


# Development of Microsatellite Markers for a Dioecious *Herpetospermum pedunculosum* (Cucurbitaceae)

Zhu-Qing Chen<sup>1\*</sup>, Zhi-Li Zhou<sup>2\*</sup>, Lin-Lin Wang<sup>2</sup>, Li-Hua Meng<sup>1</sup>   
and Yuan-Wen Duan<sup>2</sup>

<sup>1</sup>School of Life Sciences, Key Laboratory of Yunnan for Biomass Energy and Biotechnology of Environment, Yunnan Normal University, Kunming, P.R. China. <sup>2</sup>The Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, P.R. China.

Evolutionary Bioinformatics  
Volume 16: 1–6  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1176934320908261



**ABSTRACT:** Understanding the evolution of flower diversity is a central topic in plant evolutionary ecology, and natural selection on floral traits via male fitness could be estimated quantitatively using microsatellites. Here, based on RNA sequencing, we developed simple sequence repeat primers and verified polymorphisms in 2 wild populations of *Herpetospermum pedunculosum* (Cucurbitaceae), a dioecious annual plants native to the Himalaya Mountains. A total of 131 paired primers were designed; 15 paired primers were found to be polymorphic, with the expected heterozygosity varying between 0.280 and 0.767. We also identified 58 genotypes in 20 plants from the 2 populations. Conclusively, these primers could be effective in examining male fitness and population genetic structure of *H. pedunculosum* in future studies.

**KEYWORDS:** transcriptome sequencing, SSR, paternity analysis, male fitness, floral traits

**RECEIVED:** January 8, 2020. **ACCEPTED:** January 24, 2020.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the National Natural Science Foundation of China (31660109).

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Li-Hua Meng, School of Life Sciences, Key Laboratory of Yunnan for Biomass Energy and Biotechnology of Environment, Yunnan Normal University, Kunming 650500, P.R. China. Email: menglihua@mails.ucas.ac.cn

## Introduction

The incredible diversity of flowers is attractive to plant evolutionary biologists since Darwin<sup>1</sup> proposed the theory of natural selection in 1859, and, since then, floral evolution of plant species became one of the central topics in plant evolutionary ecology. Generally, floral diversity of flowering plants is strongly associated with their dominant pollinators, with aims of pollen dispersal from anthers and pollen receipt to stigmas.<sup>2,3</sup> Thus, floral diversity could be attributed to the achievements of the maximum of male and female reproductive success of plant species. However, to the present, most studies measuring natural selection on floral traits are based on female reproductive success (seed set),<sup>4</sup> which could result from the easy estimations of seed set. In contrast, examinations of natural selection are still rare on floral traits based on male reproductive success, although achieving maximum of male reproductive success is considered the primary driver to the selection of floral traits.<sup>5,6</sup> Estimations of male reproductive success could be performed by tracing pollen flows with fluorescent powder and dying pollen grains.<sup>7</sup> However, the number of pollen grains that deposited on the stigma might not be associated with seed production.<sup>8</sup> Accordingly, by tracing pollen grains, male reproductive success could be estimated only but not be quantified. The development of molecular markers could help us quantify male reproductive success.<sup>9,10</sup> For example, a significant stabilizing selection was found for corona width and flower length through male function based on microsatellite markers in the tristylous daffodil *Narcissus triandrus*.<sup>11</sup> In *Polemonium brandegeei*, plants with more nectar sugar and narrow corolla tubes had high siring success via male fitness.<sup>12</sup> Therefore, microsatellites are the most powerful markers to determine male reproductive success quantitatively.<sup>13</sup>

\* These authors contributed equally to this work.

*Herpetospermum pedunculosum* is an annual liana native to Himalaya Mountains inhabiting altitudes ranging from 2300 to 3500 m.<sup>14</sup> This species is dioecious, and thus pollinators are necessary for seed production. Seeds of *H. pedunculosum* are widely used in traditional Tibetan medicine. Infield populations, sex ratios are strongly male-biased (ca. 70%),<sup>15</sup> indicating that male-male competition might occur in pollinating female flowers. In our field observations, we found that flower size changed significantly among different plants, indicating that large flowers could attract more pollinators than small flowers. Therefore, to measure the effects of flower size on male reproductive success quantitatively in future researches, we developed microsatellite markers of *H. pedunculosum* based on RNA sequencing in this research.

## Materials and Methods

### Materials and sample collections

Seeds of *H. pedunculosum* were collected from Shangri-La Alpine Botanical Garden in 2017 and 50 of them were sowed in pots separately in the greenhouse of Yunnan Normal University in 2018. Seedlings with 2 euphylla were transplanted to the experimental lands in Yunnan Normal University, and frames were built separately for each seedling to meet their climbing habits. All seedlings were watered periodically to prevent drought-induced death. When plants were in the flowering time, we collected fresh leaves from 2 male plants and 2 female plants and kept them in liquid nitrogen separately. Besides, in each of the 2 wild populations in Shangri-La (Yunnan) and Nyingchi (Tibet), respectively, fresh leaves of 10 plants (5 male and 5 female plants) were collected and kept in silica gel separately.

RNA of each of the 4 individuals was extracted using a CTAB method,<sup>16</sup> and the integrity of RNA was measured



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA integrity number (RIN) of each of the 4 individuals ranged from 9.0 to 10.0, showing no sign of degradation. Total DNA of dried leaves from 20 plants was extracted separately using a modified CTAB method for validation of simple sequence repeat (SSR) markers.

### RNA sequencing and assembly

The cDNA library of each individual was built following Illumina's recommendations (San Diego, CA, USA). Random hexamer primers were used to synthesize the double strands, and the QIAquick PCR Purification kit (Qiagen Inc) was employed to purify the short fragments. Then paired-end reads were generated on the Illumina HiSeq2500 platform at the Biomarker Technologies Co, China.

After filtering the adaptors and reads with  $>8$  ambiguous bases and  $>50\%$  of the bases (quality score  $\leq 5$ ) in raw data of each individual, all clean reads of 4 individuals were pooled and then de novo assembled to transcripts using Trinity software.<sup>17</sup> To evaluate the quality of transcriptome assembly, we used BUSCO v4<sup>18</sup> software with plant ortholog data sets from OrthoDB v10<sup>19</sup> to assess completeness. The clean sequence data reported in this paper have been deposited in the Genome Sequence Archive<sup>20</sup> in BIG Data Center,<sup>21</sup> Beijing Institute of Genomics (BIG), Chinese Academy of Sciences.

### SSR locus search, primer design, and validation

Before detecting SSR loci in transcripts, we followed several criteria in light of SSR's traits. (1) There should be 5 and 3 repeats at least for simple and complex repeats, respectively. (2) The length of motif should be 2 to 10bp, and the interruption distance between different SSRs should be a maximum of 100bp for complex SSRs. (3) Only 2 to 6 nucleotide motifs were considered, and the minimum repeat unit was defined as 6 for di-, and 5 for tri-, tetra-, penta-, and hexanucleotides. Then, potential SSR loci were scanned in the transcripts using Micro-Satellite (MISA).<sup>22</sup>

By using Primer 3.0,<sup>23</sup> we designed paired primers for each unique SSR containing at least 5 repeats. We then performed polymerase chain reaction (PCR) in a 25- $\mu$ L volume containing 20 to 30 ng DNA. The PCR reactions were carried out with the following conditions: DNA initial denaturation at 94°C for 4 minutes, 35 cycles of 94°C for 1 minute 30 seconds, annealing temperature ranging from 45°C to 60°C for 50 seconds, 72°C for 50 seconds, and a final extension at 72°C for 7 minutes. By using a TIAN quick Midi Purification Kit (Tiangen Biotech (Beijing) Co, Ltd, China), excess primers and deoxynucleotide triphosphates were removed to purify the PCR products. We employed those paired primers that were successful in the PCR amplification to detect the polymorphism among 18 to 20 individuals from 2 populations with POPGEN v1.32.<sup>24</sup> Sequencing reactions were then performed under the instruction of ABI Prism Sequencing Ready Reaction Kit with the same primers as PCRs and analyzed on the ABI 3730 genetic analyzer (Applied Biosystems).

**Table 1.** Characteristics of de novo samples and clustered transcripts.

ITEM	CLEAN READS	Q30	GC CONTENT, %
Female plant 1	30 620 799	95.33	46.13
Female plant 2	29 180 298	95.50	46.22
Male plant 1	30 598 513	95.65	46.53
Male plant 2	26 227 138	95.52	45.73
N50	1860		
No. of transcripts after assembly	254 706		
Mean length of transcripts	848.39		
Total identified SSRs	18 510		
Sequences containing more than 1 SSR	2434		
SSR-containing sequences with BLASTx hit	12 516		
SSR-containing sequences with GO annotation	9183		
SSR-containing sequences with KEGG annotation	924		

Abbreviation: SSR, simple sequence repeat.

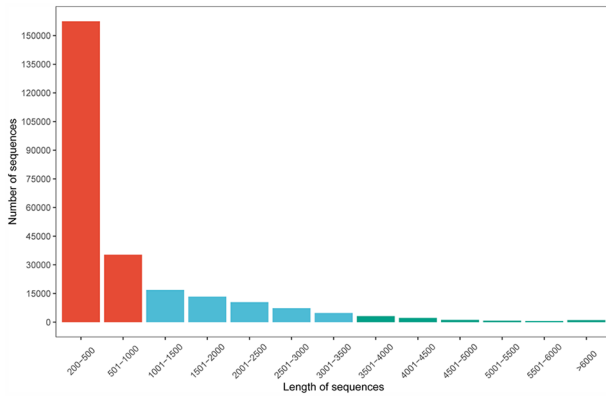
### Annotation for transcripts containing SSRs

We determined the objective sequences and gene names by the transcripts that included SSR in the National Center for Biotechnology Information's (NCBI) NR protein database using BLASTx by setting the E-value threshold as  $1e-6$ . We then conducted functional annotation of transcripts using the programs Blast2GO<sup>25</sup> and KEGG,<sup>26</sup> and classified the functional categories using the program WEGO.<sup>27</sup>

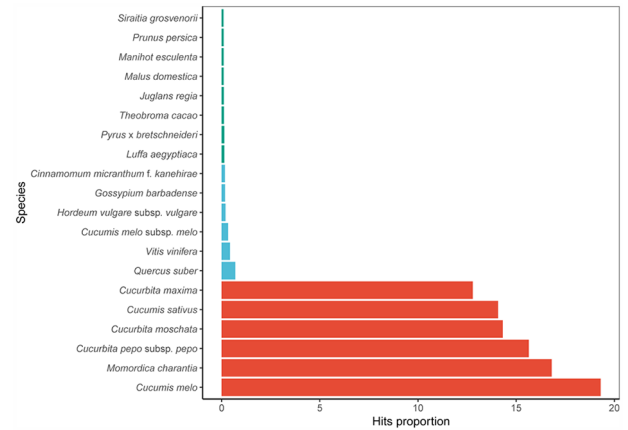
## Results and Discussion

### RNA sequencing and assembly

After filtering, 30 598 513 and 26 227 138 clean reads were acquired for 2 male plants, and 30 620 799 and 29 180 298 clean reads were obtained for 2 female plants (Table 1). Assembly resulted in 254 706 transcripts, and the length of assembled transcripts ranged from 200 to above 6000bp, with an average of 848bp (Figure 1, Table 1). The N50 value of our assembly is 1860bp and is similar to those from other species by transcriptome sequencing (Table 1), including *Veratrum baillonii* Franch<sup>28</sup> and *Halenia elliptica* D. Don.<sup>29</sup> The GC content is more than 45% for both male and female plants (Table 1). With the increase in transcripts size, the number of transcripts decreased, indicating a power-law-like distribution (Figure 1), which is a common trait of transcripts in many plant species. Furthermore, the transcriptome of *H. pedunculatum* included 350 of the 425 (82.35%) complete BUSCO genes, indicating the high quality of our assembly.



**Figure 1.** Length distribution of all transcripts in *Herpetospermum pedunculosum*.



**Figure 2.** Matching results by BLASTx for *Herpetospermum pedunculosum*.

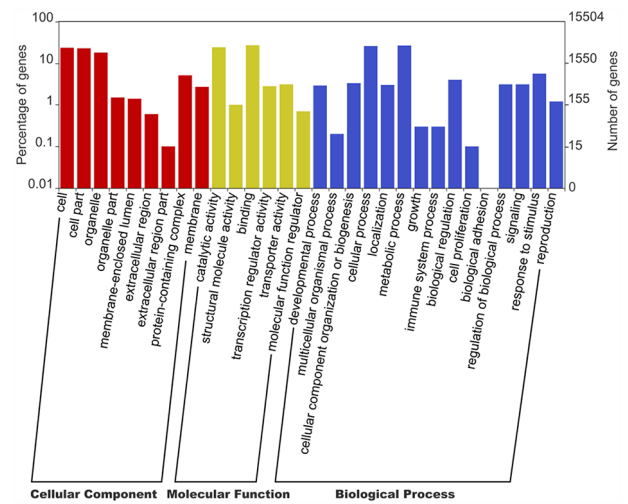
**Table 2.** Characteristics of SSR.

TYPE OF REPEAT	MAIN REPEAT MOTIF	NO.	TOTAL	FREQUENCY, %
Dinucleotide	AC/GT	227	4780	25.8
	AG/CT	2485		
	AT/AT	2063		
	CG/CG	5		
Trinucleotide	AAC/GTT	888	12 777	69.0
	AAG/CTT	5735		
	AAT/ATT	1820		
	ACC/GGT	742		
	ACG/CGT	335		
	ACT/AGT	146		
	AGC/CTG	739		
	AGG/CCT	855		
	ATC/ATG	829		
	CCG/CGG	688		
	Tetranucleotide			
Pentanucleotide			198	1.1
Hexanucleotide			221	1.2
Total			18 510	

Abbreviation: SSR, simple sequence repeat.

### Distribution and annotation of SSRs

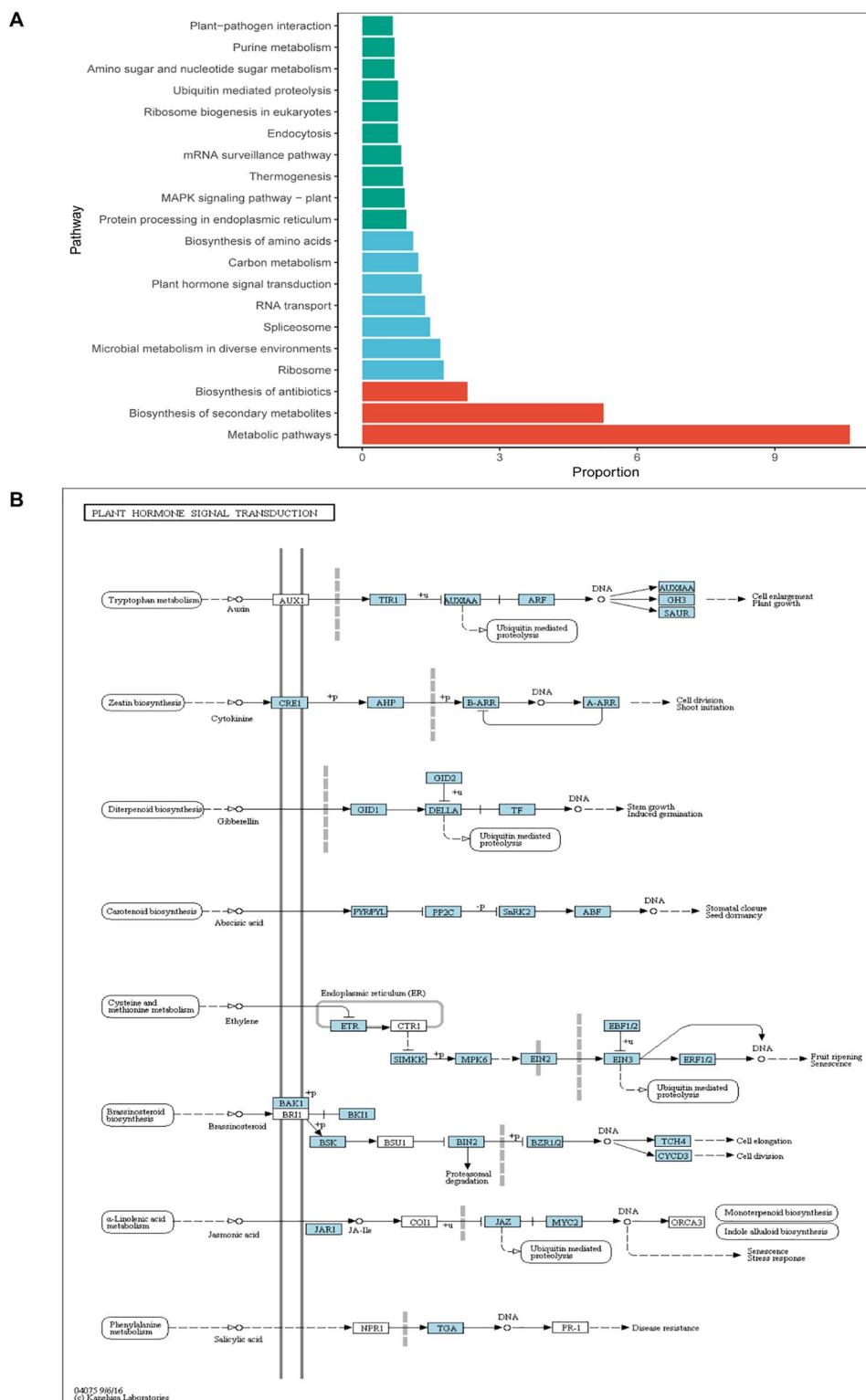
In total, 18 510 potential SSRs were identified, and 95% of them were trinucleotide and dinucleotide as we did not consider mononucleotide in SSR searching (Tables 1 and 2). Generally, all potential SSR loci could be classified into 2 categories in light of the size of genetic markers. Those with length more



**Figure 3.** GO classification of SSRs for *Herpetospermum pedunculosum*. SSRs indicate simple sequence repeat.

than 20bp were hypervariable markers and could be considered as class I. Those with a length between 12 and 20bp might be variable markers potentially and could be considered as class II. SSRs of class I could be more variable than those of class II and thus might be with high polymorphism. In the potential SSRs of *H pedunculosum*, 32.6% belonged to class I.

Results of BLAST2GO suggested that 9183 transcripts containing SSRs of *H pedunculosum* could be annotated against the GO database (Table 1, Figure 2) and classified them into 3 categories: cellular components, molecular functions, and biological process. Specifically, of the cellular components, cell, cell part and organelle part were 3 highly represented classes. Catalytic activity and binding were the most matched among 6 molecular functions. In contrast, cell and metabolic processes were the most matched 2 classes of biological processes (Figure 3). However, results of KEGG annotation suggested that only 924 transcripts including SSRs of *H pedunculosum* could be annotated. In the top 20 KEGG pathways, genes including SSRs associated with metabolic pathways (10.63%) were most representative, followed by biosynthesis of secondary



**Figure 4.** The KEGG pathway of *Herpetospermum pedunculosum*: (A) frequency distribution of the KEGG pathway functions and (B) the plant hormone signal transduction pathway.

metabolites (5.26%) and biosynthesis of antibiotics (2.30%) (Figure 4A). Besides, several genes containing SSRs participated in ethylene and jasmonic acid signal transduction pathways, which may involve in flower sex determination of *H pedunculosum*<sup>30-32</sup>(Figure 4B).

### Polymorphism of microsatellites

To validate the polymorphism of screened SSRs using transcriptome sequencing, we designed 131 paired primers of *H pedunculosum* and performed PCR in 20 plants from 2 distinct

**Table 3.** Polymorphism of microsatellites screening between 2 populations in *Herpetospermum pedunculatum*.

LOCUS	REPEAT	NO.	NA	NE	HO	HE	PIC
BLG16	(TCT)n	20	2	1.406	0.250	0.296	0.2471
BLG34	(AGA)n	20	2	1.835	0.300	0.467	0.3515
BLG37	(CAA)n	19	3	1.819	0.263	0.462	0.3488
BLG46	(TC)n	20	3	2.204	0.050	0.560	0.4422
BLG81	(ATC)n	19	2	1.634	0.211	0.398	0.3126
BLG82	(CCGCCT)n	19	3	1.375	0.211	0.280	0.2473
BLG84	(AGA)n(GGC)n	20	4	2.614	0.300	0.633	0.5421
BLG90	(ATTGG)n	20	2	1.956	0.150	0.501	0.3693
BLG101	(TC)n	18	5	3.224	0.278	0.710	0.6365
BLG108	(CTT)n(AATT)n	19	4	3.945	0.579	0.767	0.6992
BLG110	(TTG)n	20	2	1.956	0.150	0.501	0.3693
BLG111	(AGA)n	20	4	1.507	0.300	0.345	0.3058
BLG119	(GAA)n	19	3	2.087	0.316	0.535	0.4552
BLG122	(CT)n	20	2	1.882	0.450	0.481	0.3589
BLG126	(GAA)n	20	2	1.780	0.250	0.450	0.3425
Mean		19.5	2.8	2.082	0.271	0.492	0.4019
Total		292	42				

Abbreviations: He, expected heterozygosity; Ho, observed heterozygosity; No., number of samples; Na, number of alleles; Ne, effective number of alleles; PIC, polymorphism information content.

populations (Supplementary Table 1). We then sequenced the PCR products and found that 15 paired primers were polymorphic. The expected heterozygosity per locus varied between 0.280 and 0.767 in total, with a range between 0 and 0.621 for the Yunnan population and a range between 0 and 0.574 for the Tibet population, respectively. We also identified 58 genotypes in 20 plants from the 2 populations (Table 3).

### Conclusions

Microsatellites have increasingly been used in population genetics analysis and mating system estimation for their abundance in genomes, genetic codominance, high reproducibility and polymorphism,<sup>22</sup> and rapid development of next-generation sequencing advance the wide uses of SSRs. Here, we developed 15 paired SSR primers using transcriptome sequencing that showed polymorphic in 2 populations of *H. pedunculatum*. These SSR primers could be useful in future quantification of male reproductive success and paternity analysis of this dioecious plant.

### Author Contributions

ZQC and ZLZ performed the laboratory experiments and statistical analyses. LLW and YWD collected plant materials in the field and cultivated plants in the green house. LHM

designed the research and wrote the manuscript. All authors read and reviewed the final manuscript.

### Data Availability

The clean sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers CRA002098 that are publicly accessible at <http://bigd.big.ac.cn/gsa>.

### ORCID iD

Li-Hua Meng  <https://orcid.org/0000-0002-7098-4389>

### Supplemental Material

Supplemental material for this article is available online.

### REFERENCES

1. Darwin CR. *On the Origin of Species by Means of Natural Selection*. London, England: John Murray; 1859.
2. Harder LD, Johnson SD. Darwin's beautiful contrivances: evolutionary and functional evidence for floral adaptation. *New Phytol.* 2009;183:530-545.
3. Barrett SCH. Darwin's legacy: the forms, function and sexual diversity of flowers. *Philos Trans R Soc Lond B Biol Sci.* 2010;365:351-368.
4. Sapir Y, Armbruster WS. Pollinator-mediated selection and floral evolution: from pollination ecology to macroevolution. *New Phytol.* 2010;188:303-306.

5. Barrett SCH. Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Phil Trans R Soc B*. 2003;358:991-1004.
6. Zhang GP, Meng LH, Wu ZK, et al. Natural selection on floral traits of *Caltha scapoza* (Ranunculaceae), an alpine perennial with generalized pollination system from Northwest Yunnan. *Plant Diversity*. 2017;39:202-207.
7. Sage TL, Husband BCB. *Plant breeding system and pollen dispersal*. In: Dafni A, Kevan PG, Husband BC, eds. *Practical Pollination Biology*. Cambridge, ON, Canada: Enviroquest Ltd; 2005:27-82.
8. Wang H, Cao GX, Wang LL, Yang YP, Zhang ZQ, Duan YW. Evaluation of pollinator effectiveness based on pollen deposition and seed production in a gynodioecious alpine plant, *Cyananthus delavayi*. *Ecol Evol*. 2017;7:8156-8160.
9. Shimono A, Wang XR, Torimaru T, Lindgren D, Karlsson B. Spatial variation in local pollen flow and mating success in a *Picea abies* clone archive and their implications for a novel "breeding without breeding" strategy. *Tree Genet Genomes*. 2011;7:499-509.
10. Austen EJ, Weis AE. The causes of selection on flowering time through male fitness in a hermaphroditic annual plant. *Evolution*. 2016;70:111-125.
11. Hodgins KA, Barrett SCH. Natural selection on floral traits through male and female function in wild populations of the heterostylous daffodil *Narcissus triandrus*. *Evolution*. 2008;62:1751-1763.
12. Kulbaba MW, Worley AC. Selection on floral design in *Polemonium brandegeei* (Polemoniaceae): female and male fitness under hawkmoth pollination. *Evolution* 2012;66:1344-1359.
13. Christopher DA, Mitchell RJ, Karron JD. Pollination intensity and paternity in flowering plants. *Ann Bot*. 2019;125:1-9. doi:10.1093/aob/mcz159.
14. Lu A, Huang L, Chen S, Jeffrey C. Cucurbitaceae. In: Wu ZY, Raven PH, eds. *Flora of China*. Beijing, China and St. Louis, MI: Science Press and Missouri Botanical Garden; 2011:1-56.
15. Lan X, Guan F, Chao W, Zhong G, Bao L. Effect of sowing-date and marginal utility on sexual differentiation and yield of *Herpetospermum pedunculosum*. *J Northeast Agr Univ*. 2011;42:79-82.
16. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep*. 1993;11:113-116.
17. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644-652.
18. Seppy M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, ed. *Gene Prediction: Methods in Molecular Biology*. New York, NY: Humana Press; 2019:227-245.
19. Kriventseva EV, Kuznetsov D, Tegenfeldt F, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. 2018; 47:D807-D811.
20. Wang YQ, Song FH, Zhu JW, et al. GSA: genome sequence archive. *Genom Proteom Bioinform*. 2017;15:14-18.
21. Zhang Z. Database resources of the BIG data center in 2019. *Nucleic Acids Res*. 2019;47:D8-D14.
22. Sharma PC, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol*. 2007;25:490-498.
23. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000;132:365-386.
24. Yeh FC, Yang RC, Boyle T, Ye ZH, Mao JX. *POPGENE, Version 1.32: The User Friendly Software for Population Genetic Analysis*. Edmonton, AB, Canada: Molecular Biology and Biotechnology Centre, University of Alberta; 1999.
25. Conesa A, Gotz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008;2008:619832.
26. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007; 35:W182-W185.
27. Ye J, Fang L, Zheng H, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006;34:W293-W297.
28. Wang L, Wang ZK, Chen JB, et al. De novo transcriptome assembly and development of novel microsatellite markers for the traditional Chinese medicinal herb, *Veratrum baillonii* Franch (Gentianaceae). *Evol Bioinform Online*. 2015;11: 39-45.
29. Yang ML, Han NY, Li H, Meng LH. Transcriptome analysis and microsatellite markers development of a traditional Chinese medicinal herb *Halenia elliptica* D. Don (Gentianaceae) [published online ahead of print July 25, 2018]. *Evol Bioinform Online*. doi:10.1177/1176934318790263.
30. Trebitsh T, Staub JE, O'Neill SD. Identification of a 1-aminocyclopropane-1-carboxylic acid synthase gene linked to the female (F) locus that enhances female sex expression in cucumber. *Plant Physiol*. 1997;113:987-995.
31. Yin T, Quinn JA. Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am J Bot*. 1995;82:1537-1546.
32. Acosta IF, Laparra H, Romero SP, et al. tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science*. 2009;323: 262-265.