Zone IV (0.0–0.8 m), dated modern, also shows *increased* levels of these metals, indicating again an unstable period. The average concentration values of these elements are 0.87%, 0.91% and 0.16%. This is also supported by the sharp decline of mangrove vegetation. The unfavourable landscape condition may have been caused by higher biotic activity (palynological zone IV). Large areas were cleared during 1955–60 for the construction of Paradip port. Its effect was noted by Gupta and Yadav in the palynological analyses of the sediments from Paradip Lake. The decreased percentage of mangrove pollen can therefore be linked to the higher erosion rate from poor vegetational cover.

The environment around Paradip, changed from stable landform conditions around 450 years to unstable landform conditions around 290 years to present time on the basis of elemental variation and mangrove vegetation.

1. Mackereth, F. J. H., *Philos. Trans.*, 1965, **B250**, 765.
2. Mackereth, F. J. H, *Proc. R. Soc., London, Ser. B.*, 1965, **161**, 295.
3. Engstrom, D. R. and Wright, H. E., Jr., in *Lake Sediments and Environmental History* (eds. Haworth, E. Y. and Lund, J. W. G.), Leicester University Press, 1984, pp. 11–67.
4. Engstrom, D. R. and Swain, E. B., *Hydrobiologia*, 1986, **143**, 37–44.
5. Gupta, H. P. and Yadav, R. R., *Palaeobotanist*, 1990, **38**, 359–369.
6. Rajagopalan, G., Mitre, Vishnu, and Sekar, B., *Radiocarbon*, 1978, **20**(3), 398–404.

ACKNOWLEDGEMENTS. We are grateful to the Head, Quaternary Laboratory, BSIP, for providing the samples of Paradip for chemical analysis. Our thanks are due to Prof. C. P. Sharma, Head, Botany Department, Lucknow University for providing analytical facilities.

9 March 1992; revised accepted 29 July 1992

Comparison of codon usage in genes of plant viruses and their hosts

S. N. Sudha, S. Krishnaswamy* and Vaithilingam Sekar
Department of Biotechnology, *Bioinformatics Centre, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India

A preliminary analysis of plant viral genes was made using all the available 85 viral sequences from GenEMBL database. It is found that in plant viruses and their hosts, amino acids with a high frequency of occurrence have a similar distribution of codons (codon usage profile) within their set of synonymous codons as indicated by the high match coefficient. However, the codon bias indices of the plant viruses with respect to their hosts are low, indicating that the preferred codons of the host are not used often in the plant viruses.

CODON usage in an organism has been thought of as one of the strategies used for regulating gene expression[1,2]. Detailed analyses on the patterns of codon usage in several organisms including bacteria, bacteriophage, yeast, fruitfly, mammalian viruses, man and plants have been reported. However, no information is available till date on the codon usage pattern of plant viral genes. Many viruses capable of infecting a wide variety of plant species are known. Of these, some viruses (e.g. barley yellow mosaic virus) are known to infect only one plant species; while others (e.g. tobacco rattle virus) could infect more than 400 species of plants[3]. As considerable molecular information is available for the plant and the viral genomes, a comparison of codon usage in genes of plants and the plant viruses was made to see if the virus–host relationship can be explained in terms of codon usage. It has been established that monocot and dicot plant genes differ considerably in their codon usage[4]. Recently, Wada *et al.*[5] have shown that the codon usage pattern remains unaltered even when codons from several genes (with varying functions) of an organism are considered in summation. Hence, patterns of the codon usage of predominantly monocot infecting viruses (PMIVs) as a group and predominantly dicot infecting viruses (PDIVs) as another group were compared with those of the monocot and dicot plants respectively. From such an analysis we have attempted to find out whether the codon usage of plant viruses has any role in determining host range.

The 85 plant viral genes included in the sample are from the database GenEMBL (release 24.0). Coat protein genes were primarily analysed here as they are known to be highly expressed[6,7]. The data for codon usage by dicot and monocot plants were from Murray *et al*[4].

Relevant sequences from GenEMBL were extracted using the University of Wisconsin Genetics Computer Group program (version 6.2). Codon usage tables were compiled using the program CODONFREQUENCY from the UWGCG package[8]. The program CODON was developed by us to determine the choice of preferred codons and analyse the correlation between the distribution of codons among codon usage tables.

The codon bias index which determines the level of usage of preferred codons in a gene has been defined earlier[9]. In this study, the same definition of the codon bias index is used. However, the selection of preferred codons is not based on the percentage occurrence of codons within a set of synonymous codons. The significance of a percentage occurrence depends on the number of synonymous codons. For example, 50% occurrence for a codon is not significant when there are only two synonymous codons but can be considered preferred if there are three or more synonymous codons in the set. Hence, we have calculated the standard deviation in percentage occurrence within the set of synonymous codons for each amino acid. Those codons in a synonymous set whose percentage occurrence is above a given cut-off limit times the standard deviation are flagged

as 'preferred'. These preferred codons of the plant, for example, are used to determine if the codons of the plant virus are biased towards the plant codon usage. The singly occurring codons of Met and Trp, and the equally distributed codons in a synonymous set are flagged as preferred, since the standard deviation is zero in these cases. The effect of flagging all the codons of an amino acid as preferred is equivalent to not including its contribution to the codon bias index. Test calculations (data not shown) based on genes of the yeast *Saccharomyces cerevisiae* indicate that (i) the codon bias indices obtained by this procedure are similar to those calculated earlier[9] and (ii) increasing the cut-off limit for flagging the preferred codons does not reduce the discrimination between highly and poorly expressed proteins[9]. The best choice of the cut-off limit is one where the minimum number of codons is chosen as preferred while no amino acid is left without at least one preferred codon (Table 1).

In order to compare two codon usage profiles, for each amino acid a correlation coefficient is calculated using the percentage occurrence of codons within the synonymous set. The match coefficient for the entire set of amino acids is then defined as the sum of the weighted correlation coefficients. The weight for each amino acid is calculated as the ratio of the total number of codons for that amino acid to the total number of codons for all the amino acids of the two tables. This is done to take into account the differences in amino acid composition.

$$\text{Thus, the match coefficient} = \sum_{i}^{N} w_i c_i,$$

where N = all amino acids except Met, Trp,

w_i = weights

$$= \frac{\text{total number of codons for } i\text{th amino acid}}{\text{total number of codons}}$$

c_i = correlation coefficient of synonymous codon usage

$$= \frac{\sum_{j}^{n_i} (p_{ji} - \bar{p}_i)(q_{ji} - \bar{q}_i)}{\left[\sum_{j}^{n_i} (p_{ji} - \bar{p}_i)^2\right]^{\frac{1}{2}} \left[\sum_{j}^{n_i} (q_{ji} - \bar{q}_i)^2\right]^{\frac{1}{2}}}$$

where p_{ji} is percentage occurrence of jth codon of ith amino acid of one table, q_{ji} the percentage occurrence of jth codon of ith amino acid of the other table, $\bar{p}_i$, $\bar{q}_i$ are the mean and n_i the number of codons of the ith amino acid.

The match coefficient as defined here can vary from –1 (for anticorrelated profiles) through 0 (for profiles without correlation) to +1 (for identical profiles). A high-match coefficient indicates that amino acids with a higher frequency of occurrence have a similar distribution of

Table 1. Percentage occurrence of synonymous codons for plant viruses and plants. Preferred codons are marked by (*). (The cut-off limit in standard deviation units for flagging preferred codons is given in parentheses)

Plants					Plant viruses		
MP (2.0)	DP (2.0)	PL (2.5)	Amino acid	Codon	PV (3.5)	PDIV (3)	PMIV (3.5)
21	12	15	GLY	GGG	18	18	15
17	38*	32*	GLY	GGA	31*	32*	32*
18	34*	29*	GLY	GGT	31*	32*	30*
43*	16	24*	GLY	GGC	20	20*	23
75*	51*	57*	GLU	GAG	43*	43*	46*
25	49*	43*	GLU	GAA	57*	57*	54*
27	58*	50*	ASP	GAT	57*	58*	49*
73*	42*	50*	ASP	GAC	43*	42*	51*
36*	29*	31	VAL	GTG	27*	27*	31*
8	12	11	VAL	GTA	16	16	18
19	39*	34*	VAL	GTT	34*	34*	21*
37*	20	24*	VAL	GTC	23*	23*	21
22*	6	11	ALA	GCG	14	13	19
16	25	22	ALA	GCA	25	26*	21*
24*	42*	37*	ALA	GCT	36*	37*	32*
38*	27*	30*	ALA	GCC	25	24	28*
26*	25*	25*	ARG	AGG	20	21*	18*
9	30*	24*	ARG	AGA	28*	29*	23*
13	4	7	ARG	CGG	9	8	11
4	8	7	ARG	CGA	14	14	16*
12	21*	18	ARG	CGT	15	15	18*
36*	11	18	ARG	CGC	13	13	14*
8	14*	13*	SER	AGT	17*	17*	16*
26*	18*	20	SER	AGC	13	13	14*
14*	6	8	SER	TCG	10	9	8
11	19*	17*	SER	TCA	19*	19*	19*
15*	25*	22*	SER	TCT	23*	24*	22*
25*	18*	20*	SER	TCC	18*	18*	18*
86*	61*	66*	LYS	AAG	48*	48*	50*
14	39*	34*	LYS	AAA	52*	52*	50*
25	45*	41*	ASN	AAT	55*	55*	52*
75*	55*	59*	ASN	AAC	45*	45*	48*
11	18	16	ILE	ATA	24	24*	19
24	45*	40*	ILE	ATT	41*	41*	40*
64*	37*	44*	ILE	ATC	36*	35*	41*
20	8	11	THR	ACG	14	13	18
14	27*	24*	THR	ACA	28*	29*	23*
19	35*	31*	THR	ACT	33*	33*	34*
46*	30*	34*	THR	ACC	25*	26*	25*
30	44*	40*	CYS	TGT	55*	54*	57*
70*	56*	60*	CYS	TGC	45*	46*	43*
25	45*	40*	PHE	TTT	52*	52*	52*
75*	55*	60*	PHE	TTC	48*	48*	48*
46*	41*	43*	GLN	CAG	41*	41*	43*
54*	59*	57*	GLN	CAA	59*	59*	57*
33	54*	48*	HIS	CAT	50*	50*	52*
67*	46*	52*	HIS	CAC	50*	50*	48*
23*	9	13	PRO	CCG	14	14	19*
34*	42*	39*	PRO	CCA	32*	32*	30*
17*	31*	28*	PRO	CCT	30*	30*	28*
26*	18	20	PRO	CCC	24*	25*	23*

PL, plants; MP, monocot plants; DP, dicot plants; PV, plant viruses; PDIV, predominantly dicot infecting viruses; PMIV, predominantly monocot infecting viruses.

Table 2. Match coefficients and codon bias indices of plant viruses with respect to plants. (The cut-off limit in standard deviation units for flagging preferred/codons is given in parentheses.)

Plants	Plant viruses	PV (3.5)	PDIV (3)	PMIV (3.5)
PL	M.C	0.44	0.43	0.50
(2.5)	C.B	0.23	0.23	0.24
DP	M C	0.52	0.50	0.40
(2)	C.B	0.25	0.26	0.15
MP	M C	−0.15	−0.16	0.45
(2)	C.B	−0.03	−0.03	0 01

MP, monocot plants; DP, dicot plants; PL, plants; PV, plant viruses; PDIV, predominantly dicot infecting viruses; PMIV, predominantly monocot infecting viruses.

codons within their set of synonymous codons. The match coefficient and codon bias index need not relate to each other as can be seen from Table 2. This is because the preferred codons among the sets of synonymous codons being compared need not be the same even when the distribution of codons is similar. For example, the percentage occurrences for preferred codons of the amino acid alanine are: 36% (GCT) for plant viruses and 37%, 30% (GCT, GCC) for plants; whereas the pattern of variations of percentage occurrence of the four synonymous codons follows a similar pattern for both plant viruses and plants (Table 1).

Results of *preliminary* analysis based on codon bias index and match coefficient are summarized in Tables 1 and 2. Plant viruses differ from plants in their choice of preferred codons as reflected by the consistently low codon bias index. However, the codon usage profiles of PDIVs and dicot plants are similar as seen from the high values of the match coefficients. The codon usage profiles of PMIVs, on the other hand, do not show any distinction with respect to monocot or dicot plants. Such a correlation of the codon usage profiles of the plant viruses and their hosts suggests a possible common mutational bias. Since infectivity of a virus could be related to the levels of viral gene expression, it would be necessary to experimentally determine whether alteration of the viral codon bias leads to changes in infectivity through variation in the levels of viral gene expression. Our present analysis suggests that the predominantly dicot infecting viruses might be better candidates for such an experiment as they show a greater similarity in codon profile to their hosts.

1. de Boer, H. A. and Kastelein, R. A., in *Biased Codon Usage: An Exploration of its Role in Optimization of Translation* (eds. Reznikoff, W. and Gold, L.) Buttersworth, Boston, 1986, pp. 225–286.
2. Andersson, G. E. and Kurland, C. G., *Microbiol. Rev.*, 1990, **54**, 198–210.
3. AAB Descriptions of Plant Viruses, December 1989, no. 346.
4. Murray, E. E., Lotzer, J. and Eberle, M., *Nucleic Acids Res.*, 1989, **17**, 477–498.
5. Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T. and Ikemura, T., *Nucleic Acids Res.*, 1991, **19**, Supplement, 1981–1986.
6. Abel, P. P., Nelson, R. S., De, B., Hoffman, N., Rogers, S. G., Fraley, R. T. and Beachy, R. N., *Science*, 1986, **232**, 738–743.
7. Beachy, R. N., Loesch-Fries, S. and Turner, N. E., *Annu. Rev. Phytopathol.*, 1990, **28**, 451–474.
8. Devereux, J., Haberli, P. and Smithies, O., *Nucleic Acids Res.*, 1984, **12**, 387–395.
9. Bennetzen, J. L. and Hall, B. D., *J. Biol. Chem.*, 1982, **257**, 3026–3031.

ACKNOWLEDGEMENTS. We are grateful to Elizabeth Murray, Promega Corp., for suggesting this problem and for her help during the initial stages of this work. We thank R. Usha and D. Banhatti for valuable discussions. Assistance offered by R. Neela in the preparation of this manuscript is acknowledged. The program CODON is written in FORTRAN and is available upon request from S. K.

Received 9 April 1992; revised accepted 10 August 1992