UNIVERSITAT DE BARCELONA
FACULTAT DE BIOLOGIA
Departament de Biologia Vegetal, Unitat de Botànica

Programa de Doctorat: Biologia Vegetal
Bienni: 2002-2004

INSTITUT BOTÀNIC DE BARCELONA (CSIC-ICUB)

EVOLUCIÓ, FILOGÈNIA I SISTEMÀTICA DEL COMPLEX *ARCTIUM-COUSINIA*

Memòria presentada per Sara López Viñallonga per a optar al títol de Doctor per la Universitat de Barcelona

Amb el vist-i-plau dels directors de tesi:

Dra. Núria Garcia Jacas

Dr. Alfonso Susanna de la Serna

i del tutor de tesi:

Dr. Ramon M. Masalles i Saumell

Sara López Viñallonga
Barcelona, 2009

7. 4. Isolation and characterization of novel microsatellite markers for *Arctium minus* (Hill) Bernh. (Compositae)

S. López-Vinyallonga[*], M. Arakaki[†], N. Garcia-Jacas[*], A. Susanna[*], M. A. Gitzendanner[†], D. E. Soltis[†] and P. S. Soltis[††]

[*]*Botanic Institute of Barcelona (CSIC-ICUB), Pg. del Migdia s. n., 08038 Barcelona, Spain* [†] *Department of Botany, University of Florida, Gainesville, FL 32611, USA*
[††]*Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA*

RESUM. En aquest treball es presenten nou microsatèl·lits polimòrfics dissenyats per *Arctium minus* a partir d'una llibreria genòmica enriquida. La caracterització d'aquests loci es basa en sis poblacions, de més de 20 individus cadascuna, procedents d'Eslovàquia, Espanya (2 poblacions), França, Polònia i Turquia. Pel que fa a la caracterització dels loci, el nombre d'al·lels per locus va de dos a 10, els valors d'heterozigositat esperada i observada varien de 0.015 a 0.487 i de 0.016 a 0.694, respectivament i el valor mitjà de F_{IS} (indicador de consanguinitat) és de 0.316. Sis loci mostren una desviació estadísticament significativa de l'equilibri Hardy-Weinberg (HW a partir d'ara) degut a una deficiència d'heterozigots, i nou de 36 comparacions entre parelles de locus mostren un significatiu desequilibri en el lligament. S'ha detectat la presència d'al·lels nuls, amb freqüències baixes o moderades, als loci Am31 i Am34 per cinc i tres poblacions respectivament. Pel que fa a la caracterització de les poblacions considerades, només la de Turquia està en equilibri HW i les de Polònia, Eslovàquia i nord-oest d'Espanya mostren un significatiu desequilibri en el lligament. Complementàriament, s'aporten les seqüències de cinc parells de primers que amplifiquen loci monomòrfics, potencialment útils per a altres espècies o per poblacions amb més variabilitat i de quatre parells de primers assajats en menys de 20 individus per població però que amplifiquen loci polimòrfics. Els microsatèl·lits presentats a aquest treball poden ser d'utilitat per a examinar l'estructura genètica de les poblacions d'*A. minus* i per a investigar aspectes relacionats amb la colonització del continent Americà per part d'aquesta espècie.

Article acceptat a la revista *American Journal of Botany (Notes)*.

Isolation and characterization of novel microsatellite markers for *Arctium minus* (Hill) Bernh. (Compositae)

S. LÓPEZ-VINYALLONGA[*], M. ARAKAKI[†], N. GARCIA-JACAS[*], A. SUSANNA[*], M. A. GITZENDANNER[†], D. E. SOLTIS[†] and P. S. SOLTIS[††]

[*]Botanic Institute of Barcelona (CSIC-ICUB), Pg. del Migdia s. n., 08038 Barcelona, Spain [†] Department of Botany, University of Florida, Gainesville, FL 32611, USA
[††]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

Keywords: *Arctium minus*, enriched library, lesser burdock, microsatellites, population genetics, SSR

Corresponding author: Sara López-Vinyallonga, Botanical Institute of Barcelona, Pg. del Migdia s. n., 08038 Barcelona, Spain. Tel.: 34 93 2890611. Fax: 34 93 2890614. e-mail: vinyallonga@gmail.com

Running title: New microsatellite markers for *Arctium minus*

Abstract

Nine polymorphic microsatellite markers were developed in lesser burdock (*Arctium minus* L.) from an enriched genomic library. The number of alleles per locus ranged from two to 10, expected and observed heterozygosities ranged from 0.015 to 0.487 and from 0.016 to 0.694, respectively, and the mean value of F_{IS} was 0.316. Six loci showed a significant departure from Hardy-Weinberg equilibrium, and nine out of 36 pairwise locus comparisons showed significant linkage disequilibrium. Presence of null alleles was found in some populations at two loci with low or moderate frequencies. In addition, sequences for four primer pairs that were tested in fewer than 20 individuals per population but yielded polymorphic loci and for five primer pairs that amplified monomorphic loci for the samples examined are also provided. Microsatellite markers reported here will be useful for examining population genetic structure and for addressing questions regarding the colonization of the Americas by this species.

Arctium minus (Hill) Bernh.is a diploid biennial plant, native to Eurasia. It also grows in Africa, where it is rare, and it is widespread and often naturalized as a weed in many parts of North and South America. This species is monocarpic, growing vegetatively as broad-leaved rosettes and then dying after producing a tall (0.5-2 m) flowering stalk (Gross & Werner 1983). *Arctium minus* is allogamous, although it can be self-pollinated if insect pollination fails (Fenner *et al.* 2002). In previous phylogenetic studies (López-Vinyallonga *et al.* 2009), little sequence divergence among species of the *Arctium-Cousinia* complex was found, most likely due to rapid and recent speciation in the group. The high polymorphism found in microsatellite markers can be a helpful tool for the study of the population genetics of this complex, as it has been for the study of other recently speciated groups (e. g. Chirhart *et al.* 2005, Gugerli *et al.* 2001, Edwards *et al*. in press). Hence, we report here a set of novel polymorphic microsatellites that will be useful for assessing genetic variability and divergence as well as gene flow within and among populations of *A. minus*.

Genomic DNA was extracted from dried leaf tissue of specimens collected from wild populations, representing the majority of the distribution of *A. minus*, using the NucleoSpin® Plant Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Two genomic libraries enriched for 1) (AAG/TTC) and 2) (CA/GT) were constructed from *A. minus* following Symonds (pers. comm.). The genomic DNA was digested with *Sau*3AI and enriched for repeats by hybridization to 3'-biotinylated $(AAG)_{14}$ and $(CA)_{15}$ TATAAGATA oligonucleotides, respectively, followed by magnetic capture with streptavidin-coated magnetic beads (Promega Corp, Madison, WI, USA). Enriched fragments were made double-stranded by polymerase chain reaction (PCR) and were ligated into a TOPO TA pPCR 4.0 vector, transformed into One Shot *Escherichia coli* competent cells (Invitrogen, Carlsbad, California, USA) and grown on Luria Broth (LB) agar plates with Kanamycin. Replicas were transferred to LB/Kanamycin 96-well plates and grown overnight. Screening for positives was PCR-based, involving two PCRs per sample, with a repeat primer (either AAG or CA) and a primer for the vector (M13F or M13R).

A total of 264 positive clones for AAG and 235 for CA were detected; 192 and 160 of these, respectively, were sequenced in an ABI 3730xl DNA Analyzer (Applied

Biosystems, Foster City, California, USA). Using the criterion of at least six repeat units in the target sequence, we were able to design primers from 28 (for AAG) and 14 (for CA) of these sequences using PRIMER3 version 0.4.0 (Rozen & Skaletsky 2000) and Operon (Operon Biotechnologies, Inc., Huntsville, Alabama, USA). Forward primers had universal M13 tails added to their 5'ends following Boutin-Ganache *et al*. (2000). Amplifications were performed in 10 µl reactions containing 0.5 U of Go*Taq* Flexi DNA polymerase (Promega), 1x Promega colorless Go*Taq* Flexi Buffer, 2.0 mM $MgCl_2$, 50 µM of each dNTP, 0.45 µM of the reverse primer, 0.012 µM of the extended forward primer, 0.45 µM of the labelled M13 primer (6-FAM, VIC, NED or PET, Applied Biosystems) and 25 ng/µl of template DNA. The profile used for amplifications consisted of a denaturation step of 94 ºC for 3 min, followed by 34 cycles of 94 ºC for 30 s, 52 ºC for 30 s, and 72 ºC for 45 s and a final elongation time of 20 min at 72 ºC. Labeled PCR products were diluted 1:20 and pool-plexed by using 1-4 µl of each PCR (up to four loci). One µl of the diluted PCR mixture, 9.9 µl of formamide, and 0.1 µl LIZ 600 size standard (Applied Biosystems) was loaded and run on an ABI 3730xl DNA Analyzer (Applied Biosystems) at the Interdisciplinary Center for Biotechnology Research (ICBR) facility at the University of Florida. Four loci were monomorphic and 12 loci required further optimization; therefore, we did not pursue them further. Specific amplification and polymorphism were achieved for nine primer sets. We genotyped at least 20 individuals per population, belonging to six different populations. Fragment analysis was performed using GENEMARKER 1.5 (Soft Genetics, LLC) and Peak Scanner Software 1.0 (Applied Biosystems) and scored manually in both cases. The data were analyzed with GenAlEx6 (Peakall & Smouse 2006), and the results are shown in Table 1.

The number of alleles (*N*a) observed at each locus ranged from two to 10, with observed heterozygosities (H_o) ranging from 0.016 to 0.694, and expected heterozygosities (H_e) from 0.015 to 0.487 with a mean value of 0.173, showing low allelic diversity for *A. minus*. Estimates of the inbreeding coefficient F_{IS} (Weir & Cockerham 1984) ranged from -0.752 to 0.890, with a mean value of 0.316 for the studied populations.

GENEPOP version 3.4 (Raymond & Rousset 1995) was used to test for departure from Hardy-Weinberg equilibrium and linkage disequilibrium. Six loci showed a significant

departure from Hardy-Weinberg equilibrium ($P < 0.005$) due to heterozygote deficiency, and nine out of 36 pairwise locus comparisons showed significant linkage disequilibrium ($P < 0.005$).

Given the high levels of F_{IS} and linkage disequilibrium detected in our data, we performed additional analyses at the population level in order to investigate the reason for these high values. All populations but the one from Turkey showed a significant departure from Hardy-Weinberg equilibrium ($P < 0.005$). In addition, we observed significant linkage disequilibrium ($P < 0.005$) in the populations from Poland (seven out of 36 pairwise locus comparisons), Slovakia and Spain (one out of 36 pairwise locus comparisons each). There was no significant linkage disequilibrium in any of the other populations.

Furthermore, tests for the presence of null alleles were performed using MICRO-CHECKER 2.2.3 (van Oosterhout *et al.* 2004). There was no evidence for scoring errors due to stuttering and no evidence for large allele dropout at any of the nine loci tested. In addition, seven out of these nine loci did not show any presence of null alleles; null alleles were detected at two loci, but in low or moderate frequency and only in some of the populations tested. Locus Am31 showed evidence of null alleles in the five populations that are not under HWE, with estimated frequencies lower than 0.45. Locus Am34 showed evidence of null alleles in three populations, one of them under HWE, with estimated frequencies lower than 0.34.

The results of all these HWE and LD tests are consistent with the presence of null alleles and the facultative ability of self-pollination in *A. minus* as well.

Given the low diversity detected in the six populations studied, we also give details for four additional primer pairs that yielded polymorphic loci when tested in populations with fewer than 20 individuals (Appendix 1). For those loci that we genotyped and scored, no significant departure from HWE ($P < 0.005$) nor significant LD was found ($P < 0.005$).

We also provide details of five additional primer pairs that appeared to generate monomorphic loci (Appendix 2), although they might be polymorphic when tested in more populations of *A. minus*, or more individuals, or in closely related species.

Microsatellite markers reported here for *A. minus* are suitable for population genetic studies, such as understanding past evolutionary and demographic events. Extensive sampling of more populations has been carried out, and genetic analyses of these populations are in progress in order to answer questions regarding the colonization of the Americas and other population processes in this species.

Acknowledgements

This work was supported by a grant from the Spanish Government through the Ministery of Education and Science (CGL 2004-04563-C02-01/BOS and CGL 2007-60781/BOS) to S. López-Vinyallonga. We acknowledge I. Borşič, A. K. Brysting, D. Gutiérrez, A. Hilpold, S. Margaritescu, N. Montes, K. Romashchenko, A. Romo and R. Vilatersana for their help with plant collection.

References

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M-13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele sizing methods. *BioTechniques*, **31**, 24-28.

Chirhart SE, Honeycutt RL, Greenbaum IF (2005) Microsatellite variation and evolution in the *Peromyscus maniculatus* species group. *Molecular Phylogenetics and Evolution*, **34**, 408–415.

Edwards CE, Soltis DE, Soltis PS (in press) Testing hypotheses of current hybridization, ancient hybridization and incomplete lineage sorting in *Conradina* (Lamiaceae). *Molecular Ecology*.

Gross RS, Werner PA (1983) Probabilities of survival and reproduction relative to rosette size in the common burdock (*Arctium minus*: Compositae). *American Midland Naturalist*, **109**, 184—193.

Gugerli F, Senn J, Anzidei M, Madaghiele A, Büchler U, Sperisen C, Vendramin GG (2001) Chloroplast microsatellites and mitochondrial *nad*1 intron 2 sequences indicate congruent phylogenetic relationships among Swiss stone pine (*Pinus cembra*), Siberian stone pine (*Pinus sibirica*), and Siberian dwarf pine (*Pinus pumila*). *Molecular Ecology*, **10**, 1489–1497.

Fenner M, Cresswell E, Hurley A, Baldwin T (2002) Relationship between capitulum size and pre-dispersal seed predation by insect larvae in common Asteraceae. *Oecologia*, **130**, 72–77.

López-Vinyallonga S, Mehregan I, Garcia-Jacas N, Tscherneva O, Susanna A, Kadereit JW (2009) Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae-Carduinae). *Taxon*, **58,** 153–171.

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, **86**, 248-249.

Rozen S, Skaletsky HJ (2000) PRIMER3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener s), pp. 365-386. Humana Press, Totowa, New Jersey.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Table1 Primer sequence and configuration of 9 microsatellite loci in *Arctium minus*.

Locus	Primer Sequence (5'→3')	T_a (ºC)	Repeat Motif	Size Range (bp)	*N*	N_a	H_o	H_e	F_{IS}
Am8	F: *ATCGCCATCGAAGACAAGAC R: AATAAAATCCTGAGCCGGAAA	54	$(AAG)_9$	108-152	162	4	0.074	0.134 †	0.448
Am26	F: *TTGGGGTGAGGTAGAAGCAG R: TTCCACCAAGTTGGGTTAGC	52	$(AAG)_6$	260-265	131	2	0.016	0.015	-0.050
Am30	F: *GCAAAGGGGTCCTAGAGCAT R: TCGAAGTGTTATCGGTTGCT	52	$(CA)_8$	180-189	186	3	0.033	0.077 †	0.572
Am31	F: *TGTGCAACTGCTCCTTCAGT R: CTCCAACAATGCAGAAACCA	52	$(CA)_9$	182-272	142	8	0.053	0.487 †	0.890
Am32	F: *GCTGTTGCCATGACTCTAAGG R: CGGAAAAGGACGACAAAGAA	52	$(CA)_8$	212-226	157	4	0.334	0.304	-0.099
Am33	F: *TCCCTTGTGAAACGCAATTT R: CGGTCAGATCCATTTATCACG	52	$(CA)_9$	197-203	154	3	0.038	0.210 †	0.820
Am34	F: *CCATGCTCACCTCCATTTCT R: CAGCATAATGATACGGCAACA	52	$(GA)_9$	158-185	133	4	0.117	0.262 †	0.555
Am35	F: *AGTTAGTGTCATTGTTGAGAGAACTTA R: TGTGATAGCATCCAAACTCCA	50	$(TATG)_6(TG)_4[(TATG)_2(TG)_4]_4$	174-239	158	10	0.201	0.374 †	0.463
Am39	F: *TGCAGACACGCATTACAACA R: TGGCCCTAGAATGATGGAAA	52	$(CA)_3CGC(CA)_5$	248-258	170	4	0.694	0.396	-0.752

* M13 tail (CACGACGTTGTAAAACGAC), optimized annealing temperature (T_a), repeat motifs, size ranges of PCR products, number of individuals genotyped (*N*), number of alleles observed (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), fixation index (F_{IS}). Significant departures from HWE: † $P < 0.005$.

Appendix 1 Primer pairs polymorphic for *Arctium minus* tested in fewer than 20 individuals per population.

Locus	Primer Sequence (5'→3')	T_a (ºC)	Repeat Motif	Size Range (bp)	N_a	H_o	H_e	F_{IS}
Am2	F:* AGAAAGGAAAGGGGGAGCTT R: TCTTCTGGATCTGCCTCGAT	50	$(AAG)_6$	212-247	2	0.020	0.018	-0.111
Am17	F:* TCGTGGGACTCTACCACCTC R: TCCTGGACCCAGATCGTACT	52	$(AAG)_6$	254-288	3	0.076	0.054	-0.421
Am37	F:* TCTCACCGGCGATAGAAACT R: ATACCGGAAGACGGAATGTG	50	$(TC)_9$	175-204	3	0.107	0.078	-0.362
Am40	F:* CACTGTTGTGGTGGTGGTTT R: GTGGGTGGAGCTAATGTGGA	52	$(CA)_3G(CA)_2N(CA)_5$	182-266	8	0.402	0.409	0.016

* M13 tail (CACGACGTTGTAAAACGAC), optimized annealing temperature (T_a), repeat motifs, size ranges of PCR products, number of alleles observed (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), fixation index (F_{IS}).

Appendix 2 Primer pairs that successfully amplified microsatellite-containing loci in *Arctium minus*. Given the low genetic diversity in the populations used in this study, these monomorphic loci could be assessed for polymorphism in genetically more diverse populations of *A. minus* or in closely related species.

Locus	Primer Sequence (5'→3')	T_a (ºC)	Repeat Motif	Expected size (bp)
Am16	F:* TGTTTCTGCATAGTTCCAAGGTT R: AGGCATTCAAATCAACAATCC	45	$(AAG)_6$	216
Am23	F:* ATCGCCATCGAAGACAAGAC R: AATAAAATCCTGAGCCGGAAA	54	$(AAG)_9$	153
Am25	F:* GGTTTGGTTCTCCCTCAGGT R: ATCAAGCCGGTGACCATATC	45	$(TC)_8$	278
Am27	F:* CGGGTTCTGACTTAGCTTGC R: GGTAACATCCGTTTCGTTCG	50	$(AAG)_6$	361
Am36	F:* TGTTATTTCAGCCCTGGATTG R: CAACTTACAATTTCAATGGTATCTCTC	54	$(CA)_{11}$-$(TA)_6$-$(GA)_5$	204

* M13 tail (CACGACGTTGTAAAACGAC), optimized annealing temperature (T_a), repeat motifs, expected size ranges of PCR products.