grains of Zhangshugang and 9704A stained with I_2-KI. Scale bars, 50 μm; c Scanning electron microscopic analysis of anthers and pollen grains from Zhangshugang and 9704A at the mature pollen stage. d Sectional observation of anthers from Zhangshugang and 9704A at five different development stages. Scale bars, 10 μm. Sp, sporogenous cells; Ep, epidermis; En, endothecium; ML, middle layer; T, tapetum; MMC, microspore mother cells; Tds, tetrads; dTds, death tetrads; Msp, microspores; dTM, dead tapetum and microspores; PG, pollen grains.

maize, highlighting their classification within the PPR protein family. This analysis demonstrated that the protein sequences of *C* are closely related to those in maize (Fig. 3b).

Gene expression analysis and subcellular localization a strong candidate male fertility restoration gene

To further identify candidate *C* , we conducted transcriptome analysis of flower buds at the tetrad stage from Zhangshugang and 9704A using RNA sequencing technology. Differential expression analysis of the entire pepper gene repertoire between Zhangshugang and 9704A revealed 2048 differentially expressed genes (DEGs). Among them, 1004 were up-regulated, and 1044 were down-regulated in Zhangshugang compared with 9704A (Fig. 4a). We examined the distribution of these DEGs across various chromosomes, noting Chr1 had the highest number and Chr11 the lowest. Gene Ontology (GO) functional annotation and enrichment analyses indicated that the majority of these DEGs were enriched in terms associated with CMS, including hormone level regulation, oxidoreductase activity, and cell wall macromolecular metabolism. CMS can arise from disruptions in nutrient synthesis and metabolism, energy metabolism, reactive oxygen species metabolism, and hormone synthesis and metabolism. GO enrichment analysis provided insight into the specific biological processes influencing pollen fertility and highlighted key DEGs responsible for 9704A's sterile phenotype (Fig. 4b). Kyoto Encyclopedia of Genes and Genomes (KEGG)

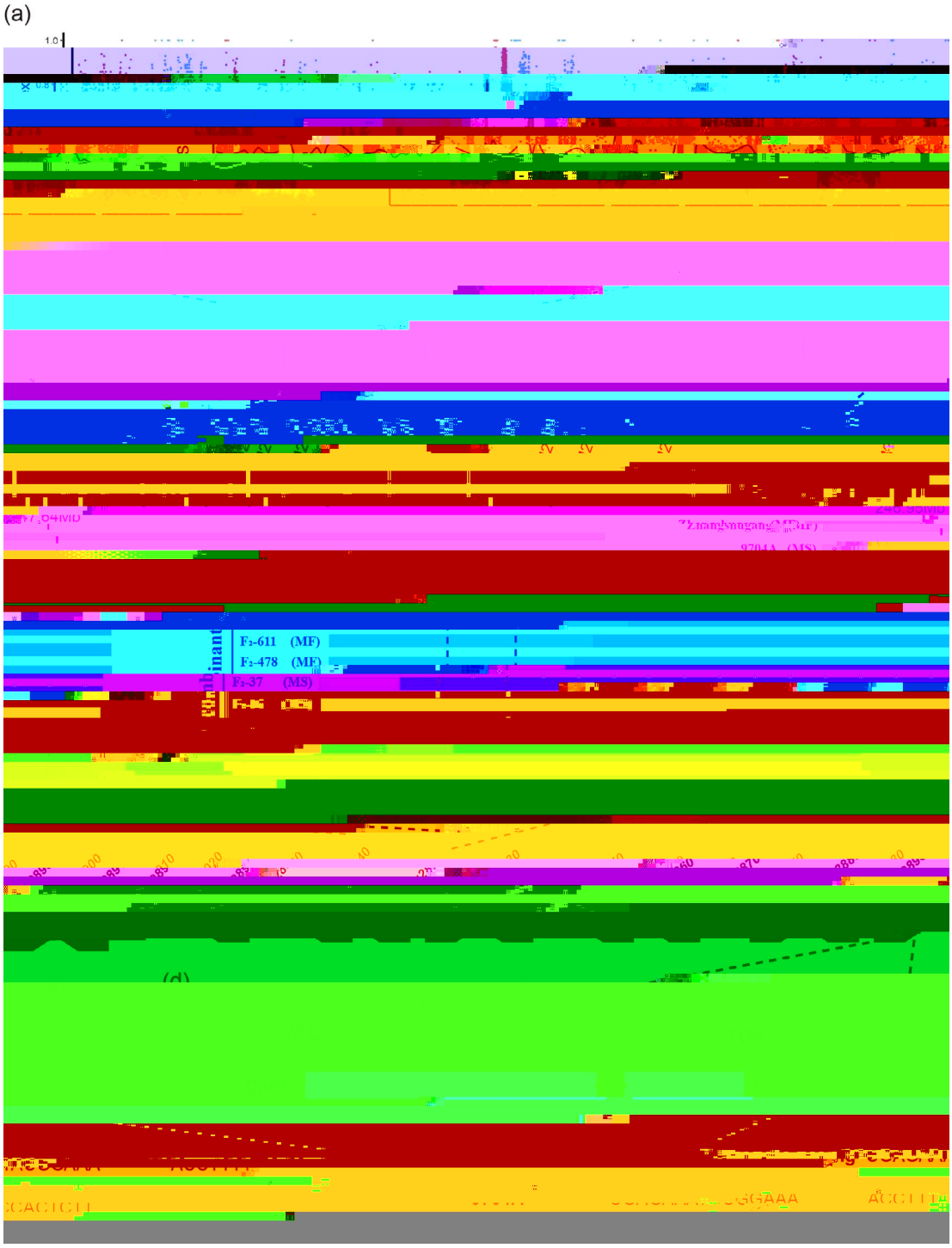

Figure 2. Mapping of the C gene. a Distribution of SNP index on chromosomes. b Fine mapping of C locus. MS, male-sterile; MF, male-fertile; c Genes within the candidate interval. Genes on the negative and positive strands are indicated on the left and right, respectively. d Predicted structure of the CaRf-coding protein sequence and alignment of the CaRf-coding protein sequence in Zhangshugang and 9704A.

Table 1. Statistics of DEGs in the candidate interval.

Gene ID	Annotation	$\log_2 F$ (MF/MS)	*P*-value
C 06 28850	Protein NTM1-like 9	3.01	1.54E−04
C 06 28880	Putative late blight resistance protein homolog R1A-3	2.04	2.24E−06
C 06 28920	Pentatricopeptide repeat protein	4.01	5.26E−11

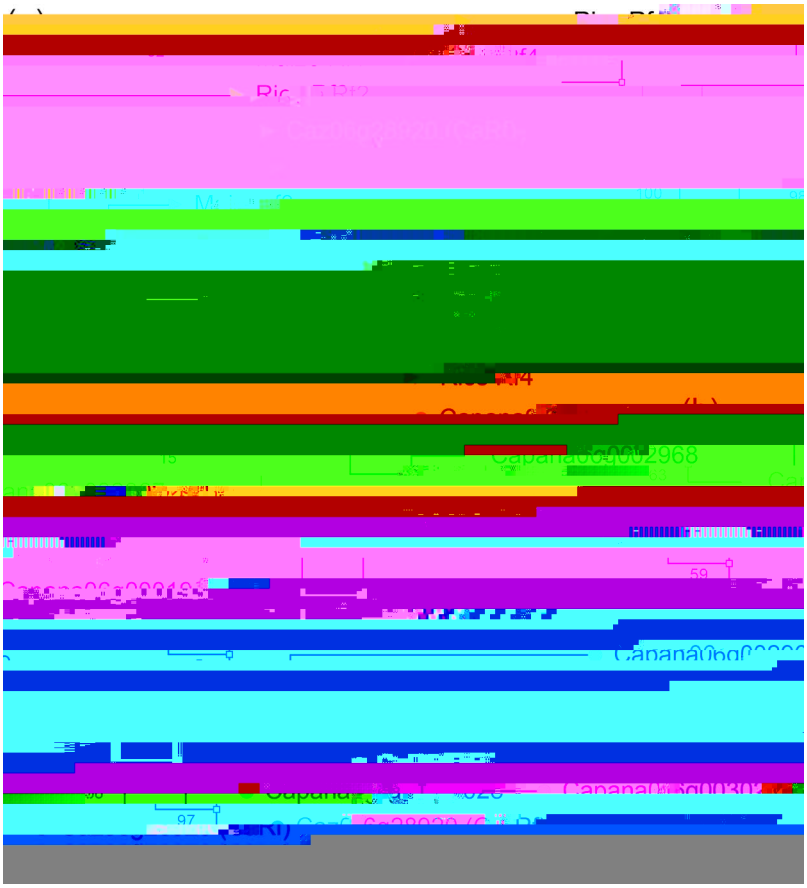

Figure 3. Candidate genes reported in pepper and cloned genes in rice and maize. a Sequence similarity between *C* and cloned genes in rice and maize. b Sequence similarity between *C* and previous candidate genes in pepper.

The Zhangshugang genome annotation placed this gene on the negative strand of chromosome 6 (Chr06:247554451–247574842), spanning a complex structure with a length of 20 392 bp (Supplementary Data Fig. S3). Analysis based on the typical characteristics of the PPR protein family, which lacks introns and exhibits highly conserved domains [48], raised doubts about the validity of *C 06 28910* as a genuine gene. Global alignment with the gapless genome CaT2T [41] revealed that *C 06 28910* aligned with seven genes in the CaT2T genome (Supplementary Data Fig. S4), none of which matched a PPR protein gene of similar length in the CaT2T genome annotation. This discrepancy suggests that *C 06 28910* might be an annotation error in the Zhangshugang genome, rectified in the current CaT2T genome annotation.

To verify the subcellular localization of the remaining two PPR genes within the interval, we constructed pCAMBIA1300 vectors expressing Caz06g28920-GFP and Caz06g28930-GFP fusion proteins under the control of the 35S promoter. These constructs were introduced into rice protoplasts along with mitochondria-specific marker Mito-Tracker Red CMXRos-mCherry (RFP) or with nucleus-specific marker mCherry-fused Ghd7. As shown in Fig. 5c, the fluorescent microscopy results confirmed that Caz06g28920-GFP was co-localized well with the mitochondria431.1(m).4(itoc)11.3(2Vtoc)1125fMito-T7723.5(r)11.8(4(itoc)3(2Vtoc)112r)-5R0.8(iear expresnotati35ozote33y adati35ubcellu33d l6.7(localiz4(the)2592.1(r)11.8(esu33d)-74t.4(icr)11.8(ng4(ighl)11ion.)]TJ0-1.4040k.0001sc[ortidate)Tj/F74

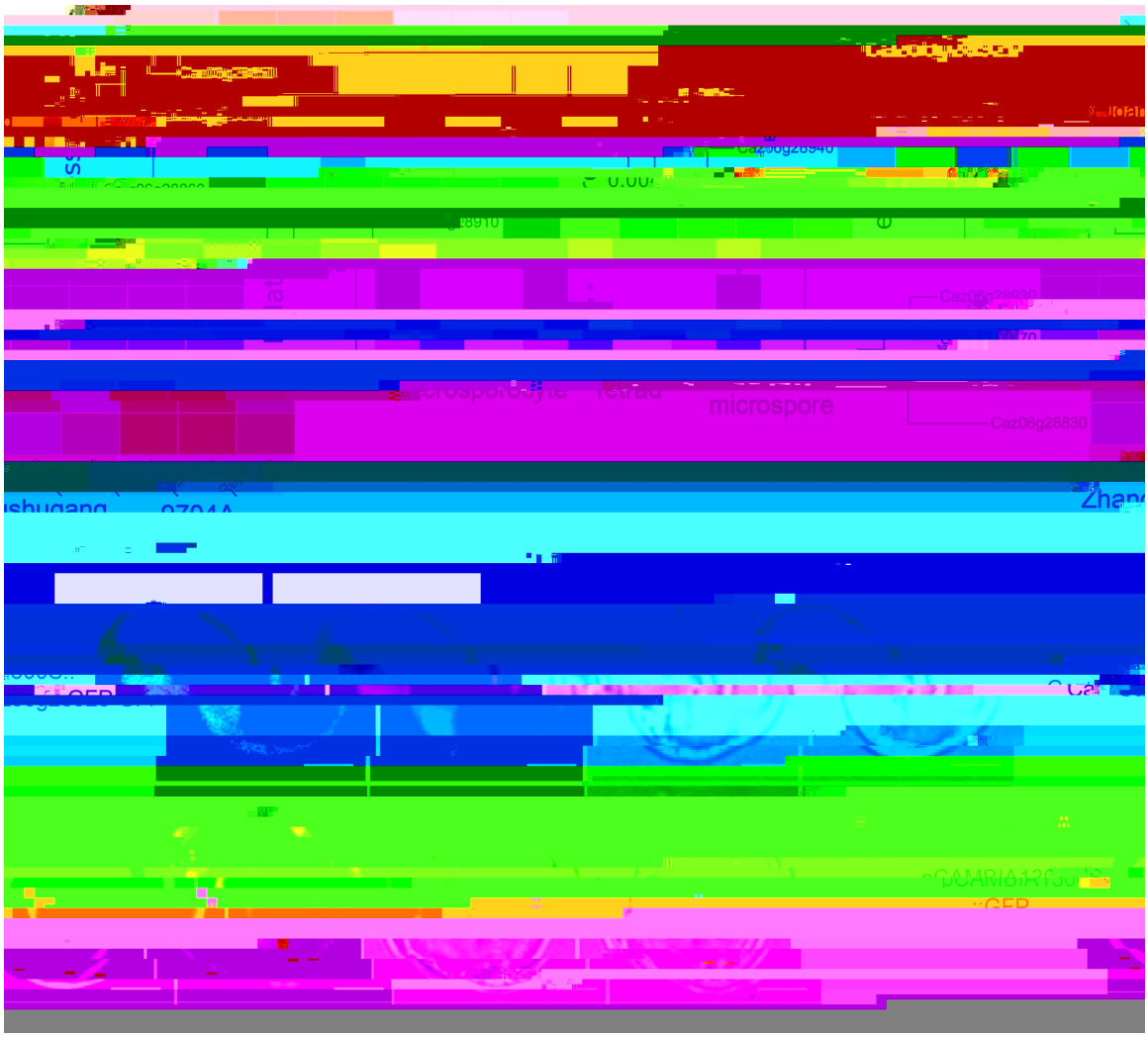

Figure 5. Gene expression and subcellular localization analysis. a Expression heat map of genes in the candidate interval at the tetrad stage of bud development in 9704A and Zhangshugang. b Relative *C 06 28920* expression levels at different stages of flower bud development: microsporocyte, tetrad, and uninucleate microspore stages, analyzed using quantitative RT–PCR. c Subcellular localization of Caz06g28920 in rice protoplasts: co-localization of Caz06g28920-GFP and Mito-Tracker Red CMXRos-RFP in the mitochondrion. $*$ < 0.05, $**$ < 0.01. GFP, green fluorescent protein; RFP, Mito-Tracker Red CMXRos-RFP (red). Scale bars, 10 μm.

distribution across the chromosome (Fig. 7a). Statistical details of SNP positions are cataloged in Supplementary Data Table S7. Mapping these SNPs to the pepper Zhangshugang genome assembly revealed an average inter-SNP distance of 60 kb. Analysis of SNPs within PepperSNP50K yielded an average PIC of 0.30, crucial for discerning genetic variations (Supplementary Data Table S2). The distribution of PIC values is illustrated in Fig. 7b, underscoring the array's utility in diverse breeding contexts.

Within PepperSNP50K, 21.3% of SNPs were located within the gene region, encompassing the upstream and downstream 2 kb of the gene, UTRs, exons, and introns (Fig. 7c). These SNPs hold significant potential to influence gene function and are usually more informative. While this percentage is lower compared with 68% in RiceSNP50 [42], SNPs in PepperSNP50K are evenly distributed across chromosomes. This distribution ratio is optimal considering the larger size of the pepper reference genome, which is 7–8 times that of the rice reference genome, and the relatively smaller proportion of gene regions in peppers. To ensure

co-b1.5(2r)11.ysualld(e)12.1tle2r4(3()-.5liqu9d-pha--5029).1((.)]TJETEMC-r)11.8(eg).9(y's)-278.1(s)-401.8iation9t8374.5(PIy)-2321.8(o).5(n22e)-3enabl8i

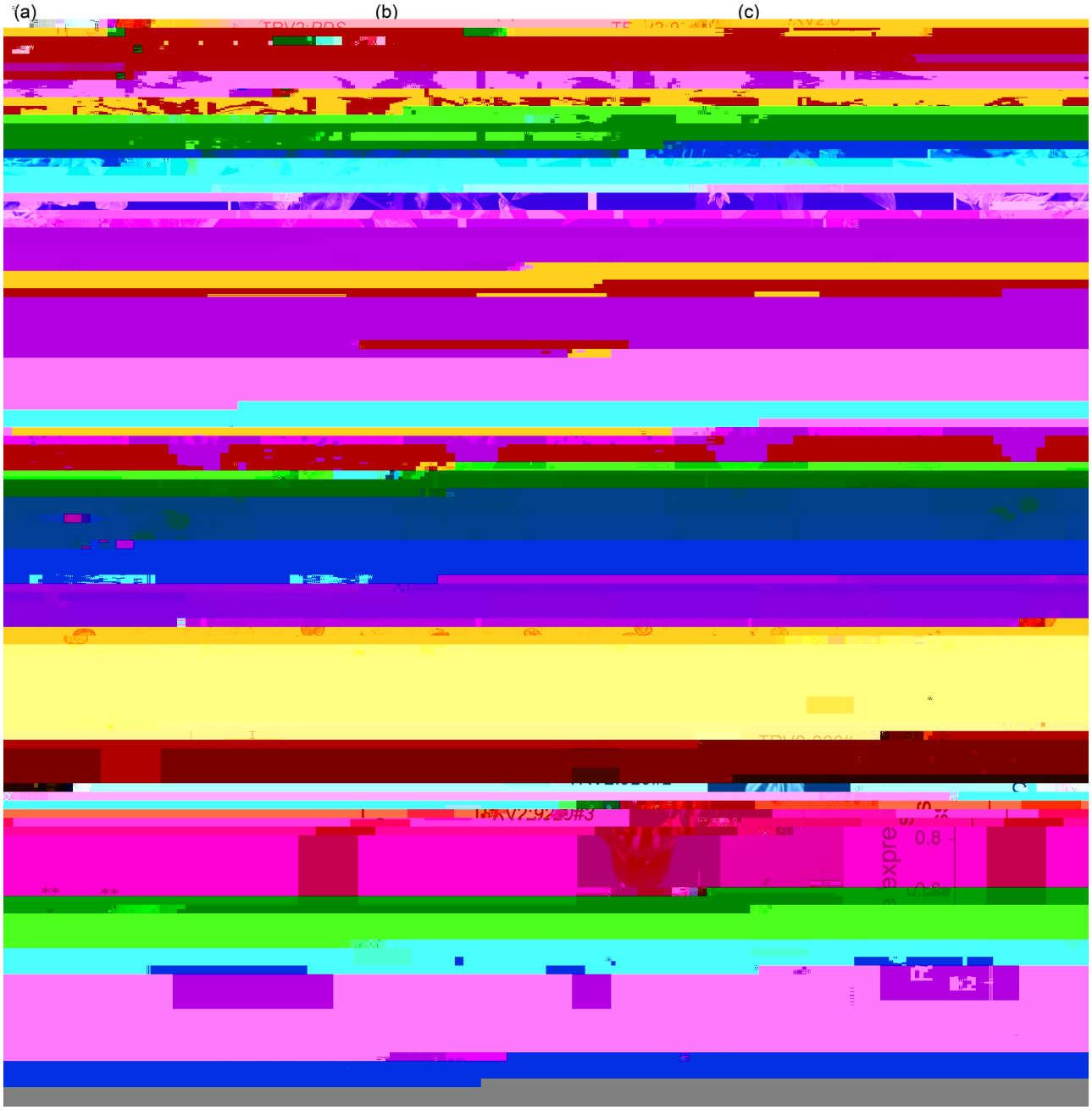

Figure 6. Functional verification of *C 06 28920* through VIGS in F_1 (9704A × Zhangshugang) plants. a–c Phenotypic comparison of TRV2:*0* (empty vector control), TRV2: *D* (phytoene desaturase), and TRV2:*920* plants. Scale bars: 1.0 cm (a, b) and 0.5 cm (c). d, e Anther dehiscence and pollen grains of TRV2:*0* and TRV2: *D* plants in F_1 generation. Scale bars: 1 mm. f–h Anther dehiscence and pollen grains of *C 06 28920* down-regulated TRV2:*920#1*, TRV2:*920#2*, and TRV2:*920#3* plants in F_1 generation. Scale bars for dehisced anthers and pollen grains: 1 mm. i *C 06 28920* expression levels in virus-induced F_1 plants; gene expression was analyzed using qRT–PCR on flower buds at tetrad and uninucleate microspore stages. Relative expression levels in TRV2:*920#1*, TRV2:*920#2*, and TRV2:*920#3* plants were compared with TRV2:*0*. ∗ <0.05, ∗∗ <0.01 (-test). Relative expression levels are presented as mean ± standard deviation, =3.

due to their superior traits. To improve the efficiency of crossing between XY21 and QN49, we employed the three-line system involving 9704A/*C* to create a new sterile line, XY21A(( )). During the development of the male QN49R restorer line, we utilized the.4054-95.AiC5V9(T)-.9(R)74.3(V)-.9(2:)]TJ/F71Tf15.57540(b)).

Table 2. Agronomic performance of various varieties.

Variety	Type	Germination rate (%)	First flower node (cm)	Plant height(cm)	Effective number of branches	Single fruit quality (g)	Longitudinal diameter of fruit (cm)	Fruit cross diameter (cm)	Fruit type index	Pulp thickness (cm)	Yield per plant (g)	Capsanthin content
XY21	Maintainer line	95 ± 0.8	12 ± 0.5	65 ± 3.3	10 ± 1.5	22 ± 1.1	13.5 ± 0.5	2.5 ± 0.1	5.4 ± 0.5	0.3 ± 0.04	500 ± 10	21 ± 1.3
XY21A	Male-sterile line	95 ± 0.6	12 ± 0.8	65 ± 4.2	10 ± 1.5							
QC65	Restorer line	95 ± 0.9	15 ± 0.6	62 ± 2.7	10 ± 1	5 ± 0.3	5.1 ± 0.2	1 ± 0.1	5.1 ± 0.5	0.15 ± 0.01	150 ± 6.3	10 ± 0.6
QN49	Maintainer line	95 ± 1.1	15 ± 0.7	85 ± 3.6	7 ± 1.5	35 ± 2.5	18 ± 1.5	2.5 ± 0.1	7.4 ± 0.5	0.3 ± 0.03	585 ± 12.6	24 ± 1.1
QN49R	Restorer line	95 ± 1.0	15 ± 0.5	85 ± 3.1	7 ± 1.5	36 ± 1.8	19 ± 1.2	2.5 ± 0.2	7.6 ± 0.6	0.3 ± 0.04	590 ± 15.1	25 ± 1.5
F_1		95 ± 10.8	14 ± 0.6	77 ± 4.1	9 ± 1	33 ± 1.2	16 ± 0.8	3 ± 0.2	5.3 ± 0.5	0.3 ± 0.03	610 ± 12.3	17.51 ± 1.350 T295

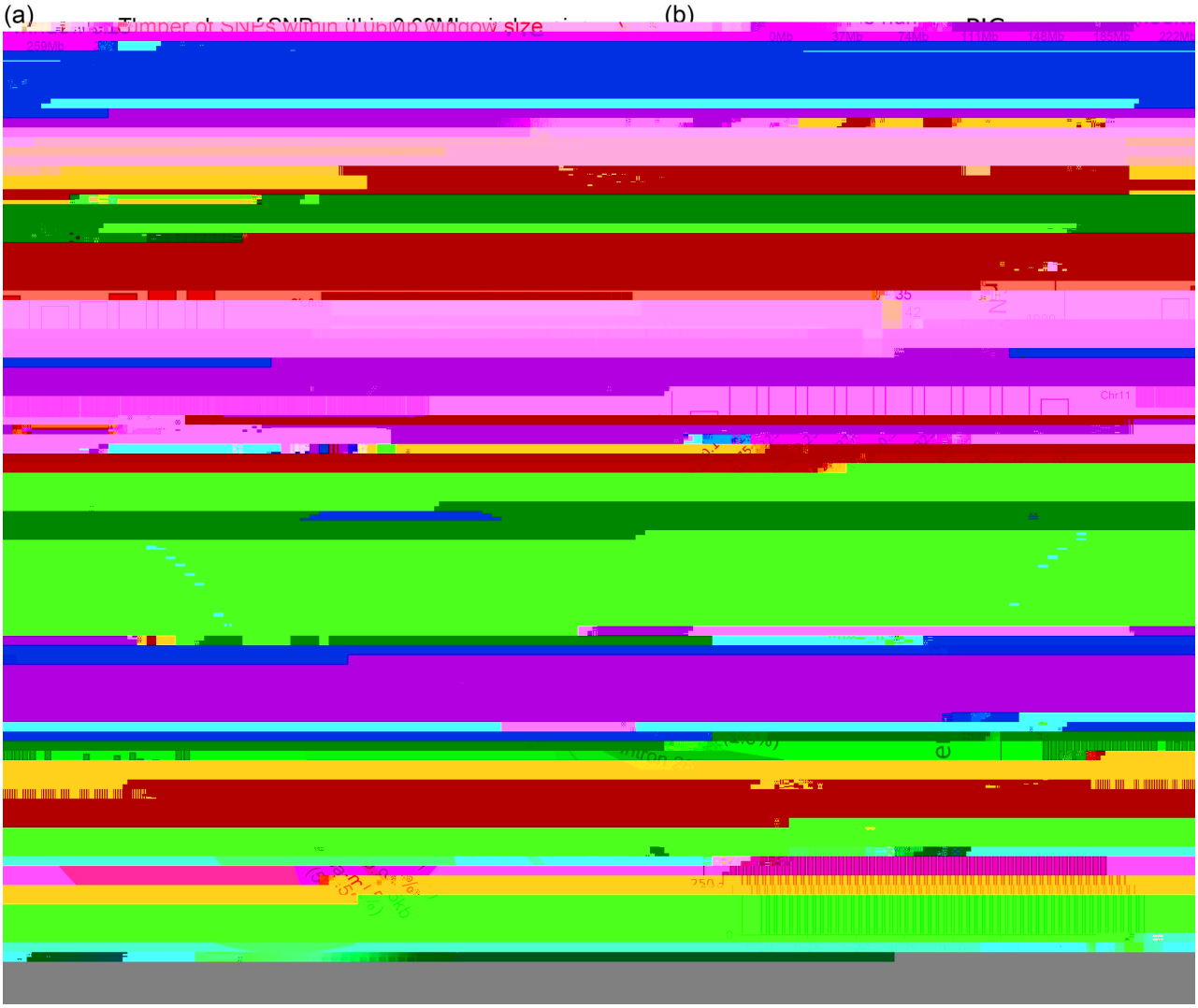

Figure 7. Design and characteristics of PepperSNP50K. a Distribution of SNPs across the entire pepper genome. b PICs of the selected SNPs. c Localization of selected SNPs within gene regions and across chromosomes. d MAFs of the selected SNPs.

illustrating that the mechanism of these F_1 hybrids can efficiently verify the reliability of restorer genes (Fig. 9). In our VIGS experiment, we silenced the expression of the putative restorer gene in the nucleus in the F_1 hybrid (9704A × Zhangshugang) (Fig. 9). Compared with control plants, the silenced plants exhibited a substantial reduction in pollen viability and a high proportion of malformed pollen grains. These findings strongly support the notion that *C* indeed functions as a credible restorer gene (Fig. 6).

There are missense mutations in the *C 06 28920* gene in the restorer line Zhangshugang and the sterile line 9704A, which may lead to decreased expression of the *C 06 28920* gene in the sterile line 9704A or differences in protein function. However, the causes of Caz06g28920 protein dysfunction in 9704A remain unclear. We conducted a VIGS experiment targeting the *C 06 28920* gene in the first-generation hybrid of sterile line 9704A and restorer line Zhangshugang. The results showed that plants with effective silencing of *C 06 28920* showed obvious pollen growth deformities and reduced pollen vitality compared with controls, suggesting that silencing *C 06 28920* interrupts fertility recovery in the first-generation hybrid. Nevertheless, the specific underlying mechanism is complex and requires further elucidation. Future efforts will focus on creating a stable transgenic pepper strain carrying the *C 06 28920* gene and identifying the sterility gene in sterile line 9704A to facilitate deeper molecular analysis.

Molecular marker-assisted selection technology significantly accelerates breeding timelines and improves breeding efficiency. The PARMS markers developed in this study, closely linked to the gene, offer substantial benefits for molecular-assisted breed-

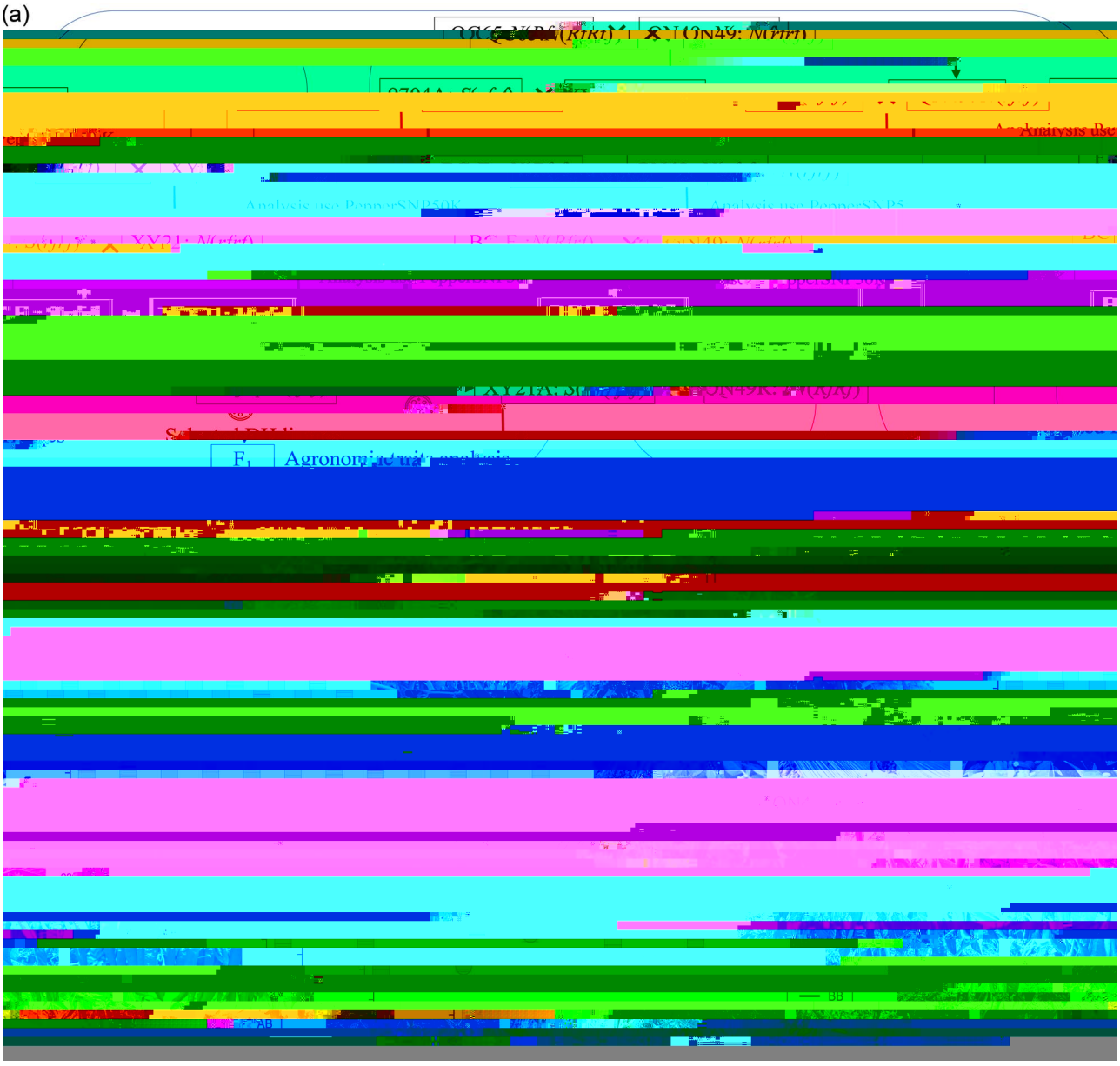

Figure 8. Breeding application of CMS/C system. a Development of CMS hybrids for breeding programs. XY21((

heterozygous genotype loci (Aa), and denotes the number of homozygous genotype loci (aa) identical to the recipient.

Double haploid technical method

The pepper materials were provided by the Pepper Research Group of Hunan Xiangyan Seed Industry. The seedlings were cultivated in a greenhouse with temperature controlled at 26–30°C during the day and 15–20°C at night. Healthy plants were selected during the flowering period, typically when the plants had bloomed to four fruits, harvested between 8 and 10.00 a.m. on sunny days. Buds with petal length equal to the sepals and in the uninucleate stage during the microspore development period were refrigerated at 4°C for 48 h. On a clean bench, the flower buds were surface-disinfected with 70% alcohol for 30 s followed by 5% sodium hypochlorite for 10–12 min and rinsed three times with sterile water. Using tweezers, anthers were carefully extracted from the buds, ensuring complete removal of filaments, and inoculated into the induction medium (NTH basic medium +0.2 mg/l NAA + 1.0 mg/l KT + 30 g/l sucrose +8 g/l agar powder). The culture dishes containing inoculated anthers were kept in the dark at 28°C until embryoids appeared. Embryoids were then transferred to the rooting medium (1/2 MS + 0.1 mg/l NAA + 20 g/l sucrose +8 g/l agar powder) for rooting. Tissue-cultured seedlings were hardened and transplanted into the substrate. After 2 weeks, 0.2% colchicine was applied to induce diploid formation. The diploid seedlings were transplanted into a field greenhouse and managed similarly to field-grown plants. Fruits were harvested upon ripening.

Acknowledgements

This work was supported by the Construction of Innovative Provinces in Hunan Province (Grant No. 2021NK1006), the National Natural Science Foundation of China (Grant No. 32402571), the Hunan Provincial Natural Science Foundation of China (Grant No. 2024JJ6239), the China Postdoctoral Science Foundation (Grant No. 2023M741144), and the Postdoctoral Fellowship Program of CPSF (Grant No. GZC20230777).

Author contributions

T.B.Q. and Y.H.P. contributed to population generation, phenotypic analysis, data curation, marker development, and manuscript writing. C.Y., L.X.M., C.Q.Z., W.M.Q., L.J.Y., X.F.L., X.L.L., C.L., and D.J. participated in the phenotypic identification. Z.X.H., Y.Q.B., and

18. Cui X, Wise RP, Schnable PS. The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. . 1996;272:1334–6
19. Qin X, Tian S, Zhang W. . The main restorer Rf3 of maize S type cytoplasmic male sterility encodes a PPR protein that functions in reduction of the transcripts of orf355. . 2021;14:1961–4
20. Jaqueth JS, Hou Z, Zheng P. . Fertility restoration of maize CMS-C altered by a single amino acid substitution within the Rf4 bHLH transcription factor. . 2020;101:101–11
21. Lin Y, Yang H, Liu H. . A P-type pentatricopeptide repeat protein ZmRF5 promotes 5′ region partial cleavages of atp6c transcripts to restore the fertility of CMS-C maize by recruiting a splicing factor. *B* . 2024;22:1269–81
22. Castillo A, Atienza SG, Martín AC. Fertility of CMS wheat is restored by two Rf loci located on a recombined acrocentric chromosome. *E B* . 2014;65:6667–77
23. Ui H, Sameri M, Pourkheirandish M. . High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley. *A G* . 2015;128:283–90
24. Uyttewaal M, Arnal N, Quadrado M. . Characterization of pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *C* . 2008;20:3331–45
25. Yin J, Guo W, Yang L. . Physical mapping of the Rf1 fertility-restoring gene to a 100 kb region in cotton. *A G* . 2006;112:1318–25
26. Matsuhira H, Kagami H, Kurata M. . Unusual and typical features of a novel restorer-of-fertility gene of sugar beet (*B* L.). *G* . 2012;192:1347–58
27. Zhang H, Wu J, Dai Z. . Allelism analysis of BrRfp locus in different restorer lines and map-based cloning of a fertility restorer gene, BrRfp1, for pol CMS in Chinese cabbage (*B* L.). *A G* . 2017;130:539–47.
28. Koizuka N, Imai R, Fujimoto H. . Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. . 2003;34:407–15
29. Kim S, Kim CW, Park M. . Identification of candidate genes associated with fertility restoration of cytoplasmic male-sterility in onion (*A* L.) using a combination of bulked segregant analysis and RNA-seq. *A G* . 2015;128: 2289–99
30. Zhang Z, Zhu Y, Cao Y. . Fine mapping of the male fertility restoration gene CaRf032 in *C* L. *A G* . 2020;133:1177–87
31. Kim S, Park M, Yeom SI. . Genome sequence of the hot pepper provides insights into the evolution of pungency in *C* species. *G* . 2014;46:270–8
32. Qin C, Yu C, Shen Y. . Whole-genome sequencing of cultivated and wild peppers provides insights into *C* domestication and specialization. *A A*. 2014;111: 5135–40
33. Wu L, Wang P, Wang Y. . Genome-wide correlation of 36 agronomic traits in the 287 pepper (*C*) accessions obtained from the SLAF-seq-based GWAS. *I* . 2019;20: 5675
34. Cheng J, Chen Y, Hu Y. . Fine mapping of restorer-of-fertility

ceu480TD-.00025Tm0.17-1.4616TD-.0001Tc[ialization.7(cad6e2.19888-1.4604TD-.00049;)Tj/F11Tf2.50080.5008CID51>>BDCBT7.52e46P<</MCID11TEMC

57. Qu Y, Fernie AR, Liu J. . Doubled haploid technology and synthetic apomixis: recent advances and applications in future crop breeding. . 2024;17:1005–18
58. Kabade PG, Dixit S, Singh UM. . SpeedFlower: a comprehensive speed breeding protocol for indica and japonica rice. *B* . 2024;22:1051–66
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *B* . 2014;30:2114–20
60. Allen GC, Flores-Vergara MA, Krasynanski S. . A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. . 2006;1:2320–5
61. Tang B, Xie L, Li X. . Novel structural annotation and functional expression analysis of GTP_EFTU conserved genes in pepper based on the PacBio sequencing data. *H* . 2021;7:443–56
62. Wan H, Yuan W, Ruan M. . Identification of reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*C* L.). *B* *B* *C* . 2011;416:24–30
63. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *A* . 2004;32:1792–7
64. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *A* . 2016;44:W242–5
65. Tang H, Krishnakumar V, Zeng X. . JCVI: a versatile toolkit for comparative genomics analysis. . 2024;3:e211
66. Minamikawa T, Sriratana A, Williams DA. . Chloromethyl-X-rosamine (MitoTracker red) photosensitises mitochondria and induces apoptosis in intact human cells. *C* . 1999;112:2419–30
67. Xue W, Xing Y, Weng X. . Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *G* . 2008;40:761–7
68. Nelson BK, Cai X, Nebenführ A. A multicolored set of in vivo organelle markers for co-localization studies in *A* and other plants. . 2007;51:1126–36
69. Chen S, Zhou Y, Chen Y. . Fastp: an ultra-fast all-in-one FASTQ preprocessor. *B* . 2018;34:i884–90
70. Andrews S. *F* *C: A* *C* *H* *D* . 2010.
71. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. . 2015;12:357–60
72. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *G* *B* . 2014;15:550
73. Klopfenstein D V, Zhang L, Pedersen B S. .. GOATOOLS: A Python library for Gene Ontology analyses. . 2018;8:1-17.
74. Aramaki T, Blanc-Mathieu R, Endo H. . KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *B* . 2020;36:2251–2
75. Yu G, Wang LG, Han Y. . clusterProfiler: an R package for comparing biological themes among gene clusters. *IC* . 2012;16:284–7
76. Zhou Y, Deng Y, Liu D. . Promoting virus-induced gene silencing of pepper genes by a heterologous viral silencing suppressor. *B* . 2021;19:2398–400
77. Kendig KI, Baheti S, Bockol MA. . Sentieon DNASeq variant calling workflow demonstrates strong computational performance and accuracy. *F* *G* . 2019;10:736
78. Danecek P, Auton A, Abecasis G. . The variant call format and VCFtools. *B* . 2011;27:2156–8
79. Hill JT, Demarest BL, Bisgrove BW. . MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *G* . 2013;23:687–97