e

gene, 2 1 42 5, that encodes a kinase-related protein and functions in PH regulation. Ma . [1] identified 10 QTLs and a novel gene, ♪ , for PH by linkage mapping in a backcross inbred line (BIL) population of cotton. Currently, the relevant QTLs
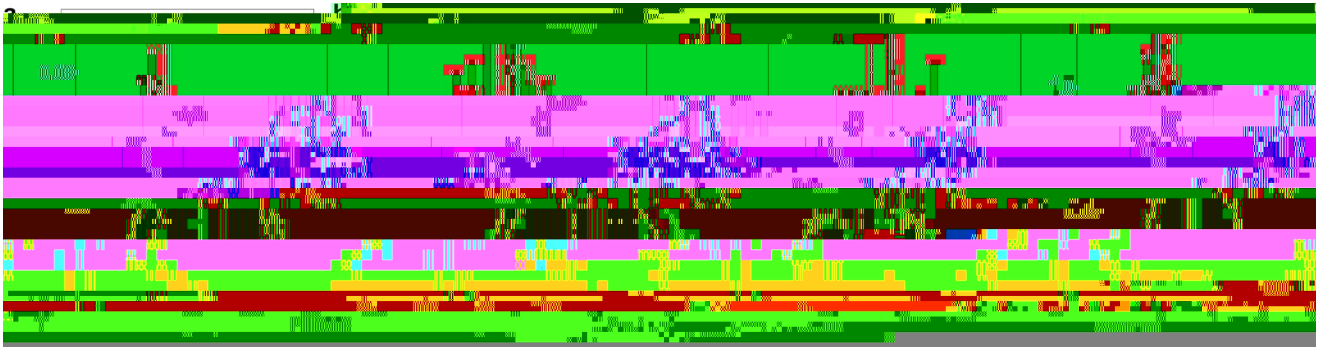
Figure 1. Distribution of PH ( ), IL ( ), IN ( ), and SD ( ) in two environments (E1 and E2). The dashed lines represent the mean values of the traits.
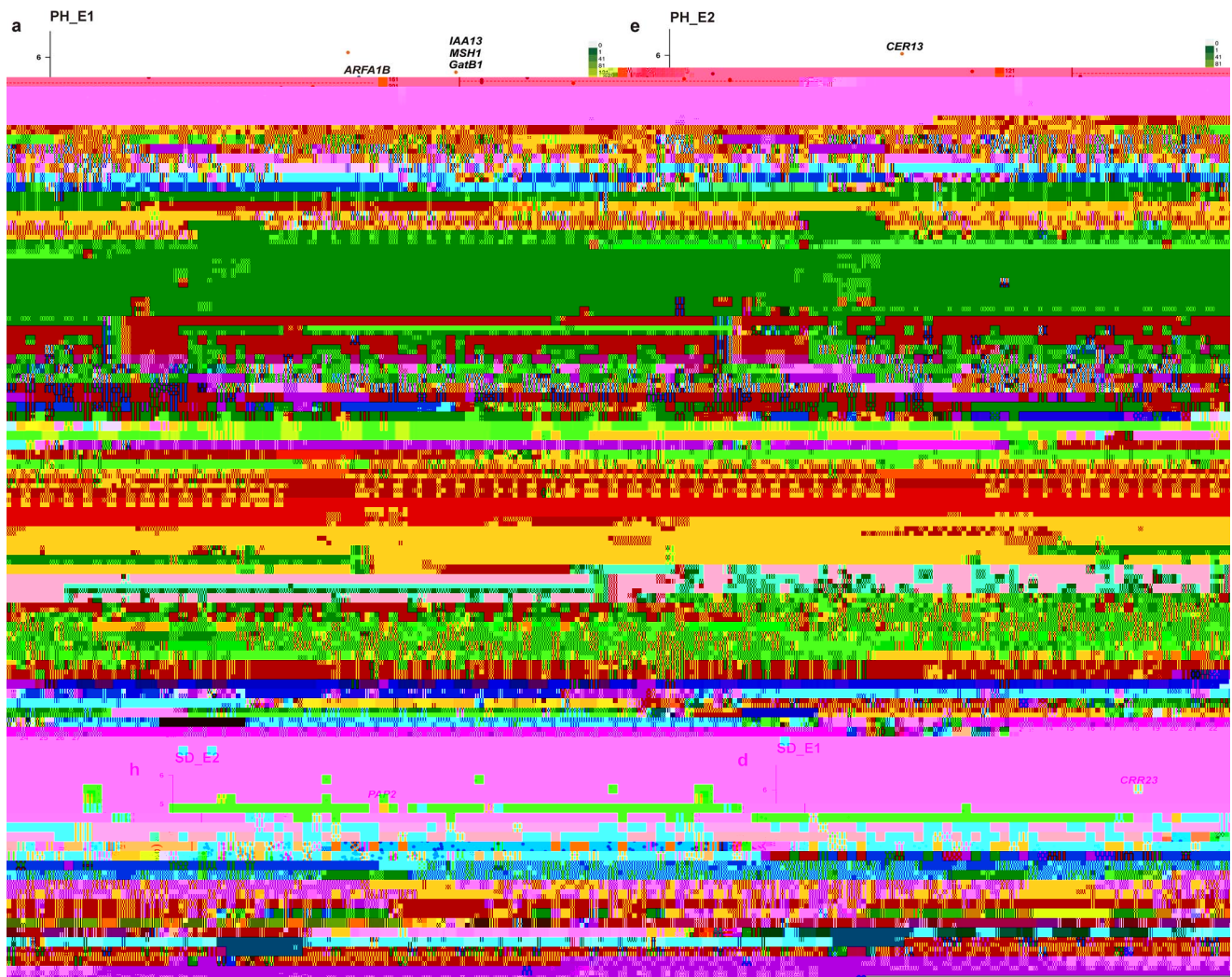
e 1.

Figure 2. Manhattan plots of GWAS for PH-related traits in E1 (–) and E2 (**e**–). The dashed line represents the significance threshold ($-\log10(p) = 5$), and proposed candidate genes for PH-related traits are marked above the corresponding SNPs.

illustrate that the SNPs Chr1_339370594 and Chr18_230810045 are associated considerably with PH-related traits, which might be selected in the process of chrysanthemum breeding. To further exclude false-positive rates, we then recruited XP-CLR analysis to define potential selective sweeps. Sixteen of the 130 differentiated genomic regions were shared by three approaches (Supplementary Data Table S8), covering 2.57 Mb of the whole genome and containing 221 genes. These genes were mainly enriched in biological processes through GO enrichment analysis, including cellular, metabolic, signaling, and developmental processes (Supplementary Data Table S9, Supplementary Data Fig. S8). We speculate that these genes may be involved in the domestication of chrysanthemum for PH.

e   e   - e   e

Genomic prediction of PH-related traits was carried out with 5-fold cross-validation using the rrBLUP model and different marker sets (Fig. 5, Supplementary Data Table S10). The prediction accuracy of the 330 710 SNPs for PH-related traits was significantly lower than that of sigGWAS, associated500, and associated1000 SNPs under the two environments, except for PH in E1. The prediction accuracy of the random1000 SNPs for PH-related traits was higher than that of the random500 SNPs in the two

environments. In addition, the prediction accuracy of the two random SNP sets was significantly lower than that of the two corresponding associated SNP sets, and the prediction accuracy of the sigGWAS SNPs containing 8–12 marker–trait associations was higher than or close to that of the random1000 and random500 SNPs, highlighting the efficiency of significant SNPs in the prediction of PH-related traits. Whether in E1 or E2, the associated1000 SNPs exhibited the highest prediction accuracy for four PH-related traits, with a range of 0.94–0.97. Thus the associated1000 marker set was more suitable for genomic selection. Meanwhile, we observed that the prediction accuracy of IN was slightly higher than that of other traits with different marker sets in the two environments.

Discussion

PH is one of the crucial components of plant architecture, affecting chrysanthemum quality and production efficiency. In chrysanthemum cultivation, plant growth regulators are often used to control the PH, which consumes a lot of human and material resources and increases potential risks to the environment and human health. Breeders need to pay more attention to plant architecture in chrysanthemums; however,

e . Significant SNPs associated with PH-related traits.

		( )	−$_{10}$(P)	e e		a_i		e
PH	Chr1_339370594	339 370 594	5.06	G/T	0.20	10.60**	E2	
	Chr4_13976903	13 976 903	5.13	A/G	0.23	11.25**	E2	
	Chr12_124163766	124 163 766	5.09	A/T	0.22	8.20**	E2	
	1 _ 0808 ♂	270 808 434	6.06	A/G	0.40	11.97**	E2	
	Chr12_270808478	270 808 478	5.36	G/A	0.40	10.83**	E2	
	Chr13_269150736	269 150 736	5.29	G/A	0.15	11.27**	E2	
	Chr14_183595760	183 595 760	5.03	G/A	0.10	11.92**	E2	
	Chr14_183595863	183 595 863	5.06	A/T	0.10	12.10**	E2	
	1 _ 81 ♂	281 792 329	6.18	T/A	0.06	12.12*	E1	
	Chr16_103728540	103 728 540	5.22	T/G	0.06	−13.06**	E1	
	1_ ♂ 188	9 532 188	5.42	G/T	0.10	−10.43**	E1	
	_18 1 1	184 971 991	5.38	C/A	0.10	17.40**	E2	
IL	Chr1_1200392	1 200 392	5.17	C/T	0.16	0.20*	E2	
	Chr2_345676	345 676	5.12	T/A	0.22	0.20**	E2	
	Chr2_345792	345 792	5.07	A/G	0.22	0.20**	E2	
	Chr2_345978	345 978	5.08	C/T	0.20	0.23**	E2	
	_110 ♂	110 756 763	5.96	C/A	0.12	0.28**	E2	
	1 _1 08 0	155 208 240	5.63	C/A	0.08	0.31**	E1	
	Chr19_155208347	155 208 347	5.12	G/A	0.09	0.29**	E1	
	_10 ♂	105 773 942	5.71	G/T	0.09	0.45**	E1	
	_8 ♂	82 362 276	6.28	C/A	0.07	0.40**	E1	
	_1 ♂ 01	124 735 601	6.32	G/A	0.06	0.29**	E1	
IN	Chr1_255259266	255 259 266	5.12	C/T	0.21	−5.91**	E1	
	Chr3_264364249	264 364 249	5.12	G/A	0.36	2.93^{ns}	E1	
	Chr3_264364290	264 364 290	5.21	C/T	0.36	2.89^{ns}	E1	
	♂_ ♂♂ ♂	233 732 224	5.55	A/T	0.23	−4.69**	E2	
	Chr5_70145757	70 145 757	5.22	G/A	0.06	8.95*	E1	
	/ _1 ♂ 0 ♂8	142 340 238	5.80	G/A	0.26	−5.78**	E1	
	1 _ 0808 ♂	270 808 434	6.80	A/G	0.40	6.34**	E2	
	1 _ 0808 8	270 808 478	6.33	G/A	0.40	5.11**	E1	
	1 _881 81	88 122 819	5.39	G/T	0.12	6.80**	E1	
	Chr18_244242678	244 242 678	5.14	C/T	0.16	3.58*	E1	
	Chr22_61447919	61 447 919	5.23	C/T	0.43	−9.27**	E1	
	Chr22_153548117	153 548 117	5.04	C/T	0.18	6.77		

e f. Information on proposed candidate genes for PH-related traits identified by GWAS.
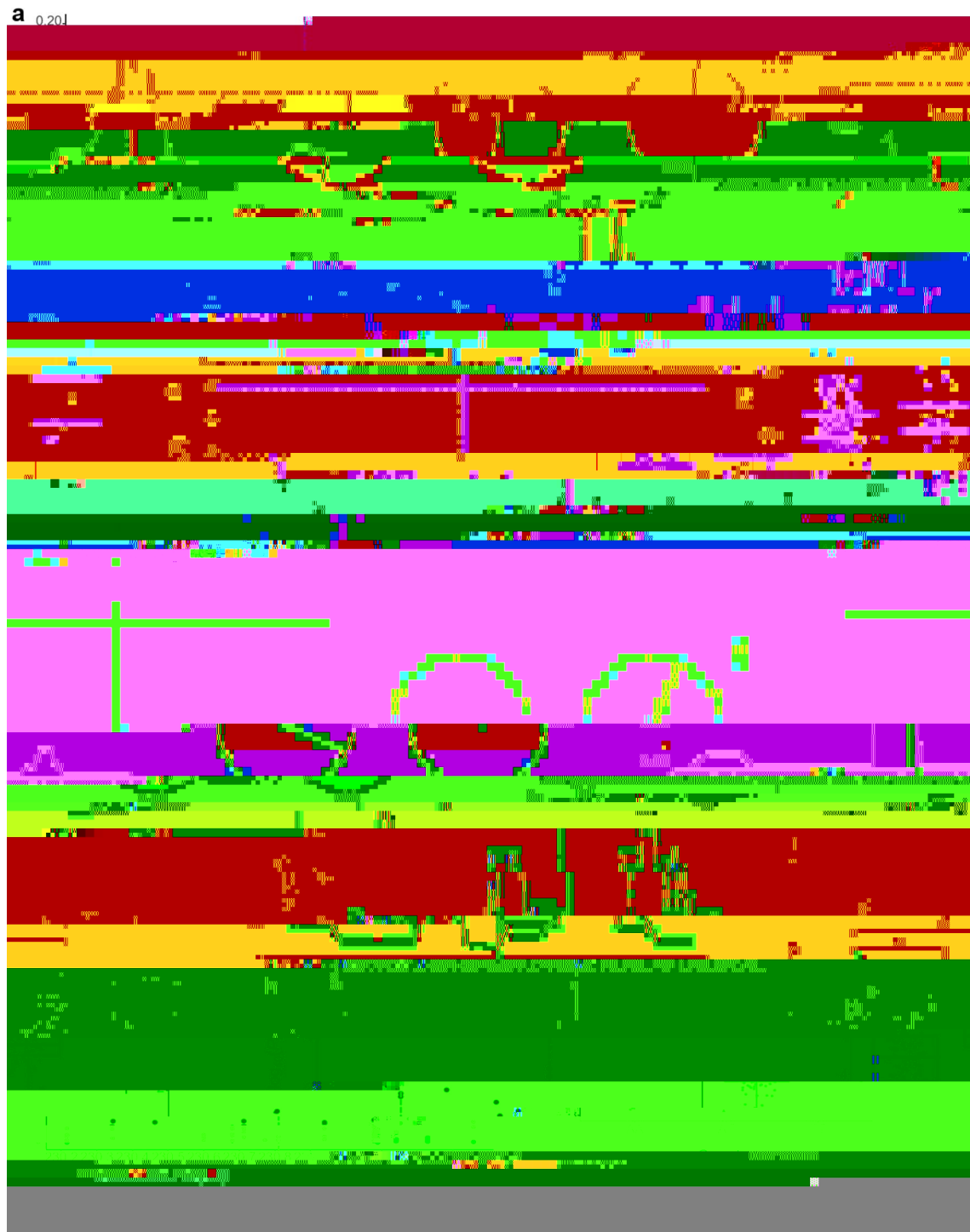
e e e

Figure 4. Selective signature and phenotypic difference between individuals with different alleles of Chr1_33937059 associated with PH (–**e**) and Chr18_230810045 associated with SD (–**j**). , F_{ST} values of the tall group versus the short group. , Distribution of π in the tall and short group of the 1.1 Mb genetic regions surrounding the SNP. , Significant GWAS signals coincident with the selective sweep regions. , Allele frequencies of the SNP between the tall and short groups. **e**, **j** Boxplot of PH and SD in accessions with different genotypes. *P < 0.05; **P < 0.01; ns not significant.

several novel candidate genes involved in various pathways, including vesicle-mediated transport, transcriptional regulation, metabolism, methylation, redox, and growth with stress. Although no reports demonstrate that these genes are directly related to the regulation of PH development, the candidate genes underlying associated loci through forward genetics provide important genetic resources for further research on chrysanthemum PH breeding.

Chrysanthemum is a complex species with high heterozygosity (>3%), high repetitiveness (>80%), huge genome size (8.47–9.02 Gb) [9], rapid decay rate, and a relatively low gene density of approximately one gene per 58.74 kb, which makes it rather challenging to identify causal genes underlying the detected association loci within a relatively narrow genomic region. It has been documented that a wide range of 100 kb region centered on the peak SNPs has been employed to explore candidate genes
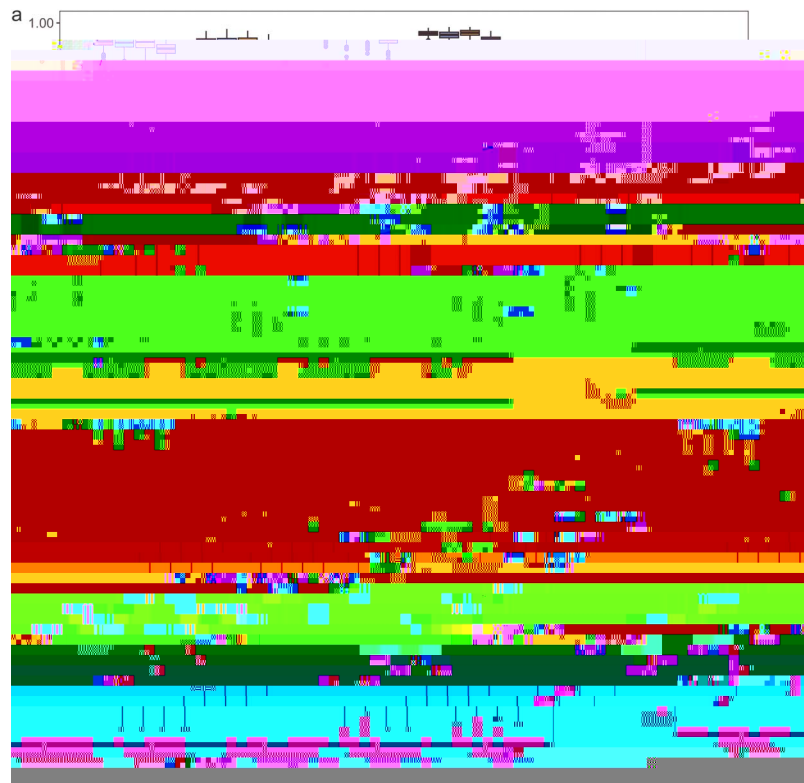
Figure 5. Genomic prediction accuracy for the four PH-related traits evaluated using the rrBLUP model in E1 ( ) and E2 ( ).

in species with rapid decay rate, such as ginkgo [55] and apple [56]. Therefore, to overcome the challenge of candidate gene identification in chrysanthemum GWAS, a relatively less stringent

the total number of internodes from the bottom of the stem to the primary inflorescence. IL was measured as the average internode length in the middle and upper parts. SD was measured as the diameter of the middle and upper internode. The mean value of

associated500 and associated1000, included the top 500 and 1000
associations, respectively. Similarly, the two sets random500 and