Variation in Mutation Rate and Polymorphism Among Mitochondrial Genes of *Silene vulgaris*

Camille M. Barr,† Stephen R. Keller,* Pär K. Ingvarsson,‡ Daniel B. Sloan,* and Douglas R. Taylor**

*Department of Biology, University of Virginia; †Division of Biological Sciences, University of Montana; and ‡Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden

The prevailing wisdom of the plant mitochondrial genome is that it has very low substitution rates, thus it is generally assumed that nucleotide diversity within species will also be low. However, recent evidence suggests plant mitochondrial genes may harbor variable and sometimes high levels of within-species polymorphism, a result attributed to variance in the influence of selection. However, insufficient attention has been paid to the effect of among-gene variation in mutation rate on varying levels of polymorphism across loci. We measured levels of polymorphism in seven mitochondrial gene regions across a geographically wide sample of the plant *Silene vulgaris* to investigate whether individual mitochondrial genes accumulate polymorphisms equally. We found that genes vary significantly in polymorphism. Tests based on coalescence theory show that the genes vary significantly in their scaled mutation rate, which, in the absence of differences among genes in effective population size, suggests these genes vary in their underlying mutation rate. Further evidence that among-gene variance in polymorphism is due to variation in the underlying mutation rate comes from a significant positive relationship between the number of segregating sites and silent site divergence from an outgroup. Contrary to recent studies, we found unconvincing evidence of recombination in the mitochondrial genome, and generally confirm the standard model of plant mitochondria characterized by low substitution rates and no recombination. We also show no evidence of significant variation in the strength or direction of selection among genes; this result may be expected if there is no recombination. The present study provides some of the most thorough data on plant mitochondrial polymorphism, and provides compelling evidence for mutation rate variation among genes. The study also demonstrates the difficulty in establishing a null model of mitochondrial genome polymorphism, and thus the difficulty, in the absence of a comparative approach, in testing the assumption that low substitution rates in plant mitochondria lead to low polymorphism.

Introduction

The prevailing wisdom of the plant mitochondrial genome, based on data among distantly-related taxa, is that it has very low substitution rates, high rearrangement rates, maternal inheritance, and no recombination (Wolfe et al. 1987, 1989; Palmer and Herbon 1988; Gaut 1998). For these reasons, it has been assumed that plant mitochondrial genes have low diversity both within species and between closely-related taxa. Thus, in contrast to animal mitochondria which evolve much more rapidly and do not incur frequent rearrangements, plant mitochondria have generally been avoided for phylogenetic reconstruction and population genetic studies (Muse 2000).

Because there are few species-level data on plant mitochondrial gene diversity, it is unclear whether the observed low substitution rates, estimated from divergences, widely translate into low polymorphism within species. Within-species polymorphism is a function of a number of factors including mutation rate, effective population size, strength of selection (Bazin et al. 2006), population structure (Ingvarsson 2004), and mating system, yet the relative contribution of each of these factors to mitochondrial sequence diversity within species is unclear. In addition, recent work has shown very high mitochondrial substitution rates in some plant lineages (Cho et al. 2004; Parkinson et al. 2005). There is also recent evidence for paternal transmission (McCauley et al. 2005), and recombination (Stadler and Delph 2002), all of which can contribute to polymorphism. The assumptions of general low substitution rates and low polymorphism in the mitochondrial genome may therefore not apply across all plant taxa.

Two recent studies have investigated how much mitochondrial/cytoplasmic polymorphism segregates within plant species, and what processes are responsible for shaping this polymorphism (Ingvarsson and Taylor 2002; Stadler and Delph 2002). However, while both studies reject a neutral model of evolution, they essentially reach opposite conclusions regarding whether polymorphism is elevated or reduced within species. Given the paucity of data addressing mitochondrial polymorphism within plant species, the different outcomes from these 2 studies are difficult to interpret (Charlesworth 2002). These studies highlight the more general problem that our current understanding of mitochondrial or cytoplasmic polymorphism is based on few genes, and because genes vary in their accumulation of mutations, these genes may not be representative of mitochondrial evolution as a whole.

In the present paper, we estimate species-wide polymorphism and substitution rates in six of the putative 27–28 protein-coding genes (Adams et al. 2002), and one intergenic spacer, in the herbaceous plant *Silene vulgaris*. Variation in substitution rate among mitochondrial genes is commonly found in animal taxa, and is also known in plant taxa (Wolfe et al. 1987, 1989). We examine whether similar variation exists among plant mitochondrial genes in our system. We first test whether the observed variance in polymorphism among genes can be explained by the variance in mutations expected from a single coalescent process. We then test if variance in polymorphism can be attributed to variance in the underlying mutation rate (c.f. the Hudson Kreitman Aguade (HKA) test). Only when we reject these null models do we then test whether variation in

Key words: plant mitochondria, polymorphism, mutation rate, recombination, *Silene vulgaris*.

E-mail: camille.barr@mso.umt.edu.

Mol. Biol. Evol. 24(8):1783–1791. 2007
doi:10.1093/molbev/msm106
Advance Access publication May 28, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

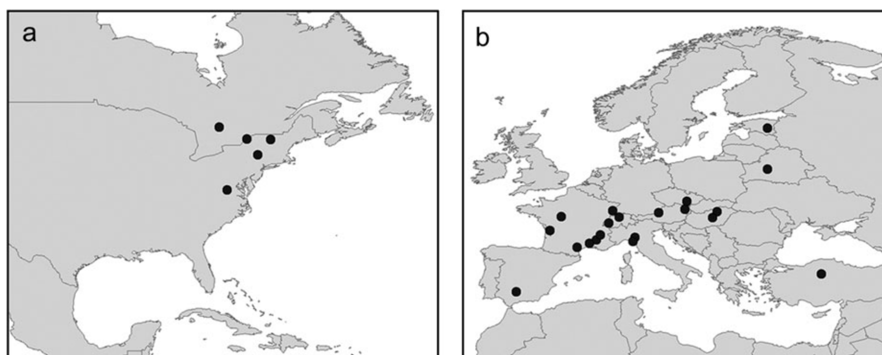

FIG. 1.—Map of sampling locations in (*a*) North America and (*b*) Europe.

the history of selection can explain variance in polymorphism among genes.

We also test for evidence of recombination in our full sample of mitochondrial genes. Recombination is an important force facilitating the differential effects of selection or demography on genes. Whether or not plant mitochondria recombine is hotly debated (Barr et al. 2005). Most evidence suggests that they do not, but a few studies suggest they might (Stadler and Delph 2002; McCauley et al. 2005). Detecting recombination, however, is highly dependent on the method used, and we discuss pitfalls of various approaches in relation to our data.

Materials and Methods

Silene vulgaris (Caryophyllaceae) is an herbaceous perennial that grows in open areas throughout Europe, and has been introduced to North America. We collected seeds from 25 populations throughout Europe and Eastern North America (fig. 1), and grew 1 plant per population in the greenhouse. We extracted DNA using the DNeasy plant miniprep kit (Qiagen, Valencia, CA).

We amplified and sequenced fragments of coding regions from 6 protein-coding mitochondrial genes: NADH dehydrogenase subunit 9 (nad9), cytochrome b (cob), cytochrome c oxidase subunit 3 (cox3), ATP synthase subunit 1 (atp1), ATP synthase subunit 4 (atp4), and ATP synthase subunit 6 (atp6) (Accession #s: EF139460-634). We also amplified and sequenced one non-coding intergenic spacer between the genes NADH dehydrogenase subunit 4L and atp4 (nad4L-atp4), and one chloroplast intron of the gene for the tRNA of leucine (trnL) (accession #s: AF518985, AF518992, AF518997-8, AF519000-5, AF519007, AF519009-10, EF139642-53). Primer sequences were either designed by us (atp6), designed by colleagues (atp1, C. dePamphilis), or obtained from published sources (nad9, cox3, atp4, nad4L-atp4, Duminil, Pemonge, and Petit 2002; cob, Stadler and Delph 2002; trnL, Taberlet et al. 1991). We cleaned PCR products using either the QIAquick PCR purification kit (Qiagen, Valencia, CA) or Exonuclease (Exo1) and Shrimp Alkaline Phosphatase (SAP) (USB Scientific), and ran cycle-sequenced samples on an automated sequencer (ABI377 and ABI3130, Applied Biosystems, Foster City, CA). We verified sequences manually and assembled contigs in the program SEQUENCHER (Gene Codes, Ann Arbor, MI), and ran multiple sequence alignments in MEGA (Kumar, Tamura, and Nei 2004). We used the program DnaSP v4 (Rozas et al. 2003) to calculate numbers of silent and replacement sites and estimate pairwise divergences using the Nei-Gojobori method (Nei and Gojobori 1986) with the Jukes-Cantor correction (Jukes and Cantor 1969).

The presence of heteroplasmy or paralogy can have a significant impact on the estimation and interpretation of gene polymorphism by inflating estimates of polymorphism. While collecting our polymorphism data, we screened a sample of individuals for evidence of multiple gene copies within individuals. Previous work has suggested that multiple copies are possible, with both paternal inheritance and heteroplasmy detected in *S. vulgaris* (McCauley et al. 2005), as well as paralogy of mitochondrial genes in other plant species (Nugent and Palmer 1991). High levels of polymorphism can also increase the chances of detecting heteroplasmy. We performed molecular cloning (Invitrogen TOPO cloning kit) on all or part of the genomic region for several genes, and then sequenced these clones (3–5 colonies sequenced per clone). For cob and atp1, we found variant sequences within single individuals. In cob, 2 of the 6 segregating sites in our full sample showed the alternative bases between clones from a single individual. It is possible that errors in *Taq* polymerase could lead to data that look like multiple copies; however, these sites that indicate multiple copies were segregating in the full sample of plants, and thus are not likely to be single instances of polymerase error. Despite the presence of multiple copies, we find divergence times that fall well within the norm. Thus, in our sample, multiple copies are not likely inflating polymorphism estimates. However, given our confirmation of multiple copies of 2 different mitochondrial genes, we suggest that in general, appropriate caution should be exercised to control for this possibility, particularly in cases in which unusual results appear.

Mutation Rate Variation

We measured divergence from 2 outgroups: *Beta vulgaris ssp. vulgaris* (Accession #: NC_002511) and *Silene paradoxa* (Accession #s: EF139635-41). *Beta vulgaris* is

in the family Amaranthaceae, which is a sister family of the Caryophyllaceae (Soltis et al. 2000) and has a complete mitochondrial genome sequence available (Kubo et al. 2000). The 2 families diverged from each other approximately 40.5 mya (Davies et al. 2004). Average pairwise divergences for silent sites (dS) (Nei and Gojobori 1986) among genes between the published sequences of *B. vulgaris* and *S. vulgaris* ranged between 0.016 and 0.078. These values are well below 1, often considered to be the threshold above which multiple substitutions per nucleotide site are likely to saturate (Endo, Ikeo, and Gojobori 1996; Conticello et al. 2001). For the congener *S. paradoxa*, we sequenced all genes used in the present analysis from an F2 progeny of individuals collected in Italy. *S. paradoxa* falls into a clade separate from *S. vulgaris* within the genus and thus appeared to be a suitable outgroup. Synonymous pairwise divergences between *S. vulgaris* and *S. paradoxa* were also well below 1, ranging from 0.00 to 0.13.

To test whether the observed polymorphism among genes is the result of significant variation in the scaled mutation rate among genes, we used 2 tests. For each, we considered only silent sites to conform to the assumptions of neutral coalescent models.

First we used a maximum likelihood version of the HKA (Hudson-Kreitman-Aguade) test developed by Wright and Charlesworth (2004). The HKA test operates under the principle that the level of polymorphism within a species for a haploid genome is a function of $\theta = 2N\mu$, the scaled mutation rate, and that from an outgroup is a function of θ and T, the divergence time between the focal species and the outgroup. Therefore, under the neutral model we expect polymorphism and divergence to be positively correlated. Using the ideas outlined in Wright and Charlesworth (2004), we fit 2 separate models to our data set. The first model allows for gene specific estimates of the scaled mutation rate, θ, whereas the second model only allows for a single, genome-wide estimate of θ. Since model 2 represents a nested subset of model 1, the difference in the log-likelihood between the 2 models is approximately chi-square distributed with degrees of freedom equal to the difference in the number of parameters estimated by the two models. The chain length used in running the mlHKA program was 200.000 in all cases. Model 1 estimates 7 parameters (6 θ estimates and an estimate of T, the divergence time between *S. vulgaris* and the ougroup) where as Model 2 estimates 2 parameters (a single estimate of θ and T). Our test therefore had 5 df.

Second we used the maximum likelihood method developed by Hudson (1991) to provide gene specific estimates of θ. This method uses recursion equations to generate the likelihood of the numbers of segregating sites in a sample, given the sample size and a specific value of θ, assuming a strictly neutral coalescent process.

We calculated this likelihood function for each gene over a range of θ values per silent site ranging from 0 to 0.05 and with a sample size of n = 25. We estimated approximate likelihood-based 95% confidence regions by identifying upper and lower values of θ for which $\log L(\theta) - \log L(\theta_{max}) < -1.92$, where $L(\theta_{max})$ is the maximum likelihood estimate of θ. Pairs of genes for which these approximate confidence regions did not overlap were considered to have θ values that were significantly different. This test is conservative, with type I error rates that are 10–100 times smaller than expected (0.05 for a 95% CI) (data not shown), and identifies those genes that differ most in θ.

We visualized the results of the HKA test by plotting the proportion of silent segregating sites (calculated as the number of segregating sites per site) as a function of divergence from outgroups *S. paradoxa* and *B. vulgaris* (calculated as the number of fixed differences between *S. vulgaris* and the outgroup, per number of sites). If silent sites represent sites under minimal selection, and variation in mutation accumulation is due solely to variation in mutation rate, then a positive relationship between divergence from a common ancestor and the amount of existing polymorphism in silent sites should exist. If mutations are accumulating according to a neutral process, those genes that have higher mutation rates should have higher silent site divergence from a common ancestor, and higher silent site polymorphism. If replacement sites are under greater selection than silent sites, as expected, we predict a weak or non-existent relationship between divergence and polymorphism.

Recombination

We tested for evidence of recombination using nine methods. Tests for recombination tend to yield highly variable results, and no single method performs well in all cases (Posada and Crandall 2001). Six methods were run using the program RDP2 (Recombination Detection Program) developed by Darren Martin (http://darwin.uvigo.es/) (Martin et al. 2005): RDP, GENECONV (Padidam et al. 1999), Bootscan (Salminien et al. 1995), Chimaera (Posada and Crandall 2001), Maximum Chi-square (MaxChi) (Maynard Smith 1992), and Sister Scanning (SiScan) (Gibbs et al. 2000). All methods except GENECONV use a sliding-window approach to compare portions of sequence for incongruities (Martin et al. 2005). We also ran a 4-gamete test (Hudson and Kaplan 1985), which scans the data for evidence of all combinations of bases between pairs of segregating sites, in the program DnaSP v4 (Rozas et al. 2003). Finally, we ran the Homoplasy test which compares the number of steps required to generate observed polymorphisms in a most parsimonious tree to the number of homoplasies generated, and is suitable for datasets with low divergence between samples (Maynard Smith and Smith 1998). We implemented the Homoplasy test using the START software made available by Keith Jolley (Jolley et al. 2001) (http://pubmlst.org/software/analysis/start/). All tests were run on a concatenated dataset of third position sites of coding genes and all base pairs for the non-coding gene nad4L-atp4, to avoid the potential for selective constraints reducing recombination detection (Eyre-Walker et al. 1999). We also ran all tests on the full dataset of 3093 bp (except for the Homoplasy test), which allows comparison with other studies which have used all nucleotides. Because low frequency alleles have been shown to significantly affect the detection of recombination (McVean et al. 2002), for some analyses we explored the effect of low

Table 1
Polymorphism in Genes of *Silene vulgaris*

Gene Region	Sequence Length	#Total Segregating Sites	% Total Segregating Sites	# Silent Segregating Sites	% Silent Segregating Sites
nad9	393	1	0.25	0	0
cob	915	7	0.77	3	1.39
cox3	603	7	1.16	1	0.69
atp1	969	22	22.7	18	7.35
atp4	291	0	0.00	0	0
atp6	693	8	1.15	4	2.52
nad4L-atp4 (non-coding)	134	4	2.99	3	2.16

frequency alleles by filtering out sites with the minor allele at a frequency of <0.1. Significant results were then followed up by tests on individual genes.

Variation in Selective Regimes

We ran a multilocus HKA test using software from J. Hey (http://lifesci.rutgers.edu/~heylab) to test whether patterns in polymorphism and divergence are consistent with neutral evolution, or, alternatively, indicate evidence of selection. This is a different application of the HKA test from the one above, which uses assumptions from the HKA test to examine whether genes vary significantly in their values of θ. Here, the HKA test generates expectations for polymorphism and divergence between loci of *S. vulgaris* and our two outgroup taxa. These expectations are based on locus specific estimates of θ and a genome-wide estimate of T, derived from the polymorphism and divergence data, respectively and Watterson's formula for neutral evolution $\theta = 2N\mu$ (given a haploid model as with mitochondria) (Watterson 1975; Hudson et al. 1987). A chi-square goodness of fit test was then run to test for deviations of the observed data from neutral expectations. It is important to note that the HKA test assumes unlinked loci, which may not be the case in a mitochondrial genome.

A second test for evidence of selection on the mitochondrial genes is the McDonald-Kreitman test (McDonald and Kreitman 1991). This test compares the ratios of variable silent and replacement sites within-species (polymorphism) to that between species (divergence). If both silent and replacement mutations are neutral or nearly-neutral, then the ratio of silent to replacement substitutions should not differ significantly when comparing individuals on a short time scale (within-species) to individuals that have accumulated substitutions on a longer time scale (between-species). If, however, replacement substitutions are not neutral, and there has been selection affecting their frequency, then these 2 ratios will not be similar. Although silent sites may not always be neutral (Comeron 2006), the McDonald-Kreitman test is conservative because replacement sites are under relatively stronger selection than silent sites (Charlesworth 1994). In addition, the McDonald-Kreitman test is robust to the effects of recombination and demography because silent and replacement sites of the same gene have the same demographic history (Nielsen 2001). Because the McDonald-Kreitman test is comparing sites within loci, it can only detect the effects of selection if selection has operated directly on that locus, whereas the HKA test can detect the effects of selection on a locus linked to the target locus.

Results

Polymorphism

Both the number and proportion of total and silent segregating sites in our sample varied among genes, with atp1 being the most polymorphic, and atp4 being the least polymorphic (table 1). The genealogies generally had 2 major clades which were well supported (fig. 2). There is no evidence of population structure in the sample, consistent with genealogical results based on the chloroplast genome in *S. vulgaris* (Taylor and Keller 2007).

Nucleotide silent substitution rates estimated from pairwise divergences between *S. vulgaris* and outgroup *B. vulgaris* are also variable among genes, and range from 0.202 to 0.980 substitutions per billion years (table 2). Interestingly, divergences between *S. vulgaris* and its congener *S. paradoxa* are higher for genes atp1, atp4, and atp6 than between *S. vulgaris* and *B. vulgaris* which are in closely related, but different families (data not shown). The silent substitution rate of the chloroplast gene trnL is 2 to 11 times higher than that of the mitochondrial genes (table 2).

Mutation Rate Variation

The variance in silent site polymorphism among loci was higher than would be expected from a single neutral coalescent process with a common θ across genes. A model with independent estimates of θ provided a significantly better fit to the data than a model with a single multi-locus estimate of θ, both for *S. paradoxa* ($2\Delta L = 103.09$, $p < 0.001$) and *B. vulgaris* ($2\Delta L = 38.23$, $p < 0.001$). Maximum likelihood estimates of θ (per base pair) also varied among genes, as expected given the variation in proportion of segregating sites (fig. 3). Based on the 95% confidence regions, nad9 and atp1 have estimates of θ that are significantly different.

There is a significantly positive relationship between the frequency of silent segregating sites within *S. vulgaris*, and the proportion of silent fixed differences between *S. vulgaris* and its outgroup *S. paradoxa* ($t_6 = 4.87$, $P = 0.005$, $r^2 = 0.83$) (Figure 4*a*,*b*). The same relationship between *S. vulgaris* and *B. vulgaris* is weakly positive, though

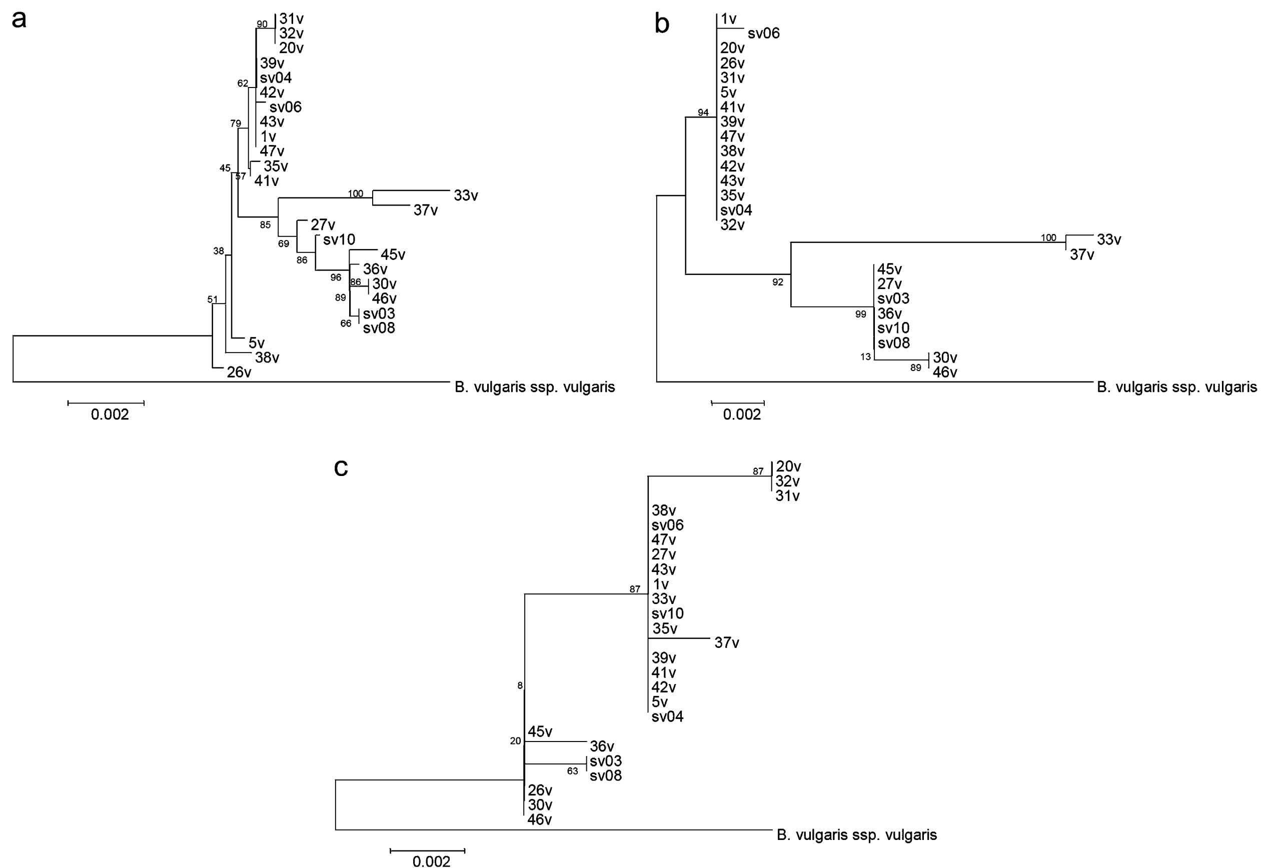

FIG. 2.—Neighbor-joining trees of (*a*) all genes, and the two genes which had clades with bootstrap support of >80%: (*b*) atp1 and (*c*) cox3. Samples labeled #v are from Europe; samples labeled sv0# are from North America.

not significant ($t_6 = 1.56$, $P = 0.18$, $r^2 = 0.33$) (Figure 4*b*,*c*). There was no significant relationship between segregating sites and fixed differences for replacement sites using either outgroup (*S. paradoxa:* $t_5 = 0.26$, $P = 0.81$, $r^2 = 0.02$, *B. vulgaris*: $t_5 = -0.83$, $P = 0.46$, $r^2 = 0.15$).

Recombination

No potential recombination events were detected with RDP, GENECONV, Bootscan, Chimaera, MaxChi, or SiScan with third position sites. Interestingly, when MaxChi is run on third position sites using the implementation of Piganeau et al. (2004), a significant signal of recombination is detected. This is the opposite of what would be predicted when comparing these 2 implementations because that of Piganeau et al. (2004) prevents large chi-square values from being generated when the expected values are very small, and thus should reduce the number of potential recombination events detected. When tests were run on the full dataset, the MaxChi test detected a significant signal of recombination with 22 potential sites of recombination ($X^2 = 14.95$, $p < 0.024$). However, Chimaera, which is a modification of MaxChi that uses triplets of sequences instead of pairs, as well as RDP, GENECONV, Bootscan, and SiScan, did not find significant evidence of recombination. The Homoplasy test (run only on third position sites), detected significant evidence of recombination (Homoplasy ratio = 0.094, p = 0.000). A minimum number of 5 recombination events was detected with the 4 gamete test both for the third position and full datasets (table 4). When we removed sites with low frequency minor alleles, the minimum number of recombination events was reduced to 3 (table 4). Two of the detected events were between sites in different genes, and one of them was between sites within the non-coding region of nad4L-atp4. Although blocks of recombinant sequences might be expected with low levels of homologous recombination, no obvious blocks were seen in our data (table 4).

Table 2
Silent Substitution Rates Between *Silene vulgaris and* Outgroup *Beta vulgaris*

Gene Region	Silent Substitution Rate (substitutions per site per 10^9 years)
	Beta vulgaris
nad9	0.519
cob	0.508
cox3	0.630
apt1	0.980
apt4	0.202
apt6	0.468
nad4L-atp4 (non-coding)	0.941
trnL intron (cp, non-coding)	2.158

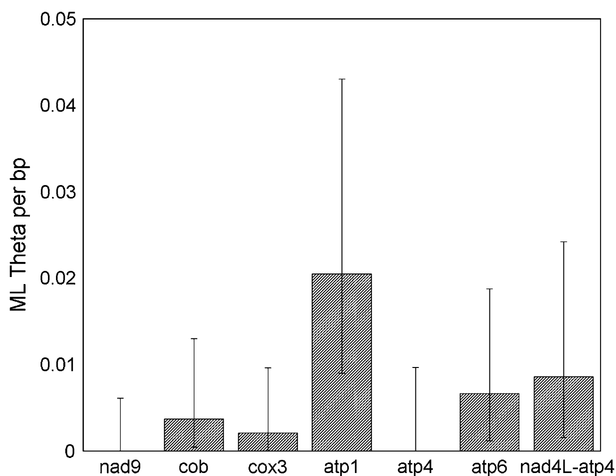

FIG. 3.—Maximum-likelihood estimates and 95% confidence regions of θ.

Variation in Selective Regimes

The multilocus HKA test using both *S. paradoxa* and *B. vulgaris* as outgroups showed no evidence of selection on the 7 loci (*S. paradoxa*: $X^2_6 = 2.37$, p = 0.88, *B. vulgaris*: $X^2_6 = 2.36$, p = 0.88), suggesting that variance in the underlying mutation rate cannot be rejected as an explanation of the variance in silent site polymorphism.

There was little evidence for selection on the individual genes tested with the McDonald-Kreitman test using *B. vulgaris* as an outgroup; the only exception was cox3 in which evidence for purifying or balancing selection was at the upper limit of significance (P = 0.05) (table 3) (results using *S. paradoxa* were not qualitatively different). In all cases, dN was smaller than dS (data not shown) which is an indication of purifying selection averaged over all sites within a gene (Nielsen 2001), though our low values of dN and dS do not provide sufficient data to perform a test of significance on dN/dS ratios.

Discussion

The overall levels of both total and silent site polymorphism in mitochondrial genes of *S. vulgaris* varied substantially among genes, but generally fell within the range of values reported in the few other published reports of plant mitochondrial genes. For example, Stadler and Delph (2002) found 22 segregating sites (2.1%, of 1041 bp) in the *cob* gene of *Silene acaulis*. However, to maximize their ability to detect polymorphism, they used RFLP analysis to enrich their sample for divergent haplotypes prior to sequencing. Because of the enrichment of divergent haplotypes, the estimate of 2.1% is likely higher than that from a random sample of plants; we found 0.83% of *cob*-nucleotides were polymorphic in our sample of *S. vulgaris*. However, even with the enrichment, the polymorphism found in *cob* for *S. acaulis* falls below that for *atp1* in *S. vulgaris*.

Nucleotide silent substitution rates estimated from pairwise divergences between *S. vulgaris* and *B. vulgaris* are also variable among genes (table 2). These values are similar to those calculated previously for other plant mitochondrial genes (Muse 2000, Gaut 1988, Wolfe, Li, and

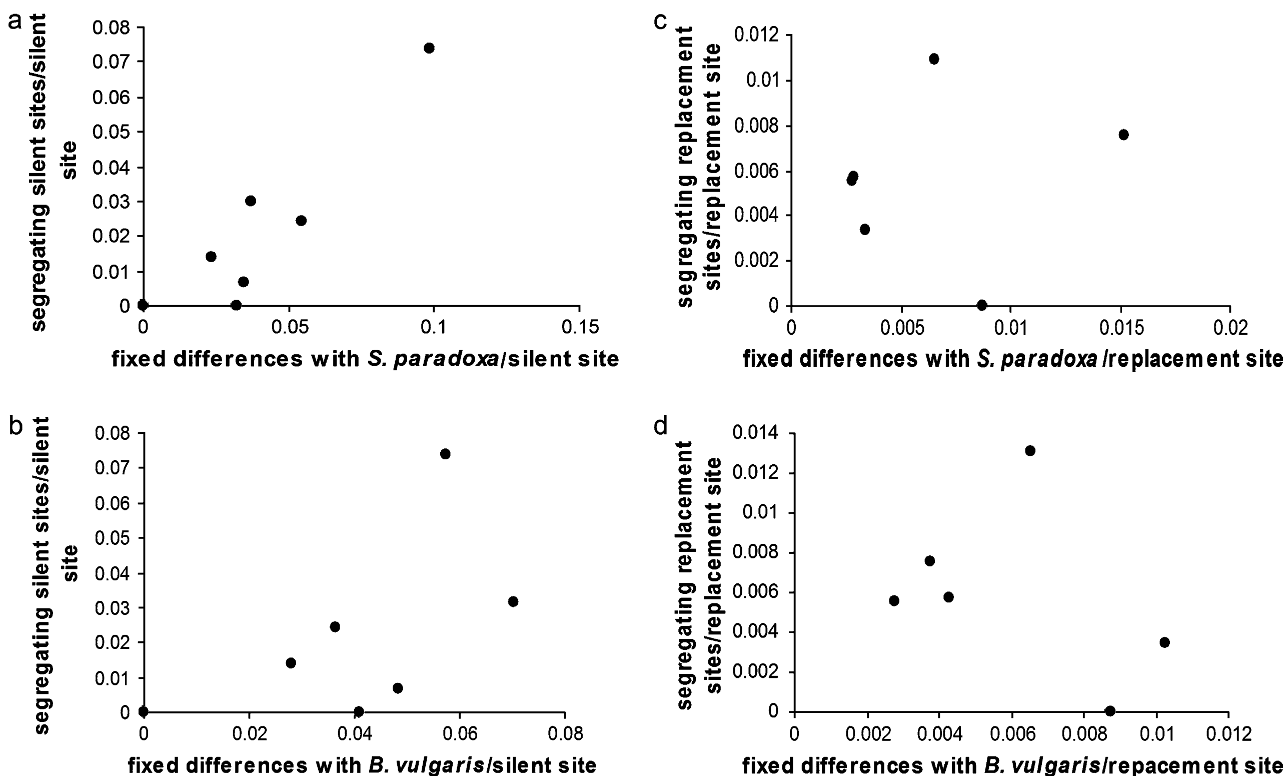

FIG. 4.—Relationship between fixed differences between *Silene vulgaris* and outgroups *Silene paradoxa*(*a*, *b*) and *Beta vulgaris* (*c*, *d*) and segregating silent (*a*, *c*) and replacement (*b*, *d*) sites.

Table 3
Values of Replacement and Silent Site Differences Within- and Between-Species (Using *Beta vulgaris* as the Outgroup) and Significance of Ratios for the McDonald-Kreitman Test

Gene		Between-Species	Within-Species	P, Fisher's Exact Test
nad9	Replacement	3	1	1.00
	Silent	4	0	
cob	Replacement	3	4	0.33
	Silent	8	3	
cox3	Replacement	3	6	0.05*
	Silent	7	1	
atp1	Replacement	2	4	0.69
	Silent	14	18	
atp4	Replacement	2	0	NA
	Silent	1	0	
atp6	Replacement	2	4	0.61
	Silent	6	4	

Sharp 1987, though see Cho et al. 2004 and Parkinson et al. 2005). The 2- to 4-fold smaller substitution rates in all mitochondrial genes (except monomorphic atp4) compared with the chloroplast region is also in line with previous observations (Gaut 1988, Wolfe, Li, and Sharp 1987). Interestingly, in a sample of 6 mitochondrial genes, 2 of which are in our sample, Wolfe, Li, and Sharp (1987) found atp1 had the second highest substitution rate. In addition, Laroche et al. (1997) report substitution rates for 5 of our 7 genes, and found that atp4 and atp1 had the highest rates, followed by cox3, and atp6 and cob. Except for atp4, the similarity of these data to ours suggests that mutation rate differences among genes (see below) may be similar across taxa. Interestingly, this comes in contrast to work done on mammalian mitochondrial DNA by Galtier et al. (2006) in which the location of mutational hot spots appears to be inconstant over evolutionary time.

The maximum likelihood HKA-test indicates that the mitochondrial genes vary in their scaled mutation rate, with separate estimates of θ leading to a significantly better model fit than a single genome-wide estimate. In addition, some of the genes vary significantly from each other in estimates of θ based on the maximum-likelihood coalescent recursions. The significantly positive relationship between the frequency of silent segregating sites within *S. vulgaris*, and the proportion of silent fixed differences between *S. vulgaris* and its outgroup *S. paradoxa* provides further and substantial evidence that variation in silent site polymorphism among genes is a function of variation in underlying mutation rate. These results support the idea that there is not a time-calibrated molecular clock for the mitochondrial genome (Muse 2000).

The tests for recombination gave mixed results, which is not an uncommon outcome (Posada 2002; Piganeau and Eyre-Walker 2004). RDP and MaxChi, the two tests determined by Posada (2002) and Posada and Crandall (2001) to have the highest power of detecting recombination and to yield the fewest false positives, came to opposing conclusions on the full dataset, with RDP finding no instances of recombination, and MaxChi finding 22. In a dataset of only third position sites, MaxChi detects no incidences of recombination. The 4-gamete test detected 3 to 5 pairs of sites of potential recombination. However, there are no obvious blocks of recombinants, which might have been expected under a model of infrequent homologous recombination (table 4).

These results emphasize a critical point that has often been overlooked in studies attempting to detect recombination: all of the tests used, which represent the variety of possible tests for recombination, contain assumptions that may be violated by plant mitochondria. These include the infinite sites model, single gene copies, and no mutation rate variation among sites (Posada 2002; Piganeau and Eyre-Walker 2004). Our data clearly show that mutation rate variation may occur among genes, and thus likely among sites, and that at least several of the genes have some presence of multiple copies. In addition to violations of critical assumptions, all of the tests are sensitive to various other properties of the data such as the degree of divergence between the sequences, the amount of recombination that has occurred, the type of recombination that has occurred, and the presence of low frequency alleles (Posada 2002). We conclude that it is possible that recombination may be occurring in plant mitochondria, but given (1) the conflicting results obtained by different tests for recombination, (2) the sensitivity of current tests to multiple assumptions and (3) our knowledge of how plant mitochondria might violate these assumptions, we argue that the evidence is not sufficient to definitively conclude that the mitochondrial genes in our dataset on *S. vulgaris* are recombining.

It is not likely that the majority of the total or silent site variation in polymorphism among genes is due to differences in selective effects acting directly on the genes, as neither the multilocus HKA nor McDonald-Kreitman tests were significant. There was marginal evidence for purifying or balancing selection on cox3 (though not significant after correcting for multiple comparisons). It is not obvious why cox3 might be different from the other genes, because all of them (except for the non-coding nad4L-atp4) code for important components of oxidative phosphorylation (Wolstenholme and Fauron 1995). In addition, although our data do not rule out the possibility of mitochondrial recombination, it is probable that it is rare, and the effect of differential selection on individual genes should decrease with decreasing recombination rate.

It is important to emphasize that our data cannot reject the hypothesis that selection is an important factor influencing the level of polymorphism in mitochondrial genes, only that selection is not necessary to explain the observed variation in levels of polymorphism between loci. Selection could contribute some additional among-gene variation, as suggested by the different relationship between polymorphism and divergence for replacement versus silent sites, although the McDonald-Kreitman test indicates that even with all sites considered, there is not significant evidence of selection. In addition, our analyses cannot speak to any effect of selection that operates equally across all loci, as might be expected in a genome with no or infrequent recombination. We are also not able to comment on whether polymorphism in the mitochondrial genome is low or high, because we lack an appropriate null model. A more powerful study would compare data from multiple loci with a comparative approach. In that case, mitochondrial diversity could be compared among species that might

Table 4
Segregating Sites from all Concatenated Genes Showing Evidence of Four Gametes Between Five Pairs of Sites.

site	18	21	684	1206	2167	3265	3898	3905
1v	C	G	A	G	G	G	A	C
5v	C	G	A	G	T	G	A	C
20v	C	G	A	G	G	G	A	C
26v	C	G	A	G	T	A	A	C
27v	C	G	G	T	G	G	A	C
30v	C	G	G	T	T	A	A	A
31v	C	G	A	G	G	G	A	C
32v	C	G	A	G	G	G	A	C
33v	G	G	A	T	T	G	A	C
35v	C	G	A	G	T	G	A	C
36v	C	G	G	T	T	A	A	A
37v	C	T	A	T	T	G	T	C
38v	C	G	A	G	T	A	A	A
39v	C	G	A	G	G	G	A	C
41v	C	G	A	G	T	G	A	C
42v	C	G	A	G	G	G	A	C
43v	C	G	A	G	G	G	A	C
45v	G	T	G	T	T	A	A	A
46v	C	G	G	T	T	A	A	A
47v	C	G	A	G	G	G	A	C
sv03	C	G	G	T	T	A	T	A
sv04	C	G	A	G	G	G	A	C
sv06	C	G	A	G	G	G	A	C
sv08	C	G	G	T	T	A	T	A
sv10	C	G	G	T	T	G	A	C

NOTE.—Boxes between pairs of sites indicate each gametic type. Sites with a low frequency minority allele are 18 and 21, and removal of these sites reduces the minimum number of recombination events to three.

reasonably be expected to experience different histories of selection (c.f. Ingvarsson and Taylor 2002, although there will necessarily be practical problems associated with such an approach as outlined in Ingvarsson 2004).

When combined, these data provide compelling evidence that variation in mutation rates among genes is sufficient to explain the variation in polymorphism among plant mitochondrial genes. Variation in mutation rate is thought to explain the variation in substitution rates between chloroplast, mitochondrial, and nuclear genomes (Yang et al. 1999), and is also found in mammalian genomes (Wolfe et al. 1989, Galtier 2006). Wolfe et al. (1989) proposed that variation in mutation rate of mammalian nuclear genomes is due to differences in timing of replication along the genome; it would be interesting to see if this mechanism can apply to a system with a circular or concatenated genome. Laroche et al. (1997) found that mitochondrial genes varied more in replacement site substitution rates than silent site rates, which they interpret as potential evidence of differential selection, assuming constant mutation rates. Our likelihood results rejecting a single genome-wide value of θ, the significantly positive relationship between divergence and polymorphism among genes, and the fact that the relative order of higher to lower substitution rates approximately matches that of Laroche et al. (1997), suggest that mutation rate variation is a major determinant of within-species polymorphism.

Previous studies of plant mitochondrial sequence polymorphism generally invoked selection to explain patterns of polymorphism. These studies often focused on the selective dynamics driven by a class of selfish mitochondrial elements known as cytoplasmic male sterility (CMS) (Ingvarsson and Taylor 2002, Stadler and Delph 2002). These selfish genes alter the breeding system in the affected species by eliminating male fitness through pollen. If this type of CMS-driven selection is occurring, and there are significant differences in polymorphism among genes, then recombination must also occur frequently enough to allow for independent evolution of mitochondrial genes. We find unconvincing evidence for recombination in our sample, and no evidence of selection on any of the genes studied, and generally confirm the standard model of plant mitochondria characterized by low substitution rates and no recombination. We show that for *S. vulgaris*, a species known to have CMS, mutation rate variation is sufficient to explain observed mitochondrial polymorphism.

Acknowledgments

We are very grateful to Whit Farnum, Ellen McRae, and Rachel Prunier for dedicated assistance with sequencing and data analysis. We would also like to thank Spencer Muse, Nicolas Galtier, and all members of the Taylor lab for helpful comments. This work was funded by the National Science Foundation DBI-0305927 to CMB and DEB-0349558 to DRT.

Literature Cited

Adams KL, Qui Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci. 99:9905–9912.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 168:39–50.

Bazin E, Glemin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science. 312:570–572.

Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res. 63:213–227.

Charlesworth D. 2002. What maintains male-sterility factors in plant populations? Heredity. 89:408–409.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci. 101: 17741–17746.

Comeron JM. 2006. Weak selection and recent mutational changes influence polymorphic synonymous mutations in humans. Proc Natl Acad Sci. 103:6940–6945.

Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M. 2001. Mechanisms for evolving hypervariability: the case of conopeptides. Mol Biol Evol. 18:120–131.

Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V. 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. Proc Natl Acad Sci. 101:1904–1909.

Duminil J, Pemonge M-H, Petit RJ. 2002. A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. Mol Ecol Notes. 2:428–430.

Endo T, Ikeo K, Gojobori T. 1996. Large-scale search for genes on which positive selection may operate. Mol Biol Evol. 13:685–690.

Eyre-Walker A, Smith NH, Maynard Smith J. 1999. How clonal are human mitochondria? Proc R Soc Lond., B. 266:477–483.

Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. 2006. Mutation hot spots in mammalian mitochondrial DNA. Genome Res. 16:215–222.

Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK, MacIntyre RJ, Clegg MT, editors. Evolutionary Biology. New York: Plenum Press. Vol. 30p. 93–120.

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics. 16:573–582.

Hudson RR. 1991. Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J, editors. Oxford Surveys of Evolutionary Biology. Oxford: Oxford University Press. Vol. 7p. 1–44.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 111:147–164.

Ingvarsson PK. 2004. Population subdivision and the Hudson-Kreitman-Aguade test: testing for deviations from the neutral model in organelle genomes. Genet Res. Camb. 83:31–39.

Ingvarsson PK, Taylor DR. 2002. Genealogical evidence for epidemics of selfish genes. Proc Natl Acad Sci. 99:11265–11269.

Jolley KA, Feil EJ, Chan MS, Maiden MC. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics. 17:1230–1231.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian Protein Metabolism. New York: Academic Press. p. 21–132.

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNACys(GCA). Nucleic Acids Res. 28:2571–2576.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. Briefings in Bioinformatics. 5:150–163.

Laroche J, Li P, Maggia L, Bousquet J. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci. 94:5722–5727.

Martin DP, Williamson C, Posada D. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics. 21:260–262.

Maynard Smith J. 1992. Analyzing the mosaic structure of genes. J Mol Evol. 34:126–129.

Maynard Smith J, Smith NH. 1998. Detecting recombination from gene trees. Mol Biol Evol. 15:590–599.

McCauley DE, Bailey MF, Sherman NA, Darnell MZ. 2005. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. Heredity. 95:50–58.

McDonald J, Kreitman M. 1991. Adaptive protein evolution at *adh* locus in *Drosophila*. Nature. 351:652–654.

McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 160:1231–1241.

Muse SV. 2000. Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol. 42:25–43.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418–426.

Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. Heredity. 86:641–647.

Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell. 66:473–481.

Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology. 265:218–225.

Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 28:87–97.

Parkinson CL, Mower JP, Qiu Y-L, Shirk AJ, Song K, Young ND, dePamphilis CW, Palmer JD. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol. 5:73–84.

Piganeau G, Eyre-Walker A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. Heredity. 92:282–288.

Piganeau G, Gardner M, Eyre-Walker A. 2004. A broad survey of recombination in animal mitochondria. Mol Biol Evol. 21:2319–2325.

Posada D. 2002. Evaluation of methods for detecting recombination from DNA sequences: Empirical Data. Mol Biol Evol. 19:708–717.

Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc Natl Acad Sci. 98:13757–13762.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Salminen MO, Carr JK, Burke DS, McCutchan FE. 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning. AIDS Res Hum Retroviruses. 11:1423–1425.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. 2000. Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. Bot J Linn Soc. 133:381–461.

Stadler T, Delph LF. 2002. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. Proc Natl Acad Sci. 99:11730–11735.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. Plant Mol Biol. 17:1105–1109.

Taylor DR, Keller SR. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. Evolution. 61:334–345.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Pop Biol. 7: 256–276.

Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast. nuclear DNAs. Proc. Natl. Acad. Sci. 84:9054–9058.

Wolfe KH, Li W-H, Sharp PM. 1989. Mutation rates differ among regions of the mammalian genome. Nature. 0: 283–285.

Wolstenholme DR, Fauron CM-R. 1995. Mitochondrial genome organization. In: Levings CS III, Vasil IK, editors. The Molecular Biology of Plant Mitochondria. The Netherlands: Kluwer Academic Publishers. p. 1–59.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics. 168:1071–1076.

Yang Y-W, Lai K-N, Tai P-Y, Li W-H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J Mol Evol. 48:597–604.

Spencer V. Muse, Associate Editor

Accepted May 11, 2007