A Low Mutation Rate For Chloroplast Microsatellites

Jim Provan,* Nicole Soranzo,* Neil J. Wilson,* David B. Goldstein† and Wayne Powell*

**Department of Cell and Molecular Genetics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, United Kingdom and †Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom*

Manuscript received May 18, 1999
Accepted for publication June 17, 1999

ABSTRACT

We used chloroplast simple sequence repeats (cpSSRs) to examine whether there is any variation present in the chloroplast genome of *Pinus torreyana* (Parry ex Carrière) that may previously not have been detected using RFLPs. Analysis of 17 cpSSR loci showed no variation, which is consistent with previous cpRFLP work and confirms that the species is descended from an original, highly monomorphic population following a bottleneck. This lack of biological variation in the chloroplast genome of *P. torreyana* allowed us to estimate the mutation rates at cpSSR loci as between 3.2×10^{-5} and 7.9×10^{-5}. This estimate is lower than published mutation rates at nuclear SSR loci but higher than substitution rates elsewhere in the chloroplast genome.

TORREY pine (*Pinus torreyana* Parry ex Carrière) is a rare endemic conifer with <10,000 individuals existing as two extant populations in southern California. Previous studies on morphological (Haller 1986), isozyme (Ledig and Conkle 1983), and chloroplast restriction fragment length polymorphism (RFLP) (Waters and Schaal 1991) variation in *P. torreyana* have shown that there is very little genetic variation present in the species. This is in contrast to most other tree species, which generally exhibit the highest levels of polymorphism within the plant kingdom (Hamrick *et al.* 1979). Ledig and Conkle (1983) have suggested that numbers of *P. torreyana* were reduced to a few (<50) individuals during the Xerothermic period (8500–3500 YBP), an event that would result in the observed depletion of genetic variation in the species.

Simple sequence repeats (SSRs) are rapidly becoming established as an extremely useful tool in population genetics due to their high levels of variability and codominance, which have allowed the analysis of natural populations to be carried out at a higher degree of resolution than had previously been possible using isozymes (Jarne and Lagoda 1996; Powell *et al.* 1996). SSR loci have also been detected in the chloroplast genomes of plants and chloroplast SSRs (cpSSRs) have been used in population and systematic studies in a variety of species (Powell *et al.* 1996; Provan *et al.* 1999a). Knowledge of mutation rates at SSR loci is important because they determine levels of variability within populations and hence greatly influence estimates of population structure. In animals, observed SSR mutation rates have been of the order of 10^{-3}–10^{-4} for autosomal repeat loci (Wiessenbach *et al.* 1992; Weber and Wong 1993) but to date there have been no figures published for mutation rates at simple repeat loci in the chloroplast genome—indeed, to date no SSR mutation rates for any plant genomes have been published. Here we present the first calculation of a mutation rate for cpSSRs based on the analysis of cpSSR loci in Torrey pine.

Corresponding author: Jim Provan, Department of Cell and Molecular Genetics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, United Kingdom. E-mail: j.provan@scri.sari.ac.uk

MATERIALS AND METHODS

Genetic material: Needles were collected from 64 individual trees from the mainland (San Diego) population, which were selected to provide even coverage of the entire range of the population (*ca.* 8000 trees). DNA was extracted from needles using the modified CTAB procedure described by Wagner *et al.* (1987).

Polymerase chain reaction: Primers used for the amplification of 17 cpSSR loci in pines were designed from the complete chloroplast sequence of *P. thunbergii* (Wakasugi *et al.* 1994) and are given in Table 1. PCR was carried out in a total volume of 10 ml containing 1× PCR buffer (10 mm Tris-HCl, 1.5 mm $MgCl_2$, 50 mm KCl, pH 8.3), 200 mm dNTPs, 10 pmol ^{32}P end-labeled forward primer, 10 pmol reverse primer, 0.1 units Taq polymerase (Boehringer Mannheim, Indianapolis), and 50 ng genomic DNA. Reactions were carried out on a Perkin Elmer (Norwalk, CT) 9600 thermal cycler using the following parameters: (1) initial denaturation at 94° for 3 min; (2) 30 cycles of denaturation at 94° for 15 sec, annealing at 60° for 15 sec, and extension at 72° for 60 sec; and (3) final extension at 72° for 5 min. After addition of 10 ml loading buffer (95% formamide), products were resolved on 6% denaturing polyacrylamide gels containing 1× TBE buffer and 8 m urea at 80 W constant power for 2 hr. Gels were transferred onto 3MM blotting paper (Whatman), dried, and exposed to X-ray film overnight without intensification screens.

TABLE 1

***Pinus sylvestris* chloroplast simple sequence repeat primers**

Locus	Repeat	Location	Primers	Size
PCP1289	$(T)_{17}$	*psbA*/ORF22 intergenic region	TCCTGGTTCCAGAAATGGAG TAATTTGGTTCCAGAATTGCG	113
PCP9434	$(A)_{10}$	ORF59a within *trnG* intron	AAACTGACGTAGATGCCATGG GCGGTATGAGGGAAGAAGC	131
PCP26106	$(A)_{14}$	*rpo*B/ORF46a intergenic region	AATCCGACAAAAAAGATTCGG GCTCCATTTCACGTGGTTG	149
PCP30277	$(A)_{12}(G)_{10}$	ORF46b	TGTTGATGTCGTAGCGGAAG ATGAAATGAATCACTTCCCCC	138
PCP36567	$(T)_{11}$	*psbJ*/ORF42b intergenic region	AAAAGAGGAGGAAAAACACCTT AAGAGCAGACAAGTAAGGGGC	115
PCP41131	$(T)_{11}$	ORF119	AAAGCATTTCCAGTTGGGG GGTCAGGATTCATGTTCTTCC	138
PCP45071	$(T)_{15}$	ORF46d/*atp*B intergenic region	ACTGGTCTGATCGACCCAAT TTCTACACTTGCGGAAACCC	149
PCP48256	$(T)_{10}$	ORF64a	ACGTTGGACCAGAGCAGG CGAATTTTTCGAAGAACTAGCG	120
PCP51928	$(T)_{10}$	*trn*S/ORF57b intergenic region	CTTTCTACGGAACGGAAAAGG GCACTGCGGGAAAAAAATAA	141
PCP63771	$(T)_{10}$	*rps*19/*rpl*2 intergenic region	TGAACGTGCCATGATCAATT GGGGCTATAGTGCACTTGGAA	140
PCP71987	$(T)_{16}$	IRF169 intron 1	TCTTTGCAAGAAGGATGGCT GGGGAGTAATCCGTGGAATT	113
PCP79987	$(A)_{12}$	*trn*S/ORF49b intergenic region	TTTTCAACAATTGCATTTACCG GGCGGGATAGGAGTCTTTTC	120
PCP87314	$(T)_{14}$	*trn*I/*trn*A intergenic region	TCCAGGATAGCCCAGCTG TATATCCCCCGTACTTGGACC	116
PCP100842	$(A)_{12}$	ORF1756/ORF64c intergenic region	TCAATACAAATGATGGGAGTCG TTTTGCCATATCCTGAAACTCC	146
PCP102652	$(T)_{11}$	*jndh*I/ORF43d intergenic region	TTCCCAGATCCATTGAAATACA TATGTGCGCGATAATTTCCA	117
PCP107165	$(A)_{10}$	ORF44d/ORF40f intergenic region	GTTTTGGATCGGAATGGATG CTATCCATTCTGCCTTCCCA	148
PCP109612	$(A)_{11}$	ORF49c/*rps*12 intergenic region	ATCGAACAACGAGAATAATCCA TTGGGGGTGATAGTGGAAAA	150

RESULTS AND DISCUSSION

We analyzed 17 loci in a sample of 64 *P. torreyana* individuals using primers previously used in a study of cpSSR variation in populations of Scots pine (*P. sylvestris* L.; Provan *et al.* 1998) and found no variation. Previous studies on other species in the genus Pinus using the same cpSSR loci have shown high levels of intrapopulation diversity (Tables 2 and 3). It can be seen from Table 3 that all 7 of the 17 loci tested in three other members of the genus (*P. pinaster*, *P. resinosa*, and *P. halepensis*) were polymorphic. This suggests that lack of polymorphism in *P. torreyana* is not due to loss or interruption of the mononucleotide repeat motif across species, a fact confirmed by sequence analysis (N. Soranzo and J. Provan, unpublished results). The relatively low numbers of repeats are typical of cpSSRs in other plant species, where long stretches of mononucleotide repeats are very rare (Powell *et al.* 1996; Provan *et al.* 1999a). Despite these relatively short repeat lengths, it has been shown that cpSSR primers designed in one species will produce polymorphic products in other species and even in different genera (*e.g.*, primers derived from *Nicotiana tabacum* revealing cpSSR polymorphism in *Solanum* spp.; Provan *et al.* 1999b).

Earlier studies examining chloroplast diversity in *P. torreyana* using RFLPs also found no variation at over 150 restriction sites within the chloroplast genome, which is consistent with an ancient bottleneck believed to have drastically reduced tree numbers during the Xerothermic period (8500–3500 YBP; Ledig and Conkle 1983). Because the chloroplast molecule is paternally inherited in conifers, this suggests that the bottleneck resulted in the existence of a single paternal genotype in the few trees that survived and that the present population has resulted from expansion in the 35–85 generations since the end of the bottleneck (assuming a generation time of ~100 yr). As far as it is known, this represents the last major bottleneck in the history of the species

TABLE 2

Published values for cpSSR diversity in Pinus species

Species	Reference	Diversity
P. leucodermis	Powell *et al.* (1995)	0.000–0.629 (0.316)
P. halepensis	Morgante *et al.* (1997)	0.236–0.925 (0.596)
	Bucci *et al.* (1998)	0.034–0.422 (0.222)
P. resinosa	Echt *et al.* (1998)	0.314–0.920 (0.568)
P. sylvestris	Provan *et al.* (1998)	0.950–0.987 (0.978)
P. pinaster	Vendramin *et al.* (1998)	0.500–0.833 (0.737)

TABLE 3

Numbers of alleles found at cpSSR loci in various Pinus species

	Species				
Locus	*P. sylvestris*[a] (n = 330)	*P. pinaster*[b] (n = 300)	*P. resinosa*[c] (n = 159)	*P. halepensis*[d] (n = 247)	*P. torreyana* (n = 64)
PCP1289	5	—	—	—	1
PCP9434	1	2	3	2	1
PCP26106	3	1	3	8	1
PCP30277	7	6	3	3	1
PCP36567	3	4	2	4	1
PCP41131	4	2	2	5	1
PCP45071	6	—	—	—	1
PCP48256	3	—	—	—	1
PCP51928	1	—	—	—	1
PCP63771	2	—	—	—	1
PCP71987	7	3	3	8	1
PCP79987	3	—	—	—	1
PCP87314	5	3	3	7	1
PCP100842	2	—	—	—	1
PCP102652	3	—	—	—	1
PCP107165	1	—	—	—	1
PCP109612	1	—	—	—	1

—, The locus was not tested.
[a] Provan *et al.* (1998).
[b] Vendramin *et al.* (1988).
[c] Echt *et al.* (1998).
[d] Bucci *et al.* (1998).

TABLE 4

Published values for mutation rates in plant and animal nuclear and organellar genomes

Kingdom	Genome	Type	Rate	Reference
Animal	Nucleus	Substitution	9×10^{-10}–1×10^{-8}	Wolfe *et al.* (1987)
		SSR length polymorphism	10^{-5}–10^{-2}	Jarne and Lagoda (1996)
			1×10^{-3}	Wiessenbach *et al.* (1992)
			6×10^{-4}	Weber and Wong (1993)
	Mitochondrion	Substitution	2×10^{-8}–5×10^{-8}	Wolfe *et al.* (1987)
Plant	Nucleus	Substitution	3×10^{-9}–5×10^{-8}	Wolfe *et al.* (1987)
	Mitochondrion	Substitution	2×10^{-10}–1×10^{-9}	Wolfe *et al.* (1987)
	Chloroplast	Substitution	1×10^{-9}–3×10^{-9}	Wolfe *et al.* (1987)
		SSR length polymorphism	3.2–7.9×10^{-5}	This study

(F. T. Ledig, personal communication). The fact that the San Diego population now contains ~8000 extant *P. torreyana* individuals, compared with the few tens of individuals left after the bottleneck, suggests that regrowth has been fairly uniform.

Because the mutation rate at SSR loci is much higher than elsewhere in the genome, with the number of mutations per generation (μ) believed to be around 10^{-3} (Goldstein and Pollock 1997), we would expect to find variation due to random mutation. The smaller numbers of alleles reported at cpSSR loci compared with nuclear SSRs (Powell *et al.* 1996; Provan *et al.* 1999a), however, would appear to suggest a lower mutation rate at SSR loci in the chloroplast than in the nucleus. If we assume that the genealogy connecting the current chloroplasts is the sort that emerges from a growing population, that is, all branch lengths trace back independently to the common origin 35 generations ago (assuming the bottleneck ended 3500 YBP; the same calculation can be carried out for 8500 YBP), a boundary on the maximum mutation rate consistent with no variation can be calculated as follows. As lineages trace back to the origin of the genealogy, the total branch length in the tree is 64*35. The probability of no mutation occurring, therefore, anywhere in the genealogical tree over 17 loci is taken from the zero term of a Poisson distribution with parameter 64*35*17*μ, where μ is the mutation rate, which we assume is the same across loci. The probability of no mutations occurring is $P = \text{Exp}[^{-38080\mu}]$, and to find a boundary on μ we need to solve $P = e^{-38080\mu}$ for $P = 0.05$. This calculation gives an upper bound on μ, at the 0.05 level, of $\sim 7.9 \times 10^{-5}$. If the bottleneck ended 8500 YBP, the figure would be 3.2×10^{-5}. This approach ignores the possibility that mutations have gone undetected. This is extremely improbable, however, with no mutations observed over 17 loci, because it would depend on two "oppositely directed" mutations at the same locus.

The lower substitution rate of cpDNA compared to the nuclear genome has been documented (Wolfe *et al.* 1987) and it would appear from this study that the mutational processes at simple repeat loci in the chloroplast genome also occur less frequently than those in the nuclear genome. Published values for mutation rates in nuclear and organellar genomes are summarized in Table 4 and show that while cpSSR loci are less variable than nuclear SSRs, at least in mammals because no SSR mutation rates have yet been published in plants, they can be expected to exhibit more variation than cpRFLPs and thus be useful at the intraspecific level (Powell *et al.* 1996; Provan *et al.* 1999a). While it is still unclear whether nuclear SSRs can be applied to phylogenetic studies due to apparent constraints on allele sizes at SSR loci and subsequent discrepancies between allele size and genetic divergence (Bowcock *et al.* 1994; Goldstein and Pollock 1997), the lower mutation rates calculated here for chloroplast simple repeats along with lower observed variances in allele size (Powell *et al.* 1996; Provan *et al.* 1998) suggest that the problem of allele size limitation may not be a factor when cpSSRs are used. Despite this, further examination of cpSSR variation is needed before it can be determined whether cpSSRs may have value in plant phylogenetic studies.

The authors thank Pete Hollingsworth and two anonymous referees for their comments on this manuscript. This work was supported by the Scottish Office Agriculture, Environment and Fisheries Department.

LITERATURE CITED

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd *et al.*, 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

Bucci, G., M., Anzidei, A. Madaghiele and G. G. Vendramin, 1998 Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol. Ecol. **7:** 1633–1643.

Echt, C. S., L. L. DeVerno, M. Anzidei and G. G. Vendramin, 1998 Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol. Ecol. **7:** 307–316.

Goldstein, D. B., and D. D. Pollock, 1997 Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. **88:** 335–342.

Haller, J. R., 1986 Taxonomy and relationship of the mainland and island populations of *Pinus torreyana* (Pinaceae). Syst. Bot. **11:** 39–50.

Hamrick, J. L., Y. B. Linhart and J. B. Mitton, 1979 Relationships between life history characteristics and electrophoretically detectable variation in plants. Annu. Rev. Ecol. Syst. **10:** 173–200.

Jarne, P., and P. J. L. Lagoda, 1996 Microsatellites, from molecules to populations and back. Trends Ecol. Evol. **11:** 424–430.

Ledig, F. T., and M. T. Conkle, 1983 Gene diversity and genetic structure in a narrow endemic Torrey pine (*Pinus torreyana* Parry ex Carr). Evolution **37:** 79–85.

Morgante, M., N. Felice and G. G. Vendramin, 1997 Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck, pp. 407–412 in *Molecular Tools for Screening Biodiversity: Plants and Animals*, edited by A. Karp, P. O. Issac and D. S. Ingrams. Chapman and Hall, London.

Powell, W., M. Morgante, R. McDevitt, G.G. Vendramin and J. A. Rafalski, 1995 Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc. Natl. Acad. Sci. USA **92:** 7759–7763.

Powell, W., G. Machray and J. Provan, 1996 Polymorphism revealed by simple sequence repeats. Trends Plant Sci. **1:** 215–222.

Provan, J., N. Soranzo, N. J. Wilson, J. W. McNicol, G. I. Forrest *et al.*, 1998 Gene pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple sequence repeats. Proc. R. Soc. Lond. Ser. B **265:** 1697–1705.

Provan, J., N. Soranzo, N. J. Wilson, J. W. McNicol, M. Morgante *et al.*, 1999a The use of uniparentally inherited simple sequence repeat markers in plant population studies and systematics, pp. 35–50 in *Molecular Systematics and Plant Evolution*, edited by P. M. Hollingsworth, R. M. Bateman and R. J. Gornall. Taylor and Francis, London.

Provan, J., W. Powell, H. Dewar, G. Bryan, G. C. Machray *et al.*, 1999b An extreme cytoplasmic bottleneck in the modem European cultivated potato (*Solanum tuberosum*) is not reflected in decreased levels of nuclear diversity. Proc. R. Soc. Lond. Ser. B **266:** 633–639.

Vendramin, G. G., M. Anzidei, A. Madaghiele and G. Bucci, 1998 Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. Theor. Appl. Genet. **97:** 456–463.

Wagner, D. B., G. R. Furnier, M. A. Sagmai-Maroof, S. M. Williams, B. P. Dancik *et al.*, 1987 Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA **84:** 2097–2100.

Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki *et*

al., 1994 Loss of all NDH genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA **91:** 9794–9798.

Waters, E. R., and B. A. Schaal, 1991 No variation is detected in the chloroplast genome of *Pinus torreyana*. Can. J. For. Res. **21:** 1832–1835.

Weber, J., and C. Wong, 1993 Mutation of human short tandem repeats. Hum. Mol. Genet. **2:** 1123–1128.

Wiessenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette *et al.*, 1992 A second generation map of the human genome. Nature **359:** 794–801.

Wolfe, K. H., W. H. Li and P. M. Sharp, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNA. Proc. Natl. Acad. Sci. USA **84:** 9054–9058.

Communicating editor: C. Haley