Plant Physiol. (1990) 92, 1-11
0032-0889/90/92/0001/11/$01.00/0

Received for publication May 25, 1989
and in revised form July 27, 1989

Review

Codon Usage in Higher Plants, Green Algae, and Cyanobacteria[1]

Wilbur H. Campbell* and G. Gowri

Department of Biological Sciences, Michigan Technological University, Houghton, Michigan 49931

ABSTRACT

Codon usage is the selective and nonrandom use of synonymous codons by an organism to encode the amino acids in the genes for its proteins. During the last few years, a large number of plant genes have been cloned and sequenced, which now permits a meaningful comparison of codon usage in higher plants, algae, and cyanobacteria. For the nuclear and organellar genes of these organisms, a small set of preferred codons are used for encoding proteins. Codon usage is different for each genome type with the variation mainly occurring in choices between codons ending in cytidine (C) or guanosine (G) *versus* those ending in adenosine (A) or uridine (U). For organellar genomes, chloroplastic and mitochrondrial proteins are encoded mainly with codons ending in A or U. In most cyanobacteria and the nuclei of green algae, proteins are encoded preferentially with codons ending in C or G. Although only a few nuclear genes of higher plants have been sequenced, a clear distinction between Magnoliopsida (dicot) and Liliopsida (monocot) codon usage is evident. Dicot genes use a set of 44 preferred codons with a slight preference for codons ending in A or U. Monocot codon usage is more restricted with an average of 38 codons preferred, which are predominantly those ending in C or G. But two classes of genes can be recognized in monocots. One set of monocot genes uses codons similar to those in dicots, while the other genes are highly biased toward codons ending in C or G with a pattern similar to nuclear genes of green algae. Codon usage is discussed in relation to evolution of plants and prospects for intergenic transfer of particular genes.

The 64 codons found in the universal genetic code provide the information for controlling expression of the 20 amino acids in proteins and terminating message translation via stop signals. Eighteen of the amino acids are encoded by more than one codon, but Met and Trp have only one codon. The use of the synonymous codons in bacteria, yeast, and higher eukaryotes has been extensively analyzed (12). For genes encoding abundant proteins in *Escherichia coli* and *Saccharomyces cerevisiae*, a set of preferred codons has been identified (Table I). For both organisms, genes for less abundant proteins use a larger set of codons and show less preference toward the set encoding abundant proteins (4,12). Grantham (14) hypothesized that codon usage is genome specific and related to taxonomic order. Until recently, too few higher plant genes had been sequenced to draw significant conclusions on their codon usage. Interest in higher plant codon usage has been heightened by the recognition that monocots differ from dicots in codons used to encode proteins with the same function in metabolism. Cerff and coworkers, prompted by their recognition of large differences in codon usage between the chloroplastic *Gap*[2] genes of maize and dicots, were the first to systematically analyze codon bias of monocot and dicot genes (8,28). Corn *Nir* was found to be encoded by a very different set of codons than the *Nir* gene of spinach (2,17). We found that corn *Nar* codon usage was narrowly biased as compared to a *Nar* gene of Arabidopsis (10,13). A recent study of 207 plant genes confirmed that codon usage in nuclear genes differed between monocots and dicots (22).

Our objective here is to take a broader view of codon bias in plants and to compare codon usage in the genomes of cyanobacteria, green algae, and higher plants. In analyzing codon usage in yeast, Bennetzen and Hall (4) considered a codon preferred if it is used in the set encoding 85% of the amino acids of the proteins. We have utilized their 85% criterion in this review to identify the preferred codons of higher plant and algal genes. We utilized Release 57 of GenBank as the source for nucleotide sequences, which are identified by their entry names. Plant genes sequenced too recently to be in GenBank have been assigned typical entry names. For those unfamilar with GenBank, Murray *et al.* (22) provided a detailed list of plant genes indexed to GenBank entry names.

CODON USAGE IN CYANOBACTERIA, GREEN ALGAL, AND HIGHER PLANT ORGANELLES

Other than the chloroplast genes of *Chlamydomonas,* only a few genes have been cloned and sequenced from green algae. These gene sequences were analyzed for codon usage and examples are presented in Table I. *Chlamydomonas* nuclear

[1] Supported by U. S. Department of Agriculture, Competitive Research Grants Office grant 88-37262-3896 and National Science Foundation grant DMB 88-03998.

[2] Nomenclature: for prokaryote-type genes, Plant Mol Biol Reporter **1**: 38-43 (1983); and for eukaryotic genes, *Adh*, alcohol dehydrogenase; *Cab*, Chl *a/b* binding protein; *Cat*, catalase; *Fed*, ferredoxin; *Gap*, glyceraldehyde-3-phosphate dehydrogenase; *Glb1*, major globulin of maize embryo; *Gls*, glutamine synthetase; *His*, histone; *Nar*, nitrate reductase; *Nir*, nitrite reductase; *RbcS*, small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase; *Pcy*, plastocyanin; *Phy*, phytochrome; *Sod*, superoxide dismutase; *Suc*, sucrose synthetase.

genes have a strong bias toward XXC/G[3] codons (more than 90%). However, chloroplast genes of green algae, which are represented here by *rbc*L, have a preference for codons ending in A or U (Table I). For the *Chlamydomonas* mitochondrial genome, only *cox*I has been sequenced and it shows a preference for 34 codons and 49% XXC/G (data not shown).

For higher plant chloroplasts, a large number of genes have been cloned and sequenced. Fifty preferred codons can be identified from a compilation of the codon usage of 39 protein genes of the tobacco chloroplast, for which the entire genome has been sequenced (32). Using GenBank entries for higher plant chloroplast genes, we calculated the codon usage for *rbc*L genes of nine species (4 monocots and 5 dicots), as well as for a number of *atp* and *psa/b* genes. The codon usage of *Zea mays* chloroplast *rbc*L with 49 preferred codons is representative of our results (Table II). However, when a larger number of chloroplast protein genes are considered, which has been done for 136 open reading frames of the *Marchantia polymorpha* chloroplast genome (24), 31 preferred codons are found (Table II). These preferred codons correspond to the 31 tRNA species encoded by the liverwort chloroplast genome, which are nearly the same as the tRNA species encoded in the tobacco chloroplast (24,32). Since the codon frequency of chloroplast genes is related to the concentrations of tRNA species in the chloroplast, the tRNAs encoded by the organelle's genome appear to be adequate to support protein synthesis by acting as isoacceptors for all codons (26). Hence, it appears that the protein synthesis system of the chloroplast differs from *E. coli* in two respects: fewer tRNA species are used in chloroplast, and abundant chloroplast proteins are encoded with a larger number of codons than less abundant proteins, which is the opposite of the situation in *E. coli* (12).

For cyanobacteria, a fairly large number of genes have been cloned and sequenced. The codon usage for *rbc*L genes of *Anabaena* and *Anacystis nidulans* is presented in Table II. These genes have a similar preferred codon set, but have a very different percentage of XXC/G. These differences were found in other genes of these two species including *rbc*S. The *rbc*L gene codon usage of *Synechococcus* was similar to that of *Anacystis*. The *rbc*L gene codon usage of *Alcaligenes eutrophus* was more biased with a set of 33 preferred codons and 92% XXC/G. Thus, it appears that codon usage in *Anabeana* and chloroplasts is similar, whereas the other cyanobacteria have a codon usage pattern more biased toward XXC/G codons, which is quite different from the chloroplast.

For the higher plant mitochondrial genome, genes for *cox* (4 monocots and 1 dicot) and *atp* (1 monocot and 2 dicots) as well as a few other genes have been cloned and sequenced. The codon usage of mitochondrial *cox* genes reveals a preferred set of 40 to 46 codons and 31 to 37% XXC/G (data not shown). While maize mitochondrial *atp*9 gene codon usage is narrower with a total of 37 codons used and a preferred set of 32 codons, it uses 33% XXC/G codons. Dicot mitochondrial *atp* genes have virtually identical codon usage patterns to those of maize. Thus, mitochondrial codon usage in higher plants appears to most resemble that found in bacteria and yeast, but with a lower percentage of total XXC/G. It does not appear to be similar to chloroplast codon usage in that mitochondrial-abundant proteins are encoded with a smaller number of preferred codons; however, the percentage of total XXC/G is similar to the chloroplast.

[3] Abbreviations: XXC/G, all codons ending in C or G, excluding the stop codon UAG; XCG, the four codons ending in CG; XUA, the four codons ending in UA.

CODON USAGE FOR HIGHER PLANT NUCLEAR GENES

The codon usage patterns for a large number of representative Magnoliopsida (dicot) and Liliopsida (monocot) nuclear genes are collected in Tables III and IV, respectively. The system of Cronquist (11) was used to classify these higher plant nuclear genes. In addition to the basic characteristics of codon usage, we have provided a functional category for each gene, the total number of codons used to encode the polypeptide, and the percent of total codons encoded by XCG and XUA, which have been found to be avoided in encoding genes in eukaryotes (6,14,22). Each species appears to have a unique codon usage pattern, but too few genes from individual species have been sequenced to allow this conclusion to be verified. Hence, we chose to place emphasis on the codon usage characteristics of subclasses and families rather than individual plant species.

For dicots, the codon usage of the 100 genes in Table III can be summarized by mean values of 52 total codons used and a preferred set of 44 codons. The mean percentage of XXC/G is 45.0, while the mean values for the avoided codons, XCG and XUA, are 1.8 and 3.2%, respectively. The distribution of the percentages of XXC/G about the mean are modal or normal (Fig. 1A). The codon usage of Fabaceae genes, which comprise one-third of the examples in Table III, do not differ significantly from these mean values. For Solanaceae, which also comprise about one-third of the examples, codon usage is slightly more biased than the average with only 37.4% XXC/G codons used to encode their genes. Genes of Caryphyllidae, which are all from the family Chenopodiaceae, and of Dilleniidae, which are predominantly Brassicaceae, use the highest percentage of XXC/G codons, 51.3 and 50.7, respectively. Differences among subclasses and families are relatively small and the codon usage of all the dicots appears to be very similar.

For monocots, the codon usage of the 63 genes in Table IV can be summarized by mean values of 47 total codons used and 38 preferred codons. The mean percentage of XXC/G is 73.5, while the mean values for the avoided codons, XCG and XUA, are 6.3 and 1.4%, respectively. In contrast to the dicots, the distribution of the percentages of XXC/G codons for monocot genes is bimodal (Fig. 1B). For corn genes, which comprise more than one-half of the examples in Table IV, the mean percentage of XXC/G is higher than the average, while for all the other grasses, it is lower. For the other subclasses Liliopsida, only three genes for two species have been sequenced, which have a much narrower codon bias than the grasses. It appears that the genes of grasses fall into two groups: (a) those with a narrow codon bias and a high percentage of XXC/G like the three non-Poaceae genes, and (b) those with broader codon usage and a lower percentage of XXC/G codons (Fig. 1B). Clearly, when compared to the dicot genes, monocot genes use smaller sets of total and preferred codons with a much higher percentage of XXC/G codons. The percentages of the avoided codons also differ,

Table I. *Comparison of Codon Bias for Green Algae Genes with E. coli and Yeast*

Amino Acid[a]	*E. coli*[b]	Yeast	*Chlamydomonas*		*Chlorella* Chloroplast	*Euglena* Chloroplast
			Nuclear	Chloroplast		
A. Preferred codon						
Leu [6]	CUG	UUG	CUG	UUA/CUU/CUA	UUA/CUU	UUA/UUG/CUU
Arg [6]	CGC	AGA	CGC	CGU	CGU	CGU/CGC/AGA
Ser [6]	UCU/ UCC/AGC	UCU/UCC	UCC/AGC	UCU/UCA	UCU/UCA/AGU	UCU/UCA/AGU
Val [4]	GUU/GUA	GUU/GUC	GUC/GUG	GUU/GUA	GUU/GUA	GUU/GUA
Pro [4]	CCG	CCA	CCC/CCG	CCU/CCA	CCU/CCA	CCU/CCA
Thr [4]	ACU/ACC	ACU/ACC	ACC	ACU/ACA	ACU/ACA	ACU/ACA
Ala [4]	GCU/GCA	GCU/GCC	GCC	GCU/GCA	GCU/GAU	GCU/GCA
Gly [4]	GGU/ GGC	GGU	GGU/GGC	GGU	GGU	GGU/GGA
Ile [3]	AUC	AUU/AUC	AUC	AUU/AUC	AUU/AUC	AUU
Tyr [2]	UAC	UAC	UAC	UAC	UAU/UAC	UAU/UAC
His [2]	CAC	CAC	CAC	CAC	CAU/CAC	CAU
Gln [2]	CAG	CAA	CAG	CAA/CAG	CAA	CAA
Asn [2]	AAC	AAC	AAC	AAC	AAU/AAC	AAU/AAC
Lys [2]	AAA	AAG	AAG	AAA	AAA	AAA
Asp [2]	GAU/GAC	GAC	GAU/GAC	GAU/GAC	GAU/GAC	GAU/GAC
Glu [2]	GAA	GAA	GAG	GAA	GAA	GAA/GAG
Cys [2]	UGU/UGC	UGU	UGC	UGU	UGU	UGU/UGC
Phe [2]	UUC	UUC	UUC	UUC	UUU/UUC	UUU/UUC
Trp [1]	UGG	UGG	UGG	UGG	UGG	UGG
Met [1]	AUG	AUG	AUG	AUG	AUG	AUG
Total [61]	28	25	25	30	33	37
% XXG/C	53.5	48.2	91.9	24.9	19.5	16.1
B. Differences in preferred codons:[c]						
E. coli	—	8	6	10	15	18
Yeast	11	—	12	12	15	19
Chlamydomonas nuclear	9	11	—	19	24	27
Chlamydomonas chloroplast	7	7	13	—	6	10
Chlorella chloroplast	11	7	15	3	—	7
Euglena chloroplast	7	7	14	5	3	—

[a] Number of synonymous codons for each amino acid are shown in brackets. [b] References: *E. coli* and yeast (12); *Chlorella* chloroplast (37); all others from GenBank: *Chlamydomonas* nuclear, GRETBA1A; *Chlamydomonas* chloroplast, CRECPRUBP; *Euglena* chloroplast, EGRCPRBCL. [c] Number of preferred codons used by this species which are not used by the other species.

with the monocots having a higher percentage of XCG and a lower percentage of XUA. Similar conclusions were reached in earlier comparisons of monocot and dicot codon usage (8,22). Most interestingly, the codon usage of monocots is most similar to that of mammalian nuclear genes, where two groups of genes are also recognized (1). The monocot genes of group 1, which are highly biased toward XXC/G codons, are similar to the nuclear genes of *Chlamydomonas* (Table I). Monocot genes of group 2 are more like dicot genes, but with a slightly higher percentage of XXC/G codons (Fig. 1).

In comparing nuclear genes to organellar genes, it is obvious that the nuclear genes of both classes of higher plants use more codons ending in C and G than the chloroplast or mitochondrion. All 61 codons are used to encode the amino acids of some genes in nuclear, chloroplast, and mitochondrial genomes. The number of preferred codons is 31 for chloroplasts, 43 for mitochondria, 44 for dicots, and 38 for monocots. The preferred codons and percentage of XXC/G codons distinguish the nuclear and organellar codon usage patterns of higher plants.

For the nuclear genomes of higher plants, some genes encoding abundant proteins are well identified, but abundant proteins differ among tissues and cells of a plant as well as under different metabolic conditions. Keeping these considerations in mind, the genes for *Rbc*S (10 dicot and 3 monocot sequences) and *Cab* (5 dicot and 4 monocot sequences) can be taken as representative of the most abundant nuclear-encoded proteins in leaves (Table III and IV). These dicot genes have a mean of 42 codons in their preferred set and 54.1% XXC/G codons, while the monocot genes have a set of 32 preferred codons and 92.8% XXC/G codons. These nuclear genes for abundant leaf proteins are more biased than other higher plant nuclear genes, a situation which resembles codon usage in microorganisms where genes for abundant proteins are more biased than the average gene (4,12). Since corn is a C_4 plant, two other genes encode abundant proteins of leaves, namely, phospho*enol*pyruvate carboxylase and pyruvate, phosphate dikinase (MZEPEPCR and MZEPPDK, respectively, in Table IV). Their codon usage is different from the other abundant leaf proteins in that they have a larger set of preferred codons and a lower percentage of XXC/G codons (Table IV). These are large proteins containing over 900 amino acids, and their codon usage illustrates a general trend that larger genes use more codons and have a larger set of

Table II. *Preferred Codons for the Liverwort Chloroplast Genome Compared to the Large Subunit of RuBP Carboxylase/Oxygenase Genes of Z. mays and Cyanobacteria*

Preferred codons of the lsuRuBPC genes are underlined if not among the liverwort preferred codons. The number of extra preferred codons is shown as the difference.

Amino acid[a]	Liverwort[b] All ORFs	Zea mays MZECPRUBP	Anabaena ANARUBP	Anacystis nidulans ANIRUBPL
Leu	UUA/UUG/CUA	UUA/UUG/CUU/CUA	UUA/UUG/CUA/CUG	UUG/CUC/CUG
Arg	CGU/CGG/AGA	CGU/CGC/CGA/AGA	CGU/CGC	CGU/CGC
Ser	UCC/UCA/AGC	UCU/UCC/UCA/AGU	UCU/UCC	UCC/UCG/UCU/AGC
Val	GUC/GUA	GUU/GUA	GUU/GUA	GUC/GUG
Pro	CCC/CCA	CCU/CCA/CCG	CCU/CCC/CCA	CCU/CCC/CCG
Thr	ACC/ACA	ACU/ACC/ACA	ACU/ACC/ACA	AAC/ACG
Ala	GCA	GCU/GCC/GCA	GCU/GCA/GCG	GCU/GCC/GCA/GCG
Gly	GGC/GGA	GGU/GGA/GGG	GGU/GGC/GGA	GGU/GGC
Ile	AUC/AUA	AUU/AUC/AUA	AUU/AUC	AUC
Tyr	UAC	UAU/UAC	UAU/UAC	UAC
His	CAC	CAU/CAC	CAC	CAC
Gln	CAA	CAA/CAG	CAA/CAG	CAA/CAG
Asn	AAC	AAU/AAC	AAC	AAC
Lys	AAA	AAA/AAG	AAA/AAG	AAA/AAG
Asp	GAC	GAU/GAC	GAU/GAC	GAU/GAC
Glu	GAA	GAA/GAG	GAA/GAG	GAA/GAG
Cys	UGC	UGU/UGC	UGU/UGC	UGU/UGC
Phe	UUC	UUU/UUC	UUU/UUC	UUU/UUC
Codons used	61	58	52	50
Preferred	31	49	42	40
Percent XXC/G	12	33	44	69
Difference	0	22	16	20

[a] Trp and Met are not shown since each has only one codon, but are included in the totals shown. [b] References: Liverwort chloroplast (24); all others GenBank.

preferred codons. Another interesting feature of pyruvate,phosphate dikinase is that the first 400 amino acids are encoded by 87% XXC/G codons, while the remaining 547 amino acids are encoded by only 58% XXC/G codons (21).

CODON USAGE IN MAIZE

To illustrate the uniqueness of codon usage in monocots, codon usage patterns of several maize nuclear genes were compared (Table V). The chloroplast *Gap* gene can be used to define the minimum set of codons preferred by the most biased genes in maize. Maize *Nir* and *Cat3* use the minimum set plus one. *Nir* appears to be an exception to the general trend that larger genes use more codons. *Nar*, for which only about two-thirds of the sequence is known, has a slightly larger preferred set of codons, but retains a strong bias toward XXC/G codons. At least 10 other corn genes have this type of strong codon bias, which can also be recognized among the genes from other monocots (Table IV). A second group of corn genes are represented by cytoplasmic *Gap* and *Cat1*, which use a much larger set of preferred codons with less bias toward the XXC/G codons (Table V). At least 10 other corn genes belong to this group and a number of genes from other monocots are similar (Table IV). The codon usage of these genes is like that of dicots as illustrated by the codon usage of chloroplastic *Gap* from tobacco (TOBGAPA in Table V). A related observation concerning the division of monocot genes into two groups comes from comparison of amino acid and nucleotide sequences for homologues (i.e. enzymes or proteins with the same function). For example, genes from the highly biased monocot group, such as *Cab*, *His3*, *His4*, *Nar*, *Nir*, *Rbc*S, and chloroplastic *Gap*, are more similar in amino acid sequence than nucleotide sequence (data not shown). Genes from the other group of less biased monocot genes such as cytoplasmic *Gap* and *Phy* are about equally similar in amino acid and nucleotide sequences. The genes for some other homologues such as *Adh*, *Cat*, and *Sod* are less easily categorized. Corn *Cat1* is similar to *Cat* from sweet potato (IPBCATR in Table III), while corn *Cat3* is very different in codon usage (29). This constrasts with the finding that the four catalases have about the same amount of difference in amino acid sequence.

Although it may be premature to draw conclusions, some corn nuclear genes for proteins targeted for the chloroplast appear to have a much stronger bias toward XXC/G codons than genes for cytoplasmic enzymes. This is illustrated in Table V by cytoplasmic *Gap* (MZEG3PD2) and chloroplastic *Gap* (MZEG3PD1). The genes for two other cytoplasmic enzymes, aldolase and triose phosphate isomerase, have a codon bias similar to cytoplasmic *Gap* (MZEALD and MZETPI1+2, Table IV). Adenine nucleotide translocator of the maize mitochondrion is a nuclear-encoded protein targeted to the organelle's inner membrane, and its gene has a codon usage pattern similar to the genes for cytoplasmic enzymes (MZEANT1, Table IV). *Sod3*, whose protein is targeted to the mitochondrial matrix, has a similar codon usage pattern (MZESOD3, Table IV). Thus, and intracellular codon bias is found in corn, although chloroplastic *Sod2* is

Table III. *Codon Bias of Representative Magnoliopsida (Dicot) Nuclear Genes*

Gene[a]	Function[b]	Codons	Codon Usage		Percent of Codons		
			Total	Preferred	XXC/G	XCG	XUA
Subclass I. Magnoliidae, Magnoliales, Lauraceae							
AVOCEL	ENZ	494	61	57	43.2	2.8	5.6
Subclass III. Caryophyllidae, Chenopodiales, Chenopodiaceae							
SIPFDX	CAR	146	47	43	52.7	6.1	1.4
SIPPCY	CAR	165	47	36	67.3	2.4	2.4
SPIACPI	CAR	138	53	48	40.1	3.5	4.3
SPINIR	ENZ	594	61	50	44.4	2.8	5.0
SPIOEC16	ENZ	232	56	45	45.6	1.7	3.9
SPIOEC23	ENZ	267	54	46	45.9	1.5	3.0
SPIPCG	CAR	169	45	37	62.4	1.8	4.2
SPIPS33	ENZ	331	45	38	52.2	3.6	1.8
Subclass IV. Dilleniidae							
A. Capparales, Brassicaceae							
ATHACP	CAR	129	49	43	45.4	4.0	5.5
ATHADH	ENZ	379	54	48	50.4	2.9	0.5
ATHATPPM	TRA	959	61	48	45.9	1.7	2.2
ATHATPTP	TRA	492	60	49	38.7	1.8	3.0
ATHH3GBC	STR	136	44	41	56.9	5.1	0.7
ATHH4GA	STR	103	36	31	52.8	2.0	0.0
ATHLHCP2	STR	267	50	42	60.9	2.2	1.1
ATHNR2	ENZ	917	61	51	56.7	5.4	3.1
ATHRBCS1B	ENZ	181	50	44	60.2	1.7	1.1
ATHRUBPA	ENZ	471	57	44	54.3	1.2	0.8
ATHTUBA	STR	450	61	50	44.7	1.7	2.7
BNANAP	STO	178	53	47	55.9	4.4	1.7
BOLACPBC	CAR	144	46	41	50.2	4.4	2.9
BOLSLSGR	STR	418	60	52	48.6	4.6	2.9
SALGAPDHR	ENZ	338	50	43	56.3	1.2	0.3
B. Malvales, Malvaceae							
COTSPA	STO	605	60	53	53.3	1.8	3.5
C. Violales, Caricaceae							
CPAPAP	ENZ	345	57	47	28.3	0.6	5.7
D. Violales, Cucurbitaceae							
CUCPHT	REG	1124	61	53	42.2	1.9	6.6
CUSLHCPA	STR	255	52	44	56.1	0.8	0.4
CUSASOX	ENZ	587	61	54	50.0	4.0	4.7
CUSSSU	ENZ	189	49	42	57.9	0.0	1.5
Subclass V. Rosidae							
A. Apiales, Apiaceae							
DAREXT	STR	306	36	32	33.6	5.3	1.0
PHOCHL	ENZ	398	60	53	49.0	2.0	4.8
B. Euphorbiales, Euphorbiaceae							
RCCAGG	LEC	564	61	52	32.4	1.6	6.0
RCCICL4	ENZ	576	58	46	40.2	0.7	2.0
RCCRICIN	TOX	586	61	52	33.5	1.7	6.0
C. Fabales, Fabacceae (Legumes)							
ALFGLNAG	ENZ	376	55	45	41.0	0.6	1.5
ALFLB3R	CAR	146	46	41	32.2	0.7	5.4
CENCONA	STO	290	54	45	42.9	0.6	2.1
DBILECS	LEC	275	52	44	70.1	2.2	1.5
LUPLBR	CAR	154	48	44	40.4	1.2	4.4
PEAABN2	STO	231	51	44	32.7	1.8	3.9
PEACAB80	STR	269	52	42	49.2	0.0	2.2
PEAGSR1	ENZ	357	55	46	32.1	0.6	3.0
PEAGSR2	ENZ	373	57	48	38.1	0.6	4.0
PEALECA	LEC	275	55	45	40.6	1.9	4.7

Table III. *Continued*

Gene[a]	Function[b]	Codons	Codon Usage		Percent of Codons		
			Total	Preferred	XXC/G	XCG	XUA
PEALEGA	STO	517	58	52	46.0	1.4	3.7
PEARUBPS	ENZ	180	50	42	49.1	0.6	1.2
PEAVIC7	STO	357	53	46	35.9	0.0	6.0
PHVCHM	ENZ	329	58	47	71.0	2.4	1.8
PHVDLECA	LEC	275	54	45	65.2	3.4	1.8
PHVGSR2	ENZ	365	54	45	38.1	0.4	1.7
PHVLBA	CAR	147	49	43	35.0	1.4	4.1
PHVLECT	LEC	246	51	43	67.0	2.4	2.0
PHVPAL	ENZ	505	57	47	43.0	0.8	3.0
PHVPHASBR	STO	422	58	50	46.5	1.1	3.6
SOYACT3G	STR	376	55	50	41.0	0.0	1.9
SOYCIIPI	INH	104	43	40	47.2	2.9	1.0
SOYGLYA1A	STO	496	60	49	53.4	1.2	3.0
SOYGLYR	STO	486	59	50	52.4	1.4	2.6
SOYHSP176	UNK	154	48	44	47.0	0.6	1.2
SOYLBGI	CAR	146	47	42	36.5	1.4	3.8
SOYLEA	LEC	285	57	48	53.0	2.6	3.1
SOYLOX	ENZ	838	59	49	39.0	2.2	4.5
SOYNOD26B	UNK	213	55	46	31.3	2.4	7.5
SOYPRP1	STR	256	33	28	38.2	1.2	6.6
SOYRUBP	ENZ	177	49	42	63.3	0.5	1.1
SOYURA	ENZ	207	49	45	40.9	4.0	3.9
VFALEB4	STO	323	56	47	42.1	0.6	4.6
Subclass VI. Asteridae							
A. Asterales, Asteraceae							
FTRBCR	ENZ	173	48	44	64.2	2.3	0.6
HNNRBCS	ENZ	178	51	42	58.1	1.8	1.2
B. Schrophulariales, Scrophulariaceae							
AMACHS	ENZ	390	60	50	58.5	7.4	1.4
C. Solanales, Convolvulaceae							
IPBCATR	ENZ	492	61	51	63.9	4.8	1.4
D. Solanales, Solanaceae							
PETCAB91R	STR	267	50	40	48.0	0.4	1.1
PETCHSR	ENZ	389	60	48	41.8	1.4	3.6
PETGCR1	STR	385	43	34	14.7	0.6	4.0
PETRBCS1A	ENZ	180	48	42	61.3	0.0	0.0
POTINHWI	ENZ	109	44	41	25.9	0.0	6.5
POTLS1G	UNK	138	45	40	40.9	0.0	2.8
POTPATG	STO	386	51	43	23.1	1.0	5.9
POTPIIIR1	INH	154	49	43	33.7	2.4	7.1
POTRBCS	ENZ	181	45	39	52.2	1.1	0.0
TOBATP21	ENZ	560	60	49	41.2	2.1	1.6
TOBECH	ENZ	311	58	51	42.9	2.9	1.8
TOBGAPA	ENZ	392	58	48	51.2	0.6	0.9
TOBGAPB	ENZ	438	58	49	41.2	1.2	3.4
TOBGAPC	ENZ	327	51	43	45.3	0.9	1.2
TOBPRPR	UNK	168	49	45	38.3	1.2	5.4
TOBPXDLF	ENZ	324	56	41	17.7	1.2	7.1
TOBRBPCO	ENZ	180	46	38	51.9	0.4	0.4
TOBTHAUR	UNK	226	56	44	42.2	1.7	1.6
TOMBIOBR	CAR	70	34	31	40.6	1.4	11.2
TOMCAB7	STR	271	54	41	35.6	1.8	4.4
TOMETHYBR	UNK	295	53	47	39.5	0.0	4.4
TOMHRGP	STR	92	16	13	18.5	3.3	5.4
TOMPGR	ENZ	457	55	42	21.8	0.4	4.2
TOMPSI	STR	246	54	47	37.3	0.8	4.0
TOMRBCSB	ENZ	180	49	46	47.1	0.0	0.0
TOMSOD1	ENZ	217	55	46	32.2	1.5	4.1
TOMSOD2	ENZ	152	43	40	32.2	0.0	2.0
TOMWIPIG	INH	111	45	42	28.8	0.0	9.0

Table III. *Continued*

Gene[a]	Function[b]	Codons	Codon Usage		Percent of Codons		
			Total	Preferred	XXC/G	XCG	XUA
Summary of Mean Values for							
Magnoliopsida	(n = 100)		52	44	45.0	1.8	3.2
Caryophyllidae	(n = 8)		51	43	51.3	2.9	3.3
Dilleniidae	(n = 21)		54	46	50.7	2.5	2.4
Brassicaceae	(n = 15)		53	45	51.9	3.0	1.9
Rosidae	(n = 38)		53	45	44.5	1.5	3.3
Fabaceae	(n = 30)		53	45	45.5	1.4	3.2
Asteridae	(n = 33)		50	43	40.4	1.4	3.7
Solanaceae	(n = 28)		49	42	37.4	1.0	3.7

[a] References: GENBANK entry names are used for all genes found in Release 57; ATHACP (27), ATHATPPM (15), ATHATPTP (19), ATHNR2 (10), ATHRBCS1B (16), ATHRUBPA (33), CUSASOX (23), PEAGSR2 (31), SPINIR (2), TOMSOD1 and TOMSOD2 (25). [b] Abbreviations: CAR, carrier protein; ENZ, enzyme; INH, inhibitor protein; LEC, lectin; REG, regulatory protein; STO, storage protein; STR, structural protein; TOX, toxin protein; TRA, transport protein; UNK, protein of unknown function.

not highly biased and some genes for proteins not targeted to chloroplasts, such as *His*3, *His*4, and *Nar*, are highly biased. For dicots, some nuclear genes for proteins targeted to the chloroplast, such as *Cab*, *Feb*, *Pcy*, and *Rbc*S, have a greater bias toward XXC/G codons than other dicot genes, while others such as *Gap* and *Sod* do not differ in codon usage between cytoplasmic and chloroplastic forms (Table III).

For corn *Cat* genes, tissue-specific expression appears to be related to codon usage (29). *Cat*3, which is expressed strongly in leaves but is absent from the kernel, has a strong bias toward XXC/G codons (Table V). *Cat*2, which is strongly expressed in scutellum but is absent from the kernel, has an intermediate codon bias (Table IV). *Cat*1, which is expressed in the kernel as well as the scutellum, is the least biased toward XXC/G and has codon usage similar to genes for root enzymes such as *Adh* and *Suc* (Table IV and V). The storage proteins of corn are, of course, tissue specific in their expression and abundant proteins in the endosperm. The codon usage in genes for the corn storage proteins also seems to be of two types: some strongly biased toward XXC/G codons, while others are much less biased (Table IV). The *Glb*1 gene for the major globulin of corn embryos, which is believed to be a storage protein, has highly biased codon usage (3). Some of the storage proteins have a biased amino acid composition and, consequently, use fewer codons. Tissue-specific expression in dicots appears unrelated to differences in codon usage. For example, *Gls* genes of pea, which are differentially expressed in leaves and nodules, have the same codon usage pattern (31).

SIGNIFICANCE OF BIASED CODON USAGE IN HIGHER PLANTS

Higher plants are like other organisms in that each species has a unique codon bias with plants of the same taxonomic class maintaining a similar codon usage pattern. This is consistent with Grantham's genome hypothesis for codon usage (14). Organelles of higher plants have codon usage patterns, which differ from those of the nuclear encoded genes. It has been stated that the codon usage in the chloroplast and unicellular organisms is similar (22). However, when the *rbc*L gene for the most abundant protein of the chloroplast is considered, the resemblance to unicellular organisms such as *E. coli* and yeast is less evident (Tables I and II). Codon usage for *rbc*L genes of cyanobacteria most resembles that of the *rbc*L gene in chloroplasts of higher plants, except that a higher percentage of XXC/G codons is used in these unicellular organisms (Table II). In summary, higher plants, algae, and their organelles as well as the cyanobacteria fit the general hypothesis that an organism has a unique codon usage pattern with closely related organisms having similar patterns.

It is also clear that the codon usage patterns of the two major classes of flowering plants are more different than might be expected (Fig. 1). For gymnosperms, only the *Cab* and *Rbc*S genes of one species have been cloned (35,36). These genes have codon usage patterns very similar to the same genes in dicots, but with a higher percentage of XXC/G codons. Hence, codon usage in monocots has evolved to be very different from the other higher plants. However, it is not clear which of the higher plants is more like the ancestral genome (11). Despite this lack of knowledge of the ancestral genome, we can still suggest that a mechanism has been operative in evolution which is influencing coding sequences without affecting the gene product. This was first recognized by Grantham (14).

What has driven this change in the coding sequence where differences are mainly found in the third base of codons? Brinkmann *et al.* (8) suggested that monocot genes with a high percentage of XXC/G codons were those induced by internal and external stimuli, but also recognized that some inducible genes in dicots were not unique in codon usage. For microorganisms, it has been suggested that codon usage and concentrations of isoaccepting tRNA have been balanced to optimize the synthesis of abundant proteins (4,12). It seems difficult to apply this concept to monocot codon usage, since some genes of abundant proteins are less biased than others and some rare ones just as biased as the abundant ones. Several other possible explanations for extreme codon bias have been offered including stability of mRNA, optimization of mRNA secondary structure, and optimization of the reading context for codons (12,30). But the manner in which these or other factors operate differently in monocots and dicots is not evident. Both the underlying driving force resulting in the

Table IV. *Codon Bias of Representative Liliopsida (Monocot) Nuclear Genes*

Gene[a]	Function[b]	Codons	Codon Usage		Percent of Codons		
			Total	Preferred	XXC/G	XCG	XUA
Subclass II. Arecidae, Arales, Lemnaceae							
LGIAB19	STR	357	31	29	99.1	6.9	0.0
LGIR15BPC	ENZ	172	38	30	93.4	3.6	0.0
Subclass III. Commelinidae, Cyperales, Poaceae (Grasses)							
ASTAP3R	REG	1129	61	54	46.8	1.7	4.6
BLYALR	STO	362	56	46	79.3	6.6	0.3
BLYAMY1	ENZ	427	52	36	89.2	6.3	1.1
BLYB1HORD	STO	243	53	45	40.0	4.0	4.1
BLYB3HORD	STO	264	56	47	36.4	3.4	4.6
BLYCHORD1	STO	105	31	30	32.6	0.9	3.8
BLYGLUCB	ENZ	290	41	33	93.4	11.7	0.0
BLYLEU	ENZ	362	56	46	79.4	6.6	0.3
BLYPAPI	INH	118	40	34	83.5	5.0	0.8
BLYTH1AR	TOX	127	47	43	53.2	0.0	4.0
BLYUBIQR	REG	77	33	32	82.1	2.6	0.0
MZEA1G	ENZ	357	51	34	90.2	11.5	0.9
MZEACT1G	STR	375	60	53	59.6	2.4	1.6
MZEADH1F	ENZ	379	57	48	65.6	4.5	1.4
MZEADH2N	ENZ	379	54	38	81.9	6.8	1.1
MZEALD	ENZ	355	48	39	71.9	1.2	0.3
MZEANT1	TRA	318	52	43	53.2	1.5	0.3
MZECAT1	ENZ	492	59	47	59.1	2.6	1.4
MZECAT2	ENZ	493	57	43	82.1	8.4	2.0
MZECAT3	ENZ	494	46	30	92.7	12.3	0.4
MZEEG2R	STO	224	40	33	77.6	18.8	0.8
MZEG3PD1	ENZ	403	39	33	96.2	7.6	0.0
MZEG3PD2	ENZ	338	51	40	67.0	2.1	0.0
MZEGLB1S	STO	565	52	34	89.7	12.4	0.2
MZEGLUT2E	STO	224	40	35	73.4	18.2	0.8
MZEGST3A	ENZ	220	42	33	88.2	14.6	0.5
MZEH3C2	STR	137	36	32	94.2	11.7	0.0
MZEH4C14	STR	104	29	26	94.0	5.8	0.0
MZEHSP70I + 2	UNK	646	58	43	77.7	5.0	0.6
MZELHCP	STR	265	36	28	96.2	11.3	0.0
MZEMPL3	STR	147	35	33	95.2	13.5	0.0
MZENIR	ENZ	507	38	30	98.5	9.8	0.0
MZENR1	ENZ	617	56	33	87.2	8.4	2.1
MZEPEPCR	ENZ	935	60	39	83.3	7.0	0.9
MZEPPDK	ENZ	947	61	47	70.2	6.3	1.6
MZERBCS	ENZ	169	35	32	97.0	8.9	0.6
MZESOD2	ENZ	151	48	44	50.8	4.6	0.7
MZESOD3	ENZ	235	53	47	66.9	8.8	1.2
MZESUSYSG	ENZ	812	60	49	67.7	3.7	0.9
MZETPI1 + 2	ENZ	253	53	47	50.4	1.2	2.0
MZEZE15A3	STO	180	41	34	87.2	8.4	0.0
MZEZE22A	STO	263	49	42	47.9	4.2	7.8
MZEZEA20M	STO	240	46	41	43.5	2.1	7.3
RICCAB1R	STR	266	32	27	99.6	13.9	0.0
RICCPI	INH	102	41	36	75.6	3.8	0.0
RICGLUII1	STO	497	60	51	42.2	1.4	5.4
RYESECGSR	STO	194	47	41	51.8	3.6	4.6
WHTAGGTD	LEC	187	39	35	88.7	2.1	0.0
WHTAMYA	ENZ	413	55	40	81.4	7.3	0.7
WHTCAB	STR	266	49	43	74.6	5.5	0.4
WHTEMR	UNK	93	29	26	92.7	2.2	0.0
WHTGIR	UNK	500	59	53	70.8	6.4	2.0
WHTGLGB	STO	291	54	48	47.4	1.6	2.6

Table IV. *Continued*

Gene[a]	Function[b]	Codons	Codon Usage		Percent of Codons		
			Total	Preferred	XXC/G	XCG	XUA
WHTGLIA	STO	318	53	47	36.2	5.4	3.4
WHTGLU1DG	STO	660	55	44	42.6	2.2	2.6
WHTGLUMRA	STO	101	29	28	54.0	5.8	4.0
WHTGLUT1	STO	838	56	46	43.2	4.1	2.0
WHTH3	STR	136	30	27	91.0	6.6	0.7
WHTH4	STR	103	27	25	95.0	4.9	1.0
WHTRBCA	ENZ	163	39	35	90.0	2.4	0.0
Subclass IV, Zingiberidae, Zingiberales, Marantaceae							
TDATHAU2	UNK	235	42	33	89.2	9.3	0.4
Summary of Mean Values for							
Liliopsida	(n = 63)		47	38	73.5	6.3	1.4
Z. mays	(n = 32)		48	38	76.8	7.7	1.6
All other Poaceae	(n = 28)		46	39	67.5	4.8	1.9

[a] References: GenBank entry names for all genes found in Realease 57; MZECAT1 and MZECAT3 (29), MZECAT2 (5), MZEG3PD2 (8), MZEGLB1S (3), MZENIR (17), MZENR1 (13), MZEPPDK (21), MZESOD2 (9), MZESOD3 (34), RICCAB1R (18). [b] Abbreviations: See Table III.

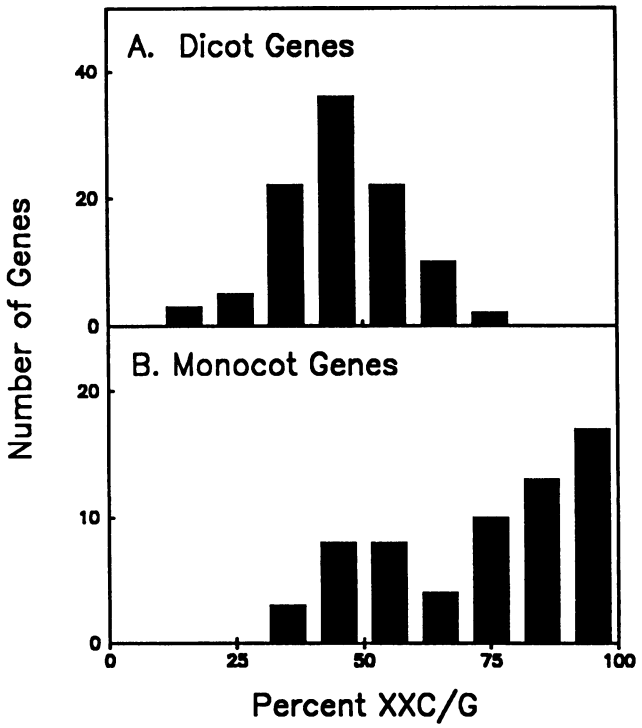

Figure 1. Percentage of codons with C and G in the third position of nuclear genes of dicots and monocots. Data were taken from Tables III and IV.

extreme codon bias found in some monocot genes and its utility in plant metabolism and development are not clear at this time.

Mammalian nuclear genomes have a codon usage pattern with some resemblance to that of the nuclear genes of monocots. The human genome appears to be composed of patches of G + C rich regions among more A + T rich ones (1). Thus, genes encoded in the G + C rich patches have a codon bias toward those codons with high content of G + C, but their introns and flanking regions also are G + C rich (1). Apparently, this may not be the case in monocots, especially maize, where introns and flanking regions of the DNA for genes highly biased toward XXC/G codons are less rich in G + C (28). Furthermore, the monocot genome does not appear to be composed of patches of DNA differing in base composition (28). However, more of the maize genome needs to be sequenced before this conclusion is firmly established.

Another observation made in relation to codon usage is the avoidance of the codons XCG and XUA (6,22). The reason for avoiding the codons XCG and XUA is not known (22). For the XCG codons, it may be related to the frequent methylation of cytosine in CG dinucleotides and the tendency for methyl-CG to mutate by deamination to TG, which may not be detected by the DNA repair system (6,7). Thus, in portions of the genome encoding expressed genes, CG is avoided or not methylated, which may be prevented by specific CG-binding proteins (7). For the XUA codons, avoidance may be related to the use of UA as a stop codon in a simpler genetic code used in primordial organisms or it may also be due to the existence of UA-selective ribonucleases (6). Thus, the avoidance of XCG and XUA codons may be due to different pressures on genomes (6). In monocots, XCG codons occur with a very high frequency in some genes, while in dicots XCG codons are generally avoided. In contrast, XUA codons are more common in dicots than monocots.

Our analysis of plant nuclear codon usage included calculation of the frequencies for these codons, which appear to be significantly different between Magnoliopsida and Liliopsida genes as well as their subclasses (Tables III and IV). These parameters may be useful aids in classification studies, when combined with the set of preferred codons and percentage of XXC/G codons. It should be noted that it is now possible to take advantage of even the limited number of genes that have been sequenced for analysis of plant evolution (20). As can be seen from the analysis of codon usage presented here,

Table V. *Comparison of Preferred Codons of Six Z. mays Nuclear Encoded Genes with a Tobacco Nuclear Gene*

A minimum set of preferred codons is defined by the codon usage of MZEG3PD1 and the preferred codons of the other genes are defined by additions or substitutions to the minimum set. For each amino acid, if the minimum set is used, no additions are shown; additions are indicated by +, while substituted codons are shown in brackets.

Amino Acid[a]	MZEG3PD1[b]	MZENIR	MZECAT3	MZENR1	MZEG3PD2	MZECAT1	TOBGAPA
Leu	CUC/CUG				+CUU	+CUU	[CUA/CUG/UUA]
Arg	CGC/CGG/AGG				[CGC/AGG/AGA]	[CGU/AGG/AGA]	[CGU/CGC/AGA/AGG]
Ser	UCC/AGC	+UCG	+UCG	+UCG	+UCG	[UCU/UCG/AGC]	+UCU/UCA
Val	GUC/GUG				[GUU/GUC]	+GUU	+GUU
Pro	CCC/CCG				[CCC/CCU]	+CCU/CCA	[CCU/CCC/CCA]
Thr	ACC/ACG			+ACA	[ACU/ACC/ACA]	[ACU/ACC/ACA]	[ACU/ACC]
Ala	GCC/GCG			+GCA	[GCU/GCC]	+GCA	[GCU/GCC/GCA]
Gly	GGC/GGG			[GGC/GGU]	[GGC/GGU]	+GGU/GGA	[GGU/GGC/GGA]
Ile	AUC				+AUU	+AUU	+AUU
Tyr	UAC				+UAU	+UAU	+UAU
His	CAC				+CAU	+CAU	+CAU
Gln	CAG						+CAA
Asn	AAC			+AAU	+AAU	+AAU	+AAU
Lys	AAG					+AAA	+AAA
Asp	GAC				+GAU	+GAU	+GAU
Glu	GAG					+GAA	+GAA
Cys	UGC					+UGU	+UGU
Phe	UUC				+UUU	+UUU	+UUU
Codons used	39	38	46	56	51	59	58
Preferred	29	30	30	33	40	47	48
Percent XXC/G	96.2	98.5	92.7	87.2	67.0	59.1	51.2
Difference	0	1	1	5	15	22	24

[a] Trp and Met are not shown since each has only one codon, but are included in the totals shown. [b] References: MZEG3PD1 and TOBGAPA, GenBank; MZENIR (17), MZECAT1 and MZECAT3 (29), MZENR1 (13), MZEG3PD2 (8).

species may be more easily distinguished and relatedness established by the nucleotide sequences of their genes than they are by the differences in the amino acid sequences of their homologous proteins.

Finally, some practical considerations should be discussed. A common practice in molecular biology is to transform a microorganism with a recently cloned gene in order to facilitate production of its gene product. This has been done for higher plant genes with some degree of success (22), but the codon usage of the plant genes may result in low levels of expression. For sperm whale myoglobin, a high level of expression in *E. coli* was obtained by inserting a totally synthetic gene which had been designed with the bacterium's preferred codons (30). Since the native gene for myoglobin has many codons not preferred by *E. coli* and was poorly expressed, it was concluded that codon usage was the most probable explanation for its limited expression (30). However, it was suggested that a few nonpreferred codons can be tolerated, but as the frequency of the poor codons increases, expression of a gene will decrease in *E. coli* (30). A similar conclusion was reached in studies of codon usage in yeast, which has more biased codon usage than *E. coli* (12). However, it is impractical to synthesize large genes, and the best solution for expression higher plant genes may be the development of more compatible hosts such as cyanobacteria. Perhaps, the differences in codon usage between the two classes of flowering plants are of greater concern at this time, since these differences have implications for intergenic transfer of genes between monocots and dicots. Since monocots express both highly biased and less biased genes among their abundant proteins, it would appear that the codon usage of dicots would not present a great barrier to these genes being expressed in monocots. However, the high-level expression of the highly biased genes of monocots in the dicot system may be a problem. Genes such as *Nar* and *Nir* of corn as well as the corn chloroplastic *Gap* (8,13,17,28) may be very useful for experimentally testing the limits of intergenic transfer in order to gain a better understanding of biased codon usage in higher plants.

ACKNOWLEDGEMENT

We thank Dr. John H. Adler, Department of Biological Sciences, Michigan Technological University, for assistance with the classification of plant species for Tables III and IV and other helpful comments.

LITERATURE CITED

1. **Aota S, Ikemura T** (1986) Diversity in G + C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res **14:** 6345–6355
2. **Back E, Burkhart W, Moyer M, Privalle L, Rothstein S** (1988) Isolation of cDNA clones coding for spinach nitrite reductase: complete sequence and nitrate induction. Mol Gen Genet **212:** 20–26
3. **Belanger FC, Kriz AL** (1989) Molecular characterization of the major maize embryo globulin encoded by the *Glb 1* gene. Plant Physiol **91:** 636–643
4. **Bennetzen JL, Hall BD** (1982) Codon Selection in Yeast. J Biol Chem **257:** 3026–3031
5. **Bethards LA, Skadsen RW, Scandalios JG** (1987) Isolation and characterization of a cDNA clone for the *Cat2* gene in maize and its homology with other catalases. Proc Natl Acad Sci USA **84:** 6830–6834

6. **Beutler E, Gelbart T, Han J, Koziol JA, Beutler B** (1989) Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. Proc Natl Acad Sci USA **86:** 192–196
7. **Bird AP** (1986) CpG-rich islands and the function of DNA methylation. Nature **321:** 209–213
8. **Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R** (1987) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Evol **26:** 320–328
9. **Cannon RE, White JA, Scandalios JG** (1987) Cloning of cDNA for maize superoxide dismutase 2 (SOD2). Proc Natl Acad Sci USA **84:** 179–183
10. **Crawford NM, Smith M, Bellissimo D, Davis RW** (1988) Sequence and nitrate regulation of the *Arabidopsis thaliana* mRNA encoding nitrate reductase, a metalloflavoprotein with three functional domains. Proc Natl Acad Sci USA **85:** 5006–5010
11. **Cronquist A** (1981) An Integrated System of Classification of Flowering Plants, Columbia University Press, New York
12. **de Boer HA, Kastein RA** (1986) Biased codon usage: an exploration of its role in optimization of translation. *In* W Reznikoff, L Gold, eds, Maximizing Gene Expression, Butterworths, Boston, pp 225–285
13. **Gowri G, Campbell WH** (1989) cDNA clones for corn leaf NADH:nitrate reductase and chloroplast NAD(P)$^+$:glyceraldehyde-3-phosphate dehydrogenase. Characterization of the clones and analysis of the expression of the genes in leaves as influenced by nitrate in the light and dark. Plant Physiol **90:** 792–798
14. **Grantham R** (1980) Workings of the genetic code. Trends Biochem Sci **5:** 327–331
15. **Harper JF, Surowy TK, Sussman MR** (1989) Molecular cloning and sequence of cDNA encoding the plasma membrane proton pump (H^+ -ATPase) of *Arabidopsis thaliana.* Proc Natl Acad Sci USA **86:** 1234–1238
16. **Krebbers E, Seurinck J, Herdies L, Cashmore AR, Timko MP** (1988) Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana.* Plant Mol Biol **11:** 745–759
17. **Lahners K, Kramer V, Back E, Privalle L, Rothstein S** (1988) Molecular cloning of complementary DNA encoding maize nitrite reductase. Molecular analysis and nitrate induction. Plant Physiol **88:** 741–746
18. **Luan S, Bogorad L** (1989) Nucleotide sequence of two genes encoding the light harvesting chlorophyll a/b binding protein of rice. Nucleic Acids Res **17:** 2357–2358
19. **Manolson MF, Ouellette FBF, Filion M, Poole RJ** (1988) cDNA sequence and homologies of the "57-kDa" nucleotide-binding subunit of the vacuolar ATPase from *Arabidopsis.* J Biol Chem **263:** 17987–17994
20. **Martin W, Gieri A, Saedler H** (1989) Molecular evidence for pre-Cretaceous angiosperms origins. Nature **339:** 46–48
21. **Matsuoka M, Ozeki Y, Yamamoto N, Hirano H, Kano-Murakami Y, Tanaka Y** (1988) Primary structure of maize pyruvate, orthophosphate dikinase as deduced from cDNA sequence. J Biol Chem **263:** 11080–11083
22. **Murray EE, Lotzer J, Eberle M** (1989) Codon usage in plant genes. Nucleic Acids Res **17:** 477–498
23. **Ohikawa J, Okada N, Shinmyo A, Takano M** (1989) Primary structure of cucumber (*Cucumis sativus*) ascorbate oxidase deduced from cDNA sequence: homology with blue copper proteins and tissue-specific expression. Proc Natl Acad Sci USA **86:** 1239–1243
24. **Ohyama K, Fukuzawa H, Kohchi T, Sano T, Sano S, Shirai H** (1988) Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification. J Mol Biol **203:** 281–298
25. **Perl-Treves R, Nacmias B, Aviv D, Zeelon EP, Galun E** (1988) Isolation of two cDNA clones from tomato containing two different superoxide dismutase sequences. Plant Mol Biol **11:** 609–623
26. **Pfitzinger H, Guillemaut P, Weil J-H, Pillay DTN** (1987) Adjustment of the tRNA population to the codon usage in chloroplasts. Nucleic Acids Res **15:** 1377–1386
27. **Post-Beittenmiller MA, Hlousek-Radojcic A, Ohlrogge JB** (1989) DNA sequence of a genomic clone encoding an *Arabidopsis* acyl carrier protein (ACP). Nucleic Acids Res **17:** 1777
28. **Quigley F, Martin WF, Cerff R** (1988) Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc Natl Acad Sci USA **85:** 2672–2676
29. **Redinbaugh MG, Wadsworth GJ, Scandalios JG** (1988) Characterization of catalase transcripts and differential expression in maize. Biochim Biophys Acta **951:** 104–116
30. **Springer BA, Sligar SG** (1987) High-level expression of sperm whale myoglobin in *Escherichia coli.* Proc Natl Acad Sci USA **84:** 8961–8965
31. **Tingey SV, Walker EL, Coruzzi GM** (1987) Glutamine synthetase genes of pea encode distinct polypeptides which are differently expressed in leaves, roots and nodules. EMBO J **6:** 1–9
32. **Wakasugi T, Ohme M, Shinozaki K, Sugiura M** (1986) Structures of tobacco chloroplast genes for $tRNA^{Ile}$ (CAU), $tRNA^{Leu}$ (CAA), $tRNA^{Cys}$(GCA), $tRNA^{Ser}$ (UGA) and $tRNA^{Thr}$ (GGU): a compilation of tRNA genes from tobacco chloroplasts. Plant Mol Biol **7:** 385–392
33. **Werneke JM, Ogren WL** (1989) Structure of an *Arabidopsis thaliana* cDNA encoding rubisco activase. Nucleic Acids Res **17:** 2871
34. **White JA, Scandalios JG** (1988) Isolation and characterization of a cDNA for mitochondrial manganese superoxide dismutase (SOD-3) of maize and its relation to other manganese superoxide dismutase. Biochim Biophys Acta **951:** 61–70
35. **Yamamoto N, Kano-Murakami Y, Matsuoka M, Ohashi Y, Tanaka Y** (1988) Nucleotide sequence of a full length cDNA clone of ribulose bisphosphate carboxylase small subunit gene from green dark-grown pine (*Pinus tunbergii*) seedlings. Nucleic Acids Res **16:** 11830
36. **Yamamoto N, Matsuoka M, Kano-Murakami Y, Tanaka Y, Ohashi Y** (1988) Nucleotide sequence of a full length cDNA clone of light harvesting chlorophyll a/b binding protein gene from green dark-grown pine (*Pinus tunbergii*) seedlings. Nucleic Acids Res **16:** 11829
37. **Yoshinaga K, Ohta T, Suzuki Y, Sugiura M** (1988) *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of a possible gene for the ß' subunit of RNA polymerase. Plant Mol Biol **10:** 245–250