Non-neutral Evolution of Organelle Genes in *Silene vulgaris*

Gary J. Houliston and Matthew S. Olson*

*Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska 99775*

Sequence data from this article have been deposited with the EMBL/Genbank libraries under accession nos. DQ841746-DQ841789.
Organelle genes in *Silene vulgaris*

mitochondria, chloroplast, recombination, cytoplasmic male sterility, gynodioecy

*Corresponding author:* Matthew S. Olson. Institute of Arctic Biology, University of Alaska
Fairbanks, 311 Irving 1, 902 N. Koyukuk Dr, Fairbanks, Alaska, 99775
Email: ffmso@uaf.edu Phone +1 907 474 2766 Fax +1 907 474 6967
ABSTRACT

Knowledge of mitochondrial gene evolution in angiosperms has taken a dramatic shift within the past decade, from universal slow rates of nucleotide change to a growing realization of high variation in rates among lineages. Additionally, evidence of paternal inheritance of plant mitochondria and recombination among mitochondrial genomes within heteroplasmic individuals has led to speculation about the potential for independent evolution of organellar genes. We report intraspecific mitochondrial and chloroplast sequence variation in a cosmopolitan sample of 42 *Silene vulgaris* individuals. There was remarkably high variation in two mitochondrial genes (*atp1, atp9*) and additional variation within a third gene (*cob*). Tests for patterns of non-neutral evolution were significant for *atp1* and *atp9*, indicative of the maintenance of balanced polymorphisms. Two chloroplast genes (*matK, ndhF*) possessed less, but still high, variation and no divergence from neutral expectations. Phylogenetic patterns of organelle genes in both the chloroplast and mitochondria were incongruent, indicating the potential for independent evolutionary trajectories. Evidence indicated reassociation among cytoplasmic genomes and recombination between mitochondrial genes and within *atp1*, implying transient heteroplasmy in ancestral lineages. Although the mechanisms for long-term maintenance of mitochondrial polymorphism are currently unknown, frequency-dependent selection on linked cytoplasmic male sterility genes is a potential candidate.
Until recently (Städler and Delph 2002; Cho et al. 2004, Parkinson et al. 2005), the coding regions of the organelle genomes of angiosperms were viewed as highly constrained, with low rates of mutation and strict maternal inheritance (Wolfe et al. 1987). This has resulted in the assumption that intraspecific variation in organelle genes should be low or even absent. Contradictory patterns of extremely high mtDNA variation have recently been observed in Pelargonium (Parkinson et al. 2005), Plantago (Cho et al. 2004) and Silene (Städler and Delph 2002). Two of these genera (Plantago and Silene) are model systems for studies of the evolution of gynodioecy in plants (De Haan et al. 1997; Charlesworth and Laporte 1998). The elevated rate of synonymous substitutions in Pelargonium has been attributed to an increase in mutation rate or a reduction in the efficacy of DNA repair mechanisms (Parkinson et al. 2005). In Plantago, high diversity at the mitochondrial cytochrome oxidase I (cox1), adenosine 5′-triphosphate (atp1), and small subunit RNA-encoding DNA (SSU rDNA) loci has been attributed to similar causes (Cho et al. 2004). Silene acaulis (L.) Jacq, possessed high diversity of mitochondrial apocytochrome b (cob) haplotypes, but this was attributed to ancient coalescence due to long term maintenance of a balanced polymorphism (Städler and Delph 2002).

If selection maintains a balanced polymorphism in mitochondrial types for long periods, one candidate for the target of selection is cytoplasmic male sterility (CMS) genes that influence sex expression. Sex expression in many gynodioecious species (those with females and hermaphrodites in natural populations) is controlled by a complex joint cytonuclear polymorphism between CMS genes and their associated restoration of male fertility (Rf) genes (Saumitou-Laprade et al. 1994; Schnable and Wise 1998). Theoretical studies predict that through time CMS and Rf frequency cycles are generated from negative frequency dependent selection assuming both female compensation (females produce more seeds than hermaphrodites;
DARWIN 1877) and a cost to carrying $R_f$ genes (cost of restoration, DELANNEY et al. 1981, CHARLESWORTH 1981). CMS/$R_f$ cycles are stabilized by specific relationships between the fitnesses of particular CMS types, the costs of carrying $R_f$ genes (FRANK 1989; GOUYON et al. 1991), and by low levels of paternal leakage of CMS genes (WADE and McCauley 2005), potentially providing mechanisms whereby the joint CMS/$R_f$ polymorphism is maintained over long periods. CMS genes are ubiquitous in plants and have utility within breeding programs of nearly every major crop species. For reasons that are yet to be discovered, the male-sterile phenotype associated with the expression of CMS is uncommon in natural populations of the majority of plant species, with gynodioecy found in only 7% of angiosperm taxa (RICHARDS 1997).

Even if some form of balancing selection maintains CMS/$R_f$ polymorphism on an ecological time scale, there remains considerable contention over whether polymorphism in CMS genes is maintained over longer time periods (FRANK 1989; GOUYON et al. 1991) or whether newly arisen types more likely spread as epidemics within gynodioecious populations (INGVARSSON and TAYLOR 2002). For instance, CMS and their associated $R_f$ genes may spread to fixation in the absence of costs of restoration (CHARLESWORTH 1981; FRANK 1989; GOUYON et al. 1991), or demographic effects and drift may cause the loss/fixation of a particular CMS type when CMS genes are at a low/high frequency during a cycle. Both of these outcomes are consistent with an epidemic model.

A comparative molecular population genetic approach may show promise for discriminating among the three alternatives proposed to shape mitochondrial nucleotide diversity. For instance, elevated mutation rates would be expected to increase diversity within lineages in all genes in the genome, whereas selection may target only specific genes or genomic regions. Furthermore, elevated mutation rates are not expected to influence either the relative
lengths of internal and external branches or the expected numbers of haplotypes, whereas selection will (Fu and Li 1993; Depaulis and Veuille 1998). An example of the influence of long term balancing selection is found in plant self-incompatibility alleles wherein negative frequency dependent selection may have preserved variants for 27-50 million years, leading to trans-specific polymorphism (IOERGER et al. 1990; UyenoYama 1995). Finally, epidemics generate short internal branches in gene genealogies as all alleles coalesce to the allele responsible for the epidemic (Hatcher 2000; Ingvarsson and Taylor 2002). If CMS genes were either maintained over long periods or were characterized by spreading as epidemics, a naive prediction might be for similar patterns of long or short time to coalescence, respectively, for CMS alleles. However, studies of species wherein multiple CMS genes have been identified have shown that CMS variants are not allelic, but rather each is constructed from different sets of genes to create a novel chimaeric open reading frame (Schnable and Wise 1998). Nonetheless, some researchers have posited that since the mitochondrial genome is inherited as a single linkage unit and the chloroplast and mitochondria are co-inherited, these genomes may harbour signals of past balancing selection or epidemics via hitchhiking (Olson and McCauley 2000; Ingvarsson and Taylor 2002; Cho et al. 2004).

Here we assess patterns of nucleotide diversity in a cosmopolitan sample of Silene vulgaris (Monech.) Garcke individuals, a well studied gynodioecious plant species (Marsden-Jones and Turrell 1957; Pettersson 1992; Olson et al. 2005). We estimate diversity in three mitochondrial and two chloroplast genes to determine whether genome-wide hitchhiking can indeed influence genetic diversity throughout the mitochondrial genome and potentially extends to the chloroplast genome which is often co-inherited. The presence of intraspecific diversity in the mitochondrial and chloroplast genomes also offers an opportunity to explore the transmission dynamics and role of recombination on the cytoplasmic genomic compartments. With rare
exceptions (REBOUD and ZEYL 1994), the mitochondria and chloroplast in angiosperms are thought to be primarily maternally inherited (BIRKY 2001; McCauley et al. 2005). Phylogenetic patterns among co-inherited genes or genomes are expected to be congruent, due to a process akin to linkage disequilibrium (Olson and McCauley 2000). The absence of such congruence implies independent transmission histories perhaps resulting from horizontal transfer (Bergthorsson et al. 2003) or paternal leakage (McCauley et al. 2005), producing heteroplasmic individuals. Over generations, heteroplasmy is expected to be reduced through random genetic drift (Albert et al. 1996; Roze et al. 2005). As heteroplasmy is a necessary precondition for recombination among divergent genes or genomes, the presence of recombination within or among cytoplasmic genomes not only indicates that cytoplasmic genes may have the potential to evolve as independent units, but also indicates that transmission dynamics may be much less constrained than previous studies have indicated. Silene vulgaris is an excellent candidate for better understanding the mitochondrial and chloroplast evolution as previous studies have identified high levels of mitochondrial polymorphism within small populations (Olson and McCauley 2000; Štorchová and Olson 2004), and small samples have indicated that sequence diversity at coding regions may also be elevated relative to other taxa (McCauley et al. 2005).

**MATERIALS AND METHODS**

Silene vulgaris (Moench.) Garcke, (Caryophyllaceae) seeds were collected from four regions (Virginia, USA; New York/Vermont, USA; Central Europe, Czech Republic/Germany; Siberia, Russia; Figure 1) and grown in the Institute of Arctic Biology greenhouse at the University of Alaska Fairbanks. Eight populations were located in Giles and Craig counties, Virginia USA (628, GR, CO, AJ, MV, AR, WR, EB, Figure 1; listed as populations 3, 4, 5, 6, 11,
12, 14, 15, respectively, in Olson and McCauley 2002), and four populations were located in New York (NY) and Vermont USA (Vt17b, Vt100, VtLnG, Figure 1; McCauley et al. 2003). Of the Central European populations, three were located near Prague, Czech Republic (Kovary Meadow: KM; Institute Garden: InG, Sedlac railway station: SD), two were in northern Moravia, Czech Republic (Mount Keprnik: MK4; Rychleby Mountains: RM), two were from southern Moravia (Klentnice: KL; Lednice: Ld), and the eighth population was located near Gemsteltal (GM) in the Allgauer Alps, Austria (Figure 1; Štorchová and Olson 2004). One population was collected from Krasnojarsk (Krn) in south central Russia near Lake Baikal (Figure 1). One plant was arbitrarily selected for study from each of 1-5 maternal families from each of these 21 populations of *S. vulgaris*. Two additional Silene species were included as potential outgroups and to determine the extent of interspecific divergence. Seeds from *Silene uniflora* Roth (also called *S. maritima*), a species closely related to and not entirely reproductively isolated from *S. vulgaris* (Marsden-Jones and Turrill 1957, Runyeon-Lager and Prentice 2000) originated in Saint Catherine’s Bay, UK. Seeds from *Silene latifolia* Poir. were collected in the courtyard at the Abbey of St. Thomas, Brno, Czech Republic; *S. vulgaris* and *S. latifolia* are not interfertile.

Total genomic DNA was isolated and purified from leaf tissue using a Qiagen DNeasy Plant mini kit (Qiagen Inc). We PCR-amplified three mitochondrial (*adenosine 5'-triphosphate subunit 1 – atp1*, *adenosine 5'-triphosphate subunit 9 - atp9, apocytochrome b - cob*) and two chloroplast genes (*maturase K - matK, NADH dehydrogenase F - ndhF*) from each individual. PCRs were performed in 35ul reactions, with 15-40ng of genomic DNA template and a 50°C annealing temperature using standard protocols (Štorchová and Olson 2004). Sequencing reactions were carried out using the ABI Big-Dye terminator sequencing kit, and sequenced directly on an ABI 3100 automated sequencer. Sequences had no indications of heterozygosity.
(e.g. overlapping chromatograms), suggesting that if heteroplasmy was present, we were sequencing the most common allele. PCR primer sequences and aligned datasets are available upon request.

As ancient gene transfers from the mitochondrial to nuclear genome or nuclear pseudogenes are possible in plants (Mower et al. 2004), we confirmed maternal inheritance of the \textit{atp}1 and \textit{atp}9 regions to assure that they were indeed of cytoplasmic origin. For \textit{atp}1, a reciprocal cross was conducted between two plants with types C and F (see Table 1) polymorphic for SmaI restriction sites in \textit{atp}1. \textit{Atp}1 was amplified from 16 progeny from each mother (32 total) and allelic inheritance was determined after separation of digested PCR products on a 3\% agarose gel stained with ethidium bromide. All progeny possessed the same haplotype as their mother. Under a null hypothesis of biparental inheritance as expected for nuclear copies, the probability of finding only the single haplotype in the progeny from each maternal parent was 0.00002, with a power of 99\%. For \textit{atp}9, a reciprocal cross was conducted between 62812 and a plant with type C (see Table 1) which differed in a MnvI restriction site. Ten progeny from the maternal parent with type C and 7 progeny from 62812 as a maternal parent were genotyped. \textit{Atp}9 restriction polymorphisms exhibited strict maternal inheritance; the probability of all 17 progeny inheriting the maternal haplotype under a null hypothesis of biparental inheritance and nuclear localization was 0.001, with a power of 91\%. As it is known that low levels of biparental inheritance occur in \textit{S. vulgaris}, one might also be interested in the probability of detecting biparental inheritance given our sample sizes. Using the 95\% confidence interval of the zero term of the Poisson distribution to estimate the minimum detectable frequency of biparental inheritance, a total sample size of 49 (combined across \textit{atp}1 and \textit{atp}9) allows detection of not less than 6.08\% (2.98/49) biparental progeny. In addition to verifying the maternal inheritance of these subunits, Southern blots with \textit{atp}1-specific probes hybridize
strongly, indicating that \textit{atp1} is present in high copy number as is expected for genes located in the mitochondria. \textit{Atp1}, \textit{atp9} and \textit{cob}, like most other respiratory genes, have never been found to be lost in the mitochondrial genome and transferred to the nuclear genome in plants, making this very unlikely in this taxon (ADAMS et al. 2002). Finally, a phylogeny of the \textit{atp1} dataset from \textit{S. vulgaris} and a much larger study (BERGTHORSSON et al. 2003) were combined to rule out the possibility of horizontal transfer via parasitic plants. In this analysis, \textit{atp1} haplotypes from \textit{Silene vulgaris} formed their own clade with high bootstrap support and were not associated with any other genera; thus, there is no evidence for horizontal transfer of organelles or genes as has been shown in other groups with high organelle genomic variation (MOWER et al. 2004).

We re-amplified and resequenced a subset of samples (SD6, NY006, Ld5, EB3AG, 62812, NY007, Vt17b, Gm7) to reconfirm the patterns of diversity and recombination within and between \textit{atp1} and \textit{atp9}. Although we feel it is unlikely that cross sample contamination could have produced the results reported here, it cannot be discounted with complete certainty.

**Statistical Analyses:** Sequences lacked indels, easing manual alignment. The following summary statistics were calculated using DNAsp version 4.0 (ROZAS et al. 2003): \( \theta \), per site estimated from the total number of mutations (WATTERSON 1975); \( \theta_s \), per site, including only synonymous nucleotide differences: \( \pi \), diversity, the average number of nucleotide differences per site between a pair of randomly chosen sequences (NEI 1987); \( K_s \), average number of pairwise synonymous substitutions per synonymous site; \( K_a \), average number of pairwise non-synonymous substitutions per non-synonymous site; Tajima’s D, (TAJIMA 1989), based on the differences between the number of segregating sites (\( \theta \)) and the average number of nucleotide differences per sequence (\( \pi \)); Fu and Li’s D with an outgroup [\textit{Nicotiana tabacum}] (Fu and Li 1993), the test statistic is based on the differences between the total number of mutations in the
external branches of the genealogy and the total number of mutations; K-test, (DEPAULIS and VEUILLE 1998) the observed number of haplotypes versus the range of the expected number of haplotypes (calculated using 1000 neutral coalescent simulations); and the four gamete test (for recombination) (HUDSON and KAPLAN 1985). Recombination was not included in the confidence interval simulations for the K-test to generate a conservative test for the minimum number of haplotypes. Specific hierarchical hypotheses concerning differences between chloroplast and mitochondrial genes and differences among mitochondrial genes were constructed and evaluated using log likelihood tests as implemented in MLHKA (WRIGHT and CHARLESWORTH 2004).

Nicotiana tabacum was chosen as an outgroup because its shared ancestor with S. vulgaris has little chance of also sharing Silene mitochondrial haplotypes; also it was the closest relative of Silene for which all five genes were available on Genbank. It has been demonstrated that the presence of population subdivision can lead to overestimates of deviations from neutrality using the HKA test (INGVARSSON 2004). Population subdivision within our study was minor; the same alleles were present in different sample regions. Additionally, all loci are organellar, and no comparisons are made with nuclear data which are under a different population dynamic (cf INGVARSSON and TAYLOR 2002).

Phylogenetic hypotheses were constructed using PAUP* with an HKY model of substitution (SWOFFORD 2003). The best fit substitution model was chosen using Modeltest (POSADA and CRANDALL 1998). Partition homogeneity tests, to assess the null hypothesis of congruence between the phylogenies of each gene, were carried out using PAUP* (SWOFFORD 2003). RDP version 2.0 was used to detect potential recombination, using the default settings (MARTIN and RYBICKI 2000). RDP implements both phylogenetic and substitution methods to test for recombination. Phylogenetic methods (RDP, Bootscan) compare phylogenies constructed from different parts of the genomes – if the phylogenies have different topologies then it is
assumed recombination has occurred. Substitution methods (GENECONV, Chimaera, Max-Chi, SiScan) look for clustering within sites with substitutions or whether substitutions occur within the window of analysis more frequently than a calculated statistical distribution. In addition, the four-gamete test (HUDSON and KAPLAN 1985) was applied to the data. This test relies on the infinite sites model and deems that where we find a combination of all four possible combinations of a pair of variable sites the most parsimonious explanation is recombination. It also is possible that such a pattern can be generated by recurrent mutations (homoplasy) at an individual site.

RESULTS

Patterns of inter and intra specific diversity and divergence: Considerable diversity was found within the three mitochondrial and two chloroplast genes in *Silene vulgaris* (Tables 1 & 2), with remarkably high levels of mitochondrial intraspecific variation for angiosperms (Table 2). The mean pairwise numbers of synonymous substitutions per total synonymous sites (Ks) for the eight *atp9* haplotypes found in *S. vulgaris* were similar to those found in comparisons of *atp9* haplotypes from rice and maize (Table 3), two species that are believed to have diverged 50 Mya (Gaut 1998). Ks for the seven *atp1* haplotypes in *S. vulgaris* also was high, approximately half the value found in comparisons of *atp1* haplotypes from rice and maize (Table 3). Ks for the six apocytochrome b (*cob*) haplotypes was not as high as for *atp1* and *atp9* and was an order of magnitude lower than that found both between rice and maize and the strikingly high value found among *S. acaulis* haplotypes (Table 3).

Plant mitochondrial and chloroplast transcripts are known to undergo post-transcriptional C to U editing at non-synonymous sites (GRAY 1996). Such editing may result in C-T DNA
polymorphism not being reflected as a polymorphism in the mRNA; thus although the site is predicted to be non-synonymous from the DNA, with editing the mutation will not alter the amino acid. In our dataset, four non-synonymous variable sites are potentially influenced by C-U RNA editing: site 21 in \textit{atp9}, site 796 in \textit{cob}, and sites 19 and 140 in \textit{matK}. Many of the analyses presented herein do not utilize information on the coding state of the site; however, we have presented adjusted Ks/ Ka values for these three genes by scoring these sites as synonymous rather than non-synonymous substitutions, as well as the non-adjusted values (Table 3).

For both \textit{atp1} and \textit{atp9}, one \textit{S. vulgaris} haplotype (represented by Vt17b and Ld5, Table 1) was identical to the haplotypes found in \textit{S. latifolia} and \textit{S. uniflora}, whereas other haplotypes were extremely diverged, either for \textit{atp1}, \textit{atp9}, or both (Figure 2, Table 1). The majority of mitochondrial variation occurred at synonymous codon positions. Four of the 25 variable sites in \textit{atp1} resulted in amino acid changes, with Ka/Ks = 0.0590 (ROZAS \textit{et al.} 2003 following NEI and GOJOBORI 1986). For \textit{atp9}, 4 of the 21 variable sites in \textit{S. vulgaris} resulted in amino acid changes, with Ka/Ks = 0.0791. Three of the four variable sites in \textit{cob} resulted in amino acid changes (Ka/Ks = 1.2222). Ka/Ks values presented for \textit{atp9} and \textit{cob} are unadjusted for potential RNA editing; adjusted values can be seen in Table 3.

Multivariate HKA tests indicated that intraspecific and interspecific divergence were not correlated among all 5 genes. Specifically, intraspecific diversity of mitochondrial genes was significantly greater than that of chloroplast genes (Likelihood Ratio Test = 24.22, df = 2, P < 0.0001; using \textit{Nicotiana tabacum} as an outgroup). Moreover, among mitochondrial genes intraspecific diversity for \textit{atp1} and \textit{atp9} was greater than that of \textit{cob} (Likelihood Ratio Test = 9.655, df = 1, P < 0.002).
The seven *atp*1 haplotypes observed were significantly fewer than expected assuming 25 segregating sites as determined by neutral coalescent simulations of the expected distribution of haplotypes (K test, Table 2). For *atp*9, haplotype numbers were within the 95% confidence intervals for the expected distribution given the observed numbers of segregating sites, whereas in *cob* there were more haplotypes than expected (Table 2). For these tests we assumed no recombination which leads to conservative estimates of the expected number of haplotypes. If recombination is allowed, then the number of haplotypes present in *cob* is not significantly different than neutral expectations (expected number haplotypes 4-7). Because recombination was detected in *cob* using the four gamete test (see below), we suspect that the elevated number of haplotypes results from recombination. Fu and Li’s D statistic detected a significant departure from neutral expectations for *atp*1 and *atp*9 (Table 2) as can be seen via relatively long internal branches for these genes (Figure 2). No other tests of non-neutrality using Tajima’s D and Fu and Li’s D were significant (Table 2).

Interspecific divergence in *mat*K and *ndh*F was substantial between *Silene vulgaris* and *S. latifolia* (Table 1). Intraspecific variation in synonymous sites among the distinct haplotypes of these two *S. vulgaris* chloroplast genes were both two orders of magnitude lower than that found between rice and maize (Table 3). In *mat*K, there was a bias toward non-synonymous substitutions, with unadjusted Ka/Ks = 2.9500. Two of the four variable sites in *ndh*F were non-synonymous, with Ka/Ks = 0.3030. No tests of non-neutrality (K-test, Tajima’s D, Fu and Li’s D) for either chloroplast gene were significant.

**Reassociation and Recombination:** Evidence for reassociation between genomes and recombination within genomes was detected at three levels of organization: between the chloroplast and mitochondrial genomes, between genes within the mitochondrial genome, and
between sites within genes in the mitochondria. Strict co-inheritance of genomes is expected to generate congruent phylogenetic histories between genes in different genomes, whereas the absence of congruence implies at least somewhat independent histories (Olson and McCauley 2000, Desplanque et al. 2000). Pairwise partition-homogeneity tests (Farris et al. 1994) indicated incongruence between the pairwise histories of matK and atp1 (P = 0.003), matK and atp9 (P = 0.001), and atp1 and atp9 (P = 0.001). No additional pairwise comparisons detected incongruence using a Bonferroni-adjusted critical value of 0.005 (α = 0.05/10 comparisons), although cob and ndhF phylogenies did not contain sufficient structure to allow for a powerful test.

Recombination between mitochondrial genes can be inferred from some particularly striking patterns wherein all four possible combinations of genes were observed in different individuals (Table 1). Examination of Table 1 reveals that the most common combined atp1-atp9 haplotype in our sample (represented by SD6) is differentiated from the atp1-atp9 haplotype carried by NY006 and NY007 in 12 changes localized in atp9. Ld5 and Vt17b possess the atp9 haplotype of NY006 and NY007, but differ from SD6, NY006 and NY007 by 8 changes in atp1. Finally, individuals such as EB3AG possess the atp1 haplotype of Ld5 and Vt17b, but have the atp9 type of SD6. Given the large numbers of differences between the two atp1 haplotypes [8] and the two atp9 haplotypes [12], it is highly unlikely that all four combinations of atp1-atp9 haplotypes arose via independent mutational events. This conclusion is supported by the detection of a recombination event between atp1 and atp9 haplotypes by all three non-phylogenetic substitution models implemented in RDP (Chimaera, Max-Chi, and SiScan; RDP version 2.0). Finally, the four gamete test was applied to all pairwise comparisons between segregating sites within each gene and detected three potential intragenic recombination events within atp1 (between positions 22 & 147, 415 & 507, and 507 & 978; Table 1), one event within
atp9 (between 34 & 148; Table 1) and one event within cob (sites 516 and 796; Table 1). The event between sites 22 and 147 of atp1 was the only recombination event that was detected using all 6 recombination detection models implemented in RDP (RDP, GENECONV, Bootscan, Chimaera, Max-Chi, and SiScan).

**DISCUSSION**

Many aspects of cytoplasmic diversity in *Silene vulgaris*, from the range of nucleotide diversity to the non-neutral patterns of diversity in atp1 and atp9, to the absence of complete linkage disequilibrium indicates that the cytoplasmic genomes are evolving in a manner unlike those of most previously studied taxa. Our results bolster the findings of LAROCHE et al. (1997), CHO et al. (2004) and PARKINSON et al. (2005), which found highly variable rates of mitochondrial gene evolution among different lineages and genes in interspecific and intergeneric comparisons and extend these observations to include the presence of high diversity within species. Here we offer several possible explanations for these widespread and enigmatic patterns in Silene.

**Diversity:** We observed extremely high intraspecific nucleotide diversity at atp9, perhaps the highest average number of pairwise differences among intraspecific mitochondrial haplotypes observed for any mitochondrial gene to date (Table 2). The most divergent atp9 haplotypes (e.g. Vt17b vs. 6283B; Ks = 0.37) exhibited five times the level of divergence found between rice and maize and one third the level found for atp1 within the entire genus Plantago, which has the highest documented rate of intrageneric mitochondrial evolution (CHO et al. 2004, MOWER et al. 2004). A comparable study of intraspecific mitochondrial nucleotide diversity showed that *Silene acaulis* also exhibited high intraspecific variation at the cob mitochondrial
coding region (STÄDLER and DELPH 2002). The common ancestor of *S. acaulis* and *S. vulgaris* is represented by the earliest members of the genus (OXELMAN et al. 1997) indicating that high mitochondrial variation may be a characteristic of the entire genus, although the amount of variation in particular genes, such as *cob* is not consistent across taxa (Table 3).

Three potential mechanisms may be responsible for the high rates of variation within plants: 1) lineage specific increased mutation rates of mitochondrial genes throughout the genome (CHO et al. 2004, PARKINSON et al. 2005); 2) divergence in allopatry or in separate species and recent hybridization or introgression; and 3) the maintenance of ancient mitotypes resulting from long term balancing selection, possibly due to the presence of cytoplasmic male sterility genes (STÄDLER and DELPH 2002). For the first hypothesis to be correct for Silene, different mitochondrial lineages within *S. vulgaris* would be predicted to have different mutation rates; an explanation that is possible but perhaps unlikely at the intraspecific scale (but see CHO et al. 2004). Also, we found that different genes within the *S. vulgaris* mitochondrial genome evolve at different rates. Such a pattern is unlikely if elevated diversity results from a genome-wide increase in mutation rate, but could result if mutation rate is elevated for particular genes and not for others.

Population structure, hybridization and introgression can generate high levels of diversity such as were exhibited in *S. vulgaris*. Although *S. vulgaris* exhibits population structure at an ecological timescale (MCCAULEY 1998; OLSON and MCCAULEY 2002), populations in the US and Europe have been isolated for only a few hundred years (MCCAULEY et al. 2003) and are found in disturbed areas that are not conducive to long term population persistence. The lack of allopatric isolation and the presence of diverged haplotypes in similar geographic locations (Figure 1, Table 1) suggests that recent divergence in allopatry was not likely to generate the observed patterns of diversity. Moreover, if more ancient allopatric isolation among European
and/or Asian populations is responsible for high contemporary variation, a genome-wide influence would be expected; so under this hypothesis it is puzzling why elevated diversity was not found among all mitochondrial and chloroplast genes. *Silene vulgaris* is widespread throughout Europe and known to hybridize with some closely related sympatric *Silene* species such as *S. uniflora* (RUNYEON-LAGER and PRENTICE 2000); thus, hybridization and introgression from different species is a potential mechanism generating some diversity. *Silene vulgaris* does not hybridize with highly divergent species such as *S. latifolia* and *S. acaulis*, however, so introgression is unlikely to explain the presence and persistence of the extremely divergent haplotypes in *S. vulgaris*.

Haplotype diversity in *atp1* and *atp9* is characterized by long internal and short external branches, a pattern consistent with the maintenance of divergent haplotypes over long time periods due to either balancing selection or population structure. As discussed above, it is unlikely that this pattern was generated by population structure. Although *atp1* and *atp9* genealogies exhibited non-neutral evolutionary patterns, it seems unlikely that the target of selection was either *atp1* or *atp9* because most polymorphisms in these genes occurred at synonymous sites (Table 1). Instead, allelic variation in mitochondrial genes may be maintained via hitchhiking effects with selection on nearby linked loci. The most likely candidates are cytopasmic male sterility (CMS) genes which theoretical models suggest may be maintained via negative frequency dependent selection (FRANK 1989, GOUYON *et al.* 1991).

In contrast to our results, studies of chloroplast intergenic spacer regions in *S. vulgaris* found patterns consistent with epidemic-like selective sweeps (INGVARSSON and TAYLOR 2002). Specifically, intraspecific variation at chloroplast intergenic spacer regions in *S. vulgaris* was depauperate relative to that in *S. latifolia* (a dioecious congener) after calibration against variation at a nuclear locus, consistent with epidemic-like selective sweeps (INGVARSSON and
This is in opposition to patterns consistent with the long term maintenance of high diversity in two of the mitochondrial genes that we studied (atp1 and atp9) in *S. vulgaris*. One explanation for these contrasting results is that chloroplast and mitochondrial genes may have sufficiently independent inheritance that selection in one genome does not influence patterns of diversity in the other.

**Congruence, heteroplasmy and recombination:** If organelles are strictly maternally inherited, we would expect the mitochondria and chloroplast genomes to contain congruent phylogenetic signals (OLSON and McCauley 2000). We found an absence of congruence between mitochondrial and chloroplast genes, and also between mitochondrial genes. The only pairs of coding regions for which congruence could not be rejected included *ndhF* or *cob*, where there was little phylogenetic signal. The absence of congruence illustrates the departure from the current paradigm of organelle gene evolution in *S. vulgaris*, and indicates the presence of at least transient heteroplasmy, non-co-inheritance of organelles, and recombination.

The data reported here contrast those by OLSON and McCauley (2000) which found no evidence of incongruence between chloroplast and mitochondrial markers in *S. vulgaris* from Virginia. Both studies included individuals from some of the same populations in Virginia and in the current study, Virginia samples exhibited chloroplast-mitochondrial incongruence; thus, the difference in outcomes of the studies did not result from the smaller geographic scale sampled in OLSON and McCauley (2000). The other major difference between the studies is that the current study included DNA sequences from coding regions only, whereas OLSON and McCauley (2000) included cpDNA non-coding sequences and mtDNA RFLP southern blots associated with regions flanking the *cytochrome oxidase* I locus. The RFLP flanking regions and intergenic spacers evolve relatively quickly compared to coding regions, and thus we suspect that each
dataset is capable of detecting re-association among the chloroplast and mitochondria at different time scales. Thus, there may be no evidence for recent chloroplast-mitochondria re-association as shown by OLSON and MCCAULEY (2000), but strong evidence for re-association of the genomes over a longer time scale (shown here).

Our data exhibit strong evidence for the exchange of genetic information among mitochondrial genomes occurring within the timescale of species. Nonetheless, reciprocal recombination between different alleles from mitochondrial homologues must still be a relatively rare event; otherwise recombination would have generated more intermediate allelic variation between the divergent alleles we observed. If the $\text{atp}1$ and $\text{atp}9$ haplotypes are ancient, even low rates of recombination would be preserved in present day haplotypes. The absence of high diversity and ancient haplotypes in $\text{cob}$ is puzzling in light of the high diversity found in $S. acaulis$ (STÄDLER and DELPH 2002), but possibly explained by the often-found close proximity of CMS genes and $\text{atp}$ subunits in species where CMS genes are well characterized (HANSON and BENTOLILA 2004). We conjecture that $\text{atp}1$ and $\text{atp}9$ may be more tightly linked than $\text{cob}$ to putative CMS genes. The influence of hitchhiking selection decreases rapidly with distance from the target of selection (HUDSON and KAPLAN 1988; SCHIERUP et al. 2000), and thus may have only weak influence on diversity at $\text{cob}$. This linkage conjecture must be tempered, however, by our incomplete knowledge of the stability of the physical proximity of plant mitochondrial genes over time. For hitchhiking to be an explanatory mechanism, the physical linkage between genes must be maintained in the face of frequent rearrangements to which mitochondrial genomes are prone (PALMER and HERBON 1988).

The larger question of why few angiosperm species that carry CMS have high frequencies of females in populations (which characterizes gynodioecy) remains unresolved. The presence of CMS and associated male fertility restorers in several agricultural crop species such
as maize and rice, despite the absence of gynodioecy in their wild relatives, suggests that the rate of mutation of new CMS genes may not differ considerably among gynodioecious and non-gynodioecious species. There appear to be two quite different fates of CMS genes after they arise within a species. In most species, such as the wild relatives of crops, CMS and their associated male fertility restorer genes probably spread as epidemics, eventually homogenizing the mitochondrial genome within the ultimately hermaphroditic populations. However, in many of the species that we recognize as gynodioecious, a confluence of associations between the fitnesses of CMS types and their restorers, and perhaps even paternal inheritance, allows the long term maintenance of CMS polymorphism within a species. A comprehensive theory of the complex evolutionary forces that lead to gynodioecy may one day unite our understanding of the evolution of CMS and mitochondrial molecular evolution within all angiosperms and promises to continue as one of the major themes in modern studies of plant breeding system evolution.

We thank D. Charlesworth and H. Štorchová for seeds. D.E. McCauley contributed seeds and inspiration. Accommodations during seed collection forays in Virginia were provided by the Mountain Lake Biological Station. Our manuscript benefited from discussions with D.E. Wolf, N. Takebayashi, the Olson lab, and Toby Bradshaw. Three anonymous reviewers provided useful comments on an earlier version of the manuscript. This work was supported by National Science Foundation Grant No. DEB 03-17115 to M.S.O.
LITERATURE CITED


DARWIN, C., 1887 The different forms of flowers on plants of the same species. University of Chicago Press, Chicago IL.


distribution of chloroplast DNA polymorphism across the introduced range of *Silene

paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a
gynodioecious plant. Heredity **95**: 50-58.

MOWER, J. P. S. STEFANOVIC, G. J. YOUNG, and J. D. PALMER, 2004  Gene transfer from
parasitic to host plants. Nature **432**: 165-166.


NEI, M., and T. GOJOBORI, 1986  Simple methods for estimating the numbers of synonymous

OLSON, M. S., and D. E. MCCAULEY, 2000  Linkage disequilibrium and phylogenetic
congruence between chloroplast and mitochondrial haplotypes of *Silene vulgaris*. Proc.

OLSON, M. S., and D. E. MCCAULEY, 2002  Mitochondrial DNA diversity, population structure,
and gender association in the gynodioecious plant *Silene vulgaris*. Evolution **56**: 253-
262.

OLSON, M.S., D.E. MCCAULEY and D. TAYLOR, 2005  Genetics and adaptation in structured
populations: sex ratio evolution in *Silene vulgaris* Genetica **123**: 49-62.

OXELMAN, B., M. LIDEN, and D. BERGLUND, 1997  Chloroplast *rps16* intron phylogeny of the

PALMER, J. D., and L. A. HERBON, 1988  Plant Mitochondrial-Dna Evolves Rapidly In


TABLE 1.
Polymorphic sites in the three mitochondrial (atp1, atp9, cob) and two chloroplast (matK, ndhF) coding regions of Silene vulgaris.

<table>
<thead>
<tr>
<th>Gene</th>
<th>atp1</th>
<th>atp9</th>
<th>cob</th>
<th>matK</th>
<th>ndhF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>557</td>
<td>1111111111111111</td>
<td>557</td>
</tr>
<tr>
<td></td>
<td>1122333444455556666790</td>
<td>1111111111111111</td>
<td>111112333444444555</td>
<td>11133455666778</td>
<td></td>
</tr>
<tr>
<td></td>
<td>112249575790159022899071</td>
<td>235568899112233344556888</td>
<td>7119</td>
<td>13347914789095593579925</td>
<td>68900556588013</td>
</tr>
<tr>
<td></td>
<td>3291275561592502765103284</td>
<td>04895251981726958149127</td>
<td>7676</td>
<td>96966510066591881105624</td>
<td>80416147522630</td>
</tr>
<tr>
<td>SD6</td>
<td>GAGCATCTAGACACCCGTAGTAACC</td>
<td>CTAGATCGACGCCTGGTGCTTC</td>
<td>ATCT</td>
<td>CATGAATCGATCGACGGACT</td>
<td>TC0GTTCTTTCTCC</td>
</tr>
<tr>
<td>Type A</td>
<td>..T.............</td>
<td>..T.............</td>
<td>A.....</td>
<td>..T.............</td>
<td>..A.............</td>
</tr>
<tr>
<td>Type B</td>
<td>..................</td>
<td>A.....</td>
<td>..C .......................</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>MV22h</td>
<td>..................</td>
<td>T.............</td>
<td>A.....</td>
<td>..T..................</td>
<td>..A.............</td>
</tr>
<tr>
<td>SD3</td>
<td>..................</td>
<td>T.............</td>
<td>A.....</td>
<td>..T..................</td>
<td>..A.............</td>
</tr>
<tr>
<td>VtLnG</td>
<td>..................</td>
<td>T.............</td>
<td>A.....</td>
<td>..T..................</td>
<td>..A.............</td>
</tr>
<tr>
<td>Type C</td>
<td>..................</td>
<td>A..................</td>
<td>G.. AT..C....ACT</td>
<td>A. ...T......G.T.........</td>
<td>..............</td>
</tr>
<tr>
<td>Type D</td>
<td>..A...............</td>
<td>CT..GCT.TA..AT..TC..</td>
<td>.G.. T......................</td>
<td>...A.............</td>
<td></td>
</tr>
<tr>
<td>NY007</td>
<td>..................</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>G.C T.............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>Ld5</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>Vt17b</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>EB3AG</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>GM7</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>MK4</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>KL1</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>KL2</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>Ld4</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>RM4</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>RM8</td>
<td>A...T.T...ATTG...</td>
<td>A...GC.T.TA..AC.T.A.C..</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>Vt100</td>
<td>AGT..CTCG.GTCG......CGTAT</td>
<td>A......................</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>InG</td>
<td>AGT..CTCG.GTCG......CGTAT</td>
<td>A......................</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
<tr>
<td>Type E</td>
<td>..T.............</td>
<td>..T.............</td>
<td>A.....</td>
<td>..T.............</td>
<td>..A.............</td>
</tr>
<tr>
<td>KM5</td>
<td>..T.............</td>
<td>..T.............</td>
<td>A.....</td>
<td>..T.............</td>
<td>..A.............</td>
</tr>
<tr>
<td>EB12A</td>
<td>..T.............</td>
<td>..T.............</td>
<td>A.....</td>
<td>..T.............</td>
<td>..A.............</td>
</tr>
<tr>
<td>Kn</td>
<td>AGT..CT.G......T...AT</td>
<td>A......................</td>
<td>..............</td>
<td>..............</td>
<td></td>
</tr>
</tbody>
</table>

28
Type F:  

Substit. type  

S. latifolia  

S. uniflora  

Type A: Ld3, CO-9F, 62812; Type B: AJ8-E, CO-5F, MVh17, GRf6c, GR6-6, CO-50; Type C: 6282B, 6283B, 6283b, 62876, 6283F; Type D: AR-34, AR-9H; Type E: KM3, KM4; Type F: WR-05, WR-07, AR-22. Dots indicate that the base at that site is identical to the one found in the reference sequence, SD6. Position of the base in the gene is indicated above the alignment.
**TABLE 2.**

Diversity measures in organelle genes of *Silene vulgaris*.

<table>
<thead>
<tr>
<th>Gene</th>
<th>$\theta$</th>
<th>$\theta$(s)</th>
<th>$\pi$</th>
<th>Tajima’s D</th>
<th>Fu &amp; Li D</th>
<th>K-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>atp1</em></td>
<td>0.005</td>
<td>0.018</td>
<td>0.006</td>
<td>0.33</td>
<td>1.47*</td>
<td>7, 9-19 *</td>
</tr>
<tr>
<td><em>atp9</em></td>
<td>0.021</td>
<td>0.070</td>
<td>0.021</td>
<td>-0.33</td>
<td>1.69*</td>
<td>8, 7-18</td>
</tr>
<tr>
<td><em>cob</em></td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.30</td>
<td>1.02</td>
<td>6, 3-5*</td>
</tr>
<tr>
<td><em>matK</em></td>
<td>0.004</td>
<td>0.002</td>
<td>0.003</td>
<td>-0.80</td>
<td>-0.54</td>
<td>8, 4-10</td>
</tr>
<tr>
<td><em>ndhF</em></td>
<td>0.001</td>
<td>0.002</td>
<td>0.0002</td>
<td>-1.76</td>
<td>-2.20</td>
<td>5, 3-5</td>
</tr>
</tbody>
</table>

$\theta$ per site estimated from the total number of mutations (Watterson 1975). $\theta$s per site, including only synonymous nucleotide differences. $\pi$ Nucleotide diversity, the average number of nucleotide differences per site between a pair of randomly chosen sequences (Nei 1987). Tajima’s D (Tajima 1989), based on the differences between the number of segregating sites and the average number of nucleotide differences per sequence. Fu & Li D with an outgroup (Fu and Li 1993), the test statistic is based on the differences between the total number of mutations in the external branches of the genealogy, and the total number of mutations. K test (Depaulis and Veuille 1998) actual no. of haplotypes, range of expected no. of haplotypes, significance). *

$*= P < 0.05$
**TABLE 3.**

Intraspecific average inter-haplotype pairwise synonymous and non-synonymous substitutions per site in organelle genes of *Silene vulgaris* relative to other taxa and the rice-maize divergence.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length Variable</th>
<th>K&lt;sub&gt;s&lt;/sub&gt;</th>
<th>K&lt;sub&gt;a&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Silene vulgaris</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>atp</em>&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1067</td>
<td>25</td>
<td>0.0373</td>
</tr>
<tr>
<td><em>atp</em>&lt;sub&gt;9&lt;/sub&gt;</td>
<td>213</td>
<td>21</td>
<td>0.0961</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1142)</td>
</tr>
<tr>
<td><em>cob</em></td>
<td>817</td>
<td>4</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0035)</td>
</tr>
<tr>
<td><em>matK</em></td>
<td>572</td>
<td>9</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0061)</td>
</tr>
<tr>
<td><em>ndhF</em></td>
<td>1067</td>
<td>4</td>
<td>0.0033</td>
</tr>
<tr>
<td><strong>Silene acaulis</strong> †</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>cob</em></td>
<td>1041</td>
<td>22</td>
<td>0.0118</td>
</tr>
<tr>
<td><strong>Plantago spp.‡</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>atp</em>&lt;sub&gt;1&lt;/sub&gt;</td>
<td>615</td>
<td>183</td>
<td>1.1249</td>
</tr>
<tr>
<td><em>ndhF</em></td>
<td>1624</td>
<td>213</td>
<td>0.0795</td>
</tr>
<tr>
<td><em>cox</em>&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1344</td>
<td>363</td>
<td>0.1908</td>
</tr>
<tr>
<td><em>rbcL</em></td>
<td>1317</td>
<td>181</td>
<td>0.0753</td>
</tr>
<tr>
<td><strong>rice-maize†</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>atp</em>&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1521</td>
<td>-</td>
<td>0.0598</td>
</tr>
<tr>
<td>Gene</td>
<td>Length</td>
<td>Divergence</td>
<td>Ks</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>------------</td>
<td>----</td>
</tr>
<tr>
<td>atp9</td>
<td>216</td>
<td>-</td>
<td>0.0748</td>
</tr>
<tr>
<td>cob</td>
<td>1155</td>
<td>-</td>
<td>0.0241</td>
</tr>
<tr>
<td>matK</td>
<td>1558</td>
<td>133</td>
<td>0.1993</td>
</tr>
<tr>
<td>ndhF</td>
<td>1697</td>
<td>185</td>
<td>0.1460</td>
</tr>
</tbody>
</table>

Ks: inter-haplotype average number of pairwise synonymous substitutions per synonymous site. Ka: inter-haplotype average number of pairwise non-synonymous substitutions per non-synonymous sites. Values in parentheses are adjusted to reflect potential RNA editing. † (Städler and Delph 2002) following (Gaut 1998). ‡ (Cho et al. 2004 – Genbank accession numbers AJ389588 – AJ389621; AY818897 – AY818951).
Figure 1 Map of the locations of *Silene vulgaris* samples collected for this study. The large dots represent multiple populations from within the same locality, whereas the small dot represents the location of a single population.

Figure 2. Maximum Likelihood Phylograms of the three mitochondrial (*atp9, cob, atp1*) and two chloroplast (*matK, ndhF*) genes. Bootstrap support values greater than 50 are shown above the branches. Individuals with five character long names are *Silene vulgaris* from North America, those with three character names are from Europe. Outgroup sequences from *Nicotiana tabacum* and *Beta vulgaris* were downloaded from Genebank (accession no.s BA00042, NC001879, BA000009, AY514832).