Applications in Plant Sciences

PRIMER NOTE

Development of simple sequence repeat markers for *Chamerion angustifolium* (Onagraceae)

Rui Mi[1], Tongcheng Wang[1], Derek W. Dunn[1], Kang Huang[1,3], and Baoguo Li[1,2]

Manuscript received 2 December 2018; revision accepted 9 January 2019.

[1] Shaanxi Key Laboratory for Animal Conservation, College of Life Sciences, Northwest University, Xi'an 710069, People's Republic of China

[2] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, People's Republic of China

[3] Author for correspondence: huangkang@nwu.edu.cn

Citation: Mi, R., T. Wang, D. W. Dunn, K. Huang, and B. Li. 2019. Development of simple sequence repeat markers for *Chamerion angustifolium* (Onagraceae). *Applications in Plant Sciences* 7(5): e1244.

doi:10.1002/aps3.1244

PREMISE: Rosebay willowherb, or fireweed (*Chamerion angustifolium*: Onagraceae), has diploid, tetraploid, and hexaploid cytotypes. There are known physiological and ecological differences among the three cytotypes, but genetic differences remain undetermined. We developed simple sequence repeat (SSR) markers for this species.

METHODS AND RESULTS: Leaf samples were collected from three hexaploid *C. angustifolium* populations. We successfully amplified 16 SSR loci, which were found to be highly polymorphic. The number of alleles, the observed heterozygosity levels, and the expected heterozygosity levels ranged from four to 13, 0.286–0.899, and 0.372–0.871, respectively. Most primers could also be amplified successfully in *C. conspersum* and the closely related species *Epilobium palustre*.

CONCLUSIONS: The 16 polymorphic SSR markers developed here will be useful for genetic studies in *C. angustifolium* and related species.

KEY WORDS allelic phenotypes; *Chamerion angustifolium*; Onagraceae; polyploidy; simple sequence repeats.

Rosebay willowherb, or fireweed (*Chamerion angustifolium* (L.) Holub: Onagraceae), is a protandrous, insect-pollinated perennial herb. It has three different cytotypes: diploids, tetraploids, and hexaploids (Mosquin, 1966, 1967). Ploidy level is known to facilitate outcrossing in this species (Routley and Husband, 2003). Hardy and drought tolerant, *C. angustifolium* is common and widely distributed across temperate Europe, Asia, and North America (Mosquin, 1966, 1967; Chen et al., 1988, 2007). Different *C. angustifolium* ploidy levels are associated with environmental factors; for example, diploid populations occur in higher latitudes and at higher elevations, whereas tetraploids occur in generally more temperate environments, and hexaploids favor warmer climates (Mosquin, 1966, 1967; Chen et al., 1988, 2007; Husband and Schemske, 1998). In *C. angustifolium*, certain morphological traits such as plant height, pollen size, and the number of pollen pores are also known to be associated with different ploidy levels (Mosquin, 1967). Due to its cytotype polymorphism, *C. angustifolium* has been used as a model organism to study divergence among cytotypes, including adaptation to variation in elevation (Martin and Husband, 2013), microclimate (Thompson et al., 2014), ecophysiological responses to drought (Guo et al., 2016), and the effects of ploidy on reproductive isolation (Husband and Schemske, 2000).

We have developed simple sequence repeat (SSR) markers for *C. angustifolium* to explore genetic differentiation among its three cytotypes. However, several genotypes for each polyploid that share the same electrophoresis band type are indistinguishable from each other. For example, when two alleles (*A* and *B*) are present at a locus in an autotetraploid individual, there are three types of heterozygotes (i.e., *AAAB*, *AABB*, and *ABBB*) that can produce the same band type (i.e., *AB*). This is referred to as "genotyping ambiguity" (Huang et al., 2014). Although there are some methods that can be used to determine the genotype of heterozygous polyploids (Gidskehaug et al., 2011; Serang et al., 2012; Uitdewilligen et al., 2013), this requires additional equipment or next-generation sequencing. Here, we report a newly developed population genetic method to enable accurate analysis for polyploids. This is directly based on the estimation of genetic diversity indices from the allelic phenotypes (i.e., electrophoresis band types).

METHODS AND RESULTS

Genomic DNA was extracted from 10 dried leaf samples of *C. angustifolium*. We used double restriction endonuclease *Eco*RI and *Mse*I digestions to break the genomic DNA into numerous short DNA fragments. From these DNA fragments, DNA segments with simple short repeats were captured by the magnetic bead enrichment protocol (MagneSphere Magnetic Separation products; Promega

Corporation, Beijing, China). A DNA library was thus established from the DNA segments. DNA fragments were used as templates for PCR amplification. The products were purified by electrophoresis and linked to the pMD18-T vector (TaKaRa Biotechnology Co., Dalian, Liaoning, China). We added the conjugates to the competent cells of *Escherichia coli*, and the DNA fragments from the ligation products were cloned. The positive clones were then isolated from the clone products, and the clones containing the target sequences were selected for sequencing. Using Chromas version 2.2.4 (Technelysium, South Brisbane, Australia), we checked and, if necessary, manually corrected each sequence peak map and then selected the sequences containing SSR motifs. We used DNAMAN version 5.2.2 (Lynnon Biosoft, San Ramon, California, USA) to remove the sequences of the adapters, and CodonCode Aligner version 3.5 (CodonCode Corporation, Centerville, Massachusetts, USA) to perform sequence alignment analysis. After removing the repetitive sequences, SSR primer pairs were designed using Primer Premier 5 (Lalitha, 2000) with the following settings: primer length 18–22 bp, product size 100–300 bp, and the remaining parameters set as default. In total, 63 pairs of primers were designed. We carried out a preliminary test using 10 individuals to exclude unamplifiable, monomorphic, or incorrectly amplified primer pairs. The alleles were segregated using polyacrylamide gel electrophoresis. Of the 63 primers, eight pairs were unamplifiable, 33 had amplified polymorphic bands, and 22 exhibited nonspecific amplification. We further selected 16 primer pairs with high numbers of alleles, high intensity of electrophoresis bands, and low stutter.

To confirm the polymorphism of these 16 primer pairs, 50 individuals of *C. angustifolium* were collected from three populations in Shaanxi Province, China (Taibai County [$n = 20$], Langao County [$n = 19$], and Chang'an County [$n = 11$]; Appendix 1). The sample leaf tissues were cut in 1-cm^2 pieces, immersed in a 500-μL nuclei extraction buffer (CyStain UV Precise P; Sysmex Partec, Münster, Germany), and quickly chopped to release nuclei using a double-edged blade. The isolated nuclei were stained with stain buffer (CyStain UV Precise P; Sysmex Partec), whose absorbance was measured using a flow cytometer (CyFlow; Sysmex Partec). By comparing the absorbance of *C. angustifolium* to that of a related diploid species (*Epilobium hirsutum* L.), the cytotypes of all individuals were calculated.

These 16 selected fluorescent primer pairs (Table 1) were amplified for capillary electrophoresis to genotype. The allelic phenotypes were manually determined from the electropherogram peaks. For reference purposes, these same 16 primer pairs were also amplified in five samples each of *E. palustre* L. and *C. conspersum* (Hausskn.) Holub (Appendix 1).

Due to genotype ambiguity (Huang et al., 2014), the true genotypes of polyploids cannot be revealed by electrophoresis band types from PCR-based codominant markers. We developed a method to

TABLE 1. Characteristics of 16 polymorphic simple sequence repeat markers developed for *Chamerion angustifolium*.

Locus	Primer sequences (5′–3′)	Repeat motif	Expected allele size (bp)	T_a (°C)	Fluorescent label[a]	GenBank accession no.
WK-P01	F: ACAGGATTAGCAGAGAAGCAA R: GAAATACCGAATAGAGGAAGA	$(CTAT)_{20}$	171	50.6	TAMRA	MK153166
WK-P02	F: ACGATACAATAGACATGGACA R: GCTAAAAATAGGCAGTAACTC	$(CTCA)_{11}$	251	48.9	HEX	MK153167
WK-P03	F: ATTCTTGGGGAGCGGTGGTAA R: TTGAACCGACCAACCGGCCAT	$(GAGT)_5$	107	60.7	FAM	MK153168
WK-P10	F: GAATTCCAAACTCAGAAGATC R: CGTGGAGTAAAAGAGAAGTGG	$(AG)_{20}$	218	50.4	HEX	MK153169
WK-P11	F: CTGAGAAATAAATGAAGGTGG R: GAGATCAAAGATACAAAGAGC	$(AG)_9$	232	47.8	HEX	MK153170
WK-P20	F: TGAACAACATTTTTGCACAGT R: CACCCCTCCCTCTAGTAGAGT	$(GA)_{13}$	134	53.8	FAM	MK153171
WK-P25	F: CCTCCATTGTCATCACAGTAGA R: ACCATATTCAAACACCTATCCA	$(GA)_{12}$	137	52.0	FAM	MK153172
WK-P27	F: ACTTTCCGAACTCTGAGGTGGG R: GGTGCATCATGCTTCAAGGTGT	$(AG)_{12}$	164	58.8	TAMRA	MK153173
WK-P32	F: CGGCGACAACAGTGATGCTTAC R: CCGATCTCTCATTTTCGTGGAA	$(AG)_{12}$	126	56.9	FAM	MK153174
WK-P34	F: AATCGTATCTGACGGACTTGGA R: CGTCTACCGAAACACCATTACA	$(GA)_{16}$	158	54.8	FAM	MK153175
WK-P35	F: GCGTTTCCAGTTTCGTAACCAC R: TCCTCCATGCTGATCTCCTCAT	$(CT)_{12}$	108	57.2	FAM	MK153176
WK-P38	F: GATCAAAGATACAAAGAGCCA R: GGTGAAGAATGAAATCAAGAA	$(CT)_{11}$	213	48.2	HEX	MK153177
WK-P41	F: GACTAGTGAATACCCGACCGA R: AGCAGAGATTTGCAGTAATGG	$(CT)_{15}$	141	54.2	TAMRA	MK153178
WK-P43	F: ACTTCGTTCCCACTTCCTTCG R: ATGTAAACCCATTCGTTCTTG	$(GA)_{12}$	171	53.7	HEX	MK153179
WK-P44	F: GAATTCACATGATAGCGGAGC R: TAGAGCGTTGATTACTGAGGC	$(CT)_{12}$	106	53.8	HEX	MK153180
WK-P58	F: ATAGACTAGAGTTTAGGGTTT R: CCACTACTGTGTGCTTGTGAC	$(AC)_{11}$	143	51.2	TAMRA	MK153181

Note: T_a = annealing temperature.
[a]Fluorescent label placed at the 5′ end of each primer.

TABLE 2. Polymorphism of the 16 simple sequence repeat sites in three populations of *Chamerion angustifolium*.[a]

	Langao County population (n = 19)					Chang'an County population (n = 11)					Taibai County population (n = 20)				
Locus	A	H_o	H_e	PIC	A_R	A	H_o	H_e	PIC	A_R	A	H_o	H_e	PIC	A_R
WK-P01	8	0.746	0.742	0.699	3.881	9	0.737	0.723	0.682	3.612	9	0.755	0.746	0.713	3.932
WK-P02	11	0.809	0.838	0.819	6.169	13	0.754	0.773	0.756	4.397	10	0.776	0.782	0.755	4.597
WK-P03	4	0.585	0.562	0.465	2.282	5	0.768	0.735	0.688	3.772	6	0.782	0.759	0.728	4.153
WK-P10	11	0.729	0.750	0.715	4.006	9	0.662	0.809	0.784	5.239	13	0.711	0.690	0.675	3.222
WK-P11	11	0.771	0.746	0.709	3.931	9	0.709	0.805	0.784	5.135	11	0.869	0.860	0.845	7.123
WK-P20	5	0.646	0.623	0.575	2.655	8	0.674	0.729	0.701	3.686	7	0.710	0.684	0.657	3.161
WK-P25	12	0.843	0.801	0.774	5.013	13	0.815	0.871	0.858	7.737	10	0.825	0.844	0.827	6.416
WK-P27	9	0.810	0.794	0.768	4.861	7	0.685	0.708	0.681	3.429	9	0.625	0.638	0.616	2.762
WK-P32	12	0.822	0.839	0.820	6.229	9	0.743	0.843	0.825	6.352	9	0.781	0.832	0.810	5.951
WK-P34	12	0.899	0.840	0.820	6.269	5	0.732	0.793	0.760	4.837	7	0.807	0.771	0.738	4.368
WK-P35	9	0.286	0.372	0.360	1.593	8	0.429	0.473	0.458	1.897	11	0.766	0.806	0.790	5.161
WK-P38	13	0.813	0.792	0.767	4.800	11	0.763	0.815	0.792	5.410	12	0.835	0.856	0.840	6.923
WK-P41	7	0.806	0.756	0.718	4.106	8	0.833	0.802	0.772	5.045	9	0.827	0.806	0.781	5.145
WK-P43	7	0.689	0.680	0.628	3.129	8	0.719	0.711	0.673	3.465	11	0.738	0.801	0.778	5.037
WK-P44	5	0.762	0.733	0.686	3.750	5	0.644	0.629	0.561	2.695	5	0.680	0.665	0.615	2.981
WK-P58	5	0.639	0.676	0.636	3.087	6	0.829	0.764	0.724	4.232	6	0.768	0.751	0.710	4.009

Note: A = number of alleles per locus per population; A_R = allelic richness per locus within a population ($1 / \sum p_i^2$); H_e = expected heterozygosity; H_o = observed heterozygosity; n = number of individuals sampled; PIC = polymorphic information content.
[a]Voucher and location information are provided in Appendix 1.

TABLE 3. Cross-amplification of 16 simple sequence repeat markers developed for *Chamerion angustifolium* in *Epilobium palustre* and *C. conspersum*.[a,b]

	Epilobium palustre (N = 5)		*Chamerion conspersum* (N = 5)	
Locus	Amplification success	Expected allele size	Amplification success	Expected allele size
WK-P01	+ (3)	100	+ (4)	100
WK-P02	+ (3)	250	+ (2)	250
WK-P03	+ (1)	175	+ (5)	100
WK-P10	+ (2)	175	+ (4)	175
WK-P11	–	–	–	–
WK-P20	+ (1)	125	+ (2)	125
WK-P25	–	–	–	–
WK-P27	+ (2)	250	–	–
WK-P32	+ (3)	125	+ (5)	125
WK-P34	+ (3)	175	+ (4)	175
WK-P35	+ (5)	100	+ (4)	100
WK-P38	+ (2)	250	+ (1)	250
WK-P41	+ (5)	250	+ (3)	125
WK-P43	+ (1)	250	–	–
WK-P44	+ (4)	100	+ (5)	100
WK-P58	+ (2)	125	+ (1)	100

Note: + = successful amplification in all individuals; – = unsuccessful amplification; N = number of samples tested.
[a]Voucher and location information are provided in Appendix 1.
[b]Numbers in parentheses represent the number of amplified individuals.

estimate the genetic diversity indices directly from the allelic phenotypes (Appendix 2). This method extracts the possible genotypes from the phenotype, then estimates the allele frequencies by an iterative algorithm developed by Kalinowski and Taper (2006); the genetic diversities are calculated from the allele frequencies or from the extracted genotypes by using their posterior probabilities as the weight.

Flow cytometry showed that all individuals in the three populations of *C. angustifolium* are hexaploids. The results of microsatellite genotyping show that the number of alleles is between four and 13 in all three sampled populations of *C. angustifolium* (Table 2), and the levels of observed and expected heterozygosity range from 0.286 to 0.899 and from 0.372 to 0.871, respectively. The polymorphism information content values of the 16 loci exceed 0.3, and allelic richness is also high in all populations, denoting high polymorphism of the developed SSR markers. These parameters also suggest high levels of polymorphism among the 16 loci. Table 3 lists the 14 primers that successfully amplifed in *E. palustre* and the 12 that successfully amplifed in *C. conspersum*.

CONCLUSIONS

We report on the development of 16 new polymorphic SSR markers for *C. angustifolium*. These markers will be useful for future population genetic studies of this species, especially those investigating the effects of both cytotype and geographical distance on genetic differentiation among populations.

ACKNOWLEDGMENTS

This study was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020302), the National Natural Science Foundation of China (31770411, 31730104, 31572278), the Young Elite Scientists Sponsorship Program by CAST (2017QNRC001), the National Key Programme of Research and Development, the Ministry of Science and Technology of China (2016YFC0503202), and the Natural Science Basic Research Plan in Shaanxi Province of China (2018JM3024). D.W.D. is supported by a Shaanxi Province Talents 100 Fellowship.

AUTHOR CONTRIBUTIONS

K.H. and B.G.L. designed the project; R.M., K.H., D.W.D., and T.C.W. collected the samples; R.M. and T.C.W. performed the experiments; and R.M., K.H., and D.W.D. wrote the manuscript.

DATA ACCESSIBILITY

Sequence information for the developed primers has been deposited to the National Center for Biotechnology Information (NCBI); GenBank accession numbers are provided in Table 1.

LITERATURE CITED

Butruille, D. V., and L. S. Boiteux. 2000. Selection-mutation balance in polysomic tetraploids: Impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. *Proceedings of the National Academy of Sciences, USA* 97: 6608–6613.

Chen, C.-J., P. C. Hoch, and P. H. Raven. 1988. Variation patterns and systematics of the widespread species *Epilobium angustifolium* L. (Onagraceae) in China. *Journal of Systematics and Evolution* 26: 81–95.

Chen, J. R., P. C. Hoch, P. H. Raven, D. E. Boufford, and W. L. Wagner. 2007. Flora of China, vol. 13. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Gidskehaug, L., M. Kent, B. J. Hayes, and S. Lien. 2011. Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics* 27: 303–310.

Guo, W., J. Yang, X. D. Sun, G. J. Chen, Y. P. Yang, and Y. W. Duan. 2016. Divergence in eco-physiological responses to drought mirrors the distinct distribution of *Chamerion angustifolium* cytotypes in the Himalaya-Hengduan Mountains Region. *Frontiers in Plant Science* 7: 1329.

Huang, K., K. Ritland, S. T. Guo, M. Shattuck, and B. G. Li. 2014. A pairwise relatedness estimator for polyploids. *Molecular Ecology Resources* 14: 734–744.

Huang, K., T. Wang, D. W. Dunn, P. Zhang, X. Cao, R. Liu, and B. Li. 2019. Genotypic frequencies at equilibrium for polysomic inheritance under double-reduction. *Genes, Genomes, Genetics* https://doi.org/10.1534/g3.119.400132.

Husband, B. C., and D. W. Schemske. 1998. Cytotype distribution at a diploid–tetraploid contact zone in *Chamerion* (*Epilobium*) *angustifolium* (Onagraceae). *American Journal of Botany* 85: 1688–1694.

Husband, B. C., and D. W. Schemske. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. *Journal of Ecology* 88: 689–701.

Kalinowski, S. T., and M. L. Taper. 2006. Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics* 7: 991–995.

Lalitha, S. 2000. Primer Premier 5. *Biotech Software and Internet Report* 1(6): 270–272.

Martin, S. L., and B. C. Husband. 2013. Adaptation of diploid and tetraploid *Chamerion angustifolium* to elevation but not local environment. *Evolution* 67: 1780–1791.

Mosquin, T. A. 1966. A new taxonomy for *Epilobium angustifolium* L. *Brittonia* 18: 167–188.

Mosquin, T. A. 1967. Evidence for autopolyploidy in *Epilobium angustifolium* (Onagraceae). *Evolution* 21: 713–719.

Routley, M. B., and B. C. Husband. 2003. The effect of protandry on siring success in *Chamerion angustifolium* (Onagraceae) with different inflorescence sizes. *Evolution* 57: 240–248.

Serang, O., M. Mollinari, and A. A. Garcia. 2012. Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. *PLoS ONE* 7: e30906.

Thompson, K. A., B. C. Husband, and H. Maherali. 2014. Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 101: 1868–1875.

Uitdewilligen, J. G., A. M. Wolters, B. B. D'Hoop, T. J. A. Borm, R. G. Visser, and H. J. van Eck. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8: e62355.

APPENDIX 1. Geographic locations and voucher information for species used in this study.

Species	Location	Population code	Voucher specimen accession no.[a]	Geographic coordinates	*n*	Ploidy
Chamerion angustifolium (L.) Holub	Langao County, China	LG	WNU-LG-2017-HK-19	32.068070°N, 108.829920°E	19	Hexaploid
	Chang'an County, China	CA	WNU-CA-2017-HK-11	33.842662°N, 108.785287°E	11	Hexaploid
	Taibai County, China	TB	WNU-TB-2017-HK-20	34.387486°N, 107.209336°E	20	Hexaploid
C. conspersum (Hausskn.) Holub	Urumqi County, China		WNU-WLMQ-2018-HK-02	43.108578°N, 87.060690°E	5	Diploid
Epilobium palustre L.	Taibai County, China		WNU-TB-2018-HK-01	34.038509°N, 107.619368°E	5	Diploid

Note: *n* = sample size.
[a]Vouchers deposited at the Herbarium of the College of Life Sciences (WNU), Northwest University, Xi'an, China. HK = Kang Huang, collector.

APPENDIX 2. Analysis of genetic diversity based on allelic phenotypes in polysomic inheritance.

The possible genotypes hidden behind each allelic phenotype are weighted by their genotypic frequencies. The posterior probability of the *i*th candidate genotype (G_i) can be derived by the Bayes equation:

$$\Pr(G_i|\mathcal{P}) = \frac{\Pr(\mathcal{P}|G_i)\,\Pr(G_i)}{\Pr(\mathcal{P})} = \frac{\Pr(G_i)}{\sum_i \Pr(G_i)},$$

where $\Pr(G_i)$ and $\Pr(\mathcal{P})$ are the frequencies of G_i and $\mathcal{P}$, respectively, $\Pr(\mathcal{P}|G_i)$ is the probability that $\mathcal{P}$ is the allelic phenotype of the individual with a genotype of G_i, then $\Pr(\mathcal{P}|G_i)$ is equal to one. We ignore any double-reduction (Butruille and Boiteux, 2000) and assume the alleles randomly appear within the genotype. Therefore:

$$\Pr(G) = \nu! \prod_{i=k}^{k} \frac{p_k^{nk}}{n_k!}.$$

Here, ν is the level of ploidy, K is the number of alleles, n_i is the number of copies of the *i*th allele in G, and p_i is the frequency of the *i*th allele.

The allele frequencies are estimated with an iterative algorithm modified from Kalinowski and Taper (2006): the frequency of each allele is assigned with an initial value of $1/K$, then iteratively updated until it is converged. The updated frequency $\hat{p}'_k$ is calculated by:

$$\hat{p}'_k = \frac{\sum_{ji} \Pr(G_i|\mathcal{P}_j)\,\Pr(A_k|G_i)}{\sum_{ji} \Pr(G_i|\mathcal{P}_j)},$$

where $\Pr(A_k \mid G_i)$ is the frequency of the *k*th alleles in G_i. Based on the estimated allele frequency, the genetic diversity indices can be calculated, including allelic richness, polymorphic information content, and both observed and expected heterozygosities. The heterozygosity in polysomic inheritance is defined as the probability of sampling two distinct alleles from a genotype without replacement. Our method has been implemented in Polygene version 1.0 (Huang et al., 2019; https://github.com/huangkang1987/polygene/).