American Journal of Botany

American Journal of Botany: e49–e52. 2012.

AJB Primer Notes & Protocols in the Plant Sciences

Nuclear microsatellite loci for *Arabidopsis halleri* (Brassicaceae), a model species to study plant adaptation to heavy metals[1]

Cécile Godé[2,7], Isabelle Decombeix[2,7], Alicja Kostecka[3], Pawel Wasowicz[4,5], Maxime Pauwels[2], Adeline Courseaux[2,6], and Pierre Saumitou-Laprade[2,8]

[2]Laboratoire de Génétique et Evolution des Populations Végétales, FRE CNRS 3268, Université des Sciences et Technologies de Lille, Lille1, F-59655 Villeneuve d'Ascq Cedex, France; [3]Institute of Botany PAS, Lubicz 46, 31-512 Krakow, Poland; and [4]Department of Plant Systematics, Faculty of Biology and Environmental Protection, University of Silesia, ul. Jagiellońska 28, 40-032 Katowice, Poland

- *Premise of the study: Arabidopsis halleri* is a model species to study the adaptation of plants to soils contaminated by zinc, cadmium, and lead. To provide a neutral genetic background with which adaptive genetic markers could be compared, we developed highly polymorphic neutral microsatellite markers.
- *Methods and Results:* Using a microsatellite-enriched library method, we identified 120 microsatellite loci for quantitative trait locus (QTL) mapping analysis, of which eight primer pairs were developed in a single multiplex for population genetic studies. Analyses were performed on 508 individuals from 26 populations. All loci were polymorphic with six to 23 alleles per locus. Genetic diversity varied between 0.56 and 0.76.
- *Conclusions:* Our results demonstrated the value of these eight microsatellite markers to investigate neutral population genetic structure in *A. halleri*. To increase the resolution of population genetic analyses, we suggest adding them to the 11 markers previously developed independently.

Key words: *Arabidopsis halleri*; Brassicaceae; demographic processes; microsatellite markers; population genetic structure; pseudometallophyte.

Arabidopsis halleri (L.) O'Kane & Al-Shehbaz (Brassicaceae) is a perennial, clonal, and self-incompatible plant that is widespread in Central Europe. Because it is both a close relative to *A. thaliana* (L.) Heynh. and a pseudometallophyte species growing on metalliferous and on nonmetalliferous soil, it has acquired the status of a model species to study plant adaptation to soil contaminated by high concentrations of zinc, cadmium, and lead (e.g., Clauss and Koch, 2006). Several studies have investigated the genetic architecture of zinc and cadmium tolerance and hyperaccumulation in *A. halleri* (e.g., Willems et al., 2007; Frérot et al., 2010). Consequently, molecular markers for candidate genes will be soon available for population genetics studies. Inferring adaptation at candidate loci would imply unraveling the molecular signature of selection (Storz and Wheat, 2010). This could require being able to disentangle the effects of selective and demographic processes on population genetics, for example, by comparing the patterns of distribution of nuclear genetic variation in natural populations among candidate and neutral loci. In this context, the availability of neutral molecular markers enables the analysis of the impact of genetic drift, gene flow among populations, or genetic bottlenecks on population genetic structure. Characterizing neutral population genetic structure could also be helpful for the study of genetic mechanisms of local adaptation. Indeed, the detection of selected "outlier" loci, likely involved in the adaptive differentiation among populations occurring in contrasted ecological conditions, usually implies comparisons of population structure with neutral loci (Luikart et al., 2003). However, to be useful at a local geographic scale, molecular markers would have to be highly polymorphic.

Four *A. lyrata* (L.) O'Kane & Al-Shehbaz and seven *A. thaliana* microsatellite markers had already been transferred to *A. halleri* in previous studies (van Rossum et al., 2004; Llaurens et al., 2008). However, some of these markers showed low genetic diversity in *A. halleri*. Moreover, genotyping protocols were not optimal because these markers have to be amplified separately. In this context, we developed and combined eight new microsatellite markers. Markers were designed to show high levels of genetic diversity and were assembled in a single multiplex reaction to reduce manipulation time and costs. We then compared these eight markers to the 11 primer pairs developed in previous studies.

[1]Manuscript received 11 July 2011; revision accepted 19 September 2011.

The authors thank A. Bourceaux, A. Créach, and H. Frérot for help in field surveys and sample collection. This research was supported the Région Nord Pas de Calais and the European Community (Arcir PLANTEQ8) to P.S.-L. Isabelle Decombeix was funded by the French Ministry of Research and Technology.

[5]Current address: Faculty of Biology and Environmental Sciences, Cardinal Stefan Wyszynski University in Warsaw, Woycickiego 1/3, 01-938 Warszawa, Poland

[6]Current address: Unité Glycobiologie Structurale et Fonctionnelle, UMR8576 du CNRS, Univ Lille 1, IFR147, 59655 Villeneuve d'Ascq Cedex, France

[7]These authors contributed equally to this work.

[8]Author for correspondence: pierre.saumitou@univ-lille1.fr

doi:10.3732/ajb.1100320

METHODS AND RESULTS

Genomic DNA extracts were obtained from 26 populations in three valleys of Lombardy (Italy): a calamine valley with populations on metalliferous and nonmetalliferous soils spread over 10 km, and two nonpolluted valleys located 40 km from the first one (see Appendix 1 for locality information). Representative specimens of the sampled populations are stored in the Genetic and Evolution of Plant Populations laboratory (GEPV) in France. In each population, we collected leaves from five to 25 individuals (Appendix 1). Sampled plants were separated by at least 3 m to avoid sampling clones (van Rossum et al., 2004). Overall, 508 genotypes were collected. DNA from each genotype was extracted from 10 to 15 mg.

Eight new markers (ah15, ah22, ah27, ah44, ah49, ah59, ah79, and ah89; Table 1) were selected from an enriched procedure for AG and TG motifs with Dynabeads (Invitrogen Dynal, Oslo, Norway) according to the protocol described in Glenn and Schable (2005). Genomic DNA from a single individual from a population in Auby (northern France) was digested with *Rsa*I (New England Biolabs, Beverly, Massachusetts, USA) and *Ssp*I (New England Biolabs) for 5 h at 37°C and ligated to SuperSNX24 (5′-GTTTAAGGCCTAGCT-AGCAGAATC-3′/5′-pGATTCTGCTAGCTAGGCCTTAAACAAAA-3′) using T4 DNA ligase (USB, Cleveland, Ohio, USA) according to Glenn and Schable (2005). For enrichment, the PCR products were denatured at 95°C for 5 min, then hybridized with two 3′ biotinylated oligonucleotides, $(AG)_{12}$ and $(TG)_{12}$. Dynabeads (Invitrogen Dynal) were used to capture DNA fragments containing targeted repeats, and the enriched DNA was cloned using the TOPO TA cloning kit (Invitrogen Dynal). Positive clones were tested according to Glenn and Schable (2005). A total of 400 clones were sequenced on an ABI Prism 3130 DNA sequencer (Applied Biosystems, Foster City, California, USA), and analyses were conducted with an inhouse program using ClustalW (Larkin et al., 2007) to eliminate sequence redundancy and MREPS (Kolpakov et al., 2003) to find microsatellite patterns. Primer sequences were designed in flanking regions of *A. halleri* microsatellites using Primer3 software (Rozen and Skaletsky, 2000; http://frodo.wi.mit.edu/) under standard conditions for three different product sizes (50–150, 150–250, and 250–350 bp). The most suitable size was then selected for multiplexing.

The eight selected markers were combined into a single multiplex for PCR and loading. Forward primers of the eight loci were labeled with FAM, PET, NED, or VIC fluorescent dye. The PCR reactions were carried out simultaneously in a total volume of 10 μL, containing 20 ng of DNA template and 2 μM of each mixed forward and reverse primers, using 0.5 U of the QIAGEN Multiplex PCR Kit (QIAGEN, Courtaboeuf, France). The cycling conditions consisted of one initial denaturation step of 15 min at 95°C, followed by two touchdowns of five cycles each: denaturation for 45 s at 95°C, annealing at 68°C (−2°C/cycle) for 5 min for the first touchdown and at 58°C (−2°C/cycle) for 1 min for the second, and extension for 60 s at 72°C; then 27 cycles of 45 s at 95°C, 30 s at 47°C, and 60 s at 72°C; and a final extension of 10 min at 72°C. PCR was conducted on a Mastercycler Pro S (Eppendorf France SARL, Le Pecq, France). Capillary electrophoresis was carried out in a mix containing 1.5 μL of PCR product, 9.75 μL of Hi-Di Formamide, and 0.25 μL of GeneScan 500 LIZ Size Standard (Applied Biosystems). Amplification products were separated on an ABI Prism 3130 DNA sequencer (Applied Biosystems). Alleles at microsatellite loci were scored using GeneMapper software version 3.7 (Applied Biosystems).

Each individual was also genotyped with 11 previously available markers that were first developed for *A. thaliana* (*Ath*) and *A. lyrata* (*Al*). Six microsatellite markers were developed and amplified separately because they cannot be multiplexed (LYR133 [*Al*], LYR132 [*Al*], LYR417 [*Al*], and GC16 [*Ath*] described in Clauss et al. [2002]; ATH-CTR1A [*Ath*] described in Bell and Ecker [1994]; and LYR104 [*Al*] [kindly provided by Thomas Mitchell-Olds]). They were previously used to investigate the neutral genetic structure in *Arabidopsis* and *Arabis* (Bell and Ecker, 1994; Clauss et al., 2002) and were used at the within-population level in *A. halleri* (van Rossum et al., 2004). Five other microsatellite markers (GC22 [*Ath*], ICE13 [*Ath*], MDC16 [*Ath*], NGA112 [*Ath*], and NGA361 [*Ath*]) were multiplexed and transferred in *A. halleri* to investigate the paternal diversity of the species (Llaurens et al., 2008). Amplification and scoring were done according to previous studies (van Rossum et al., 2004; Llaurens et al., 2008). In total, the 508 individuals were genotyped with 19 microsatellite loci scattered in the *A. halleri* genome (Tables 1, 2).

For the 19 markers, the proportion of individuals successfully amplified was high for all markers, varying from 85–98% of successful amplifications with a mean of 95% (Table 2). Those results were mainly due to a few individuals who did not amplify consistently across loci, except NGA112 for which a very low and doubtful amplification signal was specifically obtained for a part of the individuals. The proportion of individuals successfully amplified was similar for the two marker sets (95.4% for the new multiplex set and 95.6% for the other markers; Table 2). Tests for linkage and Hardy–Weinberg equilibrium and computations of expected heterozygosity were performed using FSTAT 2.9.3.2 (Goudet, 1995). Frequencies of null alleles were estimated with FreeNA (Chapuis and Estoup, 2007). We did not detect linkage disequilibrium among loci after Bonferroni correction ($\alpha = 0.05$). Estimated frequencies of null alleles were very low for all loci, varying from 1–6% with a mean of 3%. Only one marker (ah59) showed a higher frequency of null alleles (11%).

The new multiplexed loci revealed between six and 23 alleles in each locus, and the expected heterozygosity ranged from 0.56 to 0.76 with a mean of 0.64 (Table 2). In comparison, four to 18 alleles were detected at the 11 other markers and expected heterozygosity estimates ranged from 0.18 to 0.82 with a mean of 0.53. Expected heterozygosity was significantly lower for three of the six markers developed independently (0.28, 0.18, and 0.36 for ATH-CTR1A, GC16, and LYR104, respectively; Table 2). In two populations, the expected heterozygosity was higher for the new marker set (0.67 and 0.71 for I16 and I14, respectively) than for the other markers (0.55 and 0.54 for I16 and I14, respectively) whereas for I32 the expected heterozygosity was higher for the markers developed in previous studies (0.58 and 0.51 for the old and the new marker set, respectively; Table 2).

TABLE 1. Characteristics of eight microsatellite markers developed for *Arabidopsis halleri*. Shown for each marker are the forward and reverse sequence, core motif, allele size range estimated on 508 individuals, linkage group to which the marker belongs according to a genetic map of *A. halleri* × *A. lyrata* subsp. *petraeam* cross, fluorescent dye, and GenBank accession number.

Primer	Primer sequence (5′–3′)	Repeat motif	Size range (bp)	LG	Dye	GenBank Accession No.
ah15	F: CCAGGAAAGGCAAATCAAGA R: AGGGACGCACGATTTTAGTG	$(GA)_{10}$	154–254	4	NED	FR873659
Ah22	F: CCAGTTTCGATTTGTTTTACTTTG R: AGTTGTGTGATGTAATAAGGTGAAAT	$(CA)_8$	110–181	3	PET	FR873661
Ah27	F: TTGCGTTTGAGTGTGTTGTG R: CGATTGGGTAGCCAAGAGAA	$(TG)_8$	250–257	4	FAM	FR873664
Ah44	F: AGCCCTAGCATCTCCTTTCA R: TCGTGTTTGTGCTTGAGTTTG	$(TC)_8N_{26}(TC)_3(CTC)_4(CT)_4$	90–121	2	VIC	FR873665
Ah49	F: CGGCCACCTCTCTGTAAATC R: GAGAACAATGTTGAATTGATTGC	$(TC)_{12}$	65–127	7	NED	FR873660
Ah59	F: GGAAGCAATACTGGGGAAGC R: AAAAACCTTAGCCCTTTTGG	$(CA)_{10}N_6(CA)_5$	176–184	1	VIC	FR873663
Ah79	F: TGCATTGTGTTGTGTTTCTGC R: TCACATGACGCAATATGCAC	$(GA)_{16}$	82–93	5	PET	FR873662
ah89	F: TCCAACACAAGGGGAAAGTC R: CCAACCTTGCGAGTTGATTC	$(TC)_{12}$	240–266	7	NED	FR873666

TABLE 2. Characterization of 19 microsatellite markers for all the sampled populations (508 individuals) and three populations (I16 and I14 were in the same valley whereas I31 was 40 km away).

		All individuals			I16 (n = 25) (45°51′33.9N, 9°52′35.3E)		I14 (n = 20) (45°53′21.1N, 9°47′10.8E)		I31 (n = 20) (45°59′22.7N, 10°16′05.2E)	
Primer	LG	N	*A*	H_e	*A*	H_e	*A*	H_e	*A*	H_e
ah15	4	95.40	6	0.56	5	0.61	5	0.76	3	0.43
ah22	3	89.3	11	0.76	9	0.87	6	0.77	5	0.78
ah27	4	93.3	6	0.62	4	0.74	5	0.71	3	0.16
ah44	2	95.4	9	0.56	5	0.37	3	0.51	7	0.79
ah49	7	91.3	23	0.60	11	0.80	7	0.68	3	0.24
ah59	1	95.2	6	0.64	3	0.67	5	0.73	4	0.68
ah79	5	96.2	6	0.64	4	0.55	5	0.71	3	0.60
ah89	7	95.2	11	0.67	5	0.75	6	0.79	5	0.40
Mean		95.40	9.75	0.64	5.75	0.67	5.25	0.71	4.13	0.51
ATH-CTR1A	6	98.4	9	0.28	4	0.19	2	0.10	4	0.40
GC16	1	99	7	0.18	2	0.12	1	0.00	2	0.23
LYR133	2	96.6	12	0.61	5	0.64	5	0.62	4	0.55
LYR104	—	98.2	8	0.36	4	0.35	3	0.61	3	0.51
LYR132	2	96.8	5	0.73	4	0.37	4	0.26	2	0.50
LYR417	1 or 2	97.6	16	0.54	8	0.72	5	0.72	7	0.76
GC22	2	95.8	4	0.54	3	0.55	3	0.45	3	0.61
ICE13	1	95.6	12	0.56	5	0.72	5	0.59	3	0.60
MDC16	3	95.2	6	0.43	2	0.38	3	0.50	2	0.46
NGA112	5	85.2	18	0.82	9	0.86	8	0.78	10	0.85
NGA361	4	96	12	0.75	6	0.77	8	0.84	4	0.72
Mean[a]		95.60	10.00	0.56	4.80	0.55	4.50	0.54	4.00	0.58
Total mean		95.82	9.84	0.58	5.16	0.58	4.68	0.59	4.05	0.54

Note: *A* = number of alleles; H_e = expected heterozygosity; LG = linkage group; n = number of sampled genotypes per population; N = percentage of well-amplified individuals.

[a] Mean of the new multiplex and markers developed in previous studies.

CONCLUSIONS

The eight microsatellite markers newly developed in this study show both a high amplification success and a high variability. Compared to primers developed in previous studies (van Rossum et al., 2004; Llaurens et al., 2008), they show a high level of polymorphism while reducing manipulation time and costs. It seems therefore that those eight microsatellite markers would be appropriate for the analysis of neutral population genetics at a local geographic scale in *A. halleri*. Nevertheless, considering the influence of the number of loci applied in the estimation of genetic relationships among populations (Koskinen et al., 2004), we recommend the use of all the 19 microsatellite markers. To decrease manipulation time and costs, it will be helpful to multiplex all the markers, in particular the six markers developed separately. However, due to their allele range, those six markers have to be used in at least two multiplexes, using, for example, the procedure developed by Llaurens et al. (2008). We hope this microsatellite marker set will help forthcoming studies to disentangle selective and demographic processes.

LITERATURE CITED

BELL, C. J., AND J. R. ECKER. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis. Genomics* 19: 137–144.

CHAPUIS, M., AND A. ESTOUP. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

CLAUSS, M. J., AND M. A. KOCH. 2006. Poorly known relatives of *Arabidopsis thaliana. Trends in Plant Science* 11: 449–459.

CLAUSS, M. J., H. COBBAN, AND T. MITCHELL-OLDS. 2002. Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Molecular Ecology* 11: 591–601.

FREROT, L., M. P. FAUCON, G. WILLEMS, C. GODE, A. COURSEAUX, A. DARRACQ, N. VERBRUGGEN, AND P. SAUMITOU-LAPRADE. 2010. Genetic architecture of zinc hyperaccumulation in *Arabidopsis halleri*: The essential role of QTL × environment interactions. *New Phytologist* 187: 355–367.

GLENN, T. C., AND N. A. SCHABLE. 2005. Isolating microsatellite DNA loci. *Methods in Enzymology* 395: 202–222.

GOUDET, J. 1995. FSTAT: A computer program to calculate F statistics, version 2.9. *Journal of Heredity* 86: 485–486.

KOLPAKOV, R., G. BANA, AND G. KUCHEROV. 2003. MREPS: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research* 31: 3672–3678.

KOSKINEN, M. T., H. HIRVONEN, P. A. LANDRY, AND C. R. PRIMMER. 2004. The benefits of increasing the number of microsatellites utilized in genetic population studies: An empirical perspective. *Hereditas* 141: 61–67.

LARKIN, M. A., G. BLACHSHIELDS, N. P. BROWN, R. CHENNA, P. A. MCGETTIGAN, H. MCWILLIAM, F. VALENTIN, ET AL. 2007. ClustalW and ClustalX version 2. *Bioinformatics (Oxford, England)* 23: 2947–2948.

LLAURENS, V., V. CASTRIC, F. AUSTERLITZ, AND X. VEKEMANS. 2008. High paternal diversity in the self-incompatible herb *Arabidopsis halleri* despite clonal reproduction and spatially restricted pollen dispersal. *Molecular Ecology* 17: 1577–1588.

LUIKART, G., P. R. ENGLAND, D. TALLMON, S. JORDAN, AND P. TABERLET. 2003. The power and promise of population genomics: From genotyping to genome typing. *National Review* 4: 981–994.

ROZEN, S., AND H. J. SKALETSKY. 2000. Primer3 on the WWW for general users and biologist programmers. *In* S. Krawetz and S. Misener [eds.], Bioinformatics methods and protocols: Methods in molecular biology, 365–386. Humana Press, Totowa, New Jersey, USA.

STORZ, J. F., AND C. W. WHEAT. 2010. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution; International Journal of Organic Evolution* 64: 2489–2509.

VAN ROSSUM, F., I. BONNIN, S. FENART, M. PAUWELS, D. PETIT, AND P. SAUMITOU-LAPRADE. 2004. Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology* 13: 2959–2967.

WILLEMS, G., D. B. DRÄGER, M. COURBOT, C. GODE, N. VERBRUGGEN, AND P. SAUMITOU-LAPRADE. 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): An analysis of quantitative trait loci. *Genetics* 176: 659–674.

APPENDIX 1. Sampled *Arabidopsis* populations used in this study.

Population	GPS coordinates (°N, °E)	Valley[a]	*n*
I12	45°54′07.9, 9°46′54.8	1	18
I13	45°53′55.7, 9°46′31.3	1	22
I13h	45°53′55.7, 9°46′31.3	1	9
I14	45°53′21.1, 9°47′10.8	1	20
I15	45°54′44.1, 9°48′10.7	1	25
I16	45°51′33.9, 9°52′35.3	1	25
I17	45°51′57.2, 9°51′45.3	1	22
I18	45°51′49.2, 9°52′04.9	1	23
I19	45°52′14.8, 9°49′30.8	1	23
I20	45°53′13.41, 9°46′31.5	1	20
I21	45°53′23.32, 9°45′13.9	1	21
I22	45°52′53.6, 9°44′39.2	1	20
I23	45°51′04.47, 9°43′40.2	1	5
I24	45°52′57.2, 9°44′57.8	1	21
I25	45°52′56.9, 9°44′59.7	1	18
I26	45°53′29.7, 9°47′23.3	1	19
I27	45°52′26.43, 9°48′50.44	1	17
I28	46°03′25.06, 10°14′31.23	2	20
I29	46°03′29.8, 10°15′15.20	2	20
I30	45°59′28.03, 10°16′19.38	3	19
I31	45°59′22.77, 10°16′05.27	3	20
I32	45°58′46.5, 10°16′59.6	3	20
I33	45°58′50.97, 10°16′31.53	3	21
I34	45°59′18.7, 10°16′25.5	3	13
I35	45°55′01.6, 9°47′40.1	1	19
I36	45°54′47.7, 9°47′32.8	1	20

Note: n = number of sampled genotypes.

[a]The three different sampled valleys are indicated as 1, 2, or 3.