NOTE

High degree of conservation of nuclear microsatellite loci in the genus *Clusia*

M.L. Hale, A.M. Borland, and K. Wolff

Abstract: In plants, microsatellites and their flanking DNA are rarely conserved across a whole genus, let alone other genera in the same family. Therefore, the possibility of using microsatellite primers developed for a species across a large number of plant species in the same genus is often limited. Remarkably, dinucleotide nuclear microsatellites developed for *Clusia minor* and for *Clusia nemorosa* amplified homologous microsatellites in species across the whole genus *Clusia*. In this present study, we report on the DNA sequence variation across the genus of 3 microsatellite loci with varying levels of variation. Compared over the species, there was a correlation between the lengths of the microsatellite loci. Interrupts occurred multiple times and did not seem to lead to the death of the microsatellite. These highly conserved markers will be useful for studying the variable reproductive systems in the genus *Clusia*.

Key words: microsatellite, *Clusia*, cross-species amplification, microsatellite evolution.

Résumé : Chez les plantes, les microsatellites et les séquences qui les bordent sont rarement conservés au sein d'un genre et encore moins chez différents genres au sein de la même famille. Ainsi, les possibilités d'exploiter des amorces développées chez une espèce au sein d'une gamme d'espèces appartenant au même genre sont souvent limitées. De façon remarquable, les microsatellites nucléaires dinucléotidiques développés pour le *Clusia minor* et le *Clusia nemorosa* amplifient des microsatellites homologues chez d'autres espèces à travers tout le genre *Clusia*. Les auteurs rapportent ici la variation nucléotidique pour 3 locus microsatellites présentant divers niveaux de variation au sein de ce genre. En comparant les diverses espèces, une corrélation a été notée entre les longueurs des microsatellites. Des interruptions ont été observées à de nombreuses reprises, mais sans que cela mène à la mort du microsatellite. Ces marqueurs hautement conservés seront utiles pour étudier la remarquable variabilité des systèmes reproductifs au sein du genre *Clusia*.

Mots clés : microsatellites, *Clusia*, amplification interspécifique, évolution des microsatellites.

[Traduit par la Rédaction]

Microsatellites have become an important tool for many areas in plant research. For example, mapping agricultural traits, synteny studies, studying genetic diversity, and plant mating systems. We investigated the transferability of microsatellites in the genus *Clusia* and the molecular evolution in the microsatellites to help us understand their conservation. *Clusia* is a genus mainly consisting of tropical trees and shrubs. It is a biologically very interesting genus, because it shows a wide range of CO_2 fixation pathways and its flowering biology is intriguingly variable (Borland et al. 1998; Gustafsson and Bittrich 2002).

Received 7 March 2005. Accepted 7 May 2005. Published on the NRC Research Press Web site at http://genome.nrc.ca on 21 October 2005.

Corresponding Editor: P. Donini.

M.L. Hale,[1] A.M. Borland, and K. Wolff. School of Biology, Ridley Building, University of Newcastle-upon-Tyne, Newcastle, NE1 7RU, UK.

[1]Corresponding author (e-mail: marie.hale@canterbury.ac.nz).

Transferability of microsatellites between plant species has been tested in a number of cases, usually between species within a genus, but also between species from sister-genera (e.g., Whitton et al. 1997; Peakall et al. 1998). Successful transfer of microsatellites between species can be defined as the amplification of a fragment in a species that was not the source of the library and(or) sequence. In animals, the success rate of transferring primers from 1 species to another has been relatively high, with microsatellites conserved over a period of 35–50 million years, and sometimes even 450 million years (Rico et al. 1996). In plants, this has generally been less successful, with cross-amplification often within genera, but much less often between genera (Peakall et al. 1998). The majority of plant species show little transferability over more than 15–30 million years (Whitton et al. 1997; Karhu et al. 2000).

A previous study showed that 11 loci amplified across all 17 *Clusia* species that were tested, while 2 loci amplified in 10 of 17 species, whereas none amplified in a sister species in the genus *Chrysochlamys* (Hale et al. 2002). However, homology with the original fragment was not determined. In the present paper, we determined whether microsatellite markers have been preserved across the genus *Clusia*, and

 doi: 10.1139/G05-048

we obtained insights into the evolution of microsatellite repeat regions by comparing sequence data of 3 microsatellite loci among 17 congeneric species.

Three microsatellite loci (*Clm1*, *Clm5*, and *Cln2*) were amplified for a total of 26 individuals representing 17 *Clusia* species (Table 1), under the conditions described by Hale et al. (2002) (see Hale et al. 2004 for specimen details). The internal transcribed spacer (ITS) region was amplified for each individual using the primers ITS4 and ITS5 (White et al. 1990). Fragments were amplified in 25 µL reactions (1× *Taq* buffer (16 mmol $(NH_4)_2SO_4$/L, 67 mmol Tris–HCl/L, 0.01% Tween-20), 2.0 mmol $MgCl_2$/L, 0.2 mmol each dNTP/L, 0.2 µmol each primer/L, 1.0 U *Taq* (Bioline, London, UK), and 0.5 µL template DNA) under the following conditions: 94 °C for 5 min followed by 40 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 2 min; with a final extension of 72 °C for 4 min.

All PCR products were sequenced directly after purification using QIAquick® PCR columns (Qiagen, Crawley, UK). Products were sequenced using BigDye® Terminator Cycle Sequencing chemistry (Applied Biosystems, Warrington, UK) following manufacturer's recommended conditions, and sequences detected on an ABI 310 Prism® automated sequencer (Applied Biosystems). Each DNA fragment was sequenced in both directions.

An additional amplification of each microsatellite locus was conducted as above, but with the forward primer labelled with an ABI fluorescent dye (JOE™, FAM™, or TAMRA™). The labelled microsatellite fragments were detected on an ABI 310 Prism® automated sequencer using an internal size standard (GeneScan®-500 [ROX]™), and were analysed and sized with GeneScan® software (Applied Biosystems). Fragment sizes could then be compared with sequenced PCR products to provide an independent check on the accuracy of the direct sequencing of microsatellite alleles, particularly in the case of heterozygotes.

Sequences were edited and aligned using ProSequence (Filatov et al. 2000). Heterozygotes at the microsatellite loci were easily identified by the overlaid sequences after the microsatellite region and by the difference in number of repeats of the 2 alleles determined from the degree of "frameshift" detected in the sequences. The difference in number of repeats of the 2 alleles was determined from each of the 2 sequences (forward and reverse) independently and compared to test the accuracy of the ability to interpret the mixed sequences. In all cases, the 2 alleles were clearly identified. While the length of each allele in a heterozygote was relatively easy to determine, it was not possible to assign interrupts to a particular allele when a base substitution occurred only on 1 of the 2 alleles. In these cases, IUB codes have been used to indicate whether a base is either A or G (labelled R), C or T (labelled Y), or G or C (labelled S).

A strict consensus maximum parsimony tree, calculated from a total of 12 trees, was constructed for the ITS sequence data using PAUP* 4.0 (Swofford 2002). The ITS maximum parsimony tree generated was consistent with the molecular and morphological phylogeny previously constructed for this taxonomic group (Gustafsson and Bittrich 2002). *Clm5* repeat data were overlaid on this tree to investigate evolutionary processes at this locus.

All alleles sequenced for all 3 loci contained a microsatellite in all accessions (Table 1). The mean allele repeat length calculated from the sequencing data was correlated with the number of alleles detected from the fragment length data (Table 1; $r = 0.999$, $P = 0.010$, df = 1). Although we have only 3 loci to test this relationship with, it seems likely that there is a relationship between microsatellite size and polymorphism.

Clm5 shows evidence for both stepwise mutation and larger indels (Fig. 1). Stepwise mutation appears to have occurred around the 2 repeat sizes of 6 repeats and 10 repeats (Fig. 1); however there is no evidence that increases in allele size are more common than decreases. The greater variability in *Clm1* and *Cln2* makes it difficult to distinguish stepwise mutations from larger indels in these 2 loci. However, at least 1 large insertion has occurred in *Clm1* within *Clusia croatii* (Table 1).

There was a significant correlation between mean repeat length in *Clm1* and mean repeat length in *Cln2* ($r = 0.68$, $P = 0.0002$, df = 21), even when the outlier (*C. croatii*) was ignored ($r = 0.52$, $P = 0.0056$, df = 20). There was also a significant relationship between allele size in *Clm1* and *Clm5* ($r = 0.45$, $P = 0.012$, df = 21), but no relationship between allele size in *Clm5* and *Cln2* ($r = 0.013$). These 3 loci do not show linkage disequilibrium within species, suggesting that the relationship between repeat length across loci is not an artefact of loci being in the same linkage group. The 3 loci should have a similar evolutionary history, apart from perhaps, mutation rate. Mutation rate is expected to be faster for larger microsatellites and this is consistent with the polymorphism data for the 3 loci (positive relationship between microsatellite size and polymorphism). However, the direction of the mutations (i.e., towards increasing or decreasing allele length) must be similar within species across loci. This correlation suggests that certain *Clusia* species or individuals may be more susceptible to microsatellite mutations, all in the same direction, than others, as has been demonstrated in maize (Vigouroux et al. 2003).

For all 3 loci, there was no significant difference between repeat length of "perfect" alleles and of interrupted alleles (*Clm1*: $F = 0.121$, $P = 0.730$, df = 1,26. *Clm5*: $F = 2.922$, $P = 0.098$, df = 1,30. *Cln2*: $F = 0.139$, $P = 0.712$, df = 1,25). Variability still occurred in interrupted microsatellites, usually on the side of the interrupt with a greater number of repeats. For example, in the *Clusia minor* group for *Clm1*, the number of repeats on the left side of the "TS" interrupt varies from 6 to 22, but on the right side is always 4 (Table 1). Thus, it seems that interrupt mutations do not lead to the death of a microsatellite as previously suggested (Krugylak et al. 1998; Kutil and Williams 2001).

Ascertainment bias (greater polymorphism in the focal species) has been noted in some studies (Van Treuren et al. 1997; Peakall et al. 1998), but not in others (Shepherd et al. 2002; Nunome et al. 2003). We found that the focal species (the species for which the microsatellite markers were originally developed) generally has more repeats than the nonfocal species (20 versus 17.15 repeats for *Clm1*, 8 versus 6.67 repeats for *Clm5*, and 20 versus 13.67 repeats for *Cln2*, respectively). The number of repeats was calculated as the mean number of repeats over alleles, ignoring interrupts (data from Table 1). For the focal species, the number of re-

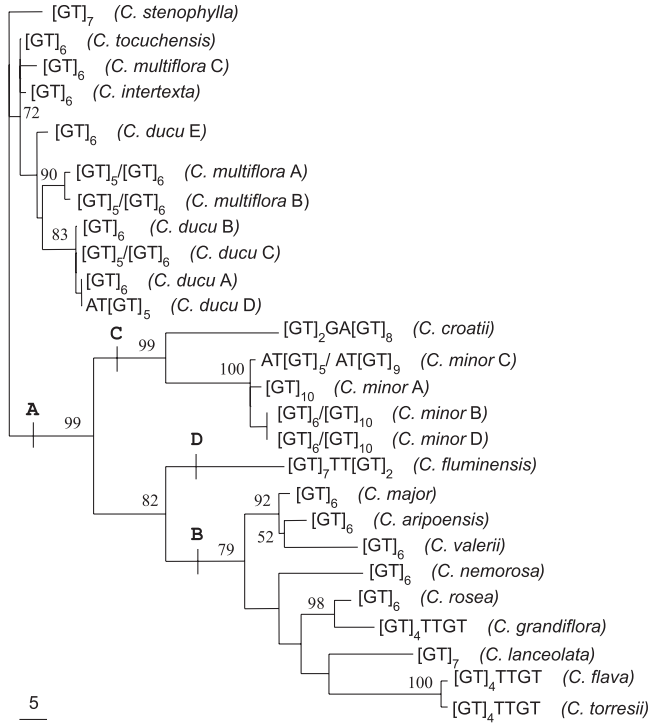
Table 1. Microsatellite composition for each individual at 3 loci. Where individuals were heterozygous for a base substitution, the position of the substitution is indicated with IUB codes: Y = C or T, S = G or C, and R = A or G. Repeat motif has been left blank for the 3 individuals where the quality of the sequence for a particular locus was too poor for accurate determination of repeat length. Mean length is the average number of repeats per allele ignoring substitutions. Length alleles represents the number of alleles detected by GeneScan® fragment length analysis. GenBank accession numbers AY371323 to AY371397.

Specimen	*Clm1*	*Clm5*	*Cln2*
C. aripoensis	$[TC]_{15}$	$[GT]_6$	$[GA]_{14}$
	—	—	—
C. croatii	$[TC]_{18}GC[TC]_4GCTC[TCCC]_6[TC]_3CC[TC]_{16}$	$[GT]_2GA[GT]_8$	$[GA]_{24}$
	—	—	—
C. ducu A	$[TC]_{11}$	$[GT]_6$	$[GA]_9$
	—	—	—
C. ducu B	$[TC]_{16}$	$[GT]_6$	$[GA]_{10}$
	—	—	—
C. ducu C	$[TC]_{10}$	$[GT]_6$	$[GA]_9$
	$[TC]_{13}$	$[GT]_5$	—
C. ducu D	$[TC]_{10}$	$AT[GT]_5$	$[GA]_2GR[GA]_7$
	—	—	$[GA]_2GR[GA]_5$
C. ducu E	$[TC]_{13}$	$[GT]_6$	$[GA]_{10}$
	—	—	—
C. flava	$[TC]_{10}$	$[GT]_4TTGT$	$[GA]_3TA[GA]_{14}$
	$[TC]_{15}$	—	—
C. fluminensis	$[TC]_{11}$	$[GT]_7TT[GT]_2$	$[GA]_7$
	—	—	—
C. grandiflora	$[TC]_{15}$	$[GT]_4TTGT$	$[GA]_9AA[GA]_6$
	—	—	—
C. intertexta	$[TC]_{10}AC[TC]_3$	$[GT]_6$	$[GA]_4CA[GA]_4$
	—	—	—
C. lanceolata	$[TC]_{13}$	$[GT]_7$	$[GA]_{13}$
	$[TC]_{17}$	—	$[GA]_{15}$
C. major	$[TC]_2YC[TC]_5$	$[GT]_6$	$[GA]_9$
	$[TC]_2YC[TC]_{15}$	—	—
C. minor A	$[TC]_{16}TS[TC]_4$	$[GT]_{10}$	—
	$[TC]_{22}TS[TC]_4$	—	—
C. minor B	$[TC]_6TS[TC]_4$	$[GT]_{10}$	$GAAA[GA]_8$
	$[TC]_{13}TS[TC]_4$	$[GT]_6$	—
C. minor C	$[TC]_6TS[TC]_4$	$AT[GT]_5$	—
	$[TC]_{18}TS[TC]_4$	$AT[GT]_9$	—
C. minor D	—	$[GT]_6$	$GAAA[GA]_8$
	—	$[GT]_{10}$	—
C. multiflora A	$[TC]_{12}$	$[GT]_5$	$[GA]_5$
	$[TC]_{11}$	$[GT]_6$	$[GA]_{20}$
C. multiflora B	$[TC]_{12}$	$[GT]_5$	$[GA]_7$
	—	$[GT]_6$	$[GA]_{20}$
C. multiflora C	$[TC]_{21}$	$[GT]_6$	$[GA]_5GG[GA]_7$
	$[TC]_{12}$	—	—
C. nemorosa	$[TC]_5TG[TC]_8$	$[GT]_6$	$[GA]_{20}$
	—	—	—
C. rosea	$[TC]_{16}$	$[GT]_6$	$[GA]_{10}$
	$[TC]_9$	—	$[GA]_{17}$
C. stenophylla	$[TC]_{11}$	$[GT]_7$	$[GA]_{15}$
	—	—	—
C. tocuchensis	$[TC]_{22}$	$[GT]_6$	$[GA]_4GC[GA]_{12}$
	$[TC]_{27}$	—	$[GA]_4GC[GA]_{15}$
C. torresii	$[TC]_{16}$	$[GT]_4TTGT$	$[GA]_{19}$
	$[TC]_{22}$	—	—
C. valerii	$[TC]_3TT[TC]_{14}$	$[GT]_6$	$[GA]_6SA[GA]_5$

Table 1. (*concluded*).

Specimen	*Clm1*	*Clm5*	*Cln2*
	—	—	$[GA]_6SA[GA]_{15}$
Mean length	15.70	6.80	13.38
Length alleles	20	7	17

Fig. 1. Maximum parsimony tree constructed from ITS data, overlaid with the microsatellite sequence for locus *Clm5* for each individual. Bootstrap values greater than 50 are indicated (100 replications). Both alleles, separated by a forward slash, are given for heterozygotes. A 4-repeat insertion appears to have occurred at position A, with loss of the insert at position B, and point mutation (G/T) at position D. Or the 4-repeat insert could have occurred twice, once at position C and once at position D. Alternatively, there could have been a 3-repeat insert at position D, along with a "TT" insert. The tree is consistent with the previous molecular and morphological phylogeny for these taxa and is rooted with members of the section Anandrogyne as the basal group in accordance with Gustafsson and Bittrich (2002).



peats was averaged over all alleles found in *C. minor* individuals for *Clm1* and *Clm5*, and over all alleles found in *Clusia nemorosa* for *Cln2*. The number of repeats was averaged over all alleles found in individuals other than the focal species for the "nonfocal species" values above. Since the number of repeats is generally positively correlated with levels of polymorphism, the finding of more repeats in the focal species than in the nonfocal species also indicates ascertainment bias.

A remarkably high percentage of dinucleotide microsatellites developed from genomic libraries for 1 species in the genus *Clusia* were conserved across the whole genus (Hale et al. 2002). In general, the evolutionary divergence time across which the microsatellites amplify in plants is up to 15–30 million years (Whitton et al. 1997; Karhu et al. 2000). However, this seems to be an upper limit and is often only applicable to trinucleotides. From the ITS data, we can tentatively calculate time since divergence of our most divergent species, if we assume the existence of a molecular clock and use rates of evolution calculated for other genera (Sang et al. 1995; Richardson et al. 2001) of between 11.9 and 19.5 million years. It seems that transferability of dinucleotide repeats in *Clusia* is on the upper side of the range generally found. These highly conserved markers will be useful for studying the variable reproductive systems in the genus *Clusia.* For example, microsatellite allele comparisons of parents and offspring will tell us whether outcrossing or selfing is more prevalent in particular taxa and which taxa reproduce apomictically.[2]

Acknowledgements

We thank Mats Gustafsson, Barry Hammel (Missouri Botanic Garden), The Royal Botanic Garden Edinburgh, and The National Herbarium of Trinidad and Tobago for providing us with samples. We also thank Mats Gustafsson and Jane Squirrell for commenting on an earlier version of the manuscript. This research was funded by Natural Environment Research Council Grant NERC B/S/2000/00147.

References

Borland, A.M., Técsi, L.I., Leegood, R.C., and Walker, R.P. 1998. Inducibility of crassulacean acid metabolism (CAM) in *Clusia* species; physiological/biochemical characterisation and intercellular localization of carboxylation and decarboxylation processes in three species which exhibit different degrees of CAM. Planta, **205**: 342–351.

Filatov, D.A., Moneger, F., Negrutiu, I., and Charlesworth, D. 2000. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. Nature (London), **404**: 388–390.

Gustafsson, M.H.G., and Bittrich, V. 2002. Evolution of morphological diversity and resin secretion in flowers of *Clusia* (Clusiaceae): insights from ITS sequence variation. Nord. J. Bot. **22**: 183–204.

Hale, M.L., Squirrell, J., Borland, A.M., and Wolff, K. 2002. Isolation of polymorphic microsatellite loci in the genus *Clusia* (Clusiaceae). Mol. Ecol. Notes, **2**: 506–508.

Hale, M.L., Borland, A.M., Gustafsson, M.H.G., and Wolff, K. 2004. Causes of size homoplasy among chloroplast micro-

[2] Supplementary data for this article are available on the Web site or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 4023. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/irm/unpub_e.shtml.

satellites in closely related *Clusia* species. J. Molec. Evol. **58**: 182–190.

Karhu, A., Dieterich, J.-H., and Savolainen, O. 2000. Rapid expansion of microsatellite sequences in pines. Mol. Biol. Evol. **17**: 259–265.

Krugylak, S., Durrett, R.T., Scug, M.D., and Aquadro, C.F. 1998. Equilibrium distributions of microsatellite length resulting from a balance between slippage events and point mutations. Proc. Natl. Acad. Sci. U.S.A. **95**: 10 774–10 778.

Kutil, B.L., and Williams, C.G. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J. Hered. **92**: 327–332.

Nunome, T., Suwabe, K., Iketani, H., and Hirai, M. 2003. Identification and characterisation of microsatellites in eggplant. Plant Breed. **122**: 256–262.

Peakall, R., Gilmore, S., Keys, W., Morgante, M., and Rafalski, A. 1998. Cross-species amplification of soybean (*Glycine max)* simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol. Biol. Evol. **15**: 1275–1287.

Richardson, J.E., Weitz, F.M., Fay, M.F., Cronk, Q.C.B., Linder, H.P., and Chase, M.W. 2001. Phylogenetic analysis of *Phylia* L. with an emphasis on island species: evidence from plastid *trn*L-F DNA and nuclear internal transcribed spacer (Ribosomal DNA) sequences. Taxon, **50**: 405–427.

Rico, C., Rico, I., and Hewitt, G. 1996. 470 millions years of conservation of microsatellite loci among fish species. Proc. Roy. Soc. Lon. B, **263**: 549–557.

Sang, T., Crawford, D.J., Kim, S.-C., and Stuessy, T.F. 1995. Radiation of the endemic genus *Dendroseris* (Asteraceae) in the Juan Fernandez Islands: evidence from sequences of the ITS region of nuclear ribosomal DNA. Am. J. Bot. **81**: 1494–1501.

Shepherd, M., Cross, M., Mahuire, T.L., Dieters, M.J., Williams, C.G., and Henry, R.J. 2002. Transpecific microsatellites for hard pines. Theor. Appl. Genet. **104**: 819–827.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. [computer manual]. Sinauer Associates, Inc., Sunderland, Mass. U.S.A.

Van Treuren, R., Kuittinen, H., Kärkkäinen, K., Baena-Gonzalez, E., and Savolainen, O. 1997. Evolution of microsatellites in *Arabis petrae* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Mol. Biol. Evol. **14**: 220–229.

Vigouroux, Y., Matsuoka, Y., and Doebley, J. 2003. Directional evolution for microsatellite size in maize. Mol. Biol. Evol. **20**: 1480–1483.

White, T.J., Bruns, T., Lee, S., and Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* PCR protocols: a guide to methods and applications. *Edited by* M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White. Academic Press, Inc., New York. pp. 315–322.

Whitton, J., Riesberg, L.H., and Ungerer, M.C. 1997. Microsatellite loci are not conserved across the Asteraceae. Mol. Biol. Evol. **14**: 204–209.