

Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*

STEPHEN I. WRIGHT*, NARDIN NANO, JOHN PAUL FOXE
AND VAQAAR-UN NISA DAR

Department of Biology, York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3

(Received 14 August 2007 and in revised form 3 October 2007)

Summary

Cytoplasmic genomes typically lack recombination, implying that genetic hitch-hiking could be a predominant force structuring nucleotide polymorphism in the chloroplast and mitochondria. We test this hypothesis by analysing nucleotide polymorphism data at 28 loci across the chloroplast and mitochondria of the outcrossing plant *Arabidopsis lyrata*, and compare patterns with multiple nuclear loci, and the highly selfing *Arabidopsis thaliana*. The maximum likelihood estimate of the ratio of effective population size at cytoplasmic relative to nuclear genes in *A. lyrata* does not depart from the neutral expectation of 0.5. Similarly, the ratio of effective size in *A. thaliana* is close to unity, the neutral expectation for a highly selfing species. The results are thus consistent with neutral organelle polymorphism in these species or with comparable effects of hitch-hiking in both cytoplasmic and nuclear genes, in contrast to the results of recent studies on gynodioecious taxa. The four-gamete test and composite likelihood estimation provide evidence for very low levels of recombination in the organelles of *A. lyrata*, although permutation tests do not suggest that adjacent polymorphic sites are more closely linked than more distant sites across the two genomes, suggesting that mutation hotspots or very low rates of gene conversion could explain the data.

1. Introduction

A central prediction from population genetics theory is that natural selection can have an important influence on levels of linked variation, a process referred to as hitch-hiking. Negative ('background selection') (Charlesworth *et al.*, 1993) and positive ('selective sweeps') (Maynard Smith & Haigh, 1974) selection both reduce the effective population size of linked sites, reducing neutral variation. Conversely, balancing selection can lead to elevated variation due to the long-term maintenance of selected alleles (Charlesworth *et al.*, 1997). Because the effects of selection on linked variation depend on the amount of linkage disequilibrium between selected and neutral sites, the effects on genetic variation will be more pronounced in genomic regions, or genomes, experiencing low levels of recombination (Braverman *et al.*, 1995; Nordborg *et al.*, 1996). Evidence for a

suppression of diversity in regions of low recombination (Begun & Aquadro, 1992), regions of high gene density per unit of recombination (Nordborg *et al.*, 2005) and self-fertilizing species (Glemin *et al.*, 2006) are broadly consistent with these predictions.

The organelle genomes of plants and animals are generally considered to be non-recombining (Barr *et al.*, 2005), experiencing uniparental inheritance as an effectively haploid genome (Birky, 1995; Birky *et al.*, 1983). Because of this, the action of natural selection at any site should influence patterns of variation genome-wide, leading to departures from neutral expectation across the genome. In most plants, predominant maternal inheritance of both the mitochondria and the chloroplast imply that they should generally be inherited as a linked unit and that, if frequent enough, selection acting at any region of these organelles can have an important influence on the amount of genetic variation.

To date, few studies have tested the assumptions of neutral equilibrium in organelle genomes. In

* Author for correspondence. Telephone: +1 (416) 7362100, ext. 20213. Fax: +1 (416) 7365698. e-mail: Stephenw@yorku.ca

Drosophila, several studies indicate a reduction in nucleotide variation in mitochondrial genomes relative to the nucleus, controlling for expected differences in effective population size due to maternal inheritance (Ballard *et al.*, 1996; Ballard & Kreitman, 1994). These patterns have been interpreted as reflecting the presence of selective sweeps of mitochondrial variation due to the recent fixation of advantageous mutations. Recent studies have also shown signatures of selection on mitochondria in other insects (Jiggins, 2003; Jiggins & Tinsley, 2005) and these have been shown to be associated with male-killing bacteria which are inherited maternally. In humans, mitochondria show an excess of low-frequency variants, in contrast to the nuclear genome (Hey, 1997). Although this difference could reflect the action of recent positive selection on the mitochondria, a population bottleneck in the species' history could also explain the data, since genomes with contrasting effective size will be at different stages of recovery from demographic events (Fay & Wu, 1999). Recently, a meta-analysis of animal polymorphism datasets suggested a general lack of correlation between nuclear and cytoplasmic diversity (Bazin *et al.*, 2006), and this was interpreted as the predominance of hitch-hiking on mitochondrial genomes (but see Wares *et al.*, 2006).

In plants, organelle genomes experience low mutation rates (Wolfe *et al.*, 1987; but see Cho *et al.*, 2004; Barr *et al.*, 2007). This has limited studies of direct nucleotide sequence variation (although restriction fragment length polymorphism in chloroplasts is often used in studies of phylogeography), and almost no data are available directly comparing nucleotide polymorphism relative to divergence at organelle versus nuclear genes. Because the mitochondria and chloroplast should behave as effectively haploid genomes, the effective population size (treated here as the effective number of gene copies, not individuals) should be half that of the nuclear genome in an outcrossing hermaphrodite, in the absence of natural selection (Birky *et al.*, 1983). However, this prediction assumes equal effective population sizes through male and female function; if the variance in fitness differs between male and female function, or migration rates through pollen and seed are different, this prediction may not hold (Laporte & Charlesworth, 2002).

In highly selfing organisms, the effective population size of the nuclear genome is reduced by homozygosity (Nordborg & Donnelly, 1997; Pollak, 1987), and under neutrality this would equalize effective sizes of the organelles and nuclear genes. Thus, under neutrality, the effective population size in selfing populations is reduced at nuclear but not cytoplasmic genomes relative to outcrossers. However, if selfing populations experience more frequent population

bottlenecks than outcrossers, both organelle and nuclear variation could be reduced in highly selfing taxa (Fenster & Ritland, 1992; Graustein *et al.*, 2002). On the other hand, nuclear genes should experience greater hitch-hiking effects in selfing taxa compared with outcrossing relatives due to lower effective recombination rates (Charlesworth & Wright, 2001), and the net result on diversity at cytoplasmic versus nuclear genes will depend on the relative rates and distributions of fitness effects of mutations, which will determine the strength of the hitch-hiking effects.

The low rates of substitution may imply that plant organelles are unlikely to experience frequent hitch-hiking, due to insufficient rates of deleterious and advantageous mutation. However, despite low nucleotide substitution rates, plant mitochondria appear to experience high rates of genome rearrangement (Palmer *et al.*, 2000), and it is possible that many of these events are subject to both frequent negative and positive selection. Indeed, many mitochondrial rearrangements have been identified that are associated with cytoplasmic male sterility (Schnable & Wise, 1998), and they can be under strong directional or balancing selection in plant populations (Saur Jacobs & Wade, 2003). Similarly, gene order has been inferred to be subject to positive and negative selection in the chloroplast (Cui *et al.*, 2006). If genome rearrangements are common and subject to natural selection, this can leave important signatures on patterns of nucleotide variation.

Despite the low rates of substitution, a number of recent studies have found evidence for significant levels of polymorphism in plant mitochondria and chloroplast genomes (Houlston & Olson, 2006; Stadler & Delph, 2002). These studies have focused on gynodioecious species that maintain polymorphisms for cytoplasmic male sterility (CMS). The patterns of nucleotide variation are suggestive of the action of long-term balancing selection associated with this sexual system, potentially reflecting frequency-dependent selection on distinct male sterility cytotypes (but see Barr *et al.*, 2007). In contrast, a recent survey of chloroplast variation in dioecious *Silene* (Muir & Filatov, 2007) suggested a reduction in chloroplast variation consistent with a selective sweep. However, few comparative data have been available from hermaphroditic taxa, and variation at multiple nuclear genes has not been compared directly with variation at organelle genes, to uncouple effects of species and population history from specific effects on cytoplasmic variation.

Despite the predominance of maternal inheritance in most plants, another open question concerns the extent of recombination occurring in cytoplasmic genomes. Paternal leakage of mitochondria and chloroplasts may be common in plants (Barr *et al.*, 2005), and recent studies using quantitative PCR

suggest that this can have significant effects on organelle genotypes in natural populations (Welch *et al.*, 2006). Paternal leakage can generate heteroplasmy, or the presence of multiple cytotypes in a single individual, and if recombination occurs in heteroplasmic individuals this can act to break down linkage disequilibrium within and between the organelle genomes. Recent analyses of polymorphism levels of chloroplast and mitochondria have provided evidence for some historical recombination in plants (Houliston & Olson, 2006; Marshall *et al.*, 2001; Stadler & Delph, 2002), although the overall frequency of recombination in organelles remains poorly characterized.

In this study we analyse nucleotide polymorphism in the chloroplast and mitochondria of the predominantly self-incompatible, outcrossing plant *Arabidopsis lyrata*, in order to quantify the effective population size and test for selection acting on these genomes. *Arabidopsis lyrata* has a circumpolar distribution, and is characterized by high levels of nucleotide polymorphism at nuclear genes (Savolainen *et al.*, 2000; Wright *et al.*, 2003; Ramos-Onsins *et al.*, 2004). Studies to date suggest elevated levels of nucleotide diversity at nuclear genes in *A. lyrata* compared with the highly selfing model plant *A. thaliana* (Savolainen *et al.*, 2000; Wright *et al.*, 2003; Ramos-Onsins *et al.*, 2004), although the relative role of population history and selection in causing these differences remains unclear. We make use of a polymorphism dataset from 53 nuclear genes in *A. lyrata* (Wright *et al.*, 2006; S. I. Wright, J. Ross-Ibarra, J. P. Foxe, A. Kawabe, L. DeRose Wilson, G. Gos, D. Charlesworth & B. S. Gaut, submitted) to quantify the relative effective population size of the organelles. We also make use of published datasets on nuclear (Nordborg *et al.*, 2005) and chloroplast (Jakobsson *et al.*, 2007) variation in the highly selfing *Arabidopsis thaliana*, to test for effects of selfing on the relative effective population sizes.

2. Materials and methods

(i) Samples

A total of 24 individuals from distinct maternal families, from four populations of *Arabidopsis lyrata*, were used in the initial survey: 12 individuals from Plech, Germany (from T. Mitchell Olds), 4 individuals from Rondeau Provincial Park, Ontario, Canada (collected by B. Mable), 4 individuals from Indiana Dunes, Indiana, USA (collected by B. Mable) and 4 individuals from Stubbsand Sweden (collected by O. Savolainen). Our sampling emphasized the Plech population, since this population has been shown to be the most highly polymorphic (Wright *et al.*, 2006; Wright *et al.*, submitted), and may be closest to

demographic equilibrium (Clauss & Mitchell-Olds, 2006), while the other populations may have experienced population bottlenecks associated with post-glacial recolonization (Wright *et al.*, 2006; Wright *et al.*, submitted). It should also be noted that a breakdown of self-incompatibility and high selfing rates have been documented in the Rondeau population (Mable *et al.*, 2005). Possible effects of this breakdown on our results are addressed in Section 4. In addition, to extend the analysis of polymorphic sites to further test for evidence of recombination and population subdivision, we sequenced all polymorphic loci (see Table 1) in three additional individuals from Sweden, and eight individuals from Karhumäki Russia (collected by O. Savolainen). Because these samples were only sequenced in a subset of the polymorphic loci for the recombination and differentiation analysis, they do not give an unbiased estimate of overall diversity, and are not included in the comparison of polymorphism levels. DNA from a single plant per maternal family was extracted using the FastDNA kit (MP Biomedicals, Solon, OH, USA) and used as template for PCR amplification. PCR conditions consisted of 35 cycles of 20 s at 94 °C, 20 s at 55 °C, 40 s at 72 °C, followed by a 4 min elongation at 72 °C.

(ii) Loci and PCR amplification

Primers were designed using PrimerQuest (Integrated DNA Technologies, <http://biotools.idtdna.com/primerquest/>) to amplify a total of 12 chloroplast fragments and 16 mitochondrial fragments, with product sizes ranging from 500 to 700 bp, using the complete organelle genomes of *Arabidopsis thaliana* (chloroplast, GenBank accession number AP000423, mitochondria, GenBank accession number Y08501). Loci include both coding and non-coding genomic regions. A summary of the lengths, positions, primer sequences and annotation of the regions is given in Supplementary Table S1. PCR products were sequenced directly on both strands by Cogenics (Houston, TX, USA).

(iii) Analysis

Summaries of polymorphism data, including Watterson's (1975) estimator of $\theta = 4N_e u$, where N_e is the effective population size and u is the mutation rate; π the average pairwise differences (Tajima, 1989); R_m , the minimum number of recombination events (Hudson & Kaplan, 1985), Wright's (1951) F_{st} , and Tajima's (1989) D were calculated using DNAsp version 4.10 (Rozas *et al.*, 2003). Maximum likelihood estimates of the population parameter $\theta = 4N_e u$ given the number of segregating sites were calculated using the equations given in Hudson (1991) as

Table 1. Table of polymorphic sites in *A. lyrata*

position	10973	16609	23695	24037	40641	41058	49239	49304	49414	43303	49462	218912	274101	274107	299238	299402
locus	c2A	C2B	c3A	c3A	c5A	c5A	c5B	c5B	c5B	m5B	m5A	m22	m27	m27	m30	m30
	syn	rep	rep	syn	syn	syn	inter	rep	rep	non	non	syn	non	non	non	non
thal	G	G	A	C	T	A	G	A	C	C	G	G	T	G	C	G
GER1	.	.	.	T	mv	mv	T	G	.	.	.
GER2	.	.	.	T	A	C	T	G	.	.	.
GER3	.	.	.	T	A	C	T	G	.	.	.
GER4	.	.	.	T	A	C	T	G	.	.	.
GER5	.	.	.	T	mv	mv	mv	mv	mv	.	.	.	G	.	.	.
GER6	.	.	.	T	.	C	T	G	.	.	.
GER7	.	.	.	T	A	C	T	G	.	.	.
GER8	.	.	.	T	mv	mv	mv	mv	mv	.	.	.	G	.	.	.
GER9	.	.	.	T	.	C	T	G	.	.	.
GER10	.	.	.	T	mv	mv	mv	mv	mv	.	.	.	G	.	A	T
GER11	.	.	.	T	.	.	.	T	.	T	.	.	G	.	A	T
GER12	.	.	.	T	A	C	T	G	.	.	.
CAN1	T	T	.	T	.	C	T	G	.	.	.
CAN2	T	.	.	T	.	C	T	T	G	.	.	.
CAN3	T	.	.	T	.	C	T	G	.	.	.
CAN4	T	T	.	T	.	C	T	.	.	.	mv	T	G	.	.	.
USA1	T	.	.	T	.	C	T	.	T	.	.	.	G	.	.	.
USA2	T	.	.	T	.	C	T	.	T	.	T	.	G	.	.	.
USA3	T	.	.	T	.	C	T	T	G	.	.	.
USA4	T	.	.	T	.	C	T	T	G	.	.	.
SWE1	.	.	G	T	.	T	.	.	G	.	A	T
SWE2	.	.	G	T	.	T	.	.	G	.	A	T
SWE3	.	.	G	T	.	T	.	.	G	.	A	T
SWE4	.	.	G	T	.	T	.	.	G	.	A	T
SWE5	.	.	G	T	.	T	.	.	G	.	A	T
SWE6	.	.	G	T	.	T	.	.	G	.	A	T
SWE7	.	.	G	T	.	T	.	.	G	.	A	T
SWE8	mv	mv	mv	mv	.	.	.	T	.	T	.	.	G	.	A	T
RUS1	G	T	.	.
RUS2	G	T	.	.
RUS3	G	T	.	.
RUS4	G	T	.	.
RUS5	G	T	.	.
RUS6	G	T	.	.
RUS7	G	T	.	.
RUS8	G	T	.	.

‘.’ indicates nucleotides identical to the top sequence, shown as the outgroup base from *A. thaliana*; ‘mv’ indicates missing values. Populations correspond to Plech, Germany (GER); Rondeau, Canada (CAN); Indiana Dunes, USA (USA), Stubbsand, Sweden (SWE), and Karhumaki, Russia (RUS).
syn, synonymous; rep, replacement, non, non-coding.

described by Wright and colleagues (2003). In addition, we used a maximum likelihood implementation of the HKA test, mlHKA (Wright & Charlesworth, 2004), to estimate the ratio of effective population sizes in the cytoplasm relative to the nucleus (N_c/N_n). Note that for the purposes of this study we considered N_n as the effective number of nuclear gene copies (not individuals), such that the neutral expectations of N_c/N_n are 0.5 under complete outcrossing, and 1 under complete selfing for a hermaphroditic plant. We used polymorphism and divergence data from a comparable multi-population sample of 53 nuclear loci from large exons (Wright *et al.*, 2006; Wright *et al.*, submitted) combined with the observed polymorphism and divergence data from the chloroplast and mitochondria. The published

nuclear sequence data were surveyed in a total of 71 diploid individuals (up to 142 chromosomes) from the same populations surveyed here, as well as a population from Esja Mountain, Iceland. Given the comparable patterns of species-wide nuclear diversity with the inclusion versus exclusion of Iceland (data not shown), the addition of another population in the nuclear dataset is unlikely to bias our results, and was retained to increase power for estimating N_c/N_n . However, we also explored within-population estimates of N_c/N_n for the German population and the pooled North American population, where sufficient sampling of individuals and samples allow. North American populations show relatively low levels of differentiation, and were pooled to allow sufficient power to estimate N_c/N_n , although it should

be noted that we are pooling across populations with distinct outcrossing rates, as addressed in Section 4.

We conducted mHKA analysis treating the chloroplast, mitochondria and the combined data as single non-recombining datasets, pooling the silent sites across loci, and using the minimum sample sizes as input. Nuclear loci were treated as independent, unlinked loci in the analysis. To avoid upward bias in estimates due to the use of minimum sample sizes for the chloroplast and mitochondrial data, we excluded loci with low sample sizes in the mHKA analysis. For the species-wide dataset, we excluded loci with sample sizes less than 18. For individual populations, we used cutoffs of 10 for the German population, and 7 for the pooled North American populations. Divergence values were corrected for multiple hits using the Jukes & Cantor (1969) correction for multiple substitutions. We generated a likelihood surface for N_c/N_n by varying the inheritance scalar for the organelle data (i.e. varying assumptions about the ratio of effective sizes in organelles versus the nucleus). Credibility intervals were obtained by assuming the X^2 distribution.

(iv) *Arabidopsis thaliana* datasets

We also applied mHKA to the estimation of N_c/N_n for *Arabidopsis thaliana*, using multilocus polymorphism data from nuclear genes (Nordborg *et al.*, 2005) and a total of 50 fragments from the chloroplast (Jakobsson *et al.*, 2007). For an equivalent comparison, we focused on synonymous diversity from the nuclear loci, since *A. lyrata* nuclear data were taken only from exons. In order to get divergence data for these fragments, nuclear and chloroplast fragments were submitted to a BLAST (Altschul *et al.*, 1997) search against the shotgun genome sequence of *A. lyrata* (<http://www.jgi.doe.gov/sequencing>) using the trace archive from NCBI (www.ncbi.nlm.nih.gov). Consensus sequences from multiple shotgun trace files were assembled from the *A. lyrata* genome dataset, and used to calculate divergence for mHKA analysis. We only retained loci where *A. lyrata* alignments included multiple shotgun trace files, using a cutoff of 10 bp in the alignment where only a single trace file was used in divergence calculation. In the interest of computation time, a subset of 99 nuclear loci was used from the survey of polymorphism at nuclear genes for the maximum likelihood analysis, chosen to have the largest numbers of total synonymous sites, to increase power in the estimation of N_c/N_n .

(v) Analysis of recombination

We used the program LDHAT 2.0 (<http://www.stats.ox.ac.uk/~mcvean/LDhat>) to estimate the

population recombination parameter $4N_er$, where r is the rate of recombination. LDHAT implements the composite likelihood approach of Hudson (2001), allowing a finite sites model (McVean *et al.*, 2002). Since the appropriate recombination model for organelles should be gene conversion (McVean *et al.*, 2002), we implemented a gene conversion model with a tract length of 300 bp. Distances between polymorphic sites were taken as the minimum distance in the circular chromosomes (McVean *et al.*, 2002), using the *A. thaliana* genome. We also used LDHAT for likelihood permutation tests of recombination, permuting the locations of segregating sites, to test for significantly greater evidence for recombination in the observed data than data permuted by location.

3. Results

(i) Patterns of polymorphism

We surveyed nucleotide variation in a total of 5827 bp (3.8%) of the chloroplast genome and 8445 bp (2.3%) of the mitochondrial genome of *A. lyrata*. Of the loci surveyed, five of 13 (38%) showed variation in the chloroplast, while five of 16 (31%) mitochondrial loci were variable. Furthermore, polymorphic loci showed uniformly low levels of variation; loci showed at most two segregating silent sites, and no highly polymorphic loci were observed (Supplementary Table S2). We identified a total of 16 segregating nucleotide sites (Table 1), five of which were synonymous, seven of which were non-coding and four of which were replacement. No indels were identified in our dataset.

Patterns of variation across populations generally conform to results from nuclear loci (Wright *et al.*, 2006; Wright *et al.*, submitted), namely reduced variation in North American populations and highest diversity in Germany (Table 1; Supplementary Table S2). However, levels of differentiation are very high; there are no shared polymorphisms across any populations with the exception of the United States and Canada, and much of the species-wide variation is characterized by fixed or nearly fixed differences between populations. The total level of differentiation, estimated by F_{st} , is 0.824, which is greater than for any nuclear locus examined in a similar population dataset (average F_{st} for nuclear genes is 0.6; Wright *et al.*, submitted), suggesting significantly elevated differentiation at cytoplasmic relative to nuclear genes. Interestingly, no chloroplast or mitochondrial variation is observed in the Swedish and Russian populations at the loci surveyed here (Table 1), although these populations were sampled only at polymorphic loci (see Section 2). At nuclear loci, these latter two populations have higher diversity

than North American populations, but lower than the German population (Wright *et al.*, 2006).

The maximum likelihood estimates of the species-wide population parameter $\theta = 4N_e u$ at the mitochondria and the chloroplast from the number of silent segregating sites are 0.00035 (95% credibility interval 0.00011 to 0.00097) and 0.001 (95% credibility interval 0.00027 to 0.00287), respectively. In contrast, the estimate of θ from the species-wide nuclear genome data is 0.02, approximately 60-fold higher than the mitochondria and 20-fold higher than the chloroplast. Calculation of Tajima's D (Tajima, 1989), a measure of the site frequency spectrum, from the combined data provides no evidence for a significant departure from neutral expectation for the species-wide sample ($D = 1.41$, $P > 0.05$), although the data show a positive skew in the frequency spectrum primarily due to sampling of fixed differences between populations (Table 1). From the nuclear genes, 18% of loci show values of Tajima's D as high as observed here, providing no evidence that this value is unusual compared with the genome average. In Germany, Tajima's D shows a non-significant negative skew ($D = -1.17$), while in the pooled North American sample Tajima's D is 0.889. Thus, no population shows evidence for an excess of rare or high-frequency variants expected under recent positive selection or long-term balancing selection, respectively.

(ii) Ratio of effective population size in *A. lyrata*

The above differences in polymorphism across the *Arabidopsis lyrata* cytoplasmic and nuclear genomes can result from differences in both mutation rates and effective population sizes. To uncouple these, we used a maximum likelihood analysis controlling for divergence between *A. lyrata* and *A. thaliana*. Fig. 1 shows estimates of the ratio of effective population sizes at the organelle versus nuclear genomes, N_c/N_n for the species-wide dataset, as well as for Plech Germany, and the North American populations pooled as a single sample. Maximum likelihood estimates from the species-wide datasets follow the neutral expectations for both organelles individually, as well as the combined dataset; in all cases, the maximum likelihood estimate of the scaled effective population size is 0.4, and shows no significant departure from the neutral expectation of 0.5. Similar patterns are observed for the German population, although the maximum likelihood estimate is slightly lower at 0.3 for the combined data, but not significantly different from 0.5. Maximum likelihood estimates of N_c/N_n were non-significantly reduced for the North American chloroplast sample, due to the lack of silent variation identified at the chloroplast loci surveyed, although the likelihood surface is flat due to an

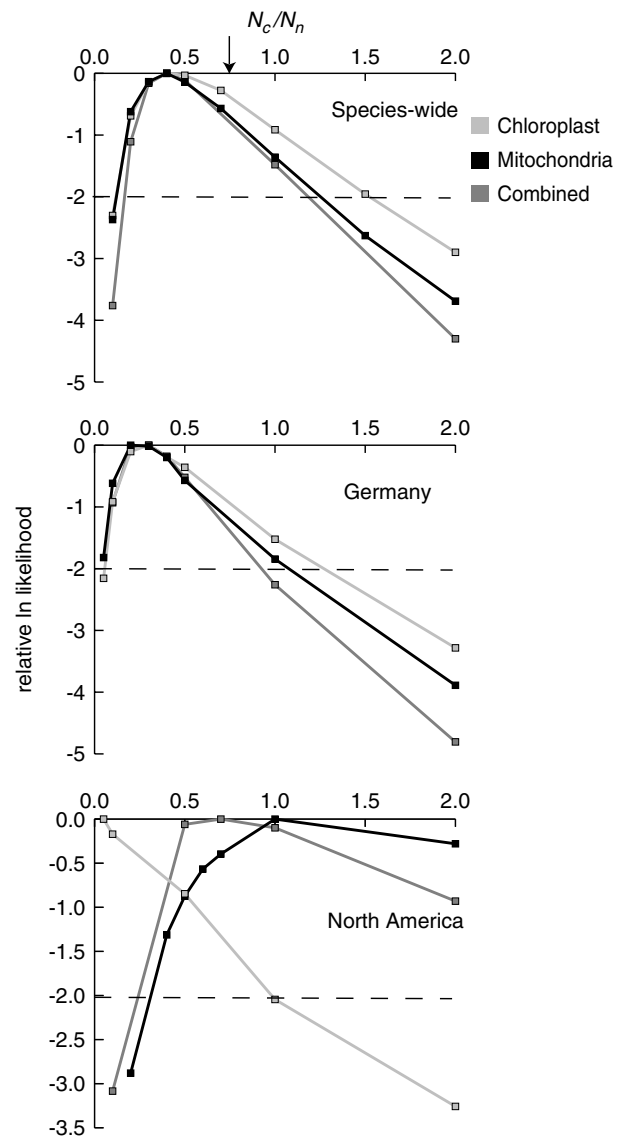


Fig. 1. Likelihood estimation of the ratio of effective population size of the cytoplasm relative to the nuclear genome (N_c/N_n) in *Arabidopsis lyrata*, using mlHKA analysis. Y-axis shows the log likelihood of the data for a given value of N_c/N_n subtracted from the maximum log likelihood. Dashed line shows the 95% credibility interval. Arrow shows the maximum likelihood estimate from *A. thaliana*.

overall low nucleotide variation in this population. Mitochondrial loci give an elevated estimate of N_c/N_n in North America, and the combined data give a maximum likelihood estimate of $N_c/N_n = 0.6$. In summary, the ratio of effective population sizes appears to be entirely consistent with neutral expectations in our *A. lyrata* dataset.

The maximum likelihood analysis also allows for a comparison of mutation rate, controlling for differences in effective population size. Our estimate of the scaled mutation rates, taking the maximum likelihood value of θ from the nuclear genes, is 8-fold elevated in the nucleus relative to the chloroplast, which is a

considerably larger difference than the 2-fold estimated by Wolfe and colleagues (1987). Our chloroplast dataset includes four loci from the inverted repeat region of the genome, which has been shown to have suppressed substitution rates, and this may in part drive our mutation rate estimates downward. Note that this should not bias our effective size estimation, since this is scaled to divergence between species. Using estimates of substitution rate from nuclear genes (1.5×10^{-8} per year) in the Brassicaceae (Koch *et al.*, 2000), this would suggest that the chloroplast has a substitution rate of 1.7×10^{-9} per year on average. For the mitochondria, maximum likelihood analysis estimates a 23-fold reduction in mutation rate relative to the nuclear genome, giving a substitution rate estimate of 6.5×10^{-10} per base pair per year for the mitochondria. Again, this value is considerably lower than the 6-fold reduction suggested by Wolfe and colleagues (1987), but both chloroplast and mitochondrial values are closer to relative rate estimates from polymorphism data in rice (Tian *et al.*, 2006). Although the latter estimates are not controlling for effective population size differences, domesticated rice is highly selfing, so these ratios may be comparable to differences in mutation rate.

(iii) Comparisons with *A. thaliana*

Maximum likelihood estimates of silent θ from *A. thaliana* are 0.00123 (95% credibility interval 0.00055 to 0.00279) for the chloroplast and 0.0055 (95% confidence interval 0.00473 to 0.00643) for the nuclear genes from this dataset. Thus, values are slightly higher than those of *A. lyrata* for the chloroplast, in contrast to patterns from nuclear genes, where a close to 4-fold reduction in polymorphism is evident. Note that controlling for silent differences in divergence between the loci surveyed in the two species in the chloroplast (0.032 for *A. thaliana* and 0.016 for *A. lyrata*), there is still a reduction in mean estimates of effective population size in *A. thaliana* in the chloroplast ($\theta_s/K_s=0.038$ for *A. thaliana*, 0.0625 for *A. lyrata*), but this is not as severe as the 4-fold reduction at nuclear loci. Consistent with these results, the maximum likelihood estimate of N_c/N_n is higher in *A. thaliana* at 0.75, and is not significantly different from 1 (95% credibility interval 0.5 to 2.0). This is consistent with a role for selfing in elevating the ratio of N_c/N_n , although the credibility intervals overlap between the two taxa.

(iv) Testing for recombination

The similarity in effective population size estimates between the chloroplast and mitochondria in *A. lyrata* is consistent with these genomes experiencing a single genealogical history due to a lack of recombination.

Furthermore, the original combined chloroplast and mitochondrial dataset provided no evidence for recombination using the four-gamete test using the species-wide *A. lyrata* dataset (Table 1; samples to SWE4); the minimum number of recombination events is inferred to be 0. However, if we include *A. thaliana* as an outgroup for the analysis, the minimum number of recombination events is inferred to be 1, and several pairs of sites are discordant based on the four-gamete test (e.g. chloroplast position 24037 and chloroplast position 49304). This pattern can only be explained by a recombination event in the history of the sample, or recurrent mutation. Although substitution rates are low, it is possible that this signature of recombination is driven by multiple substitutions since the divergence from *A. thaliana*. This prompted us to increase our sampling at polymorphic loci, adding a Russian population and larger samples from Sweden (Table 1; samples SWE5 to RUS8). With the additional data, we confirmed some of the outgroup gamete combinations segregating in Russian *A. lyrata*, and the minimum number of recombination events from within-species analysis is now 1 (Table 2). Overall, six pairs of segregating sites show incongruence using the four-gamete test, including one pair within the chloroplast, and the remaining pairs are between the chloroplast and the mitochondria (Table 2). All incongruent sites involve position 24037 of the chloroplast, which is a high frequency derived variant in our sample. Using a composite likelihood approach, the maximum likelihood estimate of the effective recombination rate ρ is 0.501 for the combined data, 0.501 for the chloroplast alone, and 0 for the mitochondria alone. Considering the length of sequence surveyed in this analysis, this implies a per base pair rate of recombination for the chloroplast of 1.3×10^{-5} and 0 for the mitochondria. The estimate of the ratio of recombination to mutation can be calculated by the ratio of estimates of ρ/θ . In the chloroplast, this implies a ratio of 0.013, which is two orders of magnitude lower than estimates from the nuclear genome (Wright *et al.*, submitted). Given that the combined dataset gives a comparable estimate of ρ , the ratio of ρ/θ overall should be orders of magnitude lower than this estimate.

If some recurrent mutation has occurred in the history of the sample, recombination rates could be even lower than those estimated here. In particular, the site 24037 in the chloroplast is the primary site involved in evidence for recombination in our data (Table 2). To explore the possibility that this could be a mutation hotspot, we submitted this region to a BLAST search, and checked for multiple substitutions at this site in the Brassicaceae. Indeed, the C-T transition identified in our data is also found to have occurred in *Sinapis alba* (accession number

Table 2. Estimates and tests for recombination using *A. lyrata* organelle polymorphism dataset

Dataset	ρ^a (/bp)	p^b	Rmin ^c	pairs ^d
chloroplast	0.505 (1.3×10^{-5})	0.209	1	c24037/c49304
mitochondria	0	0.224	0	
combined	0.505	0.164	1	c24037/c49304 c24037/m43303 c24037/m299238 c24037/m274101 c24037/m299238 c24037/m299402

^a Composite likelihood estimates of the population recombination parameter $4N_e r$. Values in parentheses are per base pair estimates.

^b Proportion of permuted data having a maximum likelihood value under recombination as high as observed.

^c Minimum number of recombination events.

^d Pairs of incompatible sites using the four-gamete test. Positions correspond to *A. thaliana* genomic positions in the chloroplast (c) or mitochondria (m), as shown in Table 1.

AJ243754), as well as *Brassica rapa* (accession number AC189190), suggesting that recurrent mutation could be important. Furthermore, a permutation test of likelihood scores provides no evidence for higher linkage disequilibrium at adjacent polymorphic sites in our dataset (Table 2), even when using the pooled chloroplast and mitochondrial data, suggesting little pattern of recombination breaking down associations for more distant sites. Thus, while we cannot rule out recombination entirely, the results are generally consistent with either a low rate of multiple substitutions due to hotspots, or with very low levels of localized gene conversion events.

4. Discussion

In *A. lyrata*, our results fit with the standard neutral prediction of a 2-fold reduction in effective population size of the chloroplast and the mitochondria. This result is most evident in the species-wide sample, as well as the Plech, Germany population, where variation at both nuclear and cytoplasmic genes tends to be highest, and our sampling was most focused. Other within-population samples showed no significant rejection of neutrality, despite a complete lack of variation identified in Sweden and Russia, and a lack of silent diversity in the chloroplast genome of North American populations. Evidence for elevated between-population differentiation at cytoplasmic loci is also consistent with the observed reduction in effective population size compared with nuclear loci. The combined data also suggest a non-significant elevation of N_c/N_n at the mitochondria in North American populations. One possible contributor to this is the effect of partial selfing increasing homozygosity in the nuclear genome in the Rondeau

population (Mable *et al.*, 2005); increased homozygosity at nuclear genes will tend to increase the ratio of N_c/N_n . More extensive sampling across individuals, populations and loci will be important to further test the importance of partial selfing and local adaptation on the effective size of these genomes.

A recent study of chloroplast variation suggested extensive haplotype sharing between *A. lyrata*, *A. arenosa* and *A. halleri* (Koch & Matschinger, 2007). This result could reflect continuing lineage sorting of ancestral variation, or recent between-species introgression. While our results cannot rule out introgression into *A. lyrata*, our inference of a neutral ratio of cytoplasmic to nuclear effective size would suggest that, if hybridization is occurring, it is not asymmetric with respect to pollen versus seed migration. More, generally, our results are inconsistent with a strong difference in the ratio of pollen to seed migration, which may reflect a general lack of gene flow in relatively fragmented, isolated populations.

In *A. thaliana*, selfing rates have been estimated at 0.97 in natural populations (Abbott & Gomes, 1989). This would imply an expected effective population size ratio of cytoplasmic relative to nuclear genes of approximately 1, and our results are consistent with this finding. This analysis is thus consistent with the conclusion from the frequency spectrum (Jakobsson *et al.*, 2007) that the *A. thaliana* chloroplast polymorphism data also follow the neutral expectation.

The overall results suggest that rates of adaptive and deleterious mutation at organelles are low enough that they are not causing extensive hitch-hiking in these species, in contrast to recent suggestions of rampant hitch-hiking of mitochondrial polymorphism in animal taxa (Bazin *et al.*, 2006). The reduction of substitution rates may be sufficient, and genomic

rearrangement rates low enough, to cause little effect of background selection and selective sweeps at organelles in many natural plant populations. Alternatively, these processes could be affecting nuclear and cytoplasmic genes equally, giving the neutral expectation when examining the ratio of effective sizes. In either case, given the unusually high diversity found in previous studies, balancing selection associated with male sterility may in fact be important in structuring cytoplasmic variation in gynodioecious species (Houliston & Olson, 2006; Stadler & Delph, 2002), although clearly more multi-locus nuclear and cytoplasmic data should be collected from other plant taxa to assess this in more detail. Our data suggesting uniformly low polymorphism levels are also in contrast with recent reports in other systems of high substitution rates at some cytoplasmic loci (Cho *et al.*, 2004; Barr *et al.*, 2007). Combined with evidence for very low to no recombination in our data, these results also suggest that rates of paternal leakage are low in these species.

Controlling for divergence, species-wide polymorphism levels in the chloroplast are slightly reduced in *A. thaliana* compared with *A. lyrata*, while nuclear variation is more extensively reduced. This differs from a study of selfing and outcrossing species in the genus *Mimulus* (Fenster & Ritland, 1992), where both chloroplast and allozyme variation were equally reduced in selfing species, but is comparable to a study in the nematode genus *Caenorhabditis* (Graustein *et al.*, 2002) where mitochondrial variation was reduced in the selfing species relative to outcrossers, although less extensively than nuclear loci. In these systems, it was inferred that frequent population bottlenecks in selfing species were contributing to the reduction in organelle diversity, explaining suppression of variation in both genomes. Given our results, we can conclude that differences in demographic history may be reducing species-wide variation at both cytoplasmic and nuclear loci, while mating system differences are specifically reducing variation at nuclear genes, leading to an increase in N_c/N_n in *A. thaliana*.

We thank Asher Cutter for helpful discussion, and Pavel Goldvasser and Neera Singal for technical assistance. This work was supported by an NSERC discovery grant and a Sloan Research Fellowship to S.I.W. S.I.W. also thanks Deborah Charlesworth for extensive guidance in the study of population genetics and molecular evolution.

References

- Abbott, R. J. & Gomes, M. F. (1989). Population genetic structure and outcrossing rate of *Arabidopsis thaliana*. *Heredity* **62**, 411–418.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.
- Ballard, J. W. & Kreitman, M. (1994). Unravelling selection in the mitochondrial genome of *Drosophila*. *Genetics* **138**, 757–772.
- Ballard, J. W., Hatzidakis, J., Karr, T. L. & Kreitman, M. (1996). Reduced variation in *Drosophila simulans* mitochondrial DNA. *Genetics* **144**, 1519–1528.
- Barr, C. M., Neiman, M. & Taylor, D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**, 39–50.
- Barr, C. M., Keller, S. R., Ingvarsson, P. K., Sloan, D. B. & Taylor, D. R. (2007). Variation in mutation rate and polymorphism among mitochondrial genes of *Silene vulgaris*. *Molecular Biology and Evolution* **24**, 1783–1791.
- Bazin, E., Glemin, S. & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572.
- Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520.
- Birky, C. W., Jr (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the USA* **92**, 11331–11338.
- Birky, C. W., Jr, Maruyama, T. & Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513–527.
- Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.
- Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effects of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.
- Charlesworth, B., Nordborg, M. & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* **70**, 155–174.
- Charlesworth, D. & Wright, S. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development* **11**, 685–690.
- Cho, Y., Mower, J. P., Qiu, Y. L. & Palmer, J. D. (2004). Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the USA* **101**, 17741–17746.
- Clauss, M. & Mitchell-Olds, T. (2006). Population genetic structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology* **15**, 2753–2766.
- Cui, L., Leebens-Mack, J., Wang, L. S., Tang, J., Rymarquis, L., *et al.* (2006). Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. *BMC Evolutionary Biology* **6**, 13.
- Fay, J. C. & Wu, C. I. (1999). A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* **16**, 1003–1005.
- Fenster, C. B. & Ritland, K. (1992). Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *American Journal of Botany* **79**, 1440–1447.
- Glemin, S., Bazin, E. & Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism

- in flowering plants. *Proceedings of the Royal Society of London, Series B* **273**, 3011–3019.
- Graustein, A., Gaspar, J. M., Walters, J. R. & Palopoli, M. F. (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**, 99–107.
- Hey, J. (1997). Mitochondrial and nuclear genes present conflicting portraits of human origins. *Molecular Biology and Evolution* **14**, 166–172.
- Houliston, G. J. & Olson, M. S. (2006). Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics* **174**, 1983–1994.
- Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1–44.
- Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics* **159**, 1805–1817.
- Hudson, R. R. & Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.
- Jakobsson, M., Sall, T., Lind-Hallden, C. & Hallden, C. (2007). The evolutionary history of the common chloroplast genome of *Arabidopsis thaliana* and *A. suecica*. *Journal of Evolutionary Biology* **20**, 104–121.
- Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics* **164**, 5–12.
- Jiggins, F. M. & Tinsley, M. C. (2005). An ancient mitochondrial polymorphism in *Adalis bipunctata* linked to a sex-ratio-distorting bacterium. *Genetics* **171**, 1115–1124.
- Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed. H. N. Munro), pp. 21–123. New York: Academic Press.
- Koch, M. A. & Matschinger, M. (2007). Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA* **104**, 6272–6277.
- Koch, M. A., Haubold, B. & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of the chalcone synthase and alcohol dehydrogenase loci among different lineages of *Arabidopsis*, *Arabidopsis* and related genera (Brassicaceae). *Molecular Biology and Evolution* **17**, 1483–1498.
- Laporte, V. & Charlesworth, B. (2002). Effective population size and population subdivision in demographically structured populations. *Genetics* **162**, 501–519.
- Mable, B. K., Robertson, A. V., Dart, S., Di Berardo, C. & Witham, L. (2005). Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* **59**, 1437–1448.
- Marshall, H. D., Newton, C. & Ritland, K. (2001). Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Molecular Biology and Evolution* **18**, 2136–2138.
- Maynard Smith, J. & Haigh, J. (1974). The hitchhiking effect of a favourable gene. *Genetical Research* **23**, 23–25.
- McVean, G., Awadalla, P. & Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.
- Muir, G. & Filatov, D. A. (2007). A selective sweep in the chloroplast DNA of dioecious *Silene* (Section *Elisanthe*). *Genetics*. epub ahead of print, 29 July 2007.
- Nordborg, M. & Donnelly, P. (1997). The coalescent process with selfing. *Genetics* **146**, 1185–1195.
- Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetical Research* **67**, 159–174.
- Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biology* **3**, 1289–1299.
- Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qiu, Y. L., et al. (2000). Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the USA* **97**, 6960–6966.
- Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.
- Ramos-Onsins, S. E., Stranger, B. E., Mitchell-Olds, T. & Aguade, M. (2004). Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**, 373–388.
- Rozas, J., Sanchez-Delbarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.
- Saur Jacobs, M. & Wade, M. J. (2003). A synthetic review of the theory of gynodioecy. *American Naturalist* **161**, 837–851.
- Savolainen, O., Langley, C. H., Lazzarro, B. P. & Fréville, H. (2000). Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Molecular Biology and Evolution* **17**, 645–655.
- Schnable, P. S. & Wise, R. P. (1998). The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends in Plant Science* **3**, 175–180.
- Stadler, T. & Delph, L. F. (2002). Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences of the USA* **99**, 11730–11735.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tian, X., Zheng, J., Hu, S. & Yu, J. (2006). The rice mitochondrial genomes and their variations. *Plant Physiology* **140**, 401–410.
- Wares, J. P., Barber, P. H., Ross-Ibarra, J., Sotka, E. E. & Toonen, R. J. (2006). Mitochondrial DNA and population size. *Science* **314**, 1388–1390.
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 188–193.
- Welch, M. E., Darnell, M. Z. & McCauley, D. E. (2006). Variable populations within variable populations: quantifying mitochondrial heteroplasmy in natural populations of the gynodioecious plant *Silene vulgaris*. *Genetics* **174**, 829–837.
- Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA* **84**, 9054–9058.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.
- Wright, S. I. & Charlesworth, B. (2004). The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**, 1071–1076.
- Wright, S. I., Lauga, B. & Charlesworth, D. (2003). Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* **12**, 1247–1263.
- Wright, S. I., Foxe, J. P., Derosé-Wilson, L., Kawabe, A., Looseley, M., et al. (2006). Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**, 1421–1430.