Chloroplast microsatellite evolution in conifers: measures of genetic diversity and the effects of homoplasy

Miguel Navascués, Brent Emerson

Chloroplast microsatellite evolution in conifers

Miguel de Navascués and Brent C. Emerson

Centre for Ecology, Evolution and Conservation, School of Biological Sciences,
University of East Anglia, Norwich NR4 7TJ, UK m.navascues@uea.ac.uk

INTRODUCTION

The chloroplast genome is a potentially powerful tool for phylogenetic studies due to its uniparental inheritance, conserved gene order, nonrecombinant nature, availability of primers and lack of heteroplasmy. However, the levels of polymorphism found at the intraspecific level have been limiting its use in population genetics.

Thus, the discovery of polymorphic mononucleotide repeats (microsatellites) in the chloroplast genome produced great expectations for its application in studies of plant population structure, differentiation and diversity [1]. The identification of such chloroplast microsatellites (cpSSRs) was done through database searches (i.e. GenBank) resulting in sets of cpSSRs for most completely or partially sequenced chloroplast genomes -*Arabidopsis*, tobacco, pines, maize, soya... [2] (e.g. Fig. 1, for cpSSRs in pines).

The ease and availability of cpSSRs has seen them used in an increasing number of scientific studies to date (Fig. 2). However, the potential levels of homoplasy that could be present in cpSSRs [2] have to be considered and could limit the confidence in some of the results. In order to study the effect of homoplasy on the analysis of population genetics of conifers with cpSSRs we have performed a series of simulation studies.

THE MODEL

The simulation model used follows a coalescent approach similar to Hudson [5]. The first step of the model is building a genealogy of a sample of *n* chloroplast genomes. Coalescent events are simulated as a function of population size, population structure and mating system (considering cpSSRs selectively neutral). Once the genealogy is completed the genetic state of the root is set. Then, random mutations are placed along the branches following a mutational model.

The conditions of the simulation were determined to match the evolution of conifers as much as possible in order to compare the genetic pattern of natural populations with the simulated populations. The stepwise mutation model (SMM) was chosen over other models for being the simplest plausible model for microsatellites (the two-phase model is probably a better fit, but requires more parameters unknown *a priori*)[6].

The output of the simulation allows the reconstruction of the complete evolutionary history of the simulated chloroplast genomes (Fig. 3). Hence, it is a simple task to score the amount of recurrent and back mutations and estimate the levels of homoplasy that are produced.

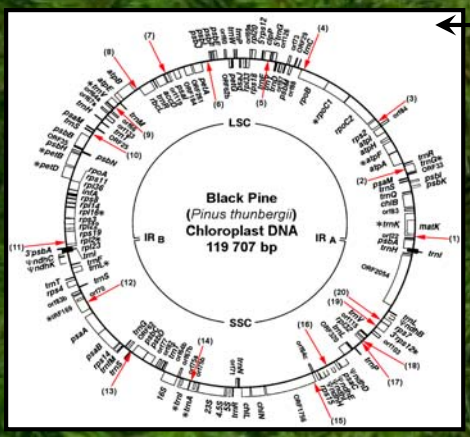
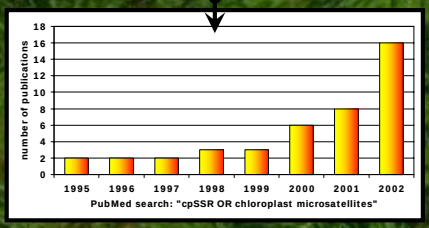


Figure 1.- Location of 20 chloroplast microsatellites described for *Pinus* [3] (modified from [4])

Figure 2.- The use of chloroplast microsatellites in scientific literature since their discovery.



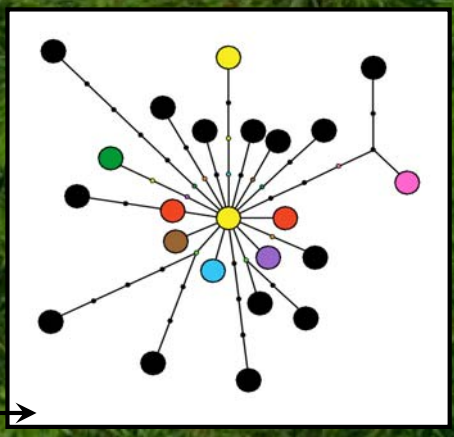
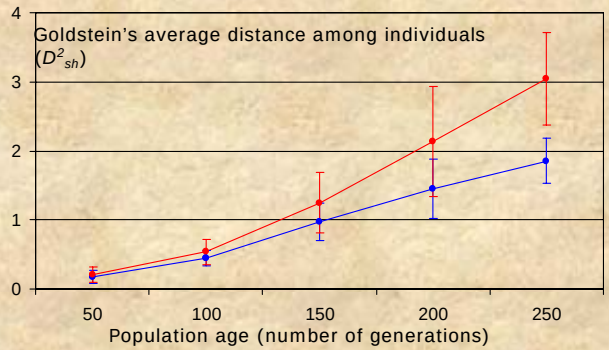
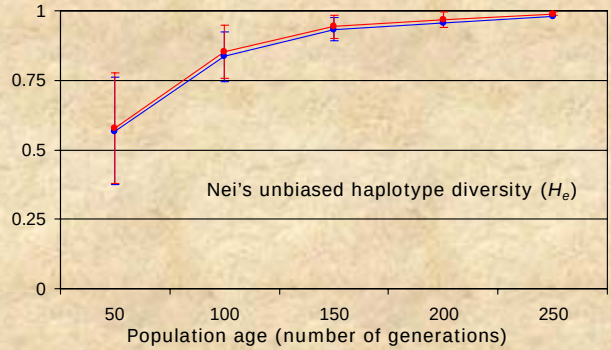


Figure 3.- Network of the simulated evolution of cpSSRs. Black nodes represent haplotypes that appeared only once through the evolution, coloured nodes represent haplotypes whose state was produced more than once during evolution. Big circles represent haplotypes present only in the last generation

EFFECT OF HOMOPLASY ON MEASURES OF GENETIC DIVERSITY

Homoplasy, due to recurrent mutation and back mutations, has the potential to reduce the genetic diversity expected for a determined mutation rate. In order to study this the genetic diversity of the simulated populations was analysed. First, the genetic diversity was studied ignoring the known genetic history and considering only the genetic structure of the last generation (blue lines in the graphs). This represents an empirical study of a natural population. Then, the genetic history (unknown in empirical studies) was considered in the analysis (red lines). This allows us separate haplotypes identical in state but with different origins. The results indicate that homoplasy has little influence in Nei's H_e, while Goldstein's D^2_{sh} significantly underestimates diversity.





BIBLIOGRAPHY

[1] Powell, W. *et al.* (1995) *PNAS* **92**:7759-7763.
[2] Provan, J. *et al.* (2001) *TREE* **16**:142-147.
[3] Vendramin, G. G. *et al.* (1996) *Mol. Ecol.* **5**:595-598.
[4] Wakasugi, T. *et al.* (1994) *PNAS* **91**:9794-9798.
[5] Hudson, R. R. (1990) *Oxf. Surv. Evol. Biol.* **7**:1-44.
[6] Di Rienzo, A. *et al.* (1994) *PNAS* **91**:3166-3170.