Euphytica (2022) 218:177
https://doi.org/10.1007/s10681-022-03124-7

Genetic analysis of anemone-type and single-type inflorescences in chrysanthemum using genotyping-by-sequencing

Ye-Ji Lee · Jae-A Jung · Na Hyeong Lee · Jung Sun Kim · So Youn Won

Received: 6 May 2022 / Accepted: 30 October 2022 / Published online: 30 November 2022
© The Author(s) 2022

Abstract Flower shape is a key trait of ornamental and commercial importance in breeding programs for chrysanthemum (*Chrysanthemum morifolium* Ramat.). Understanding the genetic basis of the phenotypic variation seen in inflorescence-related traits will contribute to genetic improvement and to the development of new varieties. In this study, we investigated the genetic determinants of inflorescence traits using an F_1 segregating population derived from a cross between two cultivars with different inflorescence types, 'Puma White' (anemone-shaped inflorescence) and 'Dancer' (single-type inflorescence). Genotyping-by-sequencing identified 26,847 single-nucleotide polymorphisms (SNPs) between 182 F_1 progenies and their parents. A genome-wide association study highlighted 17 SNPs mapping to 15 GBS-tags as being significantly associated with three inflorescence traits: flower type, number of ray florets, and disk flower diameter. No single SNP was associated with flower diameter. These SNP-harboring sequences defined ten candidate genes associated with inflorescence traits. We explored the transcript levels for nine of these in flower buds, disk florets and ray florets using publicly available genome and transcriptome data. These results will provide the genetic and genomic foundation to harness important horticultural traits and explore new avenues in chrysanthemum breeding.

Keywords Chrysanthemum · Flower shape · Genotyping-by-sequencing (GBS) · Genome-wide association study (GWAS) · Candidate genes

Introduction

Chrysanthemum (*Chrysanthemum morifolium* Ramat.) belongs to the Asteraceae family and is an economically important ornamental plant used as cut flowers or potted plants, comprising a considerable proportion of the global flower industry (Zhang et al. 2010). Phenotypic diversity in chrysanthemum is illustrated by its multiple flower types with various colors and sizes. Chrysanthemum blooms, or flower heads, are formed by many small disk florets (flowers) surrounded by ray florets (Yoshioka et al. 2010). The flower heads of chrysanthemum can take on various forms described as anemone,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10681-022-03124-7.

Y.-J. Lee · N. H. Lee · J. S. Kim (✉) · S. Y. Won
Genomics Division, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Republic of Korea
e-mail: jsnkim@korea.kr

S. Y. Won
e-mail: soyounwon@korea.kr

J.-A. Jung
Floriculture Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju 55365, Republic of Korea

pompon, single, or double inflorescence types, among others (Chen et al. 2009). Anemone-type flower heads are characterized by a prominent flower center composed of elongated and colored disk florets, making them both aesthetically and commercially attractive (Anderson 2006). Single-type flower heads consist of a central and compact disk composed of hermaphrodite florets and are considered the primitive form (Dai and Chen 1997; Dai 2004; Chen et al. 2009).

Crossbreeding has contributed to the development of modern chrysanthemum cultivars (Yang et al. 2019). Cultivated chrysanthemum show great diversity in their morphology and characteristics such as inbreeding depression, self-incompatibility, and high levels of heterozygosity. Accordingly, crosses between two parents with contrasting phenotypic traits of interest will produce F_1 progeny exhibiting extensive phenotypic variation (Yang et al. 2019). As phenotypes are affected by the underlying genotype, the surrounding environment, and genotype × environment interactions, phenotype-based selection is difficult and ineffective. These limitations are currently circumvented with the help of marker-assisted selection (MAS), whereby the genes controlling target traits, or closely linked loci, are used to select lines based solely on their genotype rather than their phenotype. Using MAS, a large collection of progeny derived from crossbreeding can be screened quickly and efficiently irrespective of any effect from the environment. This allows us to explore the influence of a target trait during earlier growth stages than those typically associated with the given trait, for example those influencing flowering time. The identification of the necessary molecular markers has been fueled by genetic analyses such as quantitative trait locus (QTL) mapping and genome-wide association study (GWAS). In chrysanthemum, many studies attempted to determine the genes associated with important traits using traditional molecular approaches such as amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNA markers (RAPD), inter simple sequence repeats (ISSRs), restriction fragment length polymorphism markers (RFLPs), expressed sequence tag-simple sequence repeats (EST-SSRs), and sequence-related amplified polymorphisms (SRAPs) (Zhang et al. 2010, 2011; Fan et al. 2020; Yang et al. 2020). However, the number of available markers for genetic analysis is limited and the resulting genomic regions were large. In addition, genetic work in chrysanthemum is challenging due to its complex genome.

Cultivated chrysanthemums have a hexaploid genome whose stable chromosome number is most frequently ($2n = 6x = 54$), although some cultivars have been described as aneuploid with a chromosome number ranging from 47 to 67 (Dowrick 1953). It is still unclear whether this event involved autopolyploid or allopolyploid. A recent cytological and molecular analysis revealed segmental allohexaploid in cultivated chrysanthemums (Klie et al. 2014). Our previous study, based on the Ks value distributions, whole-genome triplication (WGT) event ($Ks = 0.1$), as well as spices divergence ($Ks = 0.05$). At $Ks = 0.05$ divergence time, while *C. morifolium* maintained the original polyploidy status, it is likely that this particular diploid species experienced a rapid diploidization event, including a reduction in chromosome number (Won et al. 2017). Although its gigantic genome size of 6–7 Gb (Van Geest 2017) and high heterozygosity in cultivated chrysanthemum genome status caused in ambiguous of auto or allopolydization evidence for a recent WGD or WGT.

The advent of next-generation sequencing (NGS) technologies has made possible and accelerated the pace of genetic analyses and the development of molecular markers (Varshney et al. 2009). NGS platforms now allow the global identification of millions of single-nucleotide poly-morphisms (SNPs) and insertion-deletion polymorphisms over the genome, which have greatly increased the analytic power of many genetic studies (Varshney et al. 2009; Huang et al. 2010; Phan and Sim 2017). The GBS methodology has been implemented in many plants to explore their genetic diversity, construct linkage maps, map QTLs, and perform GWAS (Nguyen and Lim 2019). Besides simpler diploid species, GBS has also been widely adopted in polyploid species such as wheat (*Triticum aestivum*) and sugarcane (*Saccharum officinarum*) (Poland et al. 2012; Balsalobre et al. 2017). A recent example applied GBS to an exploration of genetic diversity and phylogenetic analysis for wild chrysanthemum species using 7,758 SNP markers (Nguyen et al. 2020).

In this study, we used an F_1 population derived from a cross between two cultivars with different inflorescence types, 'Puma White' (anemone-type) and 'Dancer' (single-type) to dissect the genetic basis of inflorescence traits in chrysanthemum. The

progeny, consisting of 182 lines, was subjected to phenotypic characterization, genome sequencing by GBS, and GWAS. We then used publicly available transcriptome data to examine the possible function of the candidate genes derived from GWAS on shaping inflorescence traits. The result of this study will provide much needed information to understand the genetic basis underlying ornamental traits and to enhance chrysanthemum breeding efficiency.

Materials and methods

Plant materials

A population of 182 F_1 progeny was derived from a cross between two hexaploid chrysanthemum (*C. morifolium*) cultivars, 'Puma White' and 'Dancer' (Park et al. 2015). To minimize selfing, 'Puma White' was selected as the maternal parent because its anemone-type inflorescence produces very little pollen. All plant materials are currently maintained at the National Institute of Horticultural and Herbal Science, Rural Development Administration, Republic of Korea.

Phenotypic evaluation and data exploration

Fresh healthy stems from the original 182 F_1 population and the parents, were grown in the 128-plug tray for 3 weeks. Rooted cuttings were then transplanted and grown in a polyethylene greenhouse in September 2018 under natural photoperiod to induce flowering in short-day conditions. Four inflorescence-related traits were investigated in the F_1 population and their parents at the flowering stage: flower type (FT), flower diameter (FD, in cm), number of ray florets (NRF), and disk flower diameter (DFD, in cm). For FT, inflorescences with cushion-like disk florets were considered to belong to the anemone type, while those with short and dense disk florets were defined as single flower heads. Traits were measured using randomly selected inflorescences for each F_1 line as three different F_1 with replicates. Exploration of the phenotypic data was performed and visualized in Microsoft Excel 2016 or with R software (R Core Team 2014). The difference in traits between the two parents was compared using a one-tailed Student's t-test. The normality of the traits in the F_1 population was examined by calculating the skewness and kurtosis of the relevant distributions and by conducting a Shapiro–Wilk normality test. The correlation between traits in the F_1 population was examined by calculating Kendall rank correlation coefficients. The difference in traits between the two inflorescence types of the F_1 population was examined by Wilcoxon rank sum test with continuity correction.

GBS library preparation and sequencing

Total genomic DNA (gDNA) was extracted from young leaves using the DNeasy® Plant Mini Kit (Qiagen) following the manufacturer's protocol. The quality and concentration of gDNA were assessed by electrophoresis on 1.2% (w/v) agarose gels, on a Dropsense96 spectrophotometer (Trinean), and with the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen). gDNA from each sample was digested with the restriction enzyme *Ape*KI (New England Biolabs) and then used as template for the construction of a GBS library, as previously described (Elshire et al. 2011). After checking the quality of libraries on a 2100 Bioanalyzer instrument (Agilent), the libraries were pooled and sequenced using a NextSeq500 platform (Illumina) as 150 bp reads in single-end mode.

Sequencing data analysis and genotyping

GBS data were analyzed using the pipeline Stacks v. 2.5 with default settings (Catchen et al. 2011, 2013). Raw reads were demultiplexed, filtered, and trimmed to 100 bp using process_rad-tags (cleaned raw read data) in Stacks and Cutadapt (Martin 2011). Stacks was run in de novo stack formation mode (Catchen et al. 2011). These assembled GBS-tags were replaced as a reference genome and named sequence tag. To track a set of loci or genotypes, the identified variants were finally filtered using VCFtools v. 0.1.15 (Danecek et al. 2011) with the following parameters: minor allele frequency $\geq 5\%$, mean read depth ≥ 5 reads, and missing genotype $< 5\%$.

GWAS of inflorescence traits

With the set of filtered SNPs defined above, GWAS was performed using TASSEL 5.0 software (Bradbury et al. 2007) with the general linear model (GLM) algorithm. Quantile–quantile (QQ) plots

indicated that the naïve model consisting of a GLM without any control for population structure (Q) or kinship matrix (K) was a suitable model for analysis. The significance threshold for a trait and marker association was set to $P \leq 0.001$. The P values were adjusted by the threshold of the Bonferroni correction for multiple tests (1/n), where n is the total number of SNPs used in the association analysis (Kan et al. 2015; Sun et al. 2017). Significant SNPs located in the same GBS sequence tag (GBS-tag) were considered as belonging to the same locus.

Identification and expression analysis of candidate genes

GBS-tag sequences carrying trait-associated SNPs were searched against the genome of diploid *Chrysanthemum nankingense* (Song et al. 2018) using Basic Local Alignment Search Tool for nucleotide sequences (BLASTN) (Altschul et al. 1997) with an E-value cutoff of 1×10^{-10}. To determine the transcript levels of candidate genes, transcriptome deep sequencing (RNA-seq) data from *C. nankingense* were downloaded for six tissues (roots, stems, leaves, flower buds, disk florets, and ray florets) (http://www.amwayabrc.com/download.htm) (Song et al. 2018). Reads were aligned to the *C. nankingense* genome and transcript levels were calculated and normalized as fragments per kilobase of transcript per million mapped reads (FPKM) using TopHat and Cufflinks (Trapnell et al. 2009, 2012). Genes with FPKM ≥ 1.0 were considered to be transcriptionally expressed. Transcript levels for each gene were visualized as a heatmap using Z-score normalization in R software (R Core Team 2014). Additionally, candidate proteins were subjected to BLASTP against the Arabidopsis (*Arabidopsis thaliana*) protein sequences (from the TAIR10 annotation) at the Ensembl Plants website (http://plants.ensembl.org).

Results

Variation in inflorescence traits

We characterized the extent of phenotypic variation in inflorescence traits in the 182 F_1 progeny derived from the two parental cultivars (Figs. 1 and 2). The cultivar 'Puma White' has anemone-type inflorescence that are composed of two layers of ray florets at the edge and cushion-like disk florets at the wide flower core (Fig. 1a), while the cultivar 'Dancer' has single inflorescence with one layer of ray florets at the edge and flattened disk florets in a narrow arrangement at the center of the flower (Fig. 1b). We classified their F_1 progeny into 60 anemone-type and 122 single-type inflorescences (Fig. 1c and Online Resource 1A). FD was the same between the two parents, while NRF and DFD were significantly greater in 'Puma White' than in 'Dancer' ($P < 0.001$) (Table 1). In the F_1 progeny, FD ranged from 2.83 to 6.83 cm, with a mean of 4.55 cm; NRF varied from 12 to 66 ray florets, with a mean of 29.24; and, DFD ranged from 1.00 to 3.00 cm, with a mean of 1.73 cm. All three phenotypic traits therefore showed strong transgressive segregation, as the range of phenotypic values extended well beyond that defined by the parental lines (Fig. 2). We next calculated the coefficient of variation for each trait: 16.04% for FD, 26.00% for NRF, and 20.80% for DFD. Shapiro–Wilk normality tests revealed that all three traits significantly deviated from normal distributions ($P < 0.05$). We observed significant correlations ($p < 0.05$) between FD and NRF, FD and DFD for the F_1 population (Online Resource 2). A comparison of traits between anemone-type and single-type inflorescence of the F_1 population indicated that DFD was significantly larger in anemone-type inflorescence than in single-type inflorescence ($P < 0.001$), while FD and NRF were not different between the two types of inflorescences (Online Resource 1B).

Identification of SNPs by GBS

As a prerequisite for genomic analysis, we genotyped all individuals in the F_1 population and their parents by GBS, resulting a total of 688,806,674 raw reads, with an average of 3,743,515 reads per individual (Table 2 and Online Resource 3). We then filtered and preprocessed these raw reads to yield 326,317,886 high-quality reads, or 47.40% of the initial number of raw reads. We then implemented the Stacks pipeline to assemble all clean reads into sequence tags covering 56,486,574 bp in length, to which we mapped the clean reads to identify SNPs. The number of mapped reads per individual ranged from 453,348 to 1,693,058, with an average of 861,374. The Stacks pipeline discovered a total of 240,117 SNPs, which



Fig. 1 Flower morphology of *C. morifolium* 'Puma White' (**a**) and 'Dancer' (**b**). **c** Representative flower head phenotypes among the population of 182 F_1 progeny at the flowering stage. Scale bar = 1 cm

we narrowed down to 26,849 SNPs for subsequent analysis based on the selection criteria of minor allele frequency of at least 5%, read support from at least five reads, and missing genotype below 5% across the F_1 population.

GWAS and candidate gene prediction

To identify the loci associated with variation in inflorescence traits in chrysanthemum, we performed GWAS using the genotype (26,849 SNPs) and phenotype (FT, FD, NRF, DFD) data from all 182 F_1 population. The association analysis yielded 17 significantly associated SNPs mapping to 15 distinct GBS-tags from the de novo assembly ($P \leq 1.1 \times 10^{-06}$) (Online Resources 4 and 5). Of these 17 SNPs, 11 mapped to 10 GBS-tags for FT, 3 mapped to 3 GBS-tags for NRF, and 3 SNPs mapped to 2 GBS-tags for DFD. We detected no significant association between genotype and FD. The GBS-tags GBS-tag31024 and GBS-tag20699 each harbored two SNPs associated with FT and DFD, respectively.

To identify the candidate genes underlying each trait, we queried the genome of diploid *C. nankingense*, one of the progenitors of *C. morifoliu*m (Song et al. 2018) with the 15 GBS-tags of significant SNPs. BLAST searches determined that 12 GBS-tags showed significant hits in the *C. nankingense* genome (E-value $\leq 2.0 \times 10^{-26}$, percent identity $\geq 92\%$, and query coverage $\geq 84\%$), while 3 GBS-tags did not match any sequence in *C. nankingense* or across the National Center for Biotechnological Information (NCBI) non-redundant (nr) database (Online Resource 6). Of the 12 GBS-tags with a match in the *C. nankingense* genome, 2 (harboring 3 SNPs) mapped to intergenic regions, but the remaining 10 GBS-tags (and their 11 SNPs) located to genes. Among these 11 genic SNPs, 1 was in a 5'

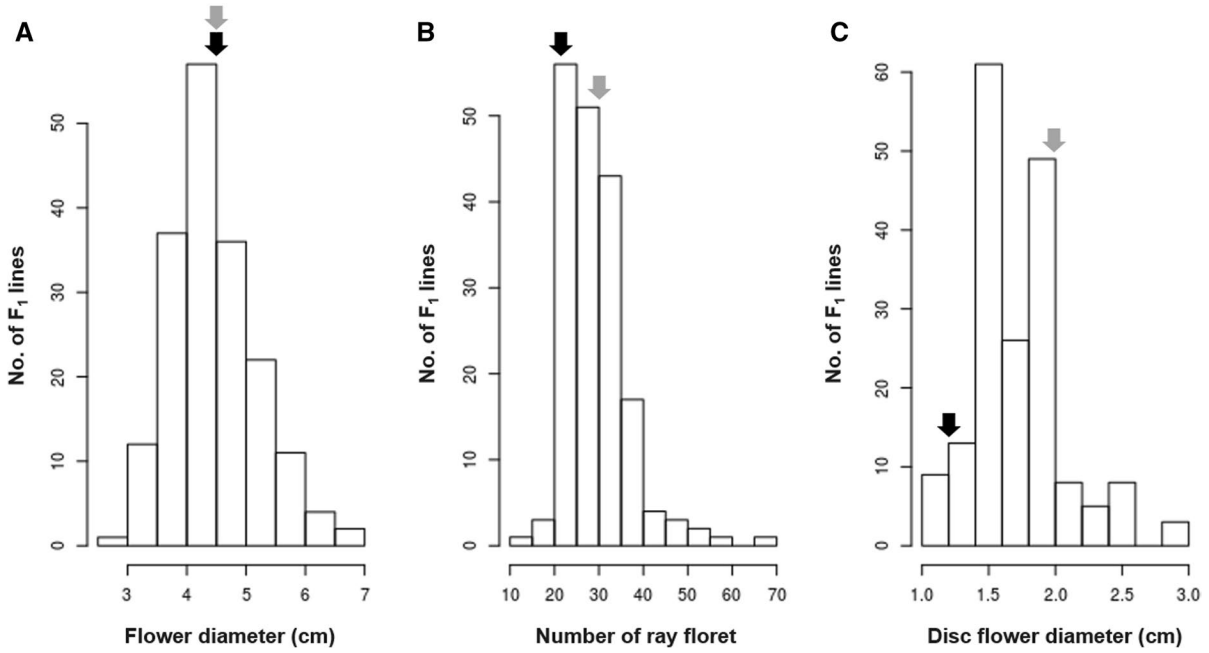


Fig. 2 Distribution of three inflorescence traits among the 182 F_1 progeny derived from a cross between *C. morifolium* 'Puma White' and 'Dancer'. The values for the two parental cultivars are indicated by grey ('Puma White') and black ('Dancer') arrowheads. Note that flower diameter is the same between the two parents. **a** Flower diameter. **b** Number of ray florets. **c** Disc flower diameter

Table 1 Descriptive statistics associated with the variation for flower-related traits based on a set of 182 F_1 population derived from the cross between 'Puma White' and 'Dancer'

Trait[a]	Parents		F_1 Population Mean ± SD[b]	Range	Skewness	Kurtosis	Coefficient of variation (%)
	'Puma White'	'Dancer'					
FD	4.50	4.50	4.55 ± 0.73	2.83–6.83	0.64	0.43	20,016.04
NRF	29.70	21.30***	29.24 ± 7.50	12.00–66.00	1.51	4.14	26.00
DFD	2.00	1.20***	1.73 ± 0.36	1.00–3.00	0.80	0.87	20.80

[a]Flower diameter (FD, cm); number of ray florets (NRF), disc flower diameter (DFD, cm)

[b]Standard deviation

***Significant between two parents at the 0.001 level according to one-tailed t-test

untranslated region (UTR) and 4 were within introns. Two other SNPs introduced a synonymous mutation that did not affect the protein sequences of their associated genes CHR00045339 and CHR00036293, encoding a homolog of ubiquitin-specific protease 16 (UBP16) and UDP-glucosyltransferase 85A1 (UGT85A1), respectively (Table 3). The remaining four SNPs introduced missense mutations in three genes and affected the encoded protein. The SNP in GBS-tag1536:83 mapped to the gene CHR00032674 and resulted in a K46Q amino acid substitution in the encoded aspartic peptidase (AP) anchored to the plasma membrane. The SNP within GBS-tag2771:23 affected gene CHR00074055 by introducing a N79K amino acid substitution in the encoded lipase class 3 family protein. Finally, two SNPs mapping to the same gene (CHR00047948) introduced T131A and G149S amino acid substitutions in the encoded octicosapeptide/Phox/Bem1p protein homolog.

Table 2 Summary of GBS approach and SNP calling of *C. morifolium* using 182 F_1 population and its parents

GBS approach[a]	Value
No. of raw reads	688,806,674
No. of clean reads after preprocessing	326,317,886
Survived reads (%)	47.4
No. of mapped reads	158,492,815
No. of an average mapped reads per sample	861,374
Max mapped reads per sample	1,693,058
Min mapped reads per sample	453,348
SNP calling	
Length of sequence tags (bp)	56,486,574
No. of raw SNPs	240,117
No. of SNPs after filteration[b]	26,849

[a]GBS libraries produced by *Ape*KI restriction enzyme to digest chrysanthemum gDNA

[b]Filtering options: a minor allele frequency of at least 5%; mean read depth with at least 5 reads; and 5% or more missing genotyping data

Table 3 Details of SNPs within GBS-tag sequences associated with flower-related traits and effects in translation

Trait[a]	SNP ID	Gene ID	Functional annotation	Position	CDS Location[b]	Protein change
FT	GBS-tag1536:83	CHR00032674	Aspartic peptidase family protein	CDS	136	K46Q
	GBS-tag2771:23	CHR00074055	Alpha/beta-hydrolases superfamily protein (Lipase class 3 family protein)	CDS	477	N79K
	GBS-tag23721:86	CHR00024623	Peptidyl-lysine demethylation	5'UTR	NA[c]	NA
	GBS-tag25938:45	CHR00045339	Ubiquitin-specific protease 16 (UBP16)	CDS	348	NA
	GBS-tag31024:39 GBS-tag31024:87	CHR00047948	Octicosapeptide/Phox/Bem1p family protein (PB1)	CDS	391,445	T131A, G149S
	GBS-tag40423:43	CHR00050395	GATA transcription factor 25 (GATA25)	Intron	NA	NA
	GBS-tag49292:32	CHR00036293	UDP- glycosyltransferase 85A1 (UGT85A1)	CDS	234	NA
	GBS-tag50705:95	NA	NA	NA	NA	NA
	GBS-tag62115:69	CHR00047943	ATGRDP1, Glycine-rich-domain protein1	Intron	NA	NA
	GBS-tag71767:11	CHR00091050	Alpha/beta-hydrolases superfamily protein (Lipase class 3 family protein)	Intron	NA	NA
NRF	GBS-tag1280:27	NA	NA	NA	NA	NA
	GBS-tag51547:43	CHR00031888	Magnesium transporter, putative (DUF803)	Intron	NA	NA
	GBS-tag52560:48	NA	NA	Intergenic	NA	NA
DFD	GBS-tag12113:80	NA	NA	NA	NA	NA
	GBS-tag20699:18	NA	NA	Intergenic	NA	NA
	GBS-tag20699:20	NA	NA	Intergenic	NA	NA

[a]Flower type (FT); number of ray florets (NRF), disc flower diameter (DFD, cm)

[b]SNP position (bp) in the coding sequence of the gene

[c]Not applicable

We next examined transcript levels for the 10 candidate genes harboring SNPs using public RNA-seq data from *C. nankingense* (Song et al. 2018). With the exception of CHR00074055, nine of our candidate

genes appeared to be expressed (expression cutoff of FPKM ≥ 1) (Online Resource 7). The expression of candidate genes involving significant SNP was detected in flower buds, disk florets, and ray florets; each candidate gene also demonstrated differential expression between analyzed tissues (Fig. 3).

Discussion

Molecular genetic analysis and MAS are limited and slow to be implemented in cultivated chrysanthemum (*C. morifolium* Ramat.) due to its large, complex and hexaploid genome, self-incompatibility, and high level of heterozygosity. Previous genetic work revealed that inflorescence traits are regulated by multiple loci rather than a single major locus in chrysanthemum (Tang et al. 2015). However, several genes were later reported to be involved in shaping inflorescence traits (Chong et al. 2016, 2019). In this study, we investigated the genetic architecture of inflorescence traits using a set of 182 F_1 segregating progeny derived from a cross between the anemone-type cultivar 'Puma White' and the single-type cultivar 'Dancer'. We used GBS to identify SNPs across the genome in all progeny and their parents, followed by GWAS to detect SNPs significantly associated with each inflorescence trait. The resulting candidate genes will provide valuable resources for chrysanthemum breeding.

Genetic analysis using segregating populations is a powerful tool to identify loci responsible for phenotypic variation in target traits. The F_1 progenies used in this study was obtained from a cross between two chrysanthemum cultivars with different inflorescence types and showed substantial variation and transgressive segregation in the flower traits examined, making this F_1 population a suitable mapping population. In the F_1 progeny, more individuals displayed single-type inflorescence than the anemone-type, suggesting that the single inflorescence morphology may be a dominant phenotype in this particular crossing combination, which should be examined further in more diverse germplasms. Progeny with anemone-type flower heads had larger DFD than did single-type progeny, which was consistent with the differences between the parents. By contrast, FD and NRF were similar in the two groups of F_1 progeny with different inflorescence types. We observed a weak significant correlation between the phenotypic traits. A relatively negative correlation was observed between NRF and DFD, which is consistent with the result in Fan et al. (2020). For example, the correlation of − 0.055 was in our study and the value was − 0.45

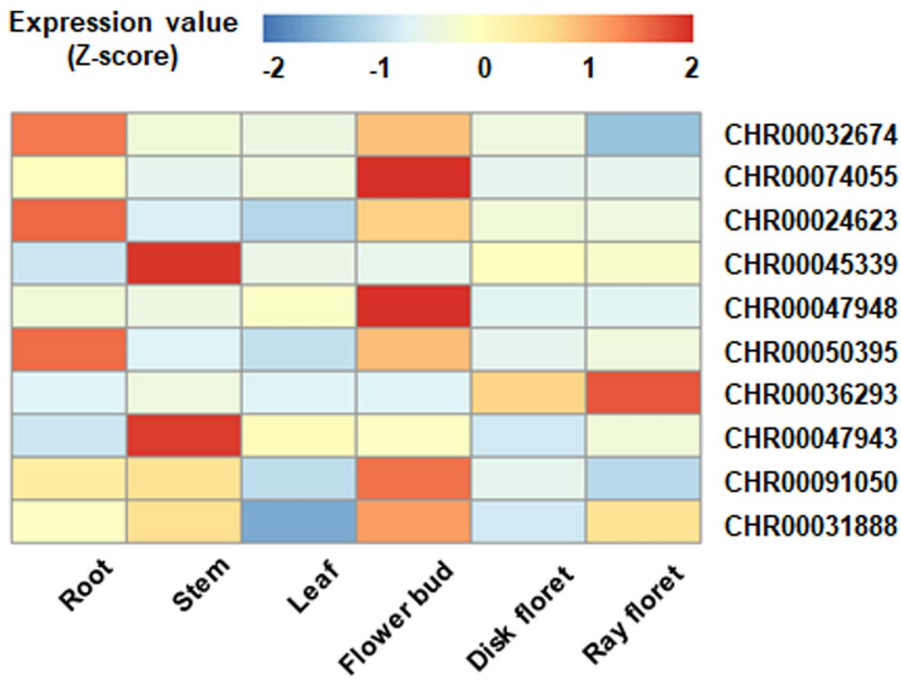


Fig. 3 Heatmap representation of the expression patterns of candidate genes for inflorescence traits in chrysanthemum. Z-score normalization of FPKM values was conducted for heatmap generation. Red represents high expression; blue indicates low expression

in Fan et al. (2020). A correlation between FD and DFD was observed within a panel of anemone-type chrysanthemum accessions (Yang et al. 2020). These findings suggest that flower traits are complex in chrysanthemum.

Traditional molecular markers have enabled genetic studies in chrysanthemum such as investigation of inheritance, linkage map construction, QTL mapping, GWAS, and phylogenetic studies, but the number of such markers is insufficient (Zhang et al. 2010, 2011; Chong et al. 2019), only reaching hundreds of markers in biparental populations and natural populations of chrysanthemum (Zhang et al. 2010, 2011; Li et al. 2016; Fan et al. 2020). With NGS technologies, however, far more SNP markers can be discovered quickly and at low cost even in species lacking a reference genome, which is expected to substantially improve the development of new molecular markers and breeding in chrysanthemum (Su et al. 2019a). GBS identified 7,758 SNPs across 80 wild chrysanthemums accessions (Nguyen et al. 2020) and 26,849 SNPs from 182 F_1 progeny and their parents in the present study. Another NGS method, double digest restriction site-associated DNA (dd-RAD-seq) identified 9,219 SNPs in an F_1 biparental population to dissect the genetic control of petal color in ray florets (Sumitomo et al. 2019). Lastly, specific locus-amplified fragment sequencing (SLAF-seq) detected approximately 92,000 SNPs, which helped identify alleles associated with inflorescence traits and waterlogging tolerance in cultivated chrysanthemum accessions (Chong et al. 2016, 2019; Su et al. 2019b).

Inflorescence traits are some of the most important factors that determine the ornamental and commercial value of chrysanthemum. GWAS identified 17 SNPs significantly associated with inflorescence-related traits mapping to 15 GBS-tags. GWAS is a powerful tool to identify loci associated with specific traits (Yu et al. 2006). In a previous study, GWAS identified DNA markers associated with petal color in a segregating F_1 mapping population of chrysanthemum, in which inheritance is typically rather complex due to its autohexaploid genome (Sumitomo et al. 2019). The random pairing of chromosomes in autopolyploids complicates the calculation of recombination frequencies during QTL analysis, although it is not impossible. Nevertheless, our result demonstrated the applicability of the allohexaploid genome by genetic and gene functional analysis in *C. morifolium*. Three GBS-tags (GBS-tag50705:95, GBS-tag1280:27, GBS-tag12113:80), specific to *C. morifolium* were newly discovered based on the *C. nankingense* genome. A similar approach to our study has been confirmed in other plants as well. In the allotetraploid rapeseed (*Brassica napus*), GWAS constituted a robust approach to identify candidate genes involved in the regulation of fatty acid composition in a biparental doubled-haploid mapping population (Gacek et al. 2017).

Twelve of the 15 GBS-tags harboring the 17 significant SNPs aligned well to the *C. nankingense* genome, supporting the notion that the two species are closely related. We hypothesize that the three remaining GBS-tags are specific to *C. morifolium*. We assigned these 12 GBS-tags to 10 candidate genes potentially associated with inflorescence traits. Publicly available RNA-seq data indicated that all but one of our candidate genes are expressed in flower buds, disk florets, and/or ray florets. Among them, two other significant SNPs affected a gene encoding an octicosapeptide/Phox/Bem1p family protein. The homologous gene in Arabidopsis was shown to be rapidly upregulated in roots within 30 min of treatment with the phytohormone auxin; its knockout mutant exhibited reduced lateral root growth in the presence of auxin (Pu et al. 2019). One member belonging to this protein family was identified during a GWAS using a collection of barrelclover (*Medicago sativa*) accessions for its association with a plant branch trait, and its overexpression resulted in an increased number of lateral branches in Arabidopsis (Wang et al. 2020). The octicosapep-tide/Phox/Bem1p protein family contains Phox/Bem1 (PB1) domains that mediate protein–protein interactions through different charges at their protein surfaces (Korasick et al. 2015). PB1 domains are also present in two classes of auxin signaling regulators, auxin response factors and auxin/indol 3-acetic acid inducible proteins and mediate their interaction (Korasick et al. 2015). We therefore hypothesize that this PB1-containing protein probably regulates inflorescence development and growth in chrysanthemum by facilitating interactions between proteins in the auxin signaling pathway. For the remaining genes containing SNPs in their 5' UTR or introns, or with synonymous mutations (UBP16 and UGT85A1), alternative splicing or differential expression might be involved in the regulation of inflorescence traits. Alternatively,

tightly linked nearby genes might be responsible for these traits.

This GBS-based association study helped the identification of potential SNPs and genes with phenotypic effects on important horticultural traits in the present study. Of them, ten candidate genes contribute to the inflorescence traits were identified in this study, and awaits the validation and direct evidence of the roles of the candidate genes to support chrysanthemum breeding. The findings of the present study provide the genetic and genomic foundation to harness important horticultural traits and explore new avenues in chrysanthemum breeding.

Author contributions Conceptualization: J.-A.J., J.S.K., and S.Y.W.; Methodology, N.H.L.; Formal analysis, Y.-J.L. and S.Y.W.; Investigation, S.Y.W.; Resources, J.-A.J.; Data curation, S.Y.W.; Writing—original draft preparation, Y.-J.L. and S.Y.W.; Writing—review and editing, J.S.K. and S.Y.W.; Visualization, Y.-J.L. All authors have read and approved the final manuscript.

Funding This work was supported by a grant from the National Institute of Agricultural Sciences (Project No. PJ013353 and PJ015780), Rural Development Administration, Republic of Korea.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anderson NO (2006) Chrysanthemum: *Dendranthema* x *grandiflora Tzvelv*. Flower breeding and genetics: issues, challenges and opportunities for the 21st century. Dordrecht, The Netherlands, pp 389–437

Balsalobre TW, da Silva PG, Margarido GR, Gazaffi R, Barreto FZ, Anoni CO, Caradoso-Silva CB, Costa EA, Mancini MC, Hoffmann HP, de Souza AP (2017) GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. BMC Genom 18:72. https://doi.org/10.1186/s12864-016-3383-x

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 1:171–182. https://doi.org/10.1534/g3.111.000240

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Chen FD, Li FT, Chen SM, Guan ZY, Fang WM (2009) Meiosis and pollen germinability in small-flowered anemone type chrysanthemum cultivars. Plant Syst Evol 280:143–151. https://doi.org/10.1007/s00606-009-0173-9

Chong X, Zhang F, Wu Y, Yang X, Zhao N, Wang H, Guan Z, Fang W, Chen F (2016) A SNP-enable assessment of genetic diversity, evolutionary relationships and the identification of candidate genes in chrysanthemum. Genome Biol Evol 8:3661–3671. https://doi.org/10.1093/gbe/evw270

Chong X, Su J, Wang F, Wang H, Song A, Guan Z, Fang W, Jiang J, Chen S, Chen F, Zhang F (2019) Identification of favorable SNP alleles and candidate genes responsible for inflorescence-related traits via GWAS in chrysanthemum. Plant Mol Biol 99:407–420. https://doi.org/10.1007/s11103-019-00826-w

Dai SL (2004) Chinese florits's chrysanthemum and the world flower horticulture. J Hebei Norm Univ Sci Tech 18:1–6

Dai SL, Chen JY (1997) A cladistics study on some *Dendranthema spp*. in China. J Wuhan Bot Res 15:27–34

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dowrick GJ (1953) The chromosomes of Chrysanthemum II: garden varieties. Heredity 7:59–72. https://doi.org/10.1038/hdy.1953.5

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Fan M, Gao Y, Wu Z, Zhang Q (2020) Linkage map development by EST-SSR markers and QTL analysis for inflorescence and leaf traits in chrysanthemum (*Chrysanthemum morifolium* Ramat.). Plants 9:1342. https://doi.org/10.3390/plants9101342

Gacek K, Bayer PE, Bartkowiak-Broda I, Szala L, Bocianowski J, Edwards D, Batley J (2017) Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. Front Plant Sci 7:2062. https://doi.org/10.3389/fpls.2016.02062

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967. https://doi.org/10.1038/ng.695

Kan G, Zhang W, Yang W, Ma D, Zhang D, Hao D, Hu Z, Yu D (2015) Association mapping of soybean seed germination under salt stress. Mol Genet Genom 290:2147–2162. https://doi.org/10.1007/s00438-015-1066-y

Korasick DA, Chatterjee S, Tonelli M, Dashti H, Lee SG, Westfall CS, Fulton DB, Andreotti AH, Amarasinghe GK, Strader LC, Jez JM (2015) Defining a two-pronged structural model for PB1 (Phox/Bem1p) domain interaction in plant auxin responses. J Biol Chem 290:12868–12878. https://doi.org/10.1074/jbc.M115.648253

Klie M, Schie S, Linde M, Debener T (2014) The type of ploidy of chrysanthemum in not black or white: a comparison of a molecular approach to published cytological methods. Front Plant Sci 5:479. https://doi.org/10.3389/fpls.2014.00479

Li P, Zhang F, Chen S, Jiang J, Wang H, Su J, Fang W, Guan Z, Chen F (2016) Genetic diversity, population structure and association analysis in cut chrysanthemum (*Chrysanthemum morifolium* Ramat.). Mol Genet Genom 291:1117–1125. https://doi.org/10.1007/s00438-016-1166-3

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200

Nguyen TK, Lim JH (2019) Tools for chrysanthemum genetic research and breeding: Is genotyping-by-sequencing (GBS) the best approach? Hortic Environ Biotechnol 60:625–635. https://doi.org/10.1007/s13580-019-00160-6

Nguyen TK, Ha ST, Lim JH (2020) Analysis of chrysanthemum genetic diversity by genotyping-by-sequencing. Hortic Environ Biotechnol 61:903–913. https://doi.org/10.1007/s13580-020-00274-2

Park SK, Arens P, Esselink D, Lim JH, Shin HK (2015) Analysis of inheritance mode in chrysanthemum using EST-derived SSR markers. Sci Hortic 192:80–88. https://doi.org/10.1016/j.scienta.2015.05.009

Phan NT, Sim SC (2017) Genomic tools and their implications for vegetable breeding. Hortic Sci Technol 35:149–164. https://doi.org/10.12972/kjhst.20170018

Poland JA, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen 5:103–113. https://doi.org/10.3835/plantgenome2012.06.0006

Pu Y, Walley JW, Shen Z, Lang MG, Briggs SP, Estelle M, Kelley DR (2019) Quantitative early auxin root proteomics identifies GAUT10, a galacturonosyltransferase, as a novel regulator of root meristem maintenance. Mol Cell Proteom 18:1157–1170. https://doi.org/10.1074/mcp.RA119.001378

R Core Team (2014) R: A language and environment for statistical computing. R Foundation Statistical Computing, Vienna

Song C, Liu Y, Song A, Dong G, Zhao H, Sun W, Ramakrishnan S, Wang Y, Wang S, Li T, Niu Y, Jiang J, Dong B, Xia Y, Chen S, Hu Z, Chen F, Chen S (2018) The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of chrysanthemum flowers and medicinal traits. Mol Plant 11:1482–1491. https://doi.org/10.1016/j.molp.2018.10.003

Su J, Jiang J, Zhang F, Liu Y, Ding L, Chen S, Chen F (2019a) Current achievements and future prospects in the genetic breeding of chrysanthemum: a review. Hortic Res 6:1–9. https://doi.org/10.1038/s41438-019-0193-8

Su J, Zhang F, Chong X, Song A, Guan Z, Fang W, Chen F (2019b) Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. Hortic Res 6:21. https://doi.org/10.1038/s41438-018-0101-7

Sumitomo K, Shirasawa K, Isobe S, Hirakawa H, Hisamatsu T, Nakano Y, Yagi M, Ohamiya A (2019) Genome-wide association study overcomes the genome complexity in autohexaploid chrysanthemum and tags SNP markers onto the flower color genes. Sci Rep 9:1–9. https://doi.org/10.1038/s41598-019-50028-z

Sun Z, Wang X, Liu Z, Gu Q, Zhang Y, Li Z, Ke H, Yang J, Wu J, Wu L (2017) Genome-wide association study discovered genetic variation and candidate genes of fiber quality traits in *Gossypium hirsutum* L. Plant Biotechnol J 15:982–996. https://doi.org/10.1111/pbi.12693

Tang HQ, Zhang F, Chen FD, Fang WM, Wang CC, Chen SM (2015) Heterosis and mixed genetic analysis of inflorescence traits of anemone-typed chrysanthemum. Acta Hortic Sin 42:907–916. https://doi.org/10.16420/j.issn.0513-353x.2014-0433

Trapnell C, Pachter L, Saizberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. https://doi.org/10.1038/nprot.2012.016

Wang Z, Wang X, Zhang H, Ma L, Zhao H, Jones CS, Chen J, Liu G (2020) A genome-wide association study approach to the identification of candidate genes underlying agronomic traits in alfalfa (*Medicago sativa* L.). Plant Biotechnol J 18:611–613. https://doi.org/10.1111/pbi.13251

Won SY, Kwon SJ, Lee TH, Jung JA, Kim JS, Kang SH, Sohn SH (2017) Comparative transcriptome analysis reveals whole-genome duplications and gene selection patterns in cultivated and wild *Chrysanthemum* species. Plant Mol Biol 95:451–461. https://doi.org/10.1007/s11103-017-0663-z

Van Geest G (2017) Disentangling hexaploid genetics: towards DNA-informed breeding for postharvest performance in chrysanthemum. Thesis, Wageningen University, The Netherlands

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications

for crop genetics and breeding. Trends Biotechnol 27:522–530. https://doi.org/10.1016/j.tibtech.2009.05.006

Yang X, Wu Y, Su J, Ao N, Guan Z, Jiang J, Chen S, Fang W, Chen F, Zhang F (2019) Genetic variation and development of a SCAR marker of anemone-type in chrysanthemum. Mol Breed 39:1–12. https://doi.org/10.1007/s11032-019-0958-7

Yang X, Ao N, Qu Y, Wu Y, Su J, Ding L, Chen S, Jiang J, Guan Z, Chen F, Fang W, Zhang F (2020) Genetic characterization of anemone-type chrysanthemum (*Chrysanthemum morifolium*) using floral morphology and SRAP markers. Plant Breed 139:419–427. https://doi.org/10.1111/pbr.12781

Yoshioka S, Sumitomo K, Fujita Y, Yamagata A, Onozaki T, Shibata M, Ohmiya A (2010) Significance of *CmCCD4a* orthologs in apetalous wild chrysanthemum species, responsible for white coloration of ray petals. Euphytica 171:295–300. https://doi.org/10.1007/s10681-009-0070-x

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Hollang JB (2006) A unified mixed-model method for association mapping that accounts for multiple levels for relatedness. Nat Genet 38:203–208. https://doi.org/10.1038/ng1702

Zhang F, Chen S, Chen F, Fang W, Li F (2010) A preliminary genetic linkage map of chrysanthemum (*Chrysanthemum morifolium*) cultivars using RAPD, ISSR and AFLP markers. Sci Hortic 125:422–428. https://doi.org/10.1016/j.scienta.2010.03.028

Zhang F, Chen S, Chen F, Fang W, Deng Y, Chang Q, Liu P (2011) Genetic analysis and associated SRAP markers for flowering traits of chrysanthemum (*Chrysanthemum morifolium*). Euphytica 177:15–24. https://doi.org/10.1007/s10681-010-0239-3

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