Development and characterization of microsatellite loci for the perennial herb *Helleborus foetidus* (Ranunculaceae)

Miyuki M. Seino*, Clara de Vega, Pilar Bazaga, Mónica Medrano and Carlos M. Herrera

Estación Biológica de Doñana, CSIC, Avda. Américo Vespucio s/n, 41092, Sevilla, Spain.

**** Corresponding author:***

Miyuki M. Seino

Estación Biológica de Doñana, CSIC

Avda. Américo Vespucio s/n, 41092, Sevilla, Spain.

Fax: + 34 954621125

e-mail: miyuki@ebd.csic.es

***Keywords*:** *Helleborus foetidus*, Ranunculaceae, microsatellites, 454 pyrosequencing, plant-pollinator mutualism

Running title: Microsatellites for *Helleborus foetidus*

Words in text: 2356

Abstract

Microsatellite loci were developed for *Helleborus foetidus* by next-generation 454 pyrosequencing of a microsatellite-enriched library. We selected 60 primer pairs, from which 13 microsatellite markers exhibited polymorphism and 13 were monomorphic. The number of alleles per locus ranged from two to eight (mean=3.92) and expected heterozygosity ranged from 0.05 to 0.77. One locus showed heterozygote deficiency due to the presence of null alleles, and no pairs of loci demonstrated linkage disequilibrium. The newly developed markers will enable population genetic studies and address specific questions involving mating system and pollen dispersal strategies in this species.

Main text

Helleborus foetidus L. (Ranunculaceae) is a diploid (Dhooghe *et al.* 2009) perennial herb widely distributed across Western Europe that grows in the understory of deciduous and mixed forests (Werner & Ebel 1994). Flowering takes place over winter and early spring. Flowers are hermaphroditic, protogynous, self-compatible and extremely long lived (up to 20 days; Vesprini & Pacini 2000). *Helleborus foetidus* is mainly pollinated by bumblebees, which forage for the abundant, sucrose-rich nectar this species offers (Vesprini *et al.* 1999). The reproductive biology and ecology of *H. foetidus* have been thoroughly investigated (e.g. Herrera *et al.* 2001; Rey *et al.* 2006 and references therein), but the lack of suitable molecular markers have so far hampered crucial studies on genetic diversity and gene flow. The microsatellite markers presented here will allow us to conduct specific research on how biotic factors, such as the presence of nectar yeasts in floral nectar (see Herrera *et al.* 2008; Herrera & Pozo 2010; Herrera *et al.* in press) may affect pollen dispersal strategies, mating pattern and seed progeny quality in this species.

For isolation of microsatellite loci, a high-throughput method coupling microsatellite DNA enrichment and 454 GS-FLX Titanium next-generation sequencing was carried out by Genoscreen (Lille, France) as described by Malausa *et al.* (2011), with modifications regarding oligonucleotide probes. Sequences of the oligonucleotide probes used were as follows: TG, TC, AAC, AAG, AGG, ACG, ACAT, ACTC. The resulting sequences were analysed according to the pipeline implemented in QDD (Méglecz *et al.* 2010), yielding primer sequences for 1010 potential microsatellite loci. We then selected 60 primer pairs for amplification trials based on primer design and suitability of expected fragment sizes for multiplexing purposes.

Total genomic DNA was extracted with Qiacube using DNeasy Plant Mini Kit (Qiagen) following manufacturer's protocol, from 20-25 mg sample of dried leaf material ground to fine powder in a Retsch MM 200 mill. Concentration and quality of DNA extracts were estimated through spectrophotometry using Nanodrop (Thermo Scientific). For initial simplex amplification tests we used 15 *H. foetidus* individuals collected from six natural populations located in our study area in Sierra de Cazorla (Jaén province, southeastern Spain): Coto del Valle, Castellón de los Cierzos, Rastrillos de la Víbora, Fuente Bermejo, La Cabrilla and Roblehondo. Fluorescent labelling of PCR products in simplex PCR was done by attaching a 'M13-tail'(5'-CAGTCGGGCGTCATCA- 3') to the 5' end of the forward primers, following the method described by Schuelke (2000). Simplex PCR reactions were performed in 20 µL volume containing 1× BSA (New England Biolabs), 1× reaction buffer, 3.5 mM $MgCl_2$, 0.5 U *Taq* DNA polymerase (Bioline), 0.25 mM dNTP (Sigma-Aldrich), 0.3 µM of each 'M13-tailed' forward primer, reverse primer and universal 'M13' primer tagged to the 5' end with one of FAM, VIC, NED or PET (Applied Biosystems), and c.45 ng of DNA extract

as template. Simplex PCR conditions consisted of 95°C for 3 min, 19 cycles of 95°C for 1 min, 58°C for 30 sec of annealing with temperature decreasing 0.5°C every cycle, 72°C for 30 sec, followed by 19 cycles of 95°C for 1 min, 48°C for 30 sec and 72°C for 30 sec. A final extension step consisted in 72°C for 10 min.

Primer pairs providing polymorphic, consistently scorable allele peaks obtained with simplex PCR conditions were further tested in 60 individuals from 3 populations of the same region (N = 20 individuals per population), distances between sites ranging from 5 to11 km: Tejerina (TEJ; 37.98° N, 2.91° W, 740 m.a.s.l.), Las Navillas (NAV; 37.94° N, 2.91° W, 1235 m.a.s.l.) and Puerto Llano (PLL; 37.81° N, 2.96° W, 1810 m.a.s.l.). Primers were fluorescently labelled and conditions optimized to perform multiplex PCR including most of the selected primers. Multiplex PCR profiles were optimized to combine 5 polymorphic microsatellite markers each in a single 20 µL reaction containing 1× reaction buffer, 3.5 mM $MgCl_2$, 0.25 mM dNTP, 2 U *Taq* DNA polymerase, 0.2 µM of 5'-labeled primer with one of FAM, VIC, NED or PET tags, 0.2 µM unlabelled primer and c.120 ng of DNA extract as template. Conditions for multiplex PCR were: 95°C for 3 min, 38 cycles of 95°C for 30 sec, 58°C for 30 sec, 72°C for 1 min and a final extension of 72°C for 3 min. Three additional primer pairs were optimized with simplex PCR reactions as described above (see Table 1). All PCR reactions were performed in a C100 Thermal Cycler (Biorad).

Diluted PCR products were separated and sized according to 500 LIZ size standard in an ABI PRISM 3130xl sequencer (Applied Biosystems). Allele patterns based on peak fluorescent intensity and morphology were visually analysed using GENEMAPPER v4.0 (Applied Biosystems). Total number of alleles, observed and expected heterozygosity and private alleles (i.e. alleles detected in one population only) were computed with GENALEX 6 (Peakall & Smouse 2006). Frequency of null alleles was estimated with INEST 1.1 (Chybicki

& Burczyk 2009; Campagne *et al.* 2012). Deviation from Hardy-Weinberg equilibrium, linkage disequilibrium between pairs of loci and fixation index (F_{IS}; Weir & Cockerham 1984) were calculated with GENEPOP v4 (Rousset 2008).

Twenty-six out of the 60 tested loci produced clear amplification products. Thirteen loci exhibited polymorphism (*Hefo1*-*Hefo13*) and enabled resolution of 60 individual genotypes, whereas other 13 loci appeared to be monomorphic (*Hefo14*-*Hefo26*). GenBank accession numbers, primer sequences, repeat motifs and PCR profiles for polymorphic loci are listed in Table 1. The remaining 34 primer pairs produced unspecific peaks or failed to amplify and therefore were dismissed.

Among polymorphic loci, the number of alleles per locus ranged from two to eight (mean = 3.92) with a total of 51 alleles in the 60 genotyped individuals. The number of private alleles in each population was two for TEJ, three for NAV and three for PLL. Chybicki-Burczyk likelihood-based method after 10000 iterations indicated high null allele frequency only at locus *Hefo12* in all populations (Table 2). All loci were at Hardy-Weinberg equilibrium except *Hefo12*, which showed a significant excess of homozygotes, due to the presence of null alleles. None of the loci pairs demonstrated linkage disequilibrium after Bonferroni correction for multiple comparisons. Population mean F_{IS} values were 0.044 for TEJ, 0.035 for NAV and 0.032 for PLL.

In conclusion, the microsatellite loci described here will be of great value for further studies on mating system, as well as assessing patterns of gene flow and levels of genetic diversity in natural populations of *H. foetidus*, and specifically, in the study of the interactions of nectar-living microbes affecting pollen dispersal and seed progeny quality in this species.

Data accessibility

GenBank accession numbers for *H. foetidus* microsatellite loci: JX905360 to JX905372 (polymorphic loci) and JX905373 to JX905385 (monomorphic loci).

Acknowledgements

We thank Rafael G. Albaladejo for helpful advice. This work was supported by the Spanish Ministerio de Economía y Competitividad through grants CGL2010-15964 to CMH, FPI program predoctoral grant BES-2011-046472 to MMS, and Juan de la Cierva Program to CdV.

Contributors

MMS and CdV carried out genetic analyses and were responsible for the writing and edition of the manuscript. PB conducted most laboratory work. MM collected field samples used in the present study and conducted some analyses, and CMH designed and supervised the study.

References

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004–1010.

Campagne P, Smouse PE, Varouchas G, Silvain JF, Leru B (2012) Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. *Molecular Ecology Resources* **12**, 975-982.

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* **100**, 106-113.

Dhooghe E, Grunewald W, Leus L, Van Labeke M-C (2009) In vitro polyploidisation of *Helleborus* species. *Euphytica* **165**, 89-95.

Herrera CM, Pozo MI (2010) Nectar yeasts warm the flowers of a winter-blooming plant. *Proceedings of the Royal Society B-Biological Sciences* **277**, 1827-1834.

Herrera CM, García IM, Pérez R (2008) Invisible floral larcenies: microbial communities degrade floral nectar of bumble bee-pollinated plants. *Ecology* **89**, 2369-2376.

Herrera CM, Pozo MI, Medrano M (in press) Yeast in nectar of an early-blooming herb: sought by bumble bees, detrimental to plant fecundity. *Ecology* (doi: 10.1890/12-0595.1)

Herrera CM, Sánchez-Lafuente AM, Medrano M, Guitián J, Cerdá X, Rey P (2001) Geographical variation in autonomous self-pollination levels unrelated to pollinator service in *Helleborus foetidus* (Ranunculaceae). *American Journal of Botany* **88**, 1025-1032.

Malausa T, Gilles A, Meglécz E *et al.* (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* **11**, 638-644.

Méglecz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.

Peakall R, Smouse P (2006) Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Rey PJ, Herrera CM, Guitián J, Cerdá X, Sánchez-Lafuente AM, Medrano M, Garrido JL (2006) The geographic mosaic in predispersal interactions and selection on *Helleborus foetidus* (Ranunculaceae). *Journal of Evolutionary Biology* **19**, 21-34.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Vesprini JL, Pacini E (2000) Breeding systems in two species of the genus *Helleborus* (Ranunculaceae). *Plant Biosystems* **134**, 193-197.

Vesprini JL, Nepi M, Pacini E (1999) Nectary structure, nectar secretion patterns and nectar composition in two *Helleborus* species. *Plant Biology* **1**, 560-568.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Werner K, Ebel F (1994) Zur Lebensgeschichte der Gattung *Helleborus* L.(Ranunculaceae). *Flora* **189**, 97-130.

Table 1. Characteristics of polymorphic loci developed for *Helleborus foetidus*. F= forward primer, R= reverse primer, M1=Multiplex set 1, M2=Multiplex set 2, S=Simplex and N_A= total number of alleles in the whole sample of 60 individuals from 3 populations (see Table 2). A 'PIGtail' (5'-GTTT-3') was added to the 5' end of reverse primers to facilitate the addition of adenosine by *Taq* polymerase (Brownstein *et al*. 1996).

Locus name	Accession number		Primer sequence (5'-3')	Repeat motif	Dye	PCR Profile	Size range (bp)	N_A
Hefo1	JX905360	F:	TTTGGAAATTTTAAAGGTTCTTGC	$(TCT)_7$	FAM	M1	197-203	3
		R:	CCAACCACATAATCATATCATAAGC					
Hefo2	JX905361	F:	AGCACTGAGTTTCTAAAAGGGC	$(GA)_7$	VIC	M1	95-97	2
		R:	TTTCTCTTCAATCAAGAATAAACCA					
Hefo3	JX905362	F:	AACATGCAAGACCGAACAACT	$(GA)_{12}$	PET	M2	101-117	8
		R:	CCGGCGAAACTTTACCTGT					
Hefo4	JX905363	F:	CTTCTTAAGTTTTATCAGAAACTTTGC	$(TCT)_8$	FAM	M1	247-265	7
		R:	AGTATCTAGGTGGATAACGCTTGA					
Hefo5	JX905364	F:	AACCACTCTATACGCTCCTCCA	$(CT)_7$	FAM	S	173-177	3
		R:	TTGAGATAGTAGCACCTATTATTGAGA					
Hefo6	JX905365	F:	ACTCACCAGTTTGGTTTTGCT	$(GTT)_5$	FAM	M1	164-167	2
		R:	TGCATACTCAATCCCATCCA					
Hefo7	JX905366	F:	CTTAACTGTACACCCTTAATGCATATC	$(CA)_8$	FAM	S	153-155	2
		R:	CAAGATACTCAAGCATGGGC					
Hefo8	JX905367	F:	GGAAGTACTCGAGGAAATTAACGA	$(GA)_9$	NED	M2	202-218	3
		R:	CCCAACTTATGATCTGCCAAA					
Hefo9	JX905368	F:	AACCATCCATTCACACCATTT	$(CT)_7$	NED	M2	266-268	2
		R:	TTTGTATTTGGCATTTCATGG					
Hefo10	JX905369	F:	AGCTTGCACAATGCTCTTCA	$(GA)_{10}$	VIC	M1	174-178	3

		R: CGATGCAAGTTGGTTCTTTTC					
Hefo11	JX905370	F: AGGTTCTAACCAAACCAATAAGG	$(TC)_8$	FAM	S	265-269	3
		R: GGATTGATATGACTTCATCACTGG					
Hefo12	JX905371	F: AGGGTAACGAAGGATATGACAGC	$(CT)_{10}$	PET	M2	185-193	5
		R: TGGGTGTTAAGAAGGTAATAAGATG					
Hefo13	JX905372	F: AATAGGCCACCAGGGTTGAT	$(AG)_8$	NED	M2	162-176	8
		R: TCTTGATAGGCCTCATTATTTGT					

179

Table 2. Population genetics statistics based on the 13 polymorphic microsatellite loci in three *Helleborus foetidus* populations (TEJ, NAV, and PLL, N = 20 individuals) located at Sierra de Cazorla (southeastern Spain). N_A=Total number of alleles, H_O =observed heterozygosity, H_E =expected heterozygosity, F_{IS} =Weir and Cockerham's (1984) fixation index, HWE=significant deviation from Hardy-Weinberg equilibrium, r = null allele frequency estimated by Chybicki-Burczyk method and S.D.=standard deviation.

	TEJ						NAV						PLL					
Locus	N_A	H_O	H_E	F_{IS}	HWE[1]	r	N_A	H_O	H_E	F_{IS}	HWE[1]	r	N_A	H_O	H_E	F_{IS}	HWE[1]	r
Hefo1	3	0.400	0.483	0.176		0.093	3	0.450	0.504	0.109		0.074	2	0.450	0.450	-0.000		0.077
Hefo2	2	0.250	0.224	-0.118		0.079	2	0.350	0.409	0.147		0.098	2	0.100	0.097	-0.027		0.104
Hefo3	5	0.800	0.773	-0.036		0.042	4	0.650	0.682	0.048		0.060	6	0.500	0.587	0.152		0.075
Hefo4	7	0.750	0.681	-0.105		0.038	6	0.550	0.690	0.207		0.080	5	0.400	0.574	0.309		0.110
Hefo5	2	0.400	0.328	-0.226		0.060	3	0.550	0.476	-0.161		0.048	2	0.350	0.409	0.147		0.098
Hefo6	2	0.500	0.385	-0.310		0.052	2	0.850	0.512	-0.691		0.030	2	0.400	0.328	-0.226		0.061
Hefo7	2	0.300	0.262	-0.152		0.071	2	0.050	0.050	n.a.		0.109	2	0.100	0.097	-0.027		0.104
Hefo8	2	0.400	0.431	0.073		0.086	3	0.350	0.465	0.253		0.106	2	0.200	0.185	-0.086		0.084
Hefo9	2	0.350	0.481	0.277		0.115	2	0.650	0.512	-0.280		0.048	2	0.600	0.508	-0.188		0.055
Hefo10	3	0.400	0.396	-0.010		0.075	3	0.200	0.344	0.424		0.142	2	0.100	0.097	-0.027		0.098
Hefo11	2	0.050	0.050	n.a.[2]		0.112	3	0.550	0.528	-0.042		0.062	3	0.400	0.347	-0.156		0.061
Hefo12	5	0.105	0.724	0.858	***	0.356	4	0.313	0.615	0.500	*	0.171	4	0.400	0.671	0.410	*	0.155
Hefo13	5	0.500	0.555	0.102		0.070	6	0.700	0.641	-0.095		0.038	6	0.650	0.747	0.133		0.063
Average	3.23	0.400	0.444	0.044			3.31	0.478	0.494	0.035			3.08	0.358	0.392	0.032		
S.D.	1.69	0.214	0.207	0.306			1.38	0.221	0.168	0.323			1.61	0.184	0.224	0.190		

[1]* ($p < 0.05$) and *** ($p < 0.001$). Empty cells indicate no deviation from HWE.
[2]n.a. = not available.

180