J Mol Evol (1996) 43:28–31

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1996

Selection on the Codon Bias of *Chlamydomonas reinhardtii* Chloroplast Genes and the Plant *psbA* Gene

Brian R. Morton

Department of Biological Sciences, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, USA

Received: 8 September 1994 / Accepted: 21 February 1996

Abstract. Plant chloroplast genes have a codon use that reflects the genome compositional bias of a high A+T content with the single exception of the highly translated *psbA* gene which codes for the photosystem II D1 protein. The codon usage of plant *psbA* corresponds more closely to the limited tRNA population of the chloroplast and is very similar to the codon use observed in the chloroplast genes of the green alga *Chlamydomonas reinhardtii.* This pattern of codon use may be an adaptation for increased translation efficiency. A correspondence between codon use of plant *psbA* and *Chlamydomonas* chloroplast genes and the tRNAs coded by the chloroplast genome, however, is not observed in all synonymous codon groups. It is shown here that the degree of correspondence between codon use and tRNA population in different synonymous groups is correlated with the second codon position composition. Synonymous groups with an A or T at the second codon position have a high representation of codons for which a complementary tRNA is coded by the chloroplast genome. Those with a G or C at the second position have an increased representation of codons that bind a chloroplast tRNA by wobble. It is proposed that the difference between synonymous groups in terms of codon adaptation to the tRNA population in plant *psbA* and *Chlamydomonas* chloroplast genes may be the result of differences in second position composition.

Key words: Codon bias — *Chlamydomonas reinhardtii* — *psbA* — Selection

Introduction

Selection for translation efficiency appears to be a major determinant of codon bias in the unicellular organisms *Escherichia coli* and yeast. In these organisms, codon use reflects the relative abundancies of the tRNAs and is thought to be the result of selection to increase the rate of translation of individual transcripts (Ikemura 1985). It has been suggested that this type of selection for translation efficiency would result in an increased representation of codons that have an intermediate strength of codon–anticodon interaction (Grosjean and Fiers 1982; Grosjean et al. 1986).

The chloroplast genome is an interesting system for studies of codon bias. Several complete chloroplast genome sequences are available from plants as well as from red, brown, and green algae. The chloroplast genome codes for a limited set of proteins involved in protein synthesis and photosynthesis (Ohyama et al. 1986; Shinozaki et al. 1986; Hiratsuka et al. 1989). The complete genome sequences also give us the entire set of tRNAs available (the tRNA population). This population consists of only 30 tRNAs and is conserved in the gymnosperm, the cyanelle, and the red, brown, and green algae that have been sequenced. In tobacco two tRNAs have been lost relative to this population. Therefore, the evidence is that chloroplasts have a tRNA population that is essentially constant in all lineages. Translation of chloroplast-coded genes occurs within the chloroplast using only these tRNAs coded by the chloroplast genome. Rec-

Table 1. Third position composition of chloroplast genes

		Twofold[a]		Twofold[b]		Fourfold[c]			
	Gene[d]	G	A*	T	C*	G	A*	T	C*
Group A	Cre[e]	5	95	28	72	2	31	67	1
	Mpo *psb*A	11	89	33	67	0	21	77	2
	Nta *psb*A	17	83	44	56	5	24	59	11
Group B	Mpo[f]	6	94	88	12	4	37	54	5
	Nta[f]	14	86	76	24	13	31	40	16

[a] Third position composition within the NNR twofold-degenerate groups. The asterisk indicates the codons that have a complementary tRNA coded by the chloroplast

[b] Third position composition within the NNY twofold-degenerate amino acids and the twofold-degenerate group of serine

[c] Third position composition in the fourfold-degenerate amino acids and the fourfold-degenerate group of serine

[d] Species are abbreviated as Mpo (*M. polymorpha*), Nta (*N. tabacum*), and Cre (*C. reinhardtii*)

[e] Combined proportional codon use of the genes *psbA, rbcL, atpB, petA, psbB, psbC,* and *psbD*

[f] Combined codon use of the 15 other major chloroplast genes greater than 750 bases in length: *rbcL, psaA, psaB, psbB, psbC, psbD, rpoA, rpoB, atpA, atpB, petA, rps2, rps3, psbG,* and *atpI*

ognition of codons for which no fully complementary tRNA is coded by the chloroplast occurs by wobble, or non–Watson-Crick recognition (Pfitzinger et al. 1990).

The codon bias of plant chloroplast genes reflects the high A+T composition of the genome (Wolfe and Sharp 1988; Morton 1993) with the single exception of the photosystem II gene *psbA*. Unlike all other plant chloroplast genes, this gene shows a preference for C over T at the third position in the twofold-degenerate amino acids (Umesono et al. 1988; Morton 1993).

Selection has been proposed to act on the codon use of *psbA* for two reasons. The first is that the protein is the most prominent translation product of the plant chloroplast due to a very high turnover rate (Mullet and Klein 1987), so any selection for translation efficiency would be expected to affect this gene most strongly. Second, for each of the twofold-degenerate groups the only tRNA coded by the chloroplast is complementary to the NNC codon (the codons with a C at the third position). The increased representation of NNC codons in *psbA* runs counter to the genome compositional bias and corresponds to the tRNA population in these amino acids (Morton 1993).

It has also been noted that the codon use of chloroplast genes from the green alga *Chlamydomonas reinhardtii* is very similar to that of the plant *psbA* gene (Morton 1993). The strong similarity in codon use between plant *psbA* genes and *Chlamydomonas* chloroplast genes is intriguing and suggests similar action by selection whether or not this selection is for translation efficiency. Explaining the codon use of these genes without selection seems difficult as there is a strong bias against the genome compositional bias in a limited set of synonymous groups (Morton 1993).

The present work attempts to explain why selection might adapt the codon use of *psbA* and *Chlamydomonas* genes to the tRNA population, and away from the compositional bias, in some synonymous groups but not in others. Although the codon use in these genes is adapted to the tRNA population in the twofold-degenerate amino acids, the codon use of the fourfold-degenerate amino acids does not correspond to the tRNA population. It is suggested here that the reason for this difference hinges on a common difference between these degeneracy classes—the composition of the second codon position. Fully complementary codons are increased in representation only in synonymous groups with an A or T at the second position, which may require complementarity to ensure efficient codon–anticodon recognition.

Materials and Methods

Gene sequences from *C. reinhardtii* were taken directly from GenBank. Sequences of plant chloroplast genes were extracted from the complete genome sequences (Ohyama et al. 1986; Shinozaki et al. 1986) which were taken from GenBank. Only those genes greater than 750 nucleotides in length were used.

Results and Discussion

Third Position Composition and Chloroplast tRNAs

Chloroplast genes from the green alga *C. reinhardtii,* liverwort (*Marchanfia polymorpha*), and tobacco (*N. tabacum*) that can be divided into two groups, referred to as group A and group B, based on the pattern of codon use. Group A consists of the *C. reinhardtii* chloroplast genes and the plant *psbA* genes while group B consists of all other plant chloroplast genes.

The difference between these two groups is apparent in the third position composition (Table 1). Group A genes have a high representation of C at the third position in the twofold-degenerate groups coded by NNY (those for which the synonymous codons both have pyrimidine at the third position) while the group B genes have a bias toward T at the third position.

The third codon position composition for the fourfold-degenerate groups is also given in Table 1. All genes from both group A and group B show the same bias toward T at the third position, as well as a lesser bias toward A, which reflects the overall genome nucleotide bias. Among the chloroplast genes of both *M. polymorpha* and *N. tabacum,* though, the *psbA* gene in each case has the highest representation of T at the third position (data not shown).

For each of the NNY twofold-degenerate groups there is only one tRNA coded in all chloroplast genomes sequenced to date, and in each case it is complementary to the codon with C at the third position. Therefore, in the twofold-degenerate NNY groups, the codon use of group A genes runs counter to the genome compositional bias to a high representation of codons which have a complementary tRNA coded in the chloroplast (Morton 1993). The opposite is observed in the fourfold-degenerate amino acids. For each of the fourfold-degenerate amino acids the chloroplast tRNAs are complementary to the codons with A or C at the third position except alanine, which has no tRNA complementary to GCC (Morton 1993). This means that the codon use of the group A genes is a high representation of codons with no complementary tRNA coded in the chloroplast.

Correlation of Second and Third Position Composition in Group A Genes

The difference in the codon use of group A genes between the twofold- and fourfold-degenerate groups in terms of adaptation to the tRNA population raises questions about the role of selection. Specifically, why, if selection is responsible for the codon bias pattern of group A genes, is there adaptation toward the tRNA population in certain synonymous groups but away from the tRNA population in others?

The first point is that the evidence is strong that selection, whatever the underlying cause, does act on the codon use of the group A genes. It is difficult to explain the strong codon bias away from the genome compositional bias in a limited set of synonymous groups, and in a specific gene in plants, without invoking selection. Given that *psbA* is the major translation product of the plant chloroplast, this selection is most likely to result in an increase in translation efficiency (Morton 1993).

One possible reason for the difference between the synonymous groups is a difference in tRNA modification between the twofold- and fourfold-degenerate groups. This would require that a codon for each fourfold-synonymous group be modified to bind the NNT codon. However, modifications that have been recorded for chloroplast tRNAs (Sprinzl et al. 1989) do not alter the composition of the anticodon in such a way that the tRNA undergoes Watson-Crick pairing with a different codon than the unmodified tRNA. The only exception to this is the ICG tRNA isolated from bean (Pfitzinger et al. 1990). It is interesting that the binding property of the arginine tRNA is modified since arginine is exceptional in that the tRNAs available to bind the fourfold-degenerate codon group (CGN) are complementary to the CGT and CGG codons. Complementarity to the T and G third position is highly unusual for chloroplast tRNAs (the TTG codon of leucine is the only other example). The lack of a general tRNA modification scheme in chloroplasts indicates that it cannot be a general explanation for the codon bias pattern observed.

Table 2. Correlation of second position composition and adaptation of codon use to the tRNA population in group A and B genes

	Fourfold[a]		Twofold	
	Second position		Second position	
Genes[b]	G or C	A or T	G or C	A or T
Cre[c]	27%	56%	4%	81%
Mpo *psbA*	17%	54%	—[d]	66%
Nta *psbA*	32%	57%	—[d]	62%
Group B	47%	48%	15%	16%

[a] Codon composition is given separately for the fourfold and twofold-degenerate amino acids where the twofold-degenerate amino acids are limited to the NNY class (his, asp, asn, tyr, phe, and cys). Composition is given as a percent codons that are complementary to a chloroplast tRNA

[b] Genes are grouped as A or B as discussed in the text

[c] Genes used were those listed in Table 1.

[d] Only two cysteine residues are present in the *psbA* coding sequence

Another possible explanation comes from a common difference between the twofold-degenerate groups and the fourfold-degenerate amino acids, and that is the composition of the second codon position. All twofold-degenerate NNY amino acids have an A or T at the second codon position with the exception of cysteine, which is coded by TGY. On the other hand, all fourfold-degenerate amino acids have a G or C at the second position, with the exception of valine, which is coded by GTN. What is of interest, then, is whether adaptation to tRNA population is specific to the twofold-degenerate amino acids, or, instead, to synonymous groups with an A or T at the second codon position.

A comparison of second codon position composition and correspondence of codon use to the tRNA population reveals that the two are correlated in group A genes but not group B genes (Table 2). Codon use in Table 2 is given as the percent representation of codons that are fully complementary to an available tRNA. Group B genes have no observed difference in codon use when comparing fourfold-degenerate amino acids with an G or C at the second position to those with an A or T. Group A genes, however, have an increased representation of codons complementary to a chloroplast tRNA when the second position is an A or T, regardless of whether it is a twofold- or fourfold-degenerate amino acid.

This increase is observed most markedly in the two-

fold-degenerate amino acids of *Chlamydomonas* but is difficult to establish strongly for the twofold-degenerate groups of plant *psbA* due to the very low number of the relevant amino acids in the coding sequence. However, relative to group B genes, *psbA* has a very high representation of complementary codons in twofold-degenerate amino acids with an A or T at the second position. It is also noted that, in Table 2, only the twofold-degenerate amino acids with T or C at the third position are considered. The three with A or G (NNR) at the third position are not comparable since all have an A at the second position and a strong compositional bias converges with the expected codon from selection (NNA). The result is that all chloroplast genes have a very high representation of the NNA codons for these amino acids (Morton 1993).

Selection on Codon Use of Group A Genes

A difference in second codon position composition could result in different selection pressure on the third codon position. The second position G:C bond in the codon–anticodon recognition of cysteine and most fourfold-degenerate amino acids could be of sufficient strength such that wobble at the third position is not only permissible but might even be preferable in order to facilitate a high rate of elongation. For the remaining twofold-degenerate amino acids and valine, however, the weaker A:U bond at the second position may not provide efficient recognition, resulting in selection for a Watson-Crick bond at the third codon position. Therefore, selection on group A genes might differentiate codons that interact with chloroplast tRNAs by wobble from those that interact by Watson-Crick bonding, with this discrimination dependent on the second codon position composition.

Although such a model helps to explain the difference in adaptation to the tRNA population between the fourfold- and twofold-degenerate amino acids, it may not be the only factor that affects codon bias of the group A genes. It may also not be consistent from one synonymous group to another, as indicated by the one exception to the pattern described. This occurs in arginine, which is coded by six codons (CGN and AGR) and has an exceptional tRNA complementarity and tRNA modification, as discussed above. All chloroplast genes, group A and B, have a high representation of CGT codons—i.e., group A genes are not adapted to a wobble codon. This unusual nature of arginine, in terms of both the tRNAs and its codon use, makes it an interesting exception that remains to be explored. One possibility is that the arginine tRNA modification noted above is widespread and affects selection on the codon bias of arginine.

Conclusions

The fact that selection is responsible for the codon use of group A genes is strongly supported. The particulars of the selection, however, remain unknown, but translation efficiency is a likely cause. It is suggested here that at least part of this role of selection is to adapt the codon use of each synonymous group to the chloroplast tRNA population, depending upon the second position composition of that group.

Acknowledgments. I thank M.T. Clegg, R.A. Morton, G.B. Golding, and B.S. Gaut for discussion of the manuscript. This research was supported in part by NIH grant GM 45144 to M.T. Clegg.

References

Grosjean H, Fiers W (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18:199–209

Grosjean H, Houssier C, Cedergren R (1986) Anticodon-anticodon interaction and tRNA sequence comparison: approaches to codon recognition. van Knippenberg PH, Hilbers CW (eds). Structure and dynamics of RNA. Plenum Press, New York

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (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

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2:13–35

Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. J Mol Evol 37:273–280

Morton BR, Clegg MT (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 JE, Klein RR (1987) Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. EMBO J 6:1571–1579

Ohyama K, Fukuzawa H, Konchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Pfitzinger H, Weil JH, Pillay DTN, Guillemaut P (1990) Codon recognition mechanisms in plant chloroplasts. Plant Mol Biol 14:805–814

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Suguira M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Sprinzl M, Hartmann T, Weber J, Blank J, Zeidler R (1989) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 17(Suppl):1–172

Umesono K, Inokuchi H, Shiki Y, Takeuchi M, Chang Z, Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Ozeki H (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 KH, Sharp PM (1988) Identification of functional open reading frames in chloroplast genomes. Gene 66:215–222