Gene, 152 (1995) 127–132

GENE 08497

Compositional properties of nuclear genes from *Plasmodium falciparum**

(Malaria; parasites; housekeeping genes; antigen genes; *Staphylococcus*)

Hector Musto**, Helena Rodriguez-Maseda** and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris, France

Received by L. Pereira da Silva: 31 July 1994; Revised/Accepted: 31 August 1994; Received at publishers: 10 October 1994

SUMMARY

We have analyzed the compositional distributions of coding sequences and their different codon positions, as well as the codon usage of the nuclear genes of *Plasmodium falciparum*, a parasite characterized by an extremely GC-poor genome. As expected, coding sequences are AT-rich, codon usage is strongly biased towards A or T in third codon positions, and some particular amino acids (aa) are especially abundant in the encoded proteins. Remarkably, however, no difference was detected between housekeeping (HK) and antigen (*Ag*) genes, in spite of differences in expression level and evolutionary constraints. Moreover, all the features found in *P. falciparum* are very similar to those found in a bacterium characterized by a very GC-poor genome, *Staphylococcus aureus*. These findings stress the importance of compositional constraints in determining codon usage and aa utilisation.

INTRODUCTION

A striking feature of *Plasmodium falciparum*, a unicellular parasite responsible for the most virulent and widespread form of human malaria, is that it hosts the GC-poorest nuclear genome known so far (Pollack et al., 1982; McCutchan et al., 1984). This genome, which only comprises 3×10^7 bp (Weber, 1988) organized in 14 chromosomes (Kemp et al., 1987a; Wellems et al., 1987), is, therefore, an excellent model to study compositional constraints and their effects.

The analysis of nt sequence data has provided useful information about the genes encoded in the nuclear genome of *P. falciparum*. The most relevant features are the following: (*i*) the coding strand is purine rich; (*ii*) A is predominant in all codon positions; (*iii*) the third codon positions are extremely AT-rich, and, as a consequence, codon usage is strongly biased (Weber, 1987; Saul and Battistutta, 1988). The most frequent dinucleotides are, as expected, those containing exclusively A and/or T, whereas the least common ones are those only composed by C and/or G. Furthermore, CG, TA and AC were lower, and TG, CC and CA were higher than the expected frequencies (Weber, 1988; Hyde and Sims, 1987).

We report here an up-to-date analysis of the nuclear coding sequences from *P. falciparum*, which now comprise 175 kb. We found that the trends described previously with more limited sets of data (see above) are still valid. Taking advantage of the increased number of sequences which are now available, we tried to understand whether the biases already noted are species-specific or deter-

Correspondence to: Dr. G. Bernardi, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France. Tel. (33-1) 4329-5824; Fax (33-1) 4427-7977;
e-mail: Bernardi@citi2.fr

*Presented at the UNESCO-WHO Meeting on Combatting Malaria, Paris, France, 19–21 January 1994.

**Permanent address: (H.M.) Departamento de Bioquimica, Facultad de Ciencias, Tristan Narvaja 1674, Montevideo 11200, Uruguay. Fax (598-2) 409-973; (H.M. and H..R.-M.) Departamento de Genética, Facultad de Medicina, Gral. Flores 2144, Montevideo, Uruguay. Fax (598-2) 949-563.

Abbreviations: *A.*, *Azotobacter*, aa, amino acid(s); Ag, antigen(s); bp, base pair(s); AT, % of adenine + thymine; GC, % of guanine + cytosine; HK, housekeeping; kb, kilobase(s) or 1000 bp; N, any nucleoside; nt, nucleotide(s); *P.*, *Plasmodium*.; R, purine (A or G); RSCU, relative synonymous codon usage; *S.*, *Staphylococcus*; *T.*, *Trypanosoma*; Y, pyrimidine (C or T).

SSDI 0378-1119(94)00708-X

mined by the composition of the genome, and whether they are different for sequences that display different expression levels and certainly are under different evolutionary pressures, like HK and *Ag* genes.

EXPERIMENTAL AND DISCUSSION

(a) Compositional distributions

Fig. 1 shows the histograms of the compositional (GC) distributions of the three codon positions of *P. falciparum*. The GC levels of first codon positions range from 15 to 82.5%, and are multimodal. The highest values are reached by antigen (Ag) coding sequences (*Ag*), while housekeeping (HK) genes only attain 50%. This is due to the fact that most antigens from this organism contain repetitive aa sequences usually encoded by codons rich in G and/or C in first position, like His (coded by CAY), Ala (GCN), Gln (CAR), etc. (for reviews, see Kemp et al., 1987b; Weber, 1988; McConkey et al., 1990). This imposes a bias on the frequency of bases, and hence, on GC levels.

The histogram of second codon positions ranges from 15 to 55%. All sequences are clustered in the histogram with the exception of three *Ag* genes, which (depending on the sequence) are rich in Ser (TCN and AGY), Thr (ACN), Ala (GCN) and Gly (GGN).

The most striking distribution is that of third codon positions. Its mean GC value is 17%, which is 6% lower than the mean value of the third codon positions of *Mycoplasma capricolum* and *Staphylococcus aureus*, two bacteria with extremely low GC levels (Muto and Osawa, 1987; D'Onofrio and Bernardi, 1992). The analysis of the histogram shows that (*i*) all sequences are comprised in a very narrow range of values (27.5%); (*ii*) HK and *Ag* sequences are roughly equally distributed; (*iii*) the distribution is asymmetrical, since it trails towards relatively higher GC levels.

Finally, the order of GC levels among the three codon positions is I > II > III as already observed in prokaryotic genomes with extremely low GC levels (D'Onofrio and Bernardi, 1992).

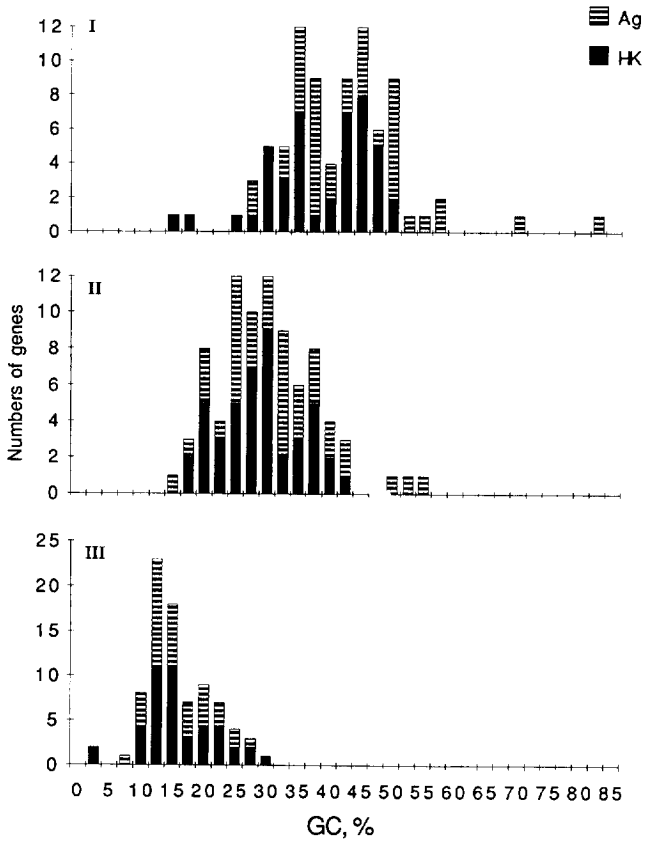

Fig. 1. Compositional patterns of the three codon positions, denoted as I, II and III in *P. falciparum*. Abscissae and ordinates display the GC% and number of sequences, respectively. *Ag* and HK genes are indicated (see also Table I).

(b) Base composition

Table I displays the percentage of bases in different codon position for HK genes, *Ag* genes and all genes. A is the most frequent base in the coding strand in all codon positions, followed by G in first codon position and by T in second and third codon positions. These results

TABLE I

Base frequencies for various codon positions in *P. falciparum*

Codon positions[a]	Nucleotide frequencies				
	A	C	G	T	GC
HK[b]					
I	38.0	10.3	29.3	22.4	39.6
II	40.7	17.0	13.1	29.3	30.1
III	41.9	8.2	8.9	41.0	17.1
Ag[c]					
I	36.8	13.1	31.8	18.3	44.9
II	45.1	19.8	13.0	22.1	32.8
III	43.0	9.9	7.2	39.9	17.1
Total[d]					
I	37.4	11.6	30.5	20.5	42.1
II	42.8	18.3	13.0	25.9	31.3
III	42.4	9.0	8.1	40.5	17.1

[a] I, II and III are first, second and third codon position, respectively. The sequences analyzed were obtained from release 74 of GenBank and the ACNUC retrieval system was used (Gouy et al., 1984). The accession numbers and mnemonics are available upon request. The genes were classified according to definition and keywords given by the authors. 83 sequences (44 HK and 39 *Ag*) with a size of no less than 210 bp were analyzed. Overall 174618 bp were analyzed.

[b] Values for housekeeping genes.

[c] Values for antigen-encoding genes.

[d] Total values.

confirm the trends described earlier (Saul and Battistutta, 1988). Both groups of sequences exhibit the same biases, the order of preference for each codon position being the same. The only difference is that in third codon position G is very slightly preferred over C in HK genes while the reverse is true in *Ag* genes.

Concerning GC levels, *Ag* sequences are slightly GC richer (+5%) in first codon positions, while in second codon positions there are no significant differences, and in third codon positions the GC levels of the two groups of sequences are the same.

Because of the homogeneously low GC levels in third codon positions, the frequencies of bases in the different codon positions (and their GC levels) are strikingly similar between HK and *Ag* sequences. Therefore, genes that have different expression levels and that code for proteins which are under different evolutionary constraints, are compositionally similar in third codon positions. This is not a characteristic of unicellular parasites, since HK and *Ag* genes of *Trypanosoma brucei* and *T. cruzi* show different biases (Musto et al., 1994).

(c) Amino-acid frequencies

Table II shows the aa frequencies in HK and Ag proteins. The order of preferences and frequencies of both groups of sequences are similar, since there are ten aa (Lys, Leu, Val, Gly, Cys, Trp, Asn, Ser, Asp and Met) occupying the same ranking position or differing by only one position between HK and Ag proteins. Furthermore, four aa differ in frequency by less than 5% (Ala, Gly, Thr and Asn) and another group of three residues (Val, Asp and Cys) display differences in frequencies between 5 and 7.5%.

TABLE II

Amino-acid frequencies[a]

aa	HK	Ag	*Pf*	*Sa*	*Av*
Lys	9.7	11.2	10.5	8.9	5.0
Asn	9.6	9.8	9.7	7.1	3.2
Glu	6.9	10.8	9.0	7.3	7.7
Leu	8.6	7.6	8.1	7.7	9.4
Ile	8.6	5.8	7.1	7.0	5.5
Ser	6.5	7.4	7.0	6.3	5.2
Asp	6.3	6.6	6.5	6.0	6.0
Val	5.1	4.8	.4.9	5.8	7.4
Thr	4.7	4.7	4.7	6.2	4.6
Gly	4.8	4.6	4.7	6.4	8.2
Tyr	5.0	3.4	4.1	4.4	2.7
Ala	4.0	4.1	4.1	5.7	9.7
Phe	4.3	3.0	3.6	4.2	3.3
Gln	2.7	4.0	3.4	3.9	3.3
Pro	3.1	3.6	3.3	3.8	4.7
Arg	3.4	2.6	2.9	3.3	5.8
His	2.4	2.6	2.5	2.4	3.3
Met	2.4	1.5	1.9	2.0	2.9
Cys	1.6	1.5	1.6	0.6	1.8
Trp	0.5	0.4	0.4	1.0	1.0

[a] The aa frequencies are given per 100 aa. *Pf*, *Sa* and *Av* are aa frequencies in *P. falciparum*, *S. aureus* and *A. vinelandii*, respectively. The data of *S. aureus* and *A. vinelandii* are from Wada et al. (1992). For HK and Ag, see Table I, footnotes b and c.

The total values (Table II, *Pf*) indicate that the preferred residues are Lys, Asn, Glu, Leu and Ile. Together, these aa constitute 44.4% of all residues.

Remarkably, all of them are encoded by A or T in second codon position. Furthermore, the two most frequent aa (Lys and Asn, which represent 20% of all aa) are encoded by AAR and AAY, respectively. Among the ten most frequent residues (representing 72.2% of the total), 77.3% are encoded by A or T in second codon position, while only 22.7% are encoded by C or G in the same position.

Two points emerge from these results. First, the two different groups of proteins tend to be constructed with rather similar frequencies of aa. This point is not trivial, since the physiological role of Ag and HK proteins are completely different, and one should expect different biases in the aa usage frequencies. Second, there is a remarkable tendency to use codons that are extremely rich in A and/or T, not only in the third codon positions, but also in first and second codon positions.

To study whether these features are species specific or due to the extremely biased composition of the genome, we compared the aa frequencies of *P. falciparum* with that of *S. aureus* (Table II, *Sa*). This showed that the most frequent residues in the two species are the same. Indeed, in *S. aureus* the order of preference is Lys, Leu, Glu, Asn and Ile; together, these aa comprise 38% of all residues. In this bacterium, the ten most frequent aa comprise 68.7% of the total, and among these, 72.5% are encoded by A or T in the second codon position and 27.5% are coded by C or G in the same position. All these figures are very similar to the ones displayed by *P. falciparum*. As a control, we investigated the aa frequencies of the GC-rich bacterium *Azotobacter vinelandii* (Table II, *Av*); in this case the most frequent residue is Ala (encoded by GCN) and Gly (GGN) is third, while Lys, the most frequent in the AT-rich genomes studied here, is only 10th.

These results clearly suggest that the features displayed by *P. falciparum* and *S. aureus* are not species specific but are the consequence of the extremely biased composition of their genomes.

(d) Codon usage

Table III shows the RSCU (Sharp et al., 1986) values for HK and Ag genes. It is evident that the values for all codons in the two groups of genes are practically the same. This has the important implication that in this

TABLE III

RSCU values of HK and *Ag* genes of *P. falciparum*[a]

aa	Codon	HK	*Ag*	aa	Codon	HK	*Ag*	aa	Codon	HK	*Ag*	aa	Codon	HK	*Ag*
Phe	TTT	1.7	1.4	Ser	TCT	1.5	1.4	Tyr	TAT	1.8	1.7	Cys	TGT	1.8	1.8
Phe	TTC	0.3	0.6	Ser	TCC	0.4	0.4	Tyr	TAC	0.2	0.3	Cys	TGC	0.2	0.3
Leu	TTA	4.4	3.8	Ser	TCA	1.9	1.9	End	TAA	*	*	End	TGA	*	*
Leu	TTG	0.7	0.8	Ser	TCG	0.2	0.1	End	TAG	*	*	Trp	TGG	1.0	1.0
Leu	CTT	0.5	0.9	Pro	CCT	1.4	1.2	His	CAT	1.6	1.5	Arg	CGT	0.8	0.8
Leu	CTC	0.1	0.2	Pro	CCC	0.3	0.3	His	CAC	0.4	0.5	Arg	CGC	0.0	0.1
Leu	CTA	0.3	0.3	Pro	CCA	2.3	2.5	Gln	CAA	1.7	1.8	Arg	CGA	0.4	0.4
Leu	CTG	0.1	0.1	Pro	CCG	0.1	0.1	Gln	CAG	0.3	0.2	Arg	CGG	0.0	0.0
Ile	ATT	1.3	1.4	Thr	ACT	1.2	1.4	Asn	AAT	1.7	1.6	Ser	AGT	1.8	1.7
Ile	ATC	0.2	0.2	Thr	ACC	0.6	0.4	Asn	AAC	0.3	0.4	Ser	AGC	0.3	0.5
Ile	ATA	1.5	1.4	Thr	ACA	1.9	2.0	Lys	AAA	1.8	1.7	Arg	AGA	4.3	4.1
Met	ATG	1.0	1.0	Thr	ACG	0.3	0.2	Lys	AAG	0.2	0.3	Arg	AGG	0.5	0.6
Val	GTT	1.9	1.7	Ala	GCT	2.0	1.9	Asp	GAT	1.8	1.8	Gly	GGT	1.9	1.8
Val	GTC	0.3	0.2	Ala	GCC	0.5	0.4	Asp	GAC	0.2	0.2	Gly	GGC	0.1	0.1
Val	GTA	1.6	1.8	Ala	GCA	1.5	1.7	Glu	GAA	1.8	1.8	Gly	GGA	1.8	1.9
Val	GTG	0.3	0.3	Ala	GCG	0.1	0.1	Glu	GAG	0.2	0.2	Gly	GGG	0.2	0.3

[a] HK and *Ag* are housekeeping and antigen sequences, respectively. RSCU were calculated according to Sharp et al. (1986). Asterisks indicate stop codons; these were not taken into account.

compositionally biased genome, the preferences among synonymous codons are not determined by the level of expression of each sequence but by the composition of the genome, since the most preferred synonymous codons always are those ending with A or T. The analysis of RSCU values of individual genes indicated that the biases are the same (data not shown).

Finally, Table IV displays the RSCU values for the total genes of *P. falciparum* (*Pf*) and *S. aureus* (*Sa*). It is evident that, in spite of minor variations, the codon usage pattern of both species is practically the same.

(e) Conclusions

(*1*) In unicellular organisms codon preferences are biased toward a group of 'major codons'. These biases have been correlated with the expression level of the protein molecules, in such a way that the higher the level of protein production, the higher the level of bias in codon usage (Grantham et al., 1981; Gouy and Gautier, 1982; Sharp et al., 1986; Shields and Sharp, 1987). Other explanations for this non-randomness of codon usage comprise the optimization of codon-anticodon interaction energy (Grosjean et al., 1978) and an adaptation of codons to the actual populations of isoaccepting t-RNAs (Ikemura, 1981a,b; 1982). A different hypothesis was that codon usage is a strategy associated with a given genome (Grantham, 1980; Grantham et al., 1980).

(*2*) Although some of these explanations may be partly correct, they are certainly incomplete since they do not take into account the composition of the genome. Indeed, multiple codon usages are found in compositionally compartmentalized genomes, like those of mammals (Bernardi et al., 1985; Bernardi and Bernardi, 1986; Bernardi, 1989; D'Onofrio et al., 1991).

(*3*) In the case of the genes analyzed here, we have noted that the biases in base composition of different codon positions, aa frequencies and codon preferences are almost identical in HK and Ag sequences. If we take into account that both the level of expression and the evolutionary constraints over these sequences are unlikely to be the same, the most obvious conclusions are (*i*) that the extremely biased composition of the genome (Pollack et al., 1982; McCutchan et al., 1984) is the major factor determining codon preferences and aa frequencies and (*ii*) that the compositional constraints (Bernardi and Bernardi, 1986) operate in the same direction over all the translated sequences and their codon positions.

(*4*) We have found that the biases displayed at the aa usage and codon preference levels are almost identical in *P. falciparum* and the bacterium *S. aureus*. These two organisms are, undoubtedly, very different, and probably the only feature that have in common is the extreme high genomic AT level. On the other hand, biases were very different when compared with *A. vinelandii*, a GC-rich bacterium. This further indicates that the biases primarily depend on the composition of the genome and not on the taxonomical proximity of two given organisms. Of course, since taxonomically closely related organisms usually display in general similar genomic compositions,

TABLE IV

RSCU values of *P. falciparum* (*Pf*) and *S. aureus* (*Sa*)[a]

aa	Codon	*Pf*	*Sa*	aa	Codon	*Pf*	*Sa*	aa	Codon	*Pf*	*Sa*	aa	Codon	*Pf*	*Sa*
Phe	TTT	1.6	1.4	Ser	TCT	1.4	1.4	Tyr	TAT	1.8	1.6	Cys	TGT	1.8	1.3
Phe	TTC	0.4	0.6	Ser	TCC	0.4	0.2	Tyr	TAC	0.2	0.4	Cys	TGC	0.2	0.7
Leu	TTA	4.1	3.3	Ser	TCA	1.9	1.8	End	TAA	*	*	End	TGA	*	*
Leu	TTG	0.7	0.9	Ser	TCG	0.2	0.3	End	TAG	*	*	Trp	TGG	1.0	1.0
Leu	CTT	0.7	0.8	Pro	CCT	1.3	1.5	His	CAT	1.6	1.5	Arg	CGT	0.8	1.8
Leu	CTC	0.1	0.2	Pro	CCC	0.3	0.2	His	CAC	0.4	0.5	Arg	CGC	0.0	0.6
Leu	CTA	0.3	0.6	Pro	CCA	2.4	1.8	Gln	CAA	1.8	1.7	Arg	CGA	0.4	0.6
Leu	CTG	0.1	0.2	Pro	CCG	0.1	0.5	Gln	CAG	0.2	0.3	Arg	CGG	0.0	0.2
Ile	ATT	1.4	1.6	Thr	ACT	1.3	1.2	Asn	AAT	1.7	1.4	Ser	AGT	1.7	1.5
Ile	ATC	0.2	0.5	Thr	ACC	0.5	0.2	Asn	AAC	0.3	0.6	Ser	AGC	0.4	0.8
Ile	ATA	1.4	0.9	Thr	ACA	2.0	2.1	Lys	AAA	1.7	1.6	Arg	AGA	4.2	2.5
Met	ATG	1.0	1.0	Thr	ACG	0.2	0.5	Lys	AAG	0.3	0.4	Arg	AGG	0.6	0.4
Val	GTT	1.8	1.6	Ala	GCT	1.9	1.4	Asp	GAT	1.8	1.6	Gly	GGT	1.9	1.8
Val	GTC	0.2	0.4	Ala	GCC	0.4	0.3	Asp	GAC	0.2	0.5	Gly	GGC	0.1	0.7
Val	GTA	1.7	1.5	Ala	GCA	1.6	1.9	Glu	GAA	1.8	1.6	Gly	GGA	1.8	1.2
Val	GTG	0.3	0.5	Ala	GCG	0.1	0.4	Glu	GAG	0.2	0.4	Gly	GGG	0.2	0.4

[a] *Pf* and *Sa* are *P. falciparum* and *S. aureus*, respectively. For other designations, see Table III, footnote a.

their codon preferences (and the aa frequencies) tend to be similar.

(5) Our results further support the hypothesis that among the various factors that certainly influence the architecture of the coding (and non-coding) sequences, the compositional constraints on the genome (or on isochores) are most relevant, as already proposed (for reviews, see Bernardi, 1993a,b). At least in some *P. falciparum* genes, however, certain coding regions do not follow the bias for AT-rich codons. A surprising third base bias has been found in various repeated regions (Scherf et al., 1988). For example, in the second position of the R6 repeated regions, 85% of the Glu codons are GAG and all of the 13 Leu codons are TTG. Based on these data it was suggested that many Ag genes consist of ancestral regions (non repetitive) and regions of recent origin (repeated blocks). The repeats might have evolved from ancestral short DNA sequences amplified as identical copies. Codons with a G and a C in the third position would be of too recent origin to be corrected by the compositional constraint to A or T.

REFERENCES

Bernardi, G.: The isochore organization of the human genome. Annu. Rev. Genet. 23 (1989) 637–661.

Bernardi, G.: The vertebrate genome: isochores and evolution. Mol. Biol. Evol. 10 (1993a) 186–204.

Bernardi, G.: The isochore organization of the human genome and its evolutionay history – a review. Gene 135 (1993b) 57–66.

Bernardi, G. and Bernardi, G.: Compositional constraints and genome evolution. J. Mol. Evol. 24 (1986) 1–11.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F.: The mosaic genome of warm-blooded vertebrates. Science 228 (1985) 953–956.

D'Onofrio, G. and Bernardi, G.: A universal compositional correlation among codon positions. Gene 110 (1992) 1–88.

D'Onofrio, G., Mouchiroud, D., Aïssani, B., Gautier, C. and Bernardi, G.: Correlations between the compositional properties of human genes, codon usage, and aminoacid composition of proteins. J. Mol. Evol. 32 (1991) 504–510.

Gouy, M. and Gautier, C.: Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10 (1982) 7055–7074.

Gouy, M., Milleret, F., Mugnier, C., Jacobzone, M. and Gautier, C.: ACNUC: a nucleic acid sequence data base and analysis system. Nucleic Acids Res. 12 (1984) 121–127.

Grantham, R.: Workings on the genetic code. Trends Biochem. Sci. 5 (1980) 327–333.

Grantham, R., Gautier, C., Gouy, M. and Paré, A.: Codon catalogue usage and the genome hypothesis. Nucleic Acids Res. 8 (1980) r49–r62.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R.: Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. 9 (1981) r43–r74.

Grosjean, H., Sankoff, D., MinJou, W., Fiers, W. and Cedergren, R.: Bacteriophage MS 2 RNA: a correlation between the stability of the codon:anticodon interaction and the choice of codewords. J. Mol. Evol. 12 (1978) 113–119.

Hyde, J.E. and Sims, P.F.G.: Anomalous dinucleotide frequencies in both coding and non-coding regions from the genome of the human malaria parasite *Plasmodium falciparum*. Gene 61 (1987) 177–187.

Ikemura, T.: Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146 (1981a) 1–21.

Ikemura, T.: Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its

protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. J. Mol. Biol. 151 (1981b) 389–409.

Ikemura, T.: Correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 158 (1982) 573–597.

Kemp, D., Thompson, J., Walliker, D. and Corcoran, L.: Molecular karyotype of *Plasmodium falciparum* conserved linkage groups and expendable histidine-rich protein genes. Proc. Natl. Acad. Sci. USA 84 (1987a) 672–7676.

Kemp, D., Coppel, R. and Anders, R.: Repetitive proteins and genes of malaria. Annu. Rev. Microbiol. 41 (1987b) 181–208.

McConkey, G., Waters, A. and McCutchan, T.: The generation of genetic diversity in malaria parasites. Annu. Rev. Microbiol. 44 (1990) 479–498.

McCutchan, T., Dame, J., Miller, L. and Barnwell, J.: Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. Science 225 (1984) 808–811.

Musto, H., Rodríguez-Maseda, H. and Bernardi, G.: The nuclear genomes of African and American trypanosomes are strikingly different. Gene 141 (1994) 63–69.

Muto, A. and Osawa, S.: The guanine and cytosine content of genomic DNA and bacterial evolution Proc. Natl. Acad. Sci. USA 84 (1987) 166–169.

Pollack, Y., Katzen, A., Spira, D. and Golenser, J.: The genome of *Plasmodium falciparum*. I: DNA composition. Nucleic Acids Res. 10 (1982) 539–546.

Saul, A. and Battistutta, D.: Codon usage in *Plasmodium falciparum*. Mol. Biochem. Parasitol. 27 (1988) 35–42.

Scherf, A., Hilbich, C., Sieg, K., Mattei, D., Mercereau-Puijalon, O. and Müller-Hill, B.: The 11–1 gene of the *Plasmodium falciparum* codes for distinct fast evolving repeats. EMBO J. 7 (1988) 1129–1137.

Sharp, P., Tuohy, T. and Mosurski, K.: Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14 (1986) 5125–5143.

Shields, X. and Sharp, P.: Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational constraints. Nucleic Acids Res. 15 (1987) 8023–8040.

Wada, K., Wada, Y., Ishibashi, F., Gojobori, T. and Ikemura, T.: Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. 20 (1992) 2111–2118.

Weber, L.: A review: molecular biology of malaria parasites. Exp. Parasitol. 66 (1988) 143–170.

Wellems, T., Walliker, D., Smith, C., do Rosario, V., Maloy, W., Howard, R., Carter, R. and McCutchan, T.: A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of *Plasmodium falciparum*. Cell 49 (1987) 33–642.