

**1 Natural interploidy hybridization among the key taxa involved in the origin of**  
**2 horticultural chrysanthemums**

3

4 Running title: Interploidy hybridization among *Chrysanthemum*

5

6 Shuai Qi<sup>1</sup>, Alex D. Twyford<sup>2,3</sup>, Junyi Ding<sup>1</sup>, James S. Borrell<sup>4</sup>, Yueping Ma<sup>5</sup>, Nian  
7 Wang<sup>1,5,7\*</sup>

8

9 <sup>1</sup>State Forestry and Grassland Administration Key Laboratory of Silviculture in  
10 Downstream Areas of the Yellow River, College of Forestry, Shandong Agricultural  
11 University, Tai'an 271018, China.

12 <sup>2</sup>Institute of Evolutionary Biology, Ashworth Laboratories, the University of  
13 Edinburgh, Edinburgh EH9 3FL, UK.

14 <sup>3</sup>Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK.

15 <sup>4</sup>Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AB, UK.

16 <sup>5</sup>College of Life and Health Sciences, Northeastern University, Shenyang 110004,  
17 China.

18 <sup>6</sup>Mountain Tai Forest Ecosystem Research Station of State Forestry and Grassland  
19 Administration, College of Forestry, Shandong Agricultural University, Tai'an  
20 271018, China.

21 <sup>7</sup>State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an,  
22 China.

23 \* Author for correspondence:

24 Nian Wang; Email: [nian.wang@sdau.edu.cn](mailto:nian.wang@sdau.edu.cn)

25

## Abstract

Understanding hybridization and introgression between natural plant populations can give important insights into the origins of cultivated species. Recent studies suggest differences in ploidy may not create such strong reproductive barriers as once thought, and thus studies into cultivated origins should examine all co-occurring taxa, including those with contrasting ploidy levels. Here, we characterized hybridization between *Chrysanthemum indicum*, *Chrysanthemum vestitum* and *Chrysanthemum vestitum* var. *latifolium*, the most important wild species involved in the origins of cultivated chrysanthemums. We analysed population structure of 317 *Chrysanthemum* accessions based on 13 microsatellite markers and sequenced chloroplast *trnL-trnF* for a subset of 103 *Chrysanthemum* accessions. We identified three distinct genetic clusters, corresponding to the three taxa. We detected 20 hybrids between species of different ploidy levels, of which 19 were between *C. indicum* (4x) and *C. vestitum* (6x) and one was between *C. indicum* and *C. vestitum* var. *latifolium* (6x). Fourteen hybrids between *C. indicum* and *C. vestitum* were from one of the five study sites. *Chrysanthemum vestitum* and *C. vestitum* var. *latifolium* share only one chloroplast haplotype. The substantially different number of hybrids between hybridizing species was likely due to different levels of reproductive isolation coupled with environmental selection against hybrids. In addition, human activities may play a role in the different patterns of hybridization among populations.

**Keywords** *Chrysanthemum*, hybrid, microsatellite marker, symmetrical introgression, *trnL-trnF*

# 1. Introduction

Hybridization has played an important role in plant domestication and diversification through human history (Arnold, 2014; Heslop-Harrison and Schwarzacher, 2007; Cornille et al., 2014). Multiple important crops have been generated through hybridization either between wild species or through introgression from crop wild relatives into cultivated lineages. Major examples include modern strawberries (*Fragaria ananassa*) (Bringhurst and Voth, 1984) and triploid bananas (Simmonds and Shepherd, 1955; Heslop-Harrison and Schwarzacher, 2007), but also ornamental species such as tree peonies (*Paeonia suffruticosa*) (Zhou et al., 2014), cherry (*Prunus yedoensis*) (Baek et al., 2018) and dahlia (*Dahlia variabilis*) (Saar et al., 2003). Understanding the frequency, phylogenetic distribution and propensity for hybridization in wild populations could not only inform breeders as to the possible range of interspecific hybrids, but could also reduce the laborious, time consuming and frequently unsuccessful process of artificial crossing and *de novo* hybrid generation (Lim et al., 2008; Kuligowska et al., 2016).

Hybridization occurs more easily between species of the same ploidy level than differing ploidy levels. For example, hybridization between diploid and tetraploid species is often limited as triploid hybrids are usually inviable and less fit, preventing backcross formation (Wang et al., 2014; Zohren et al., 2016; Husband and Sabara, 2003). However, species of contrasting higher ploidy levels appear to have weaker reproductive barriers and hybridize more easily than diploids and tetraploids (Sonnleitner et al., 2013; Sutherland and Galloway, 2017). For example, within the *Campanula rotundifolia* polyploid complex, postzygotic isolation was lower in tetraploid–hexaploid species than in diploid–tetraploid crosses (Sutherland and Galloway, 2017). To date only a small number of studies have investigated

hybridization between higher ploidy levels, and the prevalence of higher level

cross-ploidy hybridization across plant families remains unclear.

*Chrysanthemum* L. (Asteraceae) provides an excellent model for studying

hybridization between high ploidy levels, with ploidy ranging from diploid ( $2n = 2x =$

18) to decaploid ( $2n = 10x = 90$ ) (Ma et al., 2015; Zhou and Wang, 1997; Tahara,

1915; Li et al., 2013; Luo et al., 2017) and with different cytotypes within species

(Chen, 2012; Yan et al., 2019; Dowrick, 1952; Dowrick, 1953; Liu et al., 2012).

*Chrysanthemum* includes approximately 37 wild species, of which 17 occur in China

(Shih and Fu, 1983) where they have captured great public interest. Multiple wild

species have been crossed by humans to generate numerous cultivars, and they are

among the most famous Chinese flowers, with significant commercial and medicinal

value (Kim and Lee, 2005; Shahrajabian et al., 2019). The evolutionary history of

polyploidy within *Chrysanthemum* is currently unknown, though some polyploids are

thought to be allopolyploid in origin (Liu et al., 2012; Chen et al., 1996) and subject

to multiple historical polyploidization events (Yang et al., 2006).

Chrysanthemums were first cultivated in China ~1600 years ago, and were later

introduced to Japan and Europe (Chen, 2012; Shih et al., 2011; Chen, 1985). Modern

cultivated chrysanthemums are mainly hexaploids and hybridization and subsequent

artificial selection are thought to give rise to numerous cultivars (Chen, 1985; Dai et

al., 2002). The ancestry of modern chrysanthemums remains elusive, but several wild

species are thought to be involved, including *C. indicum* (4x), *C. vestitum* (6x), *C.*

*lavandulifolium* (2x), *C. nankingense* (2x) and *C. zawadskii* (2x) (Chen, 1985; Dai

and Chen, 1997; Fukai, 2003; Ma et al., 2016; Ma et al., 2020). *Chrysanthemum*

*indicum* and *C. vestitum* are key species in the origin and evolution of cultivated

chrysanthemums (Dai et al., 2002; Dai et al., 1998) based on two lines of evidence.



First, ancient literature documents multiple uses for *C. indicum* in central China, which is consistent with the geographic origin of modern cultivars (see Chen, 2012). Second, artificial hybridization between *C. indicum* and *C. vestitum* can generate hybrids resembling the prototype of modern chrysanthemums (Chen, 2012).

In this study, we investigate whether hybridization occurs between *Chrysanthemum* species with different ploidy levels, focusing on *C. indicum* and *C. vestitum* as well as varieties of those species. We ask: (1) Does interploidy hybridization naturally occur between tetraploid and hexaploid *Chrysanthemum* species, which are likely involved in origin of modern cultivated chrysanthemums? (2) If hybridization occurs, is there evidence for a greater propensity at high ploidy levels? (3) Do some modern chrysanthemum cultivars share chloroplast haplotypes with the three wild *Chrysanthemum* taxa? To this end, we genotyped 317 samples at 13 microsatellite markers and sequenced chloroplast *trnL-trnF* for a subset of 103 samples. In addition, we extracted *trnL-trnF* from the chloroplast genomes of 28 taxa, representing 15 wild species of *Chrysanthemum*, 12 cultivars and one sample of *Ajania varifolium*. We analyze these data in a population genetic and phylogeography context, and use them to investigate a case where traditionally reproductive isolation caused by differences in ploidy levels would be expected to be strong.

## 2. Material and Methods

### 2.1 Study species

*Chrysanthemum indicum* is tetraploid with a wide distribution across China, though narrowly distributed diploid and hexaploid cytotypes have also been reported (Liu et al., 2012). *Chrysanthemum vestitum* is a hexaploid distributed across Hubei and Henan provinces in central China (Zhao and Chen, 1999) though the variety *C. vestitum* var. *latifolium* is hexaploid with a restricted distribution in the Dabie

Mountains in Anhui province.

The three taxa are outcrossing perennials (Chen, 2012) and are likely to be pollinated by bees (personal observations, N. Wang). Morphologically, *C. indicum* has yellow florets and smooth leaves with deep serrations. Both *C. vestitum* and *C. vestitum* var. *latifolium* have white floret and pubescent leaves and stems (Fig. 1). Putative hybrids either with an intermediate morphology or ploidy level between *C. indicum* and *C. vestitum* have been found in localities where the species co-occur (Nakata et al., 1992; Zhao and Chen, 1999). These hybrids exhibit continuous phenotypes between *C. indicum* and *C. vestitum* (Zhao and Chen, 1999), indicating the existence of putative hybrid swarms.

## 2.2 Sampling across hybridizing populations

To understand the extent and distribution of hybridization, we surveyed sympatric populations of the three focal taxa: *C. indicum*, *C. vestitum*, and *C. vestitum* var. *latifolium*. We identified and sampled five populations where *C. indicum* and *C. vestitum* co-occur and two where *C. indicum* and *C. vestitum* var. *latifolium* co-occur (Table S1). In addition, we collected *C. indicum* from two allopatric populations, TA and ZP (Table S1). Samples were collected at random within populations, ensuring at least ten meter spacing between individuals. Healthy and pest free leaf tissue was collected and stored in silica gel. A total of 317 samples were collected including between 14 and 74 from each of the seven hybridizing populations, five from TA and three from ZP (Table S1). A Global Position System (GPS, Unistrong) was used to record the coordinates of each population. Sampling locations are illustrated in Fig. 2.

## 2.3 Microsatellite genotyping

Genomic DNA was isolated from dried leaves of all individuals, following a modified 2x CTAB (cetyltrimethylammoniumbromide) protocol (Wang et al., 2013). The

quality of isolated genomic DNA was assessed on 1.0 % agarose gels, and then diluted to a concentration of 10-20 ng/ul for genotyping and sequencing. Thirteen microsatellite loci were used for genotyping (Zhang et al., 2014; Jo et al., 2015). The 5' terminus of the forward primer was labeled with FAM, HEX or TAM fluorescent probes. Each microsatellite locus was amplified individually prior to being combined into four multiplexes (Table S2). The PCR protocol follows Hu et al. (2019).

## 2.4 Population genetic analysis

It is often difficult to assign microsatellite genotypes for mixed ploidy species, as the frequency of different alleles can be difficult to quantify. In hexaploids, each microsatellite locus would be expected to have up to six alleles per individual. We chose to score each allele separately using the software GENEMARKER 2.4.0 (Softgenetics), and checked each genotype manually. We then calculated allele richness for each population using FSTAT 2.9.4 (Goudet, 1995) and performed principal coordinate (PCO) analysis in POLYSAT 1.7-4 (Clark and Jasieniuk, 2011), based on Bruvo's pairwise genetic distances (Bruvo et al., 2004).

We performed STRUCTURE analysis for each hybridizing population separately using STRUCTURE 2.3.4 (Pritchard et al., 2000) with ploidy specified as 6n. We combined XG, NX and PH into one hybridizing population as the three localities are separated by only a few kilometers. We set the number of genetic clusters (K) to 2 when analyzing the genetic structure of each hybridizing population as only two parental species are present. The allopatric *C. indicum* populations (TA and ZP) were used as a reference population.

In addition, to identify the most likely K value across populations we included all populations in a combined STRUCTURE analysis, testing K values from 1 to 10. The number of genetic clusters was estimated using the Evanno test (Evanno et al., 2005)

in the program Structure Harvester 0.6.94 (Earl and vonHoldt, 2012). Ten replicates of the STRUCTURE analysis were performed with 1,000,000 iterations and a burn-in of 100,000 for each run. The admixture model, with an assumption of correlated allele frequencies, was used. Individuals were assigned to clusters based on the highest membership coefficient averaged over the ten independent runs. Replicate runs were grouped based on a symmetrical similarity coefficient of >0.9 using the Greedy algorithm in CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007) and visualized in DISTRICT 1.1 (Rosenberg, 2004). In populations where *C. indicum* and *C. vestitum* or *C. indicum* and *C. vestitum* var. *latifolium* co-occur, we estimated Q scores in STRUCTURE with 95% confidence intervals to define pure *C. indicum*, pure *C. vestitum*, pure *C. vestitum* var. *latifolium* or putative hybrids. We distinguish individuals with confidence intervals overlapping 1 as pure *C. indicum*, with 0 as pure *C. vestitum* and those remaining as putative hybrids, in populations where the two species co-occur. Similarly, in populations where *C. indicum* and *C. vestitum* var. *latifolium* co-occur, we distinguish individuals with confidence intervals overlapping with 1 as pure *C. indicum*, with 0 as pure *C. vestitum* var. *latifolium* and those remaining as hybrids.

We compared the average allele number per individual between *C. indicum*, *C. vestitum*, *C. vestitum* var. *latifolium* and their hybrids at each microsatellite locus, using the Kruskal–Wallis test in the R package agricolae v1.3-3 (de Mendiburu, 2020). We would expect that the average allele number is higher for *C. vestitum* and *C. vestitum* var. *latifolium* than *C. indicum*. In addition, we tested if introgression was symmetric between *C. indicum* and *C. vestitum* or between *C. indicum* and *C. vestitum* var. *latifolium* in each hybridizing population, using the function ‘wilcox.test’ in R v4.0.3 (R Core Team, 2020). The putative hybrids identified as above excluded

from such comparisons.

## 2.6 *trnL-trnF* sequencing and analysis

In order to detect the potential maternal parents of the hybrids, and to estimate the haplotype diversity of these taxa, we amplified *trnL-trnF* for a subset of 103 samples, including between five and 34 individuals from each population. Reactions were performed in 20 ul volumes containing 13 ul ddH<sub>2</sub>O, 5 ul 2×Taq PCR mix (TIANGEN, China), 0.5 ul each primer (*trnL* and *trnF* (Taberlet et al., 1991)) and 1 ul DNA template. PCR products were outsourced for purification and sequencing, to Qingdao, China. Sequences were manually edited and aligned using BioEdit v7.2.5 (Hall, 1999). The R package pegas v0.14 (Paradis, 2010) was used to construct haplotype networks, using default settings, with gaps treated as a fifth state. The total number of sites, polymorphic sites, parsimony informative sites, and nucleotide diversity were computed using DnaSPv6.12.03 (Rozas et al., 2017). All sequences obtained in this study were submitted to GenBank with accession number MZ032043 - MZ032145.

To aid in a broader phylogenetic analysis, we extracted the *trnL-trnF* region from the available whole chloroplast genomes for 15 of the 17 wild species of *Chrysanthemum* occurring in China, 12 cultivars and *Ajania varifolium*. A phylogenetic tree was estimated using the maximum-likelihood method (ML) in RAxML v. 8.1.16 (Stamatakis, 2006). *Ajania varifolium* was selected as the outgroup. A rapid bootstrap analysis with 100 bootstrap replicates and 10 tree searches was performed under the GTR + GAMMA nucleotide substitution model. The consensus tree generated from the bootstrap replicates was visualized in FigTree v.1.3.1 (Rambaut and Drummond, 2009).

## 3. Results

### 3.1 Hybridization across ploidy levels inferred from microsatellites

Genetic diversity estimates were similar among the three taxa and hybrids, with allelic richness ranging from 3.74 to 4.26 and gene diversity from 0.75 to 0.84 (Table S3). On average, the number of alleles scored in *C. vestitum* and *C. vestitum* var. *latifolium* was significantly higher than *C. indicum* at eight and seven loci, respectively ( $P < 0.05$ ); this was expected for a hexaploid possessing more chromosome copies than a tetraploid species (Fig. S1).

Principal Coordinate (PCO) analysis based on Bruvo's genetic distances among all samples revealed three clusters, with coordinates 1 and 2 explaining 13.8% and 7.8% of the total variation, respectively (Fig. 2). Coordinate 1 separated *C. indicum* from *C. vestitum* and *C. vestitum* var. *latifolium* and coordinate 2 separated *C. vestitum* from *C. vestitum* var. *latifolium* (Fig. 2). Most hybrids identified by the STRUCTURE analysis fell between *C. indicum* and *C. vestitum* in the PCO plot (Fig. 2).

The combined STRUCTURE analysis across all samples identified  $K = 3$  as the optimal  $K$  value (Fig. S2a), with the three clusters corresponding to *C. indicum*, *C. vestitum* and *C. vestitum* var. *latifolium* (Fig. 3a). At  $K = 2$ , *C. indicum* formed one cluster and *C. vestitum* and *C. vestitum* var. *latifolium* formed another cluster (Fig. S2b), supporting the close relationship between these two taxa and the inference that *C. vestitum* var. *latifolium* is a subspecies of *C. vestitum*. Interestingly, there are three *C. vestitum* individuals possessing a considerable level of introgression from *C. vestitum* var. *latifolium* and one *C. indicum* individual in population TZ with substantial introgression from *C. vestitum* (Fig. 3a).

A total of 20 hybrids were detected when STRUCTURE was performed for each hybridizing population separately, of which 19 were hybrids between *C. indicum* and

*C. vestitum* and one was a hybrid between *C. indicum* and *C. vestitum* var. *latifolium*. Fourteen hybrids between *C. indicum* and *C. vestitum* were from population XG, two from each of NX and PH and one from WFS. The only hybrid between *C. indicum* and *C. vestitum* var. *latifolium* was from population TZ. No hybrids were detected from population LJS where *C. indicum* and *C. vestitum* co-occur and from BMJ where *C. indicum* and *C. vestitum* var. *latifolium* co-occur (Fig. 3b).

Introgression occurred symmetrically between *C. indicum* and *C. vestitum* and between *C. indicum* and *C. vestitum* var. *latifolium* in all hybridizing populations except WFS where only three *C. indicum* were collected (Fig. 3c). Introgression was limited in populations TZ, BMJ and LJS but was extensive in population XG (Fig. 3c).

### 3.2 Plastid diversity and directionality of hybrid formation

A total of 17 chloroplast haplotypes were detected across samples, with 12, nine, and four in *C. indicum*, *C. vestitum* and *C. vestitum* var. *latifolium*, respectively (Fig. 4; Table S4). Both *C. indicum* and *C. vestitum* harbored three private haplotypes and *C. vestitum* var. *latifolium* harbored one (haplotype H3). *Chrysanthemum indicum* shared six haplotypes with *C. vestitum* and three haplotypes with *C. vestitum* var. *latifolium*. By contrast, *C. vestitum* only shared the most common haplotype, H6, with *C. vestitum* var. *latifolium*. Frequencies of particular haplotypes varied considerably between species as some haplotypes found in *C. indicum* or *C. vestitum* were absent in *C. vestitum* var. *latifolium* (Fig. 4). Three haplotypes (H6, H16 and H17) were found in hybrids, with haplotype H6 shared among the three taxa, and haplotype H16 found to be private to the hybrid (Fig. 4). One hybrid shared haplotype H17 with *C. indicum*, whereas this haplotype is absent from all *C. vestitum* and *C. vestitum* var. *latifolium* samples (Fig. 4).



All modern *C. morifolium* cultivars formed a clade with full support, and this clade was nested in a large monophyletic clade including *C. dichrum*, *C. zawadskii*, *C. chaneitii* (Fig. 5). Five haplotypes (1, 3, 6, 16 and 17) from the three taxa and hybrid were also nested in this monophyletic clade (Fig. 5).

#### 4. Discussion

In this study, we provide genetic evidence of natural interploidy hybridization between two pairs of taxa involved in the formation of modern *Chrysanthemum* horticultural hybrids. We detected more hybrids between *C. indicum* and *C. vestitum* than between *C. indicum* and *C. vestitum* var. *latifolium*, possibly due to different levels of reproductive isolation. In addition, we show that *C. vestitum* var. *latifolium* formed a genetic cluster distinct from *C. vestitum*, and as such deserves its varietal status. Here, we first discuss the importance of hybridization between different ploidy levels in natural *Chrysanthemum* populations, before considering the dynamics of different hybrid swarms. We finish with the wider implications of our findings for understanding the origin of horticultural chrysanthemum hybrids.

##### 4.1 Tetraploid-hexaploid hybridization and symmetrical introgression

Interploidy hybridization is more common in genera containing many polyploids, and where species readily co-occur and hybridize, such as *Betula* (Hu et al., 2019; Zohren et al., 2016) and *Spartina* (Ainouche et al., 2003). Diploid-tetraploid hybridization produces mostly sterile triploids, though where hybrids are fertile introgression usually occurs preferentially from diploids to tetraploids (Moraes et al., 2013; Pinheiro et al., 2010). In contrast, hybridization between tetraploids and hexaploids has been proposed to be easier, as pentaploid hybrids are formed frequently (Hülber et al., 2015) and are more fertile than triploids (Sutherland and Galloway, 2017).

Consistent with this, we detected 20 hybrids among *C. indicum*, *C. vestitum* and *C.*



*vestitum* var. *latifolium*, indicating incomplete reproductive isolation. The number of hybrids is likely to be underestimated in our study, because of the stringent confidence threshold applied in the STRUCTURE analysis. Three individuals with admixture between 26.2%-31.4 from population XG and two individuals with admixture between 34.5%-48.7% were not supported by Q scores with 95% CIs, but may prove to be hybrids such as later generation backcrosses.

Nineteen out of the 20 hybrids were between *C. indicum* and *C. vestitum* and one was between *C. indicum* and *C. vestitum* var. *latifolium*. This difference may reflect different levels of reproductive isolation. Higher fruit set in artificial *C. indicum*-*C. vestitum* crosses than artificial *C. indicum*-*C. vestitum* var. *latifolium* crosses, partially supports this hypothesis (Zhou, 2009). However, environmental selection against hybrids between *C. indicum* and *C. vestitum* var. *latifolium* may also account for its rarity. *Chrysanthemum indicum* and *C. vestitum* occupy similar habitats and are usually intermixed in sympatric populations (Shuai Qi, personal observations). This may enhance their opportunity for hybridization, while their similar habitat preferences may reduce the chance of ecological selection on the hybrids. In contrast, *C. vestitum* var. *latifolium* and *C. indicum* are adapted to different conditions, and hybrids may fail to survive due to the breakdown of suites of co-adapted genes. Within some diploid-tetraploid systems, introgression is more common from diploids to tetraploids (Zohren et al., 2016; Wang et al., 2020). However, we observe symmetrical introgression between *C. indicum* and *C. vestitum* and between *C. indicum* and *C. vestitum* var. *latifolium*, indicating that hybrids can backcross with both parents. This is in line with recent studies showing that pentaploids can mediate gene flow between species with different ploidy levels (Peskoller et al., 2021). We couldn't distinguish clearly between male and female parental taxa because *C.*

*indicum* and *C. vestitum* share some plastid haplotypes. However, *C. indicum* can serve as maternal parent as one hybrid shared a haplotype with *C. indicum* whereas this haplotype is absent from all *C. vestitum* and *C. vestitum* var. *latifolium* samples. In addition, one hybrid had a unique haplotype (H16), which is possibly from unsampled *C. indicum* and *C. vestitum* or introgressed from other *Chrysanthemum* species.

## 4.2 Variable hybridizing among populations

The number of hybrids between *C. indicum* and *C. vestitum* varied substantially among the hybridizing populations (Fig. 3b). Fourteen out of 19 hybrids are in population XG, and no hybrids are found in population LJS. These populations are approximately 50 km apart, and differential reproductive isolation seems unlikely to account for such differences. However, we note that population LJS is closer to human habitation and human activities may have an impact on the persistence of hybrids.

Unexpectedly, one hybrid between *C. indicum* and *C. vestitum* var. *latifolium* is from population TZ and none are from BMJ. Moreover, the extent of genetic admixture in TZ seems to be higher than in BMJ (Fig. S1). In TZ, *C. indicum* and *C. vestitum* var. *latifolium* grow closely together, meaning there are enhanced opportunities for hybridization, and likely relaxed selection against hybrids. However, in BMJ, *C. indicum* and *C. vestitum* var. *latifolium* are segregated by altitude; this may limit the survival of hybrids due to environmental selection.

However, under future climate change, *C. indicum* may move to higher altitudes and come into closer contact with *C. vestitum* var. *latifolium*, producing more hybrids, as seen in population TZ. This has implications for conserving *C. vestitum* var. *latifolium*. *Chrysanthemum indicum* is widespread and abundant whereas *C. vestitum*

var. *latifolium* is restricted to the Dabie Mountains. Hybridization between abundant *C. indicum* and rare *C. vestitum* var. *latifolium* may be predicted to drive the rare species to extinction through genetic or demographic swamping (Todesco et al., 2016).

### 4.3 The presence of orphan hybrids

Hybrids usually occur in sympatry with their parental species, but sometimes they occur separately, due to (often human-mediated) long-distance dispersal of hybrids, or natural colonization of sterile hybrid taxa (James and Abbott, 2005). Alternatively parental species may die out due to genetic swamping or competitive exclusion (Huxel, 1999; Levin et al., 1996). Regardless of the mechanism, these result in orphan hybrids (Marques et al., 2010; Groh et al., 2019).

We detected a few individuals showing considerable admixture between *C. vestitum* var. *latifolium* and *C. vestitum* in populations WFS and NX where *C. vestitum* var. *latifolium* is not known to occur. We also detected one *C. indicum* individual showing considerable admixture from *C. vestitum* in population TS in Shandong province. This indicates the presence of hybrids in the absence of one or both parental species, which has been demonstrated in some plant species, such as oaks (Dodd and Afzal-Rafii, 2004) and pines (Lanner and Phillips, 1992). A plausible explanation is the existence of undetected *C. vestitum* var. *latifolium* near populations WFS and NX, or within travelling distance of pollinators or seed dispersal.

### 4.4 Implications for the origins of cultivated chrysanthemums

Our results based on an analysis of plastid *trnL-trnF* showed a monophyletic clade composed of *C. lavandulifolium*, *C. chonetii*, *C. zawadskii* and five haplotypes of *C. indicum*, *C. vestitum*, *C. vestitum* var. *latifolium* and the hybrid between *C. indicum* and *C. vestitum* (Fig. 5). This indicates that either of these species potentially acted as

the maternal parent of cultivated chrysanthemums. This is consistent with previous research implicating their involvement (Dai et al., 1998; Dai et al., 2005; Fukai, 2003). However, the specific maternal parental progenitor of modern cultivated chrysanthemums remains elusive. The modern cultivated chrysanthemums were placed in a monophyletic clade that was sister to *C. lavandulifolium*. However, the chloroplast genome of *C. lavandulifolium* has some unique mutations compared with cultivated chrysanthemums. This leads to the hypothesis that the maternal progenitor of modern cultivated chrysanthemums has gone extinct (Ma et al., 2020). However, this requires further evaluation as only one or two whole chloroplast genomes were included for each wild *Chrysanthemum* species and cultivar. This means the diversity of chloroplast genomes was not sufficiently represented. Given the high haplotype diversity of the three taxa and the fact that these haplotypes did not form a monophyletic clade (Fig. 5), we predict that the ultimate maternal progenitor of modern cultivated chrysanthemums may potentially be any of the wild *Chrysanthemum* species, and there may be multiple maternal progenitors for chrysanthemum cultivars.

### **Acknowledgements**

This work was funded by the National Natural Science Foundation of China (31770230, 31600295 and 31872710), Shandong Provincial Natural Science Foundation, China (ZR2018PC022) and Funds of Shandong “Double Tops” Program (SYL2017XTTD13).

## References

- Ainouche ML, Baumel A, Salmon A, Yannic G. 2003.** Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161**: 165–172.
- Arnold ML. 2014.** Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology* **13**: 97–1007.
- Baek S, Choi K, Kim GB, Yu HJ, Cho A, Jang H, Kim C, Kim HJ, Chang KS, Kim JH, Mun JH. 2018.** Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biology* **19**: 127.
- Bringhurst RS, Voth V. 1984.** Breeding octoploid strawberries. *Iowa State Journal of Research* **58**: 371–382.
- Bruvo R, Michiels NK, D'Souza TG, Schulenburg H. 2004.** A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* **13**: 2101–2106.
- Chen F, Chen P, Li H. 1996.** Genome analysis and their phylogenetic relationships of several wild species of *Dendranthema* in China. *Acta Horticulturae Sinica* **23**: 67–72.
- Chen JY. 1985.** Studies on the origin of Chinese florist's chrysanthemum. *Acta Horticulturae* **167**: 349–361.
- Chen JY. 2012.** *The Origin of Garden Chrysanthemum*. Hefei: Anhui Science & Technology Publishing House.
- Clark LV, Jasieniuk M. 2011.** POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources* **11**: 562–566.

421 **Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. 2014.** The  
422 domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57–  
423 65.

424 **Dai SL, Chen JY. 1997.** Cladistic study on some Chrysanthemum spp. in China.  
425 *Journal of Wuhan Botanical Research* **15**: 27– 34.

426 **Dai SL, Chen JY, Li WB. 1998.** Application of RAPD analysis in the study on the  
427 origin of Chinese cultivated chrysanthemum. *Acta Botanica Sinica* **40**: 1053–  
428 1059.

429 **Dai SL, Wang WK, Huang JP. 2002.** Advance of researches on phylogeny of  
430 *Dendranthema* and origin of chrysanthemum. *Acta Scent Nat Univ Pekinensis*  
431 **24**: 230–234.

432 **Dai SL, Wang WK, Xu YX. 2005.** Phylogenetic relationship of *Dendranthema* (DC.)  
433 Des Moul. revealed by fluorescent in situ hybridization. *Journal of Integrative*  
434 *Plant Biology* **7**: 783–791.

435 **de Mendiburu F. 2020.** agricolae: Statistical procedures for agricultural research. R  
436 package version 1.3 - 3. Available from  
437 <https://CRAN.r-project.org/package=agricolae> [accessed 30 March 2021].

438 **Dodd RS, Afzal-Rafii Z. 2004.** Selection and dispersal in a multispecies oak hybrid  
439 zone. *Evolution* **58**: 261–269.

440 **Dowrick GJ. 1952.** The chromosome of chrysanthemum (I). *Heredity* **6**: 365–375.

441 **Dowrick GJ. 1953.** The chromosomes of chrysanthemum (II) garden varieties.  
442 *Heredity* **7**: 59–72.

443 **Earl DA, vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program  
444 for visualizing STRUCTURE output and implementing the Evanno method.  
445 *Conservation Genetics Resources* **4**: 359–361.

446 **Evanno G, Regnaut S, Goudet J. 2005.** Detecting the number of clusters of  
447 individuals using the software STRUCTURE: a simulation study. *Molecular*  
448 *Ecology* **14**: 2611–2620.

449 **Fukai S. 2003.** Dendranthema species as chrysanthemum genetic resources. *Acta*  
450 *Horticulturae* **620**: 223–230.

451 **Goudet J. 1995.** FSTAT (version 1.2): a computer program to calculate F-statistics.  
452 *Journal of Heredity* **86**: 485–486.

453 **Groh JS, Percy DM, Björk CR, Cronk QCB. 2019.** On the origin of orphan hybrids  
454 between *Aquilegia formosa* and *Aquilegia flavescens*. *AoB Plants* **11**: ply071.

455 **Hülber K, Sonnleitner M, Suda J, Krejčíková J, Schönschwetter P, Schneeweiss**  
456 **GM, Winkler M. 2015.** Ecological differentiation, lack of hybrids involving  
457 diploids, and asymmetric gene flow between polyploids in narrow contact  
458 zones of *Senecio carniolicus* (syn. *Jacobaea carniolica*, Asteraceae). *Ecology*  
459 *and Evolution* **5**: 1224–1234.

460 **Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and  
461 analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:  
462 95–98.

463 **Heslop-Harrison JS, Schwarzacher T. 2007.** Domestication, genomics, and the  
464 future for banana. *Annals of Botany* **100**: 1073–1084.

465 **Hu YN, Zhao L, Buggs RJA, Zhang XM, Li J, Wang N. 2019.** Population structure  
466 of *Betula albosinensis* and *Betula platyphylla*: evidence for hybridization and  
467 a cryptic lineage. *Annals of Botany* **123**: 1179–1189.

468 **Husband BC, Sabara HA. 2003.** Reproductive isolation between autotetraploids and  
469 their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae).  
470 *New Phytologist* **161**: 703–713.

471 **Huxel GR. 1999.** Rapid displacement of native species by invasive species: effects of  
472 hybridization. *Biological Conservation* **89**: 143–152.

473 **Jakobsson M, Rosenberg NA. 2007.** CLUMPP: a cluster matching and permutation  
474 program for dealing with label switching and multimodality in analysis of  
475 population structure. *Bioinformatics* **23**: 1801–1806.

476 **James JK, Abbott RJ. 2005.** Recent, allopatric, homoploid hybrid speciation: the  
477 origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid  
478 zone on Mount Etna, Sicily. *Evolution* **59**: 2533–2547.

479 **Jo KM, Jo Y, Chu H, Lian S, Cho WK. 2015.** Development of EST-derived SSR  
480 markers using next-generation sequencing to reveal the genetic diversity of 50  
481 chrysanthemum cultivars. *Biochemical Systematics and Ecology* **60**: 37–45.

482 **Kim HJ, Lee YS. 2005.** Identification of new dicaffeoylquinic acids from  
483 *Chrysanthemum morifolium* and their antioxidant activities. *Planta Medica* **71**:  
484 871–876.

485 **Kuligowska K, Lütken H, Müller R. 2016.** Towards development of new  
486 ornamental plants: status and progress in wide hybridization. *Planta* **244**: 1–



487 17.

488 **Lanner RM, Phillips AMI. 1992.** Natural hybridization and introgression of pinyon

489 pines in northwestern Arizona. *International Journal of Plant Sciences* **153**:

490 250–257.

491 **Levin DA, Francisco-Ortega J, Jansen RK. 1996.** Hybridization and the extinction

492 of rare plant species. *Conservation Biology* **10**: 10–16.

493 **Li J, Wan Q, Abbott RJ, Rao GY. 2013.** Geographical distribution of cytotypes in

494 the *Chrysanthemum indicum* complex as evidenced by ploidy level and

495 genome - size variation. *Journal of Systematics and Evolution* **51**: 196–204.

496 **Lim KB, Barba-Gonzalez R, Zhou S, Ramanna MS, van Tuyl JM. 2008.**

497 Interspecific hybridization in Lily (*Lilium*): Taxonomy and commercial

498 aspects of using species hybrids in breeding. In: da Silva JAT, ed. *Floriculture,*

499 *ornamental and plant biotechnology*. Japan: Global Science Books.

500 **Liu PL, Wan Q, Guo YP, Yang J, Rao GY. 2012.** Phylogeny of the genus

501 *Chrysanthemum* L.: evidence from single-copy nuclear gene and chloroplast

502 DNA sequences. *PloS ONE* **7**: e48970.

503 **Luo C, Chen D, Cheng X, Zhao H, Huang C. 2017.** Genome size estimations in

504 *Chrysanthemum* and correlations with molecular phylogenies. *Genetic*

505 *Resources and Crop Evolution* **64**: 1451–1463.

506 **Ma Y, Wei J, Yu Z, B. Q, Dai S. 2015.** Characterization of ploidy levels in

507 *Chrysanthemum* L. by flow cytometry. *Journal of Forestry Research* **26**: 771–

508 775.

509 **Ma Y, Zhao L, Zhang W, Zhang Y, Xing X, Duan X, Hu J, Harris AJ, Liu P, Dai**  
510 **S, Wen J. 2020.** Origins of cultivars of *Chrysanthemum* - evidence from the  
511 chloroplast genome and nuclear *LFY* gene. *Journal of Systematics and*  
512 *Evolution* **58**: 925–944.

513 **Ma YP, Chen MM, Wei JX, Zhao L, Liu PL, Dai SL, Wen J. 2016.** Origin of  
514 *Chrysanthemum* cultivars — Evidence from nuclear low-copy *LFY* gene  
515 sequences. *Biochemical Systematics and Ecology* **65**: 129–136.

516 **Marques I, Feliner GN, Draper Munt D, Martins-Loução MA, Aguilar JF. 2010.**  
517 Unraveling cryptic reticulate relationships and the origin of orphan hybrid  
518 disjunct populations in *Narcissus*. *Evolution* **64**: 2353–2368.

519 **Moraes AP, Chinaglia M, Clarisse Palma - Silva C, Pinheiro F. 2013.** Interploidy  
520 hybridization in sympatric zones: the formation of *Epidendrum fulgens*  $\times$  *E.*  
521 *puniceoluteum* hybrids (Epidendroideae, Orchidaceae). *Ecology and Evolution*  
522 **3**: 3824 – 3837.

523 **Nakata M, Hong D, Qiu J, Uchiyama H, Tanaka R, Chen X. 1992.** Cytogenetic  
524 studies on wild *Chrysanthemum* sensu lato in China. II. A natural hybrid  
525 between *Dendranthema indicum* ( $2n = 36$ ) and *D. vestitum* ( $2n = 54$ ) from  
526 Hubei Province. *Journal of Japan Botany* **67**: 92–100.

527 **Paradis E. 2010.** pegas: an R package for population genetics with an integrated–  
528 modular approach. *Bioinformatics* **26**: 419–420.

529 **Peskoller A, Silbernagl L, Hülber K, Sonnleitner M, Schönswetter P. 2021.** Do  
530 pentaploid hybrids mediate gene flow between tetraploid *Senecio disjunctus*

- 531 and hexaploid *S. carniolicus* s. str. (*S. carniolicus* aggregate, Asteraceae)?
- 532 *Alpine Botany*.
- 533 **Pinheiro F, de Barros F, Palma-Silva C, Meyer D, Fay MF, Suzuki RM, Lexer C,**
- 534 **Cozzolino S. 2010.** Hybridization and introgression across different ploidy
- 535 levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum*
- 536 (Orchidaceae). *Molecular Ecology* **19**: 3981–3994.
- 537 **Pritchard JK, Stephens M, Donnelly P. 2000.** Inference of population structure
- 538 using multilocus genotype data. *Genetics* **155**: 945–959.
- 539 **R Core Team. 2020.** R: a language and environment for statistical computing.
- 540 Version 4.0.3. Vienna, Austria: R Foundation for Statistical Computing.
- 541 **Rambaut A, Drummond AJ. 2009.** FigTree version 1.3.1 [computer program].
- 542 <http://tree.bio.ed.ac.uk>.
- 543 **Rosenberg NA. 2004.** DISTRUCT: a program for the graphical display of population
- 544 structure. *Molecular Ecology Notes* **4**: 137–138.
- 545 **Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P,**
- 546 **Ramos-Onsins SE, Sánchez-Gracia A. 2017.** DnaSP 6: DNA sequence
- 547 polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34**:
- 548 3299–3302.
- 549 **Saar DE, Polans NO, Sørensen PD. 2003.** A phylogenetic analysis of the genus
- 550 *Dahlia* (Asteraceae) based on internal and external transcribed spacer regions
- 551 of nuclear ribosomal DNA. *Systematic Botany* **28**: 627–639.
- 552 **Shahrajabian MH, Sun W, Zandi P, Cheng Q. 2019.** A review of chrysanthemum,

553 the eastern queen in traditional chinese medicine with healing power in  
554 modern pharmaceutical sciences. *Applied Ecology and Environmental*  
555 *Research* **17**: 13355–13369.

556 **Shih C, Christopher JH, Michael GG. 2011.** *Chrysanthemum*. In: Wu ZY, Raven  
557 PH, Hong DY, eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri  
558 Botanical Garden Press.

559 **Shih C, Fu GX. 1983.** *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press.

560 **Simmonds NW, Shepherd K. 1955.** The taxonomy and origins of the cultivated  
561 bananas. *Botanical Journal of the Linnean Society* **55**: 302–312.

562 **Sonnleitner M, Weis B, Flatscher R, Escobar Garcia P, Suda J, Krejčíková J,**  
563 **Schneeweiss GM, Winkler M, Schönswetter P, Hülber K. 2013.** Parental  
564 ploidy strongly affects offspring fitness in heteroploid crosses among three  
565 cytotypes of autopolyploid *Jacobaea carniolica* (Asteraceae). *PloS ONE* **8**:  
566 e78959.

567 **Stamatakis A. 2006.** RAxML-VI-HPC: maximum likelihood-based phylogenetic  
568 analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–  
569 2690.

570 **Sutherland BL, Galloway LF. 2017.** Postzygotic isolation varies by ploidy level  
571 within a polyploid complex. *New Phytologist* **213**: 404–412.

572 **Taberlet P, Gielly L, Pautou G, Bouvet J. 1991.** Universal primers for amplification  
573 of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**:  
574 1105–1109.

575 **Tahara M. 1915.** Cytological studies in Chrysanthemum (A preliminary note). *Bot*  
576 *Mag Tokyo* **29**: 48–50.

577 **Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S,**  
578 **Heredia SM, Hahn MA, Caseys C, Bock DG, Rieseberg LH. 2016.**  
579 Hybridization and extinction. *Evolutionary Applications* **9**: 892–908.

580 **Wang J, Dong S, Yang L, Harris A, Schneider H, Kang M. 2020.** Allopolyploid  
581 speciation accompanied by gene flow in a tree fern. *Molecular Biology and*  
582 *Evolution* **37**: 2487–2502.

583 **Wang N, Borrell JS, Bodles WJA, Kuttapitiya A, Nichols RA, Buggs RJA. 2014.**  
584 Molecular footprints of the Holocene retreat of dwarf birch in Britain.  
585 *Molecular Ecology* **23**: 2771–2782.

586 **Wang N, Thomson M, Bodles WJA, Crawford RMM, Hunt HV, Featherstone**  
587 **AW, Pellicer J, Buggs RJA. 2013.** Genome sequence of dwarf birch (*Betula*  
588 *nana*) and cross-species RAD markers. *Molecular Ecology* **22**: 3098–3111.

589 **Yan W, Jung JA, Lim K, Cabahug RAM, Hwang Y. 2019.** Cytogenetic studies of  
590 chrysanthemum: a review. *Flower Research Journal* **27**: 242–253.

591 **Yang W, Glover BJ, Rao GY, Yang J. 2006.** Molecular evidence for multiple  
592 polyploidization and lineage recombination in the *Chrysanthemum indicum*  
593 polyploid complex (Asteraceae). *New Phytologist* **171**: 875–886.

594 **Zhang Y, Dai SL, Hong Y, Song XB. 2014.** Application of genomic SSR locus  
595 polymorphisms on the identification and classification of Chrysanthemum  
596 cultivars in China. *PloS ONE* **9**: e104856.

597 **Zhao HE, Chen JY. 1999.** Investigation and studies on some native species of  
598 *Dendranthema*. *Chinese Landscape Architecture* **3**: 60–61.

599 **Zhou J. 2009.** *Studies on the problem of origin of Chinese garden chrysanthemum*,  
600 Beijing Forestry University, Beijing.

601 **Zhou SJ, Wang JW. 1997.** The cytological study on ten species of *Dendranthema*.  
602 *Journal of Wuhan Botanical Research* **15**: 289–292.

603 **Zhou SL, Zou XH, Zhou ZQ, Liu J, Xu C, Yu J, Wang Q, Zhang DM, Wang XQ,**  
604 **Ge S, Sang T, Pan KY, Hong DY. 2014.** Multiple species of wild tree peonies  
605 gave rise to the 'king of flowers', *Paeonia suffruticosa* Andrews. *Proceedings*  
606 *of the Royal Society B: Biological Sciences* **281**: 1687–1694.

607 **Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, Buggs RJA.**  
608 **2016.** Unidirectional diploid-tetraploid introgression among British birch trees  
609 with shifting ranges shown by restriction site-associated markers. *Molecular*  
610 *Ecology* **25**: 2413–2426.

611

612

613 **Tables**

614 **Table S1** Detailed sampling information for natural populations of *C. indicum*, *C.*  
615 *vestitum* and *C. vestitum* var. *latifolium*.

616 **Table S2** Details of microsatellite markers used in this study.

617 **Table S3** Genetic diversity of 14 populations of *C. indicum*, *C. vestitum* and *C.*  
618 *vestitum* var. *latifolium* based on microsatellite markers.

619 **Table S4** The chloroplast haplotypes and their variable sites based on *trnL-trnF*  
620 sequences.

621

## Figure legends

**Fig. 1** Diversity present in wild *Chrysanthemum* species included in this study.

Individual photographs and leaf morphology for *C. indicum* (a), *C. vestitum* (b) and *C. vestitum* var. *latifolium* (c).

**Fig. 2** Principal coordinate analysis of *Chrysanthemum* samples based on 13 microsatellite markers. Hybrids were defined based on Q scores in a STRUCTURE analysis (see text).

**Fig. 3** Hybridization among *C. indicum*, *C. vestitum* and *C. vestitum* var. *latifolium*. (a)

Hybridization across natural populations in a combined analysis of *C. indicum*, *C.*

*vestitum* and *C. vestitum* var. *latifolium* with K = 3. (b) Admixture in sympatric

population of *C. indicum* and *C. vestitum* and each sympatric population of *C.*

*indicum* and *C. vestitum* var. *latifolium*. Barplots on the left or the right of

STRUCTURE plot represent the number of individuals having different levels of

genetic admixture. Populations XG, NX and PH were analyzed together due to their

short geographic distance. Allopatric populations TA and ZP were served as a control.

Blue, green and orange represent *C. indicum*, *C. vestitum* and *C. vestitum* var.

*latifolium*, respectively. (c) Admixture value for each population among *C. indicum*, *C.*

*vestitum* and *C. vestitum* var. *latifolium*. Hybrids were excluded for such comparisons.

**Fig. 4** Haplotype network based on plastid *trnL-trnF* sequences in natural

*Chrysanthemum* populations. (a) The geographic distribution of haplotypes. Number

in brackets indicates the number of samples used for sequencing; (b) Haplotype

network graph. Each haplotype is represented by a circle with size proportional to the

number of individuals. Color within the circle represents species sharing the

haplotype.

**Fig. 5** The phylogenetic relationships of *Chrysanthemum* species inferred from the



647 chloroplast *trnL-trnF* region. Branch support values above 50% are depicted.

648 Haplotypes identified in the present study are indicated in brackets. Species names

649 before the identified haplotypes represent species sharing the haplotype. Colored in

650 blue, green, orange and purple represent *C. indicum*, *C. vestitum*, *C. vestitum* var.

651 *latifolium* and hybrids, respectively.

652 **Fig. S1** Allele number per individual at each of the 13 microsatellite loci among *C.*

653 *indicum*, *C. vestitum*, *C. vestitum* var. *latifolium* and hybrids. The difference in the

654 number of allele was assessed using Kruskal–Wallis test. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P <$

655 0.001.

656 **Fig. S2** The output of Structure Harvester showing that  $K = 3$  is the optimal value (a)

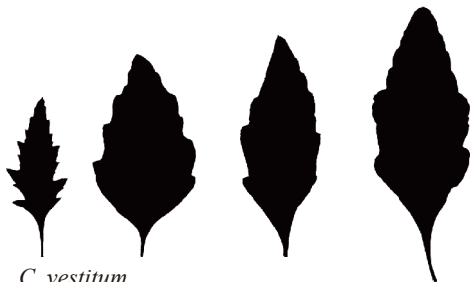
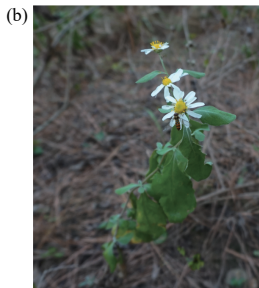
657 and STRUCTURE results at  $K = 2$  and 3 (b).

658

659



*C. indicum*



*C. vestitum*



*C. vestitum* var. *latifolium*

2 cm

