Chloroplast microsatellites: new tools for studies in plant ecology and systematics


Published in:
Trends in Ecology & Evolution

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Chloroplast microsatellites: new tools for studies in plant ecology and evolution

Jim Provan, Wayne Powell and Peter M. Hollingsworth

The nonrecombinant, uniparentally inherited nature of organelle genomes makes them useful tools for evolutionary studies. However, in plants, detecting useful polymorphism at the population level is often difficult because of the low level of substitutions in the chloroplast genome, and because of the slow substitution rates and intramolecular recombination of mtDNA. Chloroplast microsatellites represent potentially useful markers to circumvent this problem and, to date, studies have demonstrated high levels of intraspecific variability. Here, we discuss the use of these markers in ecological and evolutionary studies of plants, as well as highlighting some of the potential problems associated with such use.

Over the past 25 years, powerful insights have been gained into the ecological, historical and evolutionary dimensions of plant populations from the analysis of organelle DNA (Ref. 1). Organelle genomes are typically nonrecombinant, unparentally inherited and effectively haploid. In animals, the small size, high copy number, relatively conserved gene order, easy availability of primers and rapid substitution rates associated with mtDNA have led to its extensive use in a variety of studies. These studies have provided insights into many aspects of evolutionary population biology, including patterns of migration, population histories and levels of differentiation.

In plants, however, the vastly different mitochondrial genome, characterized by a large size, slow nucleotide substitution rates and extensive levels of intramolecular recombination, has been of limited use in similar studies. The chloroplast genome, however, shares many features with animal mtDNA and the two have been referred to as ‘natural counterparts’3. Its conserved gene order, the widespread availability of primers and a general lack of heteroplasmy and recombination, have made the chloroplast genome an attractive tool for phylogenetic studies of plants. Furthermore, its uniparental mode of inheritance (usually maternal in angiosperms and paternal in gymnosperms) makes it possible to elucidate the relative contributions of seed and pollen flow to the genetic structure of natural populations by comparing nuclear and chloroplast markers4. Recognizing this potential, various authors developed and reviewed explicit population genetic models for organelle markers in the early–mid-1990s4 5. This work provided the theoretical framework from which novel, empirical insights could be gained into the history and dynamics of plant populations. The low mutation rates associated with the chloroplast genome, however, meant that detecting sufficient variation represented a major technical barrier for the widespread application of this theory.

Two recent developments have alleviated this problem. First, a series of ‘universal primers’ have been developed, which allow the amplification of chloroplast introns and intergenic spacers in a wide range of higher plant taxa6–8. Restriction digests of these PCR-amplified regions represent a time- and cost-efficient strategy for searching for intraspecific chloroplast polymorphism. It requires only minimal amounts of plant tissue and avoids the laborious experimental procedures associated with conventional Southern hybridization-based restriction fragment length polymorphisms (RFLP).

The approach has been used successfully to detect (often extensive) intraspecific polymorphism in several species and to provide valuable information, particularly with respect to plant migrations in relation to glacial cycles10. In other species, however, few or no polymorphisms have been detected, highlighting the need for additional assays.

The second significant development has been the discovery of polymorphic mononucleotide repeats in the chloroplast genome and it is these loci that are the subject of this review. Although polymorphic microsatellites have revolutionized plant and animal

Table 1. Occurrence of mononucleotide repeats in chloroplast genomes

<table>
<thead>
<tr>
<th>Species</th>
<th>EMBL accession no.</th>
<th>No. of mononucleotide repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>AP000423</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Epilagus virginiensis</td>
<td>M81884</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Marchantia polymorpha</td>
<td>X04465</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>Z00044</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Oenothera elata</td>
<td>AJ 271079</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>X15901</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Pinus thunbergii</td>
<td>D17510</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Spinacia oleracea</td>
<td>AJ 400848</td>
<td>≥8(n) ≥8</td>
</tr>
<tr>
<td>Zea mays</td>
<td>X86563</td>
<td>≥8(n) ≥8</td>
</tr>
</tbody>
</table>

Abbreviation: EMBL, European Molecular Biology Laboratory.

*Number of repeated nucleotides ≥10.

**Number of repeated nucleotides ≥8.
Box 1. Scoring chloroplast microsatellite gels

To draw meaningful inferences from raw data, the data must be accurately recorded. Scoring of microsatellite gels can be difficult and careful experimental procedures need to be employed. With such gels, the data generally consist of the alleles themselves, and a series of shadow bands, attributable to slippage in the PCR amplification (e.g. Fig. I). To test for consistency of amplification and for interpretation, ‘standards’ representing the range of allele sizes at a given locus should be included in all PCRs and amplifications should be routinely repeated. The short length and limited number of alleles of most chloroplast microsatellites means that confirmation of allele sizes can also be obtained by sequencing (see Fig. I). This is more difficult for many nuclear microsatellites, which are often highly multiallelic (and heterozygous) and from whose long repeat arrays, clean sequences can be exceedingly difficult to obtain.

Fig. 1. An example of chloroplast microsatellite polymorphism in six individuals of the leguminous tree Caesalpinia echinata. From the sequences, it can be seen that allele size is exactly correlated with expansion and/or contraction of the mononucleotide repeat (green), ranging from (T7)1 in individuals 1_1 to 1_3 to (T9)3 in individuals 2_1 and 2_3.

Table 2. Cross-species amplification of polymorphic chloroplast microsatellites exhibiting intraspecific variation

<table>
<thead>
<tr>
<th>Source of primers</th>
<th>Species exhibiting intraspecific polymorphism*</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine soja (Leguminosae)</td>
<td>Glycine spp</td>
<td>14,17,18</td>
</tr>
<tr>
<td>Nicotiana tabacum (Solanaceae)</td>
<td>Solanum spp</td>
<td>19,20</td>
</tr>
<tr>
<td>Oryza sativa (Poaceae)</td>
<td>Oryza spp</td>
<td>21,22</td>
</tr>
<tr>
<td>Pinus thunbergii (Pinaceae)</td>
<td>Pinus spp</td>
<td>13,23-28</td>
</tr>
<tr>
<td>Abies alba</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Picea abies</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

*Weising and Gardener9 presented a set of ‘Conserved Primers for the Analysis of cpSSR Polymorphism in Dicotyledonous Angiosperms’ based around Nicotiana tabacum sequences that amplify across a broad range of plant families (Solanaceae, Actinidaeae, Brassicaceae, Fabaceae, Myrtaceae, Rosaceae, Poaceae and Agavaceae). However, little or no intraspecific polymorphism was described and the size ranges of many of the products differ far beyond that attributable to SSR repeats alone. Similarly, Cato and Richardson28 showed that primers designed from Pinus thunbergii amplified products in Pinaceae, Cupressaceae, Podocarpaceae, Anacardiaceae, Labiatae, Oleaceae, Actinidaeae and Brassicaceae. Further studies are required to establish how many of these products contain SSRs and how many show intraspecific polymorphism.

Fig. I. An example of chloroplast microsatellite polymorphism in six individuals of the leguminous tree Caesalpinia echinata. From the sequences, it can be seen that allele size is exactly correlated with expansion and/or contraction of the mononucleotide repeat (green), ranging from (T7)1 in individuals 1_1 to 1_3 to (T9)3 in individuals 2_1 and 2_3.

ecology and evolutionary biology11,12, until recently, their exploitation had been limited to the nuclear genome. Powell and co-workers, however, reported extensive levels of length polymorphism at chloroplast microsatellite loci in species of Pinus13 and Glycine14, and subsequent studies have revealed comparable levels of polymorphism in other species15. These chloroplast microsatellites (also known as simple sequence repeats: SSRs) have been found in all completely sequenced plant chloroplast genomes to date, as well as in hundreds of partial chloroplast sequences (Table I). PCR amplification of these SSRs, using primers specific to the 5’- and 3’-flanking regions, produces products that exhibit length variation corresponding to the expansion and/or contraction of the repeat region (Box 1 and Fig. I).

Availability of chloroplast microsatellites

Unlike the conventional approach to obtaining nuclear microsatellites, the nature of chloroplast microsatellites (generally <15 mononucleotide repeats), and the need for genome specificity means that it is not possible to screen libraries for these motifs. Instead, the availability of sequence data for the species of interest or for closely related species is the major factor in obtaining microsatellite markers in the chloroplast genome. For those species (and their close relatives) that have had their chloroplast genomes completely sequenced, identifying mononucleotide repeats can be achieved by simple database searches. Primers can then be designed and screened for polymorphism.

When chloroplast microsatellite primers have been designed for one species, they regularly cross-amplify in related species (Table 2). Finding true ‘universal’ primers that show widespread intraspecific polymorphism remains difficult. Although there are primers available that amplify across a broad range of taxa, in published studies to date, these have rarely shown widespread intraspecific microsatellite polymorphism16 (Table 2). This is attributable to the contrasting requirements in a short stretch of DNA of extreme stability and sequence conservation for priming sites and consistently high levels of polymorphism in the intervening region31. Even for taxa without complete chloroplast genome sequences, however, the large amount of effort currently being directed at the generation of sequence matrices for phylogenetic studies has resulted in the widespread availability (via GenBank: http://www.ncbi.nlm.nih.gov) of sequence data for at least a few chloroplast loci (e.g. the trnL intron and surrounding spacers3). Table 3 provides a summary to help target regions to sequence in the absence of any other data.

http://tree.trends.com
Levels of polymorphism

To date, studies of chloroplast microsatellites have revealed much higher levels of diversity than those of chloroplast RFLPs. An investigation of a range of rice species (Oryza), which had previously been examined using chloroplast RFLPs, revealed a fourfold increase in levels of diversity when the same species were studied using chloroplast microsatellites. At the intraspecific level, a study on barley (Hordeum vulgare) revealed six chloroplast microsatellite haplotypes among 12 H. vulgare ssp. spontaneum accessions from Israel, whereas a previous RFLP study on 245 samples generated only three haplotypes. Similarly, the analysis of natural populations of Scots pine (Pinus sylvestris) using chloroplast microsatellites resulted in within-population diversity values of >0.95 and ~40% of the 330 trees studied could be individually genotyped, therefore highlighting the power of the assay.

Applications of chloroplast microsatellites

Population genetics

The development of molecular genetic markers has had a great impact on our understanding of the processes that determine structure and variation within and among natural populations. Because organelle genomes are haploid, their effective population size in hermaphrodite outcrossing plants is half that of diploid nuclear genomes and, as a result, chloroplast-specific markers should be good indicators of historical bottlenecks, founder effects and genetic drift. The use of chloroplast microsatellites has allowed the examination of these events at a finer level of detail than was previously possible.

Echt et al. examined the genetic structure of red pine (Pinus resinosa), an endemic occurring in disjunct populations in eastern North America. Previous studies of this species detected no allozyme diversity and only limited RAPD diversity. Using nine chloroplast microsatellite loci on 159 individuals from seven widely separated populations, a total of 23 haplotypes were recovered, with the most common haplotype present in all populations. Most other haplotypes differed by a single mononucleotide, with the largest pairwise difference involving seven steps. The presence of the widespread common haplotype was invoked as being indicative of a major ancient population bottleneck. Chloroplast microsatellites have also revealed patterns of genetic diversity consistent with population bottlenecks and founder effects in Pinus halterensis, as well as in the extreme case of Torrey pine (P. torreyana), where no variation was detected in the chloroplast genome, thus confirming previous RFLP studies.

---

Table 3. Identifying chloroplast microsatellite regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Monocots</th>
<th>Dicots</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Arabidopsis</td>
<td>Nicotiana</td>
</tr>
<tr>
<td>trnT/trnL</td>
<td>144–145</td>
<td>2 (1/1)</td>
</tr>
<tr>
<td>trnL/trnF</td>
<td>374–375</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnH/trnK</td>
<td>1932–1933</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnK intron</td>
<td>2546</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnL/trnF</td>
<td>374–375</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnD/trnT</td>
<td>171–172</td>
<td>2 (1/1)</td>
</tr>
<tr>
<td>trnS/trnfM</td>
<td>1374–1375</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnF/trnV</td>
<td>2938–2939</td>
<td>5 (2/2)</td>
</tr>
<tr>
<td>trnC/trnD</td>
<td>1814</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnL intron</td>
<td>614–615</td>
<td>1 (1/1)</td>
</tr>
<tr>
<td>trnQ/trnR</td>
<td>3046–3047</td>
<td>6 (2/1)</td>
</tr>
</tbody>
</table>
| chloroplast microsatellites for the species of interest. Many of these introns and spacers contain mononucleotide repeats. This information could be used to target potentially worthwhile regions to sequence, to identify chloroplast microsatellites for the species of interest. 

---

Despite difficulties in designing ‘universal’ chloroplast microsatellite primers that show widespread intraspecific chloroplast polymorphism, several sets of universal primer pairs are available to amplify chloroplast introns or intergenic spacers. A survey of the published complete chloroplast genome sequences of higher plants reveals that many of these introns and spacers contain mononucleotide repeats. This information could be used to target potentially worthwhile regions to sequence, to identify chloroplast microsatellites for the species of interest.

The figures in the table give the number of microsatellites of eight or more mononucleotide repeats. For each taxon, the first column signifies the length in base pairs of the region and the second column gives the number of repeats in the region. Figures in parentheses show the number of repeats within 500 bases of each primer for sequencing purposes.

Because of a rearrangement, these primers amplify a region including that amplified by the psbC/trnS primers.

Because of an inversion, these genes are separated by at least 30 kb.

Because of a rearrangement in the Oenothera chloroplast genome, these two genes are separated by ~50 kb.

'–' indicates that there is no microsatellite present.

The region amplified by the trnF/trnV primer pair largely corresponds to the region amplified by the trnM/rbcL primers.
Box 2. Potential limitations of chloroplast microsatellites

Hypervariability of markers is not always a good thing\(^a\). The level of polymorphism needs to be appropriate for the question being addressed, and the distinction between precision and accuracy should always be considered. For studies estimating the partitioning of genetic diversity, sample sizes should be sufficient to represent the variability present; this often requires large numbers of individuals or loci. In addition, if there are high levels of within population variability, some distance measures can give low estimates of population differentiation, even among populations that have no alleles in common\(^b\).

Another area of uncertainty relates to homoplasmy and the recurrent generation of alleles of identical length, a consequence of which is the prediction that alleles of individual lengths might have identity in state without identity by descent\(^c\). High levels of homoplasmy can confound estimates of population differentiation and recurrent generation of alleles can mimic gene flow. If there are size constraints on allele lengths\(^d\), the problem is compounded because the number of possible allelic states becomes decidedly finite. Given that all allelic states and transfer between all allelic states might not be equally probable, and that there is usually greater the divergence between the study samples, the more likely this is to represent a problem, and difficulties might be especially acute for broad-scale phylogeographical studies and interspecific comparisons. At these levels, if sufficient PCR–RFLP variation owing to nucleotide substitutions can be detected, this type of polymorphism might be preferable to that from mononucleotide repeats.

References


With interest developing in direct measurements and spatially explicit models of gene flow, the assignment of parentage to seeds and plants is becoming increasingly important. There is considerable potential for hypervariable chloroplast microsatellites to provide paternal or maternal markers for such direct measures of seed or pollen flow. Their multiple nonrecombiant loci give profiles that exactly match those of the parent, rather than the gametic phase having to be inferred from multiple heterozygous loci, as would be the case for nuclear microsatellites. Although chloroplast microsatellites provide data on only one parent and, on their own, might not be variable enough to obtain high exclusion probabilities for parentage assignment, they might represent a useful adjunct to studies using nuclear microsatellites. Ziegenhagen et al.\(^b\) used two chloroplast microsatellite loci, each with five alleles, to investigate the paternity of 24 naturally regenerated young silver fir trees (Abies alba) in the direct neighbourhood of two relatively isolated adult trees (chloroplasts are paternally inherited in Pinaceae). Six of the 24 young plants had chloroplast microsatellite haplotypes that did not match the two adult trees, therefore indicating pollen flow into the stand. Although complete paternity analysis was not the object of the study, and the sampling of more individuals and loci might well have given greater resolution, the study demonstrates the potential use of the approach.

Similarly, a study on sympatric populations of the closely related species P. halepensis and P. brutia\(^b\) used chloroplast microsatellites to examine whether individual seeds were of hybrid origin. Species-specific haplotypes were identified and 12.5% of the embryos from P. brutia trees displayed a P. halepensis haplotype, thus indicating that cross-pollination from P. halepensis had taken place. Interestingly, no reciprocal hybrids were detected, which was consistent with results from artificial crosses that were only successful when P. halepensis was the pollen donor. These direct measures of gene flow have been carried out in conifers where paternal plastid inheritance is the rule. Studies of angiosperms, where chloroplast DNA is predominantly maternally inherited, might offer further insights, and might provide information on the patterns and extent of localized seed dispersal.

Understanding crop plant evolution and domestication

Molecular techniques have been used widely to study various aspects of crop-plant evolution. Nuclear and chloroplast markers have helped shed new light on topics such as the identification of the progenitors of cultivated species, the amounts and patterns of diversity in crop plants and their wild relatives, and the role of introgression in crop evolution\(^b\). In particular, chloroplast DNA has been used in many such studies\(^b\), and chloroplast microsatellites have a potential advantage over RFLPs in terms of ease of analysis and increased resolving power.

It has been well documented that the processes of domestication and selection involved in the cultivation of wild plants have led to a decrease in the variation available in the gene pools of many crop species\(^c\). A study of 51 accessions of the wild progenitor (Hordeum vulgare ssp. spontaneum) of cultivated barley (H. vulgare ssp. vulgare) revealed a total of 12 haplotypes. By contrast, only a single haplotype was observed in 101 accessions of H. vulgare ssp. vulgare\(^c\). This loss of chloroplast diversity in the cultivated plants is matched by a parallel loss of nuclear diversity, as measured by nuclear microsatellites\(^c\).

Studies using nuclear and cytoplasmic microsatellite markers in potato (Solanum tuberosum ssp. tuberosum) revealed somewhat different patterns of diversity\(^c\). Previously, it had been documented that most European potato cultivars were characterized by a specific type of cytoplasm (‘T-type’) found at low frequency in S. tuberosum ssp. andigena, the wild progenitor of S. tuberosum ssp.
Under various models of population structure (most simply an idealized neutral Wright–Fisher model), the shape of the genealogical tree of molecular haplotypes can be used to make inferences of the demographic history of a sample. If estimates of mutation rates are available, this can include an approximation of the time to the most recent common ancestor (MRCA). Using these coalescent approaches on a range of data sources, exciting insights have been gained into the history of the human population. An interesting parallel can be drawn between chloroplast microsatellites and studies of microsatellite variation from the human Y chromosome. Similar to chloroplast DNA, nonrecombinant regions of Y chromosomes contain microsatellite motifs and are uniparentally inherited. Recent studies using coalescent-based approaches of human Y chromosome microsatellites suggest a surprisingly recent expansion from a MRCA within the last 30 000–40 000 years (albeit with uncertainty up to ~120 000 years).

The application of similar analytical approaches to chloroplast microsatellites could be explored. Of course, to be informative about natural populations, coalescent models need biological realism. Most plant populations do not exist as idealized populations; they are often geographically subdivided and not at equilibrium, and they might experience heterogeneous levels and types of selection and have population sizes that fluctuate over time. The mutational processes at chloroplast microsatellites are not well understood and their adherence to the stepwise mutation model requires testing. For dating the time to the MRCA of a sample, solid estimates of mutation rates are required. In addition, if allele sizes are tightly constrained, but mutation rates high, genealogies might only poorly represent the ‘true’ relationships among haplotypes. The incorporation into stochastic models of information on, and assessing the sensitivity of inference to, these multiple and as yet often unknown variables, represents an ongoing challenge to the implementation of this methodology.

References


Future prospects

There is a premium on learning more about rates of mutational change in chloroplast microsatellites. As divergence times between study samples increase, so does the possibility of homoplasy. The extent to which this is a problem, however, and at which taxonomic level, is still unclear. In this respect, comparisons of chloroplast microsatellites with RFLP markers located in the same region would be informative. Because the chloroplast genome is nonrecombinant, there is an expectation for variability in one marker to be correlated with that of another. Tests for the association of allelic variants can then be made to assess the extent of recurrent origins of alleles – where homoplasy is high, linkage disequilibrium should be low.

Examining mutation rates in known pedigrees, and in in vivo experiments of allelic stability in cell
lines should also give some approximation of the potential rates of change within individual species. In addition, there are many natural experiments available for assessing mutation rates of chloroplast microsatellites. Ancient organisms, such as quaking aspen (Populus tremuloides) clones estimated to be $\pm 1$ million years old$^{30}$ might contain significant intragenet chloroplast microsatellite polymorphism. Less ancient, but easier to obtain reliable age estimates from, are long-lived individual trees, and a search for mutations between apical twigs stemming from early branching dateable dichotomies could be an informative exercise.

From a data analytical perspective, the potential exists for the application of coalescent-based models to chloroplast microsatellite data (Box 3). Coalescent approaches can be extremely useful in assessing a range of demographic histories$^{38}$ but their application to intraspecific studies in plants has been hampered by the slow mutation rate of the nonrecombiant genomes, such as chloroplast DNA and, until recently, a general paucity of sequence data from the nuclear genome$^{36}$. Although not without their limitations (Box 2), chloroplast microsatellites represent a potentially informative data source with which coalescent-based approaches can be explored. To date, most of the published work on chloroplast microsatellites concerns crop species or members of the Pinaceae. Further studies from other plant taxa would be useful. In particular, the complete chloroplast sequences of algae and bryophytes represent a potentially useful resource for the analysis of plastid polymorphism among populations of cryptogamic species. In addition, the search for microsatellites in the mitochondrial genome might also be informative, although preliminary studies on plant mitochondrial microsatellites have shown little intraspecific variability$^{39}$.

**Acknowledgements**

We thank Jane Squirrell and Richard Ennos for many helpful discussions on chloroplast microsatellites and Deborah Charlesworth for advice on coalescent theory. We are also very grateful to three anonymous referees for their constructive comments on this article. PMH’s work on chloroplast microsatellites has been supported by NERC grant GST/02/833.

**References**