

RECOMBINATION AND LINKAGE DISEQUILIBRIUM AMONG MITOCHONDRIAL GENES IN STRUCTURED POPULATIONS OF THE GYNODIOECIOUS PLANT *SILENE VULGARIS*

David E. McCauley^{1,2} and Jennifer R. Ellis^{1,3}

¹Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235

²E-mail: david.e.mccauley@vanderbilt.edu

³E-mail: jennifer.ellis@vanderbilt.edu

Received August 8, 2007

Accepted January 6, 2008

The impact of intergenic recombination on the population genetics of plant mitochondrial genomes is unknown. In an effort to study this in the gynodioecious plant *Silene vulgaris* three-locus PCR/RFLP genotypes (based on the mitochondrial genes *atpA*, *cox1*, and *cob*) were determined for 239 individuals collected from 20 North American populations. Seventeen three-locus PCR/RFLP genotypes were found. Recombination was indicated by observation of each of the four two-locus genotypes possible when the two most common alleles are considered for each of two loci. Based on these common alleles the absolute values of standardized linkage disequilibrium $|D'|$ between pairs of loci range from 0.17 to 0.78. This indicates modest disequilibrium, rather than the maximum value expected in the absence of recombination $|D'| = 1$, or the linkage equilibrium expected if recombination is pervasive ($D' = 0$). Values of D' did not depend on which pair of loci contributed alleles to the analysis. The direction of D' obtained for the common *atpA* and *cox1* alleles was comparable in sign and magnitude to that obtained by examining similar information obtained in a prior study of European samples. All three loci indicated a high degree of population structure (average $F_{ST} = 0.63$), which would limit the within-population genetic diversity required for intergenic recombination to create novel genotypes, if most mating is local. Thus, population structure acts as a constraint on the approach to linkage equilibrium.

KEY WORDS: Gynodioecy, linkage disequilibrium, plant mitochondrial genome, population structure, recombination.

The population biology of the plant mitochondrial genome has attracted the increased interest of evolutionary biologists recently for a variety of reasons. Among these is the fact that mitochondrial genetic markers have been used in studies of population structure and gene flow by seeds (Petit et al. 2005), that plant mitochondrial DNA is used more frequently in phylogenetic studies (e.g., Barkman et al. 2000), and that mutations within the mitochondrial genome are a major component of sex determination in many gynodioecious species (Frank 1989; Budar et al. 2003). One potentially important, but little studied, aspect of the evolutionary dynamics of the plant mitochondrial genome is the role of intergenic recombination.

The plant mitochondrial genome is capable of both intra- and intermolecular recombination, resulting in the potential for both intra and intergenic recombination events and a rather fluid structure (Palmer and Hebron 1988; Mackenzie and McIntosh 1999). Intergenic recombination has the potential to create novel multilocus genotypes to the degree that the loci involved exhibit intermolecular sequence differences. Such differences would not be expected when strict maternal inheritance of the mitochondrial genome enforces a highly homoplasmic state within a cell. However, biparental inheritance of the mitochondrial genome resulting from paternal leakage could result in the heteroplasmy necessary for recombination to create novel genotypes (Städler and Delp

2002; McCauley et al. 2005; Welch et al. 2006), especially if the mitochondrial fusion known to occur in plants (Logan 2003, 2006; Arimura et al. 2004; Sheahan et al. 2005) were to allow the comingling of maternally and paternally derived genomes. Even under these circumstances recombination would be effective at generating novelty only to the degree that the maternally and paternally derived genomes differ.

Little is known about the role of intergenic recombination in the evolution of plant mitochondrial genomes or in plant mitochondrial population genetics. Evidence for the existence and consequences of intergenic recombination within plant mitochondrial genomes might be obtained from a population genetic perspective by examining patterns of multilocus genotypic diversity and linkage disequilibrium among mitochondrial genes in natural populations. In the absence of recombination only a limited number of multilocus genotypes are possible at a given level of allelic diversity. For example, if one considers two alleles at each of two loci, three two-locus haploid genotypes are possible without recombination, given that under this constraint only two can evolve from a common ancestor, barring homoplasmy of at least one of the alleles. Without recombination linkage disequilibrium between the loci should be at its maximum. With recombination between them a fourth haploid genotype is possible (hence, the "four gamete test" for recombination (Hudson and Kaplan 1985) and the magnitude of linkage disequilibrium should be lower. See Posada et al. (2002) for a general discussion concerning detection of recombination and its consequences.

Any study of the opportunity for intergenic recombination would have to consider the population structure and mating system of the species involved. Several properties of natural populations would influence the rate at which recombination might generate novel genotypes. Clearly, the frequency at which biparental inheritance generates heteroplasmic collections of mitochondria is important. However, this requires that maternally and paternally derived mitochondrial genomes differ. Many plant species exhibit very high levels of local population structure for organellar genes (Petit et al. 2005). If most mating is local then this high level of population structure would increase the probability that any copies of the mitochondrial genome transmitted through pollen could well be genetically identical to maternally derived copies, to the degree that it reduces within population genetic diversity relative to panmixia. Similarly, in a species with a high rate of self-fertilization transmission through pollen would not enhance the intracellular diversity of mitochondrial DNA when compared to that expected with outcrossing.

Recent studies have suggested that heteroplasmy and recombination could play a role in the mitochondrial population genetics of natural populations of plants in the genus *Silene* (Städler and Delph 2002; McCauley et al. 2005; Houlston and Olson 2006; Welch et al. 2006), although the evidence is not yet considered

by all to be definitive (Barr et al. 2007). Evidence for recombination in the mitochondrial genome in *Silene* is particularly interesting owing to the prevalence of gynodioecy in this genus. As in many gynodioecious systems sex determination seems to be cytonuclear (Charlesworth and Laporte 1998; Taylor et al. 2001), with the assumption that cytoplasmic male sterility (CMS) elements in the mitochondrial genome play an important role. Thus understanding the evolutionary dynamics of the mitochondrial genome in this genus may play a major role in understanding the evolutionary dynamics of gynodioecy; a phenomenon thought to be a premier example of genomic evolutionary conflict (Burt and Trivers 2006).

McCauley et al. (2005) studied the distribution of PCR/RFLP genotypes in a collection of *S. vulgaris* individuals and found all four possible genotypes when individuals were characterized for the mitochondrial genes *atpA* (*atp1*) and *cox1*, an indication of intergenic recombination. Although there was evidence for recombination, the *cox1* and *atpA* genotypes were not statistically independent, indicating some linkage disequilibrium, although it was not quantified. The 89 individuals used in that study were collected from a large number of localities within the species' native range in Europe, and the pooled collection was not analyzed in regard to local population structure.

These previous results raise several questions. First, what is the magnitude of linkage disequilibrium among mitochondrial genes in *S. vulgaris*? Is it at or near the maximum level expected in the absence of recombination, or much lower as expected should intergenic recombination be common? Second, what is the role of population structure in generating recombinatorial novelty? If local populations are highly structured in their mitochondrial diversity any consequential reduction in within-population genetic diversity would greatly limit the ability of recombination to create novel multilocus genotypes, relative to panmixia. Similarly, any linkage disequilibrium would be distributed as a primarily among population, rather than within population, phenomenon. Finally, *S. vulgaris* is native to Europe but is now widespread in North America and has been used as a model system for the study of the population genetic processes accompanying the establishment and spread of an invasive species (McCauley et al. 2003; Taylor and Keller 2007). It would be interesting to compare patterns of linkage disequilibrium found in North American *S. vulgaris* populations with those found previously in samples from Europe.

This is a study of patterns of linkage disequilibrium found among three mitochondrial genes, the *cox1* and *atpA* genes studied previously by McCauley et al. (2005), plus the *cob* locus. Three-locus genotypes were obtained from *S. vulgaris* individuals collected from 20 populations in the eastern United States and analyzed with regard to both population structure and patterns of linkage disequilibrium. We ask specifically (1) whether

the magnitude of pair-wise linkage disequilibrium and evidence for recombination depends on the pair of loci or specific alleles under consideration, (2) how the observed pattern of population structure might influence the consequences of recombination, and (3) how the observed pattern of linkage disequilibrium found in the North American samples compares to those seen previously in Europe. It is hoped that these results will contribute to a more general understanding of whether intergenic recombination within the plant mitochondrial genome influences its evolution, including its relationship to the problem of Muller's Ratchet often considered in the evolution of organellar genomes (Lynch and Blanchard 1998; Rand 2001) and quite relevant to the study of gynodioecy.

Materials and Methods

Between 4 and 15 individuals (total $n = 239$) were sampled from each of 20 local populations in the eastern United States, ranging from central Tennessee to northern Vermont (Fig. 1). All individuals collected from a given population were located less than 50 m from one another. Populations 1–16 were sampled in the summer of 2006 and populations 17–20 in the summer of 2002. In most cases DNA was extracted from leaves collected in the field, although a few individuals were sampled as seed (one/maternal line) and reared in the Vanderbilt greenhouse until sufficient leaf material developed for DNA extraction. Genomic DNA was extracted by using an Applied Biosystems 6100 Nucleic Acid PrepStation and associated protocols (Foster City, California) (populations 1–16) or by the method of Doyle and Doyle (1987) (populations 17–20).



Figure 1. The location of 20 *Silene vulgaris* populations from which a total of 239 individuals were sampled.

A two-stage genotyping strategy was employed that used both PCR/RFLP genotyping of the large number of individuals required for accurate measures of linkage disequilibrium and the additional sequencing of the PCR product obtained from a much smaller subset of those. First, each of the 239 individuals was assayed for its *atpA*, *cox1*, and *cob* PCR/RFLP genotype based on prior knowledge of restriction site variation in North American *S. vulgaris* populations. Welch et al. (2006) provide information concerning *atpA* and *cox1* PCR/RFLP's, including the primer sequences and cycle conditions used in the polymerase chain reaction. As in Welch et al. (2006) the restriction enzymes *MspI* and *AluI* were used to digest the 704 bp *atpA* PCR product and *DdeI* and *MspI* to digest the 1404 bp *cox1* product. Note that the primers used for *atpA* and *cox1* amplification here and in Welch et al. (2006) differ from those reported in McCauley et al. (2005). Consequentially, representative European samples studied by McCauley et al. (2005) were also genotyped for *atpA* and *cox1* with the primers and the allelic designations used here to facilitate a comparison of *atpA/cox1* disequilibrium patterns seen in the U.S., and those documented previously in Europe.

The *cob* assay was developed as follows. Several *S. vulgaris cob* sequences obtained from Houliston and Olson (2006) (GenBank accession numbers DQ841763, DQ841766, DQ841767, and DQ841768) were examined and determined to differ by an *ApoI* restriction site. This enzyme was used to assay a 1040 bp PCR product obtained using *cob* primers described in Städler and Delph (2002) and the "touchdown" PCR protocol used for *atpA* and *cox1* amplification here, and described in Welch et al. (2006). It was later found that two rare alleles could be distinguished as a subset of one of the *ApoI* types using the restriction enzymes *MspI* and *AluI*.

For each locus 10 μ l of PCR product was digested with the appropriate restriction enzyme or enzymes and then electrophoresed on a 4% Metaphor agarose gel and visualized with ethidium bromide staining.

Note that the PCR/RFLP genotyping used here would not detect most instances of heteroplasmy, given the results and methodology of McCauley et al. (2005) and Welch et al. (2006). Although instances of *atpA* heteroplasmy were fairly common in those studies, in most cases one of the two *atpA* alleles detected within heteroplasmic individuals was relatively much more common than the other. The minority allele could only be detected following a so-called knock back treatment that eliminated the PCR amplification of the majority allele. In that case the individual was said to be "cryptically heteroplasmic." The standard PCR methodology used here would not detect this cryptic heteroplasmy, and thus, strictly speaking, the genotype of an individual is its majority haplotype, if that individual is indeed cryptically heteroplasmic.

Much of the analysis presented here is based on the assumption that a given PCR/RFLP genotype evolved only once. This

assumption could be problematic if independent but identical mutations result in independent gains of the same restriction site, or more likely, if several different nucleotide substitutions could each result in the loss of a given restriction site. For that reason the second stage of genotyping consisted of sequencing much or all of the *atpA*, *cox1* and *cob* PCR products obtained from a subset of the sampled individuals consisting of two representatives (from different populations) of each of the seven most common three-locus genotypes (see Results). Forward and reverse primers used to sequence *atpA* and *cob* products were the same as those used in the PCR/RFLP survey. For *cox1* sequencing an internal reverse primer (5'-CCA TGC CTA GAT ACC CGA AG-3') was combined with the forward primer used for the *cox1* PCR/RFLP to generate a fragment of a manageable size containing the diagnostic restriction site. Sequencing was performed by the Vanderbilt University DNA Sequencing Facility using BigDye Terminator chemistry (Applied Biosystems) and electrophoresed on ABI's 3730xl DNA Analyzer (Applied Biosystems).

Two questions were asked of the sequence data (GenBank accession numbers EU440345-EU440352). First, are the sequences of a given PCR/RFLP allele of a given locus identical regardless of the alleles at the other two loci with which it was associated in that individual? If so, independent evolution of that allele in different genotypic lineages would require independent nucleotide substitutions defining restriction sites, in an otherwise highly conserved sequence. Second, do different PCR/RFLP alleles differ at the sequence level at sites other than those detected by the restriction digests employed here? If so, convergent evolution of PCR/RFLP alleles would require independent mutations in different lineages additional to those that define the diagnostic restriction sites.

The statistical association of the PCR/RFLP alleles was then tested for each pair of loci using G-tests of Independence (Sokal and Rohlf 1995), the same tactic used by McCauley et al. (2005) in their analysis of the European sample. For the North American sample rare alleles (see Results) were pooled for each locus so that all cells contained at least the minimum expected number required

by these tests under the null hypothesis of no association (linkage equilibrium). Linkage disequilibrium (D) was also calculated from the full North American PCR/RFLP dataset for all possible pairs of alleles at each of the three pairs of loci (i.e., *atpA/cox1*, *atpA/cob*, *cox1/cob*) using the following equation from Hedrick (2005a)

$$D = x_{ij} - p_i q_j,$$

where x_{ij} is the observed frequency of individuals with a particular two-locus genotype, p_i is the frequency of the relevant allele at the first locus, and q_j is the frequency of the relevant allele at the second locus. This value was then standardized by dividing D by D_{\max} , which is the maximum value of D possible given the observed allele frequencies p_i , $(1 - p_i)$, and q_j , $(1 - q_j)$ (Hedrick 2005a, p 532), to obtain D' . The standardized linkage disequilibrium D' then ranges from -1 to 1 with $D' = 0$ indicating linkage equilibrium. Standardized linkage disequilibrium was also calculated for the European *atpA/cox1* dataset gathered previously, using the allelic nomenclature based on the primers used here.

Population structure was evaluated first by calculating population-specific allele frequencies for the three loci. A haploid form of F_{ST} was then calculated for each locus using the methods of Weir and Cockerham (1984) and software ARELEQUIN 2.0 (Schneider et al. 2000). A composite value of F_{ST} was also calculated by combining information across the three single locus estimates, recognizing that linkage disequilibrium reduces their independence. In addition F_{ST} was calculated based on the population-specific frequencies of the three-locus genotypes.

Results

Among the 239 North American individuals sampled, five PCR/RFLP alleles were found for *atpA* and four PCR/RFLP alleles were found for *cox1* and *cob*. They will be referred to here as alleles A–E, 1–4, and a–d for *atpA*, *cox1*, and *cob*, respectively (see Fig. 2 for overall allele frequencies). Among the 239 individuals 17 three-locus combinations were found, of the 80 combinations

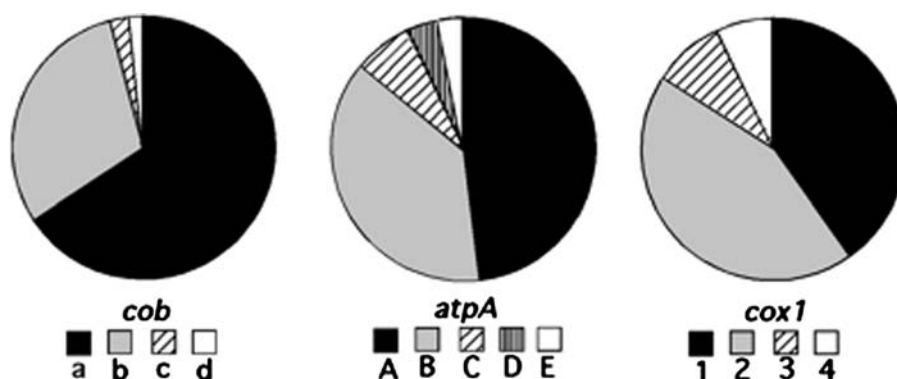


Figure 2. PCR/RFLP allele frequencies for three *Silene vulgaris* mitochondrial genes based on a sample of 239 individuals.

Table 1. Three-locus mtDNA haplotype combinations (*atpA/cox1/cob*) found among 239 *Silene vulgaris* sampled in the eastern United States. Also included are the number of individuals of each type, and the populations in which it was found (1–20, see Fig. 1).

Haplotype	No. of Individuals	Population
A1a	14	6,14,15,19
A1b	49	3,9,14,15,16,17
A2a	38	1,8,10,12,13,14,18
A2d	2	18
A3b	6	3,16
A3c	6	16
B1a	11	1,4,14,15
B1b	5	11,14,15
B2a	56	1,4,5,6,9,11,12,13,14,18,20
B2b	7	11,12
B2d	2	18
B3a	8	3,4
B3b	1	7
C1a	17	7
D4a	10	2
E4a	3	11,15
E4b	4	11,16

($5 \times 4 \times 4$) theoretically possible (Table 1). If just the two most common alleles found at each locus are considered, seven of the eight mathematically possible three-way combinations are included in the sample (A2b was not encountered). Because only four three-locus genotypes can evolve by a strictly branching process (if limited to two alleles/locus without homoplasy), the discovery of these seven combinations is evidence for recombination.

All 12 possible two-locus genotypes were found when the combinations of these common alleles are considered for the three possible pairs of loci (e.g., haplotypes Aa, Ab Ba, and Bb were found when only *atpA* and *cob* are considered). Thus, all four possible combinations were found for the common alleles at each pair of loci, a further indication of intergenic recombination. When the rare alleles are also inspected five additional cases can be seen in which all four two-locus combinations were found, if two alleles at each of two loci are considered (specifically, the combinations A1/A3/B1/B3, A2/A3/B2/B3, Aa/Ab/Ea/Eb, 1a/1b/3a/3b, and 1a/1b/4a/4b). However the alleles were not randomly associated, as indicated by the results of the G-tests of Independence conducted on each of the three sets of associations ($G = 86.9$ for *atpA/cox1*, $G = 46.4$ for *atpA/cob* and $G = 81.2$ for *cox1/cob*; $df = 4$ and $P < 0.001$ in each case). Thus, the null hypothesis of no linkage disequilibrium (all $D = 0$ or $D' = 0$) is rejected for each pair of loci.

Copies of each of the two common PCR/RFLP alleles at each locus were sequenced from genomic DNA obtained from a subset of those individuals carrying that allele in combination with each

of the common alleles at the other two loci. A common 653-bp sequence of *atpA* revealed a consistent 6 bp difference between the “A” and “B” PCR/RFLP alleles (all synonymous substitutions). Included in this difference is the diagnostic *AluI* restriction site that distinguishes the A and B restriction fragment profiles. However, no variation was found within the set of A sequences or within the set of B sequences, regardless of which *cox1* or *cob* allele *atpA* was paired with in the individual from which the DNA was obtained. A set of 973 bp sequences of *cob* consistently differed by only a 1 bp nonsynonymous substitution (the diagnostic *ApoI* restriction site), between individuals carrying the “a” and “b” alleles, although again there was no variation within the group of a alleles, or within the group of b alleles, regardless of the *atpA* and *cox1* allele with which they were paired. Finally, two synonymous substitutions, including the diagnostic *MspI* restriction site, distinguish the “1” and “2” alleles of *cox1* within the 675 bp sequenced. All copies of allele 1 were identical regardless of the *atpA* and *cob* alleles with which each was paired. All copies of allele 2 were identical save for one individual from population 1 carrying the A2a three-locus genotype. In that individual the two sequences differed by a 1 bp synonymous substitution from the other copies of allele 2, thus representing a unique 2 allele at the sequence level. However, a second A2a individual, in this case sampled from population 8, carried the common allele 2 sequence.

The sequence data, concatenated across loci, can be used to apply the four gamete rule to all variable nucleotides, rather than just those at the restriction sites (Table 2). Twenty pairs of nucleotides follow this rule.

Values of pairwise linkage disequilibrium (D), and the standardized version of this statistic (D'), were calculated by pooling the entire North American dataset and are presented in Table 3(A–C). It can be seen that for cases that include the two most common alleles at that pair of loci (in bold type) absolute values of D' vary between 0.17 and 0.78 with an average of 0.50. That is, linkage disequilibrium ranges from modest to substantial but is never at its maximum, or absent. There are no large numerical differences in the averages of the four absolute values of D' associated with each pair of loci (*atpA/cox1* = 0.40, *atpA/cob* = 0.53, *cox1/cob* = 0.56) nor are there statistically significant differences between pairs of loci based on rank $|D'|$ values (Kruskal–Wallis $H = 2.47$, $df = 2$, $P = 0.29$).

Values of $|D'|$ calculated for pairings that included one or both of the rare alleles at each locus were, on average, higher (average $|D'| = 0.82$) but more variable (Table 3). Note that 34 of the 44 values calculated for the rare alleles had a $|D'| = 1$, but also that several of these values approached $|D'| = 0$. Given that the observed (average $|D'_{\text{rare}}| - \text{average } |D'_{\text{common}}| = 0.32$) it is useful to test the null hypothesis ($|D'_{\text{rare}}| - |D'_{\text{common}}| = 0$), or whether there is statistical support for the observation the absolute values of D' associated with rare alleles tend to be greater in

Table 2. Variable nucleotide sites in the concatenated *atpA-cox1-cob* sequences of the seven most common three-locus PCR/RFLP genotypes found in a sample of 239 *Silene vulgaris* individuals. The order of the three genes is arbitrary. The restriction sites diagnostic of the PCR/RFLP genotype designation are indicated (* = *AluI*, † = *MspI*, # = *ApoI*). Twenty pairs of sites obey the four haplotype rule, with the paired sites being from separate loci in each case.

A1a	C	A	A	T	A	A	A	C	A#
A1b	C	A	A	T	A	A	A	C	C
A2a	C	A	A	T	A	A	C† . . .	T	A#
A2b	C	A	A	T	A	A	C† . . .	T	C
B1a	A* . . .	C	G	G	C	G	A	C	A#
B1b	A* . . .	C	G	G	C	G	A	C	C
B2b	A* . . .	C	G	G	C	G	C† . . .	T	C
Site	245	263	278	320	419	575	705	708	1803
Locus	<i>atpA</i>						<i>cox1</i>		<i>cob</i>

magnitude. This was tested by computing average $|D'_{rare}|$ – average $|D'_{common}|$ individually for each of the three pairs of loci and then jackknifing these values across pairs to obtain an estimated standard error of $|D'_{rare}| - |D'_{common}|$ that would allow a test of the null hypothesis. The resulting jackknife mean and standard error ($(|D'_{rare}| - |D'_{common}| = 0.34 \pm 0.18)$ led to acceptance of the null hypothesis ($t = 1.89, df = 2, P = 0.20$).

Genotypic information and standardized linkage disequilibrium (D') between *atpA* and *cox1* for the European samples originally studied by McCauley et al. (2005) are presented in Table 4.

The D' values can be compared to those presented for the same allelic pairings in Table 3A. Note that the distribution of the *atpA* and *cox1* alleles into two-locus genotypes in the European sample was subject to the same type of G Tests of Independence by McCauley et al. (2005) as presented here for the North American sample, and found to be highly significant.

Among the 20 North American populations sampled, seven populations were jointly polymorphic for *atpA* and *cox1*, nine for *atpA* and *cob*, and seven for *cox1* and *cob*. Seven populations were jointly polymorphic for all three mitochondrial genes. An

Table 3. Linkage disequilibrium values and standardized linkage disequilibrium values (D/D') estimated for North American *Silene vulgaris* for all pairwise haplotype frequencies at paired mitochondrial loci (A) *atpA/cox1*, (B) *atpA/cob*, and (C) *cox1/cob*. Pairs that include one of the two most common haplotypes at each locus are in bold.

(A) <i>atpA/cox1</i>					
	A	B	C	D	E
1	0.07/0.34	-0.08/-0.56	0.04/1	-0.02/-1	-0.01/-1
2	-0.40/-0.21	0.11/0.50	-0.03/-1	-0.02/-1	-0.01/-1
3	0.01/0.17	0.00/0.07	-0.01/-1	-0.00/-1	-0.00/-1
4	-0.03/-1	-0.03/-1	-0.01/-1	0.04/1	0.03/1
(B) <i>atpA/cob</i>					
	A	B	C	D	E
A	-0.10/-0.55	0.07/0.51	0.02/1	0.01/1	-0.01/-0.34
B	0.09/0.55	-0.06/-0.52	-0.02/-1	-0.01/-1	0.01/0.38
C	0.01/1	-0.01/-1	-0.00/-1	-0.00/-1	-0.00/-1
D	0.00/0.02	0.00/0.18	-0.00/-1	-0.00/-1	-0.00/-1
(C) <i>cox1/cob</i>					
	1	2	3	4	
A	-0.03/-0.17	0.10/0.69	-0.02/-0.43	0.01/0.30	
B	0.10/0.58	-0.10/-0.78	0.00/0.04	-0.00/-0.20	
C	-0.01/-1	-0.01/-1	0.02/1	-0.00/-1	
D	-0.01/-1	0.01/1	-0.00/-1	-0.00/-1	

Table 4. Two-locus genotypes of 89 European *Silene vulgaris* individuals sampled by McCauley et al. (2005) and the standardized linkage disequilibrium between allelic pairs (D'). For each cell the number of individuals carrying that allelic combination is followed by D' . Note that *cox1* allelic nomenclature differs from McCauley et al. (2005) to correspond to that used here for North American samples, given the change in the *cox1* primers used to generate PCR/RFLP genotypes.

<i>cox1</i> allele	<i>atpA</i> allele	
	A	B
1	28 (0.37)	7 (-0.36)
2	24 (-0.43)	20 (0.44)
3	9 (0.69)	1 (-0.69)

average of 2.6 of the 17 three-locus genotypes was found per population. However, when only the common alleles are considered only four, three, and two populations were jointly polymorphic for the *atpA/cox1*, *atpA/cob* and *cox1/cob* pairs of loci, respectively (and only two of these polymorphic for all three loci). This high level of local population structure was evidenced by the large magnitude of F_{ST} (0.627) found by combining information across loci (locus-specific values of F_{ST} were 0.686, 0.663, and 0.603 for *atpA*, *cox1*, and *cob*, respectively). An F_{ST} value based on the distribution of the 17 three-locus genotypes equaled 0.545. The impact of this population structure on the spatial aspects of the linkage disequilibrium can be seen by calculating the product-moment correlation, r , across populations, of allele frequencies at pairs of loci, considering only the common alleles in each population, and then graphing r against values of D' obtained for that allelic pairing in the pooled data (Fig. 3). It can be seen that there is nearly a 1:1 correspondence between D' and r , suggesting that the linkage disequilibrium seen in the pooled sample is largely an among-population phenomenon owing to the low level of joint polymorphism within populations.

What is linkage disequilibrium within populations? Recall that linkage disequilibrium among the common alleles cannot be calculated for 15 of the sampled populations owing to lack of joint polymorphism of the common alleles. Values of D and D' were calculated for the common alleles for each population for which those values are defined and then averaged across the population-specific values. The four haplotype rule was met in two of the four populations jointly polymorphic for *atpA* and *cox1*, two of three populations jointly polymorphic for *atpA* and *cob* and in neither of the two populations jointly polymorphic for *cox1* and *cob*. The D' averages are presented in Table 5. It can be seen that the D' values are identical in sign but more variable in magnitude to the corresponding values for the pooled data presented in Table 3. Formal statistical tests of the null hypothesis $D = 0$ within spe-

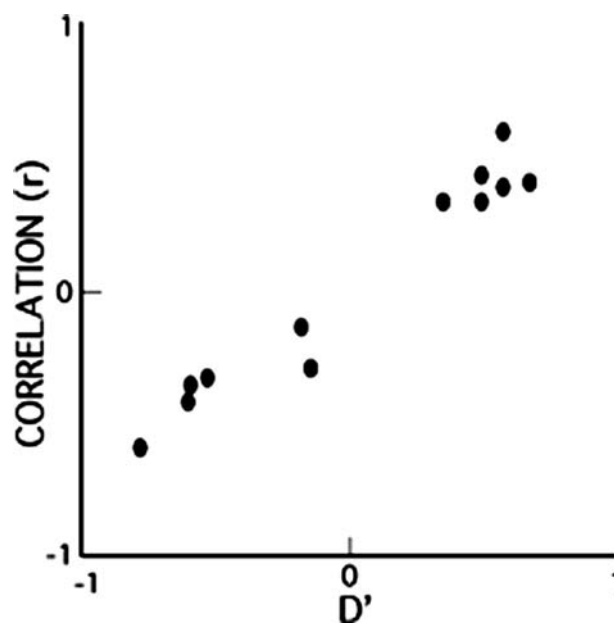


Figure 3. The relationship between the product-moment correlation (r) calculated for 12 combinations of population-specific allele frequencies and the corresponding standardized linkage disequilibrium (D') values calculated for each allelic pair without regard to population by pooling all North American individuals. Allelic pairing of *atpA/cox1*, *atpA/cob*, and *cox1/cob* considers only common alleles at each locus.

cific populations are not presented owing to the small numbers associated with individual population samples.

Discussion

Several aspects of the results indicate that intergenic recombination is an important component of the dynamics of the mitochondrial genome of *S. vulgaris*. First, when considering the two most common alleles at each of the three loci all four possible two-locus genotypic combinations were found for each pair of loci, the so-called four-gamete test (Hudson and Kaplan 1985). Although it is possible that the same restriction site could evolve twice via independent mutations at one or more of our marker genes, this seems unlikely, given the results of the sequencing study. At all three loci all sequences of a given PCR/RFLP “allele”

Table 5. Values of D' calculated within those populations jointly polymorphic for common alleles and then averaged across such populations.

	<i>atpA</i>		<i>atpA</i>			<i>cox1</i>			
	<i>cox1</i>	A	B	<i>cob</i>	A	B	<i>cob</i>	1	2
1		0.59	-0.61	a	-0.23	0.43	a	-1	1
2		-0.52	0.40	b	0.32	-0.24	b	1	-1

were identical (with one minor exception) regardless of which alleles at the other loci it was associated. Thus, convergent evolution of diagnostic restriction sites would require independent identical mutations in otherwise very highly conserved sequences. In fact, the conservative nature of these sequences can be seen at a larger phylogenetic scale by comparing the *atpA* and *cob* *S. vulgaris* sequences obtained here with *atpA* and *cob* sequences obtained for *S. latifolia* and *S. uniflora* by Houliston and Olson (2006) (GenBank accession numbers *S. latifolia*: DQ841770, *S. uniflora*: DQ841769). Those sequences are identical to those of our *atpA* "A" and *cob* "a" sequences, respectively. Although *S. vulgaris* and *S. uniflora* may hybridize under some circumstances and could be subject to introgression (Runyeon-Lager and Prentice 2000), *S. vulgaris* and *S. latifolia* are not known to hybridize. No *coxI* sequences were available from *S. latifolia* or *S. uniflora* for comparison. In addition, when rare PCR/RFLP alleles are also considered, five of the 156 remaining *atpA/coxI*, *atpA/cob* or *coxI/cob* two-locus combinations that are theoretically possible obey the four-gamete rule.

In the same vein, although the results of the G-tests give clear indication of linkage disequilibrium, standardized disequilibrium values (D') were not near the maximum in any calculation involving common pairs of alleles in the pooled North American dataset, again indicating some, but not pervasive, recombination. The observed values of D' near 1 or -1 are due to our failure to find 63 of the 80 possible three-locus genotype combinations, all but one that include the rare alleles at one or more loci. Failure to find such combinations could either indicate that they have not arisen by recombination, perhaps because some alleles are recent and of limited spatial distribution, or because, although they exist, they are too rare to have been encountered in our limited sample. It seems unlikely that all 63 were missed owing to sampling alone, given that in total they would be expected to comprise 19% of all individuals assuming three-locus linkage equilibrium.

Values of D' were comparable when compared across pairs of loci (i.e., *cob/coxI*, *cob/atpA*, *coxI/atpA*) indicating that intergenic recombination may not be restricted to certain mitochondrial genes. Because the mitochondrial genome of *S. vulgaris* has not been sequenced the physical linkage relationship among these genes is not known (if it is indeed stable), nor is its size, which could influence the probability of recombination. Such a whole mitochondrial genome sequence could prove invaluable in interpreting future studies of *S. vulgaris*, both because it would provide insight into the physical structure of the genome and because it could determine whether there are paralogous copies of the genes under study within the mitochondrial genome whose existence could complicate the study of recombination. One could assume that the structure of the genome resembles those angiosperms whose mitochondrial genomes have been sequenced, in that it is likely to include a number of noncoding repeats that facilitate re-

combination, and consist of a multipartite structure that consists of several subgenomic subcircles (Palmer and Shields 1985; Kuobo and Mikami 2007).

It is interesting to note that at least for the common alleles the direction and magnitude of linkage disequilibrium between *atpA* and *coxI* seen in this sample from the eastern U.S. are similar to those seen in a much smaller sample consisting of individuals collected from numerous localities within the native range of *S. vulgaris* in Europe (McCauley et al. 2005). Comparisons of genetic diversity and genetic structure between the introduced and native portions of species' ranges are fairly common, and previous studies have shown that introduction of *S. vulgaris* to North America from Europe has not resulted in a meaningful reduction in cpDNA diversity (McCauley et al. 2003; Taylor and Keller 2007). Similarity in patterns of linkage disequilibrium between native and introduced ranges has not been widely considered in studies of the population biology of invasive species, yet the results presented here suggest that the recombinatorial patterns seen in the *S. vulgaris* mitochondrial genome may have been imported from Europe during the establishment of the species in North America.

The high level of population structure seen in the three loci, evidenced by the combined F_{ST} value of 0.627, is very similar to the value of 0.686 seen in a recent study of cpDNA PCR/RFLP diversity in this species (McCauley et al. 2003). This is not surprising given that cpDNA and mtDNA would be expected to be largely cotransmitted maternally and therefore disperse largely by seed (Olson and McCauley 2000), and given that many of the same populations were included in both studies. The F_{ST} value obtained here by collecting samples from across a significant portion of the eastern U.S. was also comparable in magnitude to F_{ST} values computed from both mtDNA and cpDNA markers in collections restricted to populations found in southwest Virginia (McCauley 1998; Olson and McCauley 2002; McCauley et al. 2003). The significance of this population structure for the generation of novel intergenic recombinants in the *S. vulgaris* mitochondrial genome is that it could limit opportunities for any paternal leakage of mitochondrial genomes to create the polymorphic collections of mitochondrial genomes within cells that are necessary for recombination to generate novel multilocus combinations. This intracellular polymorphism is analogous to the heterozygosity required for chromosomal recombination to generate novel nuclear genotypes. Similarly, the fairly common self-fertilization by hermaphrodites known to occur in this species (Peterson 1992; Emery 2001) would limit opportunities for mixing mitochondrial genomes. The observed high level of population structure also illustrates the necessity of sampling broadly among as many populations as possible, rather than intensively among fewer, when trying to assess organellar diversity in plants.

An F_{ST} value obtained by incorporating the frequencies of the three-locus genotypes into the calculations would be greater than

that obtained using haplotype frequencies at each locus separately when linkage disequilibrium varies in direction from population to population, because under those circumstances multilocus genotypic frequencies could vary among populations more than individual single locus haplotype frequencies. That was not the case here, because the F_{ST} value based on the three-locus combinations was, in fact, less than the single locus values. This indicates that any linkage disequilibrium within populations is consistent in direction from population to population, as indicated by the averages of the few D' values that could be calculated within populations. The fact that the three-locus measure was lower than any single locus measure could be a reflection of the general observation that values of F_{ST} and G_{ST} tend to decline with increased levels of the genetic diversity of the markers used to compute them (Hedrick 2005b).

Finally, the implications of the results presented here for the evolution of gynodioecy should be discussed. It appears that gynodioecious species can often harbor multiple forms of CMS that could result from mutations at several mitochondrial genes (Budar et al. 2003). If so, intergenic recombination could create novel associations of CMS types within individuals that would require novel combinations of nuclear restorers for male fertility. Alternatively, recombination among lineages could also regenerate fertile genotypes without back mutation, effectively an escape from Muller's Ratchet.

It is also interesting to consider the possibility of statistical associations between unknown CMS genes and mitochondrial genes used as markers in population studies. Olson and McCauley (2002) and McCauley and Olson (2003) have found statistical associations between the mtDNA haplotype of *S. vulgaris* individuals and gender (i.e., female or hermaphrodite). This was ascribed to linkage disequilibrium between the regions of the mitochondrial genome used as markers in these studies and unknown CMS factors that could vary in their tendency to produce the female phenotype. Associations between mitochondrial markers and gender have been seen in other gynodioecious plant species as well (Klaas and Olson 2006 and references therein). The moderate level of linkage disequilibrium found here between the three mitochondrial genes studied suggests that the expectation of some linkage disequilibrium between mitochondrial markers and CMS factors seems reasonable, but that such associations might not be expected to be perfect.

In summary, it appears that intergenic recombination within the mitochondrial genome of *S. vulgaris* is sufficiently common to be detected by a population genetic analysis, but not so common as to result in the linkage equilibrium that is theoretically possible in recombining systems. What limits the approach to linkage equilibrium? In order for a novel multilocus recombinant to enter a population several criteria must be met. First, the physical act of intermolecular homologous recombination must occur. Second,

the sequences of the recombining molecules must differ. Third, the product of that recombination must achieve an intracellular and intraindividual frequency sufficient to ensure transmission to the next generation. Intermolecular mtDNA sequence differences within cells require heteroplasmy, most likely resulting from paternal leakage from a pollen source carrying a different mtDNA genotype. Although there is evidence of mitochondrial paternal leakage and heteroplasmy in *S. vulgaris* (McCauley et al. 2005; Houliston and Olson 2006; Welch et al. 2006), at this point it is unclear how common this might be. Whatever the frequency of leakage, if most mating is between individuals cooccurring in the same local population then the results presented here suggest that population structure would greatly reduce the probability that any leakage would produce the heterogeneous mixture of mitochondrial genotypes required for intergenic recombination to produce novel multilocus genotypes. Once any such recombinants are formed they may increase in frequency within cells and then within individuals by chance events associated with mitochondrial fission and cell division, or by intracellular and intraindividual selection if they confer some sort of advantage (Rand 2001). It must be that many recombinants are not transmitted to the next generation and have little impact on the population genetics of the species. It is not yet clear which factor, or combination of factors, listed above limits the rate at which the mitochondrial genome of *S. vulgaris* approaches linkage equilibrium.

ACKNOWLEDGMENTS

The laboratory and/or intellectual contributions of M. Welch, S. Pearl, A. Sundby, M. Olson, M. Wade, D. Taylor, D. Funk, and S. Keller are much appreciated. Financial support was provided by National Science Foundation award 0621867.

LITERATURE CITED

- Arimura, S., J. Yamamoto, G. P. Aida, M. Nakazono, and N. Tsutsumi. 2004. Frequent fusion and fission of plant mitochondria with unequal nucleotide distribution. *Proc. Natl. Acad. Sci. USA* 101:7805–7808.
- Barkman, T. J., G. Chenery, J. R. McNeal, J. Lyons-Weiler, W. J. Elisens, G. Moore, A. D. Wolfe, and C. W. dePamphilis. 2000. Independent and combined analysis of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. USA* 97:13166–13171.
- Barr, C. M., S. R. Keller, P. K. Ingvarsson, D. B. Sloan, and D. R. Taylor. 2007. Variation in mutation rate and polymorphism among mitochondrial genes of *Silene vulgaris*. *Mol. Biol. Evol.* 24:1783–1791.
- Budar, F., P. Touzet, and R. DePaepe. 2003. The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* 117:3–16.
- Burt, A., and R. Trivers. 2006. *Genes in conflict*. Harvard Univ. Press, Cambridge, MA.
- Charlesworth, D., and V. Laporte. 1998. The male-sterility polymorphism of *Silene vulgaris*: analysis of genetic data from two populations and comparison with *Thymus vulgaris*. *Genetics* 150:1267–1282.
- Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bull.* 19:11–15.
- Emery, S. N. 2001. *Inbreeding Depression and Its Consequences in Silene vulgaris*. Master's Dissertation. Vanderbilt Univ., Nashville, TN. 74 pp.

- Frank, S. A. 1989. The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* 133:345–376.
- Hedrick, P. W. 2005a. *Genetics of populations*. Jones and Bartlett, Sunbury, MA. 737 pp.
- . 2005b. A standardized genetic differentiation measure. *Evolution* 59:1633–1638.
- Houliston, G. J., and M. S. Olson. 2006. Independent and non-neutral evolution of organellar genes in *Silene vulgaris*. *Genetics* 174:1985–1994.
- Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.
- Klaas, A. L., and M. S. Olson. 2006. Spatial distributions of cytoplasmic types and sex expression in Alaskan populations of *Silene acaulis*. *Int. J. Plant Sci.* 167:179–189.
- Kuobo, T., and T. Mikami. 2007. Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum* 129:6–13.
- Logan, D. C. 2003. Mitochondrial dynamics. *New Phytol.* 160:463–473.
- . 2006. Plant mitochondrial dynamics. *Biochimica et Biophysica Acta* 1763:430–441.
- Lynch, M., and J. L. Blanchard. 1998. Deleterious mutation accumulation in organelle genomes. *Genetica* 102/103:29–39.
- McCauley, D. E. 1998. The genetic structure of a gynodioecious plant: nuclear and cytoplasmic genes. *Evolution* 52:255–260.
- McCauley, D. E., and M. S. Olson. 2003. Associations among cytoplasmic molecular markers, gender, and components of fitness in *Silene vulgaris*, a gynodioecious plant. *Mol. Ecol.* 12:777–787.
- McCauley, D. E., R. A. Smith, J. D. Lisenby, and C. Hsieh. 2003. The hierarchical spatial distribution of chloroplast DNA polymorphism across the introduced range of *Silene vulgaris*. *Mol. Ecol.* 12:3227–3235.
- McCauley, D. E., M. F. Bailey, N. A. Sherman, and M. Z. Darnell. 2005. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* 95:50–58.
- Mackenzie, S., and L. McIntosh. 1999. Higher plant mitochondria. *Plant Cell* 11:571–585.
- Olson, M. S., and D. E. McCauley. 2000. Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proc. R. Soc. Lond. B* 267:1801–1808.
- . 2002. Mitochondrial DNA diversity, population structure, and gender association in the gynodioecious plant *Silene vulgaris*. *Evolution* 56:253–262.
- Palmer, J. D., and L. A. Hebron. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* 28:87–97.
- Palmer, J. D., and C. Shields. 1985. Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* 307:437–440.
- Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvinin, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.* 14:689–701.
- Peterson, M. W. 1992. Advantages of being a specialist female in gynodioecious *Silene vulgaris* S. L. (Caryophyllaceae). *Am. J. Bot.* 79:1389–1395.
- Posada, D., K. A. Crandall, and E. C. Holmes. 2002. Recombination in evolutionary genomics. *Ann. Rev. Genetics* 36:75–97.
- Rand, D. M. 2001. The units of selection on mitochondrial DNA. *Ann. Rev. Ecol. Syst.* 32:415–448.
- Runyeon-Lager, H., and H. C. Prentice. 2000. The morphometric variation in a hybrid zone between the weed *Silene vulgaris* and the endemic *Silene uniflora ssp. petraea* (Caryophyllaceae), on the Baltic island of Oland. *Can. J. Bot.* 78:1384–1397.
- Schneider, S., D. Roessli and L. Excoffier. 2000. *Arelequin version 2.000*: A software for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology, Univ. of Geneva, Geneva, Switzerland.
- Sheahan, M.G., D. W. McCurdy, and R. J. Rose. 2005. Mitochondria as a connected population: ensuring continuity of the mitochondrial genome during plant cell differentiation through massive mitochondrial fusion. *Plant J.* 44:744–755.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry*, 3rd ed. W.H. Freeman, New York, NY.
- Städler, T., and L. F. Delph. 2002. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proc. Natl. Acad. Sci. USA* 99:11730–11735.
- Taylor, D. R. and S. R. Keller. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61:334–345.
- Taylor, D. R., M. S. Olson, and D. E. McCauley. 2001. A quantitative genetic analysis of nuclear-cytoplasmic male sterility in structured populations of *Silene vulgaris*. *Genetics* 158:833–841.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Welch, M. E., M. Z. Darnell, and D. E. McCauley. 2006. Variable populations within variable populations: quantifying mitochondrial heteroplasmy in natural populations of the gynodioecious plant *Silene vulgaris*. *Genetics* 174:829–837.

Associate Editor: S. Otto