J Mol Evol (2007) 64:136–141
DOI: 10.1007/s00239-005-0274-1

JOURNAL OF MOLECULAR EVOLUTION

Neutral Evolution of Synonymous Base Composition in the Brassicaceae

Stephen I. Wright, Gabriel Iorgovan, Sushant Misra, Mohammad Mokhtari

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

Received: 14 November 2005 / Accepted: 3 October 2006 [*Reviewing Editor:* Dr. Magnus Nordborg]

Abstract. The GC content of synonymous sites is elevated in genes from both *Brassica oleraceae* and *Arabidopsis lyrata* compared with *Arabidopsis thaliana*. However, this shift in base composition is independent of gene expression level, and there is no evidence for a similar difference in the frequency of codons preferred by translational selection. The results suggest that composition evolution is caused by a change in mutation bias or biased gene conversion, rather than by a reduction in the efficacy of natural selection in selfing *Arabidopsis*.

Key words: Codon bias — Base composition — Brassicaceae — Arabidopsis

Introduction

Synonymous codon usage in many eukaryotic genomes is thought to be the result of a balance among the forces of weak natural selection, genetic drift, and mutation (Akashi 1997). Accordingly, the evolution of codon usage among species can be driven by differences in the strength and direction of these three forces (Powell et al. 2003). For example, reductions in effective population size increase the role of genetic drift relative to natural selection, leading to lower frequencies of codons preferred by translational selection. Alternatively, codon usage may evolve due to shifts in neutral processes, rather than as a consequence of selection. For example, changes in the rate and direction of mutation bias can also cause the evolution of codon bias irrespective of the parameters of selection and population size (e.g., Arndt et al. 2005). Additionally, if gene conversion is biased with respect to base composition, changes in the strength of this bias may influence synonymous base composition (Marais 2003), and this can be caused by changes in both recombination rate and effective population size (Marais et al. 2004). To date, we have little understanding of which forces dominate the evolution of codon usage.

The model plant *Arabidopsis thaliana* and its congeners provide an excellent system for testing the causes of genome evolution. In particular, the species exhibit a variety of mating systems, and this diversity has important consequences for the effective rates of recombination and effective population size (Charlesworth and Wright 2001). Two lines of evidence suggest the action of weak natural selection on codon usage in *Arabidopsis thaliana*. First, the degree of biased codon usage shows a significant correlation with gene expression level in *A. thaliana* (Duret and Mouchiroud 1999; Wright et al. 2002), consistent with stronger selection for translational efficiency and/or accuracy in highly expressed genes. Second, the identified "preferred" codons showing an increased frequency in highly expressed genes match closely with those expected by the most abundant isoaccepting transfer RNA genes (Wright et al. 2004; Kliman and Henry 2005). However, the general level of codon usage bias is low in *A. thaliana*, and recent analyses suggest the possibility of a shift in codon usage compared with related species, particularly between *A. thaliana* and *Brassica rapa*, where *Bras-*

Correspondence to: Stephen I. Wright; *email:* stephenw@yorku.ca

sica showed a stronger preference for codons ending in G or C compared with *Arabidopsis* (Tiffin and Hahn 2002). This shift may result from a reduction in the efficacy of natural selection in *A. thaliana* driven by high levels of self-fertilization as expected by population genetic models (Charlesworth and Wright 2001), but alternative neutral explanations have not yet been ruled out (Tiffin and Hahn 2002), and a comparison of codon usage in a small sample of genes from the closely related *A. lyrata* showed no consistent evidence for a reduction in codon bias in *A. thaliana* (Wright et al. 2002).

Here we further test for a reduction in the efficacy of natural selection in *Arabidopsis thaliana* using two datasets: (1) a set of the highest- and lowest-expressed genes from *A. thaliana* aligned with orthologues from the shotgun genome sequence of *Brassica oleraceae* (estimated divergence time of approx. 25 MYA [Koch et al. 2000]), making use of the complete genome sequence of *A. thaliana*, and shotgun genome survey sequence from *Brassica oleraceae* (Ayele et al. 2005); and (2) 83 genes sequenced in *Arabidopsis lyrata*, aligned with their orthologues from the complete genome of *A. thaliana* (see Wright et al. 2004; estimated divergence time of approx. 5 MYA [Koch et al. 2000]).

Evolution of Codon Usage in A. thaliana *vs.* B. oleraceae

For the comparison with *Brassica oleraceae*, we use quantitative estimates of gene expression from massively parallel signature sequencing (MPSS) in *A. thaliana* (Meyers et al. 2004a, b) to focus on the very highest- and lowest-expressed genes. Our prediction is that if *Brassica* has a higher efficacy of selection on codon bias, the difference in GC content between the species should be most prevalent in highly expressed genes. Alternatively, a neutral shift in GC content should be equally apparent in both the highest- and the lowest-expressed gene classes. We focus our analysis on the extremes of gene expression level with the assumption that these classes are most likely to remain in similar expression classes in *Brassica* (see below).

Across the entire dataset, *B. oleraceae* genes show a consistently higher GC_3 (Wilcoxon ranked sign test $Z = -12.292$, $p < 0.001$, $n = 185$ genes), in agreement with previous results comparing *A. thaliana* and *Brassica rapa* ESTs (Tiffin and Hahn 2002). The elevated GC_3 is present in both the highest- and the lowest-expressed genes (highest expressed, Wilcoxon ranked sign test $Z = -6.004$, $p < 0.001$, $n = 96$ genes; lowest expressed, Wilcoxon ranked sign $Z = -5.662$, $p < 0.001$, $n = 85$ genes). Furthermore, for codons with conserved amino acids between species, *Brassica* shows an elevated number of synonymous bases ending in G or C over *A. thaliana* (Table 1; $\chi^2 = 245.6$, $p < 0.001$). As shown in Table 1 and illustrated in Fig. 1, there is no evidence that the elevated GC in *Brassica* is affected by gene expression level; the ratio of GC synonymous bases in *Brassica* to those in *A. thaliana* is nearly identical in the lowest- and highest-expressed genes (two-tailed Fisher's exact test, $p \gg 0.05$). Thus, while GC_3 shows a significant increase in *Brassica*, it is a similar increase in both high- and low-expression genes, providing no evidence for stronger selection on codon usage in *B. oleraceae*. Taken together, these results suggest that differences in the strength or efficacy of natural selection on codon usage are not causing the evolution of synonymous base composition in these taxa.

Table 1. Changes in GC content in high- and low-expression genes

Expression	GC_{brass}/AT_{thal}	GC_{thal}/AT_{brass}
Low	1092	724
High	2581	1723

Note. For conserved amino acid sites, the table shows the number of codons where *B. oleraceae* has a GC-ending codon and *A. thaliana* an AT-ending codon (GC_{brass}/AT_{thal}) and where *A. thaliana* has a GC-ending codon and *B. oleraceae* an AT-ending codon (GC_{thal}/AT_{brass}). Gene expression level was measured as the sum of expression level across tissues, using the MPSS database (see text), and the 100 highest- and 100 lowest-expressed genes were selected. Gene sequences from *A. thaliana* were submitted to a BLAST search (Altschul et al. 1990) against the *Brassica oleraceae* genomic survey sequence at the NCBI web site (www.ncbi.nih.nlm.gov), and the highest-scoring matches for each region (BLAST scores > 50 bits) of the coding regions were aligned with the *A. thaliana* gene sequence.

Our prediction is that a difference in the efficacy of selection should affect codon bias most strongly in the highest expressed genes, while a neutral shift in base composition should lead to changes in both expression classes. The strength of codon bias in diploid populations is determined by the equation (Bulmer 1991)

$$q = \frac{e^{4N_e s}}{e^{4N_e s} + u/v} \quad (1)$$

where u/v is the ratio of the rate of changes from preferred to unpreferred codons to the rate of changes from unpreferred to preferred. If there is no selection acting on codon bias in the lowest-expression genes, it is clear that codon bias in this class of genes would be unaffected by a change in effective population size, and so a difference in N_e would only affect highly expressed genes. However, another possibility is that lowly expressed genes are subject to weaker selection on codon bias than highly expressed genes. GC% at synonymous sites in the lowest expressed genes is on average higher than in introns (e.g., 0.49 vs. 0.32 in *B. oleraceae*), consistent with the action of weak selec-

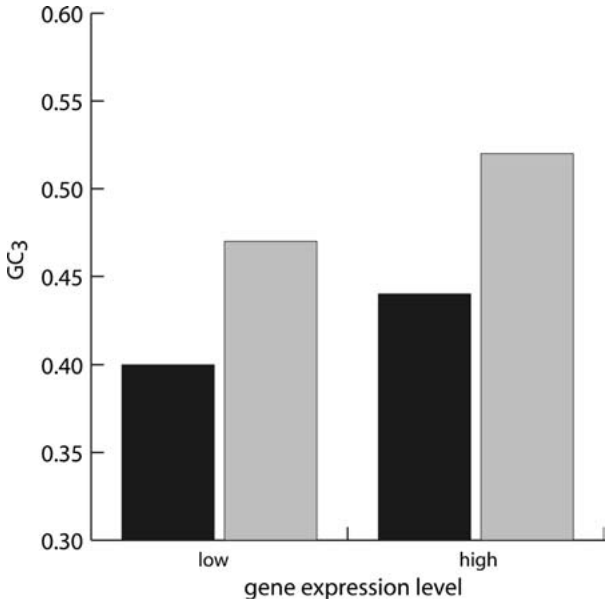

Fig. 1. Comparison of the GC content of synonymous third codon positions for high- vs. low-expression genes in *Brassica oleraceae* (gray bars) and *Arabidopsis thaliana* (black bars).

tion on synonymous sites. However, this comparison is complicated by functional evidence suggesting that selection favors a T-rich base composition and acts against a GC-rich composition in introns (Ko et al. 1998; Rose 2002) and the possible role of adjacent bases in the mutational process (Morton 2001). Nevertheless, the difference between base composition at synonymous sites and that at introns gives us an upper-bound estimate of the strength of selection on codon usage in synonymous sites for low-expression genes. Using the average GC% of introns in *Brassica* (0.32), we estimate that the ratio u/v is approximately 2.1. Based on the average GC% at twofold degenerate sites where G/C bases are expected to be favored over A/T (using tRNA abundance data [Wright et al. 2004]), this gives a rough estimate of $N_e s = 1.59$ for the highly expressed genes (GC = 0.7) and 0.9 for the lowest-expressed genes (GC = 0.54) in *Brassica*. Alternatively, using GC content at twofold degenerate sites with no codon preferences expected based on tRNA abundance, $u/v = 1.2$, and $4N_e s$ is estimated as 1.05 in highly expressed genes and 0.35 in low-expression genes. Using these parameters, we can predict how a reduction in effective population size is expected to change codon bias in the highest- and lowest-expressed genes. Figure 2 shows that the highly expressed genes are more responsive to changes in effective population size than genes subject to weaker selection, suggesting that highly expressed genes should in fact show a larger shift in codon bias than low-expression genes. While this model assumes that codon usage has returned to equilibrium, a shift in N_e is more likely to have a detectable effect on the highest-expressed genes than on low-expression genes.

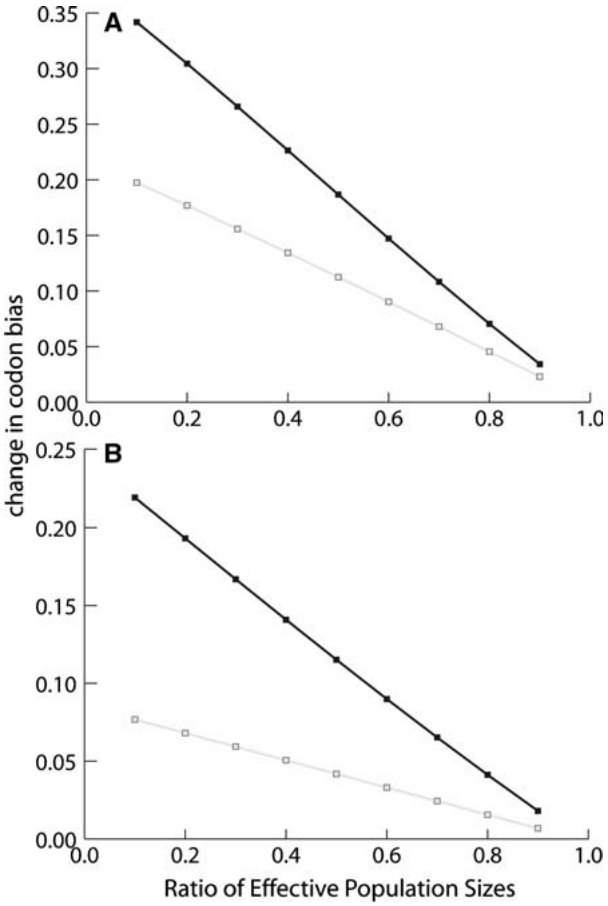

Fig. 2. Changes in codon bias expected as a result of a reduction in effective population size. The *y*-axis shows the difference in codon bias between the species with a larger effective size and the species with a reduced effective size. The *x*-axis shows the ratio of effective sizes between the species. Black lines are parameter values for highly expressed genes, and gray lines are values for low expression genes. **A**, $u/v = 2.1$; **B**, $u/v = 1.2$.

Although Fig. 2 demonstrated a larger effect of changes in N_e on highly expressed genes, it also illustrates that small changes in effective population size might not have a detectable difference on high- vs. low-expression genes, while still having an overall effect on synonymous base composition. Furthermore, the GC content of synonymous sites across all codons is only weakly associated with selection on codon usage in *Arabidopsis* (Wright et al. 2004; Kliman and Henry 2005), and it is possible that there is a residual difference in the efficacy of selection driving the evolution of codon usage. To further test for a difference in the efficacy of natural selection between species, we use the classification of codon preferences based on tRNA abundance (Wright et al. 2004). Focusing on the highly expressed genes, where selection is most likely acting on codon usage, we examine all conserved amino acid sites where there has been a synonymous substitution leading to a GC-AT change between the species, and classify changes as preferred in *A. thaliana* and unpreferred in *B. oleraceae*, unpreferred in *B. oleraceae* and pre-

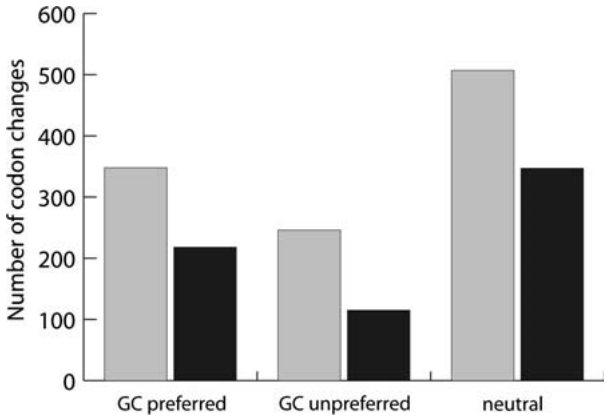

Fig. 3. Synonymous codon evolution between *A. thaliana* and *B. oleraceae*. The *y*-axis is the number of synonymous AT-GC changes between species leading to a GC-ending codon in *B. oleraceae* (gray bars) or a GC-ending codon in *A. thaliana* (black bars) for GC-ending codon changes which are preferred, unpreferred, or neutral relative to the AT-ending codon in the alternate species.

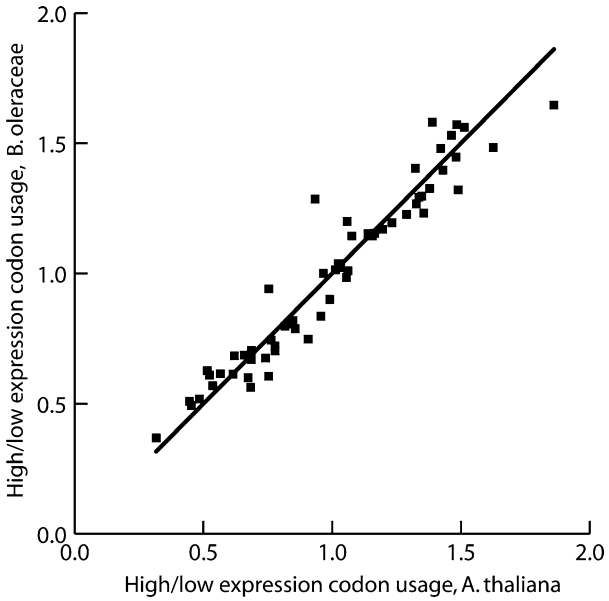

Fig. 4. Ratio of codon frequencies in high- vs. low-expression genes in *Arabidopsis thaliana* and *Brassica oleraceae*. The figure shows, for each degenerate codon, the frequency of each codon in the highest-expressed genes relative to its frequency in the lowest-expressed genes. Values >1 thus represent codons thought to be preferred by translational selection, while values <1 represent unpreferred codons. The line is the 1:1 line, for equal relative frequency in both species.

ferred in *A. thaliana*, and neutral changes (P-P or U-U). By comparing codon preferences for AT-GC changes and GC-AT changes separately, we can control for any possible neutral difference in base composition between the species. As shown in Fig. 3, there is a consistent elevation of changes where *B. oleraceae* has the GC-ending codon and *A. thaliana* has the AT-ending codon, irrespective of whether the GC-ending codon is a preferred, unpreferred, or neutral change. This pattern is inconsistent with a difference in the effectiveness of selection on codon usage between species but is entirely consistent with a neutral shift in substitution patterns.

There are two major limitations with our current analysis. First, by dividing genes into expression level classes based on *A. thaliana* MPSS data, we have assumed that gene expression levels are conserved across species. Our focus on the extremes of gene expression level make it likely that this assumption is valid, but the possibility that a number of our genes have shifted expression levels in *Brassica* has not been ruled out. Second, we have also assumed that codon preferences have remained constant between the species. Our analysis of the frequency of preferred codons relies on tRNA gene abundance data from *A. thaliana*, and changes in tRNA abundance between species may lead to shifts in codon adaptation that would obscure a change in the efficacy of natural selection. To test these possibilities further, we compared codon frequencies in high- vs. low expression genes in the two species. Our prediction is that if selection regimes and gene expression levels are conserved across species, the direction and magnitude of shifts in codon usage should be conserved between the two species. Figure 4 shows the relationship of relative codon frequencies in high- vs. low-expressed genes in *A. thaliana* and *B. oleraceae* for each degenerate codon. The codon frequency differences with gene expression appear to be remarkably conserved in both strength and magnitude, and there are no striking patterns that suggest a change in codon preferences between taxa. Thus, our data suggest that codon preferences have remained conserved, and changes in tRNA abundance or gene expression are unlikely to be responsible for the observed change in codon usage.

Evolution of Codon Usage in A. thaliana *vs.* A. lyrata

Previous analysis suggested that *A. lyrata* and *A. thaliana* maintained preferences for the same codons (Wright et al. 2004); observed differences in codon usage should thus reflect differences in either mutation or the strength of selection. In a comparison of *A. thaliana* with genes sequenced in *A. lyrata*, we also find an elevated GC3 in *A. lyrata* (Wilcoxon signed rank test $Z = -2.475$, $p < 0.05$; 83 genes), although the difference is much smaller than with the *Brassica* comparison. In contrast, however, we find no significant difference between *A. lyrata* and *A. thaliana* in a comparison of tRNA-based codon preferences (comparing the number of orthologous codons where *A. lyrata* has a preferred codon and *A. thaliana* an unpreferred codon with codons where *A. thaliana* has a preferred and *A. lyrata* an unpreferred codon, $Z = -0.4199$, $p \gg 0.05$), although there is a weak trend toward a higher number of preferred codons in *A. lyrata* (40 genes elevated pre-

ferred codons in *A. lyrata*; 32 genes elevated preferred codons in *A. thaliana*). Thus, while we find evidence for a significant difference in GC content, measures of codon bias closer to those reflecting selection show no significant difference. Both GC_3 and tRNA abundance-based measures of the frequency of preferred codons are positively correlated with MPSS-based measures of gene expression in both species (Wright et al. 2004). The correlation coefficients are nearly identical, providing no evidence to suggest that the strength of selection differs between the species.

Evolution of Noncoding Base Composition

If neutral processes are governing base composition evolution at synonymous sites, a similar elevated GC content would be expected in noncoding regions of *B. oleraceae* and *A. lyrata*, in comparison with *A. thaliana*. However, we find no significant difference in GC at introns between the species for the genes with introns sequenced in *A. lyrata* (Wilcoxon ranked sign test $Z = 0.8863$, $p \gg 0.05$; $n = 43$) and *B. oleraceae* ($Z = -0.837$, $p \gg 0.05$; $n = 94$). We thus find no evidence for a general elevation of GC content in *A. lyrata* and *B. oleraceae*. However, intron sequences are short in these species, and it is possible that a significant proportion of intron sequence is under selective constraint in such species (Andolfatto 2005). Indeed, in the complete *A. thaliana* genome, GC content is negatively correlated with the GC content at third codon positions (Kliman and Henry 2005), suggesting either general selective constraint on base composition in introns or synonymous sites. Experimental studies have shown that intron base composition has functional effects on splicing efficiency and mRNA transcript levels (Ko et al. 1998; Rose 2002), and thus these patterns may reflect the action of selection on intron base composition. Given that base composition effects are measurable in the lab, the strength of selection in introns may be sufficient to compensate for species differences in mutational biases. Additionally, the effects of adjacent bases on mutation patterns in coding vs. noncoding regions (Morton 2001) may be causing differences in base composition in the two regions. Thus, while we cannot entirely rule out the possibility that intron base composition is neutral and that our observed synonymous base composition evolution is due to differences in selection, our analyses show that the difference between species is unrelated to tRNA-based translational selection, and given the functional evidence for selection on noncoding DNA, we favor the neutral explanation for the evolution of synonymous GC content. Clearly, further tests for the action of natural selection on synonymous and noncoding sites are important, making use of both patterns of nucleotide polymorphism and divergence.

Evolution of Codon Usage in the Brassicaceae

With the caveats noted above, our results suggest that the difference in GC content in these species is not driven by differences in the efficacy of translational selection and is more likely determined by neutral differences. Note that this does not imply the absence of selection on synonymous sites; indeed, the observed positive correlation between codon bias and gene expression and the association between codon preferences and tRNA abundance do suggest the action of translational selection (Wright et al. 2004; Kliman and Henry 2005). Note that it might be expected that a neutral evolution of base composition should generate a stronger difference in GC content for low-expression genes, where selection on codon usage is weak. However, given the weak selection acting on codon usage bias, and the large proportion of amino acids which have both C- and T- ending codons expected to be preferred by translational selection (Wright et al. 2004), it is clear from our results that frequencies of preferred codons can remain constant despite shifts in base composition. In particular, we can use eq. 1 above to predict the expected effect of a change in mutation bias on high- vs. low-expression genes in *A. thaliana*. Given our estimates above of $N_e s$ and u/v at twofold degenerate sites in *Brassica*, and our observed GC frequency at twofold degenerate sites in low-expression genes in *A. thaliana* (0.47), our results would imply a shift in u/v from 2.1 to 2.8. If we use this estimate to predict GC frequency in highly expressed genes in *A. thaliana*, we predict GC to be 0.636, nearly identical to our observed average of 0.633. Our results are thus entirely consistent with a shift in the ratio of GC-AT mutations.

Our conclusion is that selective pressures have remained constant across species but that neutral differences among species in the mutation process have driven the evolution of base composition. As demonstrated in the previous paragraph, shifts in mutation bias could explain our results. An alternative possibility is that the evolution of base composition is caused by differences in the strength of biased gene conversion. Although there is little or no empirical evidence to date for biased gene conversion acting in plant genomes, gene conversion has been hypothesized to show a bias toward the replacement of A/T bases with G/C, leading to a fixation bias of G/C nucleotides (reviewed by Marais 2003). The strength of biased gene conversion is dependent on the product of the effective population size and the rate of biased gene conversion. In a highly selfing species, the effective rate of biased gene conversion is reduced dramatically, due to the very low percentage of heterozygotes in the population (Marais et al. 2004). *Arabidopsis thaliana* is highly selfing in natural populations, while *B. oleraceae* and *A. lyrata* both have

self-incompatibility and are expected to have been outcrossing over evolutionary time scales. This difference in mating system may thus have lead to an important change in the levels of biased gene conversion. In particular, base composition under biased gene conversion is determined by the equation (Marais et al. 2004)

$$q* = \frac{e^{4N_e\omega(1-S)}}{e^{4N_e\omega(1-S)} + u/v} \quad (2)$$

where ω is the strength of biased gene conversion, and S is the selfing rate. Given the selfing rate estimate from *A. thaliana* of 0.98, this would imply a strong reduction in the effective rate of biased gene conversion, where the effects of biased gene conversion are effectively eliminated. If, for example, we use our previous estimate of $u/v = 2.1$, the level of synonymous GC% at twofold degenerate codons in *Brassica* low-expression genes (0.54) would be consistent with a value of $4Ne\omega$ of 0.9. With these parameters, the increased selfing rate in *A. thaliana* would lead to an expected reduction in GC to 0.33. Alternatively, if $u/v = 1$ and base composition is only determined by the level of biased gene conversion, we would infer a value of $4Ne\omega$ in Brassica of roughly 0.15, and we would expect an average GC% of approximately 0.5 in *A. thaliana*. Although GC% is unlikely to have recovered equilibrium since the evolution of selfing following a shift in biased gene conversion, our observed value of 0.47 for GC content at twofold degenerate sites in *A. thaliana* is consistent with parameter values of biased gene conversion intermediate between these two examples.

To further test for the action of selection on noncoding and synonymous base composition, and to distinguish between mutation bias and biased gene conversion, it will be important to examine patterns of nucleotide polymorphism as well as divergence between species; under biased gene conversion, outcrossing species should show a fixation bias from AT-GC compared with polymorphism, while differences in mutation bias should not generate such a pattern. As large polymorphism datasets accumulate in the outcrossing relatives of *A. thaliana*, we will be able to examine and compare the patterns of polymorphism and divergence to test for the role of mutation bias and biased gene conversion in the evolution of base composition.

Acknowledgments. We thank Deborah Charlesworth and Brian Morton for discussion and two anonymous reviewers for comments on the manuscript. This work was supported by a Natural Sciences and Engineering Research Council of Canada to S.I.W.

References

Akashi H (1997) Codon bias evolution in Drosophila. Population genetics of mutation-selection drift. Gene 205:269–278

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Andolfatto P (2005) Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149–1152

Arndt PF, Hwa T, Petrov DA (2005) Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. J Mol Evol 60:748–763

Ayele M, Haas BJ, Kumar N, Wu H, Xiao Y, Van Aken S, Utterback TR, Wortman JR, White OR, Town CD (2005) Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis. Genome Res 15:487–495

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907

Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. Curr Opin Genet Dev 11:685–690

Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc Natl Acad Sci USA 96:4482–4487

Kliman R, Henry AM (2005) Inference of codon preferences in *Arabidopsis thaliana*. Int J Plant Sci 166(1):3–11

Ko CH, Brendel V, Taylor RD, Walbot V (1998) U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. Plant Mol Biol 36:573–583

Koch M, Haubold B, Mitchell Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498

Marais G (2003) Biased gene conversion: implications for genome and sex evolution. Trends Genet 19:330–338

Marais G, Charlesworth B, Wright SI (2004) Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. Genome Biol 5:R45

Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004a) The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. Genome Res 14:1641–1653

Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning J, Haudenschild CD (2004b) Analysis of the transcriptional complexity *of Arabidopsis thaliana* by massively parallel signature sequencing. Nat Biotechnol 22:1006–1011

Morton BR (2001) Selection at the amino acid level can influence synonymous codon usage: implications for the study of codon adaptation in plastid genes. Genetics 159:347–358

Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A (2003) Analysis of a shift in codon usage in Drosophila. J Mol Evol 57 (Suppl 1):S214–S225

Rose AB (2002) Requirements for intron-mediated enhancement of gene expression in Arabidopsis. RNA 8:1444–1453

Tiffin P, Hahn MW (2002) Coding sequence divergence between two closely related plant species: Arabidopsis thaliana and Brassica rapa ssp. pekinensis. J Mol Evol 54:746–753

Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred Arabidopsis. Mol Biol Evol 19:1407–1420

Wright SI, Yau CB, Looseley M, Meyers BC (2004) Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. Mol Biol Evol 21:1719–1726