Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations

RÉMY J. PETIT,* JÉRÔME DUMINIL ,* SILVIA FINESCHI ,† ARNDT HAMPE ,‡ DANIELA SALVINI †§ and GIOVANNI G. VENDRAMIN¶

*INRA, UMR Biodiversity, Genes and Ecosystems, 69 Route d'Arcachon, F-33612 Cestas Cedex, France, †Institute for Plant Protection, Consiglio Nazionale delle Ricerche, via Madonna del Piano, Edificio E, I-50019 Sesto Fiorentino, Italy, ‡Integrative Ecology Group, Estación Biológica de Doñana (CSIC), Apartado 1056, E-41080 Sevilla, Spain, §Present address: Danish Centre for Forest, Landscape and Planning, Royal Veterinary and Agricultural University, Hørsholm Kongevej 11, 2970 Hørsholm, Denmark, ¶Plant Genetics Institute, Florence division, Consiglio Nazionale delle Ricerche, via Madonna del Piano, Edificio D, I-50019 Sesto Fiorentino, Italy

Abstract

Plants offer excellent models to investigate how gene flow shapes the organization of genetic diversity. Their three genomes can have different modes of transmission and will hence experience varying levels of gene flow. We have compiled studies of genetic structure based on chloroplast DNA (cpDNA), mitochondrial DNA (mtDNA) and nuclear markers in seed plants. Based on a data set of 183 species belonging to 103 genera and 52 families, we show that the precision of estimates of genetic differentiation (G_{ST}) used to infer gene flow is mostly constrained by the sampling of populations. Mode of inheritance appears to have a major effect on G_{ST}. Maternally inherited genomes experience considerably more subdivision (median value of 0.67) than paternally or biparentally inherited genomes (~0.10). G_{ST} at cpDNA and mtDNA markers covary narrowly when both genomes are maternally inherited, whereas G_{ST} at paternally and biparentally inherited markers also covary positively but more loosely and G_{ST} at maternally inherited markers are largely independent of values based on nuclear markers. A model-based gross estimate suggests that, at the rangewide scale, historical levels of pollen flow are generally at least an order of magnitude larger than levels of seed flow (median of the pollen-to-seed migration ratio: 17) and that pollen and seed gene flow vary independently across species. Finally, we show that measures of subdivision that take into account the degree of similarity between haplotypes (N_{ST} or R_{ST}) make better use of the information inherent in haplotype data than standard measures based on allele frequencies only.

Keywords: cpDNA, geographical structure, microsatellites, mtDNA, phylogeography, variation

Introduction

Gene flow shapes the organization of genetic diversity within and among populations (e.g., Wright 1931). In contrast to many vertebrates, seed plants have intrinsically complex and asymmetrical dispersal behaviours. Because adults are fixed, the dispersal function is mediated by two very distinct vehicles that operate in sequence, the male gametophyte (pollen) and the young sporophyte (seed). Another peculiarity of plants is that two organelle genomes coexist in the cytoplasm of their cells: the mitochondrial genome, nearly ubiquitous in eukaryotes, and the plastid genome, specific to plants. The mode of inheritance of these two genomes is varied and not always coincident, ranging from strictly maternal to strictly paternal (Harris & Ingram 1991; Reboud & Zeyl 1994; Mogensen 1996; Röhr et al. 1998). The contrasted patterns of inheritance of organelle and nuclear genes can be used to unravel the complexity of gene flow in plants, as they are predicted to result in very different distribution of genetic diversity within and among populations (Birky et al. 1989; Petit et al. 1993). Therefore, inferences can be made as to the relative importance of the two main components of

Correspondence: Rémy J. Petit, Fax: + 33557122881; E-mail: petit@pierroton.inra.fr

dispersal, seed and pollen, thanks to the development of theoretical models that relate level of gene flow and population genetic structure (e.g. Petit 1992; Ennos 1994; Hu & Ennos 1997; Oddou-Muratorio et al. 2001; Hamilton & Miller 2002). However, these models are based on assumptions that are unlikely to be met in natural systems; hence interpretations have to be made carefully.

Relying on a large body of case studies should help evaluate to which extent these models are useful. Over more than two decades, research on plant organelle DNA (oDNA) diversity has been lagging behind that on animal mtDNA (Soltis et al. 1997; Schaal et al. 1998; Brunsfeld et al. 2001; Petit & Vendramin 2005). The situation has changed during the last 10 years, as a result of the emergence of efficient molecular techniques to identify and screen oDNA diversity. There is now sufficient data available to examine the realism of several predictions made by theoretical models and to provide some reference values for comparative purposes. A general introduction on oDNA variation in plants has been published recently (Petit & Vendramin 2005) but there has been no comprehensive review so far of the primary data generated during population surveys of oDNA variation in plants, despite some early or limited attempts (e.g. Ennos 1994; McCauley 1995; El Mousadik & Petit 1996; Ouborg et al. 1999; Morjan & Rieseberg 2004).

Here we compile studies that report the partitioning of genetic diversity within and among populations for organelle genes in seed plants and, whenever possible, identify estimates from the same species based on nuclear markers. We first examine the precision of estimates of population subdivision and test if it is more limited by the sampling of populations than by the sampling of individuals within populations, as suggested earlier (Pons & Petit 1995). The findings should be useful for planning future studies of organelle diversity. Furthermore, we ask whether biases are introduced into G_{ST} estimates by the genotyping method or by the examination of only a restricted part of the range. We then estimate mean levels of differentiation at markers having different modes of inheritance (maternal, paternal or biparental) and examine how these measures of subdivision covary across species. By relying on previously developed (neutral) equilibrium expectations, one should indeed obtain rough indirect measures of the relative importance of gene flow through pollen vs. seed and hence obtain some insight on how gene flow is generally achieved in plants. For instance, does gene flow take place predominantly through pollen, as often assumed (e.g. Levin & Kerster 1974; Ellstrand 1992)? Does some type of compensation exist between the two components of gene flow, with plants relying little on seed gene flow subjected to correspondingly higher level of pollen gene flow and vice versa? Because a minimum rate of gene flow is likely to be necessary for species cohesion and survival, some compensation might exist, particularly in species with very low dispersal rates. On the other hand, pollen and seed flow do not play an equivalent ecological role and other processes could overwhelm any compensatory effects between these two components of dispersal. Therefore, empirical data is needed to answer this quite fundamental question, which has received little if any attention so far. Finally, we test if the presence of a phylogeographical structure at organelle genes is a general feature in seed plants, comparing systematically the partitioning of organelle DNA diversity among populations when similarities between haplotypes are taken into account (N_{ST} or R_{ST}) or when they are not (G_{ST}). No clear consensus exists regarding the usefulness of taking allele similarity into account when empirically measuring population genetic structure (e.g. Gaggiotti et al. 1999; Balloux & Lugon-Moulin 2002), so this question seems relevant.

Materials and methods

Data set

The primary literature was searched until April 2004 for population studies of seed plant oDNA diversity, using bibliographical databases, checking references in published papers and contacting colleagues. Only studies with ≥ 5 populations and having ≥ 2 individuals per population analysed were considered. We sought to obtain the raw frequencies of haplotypes in populations and their full molecular characterization by examining published information or directly soliciting it from the authors. When several surveys were available for a single species, we included only those based on the highest number of populations and covering the largest fraction of the range, to avoid attributing too much weight to a few well-studied taxa. When a suitable oDNA entry was detected, we further attempted to identify a corresponding survey based on nuclear markers, by screening the literature for results from that particular species. Here as well, priority was given to range-wide studies.

The following information was recorded for each species: taxonomic group (gymnosperm or angiosperm), family, genome investigated (cpDNA, mtDNA or nuclear), its known or presumed predominant mode of inheritance (paternal, maternal or biparental), molecular technique used [the most frequent were: Probe-RFLP (restriction fragment length polymorphism, where the fragments are identified using a marked DNA fragment, the probe), Purif-RFLP (RFLP on DNA purified after isolation of organelles), PCR-RFLP (PCR amplification with specific primers followed by RFLP), SSR (single sequence repeats, also called microsatellites), sequencing, AFLP (amplified fragment length polymorphism), RAPD (random amplified polymorphic DNA) and isozymes], proportion of the range

sampled (local, regional — that is, less than half of the range covered, or species-wide), number of populations investigated, total number of individuals sampled, arithmetic and harmonic mean number of individuals per population, number of haplotypes detected (for oDNA studies), number of 'characters' studied (number of polymorphic bands, variable nucleotide sites or SSRs for oDNA, number of loci for nuclear markers), and diversity statistics. For nuclear data, these included mean expected heterozygosity H_E, F_{IS} and F_{ST}, G_{ST} or Φ_{ST}, as provided in the papers. Although it would be preferable to use identical measures of genetic subdivision, F_{ST}, G_{ST} or Φ_{ST} estimate the same quantity and the differences between them is generally quite low (Morjan & Rieseberg 2004). For oDNA data, the diversity statistics were computed by us on the basis of the raw data as described below. In the few cases where this proved impossible, the estimates provided in the paper were used instead.

Data analyses

For oDNA, the following parameters were included in the database: total diversity H_T, G_{ST} and its standard error, N_{ST} and R_{ST} and their standard errors (whenever appropriate: for N_{ST}, when there were ≥ 3 studied characters and when the description of haplotypes was available; for R_{ST}, when details on length variation at microsatellite loci were available). The program permut (by RJP, available at http://www.pierroton.inra.fr/genetics/labo/Software/) was used to estimate the parameters of population subdivision G_{ST} and N_{ST} and their standard errors. The difference between these two parameters is that N_{ST} takes similarities between haplotypes into account, contrary to G_{ST} (Pons & Petit 1996). These similarities were measured by counting the number of characters that differ between all pairs of haplotypes. The difference between N_{ST} and G_{ST} was tested through 1000 random permutations of haplotype identity (Burban et al. 1999). In the case of chloroplast microsatellites (cpSSRs), we used the program cpssr (available at http://www.pierroton.inra.fr/genetics/labo/Software/) to derive R_{ST}. This program is essentially identical to permut except that the distance between haplotypes is adapted to the mode of evolution of SSRs (it is the sum across loci of the squared differences in number of repeats). Note that the estimators of these parameters, proposed by Pons & Petit (1995, 1996), are based on a random model of population variation, instead of the fixed model procedure assumed by Nei & Chesser (1983). Hence, the variation resulting from the sampling of populations is taken into consideration, allowing comparisons across species. On the other hand, all populations are given the same weight, regardless of their sample sizes, as in Nei & Chesser (1983), sample sizes are identical in all populations, the estimate of G_{ST} proposed by Pons & Petit (1995) becomes identical to the parameter θ of Weir & Cockerham (1984).

The pollen-to-seed migration ratio ($r = m_p/m_s$) was estimated following Petit (1992) and Ennos (1994):

$$r = \frac{m_p}{m_s} = \frac{\left(\frac{1}{G_{STb}} - (F_{IS})\right)\left(\frac{1}{G_{STm}}\right)}{1\left(1 + F_{IS}\right)\left(\frac{1}{G_{STm}} - 1\right)} \qquad (1)$$

where G_{STm} and G_{STb} correspond to the estimate of subdivision at maternally inherited markers and at nuclear (biparentally inherited) markers, and F_{IS} is the heterozygote deficit estimated with nuclear codominant markers. When information from paternally and biparentally inherited markers is available (e.g. in conifers), this ratio can also be estimated as follows:

$$r = \frac{m_p}{m_s} = \frac{\left(\frac{1}{G_{STp}} - 1\right) - \left(\frac{1}{G_{STm}} - 1\right)}{\left(\frac{1}{G_{STm}} - 1\right)} \qquad (2)$$

but contrary to Weir & Cockerham (1984), where populations with higher sample sizes are given more weight. When

assuming strict allogamy, with G_{STp} corresponding to the estimate of subdivision at paternally inherited markers. Interpretation of m_p/m_s values requires particular caution because of a number of unrealistic assumptions underlying the island model used to derive Equations 1 and 2. These assumptions include (1) no mutation and no selection; (2) equal number of migrants among all populations, implying a lack of spatial genetic structure; (3) identical male and female effective population sizes; and (4) equilibrium between genetic drift and gene flow (e.g. Whitlock & McCauley 1999; Oddou-Muratorio et al. 2001).

Most statistical analyses were performed with systat version 10.2. Parametric standard tests were used after transforming data to meet normality assumptions. In particular, G_{ST} data were arcsin square root transformed before proceeding with further analyses. Because original G_{ST} and m_p/m_s values were not normally distributed, we provide medians plus their quartiles instead of arithmetic means, as this permits a more straightforward comparison of categories (see also Morjan & Rieseberg 2004).

Results

Database

Data were obtained for 183 species belonging to 103 genera and 52 families (Table 1). The raw data and list of source references is provided in Table S1. Most studies have been published recently (30% since 2003) in a variety of journals

Table 1 Number of entries in the database, classified according to taxonomy, genome and mode of inheritance

	Number of species							
	Cp			Mt maternal	Nuclear biparental	Total*	Number of genera	Number of families
	Maternal	Paternal	All					
Angiosperms	138	0	138	13	86	141	95	49
Gymnosperms	0	37	37	21	33	42	8	3
All	138	37	175	34	119	183	103	52

*Total number of species represented, which is lower than the total of the values situated on the left, because a given species might have been studied with markers from more than one genome.

(29), with a maximum of 27% of all entries in a single journal (Molecular Ecology). Population studies based on organelle markers in plants have relied more on cpDNA markers (175 species) than on mtDNA markers (34 species), as stressed elsewhere (Petit & Vendramin 2005). Twenty-six species have been investigated at both cpDNA and mtDNA markers. Gymnosperms have been relatively well studied (42 of the c. 600 existing species investigated). They are particularly interesting models for studies on gene flow as many species have paternally inherited cpDNA and maternally inherited mtDNA (whereas both genomes are maternally inherited in most angiosperms). Raw data on haplotype frequencies were obtained in a large fraction of entries (92%). The most frequent technique used to study oDNA variation has been PCR-RFLP (46%), followed by Probe-RFLP (28%), microsatellites (11%), direct sequencing (9%) and cpDNA purification followed by RFLP (4%). In terms of sampling, the median values across studies were 157 individuals screened in total, distributed in 14 populations with an average of 10 individuals per population, allowing the identification of approximately seven haplotypes on the basis of around seven polymorphic 'characters'. However, sampling strategies were very heterogeneous across species and each of these parameters varied by at least one order of magnitude. Studies of genetic structure based on nuclear markers were identified for 119 of the 183 species (65%), most of them (72%) based on isozyme markers (Table S2).

Relationships between parameters

As expected, the total number of individuals analysed per species is positively related with the number of populations sampled (Pearson $r = 0.62$, $n = 207$, $P \ll 0.001$) and, to a lesser extent, with the mean sample size within population ($r = 0.44$, $P \ll 0.001$). There is a negative correlation between the number of populations investigated and the number of individuals per population ($r = -0.42$, $P \ll 0.001$), a result of the trade-off between sampling many individuals per population or many populations. On the other hand, sampling did not affect G_{ST} (all correlations between sample sizes and G_{ST} vary between −0.1 and 0.1), as expected for an unbiased estimate. However, there was a significant negative correlation between G_{ST} and the number of haplotypes ($r = -0.44$, $n = 171$, $P \ll 0.001$). Two other variables affect G_{ST} at maternally inherited markers: the molecular technique used and the coverage of the range. The effect of the technique (one-way anova: $F = 3.03$, d.f. = 5, $n = 171$, $P = 0.01$) mostly resulted from the fact that the eight studies based on purification of cpDNA followed by restriction enzyme digestion had higher G_{ST}. Similarly, the effect of range coverage ($F = 3.93$, d.f. = 2, $P = 0.02$) resulted from the fact that the species studied in a restricted area (i.e. locally) had lower G_{ST} compared to those studied across a larger fraction of the species' range. For nuclear data, an effect of the technique was also noted ($F = 5.11$, d.f. = 6, $n = 110$, $P < 0.001$). It was the result of a higher mean subdivision in studies based on RAPD (11 cases), whereas estimates of genetic subdivision based on the other techniques were smaller and of similar magnitude (results not shown). On the other hand, no effect of range coverage was noted with these markers ($F = 1.98$, d.f. = 2, $n = 109$, $P = 0.14$).

Precision of G_{ST} estimates

The precision (standard error) of the estimates of population subdivision (G_{ST}) at organelle genes was also investigated. The 162 available estimates of G_{ST} standard error (at both maternally and paternally inherited markers) were plotted as a function of the number of populations, the total number of individuals analysed per species and the harmonic mean sample size per population (Fig. 1). The best predictor of the G_{ST} standard error (as assessed by the coefficient of determination R^2) was the total number of individuals analysed, followed by the number of populations and the mean sample size per population (all after log transformation). The relation between total diversity h_T and standard error

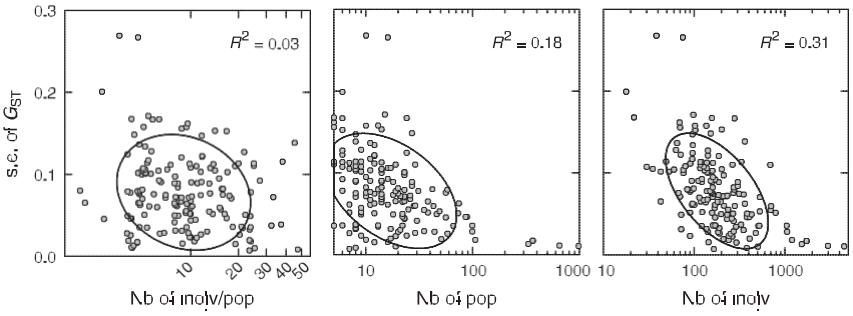

Fig. 1 Precision of G_{ST} as a function of the sampling of individuals within population, populations or total number of individuals (Nb = Number).

of G_{ST} was also estimated and was found to be negative and significantly different from zero (Pearson r = −0.22, P < 0.001).

Selection of studies for further analysis

The above results indicated two possible sources of bias: the incorporation of species sampled in a too restricted part of their range and the inclusion of studies relying on prior purification of cpDNA. In the latter case, we attribute the bias to the fact that, in five out of eight cases, sampled individuals had been pooled to minimize the number of DNA isolation and purification experiments. Although the authors of these papers claimed that they would have detected any mixture of haplotypes from the resulting banding pattern (e.g. Soltis et al. 1991), some bias seems likely. All studies in which such a procedure had been used were therefore discarded from further analysis (12 in total, all based on maternally inherited markers; as a matter of fact, in only one of these 12 studies did the authors detect some intrapopulation variation). Similarly, those seven oDNA studies where only a small part of the species range had been investigated were removed, as well as one study for which the observed level of total diversity was particularly low (cf. our finding that a low level of diversity reduces the precision of G_{ST}). This left us with a total of 152 entries based on maternally inherited markers, corresponding to 144 different species, as in eight cases, results from both maternally inherited cpDNA and mtDNA markers were available. No study involving paternally inherited markers was deleted (grand total of 166 species).

Taxonomic effects

The effect of taxonomic identity on G_{ST} was investigated using a nested anova with as main effects the factors 'family' and 'genus nested within family'. For maternally inherited markers, the model explained 73% of the variance in G_{ST}. For paternally inherited markers, only the genus effect could be tested because most species involved belonged to Pinaceae; it was not significant.

Genetic subdivision at markers with different modes of inheritance

The distributions of G_{ST} values and their means were computed separately for angiosperms and gymnosperms and for maternally, paternally or biparentally inherited markers (Table 2; Fig. 2). The mode of inheritance has a major effect on the partitioning of genetic diversity, with studies based on maternally inherited markers having considerably higher G_{ST} than those based on paternally or biparentally inherited markers for both gymnosperms and angiosperms. On the other hand, there is no significant difference between G_{ST} at biparentally inherited markers and at paternally inherited markers in gymnosperms (separate variance t-test: t = −0.91, d.f. = 54.4, P = 0.37). A significant difference between angiosperms and gymnosperms was found at maternally inherited markers (t = −2.35, d.f. = 36.1, P = 0.024), whereas at nuclear markers,

Table 2 Genetic differentiation according to mode of inheritance in angiosperms and gymnosperms (conifers)

Taxonomic group	N	Mean* ± SE	Median [Q_{25}, Q_{75}]†
Conifers			
Paternal	37	0.165^{ab} ± 0.036	0.099 [0.033, 0.163]
Maternal	20	0.764^{d} ± 0.008	0.759 [0.655, 0.890]
Biparental	33	0.116^{a} ± 0.003	0.088 [0.044, 0.152]
Angiosperms			
Paternal	—	—	—
Maternal	124	0.637^{c} ± 0.002	0.646 [0.416, 0.871]
Biparental	77	0.184^{b} ± 0.002	0.137 [0.064, 0.230]
All			
Paternal	37	0.165 ± 0.036	0.099 [0.033, 0.163]
Maternal	144	0.655 ± 0.002	0.673 [0.459, 0.879]
Biparental	110	0.163 ± 0.001	0.115 [0.057, 0.199]

*When both cpDNA and mtDNA data were available, only one data set was used in case of identical mode of inheritance (that based on the highest sample size or if similar on the most polymorphic markers, see Table S1). Superscript letters indicate significant differences between means at P < 0.05 (see text for details); †Q_{25} and Q_{75} are the first and third quartiles of the distribution of G_{ST} values.

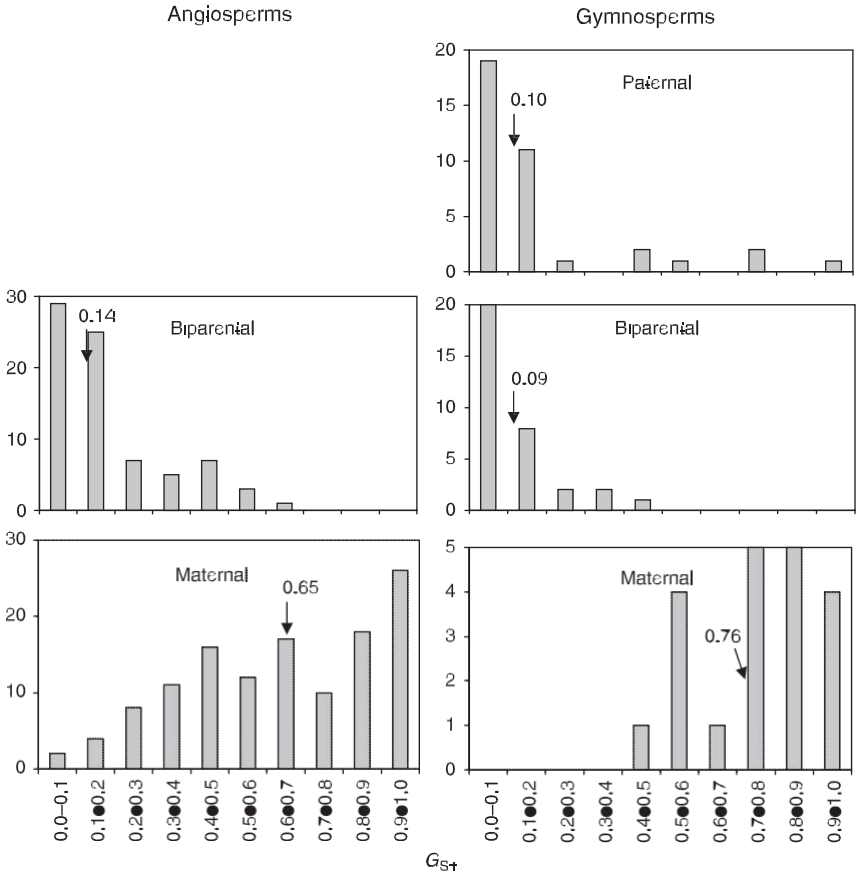

Fig. 2 Distribution of G_{ST} estimates classified by taxonomic group and mode of inheritance. Median values are provided for each case.

angiosperms also have significantly higher G_{ST} than gymnosperms (t = 2.66, d.f. = 77.2, P = 0.009).

Covariation between G_{ST} based on markers from different genomes

There were 23 species for which both cpDNA and mtDNA data are available, including eight cases in which both genomes are maternally inherited. Data from 93 species were available for the combination of maternally inherited (cpDNA, mtDNA or both) and biparentally inherited (nuclear) markers. There were 29 species with data from both paternally inherited (cpDNA) and biparentally inherited markers (all conifers); for 13 of these, data was also available from mtDNA (i.e. data was available from three differentially inherited genomes).

In conifers, G_{ST} is nearly always larger at mtDNA markers than at cpDNA markers (Fig. 3). In contrast, G_{ST} estimates in angiosperms are similar for the two genomes and covary rather narrowly, especially when they derive from the same study (i.e. are based on the same individuals; see Table S1). In both angiosperms and gymnosperms G_{ST} at nuclear markers (G_{STb}) was lower than at maternally inherited markers (G_{STm}), with only three exceptions out of 93 (Fig. 4). We also checked whether $G_{STb} < (2/G_{STm} - 1)^{-1}$ by indicating the corresponding curve on Fig. 4. This limit

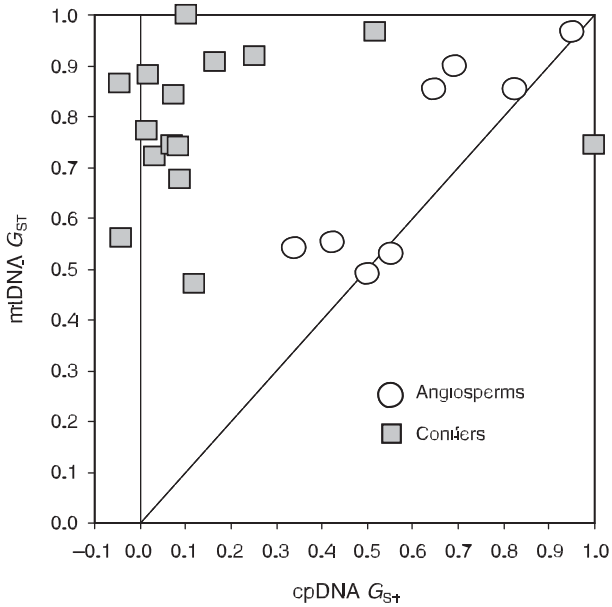

Fig. 3 Covariation between G_{ST} estimates based on markers from the two organelle genomes.

corresponds to the maximum value that G_{STb} can reach, for a given value of G_{STm} in an island model at equilibrium between migration and drift (Petit 1992; Petit & Vendramin 2005). There are seven cases out of 93 (including the three

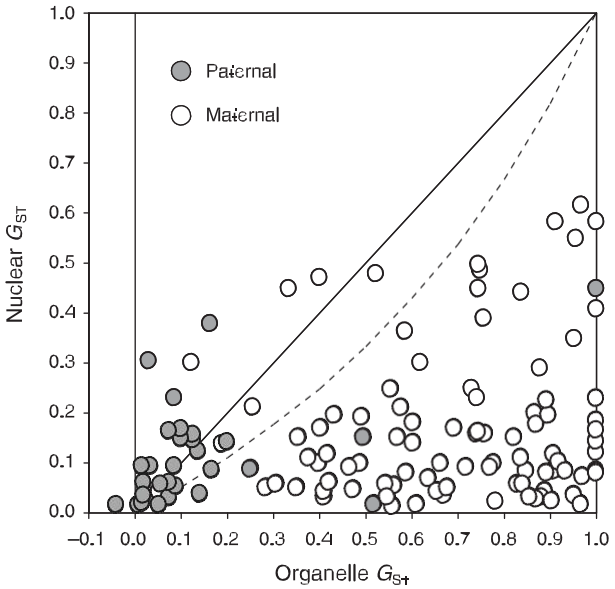

Fig. 4 Covariation between G_{ST} estimates based on markers from nuclear and organelle genomes. The dotted line below the diagonal corresponds to the maximum possible theoretical value for nuclear G_{ST} when contrasted with a maternally inherited gene and to the minimum possible value when contrasted with a paternally inherited gene.

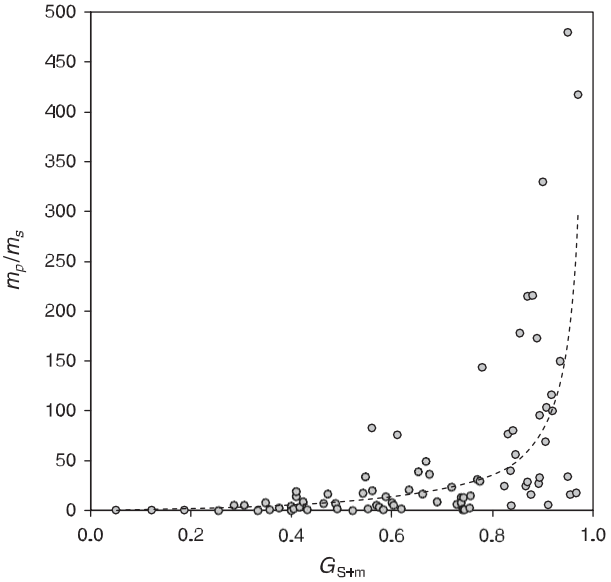

Fig. 5 Migration ratio as a function of genetic differentiation for maternally inherited genes (G_{STm}). The dotted line corresponds to the case where the migration ratio is independent of the pollen migration rate (m_p). The results indicate that the pollen migration rate is largely independent of the seed migration rate as the adjusted curve fits well to the observed data, with no trend for m_p/m_s to be either too large or too small at a given value of G_{STm}.

previous cases) that reach this limit. G_{ST} at paternally inherited markers (G_{STp}) was also compared to G_{STb} (Fig. 4). Most cases (25 out of 29) do not fall in the zone of covariation predicted by theory, which lies between the diagonal ($G_{STb} = G_{STm}$) and the previously described limit (Petit 1992; Petit & Vendramin 2005). Interestingly, in 17 out of 29 cases, G_{STb} is higher than the maximum predicted value (i.e. $> G_{STp}$), compared to only eight cases where G_{STb} is below the lower threshold [i.e. $< 1/(2/G_{STp} - 1)$]. That is, G_{STp} is often 'too low' or alternatively G_{STb} 'too high', compared to neutral equilibrium expectations. However, there remains a large and positive correlation between G_{STp} and G_{STb} (Pearson r = 0.52, P = 0.004), not observed between G_{STm} and G_{STb} (r = 0.13, P = 0.22).

Pollen-to-seed migration ratios

Equation 1 was used to derive the pollen-to-seed migration ratio, using as input values estimates of F_{IS}, G_{STb} and G_{STm}. As F_{IS} estimates were available only for a subset of studies (73%), we first checked whether results were much affected when this parameter was taken into account (Equation 1). Migration ratios were little affected (Pearson r > 0.999 between estimates based on actual F_{IS} values and those obtained by setting F_{IS} values to zero), so we decided to set F_{IS} to zero in all subsequent analyses. The median of the pollen/seed migration ratio estimates was 17, based on 93 individual measures [i.e. setting m_p/m_s to an arbitrary high value for those species where $G_{STm} = 1$, and setting it to zero for the seven species where $G_{STb} < (2/G_{STm} - 1)^{-1}$]. This result suggests an overall predominance of gene flow by pollen, although in 25 species (27% of total) seed dispersal appears to account for a large (> 20%) component of total gene flow (i.e. $m_p/m_s < 5$). Alternative estimates of the same migration ratio were obtained using Equations 1 and 2 in those conifers for which data was available from all three genomes. One species was excluded (Pinus pinaster) because there was complete fixation for mtDNA markers. For the remaining 12 species, there was no correlation between the two estimates of m_p/m_s (Pearson r = −0.12, P = 0.71), indicating that estimates of this parameter lack stability when different sources of data and/or markers are used.

Does gene flow by seeds covary with gene flow by pollen?

The m_p/m_s ratio obtained by contrasting G_{STm} and G_{STb} increases with G_{STm} (Fig. 5). However, this could result either from a reduction of seed flow across species, without any concomitant increase in pollen flow, or from an increase of pollen flow in species with lower seed flow (i.e. compensation between dispersal by pollen and by seed). A null hypothesis was constructed by assuming that m_p is

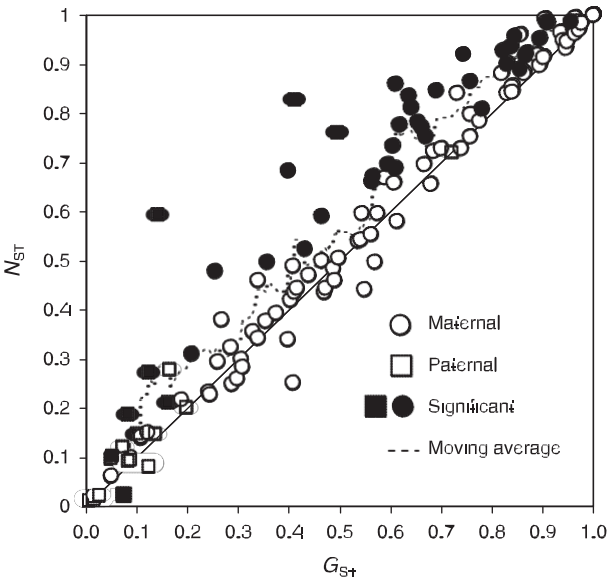

Fig. 6 Estimate of oDNA genetic subdivision based on similarities between haplotypes (N_{ST}) as a function of G_{ST}. A majority of values are found above the diagonal, including all but one of the significant differences between N_{ST} and G_{ST} (shown in black). Note the trend for N_{ST} based on maternally inherited markers to increasingly depart from the diagonal when G_{ST} increases.

independent of m_s. This implies that $m_p/m_s = K/(1/G_{STm} - 1)$, where K is a numerical constant (i.e. the migration ratio m_p/m_s is a function only of the denominator m_s). We fitted this function with the data presented in Fig. 5 by minimizing the sum of squares of the deviations, yielding K = 8.9. Observed values fall on both sides of the curve (Fig. 5), regardless of the value of the abscissa (G_{STm}). This means that m_p/m_s values are neither too large nor too low for a given value of G_{STm}, suggesting the independence of pollen flow from levels of seed flow and hence no tendency for pollen flow to compensate low levels of seed flow.

Alternative ways to measure differentiation: G_{ST} vs. N_{ST} vs. R_{ST}

For 140 entries, both G_{ST} and N_{ST} could be estimated (118 with maternally inherited markers and 22 with paternally inherited markers). On average, N_{ST} is higher than G_{ST} (overall mean is 0.69 compared to 0.65 for maternally inherited markers, and 0.23 instead of 0.15 for paternally inherited ones; the mean G_{ST} values differ slightly from those in Table 2 given that they are based on a subset of all species). There are 99 cases (71%) where $N_{ST} > G_{ST}$, and the difference between the two estimates is significant (at $P < 0.05$) in 41 cases (including 32 cases based on maternally inherited markers and nine based on paternally inherited markers; Fig. 6). By contrast, a single case study generated a G_{ST} significantly higher than its corresponding N_{ST}. For maternally inherited markers, there is a tendency for N_{ST} to depart more from G_{ST} at increasing levels of differentiation: there were 26 out of 79 (33%) significant tests when $G_{ST} > 0.5$, compared to six out of 39 (15%) for $G_{ST} < 0.5$. However, this trend did not hold for paternally inherited markers.

There were 20 data sets where R_{ST} could be estimated. All are based on chloroplast microsatellite (cpSSR) data (16 conifers and four angiosperms). The average R_{ST} is higher than N_{ST}, itself higher than G_{ST}, and differences are often great (R_{ST} higher than G_{ST} by an average of 0.1) (Table 3). R_{ST} is significantly higher than G_{ST} in nine out of the 20 cases (Table 3).

Discussion

Our survey of population studies of organelle DNA in plants includes data from 183 species, which should come close to an exhaustive compilation of the data available in spring 2004. The first population genetic studies based on oDNA that were suitable for inclusion (i.e. that had sampled different individuals within several populations without bulking them in the molecular analyses) were by Banks & Birky (1985) on Lupinus texensis (98 individuals in 15 populations) and by Neale et al. (1986) in Hordeum vulgare (229 individuals in 21 populations, excluding cultivated accessions). Although a high level of subdivision was already apparent in these studies, direct estimates of G_{ST} for oDNA markers were first calculated in the 1990s (Kremer et al. 1991; in Quercus robur and Quercus petraea), making it clear that subdivision at cpDNA markers could be considerably larger than at nuclear ones.

For this compilation, measures of oDNA population structure are based in most cases on our re-analyses of the original raw data sets to improve comparability across studies. Indeed, different estimators of genetic differentiation have often been used in the literature, somehow restricting the interest of comparative studies (e.g. Nybom & Bartish 2000; Nybom 2004; but see Morjan & Rieseberg 2004). In principle, the type of molecular technique used could also affect measures of subdivision of diversity for a given genome. For instance, selection could affect estimates of genetic subdivision at protein markers (McDonald 1994), high mutation rates could affect subdivision at SSR markers (Hedrick 1999) and anonymous detection techniques such as RAPD or AFLP could include not only fragments from the nuclear genome but also oDNA fragments, resulting in upwardly biased measures of nuclear subdivision if undetected (Aagaard et al. 1995). However, the differences in levels of subdivision across markers found in the present paper were limited and seemed to be the result of other factors (such as bulked analyses of DNA samples, which were subsequently removed from the analyses) or involved only a small fraction of the studies (in the

Table 3 Comparison of differentiation estimates for cpSSR data

Species†*	Number of loci	Inheritance‡	G_{ST}	N_{ST}§	R_{ST}§
Abies guatemalensis	2	P	0.137	0.148	0.231
Abies nordmanniana	3	P	0.007	0.012	0.020
Alyssum bertolonii	5	M	0.210	0.310*	0.475*
Arabidopsis thaliana	8	M	0.968	0.990	0.991
Caesalpinia echinata	7	M	0.911	0.980*	0.997*
Picea abies	3	P	0.089	0.095	0.063
Pinus albicaulis	3	P	0.073	0.024	0.102
Pinus brutia	8	P	0.124	0.271*	0.348*
Pinus canariensis	6	P	0.085	0.187*	0.174*
Pinus cembra	3	P	0.029	0.022	0.017
Pinus halepensis	6	P	0.163	0.210*	0.120
Pinus lambertiana	8	P	0.143	0.592*	0.840*
Pinus leucodermis	10	P	0.054	0.100*	0.130*
Pinus mugo	3	P	0.052	0.096*	0.122*
Pinus nelsonii	4	P	0.126	0.081	0.035
Pinus pinaster	7	P	0.101	0.146*	0.191*
Pinus pinceana	4	P	0.495	0.760*	0.929*
Pinus resinosa	3	P	0.720	0.721	0.681
Pinus sylvestris	6	P	0.017	0.018	0.018
Vitis vinifera	2	M	0.188	0.220	0.282
Mean	5		0.235	0.299	0.338

†References are provided in Table S1; ‡P: paternal inheritance, M: maternal inheritance; §Asterisks indicate that N_{ST} (or R_{ST}) is significantly higher than G_{ST} ($P < 0.05$).

case of nuclear differentiation). Previous comparisons have shown some discrepancy between estimates of subdivision based on different nuclear markers for a given species (e.g. Nybom 2004) but no systematic bias has been reported (Vandewoestijne & Baguette 2002; Nybom 2004). As a consequence, we do not expect that the main conclusions of our studies could be affected by the heterogeneity of techniques used, although the field would certainly gain from further harmonization and standardization of techniques and methods of data analyses.

We also computed standard errors of examined parameters to assess the effects of sampling on the precision of estimates. As expected, the total number of sampled individuals is the best predictor of the precision of G_{ST}. However, we could also confirm an earlier observation (Pons & Petit 1995) that the precision of parameter estimates is more affected by the number of populations sampled than by the sampling of individuals within populations. This suggests that future studies on the geographical structure of plant populations based on oDNA should allocate most efforts to the sampling of as many populations as possible, even at the expense of within-population sampling. Sequential approaches first analysing simultaneously (i.e. bulking) several individuals from the same population followed by separate analyses of each individual when within-population variation is detected, do not appear to be advisable, given the prevalence of intrapopulation diversity. As a matter of fact, complete fixation ($G_{ST} = 1$) was detected in only 11 out of 152 studies based on maternally inherited markers.

The finding that the portion of the range sampled affects G_{ST} estimates confirms similar results by Morjan & Rieseberg (2004) with data from both plants and animals. Several studies have shown that differentiation generally increases with distance (e.g. Dumolin-Lapègue et al. 1997; Grivet & Petit 2002a; Palmé & Vendramin 2002; Heuertz et al. 2004), so detecting higher G_{ST} values in species-wide surveys makes sense. It would be desirable that future studies provide complete curves of differentiation as a function of distance, instead of single estimates of measures of differentiation, allowing standardization of G_{ST} to a common reference distance for comparison purposes, as recently performed in an analysis of mtDNA in vertebrates (Martin & McKay 2004).

In the present paper, the genera with the highest number of species studied at maternally inherited markers were Pinus and Quercus (11 and 10, respectively, out of a total of 144 species). Although these genera might be expected to bias overall estimates, averaging across genera instead of species changed mean G_{ST} estimates by less than 1%, so the values presented here should be representative, at least taxonomically. However, the finding that the taxonomic identity of species explains a considerable amount of variation in G_{ST} does challenge the significance of previously identified relationships between levels of genetic differentiation, life history traits or ecological attributes. Because G_{ST} values of

species belonging to the same genus or family tend to be similar, analysing them independently increases the risk of statistical pseudoreplication. In fact, a recent study relying on phylogenetically informed analytical methods has demonstrated that the use of direct species comparisons without consideration of their phylogenetic relationships might result in many false positives when seeking to identify relations between G_{ST} and life history traits or ecological attributes of species (Aguinagalde et al. 2005).

The most striking contrast between G_{ST} estimates based on markers from different plant genomes is that between maternally inherited markers and markers having other modes of inheritance (biparental or paternal). Compared to biparentally inherited nuclear markers, maternally inherited oDNA markers display not only larger values of G_{ST} but also more heterogeneous values, occupying the full spectrum between little population structure (although cases below 0.1 are unusual: only two examples were found in this review) and near complete fixation (which is not rare, even in species that have been well-sampled and which display sufficient levels of total diversity, see, e.g. Grivet & Petit 2002b or Burban & Petit 2003). There is more heterogeneity in G_{ST} estimates based on paternally inherited markers than for those based on biparentally inherited ones, which might be attributed to the greater action of drift on effectively haploid genomes.

The lower levels of differentiation observed at nuclear genes in conifers compared to angiosperms merely confirm earlier studies (e.g. Hamrick et al. 1992), but the fact that this is matched by greater (and not lower) levels of differentiation at maternally inherited markers is a first indication that measures of genetic differentiation at different markers need not covary positively. In fact, no correlation was found between G_{ST} at nuclear loci and maternally inherited oDNA genes. On the other hand, a strong covariation of G_{ST} across genomes was observed between cpDNA and mtDNA in angiosperms, where both are generally maternally inherited. Strictly co-transmitted uniparentally inherited genomes represent a single 'locus' and should be characterized by similar levels of differentiation (Dumolin-Lapègue et al. 1998; Desplanque et al. 2000; Olson & McCauley 2000; Belahbib et al. 2001; Huang et al. 2001). In conifers, paternally and biparentally inherited markers also have very similar levels of genetic structure, as shown by the relatively tight positive correlation between the two sets of G_{ST} estimates. This makes sense as both the cpDNA and nuclear genomes are moved by pollen and by seeds, i.e. they use the same vehicles to achieve gene flow (note that paternal mode of transmission does not imply transmission by pollen only, as sometimes stated in the literature, e.g. Morjan & Rieseberg 2004). The only difference is that paternally inherited genes will experience dispersal by pollen at each reproductive cycle, whereas only 50% of the nuclear genes will experience gene flow by pollen in a given cycle (all genes experience gene flow by seeds). However, G_{STb} is often 'too large' (or G_{STp} 'too low'), compared to equilibrium conditions that predict that $G_{STb} \leq G_{STp}$ (e.g. Petit et al. 1993). One possible interpretation for this observation has been proposed by Petit et al. (1993; see their Fig. 2): as cpDNA is effectively haploid, its effective population size is lower than that of a nuclear gene (Birky et al. 1983, 1989). Hence, G_{ST} values at cpDNA markers will reach equilibrium faster than G_{ST} values at nuclear genes, resulting in transient situations where $G_{STp} < G_{STb}$ (assuming that for both genomes the initial situation is characterized by larger-than equilibrium G_{ST} values).

The median m_p/m_s value across all species was c. 17, suggesting that gene flow through pollen is quantitatively much more important than through seeds. From these results, one might conclude that gene flow is generally asymmetrical in seed plants, with most species relying predominantly on pollen for gene exchanges. However, there are 27% of plants with m_p/m_s estimates below five, indicating that seeds can also play a significant role in overall gene flow (i.e. > 20% of total; typical examples from this category would include insect-pollinated and fleshy-fruited species, for which pollen dispersal is limited while frugivorous vertebrates render long-distance seed dispersal relatively common, e.g. Oddou-Muratorio et al. 2001).

Comparisons between various G_{ST} estimates indicate departure from migration-drift equilibrium, from neutrality or from an island model of population structure, so these conclusions are at best very rough indications and could even be misleading. Furthermore, the absence of correlation between migration ratios derived by contrasting either G_{STp} or G_{STb} with G_{STm} indicates that these estimates do suffer from low stability. Nevertheless, given the large variation in G_{ST} values observed across species, the m_p/m_s ratio should provide an approximate idea of the relative importance of the two components of historical (rather than current) gene flow. Morjan & Rieseberg (2004), while acknowledging the unrealistic assumptions made in such studies, have similarly concluded that estimates of number of migrants (Nm) do not appear to be so biased as to mask expected interspecific trends in patterns of gene flow. Moreover, an earlier study based on two different methods to estimate this ratio yielded similar estimates, thus providing some support for this approach (Oddou-Muratorio et al. 2001).

In principle, it could be of interest to check if species having high (or low) pollen dispersal ability simultaneously tend to have high (or low) seed dispersal ability. We are not aware of the question having ever been raised so far. In the near future, direct estimates of dispersal distances obtained by parentage analysis might provide an answer to this question. In the meantime, we have used an indirect approach, susceptible to the same criticisms made previously, to check for the existence of some kind of compensation

between both forms of gene flow. For this purpose, we built a null hypothesis that m_p is independent of m_s, i.e. that it does not increase when G_{STm} increases (i.e. when m_s decreases). This scenario was fitted on the data of Fig. 5, indicating no visible trend besides the logical increase of m_p/m_s with decreasing m_s (i.e. increasing G_{STm}). In other words, no sign exists that would indicate either compensation or positive covariation between gene flow through pollen and through seeds.

Our last analysis consisted in testing whether the presence of a phylogeographical component of population structure is a common phenomenon in plants. By this we mean an additional component of geographical structure not seen when considering only differences in allelic frequencies between populations. It is obtained by subtracting standard G_{ST} (or F_{ST}) estimates from measures such as N_{ST} (e.g. Pons & Petit 1996) or R_{ST} (Slatkin 1995) that take into account similarities between haplotypes. In the frame of an island model, the finding of a significant phylogeographical component would provide support for historical (non equilibrium) population genetic structure, as R_{ST} and F_{ST} (= G_{ST}) are expected to converge to the same values under equilibrium conditions (Slatkin 1995) but it could equally point to an equilibrium situation in an isolation-by-distance model, as R_{ST} is predicted to be higher than F_{ST} in this case (Rousset 1996; see his Fig. 1).

Regardless of the cause for the difference between estimators, it appears that measuring genetic differentiation without including information on similarities between haplotypes results in the loss of useful information. In fact, in many cases $N_{ST} > G_{ST}$, and the trend is even stronger for R_{ST}. The same result was found on the basis of a smaller data set analysed by Petit et al. (2003) and Aguinagalde et al. (2005). In contrast, earlier studies by other authors had often questioned the relevance of R_{ST} estimates, in view of its higher variance compared to other estimators of F_{ST} (e.g. Streiff et al. 1998; Gaggiotti et al. 1999; Balloux et al. 2000; Balloux & Goudet 2002; Balloux & Lugon-Moulin 2002; Kalinowski 2002; Neigel 2002). Our findings clearly support the use of N_{ST} or R_{ST} for estimating organelle population genetic structure, especially at broad geographical scales. This fits with theoretical findings that have shown that 'the relative efficiency of allele size-based vs. allele identity-based statistics depends on the relative contributions of mutations vs. drift to population differentiation' (Hardy et al. 2003), and hence on the spatio-temporal scale considered.

In any case, the use of these estimators further emphasizes the high level of geographical structure found in most plant species when using maternally inherited markers, in stark contrast with biparentally inherited nuclear markers. This finding has attracted much interest for about 15 years, giving birth to a new discipline, plant phylogeography, and leading to the development of many important practical applications, such as traceability and ecocertification of forest products and the identification of plant populations for conservation.

Acknowledgements

The research was supported be the EC research program FAIR5-CT97-3795 (CYTOFOR) to RJP, SF and GGV and by a grant from the Bureau des Ressources Génétiques to RJP. AH received financial support from the FPI grant PB98-1144 of the Spanish Ministry of Science and Technology. We thank the following colleagues who have contributed unpublished data or complementary information for inclusion in the database: Richard Abbott, Itziar Aguinagalde, Erika Aguirre Planter, François Balfourier, Guillaume Besnard, Christiane Bittkau, Isabelle Bonnin, Jean Bousquet, Linda Broadhurst, Steven Brunsfeld, Margaret Byrne, Henri Caron, Stephen Cavers, Catherine Clark, Els Coart, Rosane Collevatti, Salvatore Cozzolino, Dan Crawford, Mitch Cruzan, Nicolas Devos, Alexis Ducousso, Cyril Dutech, Richard Ennos, Colin Ferris, Isabelle Gamache, Leonardo Gallo, Pauline Garnier-Géré, Felix Gugerli, Myriam Heuertz, Juan Pablo Jaramillo, Andy King, Marcus Koch, Monika Konnert, Antoine Kremer, Martin Lascoux, Sascha Liepelt, Catarina Lira, Andy Lowe, Roselyne Lumaret, Carlos Magni, Alessio Mengoni, Jose Gabriel Segarra Moragues, Nicole Muloko-Ntoutoume, Gerhard Müller-Starck, Brigitte Musch, Sylvie Oddou-Muratorio, Matt Olson, Anna Palmé, Daniel Piñero, Daniel Prat, Jim Provan, Olivier Raspé, Sarah Rendell, Bryce Richardson, Joëlle Ronfort, Stéphanie Roux, Luis Gil Sanchez, Pierre Saumitou-Laprade, Vladimir Semerikov, Tim Sharbel, Marco Soliva, Christoph Sperisen, Ivana Stehlik, Naoki Tani, Michèle Tarayre, Yoshihiko Tsumura, Lisa Wallace and Birgit Ziegenhagen.

Supplementary material

The following material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/mec/mec2410/mec2410sm.htm

Table S1. Measures of population subdivision at cpDNA and mtDNA markers in seed plants

Table S2. Measures of population subdivision at nuclear markers for plants from Table S1

References

Aagaard JE, Vollmer SS, Sorensen FC, Strauss SH (1995) Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. Molecular Ecology, 4, 441–446.

Aguinagalde I, Hampe A, Mohanty A et al. (2005) Effects of life history traits and species distribution on genetic structure at maternally inherited markers in European trees and shrubs. Journal of Biogeography, in press.

Balloux F, Brunner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatellites can be misleading: an empirical and simulation study. Evolution, 54, 1414–1422.

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Molecular Ecology, 11, 771–783.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology, 11, 155–165.

Banks JA, Birky CW Jr (1985) Chloroplast DNA diversity is low in a wild plant, Lupinus texensis. Proceedings of the National Academy of Sciences of the USA, 82, 6950–6954.

Belahbib N, Pemonge M-H, Ouassou A, Sbay H, Kremer A, Petit RJ (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: Quercus suber and Q. ilex in Morocco. Molecular Ecology, 10, 2003–2012.

Birky Jr CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics, 103, 513–527.

Birky Jr CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for gene in mitochondria and chloroplasts, and some results. Genetics, 103, 513–513.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: A synthesis. In: Integrating Ecological and Evolutionary Processes in a Spatial Context (eds Silvertown J, Antonovics J), pp. 319–339. Blackwell Science, Oxford.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. Molecular Ecology, 12, 1487–1495.

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maritime pine bast scale Matsucoccus feytaudi Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. Molecular Ecology, 8, 1593–1602.

Desplanque B, Viard F, Bernard J, et al. (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in Beta vulgaris ssp. maritima (L.): the usefulness of both genomes for population genetic studies. Molecular Ecology, 9, 141–154.

Dumolin-Lapègue S, Demesure B, Le Corre V, Fineschi S, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics, 146, 1475–1487.

Dumolin-Lapègue S, Pemonge M-H, Petit RJ (1998) Association between chloroplast and mitochondrial lineages in oaks. Molecular Biology and Evolution, 15, 1321–1331.

El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. Molecular Ecology, 5, 547–555.

Ellstrand NC (1992) Gene flow by pollen: implications for plant conservation genetics. Oikos, 63, 77–86.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. Heredity, 72, 250–259.

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Molecular Ecology, 8, 1513–1520.

Grivet D, Petit RJ (2002a) Phylogeography of the common ivy (Hedera sp.) in Europe: genetic differentiation through space and time. Molecular Ecology, 11, 1351–1362.

Grivet D, Petit RJ (2002b) Chloroplast DNA phylogeography of the hornbeam in Europe: evidence for a bottleneck at the outset of postglacial colonization. Conservation Genetics, 4, 47–56.

Hamilton MB, Miller JR (2002) Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. Genetics, 162, 1897–1909.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forests, 6, 95–124.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics, 163, 1467–1482.

Harris SA, Ingram R (1991) Chloroplast DNA and biosystematics: The effects of intraspecific diversity and plastid transmission. Taxon, 40, 393–412.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. Evolution, 53, 313–318.

Heuertz M, Fineschi S, Anzidei M et al. (2004) Chloroplast DNA variation and postglacial recolonisation of common ash (Fraxinus excelsior L.) in Europe. Molecular Ecology, 13, 3423–3438.

Hu X-S, Ennos RA (1997) On estimation of the ratio of pollen to seed flow among plant populations. Heredity, 79, 541–552.

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY (2001) Organelle DNA phylogeography of Cycas taitungensis, a relict species in Taiwan. Molecular Ecology, 10, 2669–2681.

Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. Molecular Ecology, 11, 1263–1273.

Kremer A, Petit RJ, Zanetto A, et al. (1991) Nuclear and organelle diversity in Quercus robur and Q. petraea. In: Genetic Variation in European Populations of Forest Trees (eds Müller-Starck G, Ziehe M), pp. 141–166. Sauerländer's-Verlag, Frankfurt am Main.

Levin DA, Kerster HW (1974) Gene flow in seed plants. Evolutionary Biology, 7, 139–220.

Martin PR, McKay JK (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. Evolution, 58, 938–945.

McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution, 10, 198–202.

McDonald JH (1994) Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms. In: Non-Neutral Evolution. Theories and Molecular Data (ed Golding B), pp. 88–100. Chapman & Hall, New York, NY.

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. American Journal of Botany, 83, 383–404.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Molecular Ecology, 13, 1341–1356.

Neale DB, Saghai-Maroof MA, Allard RW, Zhang Q, Jorgensen RA (1986) Chloroplast DNA diversity in populations of wild and cultivated barley. Genetics, 130, 1105–1110.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253–259.

Neigel JE (2002) Is F_{ST} obsolete ? Conservation Genetics, 3, 167–173.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology, 13, 1143–1155.

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspectives in Plant Ecology, Evolution and Systematics, 3, 93–114.

Oddou-Muratorio S, Petit RJ, Le Guerroué B, Guesnet D, Demesure B (2001) Pollen- vs. seed-mediated gene flow in a scattered forest tree species. Evolution, 55, 1123–1135.

Olson MS, McCauley DE (2000) Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in Silene vulgaris. Proceedings of the Royal Society of London, B, 267, 1801–1808.

Ouborg NJ, Piquot Y, Van Groenendael JM (1999) Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology, 87, 551–568.

Palmé A, Vendramin GG (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, Corylus avellana. Molecular Ecology, 11, 1769–1779.

Petit RJ (1992) Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Doctorate Thesis, Paris XI-Orsay.

Petit RJ, Aguinagalde I, de Beaulieu J-L et al. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science, 300, 1563–1565.

Petit RJ, Kremer A, Wagner DB (1993) Finite island model for organelle and nuclear genes in plants. Heredity, 71, 630–641.

Petit RJ, Vendramin GG (2005) Phylogeography of organelle DNA in plants: an introduction. In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N). in press. Kluwer, Amsterdam.

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. Theoretical and Applied Genetics, 90, 462–470.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered vs. unordered alleles. Genetics, 144, 1237–1245.

Reboud X, Zeyl C (1994) Organelle inheritance in plants. Heredity, 72, 132–140.

Röhr H, Kües U, Stahl U (1998) Organelle DNA of plants and fungi: Inheritance and recombination. Progress in Botany, 60, 39–87.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics, 142, 1357–1362.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. Molecular Ecology, 7, 465–474.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics, 139, 457–462.

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest. Plant Systematics and Evolution, 206, 353–373.

Soltis DE, Mayer MS, Soltis PS, Edgerton M (1991) Chloroplast-DNA variation in Tellima grandiflora (Saxifragaceae). American Journal of Botany, 78, 1379–1390.

Streiff R, Labbé T, Bacilieri R et al. (1998) Within-population genetic structure in Quercus robur L. & Quercus petraea (Matt.) Liebl. assessed with isozymes and microsatellites. Molecular Ecology, 7, 317–328.

Vandewoestijne S, Baguette M (2002) The genetic structure of endangered populations in the Cranberry Fritillary, Boloria aquilonaris (Lepidoptera, Nymphalidae): RAPDs vs allozymes. Heredity, 89, 439–445.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358–1370.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. Heredity, 82, 117–125.

Wright S (1931) Evolution in Mendelian populations. Genetics, 16, 97–159.

Rémy J. Petit is studying population genetics and evolution of trees, with a special focus on phylogeography, organelle genetics and evolution, genetic diversity, palaeogenetics, and conservation genetics of trees. Jérôme Duminil is a PhD student at INRA, working on comparative studies of genetic diversity in plants. He is particularly interested in ecological, life historical and chorological features of species that can explain the distribution of genetic diversity. He is also using comparative methods to study organelle genome and sequence evolution. Silvia Fineschi is interested in the phylogeography and ecological genetics of angiosperm tree species. Daniela Salvini is a PhD student working on ecological and spatial genetics of white oaks, with particular interest on the study of gene flow and environmental features as factors shaping the natural genetic diversity of plant populations. Giovanni G. Vendramin's studies focus on conservation and population genetics, with particular emphasis on range-wide phylogeography and fine-scale population gene dynamics. Arndt Hampe has recently completed his PhD on the phylogeography and reproductive ecology of a climatic relict tree. He is now working on plant–animal interactions using both ecological and genetic approaches. The collaboration between the authors on this topic has started in 1998, during a stay of SF in Bordeaux, when the database was initiated. Since then, numerous exchanges, especially during the EU CYTOFOR project and then during stays of DS and AH in Bordeaux and of JD in Sevilla, have given further impetus to this work.