

Letter to the Editor

Sequence-Repeat Polymorphisms Exhibit the Signature of Recombination in Lodgepole Pine Chloroplast DNA

H. Dawn Marshall,* Craig Newton,† and Kermit Ritland*

*Department of Forest Sciences, University of British Columbia, Vancouver, British Columbia, Canada; and

†B.C. Research Inc., Vancouver, British Columbia, Canada

The chloroplast genome is widely used in plant systematic studies (Olmstead and Palmer 1994), in part because it is slowly evolving and is assumed to be non-recombining (Clegg 1993). Microsatellite markers have also been identified within this genome (Powell et al. 1995; Provan et al. 1996; Vendramin et al. 1996; Newton et al. 1999), and these markers are sufficiently variable for phylogeographic studies within a species (Schaal et al. 1998; Newton et al. 1999; Marshall, Newton, and Ritland 2001). Lack of recombination reduces homoplasmy, which in turn increases the precision of phylogenetic inference in such studies.

However, recently there has emerged some evidence of recombination in another organelle, the mitochondrion. Lunt and Hyman (1997) found end products of mitochondrial genome recombination in the nematode *Meloidogyne javanica*. Saville, Kohli, and Anderson (1998) reported a discrepancy in the expected genotypic structure of mitochondrial DNA (mtDNA) sequences in the fungus *Armillaria gallica* relative to expectations under purely clonal transmission, consistent with observations of mitochondrial heteroplasmy. Using both sequence and restriction fragment length polymorphism data from humans, Awadalla, Eyre-Walker, and Maynard Smith (1999) found linkage disequilibria to decrease with increasing physical distance, consistent with mtDNA recombination. Phylogenetic trees constructed using similar data contained a larger number of homoplasies than expected on the basis of simulated data, which may indicate recombination (Eyre-Walker, Smith, and Maynard Smith 1999). Biologically, mtDNA recombination in humans is feasible, since mitochondria contain the necessary enzymes (Thyagarajan, Padua, and Campbell 1996), and a few paternal mitochondria are contributed to the egg during fertilization. The human mtDNA recombination issue is, however, still the subject of much debate (see Eyre-Walker 2000; Ingman et al. 2000).

Here, we look for signatures of recombination in the chloroplast of lodgepole pine (*Pinus contorta*). Lodgepole pine is common in the forests of western North America, with a climactically and edaphically diverse natural range covering over 2.6×10^7 ha. A wind-pollinated outcrosser and aggressive pioneer, it is thought to have colonized its present range in a north-

ward migration following the last glacial stage (MacDonald and Cwynar 1991). While the chloroplast genome is inherited maternally in most seed plants, in lodgepole pine and in the pine family (Pineaceae) it exhibits paternal inheritance (Szimdt, Aldén, and Hällgren 1987; Wagner et al. 1987).

We recently characterized a set of six hypervariable chloroplast (cp) DNA markers for this species (Stoehr and Newton 2001). These markers consist of three mononucleotide repeats (SSRs), two 10-base repeats (VNTRs), and one combination 10-base/mononucleotide locus, distributed at approximately 5–20-kb intervals around the chloroplast genome. A total of 500 trees located throughout the species' range were assayed in a phylogeography study (Marshall, Newton, and Ritland 2001). We noticed a large number (205) of chloroplast haplotypes characterized by substantial phylogenetic homoplasmy, pointing to the possible involvement of recombination in addition to mutation in generating genetic variability.

A difficulty associated with using microsatellites to detect departures from complete linkage is homoplasmy due to recurrent mutation, which obscures homoplasmy caused by recombination. Homoplasmy is expected to be substantial for microsatellites, which evolve according to a stepwise model of mutation and exhibit high mutation rates (Estoup et al. 1995; Jarne and Lagoda 1996).

In light of this problem, we examined our data from three angles. First, we computed the Hill and Robertson (1968) measure of linkage disequilibrium for each pair of loci. This measure is defined for two alleles A and B residing at different loci as

$$r^2 = \frac{(p_{AB} - p_A p_B)^2}{p_A(1 - p_A)p_B(1 - p_B)},$$

where p_A and p_B are the frequencies of the two respective alleles, and p_{AB} is the two-locus gamete frequency. Estimates were combined among alleles and loci by weighting each r^2 in proportion to $p_A(1 - p_A)p_B(1 - p_B)$. Estimates of r^2 for each pair of loci were regressed against physical distance separating loci (loci were at relative nucleotide positions 9905, 28125, 30358, 61070, 69175, and 87358, respectively; the first locus is a compound locus and was treated as two loci, but comparisons within the compound locus were omitted). Significance was assessed by bootstrapping individuals and by randomizing alleles across individuals (within loci). Any significantly negative slope provides evidence for recombination.

Second, we quantified levels of phylogenetic homoplasmy relative to expected levels either in the case of free recombination or in the absence of recombination using the phylogenetic test of linkage disequilibrium described by Burt et al. (1996). In this test, randomized

Key words: microsatellites, chloroplast, lodgepole pine, recombination.

Address for correspondence and reprints: Kermit Ritland, Department of Forest Sciences, University of British Columbia, 3041-2424 Main Mall, Vancouver, British Columbia, Canada V6T 1Z4. E-mail: ritland@interchg.ubc.ca.

Mol. Biol. Evol. 18(11):2136–2138. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

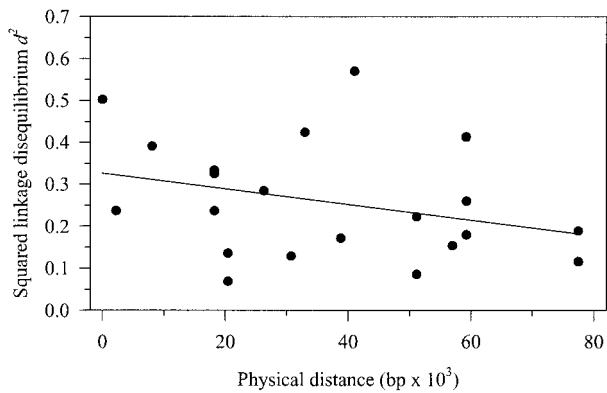


FIG. 1.—Relationship between squared linkage disequilibrium (averaged over alleles) and physical distance separating genes. The slope is significantly negative (see text).

data sets (here, 100) are generated from the original and used to construct the distribution of most-parsimonious tree lengths expected under recombination, against which the original most-parsimonious tree is compared. A tree shorter than expected from the distribution indicates a departure from complete linkage equilibrium.

Third, following Maynard Smith and Smith (1998), we calculated the probability of obtaining the observed number of haplotypes in the data set under a stepwise mutation model with no recombination. This approach is predicated on the concept that new haplotypes can be generated in the population, without the occurrence of new alleles, as the result of recurrent mutation. The distribution of data under this null hypothesis was found by Monte Carlo simulation with 1,000 replications. Starting with a common ancestor, allele and haplotype number evolve iteratively as follows. If there are k alleles, n loci, and m haplotypes at each iteration, (1) both

a new allele and a new haplotype are created with probability n/k or (2) only a new haplotype is created, with probability $(t - m)/t$, where $t = k!/[n!(k - n)!]$ is the number of possible haplotypes. This process is iterated until the number of alleles observed in the data of interest is reached, whereupon the number of haplotypes is recorded. We previously showed the stepwise mutation model to hold approximately for the loci under consideration (Marshall, Newton, and Ritland 2001).

A difficulty with the second two approaches to detecting recombination which involves their reliance on predicted levels of homoplasy must be noted. For example, with the third method, the probability that a mutational event will produce a new haplotype but not a new allele is as given above only when all possible mutational events occur with equal probability. This assumption is unrealistic biologically, and mutational bias may decrease the expected probability of a new allele and therefore increase the expected number of haplotypes. Similarly, the predicted level of homoplasy will not be accurate in the second method if the mutation process is biased.

Nonetheless, the results of all three analyses pointed to recombination. In the first analysis, the relationship between linkage disequilibrium (r^2) and map distance (fig. 1) was significantly negative ($b = -1.24$ per 10^6 bases with a 95% confidence interval of $-0.85, -1.45$; the distribution of randomized estimates were -0.20 to $+0.21$ with a mean of 0.002). In the second analysis, the observed tree length (233) greatly exceeded the expected tree length of 46 (under no recombination and no recurrent mutation) but was substantially less than would be predicted under complete linkage equilibrium ($P = 0$; fig. 2). In fact, when the loci were divided into two linkage groups based on physical proximity, the observed length fell within the predicted distribution of lengths (although barely; $P = 0.103$; fig. 2). In the third analysis, for the five simple sequence repeat loci, the observed number of haplotypes (143) exceeded the expected number for 32 alleles (average, 103) in 968 of 1,000 simulations, resulting in a low probability ($P = 0.032$) of complete linkage among loci.

For comparative purposes, we also evaluated a cpSSR data set from a monocot angiosperm, rice (Provan et al. 1996). These data consisted of 43 haplotypes, 20 alleles, and five loci and gave $P = 0.389$ in our third approach, consistent with complete linkage among loci. Thus, our finding of signatures of recombination in lodgepole may not represent a widespread phenomenon in plants; further investigation of other plant species is needed.

Any evidence for recombination must be reconciled with the apparent uniparental inheritance of the chloroplast. Uniparental inheritance of organellar DNA is a widespread phenomenon, with an array of diverse mechanistic and evolutionary explanations (Birky 1995). Mechanistically, organelles from one parent may be eliminated either (1) prezygotically via production of differentially sized gametes or by degradation of organellar DNA in the gamete, (2) during fertilization by exclusion from the zygote of the organelles of one par-

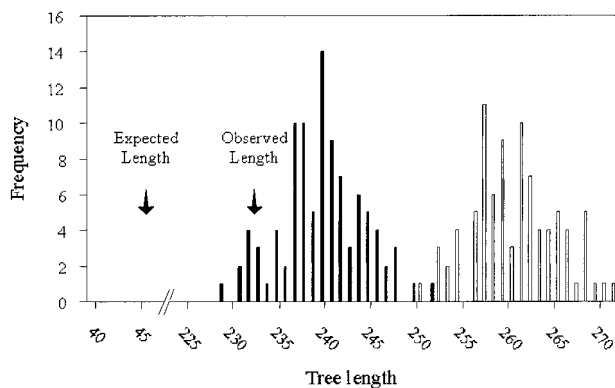


FIG. 2.—Frequency distribution of tree lengths generated from randomized data sets of lodgepole pine chloroplast sequence repeat haplotypes. Empty bars represent the distribution when the loci are treated as completely unlinked. The filled-bar distribution results from dividing the loci into two linkage groups, one containing three loci located in the small-single-copy region of the genome (spanning the region between 10 and 35 kb according to the genome coordinates of Wakasugi et al. 1994) and the other containing three loci located in the large-single-copy region (spanning the region between 60 and 90 kb). The arrows point to the expected tree length if the loci were completely linked (expected length) and the tree length obtained from the data (observed length).

ent, or (3) postzygotically by stochastic or deterministic exclusion of organelles from embryonic tissue (Birky 1995). As an example of the latter, in the fertilized egg of the gymnosperm *Larix*, embryonic cytoplasm is segregated into a region that contains paternal plastids but maternal mitochondria (Szmidt, Aldén, and Hällgren 1987). Evolutionary explanations hinge in part on the concept of the reduced importance of sexual reproduction to organellar genes because they are scarce relative to nuclear genes (Birky 1995).

Whatever the cause, Birky (1995) emphasized that strict uniparental inheritance of chloroplasts may not be as common as is generally believed. Wagner et al. (1987) noted unusual apparent recombinant cpDNA phenotypes in a zone of sympatry between lodgepole and jack pines and hypothesized that biparental inheritance occurs in this species. Furthermore, evidence of chloroplast recombination has previously been reported in species that normally exhibit maternal inheritance of cpDNA, such as *Nicotiana* (Medgyesy, Fejes, and Maliga 1985).

Here we found genetic signatures of recombination in lodgepole pine cpDNA. In light of previous reports of biparental inheritance and/or recombination in organellar DNA, further investigation on the prevalence of this phenomenon and its possible mechanism is needed. Furthermore, greater caution should be exercised about assumptions of complete linkage in phylogenetic and phylogeographic inferences from chloroplast DNA.

Acknowledgments

We thank the Natural Sciences and Engineering Research Council of Canada for postdoctoral support to H.D.M. and an operating grant to K.R. Rita Wagner and Carol Fleetham were responsible for the tree sampling related to this project.

LITERATURE CITED

- AWADALLA, P., A. EYRE-WALKER, and J. MAYNARD SMITH. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**:2524–2525.
- BIRKY, C. W. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**:11331–11338.
- BURT, A., D. A. CARTER, G. L. KOENIG, T. J. WHITE, and J. W. TAYLOR. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **93**:770–773.
- CLEGG, M. T. 1993. Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA* **90**:363–367.
- ESTOUP, A., C. TAILLIEZ, J.-M. CORNUET, and M. SOLIGNAC. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* **12**:1074–1084.
- EYRE-WALKER, A. 2000. Do mitochondria recombine in humans? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**:1573–1580.
- EYRE-WALKER, A., N. H. SMITH, and J. MAYNARD SMITH. 1999. How clonal are human mitochondria? *Proc. R. Soc. Lond. B Biol. Sci.* **266**:477–483.
- HILL, W. G., and A. ROBERTSON. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**:226–231.
- INGMAN, M., H. KAESSMANN, S. PÄÄBO, and U. GYLLENSTEN. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708–713.
- JARNE, P., and P. J. L. LAGODA. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **11**:424–429.
- LUNT, D. H., and B. C. HYMAN. 1997. Animal mitochondrial recombination. *Nature* **387**:247.
- MACDONALD, G. M., and L. C. CWCYNAR. 1991. Post-glacial population growth rates of *Pinus contorta* ssp. *latifolia* in western Canada. *J. Ecol.* **79**:417–429.
- MARSHALL, H. D., C. N. NEWTON, and K. RITLAND. 2001. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). *Theor. Appl. Genet.* (in press).
- MAYNARD SMITH, J., and N. H. SMITH. 1998. Detecting recombination from gene trees. *Mol. Biol. Evol.* **15**:590–599.
- MEDGYESY, P., E. FEJES, and P. MALIGA. 1985. Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. *Proc. Natl. Acad. Sci. USA* **82**:6960–6964.
- NEWTON, A. C., T. R. ALLNUT, A. C. M. GILLIES, A. J. LOWE, and R. A. ENNOS. 1999. Molecular phylogeography, intraspecific variation, and the conservation of tree species. *Trends Ecol. Evol.* **14**:140–145.
- OLMSTEAD, R. G., and J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.* **81**:1205–1224.
- 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.
- PROVAN, J., G. CORBETT, R. WAUGH, J. W. MCNICOL, M. MORGANTE, and W. POWELL. 1996. DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple sequence repeats. *Proc. R. Soc. Lond. B Biol. Sci.* **263**:1275–1281.
- SAVILLE, B. J., Y. KOHLI, and J. B. ANDERSON. 1998. MtDNA recombination in a natural population. *Proc. Natl. Acad. Sci. USA* **95**:1331–1335.
- SCHAAL, B. A., D. A. HAYWORTH, K. M. OLSEN, J. T. RAUSCHER, and W. A. SMITH. 1998. Phylogeographic studies in plants: problems and prospects. *Mol. Ecol.* **7**:465–474.
- STOEHR, M. U., and C. NEWTON. 2001. Evaluation of mating dynamics and pollen contamination in a lodgepole pine seed orchard using chloroplast DNA markers. *Can. J. For. Res.* (in press).
- SZMIDT, A. E., T. ALDÉN, and J.-E. HÄLLGREN. 1987. Paternal inheritance of chloroplast DNA in *Larix*. *Plant Mol. Biol.* **9**:59–64.
- THYAGARAJAN, B., R. A. PADUA, and C. CAMPBELL. 1996. Mammalian mitochondria possess homologous DNA recombination activity. *J. Biol. Chem.* **271**:27536–27543.
- VENDRAMIN, G. G., L. LELLI, P. ROSSI, and M. MORGANTE. 1996. A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Mol. Ecol.* **5**:595–598.
- WAGNER, D. B., G. R. FURNIER, M. A. SAGHAI-MAROOF, S. M. WILLIAMS, B. P. DANCIC, and R. W. ALLARD. 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, and M. SUGIURA. 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.

BRANDON GAUT, reviewing editor

Accepted July 20, 2001