

# Somatic Mutations at Microsatellite Loci in Western Redcedar (*Thuja plicata*: Cupressaceae)

L. M. O'CONNELL AND K. RITLAND

From Natural Resources Canada, Canadian Forest Service, Atlantic Forestry Centre, P.O. Box 4000, Fredericton, NB, E3B 5P7, Canada (O'Connell), and the Department of Forest Sciences, University of British Columbia, 2424 Main Mall, Vancouver, British Columbia, V6T 1Z4, Canada (Ritland). We thank S. Otto for discussion and comments on an earlier version of this manuscript and J. Russell for his help collecting samples. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) research grant to KR, and an NSERC postgraduate scholarship and Izaak Walton Killam predoctoral Fellowship to LMO.

Address correspondence to Lisa M. O'Connell at the above address, or e-mail: loconnel@nrcan.gc.ca.

A per-generation somatic mutation rate for microsatellites was estimated in western redcedar (*Thuja plicata*, Donn ex D. Don.: Cupressaceae). A total of 80 trees representative of the average size and age of reproductive trees were sampled in four natural populations in southwestern British Columbia. Samples of bulked haploid megagametophytes were collected from two or three positions on each tree, assuming that the collections were far enough apart that the same mutant sector was not sampled twice. All samples were genotyped at eight microsatellite loci. A single mutation corresponding to a stepwise increase in one dinucleotide repeat was detected. The estimated mutation rate for microsatellites was  $6.3 \times 10^{-4}$  mutations per locus per generation (or  $3.1 \times 10^{-4}$  per allele per generation), with a 95% confidence interval of  $3.0 \times 10^{-5}$  to  $4.0 \times 10^{-3}$  mutations per locus. Somatic mutations can contribute to a greater mutational load in trees, as compared to shorter lived plants, and genotypic mosaics within an individual have important implications for plant defense strategies and plant evolution.

In plants, somatic mutations (i.e., mutations arising from mitosis) can be a significant source of new genetic variation, both within and between individuals. Genetic variation within individuals offers the opportunity for cell lineage selection (Otto and Hastings 1998) and could be important in plant defense by creating a mosaic of different environments for insect pests (Antolin and Strobeck 1985; Whitham and Slobodchikoff 1981). At the population level, somatic mutations can change allele frequencies (Orive 2001). Somatic mutations are important in the evolution of plant mating systems, particularly in long-lived species such as forest trees, because they contribute to mutational load and inbreeding

depression, favoring the evolution of predominantly outcrossing mating systems (Barrett et al. 1996; Morgan 2001).

Somatic mutations can be detected at either genetic marker loci, such as RAPDs (as in aspen clones, *Populus tremuloides*; Tuskan et al. 1996), or at conspicuous morphological loci, such as chlorophyll deficiency (as in six species of Cupressaceae, Korn 2001). Microsatellites, or simple sequence repeats (SSRs), offer a special opportunity to observe and study somatic mutations, because their rate of mutation is several orders of magnitude greater than that of other DNA markers (Ellegren 2000b). Microsatellites consist of tandemly repeated units of DNA of one to six base pairs, and their tandem nature results in mutations due to replication slippage or slipped-strand mispairing during DNA replication (Levinson and Gutman 1987). Microsatellites are popular markers in population genetic studies, because they are highly variable and codominant. Ultimately, observations of microsatellite mutations will enable more accurate inferences based upon microsatellite mutation models, because such observations provide information about the sizes (change in repeat number) and rates (numbers per mitosis or per generation) of microsatellite mutations.

Western redcedar (*Thuja plicata*, Donn ex D. Don: Cupressaceae) is a conifer with a mixed mating system and low isozyme diversity (El-Kassaby et al. 1994; O'Connell et al. 2001; Yeh 1988). Individual trees can live up to 1,000 years and attain heights of more than 50 m (Minore 1990). Therefore, western redcedar provides a good opportunity to detect and characterize new microsatellite mutations arising from somatic processes. In this study, we sampled haploid megagametophyte tissue in *Thuja plicata* to detect mutations in a long-lived plant. An advantage of using megagametophytes to detect mutations is that they are part of the germ line, so the new mutations are heritable. Observations of micro-

satellite mutations will provide information on the generational somatic mutation rate and the magnitude of size changes of microsatellite mutations.

## Materials and Methods

### Estimating Mutation Rate

Mitotic mutation rates can be estimated by calculating the number of cell divisions leading to a new mutation, but this estimation involves several assumptions, such as constancy of cell sizes and the fidelity of apical meristems. In addition, when several samples are made throughout a tree, the uncertain origin of cell lineages leading to different sampled tissues further complicates these calculations. Instead, in this study we employed a simple method to estimate *per-generation* mutation rates, as opposed to *per-mitosis* rates. On a per-generation basis, the mutation rate is found by simply observing the frequency of new mutations in seed-producing tissue. This method is similar to those based on the number of genomes sampled (Schlötterer et al. 1998; Udupa and Baum 2001; Vázquez et al. 2000; Vigouroux et al. 2002).

The mutation rate per locus per generation is estimated as  $U = m/NLK$ , where  $m$  is the number of mutations observed (number of times that genetic differences among sampled tissues within a tree was observed),  $N$  is the number of tissues sampled per tree,  $L$  is the number of trees sampled, and  $K$  is the number of loci sampled. This estimator is derived as follows. If  $u$  is the expected per-generation somatic mutation rate, the distribution of the number of mutations found in a sample is the terms in the expansion of  $LK(u + (1 - u))^N$ . If  $u$  is small, only two terms predominate:  $LK(1 - u)^N$  (trees with no mutations) and  $LKNu(1 - u)^{N-1}$  at  $LKNu$  (trees with one sample of  $N$  mutant). The estimator is then obtained by equating this latter term to the observed numbers of mutants in the total sample ( $m$ ) and solving for  $u$ .

This estimator for a per-generation mutation rate makes two major assumptions: (1) that the seed-producing tissue sampled represents the historical average age of reproduction for the species and (2) that new mutations are identified in an unbiased manner. Regarding (1), trees of average mature age should be sampled. Regarding (2), one needs to sample at least two tissues per tree to detect mutational changes. We assume that mutant sectors are sufficiently small, such that not all samples are mutant; collection of tissues at points mutually separated by the largest number of mitoses should minimize this possibility of sampling only mutant tissues. Overall, to the extent that the trees represent the average age of reproduction, this estimator would slightly underestimate the true mutation rate, because of the slight possibility that all samples within a tree are new mutants.

It does not matter if a mutant sector has been missed, because the expected estimate of the frequency of mutant sectors equals the observed fraction of trees with mutant sectors; it does not matter whether all mutant sectors have been sampled, but only that they have been sampled in an unbiased manner. We also assume that multiple independent mutations do not occur in the same tree; given the relatively

low mutation rate (about one in 1,000), this event should be highly improbable and not significantly affect our estimate.

The 95% confidence interval was calculated with use of the Wilson score method, with continuity correction appropriate for samples sizes above 400 (Newcombe 1998; Wilson 1927). The lower and upper confidence limits were calculated as follows:

$$Lower = \frac{2np + \bar{z}^2 - 1 - \bar{z}\sqrt{\bar{z}^2 - 2 - 1/n + 4p(nq + 1)}}{2(n + \bar{z}^2)}, \quad (1)$$

and

$$Upper = \frac{2np + \bar{z}^2 + 1 + \bar{z}\sqrt{\bar{z}^2 + 2 - 1/n + 4p(nq - 1)}}{2(n + \bar{z}^2)}, \quad (2)$$

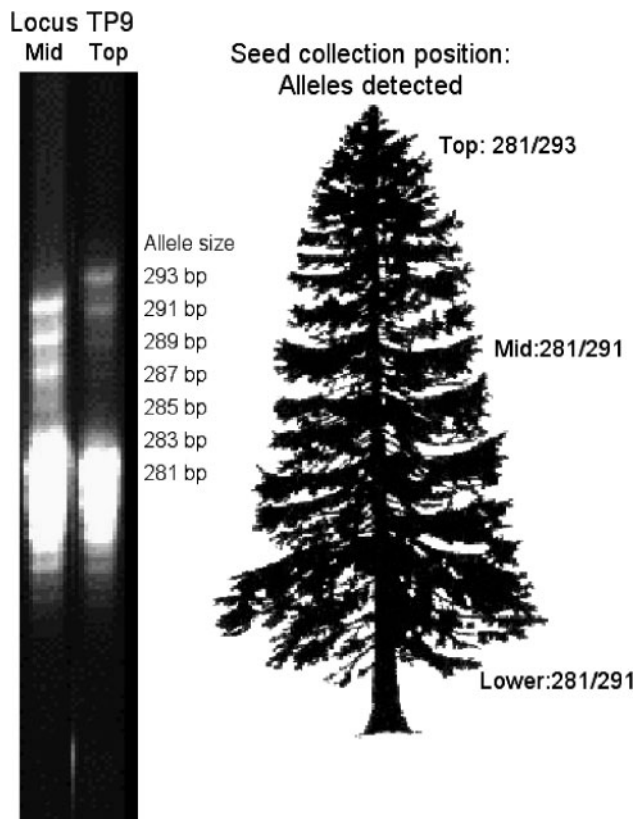
where  $n$  is the sample size,  $p$  is the proportion of mutations observed ( $U$ ),  $q = 1 - p$ , and  $\bar{z}$  is the standard normal deviate associated with a two-tailed probability (Newcombe 1998).

### Sample Collections

During the autumn of 1999 mature cones were collected from a total of 80 trees in four natural populations in southwestern British Columbia (20 trees per population; populations BC12, BC13, BC14, and BC15, O'Connell 2003). Western redcedar trees produce seed cones throughout the crown, including the lower branches that often reach the ground (O'Connell L, personal observation). Cones were collected from reproductive trees ranging from 4.8 to 36.8 m in height (average height = 20.1 m  $\pm$  6.7 SD). In three of the populations, cones were collected from two branches from three different heights (top, middle, and lower) in most trees, but only from the top and lower branches in the shorter trees (Figure 1). In one population, cone collections were made from one branch from each of two positions (top and lower). Thus our sampling attempted to minimize the probability that all samples within a tree came from the same mutant sector (if the sector existed).

To ensure that trees representative of the average age of reproduction were sampled, we also attempted to sample reproductive trees spanning all heights in a population, from the shortest and the tallest. The trees sampled are therefore representative of the range in size of mature redcedar trees in the geographical area studied, so a mutation rate per tree, based on these trees, is reasonable.

Seeds were mechanically extracted from cones and stored at 4°C until germination. Seeds were germinated following O'Connell et al. (2001), and haploid megagametophytes were separated from the seedlings. From each collection position 10 megagametophytes were bulked. A total of 20 to 60 megagametophytes per tree were sampled, and we extracted DNA, using a modified CTAB method (Doyle and Doyle 1987). This use of megagametophytes effectively captures the somatic tissue just before its entrance into the germ line (new mutations arising from meiosis are not detected in these bulks; O'Connell 2003).



**Figure 1.** Image of a microsatellite gel showing the genotype at locus *TP9* for two different heights within the same tree (left). Two collections of 10 bulked megagametophytes were made from three heights in each tree (right); 281 bp = 29 dinucleotide repeats, 291 bp = 34 repeats, and 293 bp = 35 repeats.

## Microsatellites

Each sample of 10 bulked megagametophytes per branch was genotyped at eight polymorphic microsatellite loci developed for *Thuja plicata* (O'Connell and Ritland 2000). Microsatellite repeat motifs ranged from simple dinucleotide repeats to more complex and interrupted motifs (Table 1). Polymerase chain reactions (PCRs) and allele scoring on a LI-COR 4200 sequencer (LI-COR Inc., Lincoln, Nebraska) were carried out as described in O'Connell and Ritland (2000).

## Results

### Microsatellite Mutations

After screening the material at eight microsatellite loci, we found a single new allele at locus *TP9*. Alleles 281 and 291 were found in the lower and middle part of a tree, and alleles 281 and 293 in the top part (Figure 1). The new allele (293) likely arose in the upper part of the tree. To confirm that the new allele was not a PCR artifact, all the samples from the

**Table 1.** Description of eight microsatellite loci used to genotype 80 *Thuja plicata* trees from four natural populations

Locus	Repeat motif	N	bp	A
<i>TP1</i>	(CA) <sub>N</sub>	13–34	166–208	12
<i>TP3</i>	(TG) <sub>N</sub>	9–36	178–232	15
<i>TP4</i>	(TG) <sub>N</sub>	10–23	282–308	9
<i>TP6</i>	(GC) <sub>N</sub> (GT) <sub>N</sub> (ATATGT) <sub>N</sub> ... (GT) <sub>N</sub>	78–110	231–295	28
<i>TP7</i>	(CA) <sub>N</sub>	11–29	241–277	15
<i>TP8</i>	(CA) <sub>N</sub> CG(CA) <sub>N</sub>	23–49	208–266	21
<i>TP9</i>	(AC) <sub>N</sub>	20–59	263–341	27
<i>TP11</i>	(CT) <sub>N</sub> (CA) <sub>N</sub>	25–33	220–232	7

N is number of dinucleotide repeats; bp is range of allele lengths in base pairs; A is total number of alleles detected.

tree with the new allele were reamplified and rescored from the same DNA extraction. The same allele sizes were observed each time.

### Type of Mutation

The size of the new allele corresponded to an increase in one dinucleotide repeat: from 34 to 35 repeats. The new allele size already existed in the sampled populations and was near the middle of the distribution of allele sizes at locus *TP9* (Figure 2).

### Somatic Mutation Rate Estimate

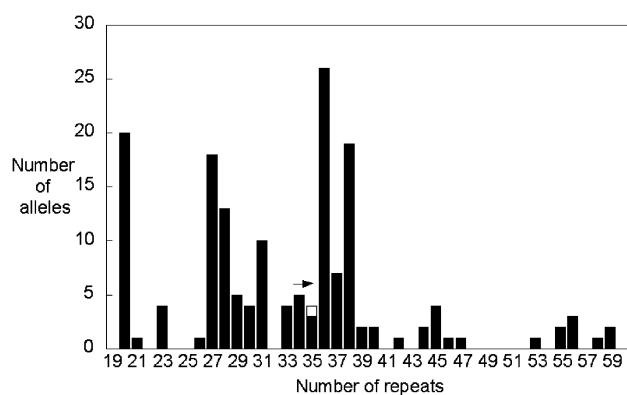
Two positions were sampled in 42 trees, and three positions were sampled in 38 trees. Using the above estimator and its assumptions, we estimated rate of somatic mutation for microsatellites to be  $U = m/NLK = 1/((2 \times 42 \times 8) + (3 \times 38 \times 8)) = 1/1,584 = 6.3 \times 10^{-4}$  mutations per locus per generation, with a 95% confidence interval of  $3.0 \times 10^{-5}$  to  $4.0 \times 10^{-3}$ .

## Discussion

### Somatic Mutation Rate

We observed a single somatic mutation occurring in the upper crown of a redcedar tree. The estimated mutation rate of  $6.3 \times 10^{-4}$  per locus per generation (or  $3.1 \times 10^{-4}$  per allele per generation) is within the expected range of  $10^{-3}$  to  $10^{-4}$  mutations per generation generally reported for microsatellites (Ellegren 2000b). In plants, microsatellite mutations rates have been estimated from mutations accumulated in inbred lines. In maize (*Zea mays* subsp. *mays*) the estimated mutation rate for 142 microsatellite loci was  $7.7 \times 10^{-4}$  mutations per allele per generation (Vigouroux et al. 2002). Udupa and Baum (2001) estimated microsatellite mutation rates of  $1.0 \times 10^{-2}$  and  $3.9 \times 10^{-3}$  mutations per allele per generation in two annual varieties of chickpea (*Cicer arietinum*: Fabaceae). These methods would have captured both somatic and meiotic mutations, so mutation rates should be higher for somatic mutations only.

The actual mutation rate per generation in western



**Figure 2.** Allele distribution at locus *TP9* over four populations of *Thuja plicata* ( $N = 80$  trees). The new allele, which increased from 34 to 35 dinucleotide repeats, is indicated by the white box, with the arrow showing the original allele.

redcedar is likely higher than reported here, because the method we used was not sensitive enough to detect meiotic mutations. Although the material sampled was part of the germ line and offered the opportunity to detect meiotic mutations, doing so would have been difficult because the megagametophyte material was bulked. Results from O'Connell (2003) showed that alleles occurring at a low frequency, such as a new mutation arising during meiosis, would likely not be detectable in a bulk of 10 megagametophytes.

Strictly speaking, a generation's worth of somatic growth should mean something quite precise if we take the demographic definition of the term "generation" into account. It should reflect the survivorship and fecundity schedule of the population. These values are quite difficult to estimate outside a controlled setting and for a long-lived species such as western redcedar. During our sampling, we attempted to sample from trees representative of the age structure of reproductive individuals in a population, in order to approximate a generation.

### Mutation Model

Information on the mutation processes of microsatellites will help provide more accurate mutation models for population genetics. Distance measures and timing of evolutionary events depend on accurate mutation models. The observed mutation was stepwise, increasing in size by one base pair (Figure 2). Similarly, of 71 observed microsatellite mutations in maize, Vigouroux et al. (2002) found that changes of a single repeat (83% of mutations) were more common than changes of multiple repeats (17%), and a higher proportion of alleles mutated to a larger allele size (79%) than to a smaller allele size (21%). The same directional biases have been reported in birds and humans (Ellegren 2000a; Primmer et al. 1996).

Microsatellites are known to exhibit extensive homoplasy (unrelated alleles of the same size), and correspondingly the new allele in this study mutated to an already existing

allele size in the populations sampled. Different loci, and even different alleles, probably mutate at different rates (Ellegren 2000b). Mutations seem to occur in longer loci or alleles, and at loci with simple repeats as opposed to more complex loci. The locus with the observed mutation, *TP9*, is one of the most variable loci in western redcedar, with 27 alleles detected in 80 individuals (Table 1). In a range-wide study, 41 alleles were detected in 620 individuals at locus *TP9* (O'Connell 2003). Changes in allele size can also be caused by changes in the DNA regions flanking the microsatellite. This is unlikely in this study, because the new allele corresponded exactly to an increase in one dinucleotide repeat (two extra base pairs) longer than the original allele.

### The Consequences of Somatic Mutations in Redcedar

Generation time can be anywhere between 30 and 1,000 years in *Thuja plicata*. A high per-generation mutation rate at microsatellite loci can account for the high diversity at these markers compared to other genetic markers, despite there being only a few hundred generations since a population bottleneck during the last glaciation about 15,000 years ago (O'Connell 2003). In a long-lived species such as redcedar, deleterious somatic mutations also have the potential of increasing the mutational load.

### Genetic Mosaicism

The new allele was found in megagametophytes collected from branches at the top of the tree, but not in any of the lower collections. Korn (2001) showed evidence of a single apical initial cell in several Cupressaceae species. The mutation probably occurred in the apical cell, and all cells above this apical meristem contained the new allele leading to a large sector with a novel genotype. That trees can be mosaics of cells of different genotypes poses interesting questions about plant defense strategies and plant evolution (Antolin and Strobeck 1985; Gill et al. 1995; Whitham and Slobodchikoff 1981).

### References

- Antolin MF and Strobeck C, 1985. The population genetics of somatic mutation in plants. *Am Nat* 126:52–62.
- Barrett SCH, Harder LD, and Worley AC, 1996. The comparative biology of pollination and mating in flowering plants. *Philos Trans R Soc Lond B Biol Sci* 351:1271–1280.
- Doyle JJ and Doyle JL, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15.
- El-Kassaby YA, Russell J, and Ritland K, 1994. Mixed mating in an experimental population of western red cedar, *Thuja plicata*. *J Hered* 85: 227–231.
- Ellegren H, 2000a. Heterogenous mutation processes in human microsatellite DNA sequences. *Nat Genet* 24:400–402.
- Ellegren H, 2000b. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* 16:551–558.

- Gill DE, Chao L, Perkins SL, and Wolf JB, 1995. Genetic mosaicism in plants and clonal animals. *Annu Rev Ecol Syst* 26:423–444.
- Korn RW, 2001. Analysis of shoot apical organization in six species of the Cupressaceae based on chimeric behavior. *Am J Bot* 88:1945–1952.
- Levinson G and Gutman GA, 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–211.
- Minore D, 1990. *Thuja plicata* Donn ex D. Don. In: *Silvics of North America*, vol. 1, conifers (Burns RM and Honkala BH, eds). Washington, DC: USDA Forest Service Agriculture Handbook; 654.
- Morgan MT, 2001. Consequences of life history for inbreeding depression and mating system evolution in plants. *Proc R Soc Lond Ser B Biol Sci* 268:1817–1824.
- Newcombe RG, 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 17:857–872.
- O'Connell LM, 2003. The evolution of inbreeding in western redcedar (*Thuja plicata*: Cupressaceae) (PhD dissertation). Vancouver: University of British Columbia.
- O'Connell LM and Ritland CE, 2000. Characterization of microsatellite loci in western redcedar (*Thuja plicata*). *Mol Ecol* 9:1920–1922.
- O'Connell LM, Viard F, Russell J, and Ritland K, 2001. The mating system in natural populations of western redcedar (*Thuja plicata*). *Can J Bot* 79: 753–756.
- Orive ME, 2001. Somatic mutations in organisms with complex life histories. *Theor Popul Biol* 59:235–249.
- Otto SP and Hastings IM, 1998. Mutation and selection within the individual. *Genetica* 102/103:507–524.
- Primmer CR, Ellegren H, Sainio N, and Møller AP, 1996. Directional evolution in germline microsatellite mutations. *Nat Genet* 13:391–393.
- Schlötterer C, Ritter R, Harr B, and Brem G, 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol Biol Evol* 15:1269–1274.
- Tuskan GA, Francis KE, Russ SL, Romme WH, and Turner MG, 1996. RAPD markers reveal diversity within and among clonal and seedling stands of aspen in Yellowstone National Park, USA. *Can J For Res* 26:2088–2098.
- Udupa SM and Baum M, 2001. High mutation rate and mutational bias at (TAA)<sub>n</sub> microsatellite loci in chickpea (*Cicer arietum* L.). *Mol Gen Genomics* 265:1097–1103.
- Vázquez JF, Pérez T, Albornoz J, and Domínguez A, 2000. Estimation of microsatellite mutation rates in *Drosophila melanogaster*. *Genet Res* 76:323–326.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, and Doebley J, 2002. Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol* 19:1251–1260.
- Whitham TG and Slobodchikoff CN, 1981. Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability: the adaptive significance of somatic mutation in plants. *Oecologia* 49:287–292.
- Wilson EB, 1927. Probable inference, the law of succession, and statistical inference. *J Amer Statist Assoc* 22:209–212.
- Yeh FC, 1988. Isozyme variation of *Thuja plicata* (Cupressaceae) in British Columbia. *Biochem Syst Ecol* 16:373–377.

Received May 3, 2003

Accepted December 11, 2003

Corresponding Editor: Prem Jauhar