Rates of Synonymous Substitution Do Not Indicate Selective Constraints on the Codon Use of the Plant *psb*A Gene

Brian R. Morton

Department of Biological Sciences, Barnard College, Columbia University

The *psb*A gene of the flowering plant chloroplast genome has a pattern of codon bias that differs from all other angiosperm chloroplast genes. In *psb*A, unlike all other chloroplast genes, the third-codon-position composition does not reflect the general genome compositional bias of a high A+T content. Instead, in specific synonymous groups, the codon use of *psb*A more closely corresponds to the tRNA population available for translation. Since it requires a composition unlike the genome composition bias, this pattern of codon use is likely to be the result of selection. Selective constraints on codon use are expected to result in decreased rates of synonymous substitution, and it has been observed that *psb*A has the lowest rate of synonymous substitution among plant chloroplast genes. In the present study, this is examined further by testing whether or not those synonymous groups that specifically have an atypical codon use in *psb*A have correspondingly low rates of silent substitution. An analysis of synonymous substitution rate, performed separately for different degeneracy classes of amino acids, shows that, contrary to the expectation, those sites that are presumed, based on their codon use, to be under selective constraint in *psb*A do not show low rates of substitution and are not responsible for the overall low rate of synonymous substitution in this gene. Instead, they actually show increased rates of substitution relative to other chloroplast genes. Two hypotheses concerning the role of selection in *psb*A are advanced to explain the results.

Introduction

A bias in synonymous codon representation is the result of an interaction between a mutational bias and selection (Sharp and Li 1987). Selection for codon use has been most thoroughly studied in *Escherichia coli* and yeast. In these organisms, highly expressed genes have been shown to have a high representation of codons that are complementary to tRNAs that are abundant in the cell (Ikemura 1985). To explain this observation, it has been proposed that selection on synonymous substitutions has adapted codon use to the cellular tRNA population in order to increase translation efficiency (Ikemura 1985; Sharp and Li 1987). This hypothesis is supported by the inverse correlation that exists between the rate of synonymous substitution and the degree of codon bias in bacteria, suggesting that selection is constraining codon use most strongly in highly expressed genes (Sharp and Li 1987; Sharp 1991).

The second factor that influences codon bias is an overall genome compositional bias. A bias in the third-codon-position composition can simply reflect a bias that exists in all sequences, both coding and noncoding (Bernardi and Bernardi 1986). Although the role selection plays in determining compositional bias remains disputed (Bernardi 1993; Sueoka 1995), selection on general composition bias is recognized as fundamentally different from selection for specific codons in order to increase translation efficiency of transcripts, which is limited to coding sequences.

The chloroplasts of plants and unicellular photosynthetic organisms contain a genome that codes for a fairly conserved set of less than 100 genes, most of which are involved in protein synthesis and photosynthesis. Among these genes is a set of 30 tRNAs. Translation of chloroplast genes occurs only with these tRNAs (Pfitzinger et al. 1990). The genes of the plant chloroplast genome have a codon bias that appears to be the result of a compositional bias toward a high genomic A+T content since synonymous codons with A or T at the third position are highly represented (Wolfe and Sharp 1988).

A noticeable exception to this general pattern is the codon use of the gene *psb*A, which codes for the central protein of photosystem II. In liverwort *Marchantia polymorpha* (Umesono et al. 1988; Morton 1993) and angiosperms (Morton 1993), *psb*A has a high representation of NNC codons (codons with a C at the third position) in the two-fold degenerate synonymous groups coded by NNY (groups that have a pyrimidine at the third position of both codons). All other plant chloroplast genes have a strong bias toward the NNT codons in these groups (Morton 1993). In each of these synonymous groups the only tRNA coded by the chloroplast genome is complementary to the NNC codon, the result being that the codon use of *psb*A is more closely adapted to the chloroplast tRNA population (Morton 1993). This pattern of codon bias found in *psb*A is very similar to the codon bias of chloroplast genes from *Chlamydomonas reinhardtii* (Morton 1993).

Based on the rapid turnover of gene product (Mullet and Klein 1987), and the tRNA population of the chloroplast, the codon use by *psb*A has been proposed to result from selection for translation efficiency (Morton 1993, 1996). Although this has yet to be tested rigorously, a role for selection is supported by the observation that *psb*A has the lowest rate of synonyomus substitution relative to other plant chloroplast genes (Morton 1994). Given the pattern observed, and its unique nature relative to other plant chloroplast genes, it is difficult to explain the codon use of *psb*A without some form of selection. Therefore, the low synonymous rate

Key words: natural selection, codon bias, nucleotide composition, chloroplast genome.

Address for correspondence and reprints: Brian R. Morton, Department of Biological Sciences, Barnard College, Columbia University, 3009 Broadway, New York, New York 10027. E-mail: bmorton@barnard.columbia.edu.

Mol. Biol. Evol. 14(4):412–419. 1997

suggests that constraining, or negative, selection is acting on this gene.

The low rate of silent substitution in *psb*A is tested further in the present study by examining synonymous rates separately for the different degeneracy classes. The unique pattern of codon use observed in *psb*A is observed predominantly in a limited set of synonymous groups, as already noted. If selection is constraining the codon use of *psb*A, those synonymous groups that show an atypical codon use relative to other chloroplast genes should also have reduced rates of silent substitution. Therefore, silent substitution rate differences between *psb*A and a number of other chloroplast genes, within specific synonymous groups, can be used to test for selective constraint on *psb*A.

This was performed by comparing substitution rates in all plant chloroplast genes of significant length from *Nicotiana* and *Oryza* (tobacco and rice). Contrary to the expectation, it was found that, although *psb*A has the lowest overall rate of silent substitution, it has the highest rate of synonymous substitution in the NNY synonymous groups—those groups that have been suggested to be under selective constraint. The same pattern is observed when rice and maize are compared, as well as when maize and tobacco are compared. These results indicate that the role of selection in maintaining the codon use of the plant *psb*A gene may be more complex than previously supposed. Alternative theories are put forward in an attempt to guide further work.

Materials and Methods

All gene sequences used in this study were taken from GenBank release 77. Chloroplast genes from *Oryza sativa* (Hiratsuka et al. 1989) *Zea mays* (Maier, Neckerman, and Igloi 1995), and *Nicotiana tabacum* (Shinozaki et al. 1986) were extracted from the complete genome sequences. Distances between genes in terms of codon use were measured by the method of Long and Gillespie (1991). Genes were clustered by the UPGMA to investigate relationships based strictly on similarity in codon use.

The rates of synonymous substitution for the different degeneracy groups were calculated using the method of Lewontin (1989) for genes more than 750 nucleotides in length. Sequences were aligned using the Gap program from the Genetics Computer Group (1991). Conserved amino acid residues were used to calculate the rate of synonymous substitution separately for the two-, four-, and six-fold degenerate groups. The number of synonymous substitutions per codon was calculated by equation (1), where p is the number of observed synonymous changes divided by the number of conserved amino acid residues for that degeneracy group.

$$d = (-b)\ln(1 - p/b) \quad (1)$$

The b value represents the effective number of alternative states and was calculated separately for each degeneracy group by equation (2), where X_i represents the third position frequency of the ith base within the group.

Table 1
Third-Position Composition of Chloroplast Genes

	Two-fold[a]		Four-fold[b]			
Gene	T	C	G	A	T	C
Cre[c,d].....	28	72	2	31	67	1
Nta[c] *psb*A...	39	61	0	26	61	11
Nta[c,e].....	76	24	13	31	40	16

[a] Percent representation at the third codon position for the NNY two-fold degenerate amino acids.

[b] Percent representation at the third codon position for the four-fold degenerate amino acids.

[c] Species are abbreviated as Cre (*Chlamydomonas reinhardtii*) and Nta (*Nicotiana tabacum*).

[d] Combined proportional codon use of the genes *psb*A, *rbc*L, *atp*B, *pet*A, *psb*B, *psb*C, and *psb*D.

[e] Combined codon use of the chloroplast genes, excluding *psb*A, greater than 750 nucleotides in length; *rbc*L, *psa*A, *psa*B, *psb*B, *psb*C, *psb*D, *rpo*A, *rpo*B, *atp*A, *atp*B, *pet*A, *rps*2, *rps*3, *psb*G, and *atp*I.

$$b = 1 - \Sigma_i X_i^2 \quad (2)$$

Variance of the rate was calculated separately for each degeneracy group using the method given by Nei (1987).

Results and Discussion

Codon Use and Synonymous Rate of Substitution in *psb*A and Other Plant Chloroplast Genes

The difference between tobacco *psb*A and other tobacco chloroplast genes in terms of codon use is shown in table 1. This codon use is common to all plant *psb*A genes that have been sequenced (Morton 1993). The *psb*A gene has a bias toward NNC codons in the two-fold degenerate NNY amino acids, whereas all other chloroplast genes have a bias toward codons with a T at the third position. The *psb*A gene is much more similar to other chloroplast genes in the pattern of codon use in four-fold degenerate groups, although it does have a higher representation of NNT codons than any other chloroplast gene (Morton 1996). The pattern of codon bias found in the plant *psb*A gene is also observed in chloroplast genes from the green alga *Chlamydomonas reinhardtii* (Morton 1993, 1996), as shown in table 1.

The unique nature of *psb*A, as well as the similarity of plant *psb*A and *Chlamydomonas* chloroplast genes, is also apparent when various chloroplast genes are clustered by similarity of codon use. When chloroplast genes from tobacco are compared to *Chlamydomonas,* the tobacco *psb*A gene clusters with the *Chlamydomonas* chloroplast genes, not other tobacco chloroplast genes (fig. 1). This atypical codon bias pattern, shown in table 1 and figure 1, is not limited to tobacco *psb*A. When the sequenced plant *psb*A genes, both monocot and dicot, are compared to other plant chloroplast genes, all of the *psb*A genes cluster in a distinct group (fig. 2). Therefore, all known plant *psb*A genes share an atypical pattern of codon use relative to other chloroplast genes.

Given the strong genome compositional bias toward A+T, the atypical codon use of *psb*A is likely to be a result of some form of selection (Morton 1993). Selective constraints on codon use are expected to result

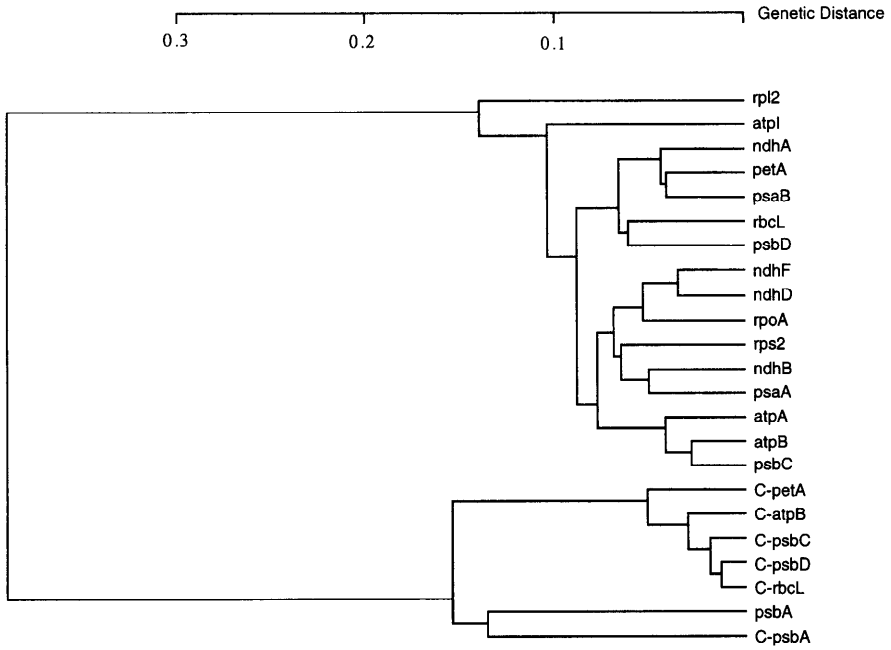

FIG. 1.—Cluster analysis of *Nicotiana* and *Chlamydomonas* chloroplast genes based on similarity of codon use. Genetic distance, in terms of codon bias, was measured using the method of Long and Gillespie (1991) and genes clustered by UPGMA. Chloroplast genes from *Chlamydomonas* are indicated by a C preceding the gene designation.

in a decreased rate of silent substitution (Sharp and Li 1987; Sharp 1991). In a previous study, it was observed that *psb*A has the lowest rate of synonymous substitution of chloroplast genes (Morton 1994), consistent with the hypothesis that selective constraints act more strongly on the *psb*A gene than on other chloroplast genes. However, the atypical codon use of *psb*A is most noticeable in the two-fold degenerate amino acids (table 1), so selective constraints are expected to be strongest in these synonymous groups. Therefore, a significant decrease in silent rate should be observed in the two-fold degenerate amino acids of *psb*A relative to other chloroplast genes.

To test this, the rates of synonymous substitution for different degeneracy groups were calculated following an adaptation of the method proposed by Lewontin (1989). Since complete genome sequences are available from tobacco, maize, and rice, chloroplast genes were

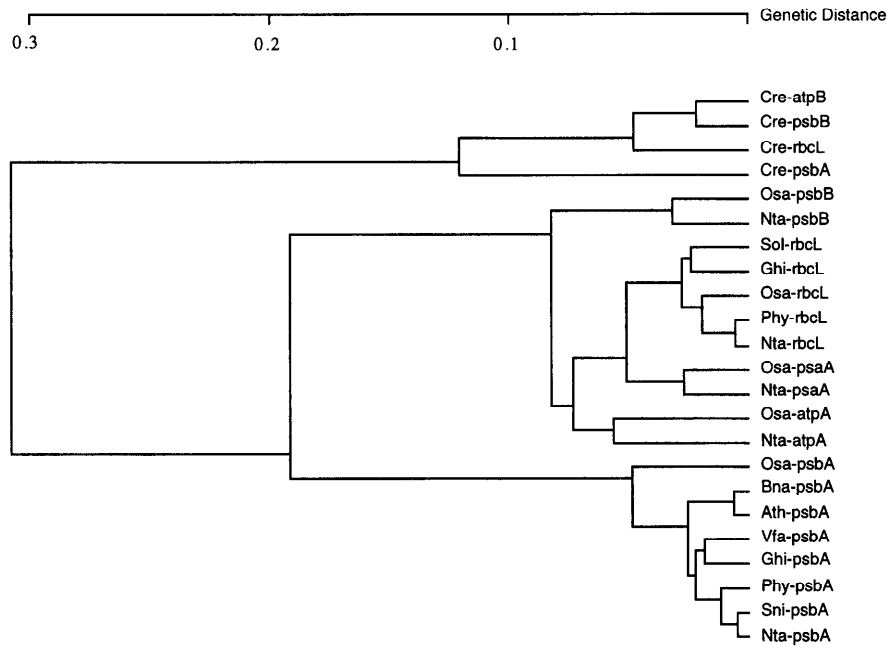

FIG. 2.—Cluster analysis of various plant *psb*A genes and other chloroplast genes. Designations are organism–gene. The three-letter organism designations are Cre (*Chlamydomonas reinhardtii*), Nta (*Nicotiana tabacum*), Osa (*Oryza sativa*), Ghi (*Gossypium hirsutum*), Sol (*Spinacia oleraceae*), Bna (*Brassica napus*), Phy (*Petunia hybrida*), Sni (*Solanum niger*), Vfa (*Vicia faba*), and Ath (*Arabidopsis thaliana*).

Table 2
Rates of Synonymous Substitution by Degeneracy Class

Gene	Two-fold Rate[a]	Four-fold Rate	Six-fold Rate
*psb*A	0.40 (0.17)	0.26 (0.09)	0.27 (0.15)
*rbc*L	0.34 (0.15)	0.47 (0.17)	0.71 (0.30)
*psb*B	0.29 (0.14)	0.53 (0.12)	0.39 (0.17)
*psb*C	0.22 (0.11)	0.46 (0.11)	0.33 (0.14)
*psb*D	0.15 (0.08)	0.55 (0.16)	0.37 (0.17)
*atp*A	0.24 (0.17)	0.54 (0.13)	0.42 (0.17)
*atp*B	0.27 (0.15)	0.55 (0.12)	0.66 (0.30)
*psa*A	0.28 (0.11)	0.50 (0.09)	0.48 (0.16)
*psa*B	0.19 (0.08)	0.53 (0.11)	0.35 (0.12)
*pet*A	0.28 (0.17)	0.76 (0.24)	0.41 (0.24)
*atp*I	0.27 (0.24)	0.53 (0.17)	0.37 (0.24)
*rps*2	0.17 (0.15)	0.60 (0.30)	0.32 (0.17)

NOTE.—Number of synonymous substitutions per conserved codon is given with SD in parentheses.

[a] Only the NNY amino acids are included in this group; NNR amino acids are excluded.

chosen from these organisms. The rates of synonymous substitution for three different degeneracy classes are given in table 2 for 12 chloroplast genes from rice and tobacco, each greater than 750 nucleotides in length. The two-fold degenerate class was limited to the synonymous groups coded by NNY since the major difference between *psb*A and other plant chloroplast genes in terms of codon use occurs in these synonymous groups.

The breakdown of substitution rate by degeneracy class gives an answer which is contrary to the expectation if selective constraints exist on the *psb*A codon bias. The synonymous substitution rate in the two-fold degenerate NNY synonymous groups of *psb*A is noticeably higher than in any other chloroplast gene (table 2). It is in the four- and six-fold degenerate codon groups that *psb*A has a low rate of silent substitution. Curiously, *psb*A is also the only chloroplast gene for which the two-fold degenerate synonymous rate is greater than the synonymous rate in the four-fold and six-fold degenerate classes of the same gene. Unlike the NNY amino acids, the synonymous rate in the other two-fold degenerate synonymous sites, those with a purine at the third position (lysine, glutamine and glutamic acid), is not greater in *psb*A than in all other chloroplast genes (data not shown). This pattern of substitution rate in *psb*A, relative to other chloroplast genes, is also apparent in the unadjusted proportional differences (tables 3 and 4) and is not a function of the method used to account for multiple hits.

To examine whether or not this rate pattern is unique to either rice or tobacco *psb*A, the same comparison was made using maize chloroplast genes. Proportional differences between each gene from the maize/rice, rice/tobacco, and maize/tobacco pairings are presented for the two-fold and four-fold degenerate groups in tables 3 and 4. The same trend is observed in all comparisons: a relatively high proportion of differences for *psb*A in the NNY groups, but a low proportion in the four-fold degenerate groups. This is also observed with distances corrected for multiple hits (data not shown). It is also noted that the increase in rate for the NNY groups in *psb*A is only a single standard deviation above the mean in table 2. However, this gene also has the highest rate along the lineages from tobacco to maize and from rice to rice-maize ancestor, using sites that can be reconstructed. A permutation test on these two lineages indicates that the increased rate in the NNY amino acids in *psb*A is significant ($P < 0.01$).

The high rate in *psb*A is also not a function of an unusual amino acid composition. Table 5 shows the two-fold degenerate (NNY only) amino acid composition of *psb*A and the number of differences observed in each synonymous group when rice and tobacco are compared. The *psb*A gene is not strikingly different in amino acid composition from the average chloroplast gene. Further, the synonymous substitutions are distributed randomly among the different groups ($\chi^2 = 0.97$, NS). Therefore the increased synonymous rate in these groups is due to a general increase, it is not an artifact of amino acid composition, nor is it limited to specific synonymous groups.

The role of Selection in *psb*A Codon Use

The high rate of synonymous substitution in the NNY synonymous groups of *psb*A is not consistent with

Table 3
Proportion of Two-fold Degenerate Sites with an Observed Difference in Chloroplast Genes

	Rice/Tobacco			Maize/Tobacco			Rice/Maize		
Gene	Sites	*D*[a]	P[b]	Sites	*D*	P	Sites	*D*	P
*psb*A	89	24	0.27	89	21	0.24	89	9	0.10
*rbc*L	101	23	0.23	101	19	0.19	104	5	0.05
*psb*B	118	23	0.19	119	27	0.23	119	4	0.03
*psb*C	99	18	0.18	99	17	0.17	102	6	0.06
*psb*D	89	11	0.12	86	15	0.17	86	8	0.09
*atp*A	60	10	0.17	60	10	0.17	71	5	0.07
*atp*B	77	15	0.19	78	19	0.24	84	8	0.10
*psa*A	183	36	0.20	184	37	0.20	186	8	0.04
*psa*B	187	27	0.14	186	26	0.14	189	9	0.04
*pet*A	61	12	0.20	60	14	0.23	65	5	0.08
*atp*I	47	9	0.19	47	8	0.17	51	3	0.06
*rps*2	45	6	0.13	44	6	0.14	51	1	0.02

[a] Number of codons with an observed synonymous difference.

[b] Proportion with a difference.

Table 4
Proportion of Four-fold Degenerate Sites with an Observed Difference in Chloroplast Genes

	Rice/Tobacco			Maize/Tobacco			Rice/Maize		
Gene	Sites	*D*[a]	P[b]	Sites	*D*	P	Sites	*D*	P
*psb*A	162	32	0.20	161	36	0.22	165	12	0.07
*rbc*L	213	69	0.32	208	71	0.34	211	19	0.09
*psb*B	234	88	0.38	236	88	0.37	241	28	0.12
*psb*C	230	71	0.31	230	77	0.33	232	36	0.16
*psb*D	168	58	0.35	161	53	0.33	169	29	0.17
*atp*A	210	76	0.36	213	77	0.36	237	39	0.16
*atp*B	234	89	0.38	228	84	0.37	240	35	0.15
*psa*A	327	113	0.35	325	116	0.36	334	46	0.14
*psa*B	308	108	0.35	306	106	0.35	316	43	0.14
*pet*A	132	58	0.44	131	56	0.43	143	16	0.11
*atp*I	105	35	0.33	105	32	0.30	112	15	0.13
*rps*2....	71	28	0.39	72	29	0.40	83	7	0.08

[a] Number of codons with an observed synonymous difference.
[b] Proportion with a difference.

the hypothesis that the codon use of this gene is being constrained by negative selection. A possible explanation for this observation is that selection has no role whatsoever with regard to the codon bias of the plant *psb*A gene. However, this seems unlikely since the codon use of *psb*A is atypical in a very specific manner. The genome composition bias of a high A+T content is countered with a high third-position C content in a limited set of synonymous groups (table 1). Such a pattern of codon use is difficult to explain solely as the result of a mutation bias specific to a single gene. Moreover, *psb*A is the most highly translated chloroplast gene, making it the most likely candidate for selection on codon use. As a result, it is assumed that selection is a factor, and alternative explanations that involve selection must be explored regarding the codon bias of *psb*A. The key point is that constraining selection assumes an equilibrium state. Selection can be considered as an explanation for the codon use of *psb*A if the possibility that this gene is not at equilibirum is examined.

If negative selection is eliminated, the two other considerations are positive selection or a recent relaxation of selection (no current selection). Based on these considerations regarding the role of selection, two hypotheses are advanced below that are consistent with the high rate of silent substitution in the two-fold degenerate amino acids, but not the four-fold degenerate groups, and which do not require negative selection. Both of these models posit a recent change in the selective pressure on codon use with the consideration that the composition of the plant *psb*A gene may not presently be at equilibrium. The key difference between the two hypotheses is that, while both propose a recent change in selective pressure, the first model is based on a recent gain, the second on a recent relaxation.

Table 5
Distribution of Substitutions Among NNY Codon Groups in *psb*A

Amino Acid	Codons	Cumulative Use[a]	*psb*A[b]	*D*[c]
His	CAT + CAC	13%	10	2
Asp.....	GAT + GAC	18%	7	2
Asn.....	AAT + AAC	18%	22	7
Tyr	TAT + TAC	15%	12	3
Phe	TTT + TTC	24%	26	6
Cys	TGT + TGC	4%	2	1
Ser.....	AGT + AGC	9%	10	3

[a] Percent representation of each amino acid in all plant chloroplast genes relative to the NNY amino acids only. The number is an average of rice and tobacco representation.
[b] Number of occurrences of each amino acid in the plant *psb*A gene. The numbers are the same for both rice and tobacco.
[c] Number of residues that have a synonymous difference in a comparison of rice and tobacco.

Hypothesis I

The first hypothesis, consistent with the results in table 2, is that codon use of *psb*A is currently under positive selection which is acting to adapt the NNY two-fold degenerate groups from a low to a high NNC representation. The basis of this model is a recent change in selective pressure such that translation efficiency of *psb*A, not under selection in the recent past, is currently being adapted to match the tRNA population in the NNY groups. While selection on codon use will constrain codon bias if it already matches an "optimal" codon use, as observed in previous studies on yeast and *E. coli,* it will affect other genes in a different manner. A gene that does not have an "optimal" codon bias will not be constrained by selection but, rather, will be adapted by positive selection.

Such a model has not yet been considered for codon bias but all models previously examined assume that codon bias is currently optimized. If this is not the case, any mutation from a nonoptimal to an optimal codon will have a positive selective value. Such mutations are fixed at a higher frequency and rate than neutral or negative mutations (Kimura 1983). In such a situation, an increased rate of silent substitution would result, since most mutations would be beneficial. This increase in rate is analogous to the increased rate of amino acid replacement that has been associated with positive se-

lection in other studies (Hughes and Nei 1988; Tanaka and Nei 1989; Hughes 1992; Riley 1993).

Considering the specific example of the two-fold degenerate amino acids of the plant *psb*A gene, we can assume, based on *Chlamydomonas* codon use, that the NNC codons are "preferred" by selection over NNT codons (Morton 1996). In any gene that is at compositional equilibrium (high A+T) at the time of a change in selective pressure, mutations from NNT to NNC will be fixed at an increased rate. If there has been a recent shift in selection on *psb*A, but not on other chloroplast genes, the result would be an increased rate of synonymous substitution in these synonymous groups as is observed.

In synonymous groups where the compositional bias closely matches the "optimal" codon bias, a change in selective pressure would have a different effect. Since most codons are already optimal, selection would constrain codon use in these synonymous groups. In the plant chloroplast genome, while the compositional bias is at odds with the preferred set of codons in two-fold degenerate groups, in the four-fold degenerate groups the two sets are very similar (Morton 1996). Therefore, the relatively high rate in the two-fold degenerate groups of *psb*A, and the relatively low rate in the four-fold degenerate groups, could be due to the different selective pressure that would result from the different relationship between compositional bias and preferred codons in the two classes.

The difficulty with this hypothesis is that, while it is consistent with the rate data, there is no obvious model for why there would be a recent gain in selective pressure, or why it would be limited to *psb*A. It should also be noted that, if this hypothesis is correct, the gain must have occurred either prior to the divergence of the angiosperm or in multiple lineages, since all plant *psb*A genes have an atypical codon bias (fig. 2). Therefore, high silent rate in the two-fold degenerate groups of *psb*A cannot be generated by a high rate in one lineage only if this model is correct.

Hypothesis II

The second hypothesis is that selection has constrained the *psb*A codon use in the past but has been recently relaxed. As a result, selection has not been constraining synonymous substitutions in this gene during the divergence of the angiosperms, and the atypical codon bias of *psb*A is a remnant of the past selection. It should be noted that the data cannot exclude the possibility that selection has been lost only in the monocots. The high rates between maize and rice and between rice and tobacco are consistent with either a loss of selection within the monocots or a loss in all angiosperms.

There are two points to consider with regard to this hypothesis. The first is that the pattern of codon use in *psb*A is similar to the pattern observed in *Chlmaydomonas* chloroplast genes (see fig. 1) but is not as strongly biased as these genes (Morton 1994). This suggests that the codon use of angiosperm *psb*A genes could reflect an ancestral codon bias that is being degraded toward the compositional bias of the chloroplast genome.

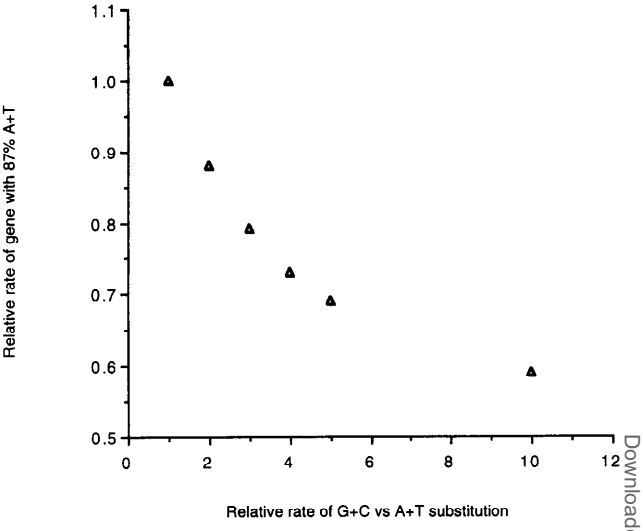

FIG. 3.—Comparison of the rate of silent substitution at four-fold degenerate sites in two genes with different A+T content, when the rate of replacement of A and T differs from the rate of replacement of G and C. The basal rate was calculated for a gene with 70% A+T, which is average for the chloroplast genome (table 1). The rate of a gene with the composition of *psb*A was calculated relative to this basal rate.

The model, therefore, would posit that the ancestral angiosperm *psb*A gene had a more strongly atypical codon use than extant *psb*A genes and was, therefore, more similar to the codon bias of *Chlmaydomonas* chloroplast genes.

The second point is that there is a significant difference in the rate of replacement of different nucleotides in the chloroplast genome (Morton and Clegg 1995). Instead of indicating an effect of selection, the high rate of *psb*A in the two-fold degenerate NNY groups and low rate in the four-fold degenerate groups (table 2) might simply be a reflection of the mutational biases of the chloroplast and the atypical composition of *psb*A (table 1) which would exist at the time of relaxation of selection.

Considering only the four-fold degenerate groups, a comparison of a gene with an 87% A+T third-position content (*psb*A) with a gene with a 70% A+T content (other chloroplast genes on average—see table 1) over various ratios of replacement of A and T to G and C shows that *psb*A would have significantly lower rates of substitution. Under a simple model that considers a single substitution per site, if substitutions are neutral and G or C mutates at a rate 10 times that of A or T, the *psb*A gene would have a rate less than 60% that of the average chloroplast gene using the composition values in table 1 (see fig. 3). Therefore, the low rate of synonymous substitution in four-fold degenerate groups of *psb*A could be a reflection of its composition (table 1). A rough estimate of this replacement ratio using data from the *rbc*L gene is about 4.5 (Morton and Clegg 1995), which would result in *psb*A having a rate about 70% that of other genes (fig. 3). This estimate, while indicating that the low rate in *psb*A could result solely from the compositional bias, is quite rough since the

ratio may vary between genes and the data from Morton and Clegg (1995) do not provide an exact measurement.

Similarly, when considering the two-fold degenerate groups, a rough estimate from *rbc*L of the replacement rate of C is that it occurs about twice the expectation based on composition, while the replacement rate of T is about half that expected from composition (Morton and Clegg 1995). If substitutions are neutral, the high C content at the third position of the NNY amino acids in *psb*A (table 1) could account for the high rate of silent substitution in this degeneracy group. It is interesting to note that the *rbc*L gene has the second highest synonymous rate in the two-fold degenerate groups and also has the second highest C content of chloroplast genes (Morton 1994).

The second hypothesis, then, is that the rate differences between *psb*A and other chloroplast genes are a reflection of compositional differences along with a mutational bias, not the result of selection. The compositional differences between *psb*A and other genes is a result of a recent loss of selection which had maintained the atypical codon bias of *psb*A. Therefore, the codon bias of *psb*A is currently being changed by the mutational bias and is approaching the compositional equilibrium.

This second hypothesis has the advantage that a potential reason for the recent loss of selection can be advanced. Selection for translation efficiency assumes that there is a requirement for rapid translation of individual mRNA molecules in order to maintain protein levels within the cell. If transcript number increases, the rate of translation of individual mRNA molecules could decrease without lowering the level of protein available. In *Chlamydomonas,* where all genes have a codon bias pattern consistent with selection, a single chloroplast genome is present, meaning that transcript level is limited by transcription rate. In contrast, flowering plants have thousands of genomes in each chloroplast. Therefore, translation rate of each transcript could decrease since it is offset by an increase in the number of transcripts available. Loss of selection on translation efficiency, and, therefore, codon bias, could have resulted from an increase in genome number in the past.

Following this model further, if genome number increases gradually over time, then it is expected that a loss of selection on translation efficiency would occur at different times for different genes. Genes with a low level of translation could lose selective pressure on the rate of translation of individual transcripts at a relatively low genome copy number. On the other hand, highly translated genes would require a much larger genome copy number to relieve selection on translation efficiency of individual transcripts. Since *psb*A is the most highly translated chloroplast gene, selection would persist for the longest period on this gene. If this is true, then mutation pressure has been degrading the codon bias of all other chloroplast genes for a longer period of time which are, therefore, currently at, or very near, compositional equilibrium. Only *psb*A maintains a remnant of its ancestral codon bias—the bias that existed while translation efficiency was still under selection. It is intriguing that the other highly translated gene of the plant chloroplast genome, *rbc*L, although very similar to other chloroplast genes, is the gene that is more like *psb*A in terms of codon use (Morton 1994). This could indicate an even fainter trace of the ancestral bias.

This hypothesis, then, proposes the following:

1. In the plant ancestor a single chloroplast genome was present and all chloroplast genes had a codon bias similar to *Chlamydomonas* chloroplast genes.
2. Genome copy number increased gradually during the evolution of plants.
3. As genome number increased, selection for translation efficiency of individual transcripts weakened. Relaxation of selection occured at different times for different genes depending upon expression level.
4. Selection for translation efficiency has been replaced by genome copy number for all chloroplast genes but was relaxed most recently in *psb*A, since this gene is the most highly translated chloroplast gene.
5. The codon bias observed in extant plant *psb*A genes is not at equilibrium, but is a remnant of this ancestral codon bias, which is currently evolving toward compositional equilibrium.
6. Differences in the relationship of substitution rate of different synonymous groups between *psb*A and other chloroplast genes is a result of compositional differences that remain.

What remains essential to test for such a model is the expected rate of change in codon bias once relaxation occurs. Such an anlysis requires an accurate model of nucleotide substitution. This is complicated by the observation that the substitution process in the chloroplast is strongly dependent upon neighboring-base composition (Morton and Clegg 1995; Morton 1995). However, given the overall low rate of nucleotide substitution in the chloroplast genome (Wolfe, Li, and Sharp 1987), a chloroplast gene could take some time to reach compositional equilibrium. It will be important to develop a very accurate model of nucleotide substitution for the chloroplast genome in order to examine the dynamics of such a model.

Conclusions

The high rate of silent substitution in the two-fold degenerate synonymous groups in *psb*A indicates that negative selection cannot be constraining codon use in this gene. Two alternative hypotheses have been advanced in an attempt to provide a direction for further research into the codon bias of the *psb*A gene. It is most likely that the situation is very complex and a number of factors play a role. Clearly, more work needs to be done to more fully resolve the issue. In particular, a more accurate model of chloroplast molecular evolution is required.

In addition, the two hypotheses advanced can be distinguished by a reconstruction of the ancestral sequence of the plant *psb*A gene. Since both propose that *psb*A is a gene that is not at compositional equilibrium, but predict opposite trends in the direction of evolution

of the codon bias, a comparison of the codon use of a reconstructed ancestral gene to extant *psb*A genes can eliminate one of the two models. The first predicts an ancestral codon bias more like other chloroplast genes. The second predicts the opposite, a codon bias more similar to the *Chlamydomonas* pattern than extant plant *psb*A genes show. The availability of more plant *psb*A genes will allow for such a test.

Acknowledgments

I would like to thank M. T. Clegg, B. S. Gaut, G. B. Golding, R. A. Morton, and G. Learn Jr. for helpful discussion of this work and Stanley Sawyer for some very helpful comments that improved the manuscript significantly. This work was supported in part by NIH grant GM 45144.

LITERATURE CITED

BERNARDI, G. 1993. The vertebrate genome: isochores and evolution. Mol. Biol. Evol. **10**:186–204.

BERNARDI, G., and G. BERNARDI. 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24**:1–11.

GENETICS COMPUTER GROUP. 1991. Program manual for the GCG package. Version 7. Madison, Wis.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER et al. (16 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217**:185–194.

HUGHES, A. L. 1992. Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. **9**:381–393.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

IKEMURA, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–35.

KIMURA, M. 1983. The neutral theory of molecular evolution. Canbridge University Press, Cambridge.

LEWONTIN, R. C. 1989. Inferring the number of evolutionary events from DNA coding sequence differences. Mol. Biol. Evol. **6**:15–32.

LONG, M., and J. H. GILLESPIE. 1991. Codon usage divergence of homologous vertebrate genes and codon usage clock. J. Mol. Evol. **32**:6–15.

MAIER, R. M., K. NECKERMANN, and G. L. IGLOI. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing J. Mol. Biol. **251**:614–628.

MORTON, B. R. 1993. Chloroplast DNA codon use: evidence for selection at the psbA locus based on tRNA availability. J. Mol. Evol. **37**:273–280.

———. 1994. Codon use and the rate of divergence of land plant chloroplast genes. Mol. Biol. Evol. **11**:231–238.

———. 1995. Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. Proc. Natl. Acad. Sci. **92**:9717–9721.

———. 1996. Selection on the codon bias of *Chlamydomonas reinhardtii* chloroplast genes and the plant *psb*A gene. J. Mol. Evol. **43**:28–31.

MORTON, B. R., and M. T. CLEGG. 1995. Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. J. Mol. Evol. **41**:597–603.

MULLET, J. E., and R. R. KLEIN. 1987. Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. EMBO J. **6**:1571–1579.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

PFITZINGER, H., J. H. WEIL, D. T. N. PILLAY, and P. GUILLEMAUT. 1990. Codon recognition mechanisms in plant chloroplasts. Plant Mol. Biol. **14**:805–814.

RILEY, M. A. 1993. Positive selection for colicin diversity in bacteria. Mol. Biol. Evol. **10**:1048–1059.

SHARP, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: Codon usage, map position, and concerted evolution. J. Mol. Evol. **33**:23–33.

SHARP, P. M., and W.-H. LI. 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222–230.

SHINOZAKI, K., M. OHME, M. TANAKA et al. (23 co-authors). 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

SUEOKA, N. 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. **40**:318–325.

TANAKA, T., and M. NEI. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobins. Mol. Biol. Evol. **6**:447–459.

UMESONO, K., H. INOKUCHI, Y. SHIKI, M. TAKEUCHI, Z. CHANG, H. FUKUZAWA, T. KOHCHI, H. SHIRAI, K. OHYAMA, and H. OZEKI. 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome II. Gene organization of the large single copy region from rps'12 to atpB. J. Mol. Biol. **203**:299–331.

WOLFE, K. H., W. H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**:9054–9058.

WOLFE, K. H., and P. M. SHARP. 1988. Identification of functional open reading frames in chloroplast genomes. Gene **66**:215–222.

STANLEY A. SAWYER, reviewing editor

Accepted December 12, 1996