Mem. Inst. Oswaldo Cruz, Rio de Janeiro, Vol. *87* (4): 517-523, oct./dec. 1992

TRYPANOSOMATIDAE CODON USAGE AND GC DISTRIBUTION

GUILLERMINA ALONSO; PALMIRA GUEVARA & JOSE L. RAMIREZ+

Centro de Biología Celular, Universidad Central de Venezuela Apartado 47525, Caracas 1041A, Venezuela

A study of Trypanosomatidae GC distribution and codon usage is presented. The codon usage patterns in coincidence with the phylogenetical data are similar in Crithidia *and* Leishmania, *whereas they are more divergent in* Trypanosoma brucei *and* T. cruzi. *The analysis of the GC mutational pressure in these organisms reveals that* T. brucei, *and to a lesser extent* T. cruzi, *have evolved towards a more balanced use of all bases, whereas* Leishmania *and* Crithidia *retain features of a primeval genetic apparatus. Tables with the approximated GC mutational pressure in homologous genes, and codon usage in Trypanosomatidae are presented.*

Key words: Trypanosomatidae – codon usage – mutational pressure – GC content

The Trypanosomatidae family (order Kinetoplastida), includes a group of parasitic protozoa with a wide spectrum of hosts, vectors, reservoirs and life cycles (Vickerman, 1976). In the genera *Trypanosoma* and *Leishmania,* a clonal population structure has been proposed, but it is likely that other members of the group follow a similar trend (Tibayrenc et al., 1990).

The studies on these organisms are mainly focused on central metabolic enzymes and gene expression, in an effort to find differences between the host and the parasites that could be exploited in a rationale drug design (Opperdoes, 1983; Berman, 1988).

In the present work, as a contribution to the understanding of basic aspects of Trypanosomatidae genome organization and genome evolution, we have made a comparative study of nuclear gene sequences reported for *Crithidia fasciculata, Leishmania, Trypanosoma cruzi* and *T. brucei.* There is a large variation in the total GC content from one group to the other that is reflected in their codon usage. In coincidence with the genome hypothesis (Grantham et al., 1980) we found that closely related organisms share similar codon usage patterns, thus *Crithidia* and *Leishmania* have an almost identical codon usage, whereas we found a lesser degree of similarity between *T. brucei* and *T. cruzi.* A discussion on the evolution of codon usage and GC content within several genes of the Trypanosomatidae family is presented.

MATERIALS AND METHODS

The DNA sequences used in this study were retrieved from GenBank and EMBL databases. The sequences were analyzed through a DNASIS program version 5.0 from LKB (Hitachi, 1988). Table I compiles these sequences.

In the analysis of nuclear sequences we observed the following premises: The ribosomal genes were excluded; in the case of tandemly repeated genes the borders between a 3' flank and the neighbor 5' were defined by the presence or not of consesus sequences such as the 5' CT tract and 3' T tract (Curotto de Lafaille et al., 1992; Dr G. Alonso unpublished results). For the remaining nuclear genes the definition of the flanking regions was in some extent arbitrary, although the presence of the mentioned tracts was checked. Finally, only sequences published in journals subjected to peer review were included.

The GC composition of mitochondrial genomes was incorporated for comparative purposes.

The analysis of the data was done with a program SigmaPlot^{R} version 4.1 (Jandel Scientific, 1991).

Supported by a grant from the Commission of the European Communities Contract No. TS2-63-V.
+Corresponding author.
Received 30 April 1992.
Accepted 10 August 1992.

TABLE I

GenBank sequences analyzed

Organism	Gen	Gen Bank No.
L. donovani	CATP	M17889-J04004
L. sp	DHFR-TS	M12734-X51735-M14330
L. major	Membrane antigen	X06555
L. sp	GP-63	Y00647-M60048
L. major	P100	J04483
L. enriettii	Transport protein	M26229-M38214
L. mexicana	B-tubulin	M23441
L. major	HSP 70	X14574-X14575
L. donovani	IMPDH	M35667
L. amazonensis	Membrane glycoprotein	M38368
L. enriettii	Histone H2B	M38215
L. donovani	ODCase	M81192
L. enriettii	Tubulin intergenic regions	M15506-M15507
L. major	Mitochondrial NADH-DH	X13751
L. tarentolae	Mitochondrial rRNA	X02354
L. tarentolae	Maxicircle	M10126
L. tarentolae	Minicircle	K01980
C. fasciculata	FGK	X07458-X07459
C. fasciculata	DHFR-TS	M22852
C. fasciculata	Pol II	X13489
C. fasciculata	Mitochondrial Cytochrome ox.	X05063
C. fasciculata	Minicircle	M19266
C. fasciculata	Maxicircle	X15081
C. fasciculata	rRNA	X03450
C. fasciculata	LSU rRNA	Y00055
T. cruzi	Antigen	M21582
T. cruzi	Surface antigen	X04186
T. cruzi	CRA	J04016
T. cruzi	FRA	J04015
T. cruzi	HSP 70	X07083-M26595-X13690
T. cruzi	85 kDa antigen (HSP)	M15346
T. cruzi	1F8	X02838
T. cruzi	Ubiquitin	J03945
T. cruzi	Cysteine protease	M27305
T. cruzi	KAP	M25364
T. cruzi	85 kDa surface antigen	J04667
T. cruzi	FUS-UBQ	X07451
T. cruzi	Neuraminidase like	M64836
T. cruzi	P Protein	X52323
T. cruzi	Trypanothione reductase	M38051
T. cruzi	Calmodulin-Ubiquitin	X52096
T. cruzi	SAPA	X57235
T. cruzi	Minicircle	X04680-M15511
T. cruzi	Mitochondrial rRNA	M19982-M19981
T. brucei	Actin	M20310
T. brucei	Aldolase	X03061
T. brucei	Calmodulin	K02944
T. brucei	HSP	M14697-X14176
T. brucei	ODCase	J02771
T. brucei	PEP-CK	M20570
T. brucei	PGK	X03370
T. brucei	TIM	X03921
T. brucei	Tubulin	K02836
T. brucei	Trypanothione reductase	M21122
T. brucei	TOP 2	M26803
T. brucei	GAPDH	M26816
T. brucei	Phospholipase C	X13292-J04124
T. brucei	Pol I	M27164-X14399-J05074
T. brucei	Pol II	J04841-J03157-X13491
T. brucei	Pol III	X12494-M27101
T. brucei	BS2	J02865
T. brucei	X92	X14820
T. brucei	PFR	X14819
T. brucei	HSP 70	M14697-M32139-M32140
T. brucei	PGI	X15540
T. brucei	PyK	X57950-X57951
T. brucei	Cytochrome oxidase	M14820
T. brucei	Cytochrome C	M23360
T. brucei	Mitochondrial rRNA	X02547
T. brucei	Apocytochrome b	M17998
T. brucei	Maxicircle	M18979-M14429
T. brucei	Minicircle	M14763
T. brucei	PARP	M17027
T. brucei	ESAG's	M20871
T. brucei	VSG's	J01232-V01387-J01221-J01224-X01843

RESULTS AND DISCUSSION

The total GC content and the GC distribution for the organisms here studied is presented in Table II. *C. fasciculata, Leishmania,* and *T. cruzi* possess GC-rich genomes when compared with vertebrates (40-45%) (Sueoka, 1964) (Table II). In coincidence with other organisms, Trypanosomatidae genes have a lower GC content in their flanking regions. However, contrary to most organisms, the 3' flanking region has a higher GC content than the 5' flanking region. The higher GC level in the 5' flanking regions of the majority of genes studied has been related (Li & Graur, 1991) to a tendency of the promoter and its surrounding regions to be GC-rich. In agreement with this speculation, the trypanosomatidae genes lacking orthodox promoters in their 5' flanking regions (Borst, 1986; Clayton, 1988) escape from this tendency.

Two general hypotheses that try to explain variations in GC content among organisms can be considered in the context of Trypanosomatidae genome. First, the selectionist hypothesis (Bernardi & Bernardi, 1986) regards a high GC content as an adaptation to environmental conditions. For example, thermophilic bacteria have a strong preference for thermally stable GC rich codons (coding alanine and arginine) and tend to avoid using thermally unstable GC poor codons (coding serine and lysine) (Argos et al., 1979). It is therefore tempting to speculate that the *Leishmania* GC-rich coding regions were selected as an adaptation to the environment of a warm blooded host. However, this explanation is not consistent with the fact that *Crithidia,* which is a parasite of cold blooded insects, has the highest GC content, whereas the vertebrate parasite *T. brucei* has the lowest. More difficult to reconcile is the coexistence of AT-rich mitochondrial genes (Table II) in these organisms. Any other selectionist explanation such as invoking UV light as the selective force will face this same contradiction (Singer & Ames, 1970).

A second hypothesis that we favor is the mutationist (Sueoka, 1964; Muto & Osawa, 1987), which invokes biased mutation patterns to explain variations in GC content. According to this view, the GC content of a given species is determined by the balance between the rate of substitution from G or C to T or A (u), and the rate of substitution in opposite direction (v). In the end, the pattern of substitution is determined by the pattern of mutation and the pattern of purifying selection against certain mutations. Since the number of published gene sequences in Trypanosomatidae is relatively small, it is difficult to make accurated calcu-

TABLE II

GC content (%) in different regions of Trypanosomatidae genome

Organism	Total	Coding	5' Flanking	3' Flanking	Codon position			Mitochondrial
					1st	2th	3th	
Crithidia	58	64.9	44.3	50.5	62.7	43.1	88.9	31.8
Leishmania	57	62.9	55.5	58.3	59.5	44.6	84.6	26.0
T. cruzi	51	56.9	41.6	41.0	59.6	43.7	67.3	30.6
T. brucei	44	51.6	40.9	43.6	57.0	40.8	56.9	23.1

TABLE III

Base substitution in Trypanosomatidae

Homologous genes	Organism	GC/AT (u)	AT/GC (v)	GC mutational pressure (u/v)
DHFR-TS	*Crith. – Leish.*	85	86	0.98
PGK	*Crith. – T. brucei*	240	52	4.6
B-Tubulina	*Leish. – T. brucei*	112	19	5.9
HSP (Fragment)	*Leish. – T. brucei*	42	15	2.8
Calmodulin	*T. cruzi – T. brucei*	52	7	7.43

TABLE IV

Codon usage in Trypanosomatidae

AA	Codons	*Crith.*	*Leish.*	*T. cruzi*	*T. brucei* 1	*T. brucei* 2	AA	Codons	*Crith.*	*Leish.*	*T. cruzi*	*T. brucei* 1	*T. brucei* 2
Phe	TTT	8.3	15.0	52.0	45.3	47.5	Tyr	TAT	5.7	7.8	14.8	35.1	43.2
	TTC	91.6	85.0	48.0	54.7	52.4		TAC	94.2	92.2	85.0	64.9	56.7
Leu	TTA	0.6	0.7	2.3	5.3	11.2	***	TAA	25.0	16.6	12.5	46.2	100.0
	TTG	3.8	4.8	10.8	15.3	13.0		TAG	50.0	66.6	43.7	30.8	0.0
	CTT	5.7	12.1	22.3	25.5	14.4		TGA	25.0	16.6	43.7	23.0	0.0
	CTC	22.3	18.5	16.6	23.6	13.3							
	CTA	1.9	2.5	4.0	7.8	28.1	His	CAT	3.3	8.9	15.0	36.2	48.3
	CTG	65.6	61.3	44.0	22.3	19.8		CAC	96.6	91.1	85.0	63.8	5.7
Ile	ATT	21.1	21.4	44.2	44.1	40.6	Gln	CAA	6.6	4.5	18.0	37.3	57.2
	ATC	78.8	76.9	47.7	42.5	22.6		CAG	93.3	95.4	82.0	62.6	42.8
	ATA	0.0	1.6	8.0	13.3	36.7							
							Asn	AAT	1.8	7.4	37.0	38.2	42.8
Met	ATG	100.0	100.0	100.0	100.0	100.0		AAC	98.2	92.6	63.0	61.7	57.2
Val	GTT	6.4	9.2	31.5	29.5	21.4	Lys	AAA	1.7	7.2	20.0	31.7	67.9
	GTC	16.5	20.0	16.6	16.9	18.1		AAG	98.3	92.8	80.0	68.2	32.1
	GTA	0.7	4.2	3.7	13.1	31.1							
	GTG	76.3	66.6	48.2	40.3	29.2	Asp	GAT	10.7	20.7	27.0	48.4	36.8
								GAC	89.3	79.3	73.0	51.6	63.2
Ser	TCT	4.8	8.3	9.8	17.1	10.9							
	TCC	19.4	18.9	16.6	21.8	8.9	Glu	GAA	3.7	7.8	34.0	39.7	68.9
	TCA	2.9	5.2	8.7	14.4	16.9		GAG	96.3	92.2	66.0	60.2	31.1
	TCG	35.9	33.4	18.8	13.0	8.4							
	AGT	1.9	5.4	19.0	15.0	20.4	Cys	TGT	3.7	13.0	26.6	40.8	24.3
	AGC	34.9	28.7	27.1	18.6	34.3		TGC	96.3	87.0	73.3	59.1	75.7
Pro	CCT	6.2	15.5	10.5	21.5	34.8	Trp	TGG	100.0	100.0	100.0	100.0	100.0
	CCC	13.4	15.8	41.8	30.5	16.5							
	CCA	9.3	9.4	14.2	27.2	30.4	Arg	CGT	16.0	13.6	17.7	26.3	9.1
	CCG	71.1	59.2	33.5	20.7	18.2		CGC	58.0	58.2	43.0	33.0	11.3
								CGA	4.9	4.0	8.7	11.5	18.2
Thr	ACT	7.4	7.6	18.8	20.0	15.8		CGG	14.8	15.5	13.5	13.4	7.9
	ACC	23.4	19.6	17.5	25.0	18.0		AGA	2.5	1.2	5.4	5.7	39.8
	ACA	3.7	5.1	12.9	25.5	44.8		AGG	3.7	7.4	11.6	10.1	13.7
	ACG	65.4	67.6	50.8	29.5	21.3							
							Gly	GGT	16.3	15.6	26.8	42.2	14.7
Ala	GCT	15.8	15.5	25.4	27.5	17.2		GGC	65.2	66.2	43.5	27.8	34.8
	GCC	26.7	25.2	28.8	26.1	22.4		GGA	0.7	3.5	14.5	17.0	34.1
	GCA	7.1	8.6	21.3	24.5	40.3		GGG	17.7	14.6	15.2	12.9	16.4
	GCG	50.3	50.7	24.4	21.8	20.0							

Crithidia: 4 genes (1679 codons).
Leishmania: 16 genes (8216 codons).
T. cruzi: 17 genes (6193 codons).
T. brucei: 1: 26 genes, protein coding (15540 codons).
2: 8 genes, VSG, PARP, ESAG'S coding (3263 codons).

lations of their GC mutational pressures (u/v). Nonetheless, in an effort to obtain an insight into the GC mutational pressure in these organisms, we calculated an approximated u/v ratio for some homologous genes (Table III).

The u/v values between any other Trypanosomatidae and *T. brucei* are larger than 1, revealing a predominance of GC to AT (u) substitutions. However, the rate of substitution varies from one gene to the other, for example, the phosphoglycerate kinase gene of *Crithidia* and *T. brucei* presented 240 GC to AT replacements, whereas only 52 changes occurred from AT to GC (u/v = 4.6). Two genes of *Leishmania,* namely β-tubulin and a fragment of HSP70, when compared with homologous genes in *T. brucei,* yielded u/v values of 5.9 and 2.8, respectively. The calmodulin gene of *T. cruzi* and *T. brucei* showed the biggest u/v ratio (7.4) (Table III), even when the difference in their GC contents is not the largest.

The GC mutational pressure for the DHFR-TS gene of *Crithidia* and *Leishmania* is small (u/v = 0.98).

These data indicate that *T. brucei* evolved towards an AT-rich genome, followed by *T. cruzi.* In *Leishmania* and *Crithidia,* on the other

hand, there seems to be a constraint to evolve towards a more balanced GC content in their genomes.

Next, we examined how purifying selection affects the use of bases at the codon level, plotting the total GC content vs the presence of G or C at the three codon positions. Figure 1A shows, as expected, that the less-constrained wobble position follows the changes in total GC as though little selection is operating. There is a more stringent selection on first and second positions, being stronger in the second position. Similar findings were reported for other organisms (Bernardi & Bernardi, 1986; Muto & Osawa, 1987).

The overall GC content at the flanking regions should follow the general pattern of GC substitution of the third coding position, providing that little or no selection is taking place. However, a plot of these parameters (Fig. 1B), reveals a departure from the theoretical slope, therefore the selection exerted over the flanking regions seems to be larger than in the third codon position.

An analysis of codon usage for the organisms under study, reveals several important facts (Table IV): out of the eight triplets where any base can occupy the third position, namely, leu, val, ser, pro, thr, ala, arg and gly, *Crithidia* and *Leishmania* preferred G in six aminoacids, *T. cruzi* in five, and *T. brucei* in two (see Table V). Since the aminoacids arg and gly, have G in the second position, it appears that *Crithidia* and *Leishmania* preferred the weaker NGC interaction over the strong NGG, the only exception being the unique non-ambiguous codon for tryptophan TGG.

A recent report by Langford et al. (1992) (published while this work was under review) supports the *Leishmania* codon usage table presented in this work.

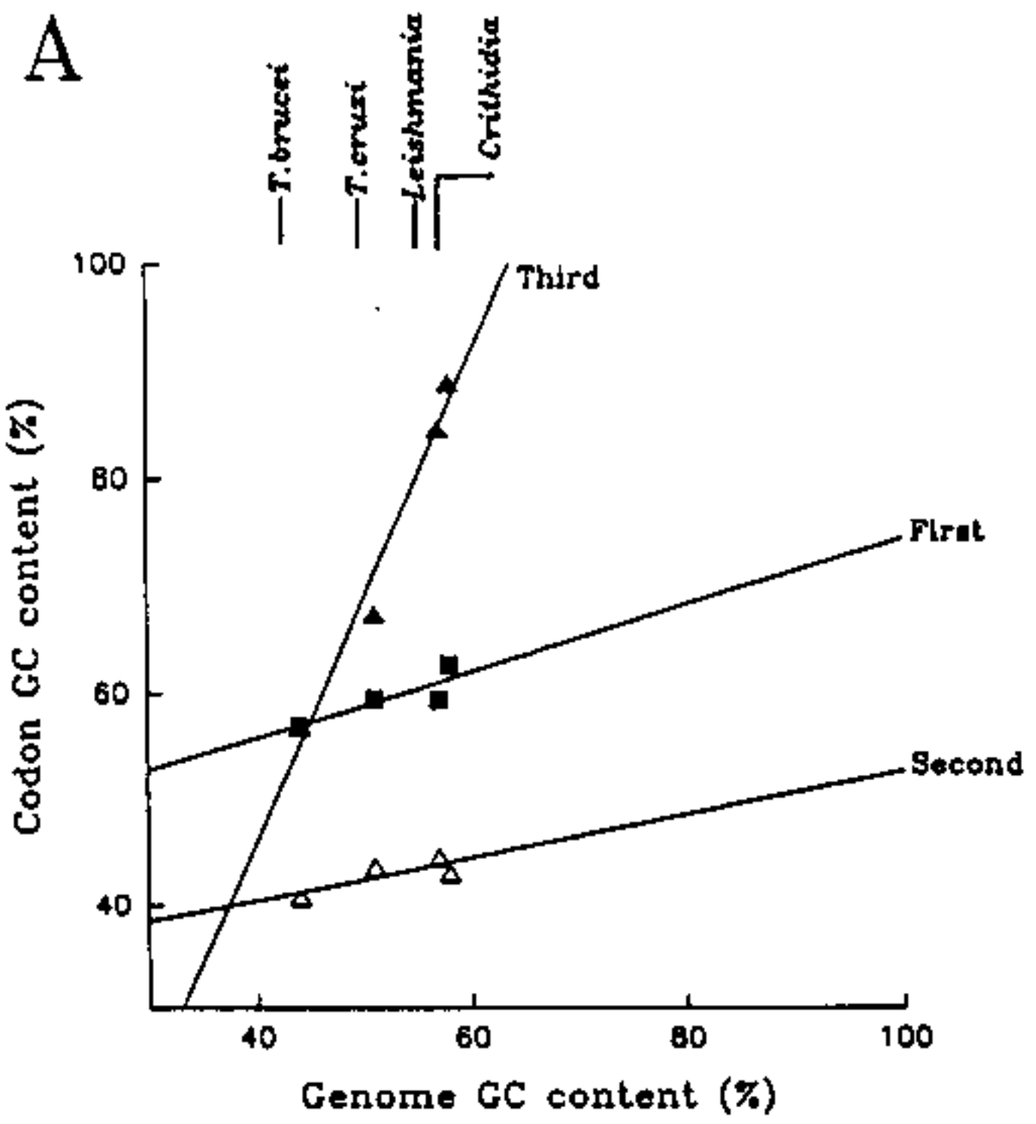

Fig. 1: Trypanosomatidae total GC content and its relationship with the codon position and GC content in gene flanking regions. A: correlation of the GC content (%) between total genomic DNA and the first, second and third codon positions. B: GC content (%) at the third codon position (%) vs GC content (%) in the 5' and 3' gene flanking regions.

A close examination of the codon usage patterns presented in Table IV reveals, in concordance with the genome hypothesis and previous phylogenetical studies (Grantham et al., 1980; Gómez et al., 1991), that *Crithidia* and *Leishmania* share similar codon usage. *T. cruzi* has some common patterns with *Crithidia* and *Leishmania,* but in general, it is more closely related to *T. brucei.* A comparison with the codon usage in other parasitic protozoa such as *Entamoeba* (Tannich & Horstmann, 1992) and *Plasmodium* (Saul & Battistutta, 1988), confirms the correlation between taxonomic status and codon preferences.

In our study, we noticed that some genes of *T. brucei,* such as variant surface glycoproteins (VSGs), procyclic acidic repetitive protein (PARP) and expression-site associated genes (ESAGs), had a strong deviation towards AT in their codon usage, and in their flanking

TABLE V

Frequency of nucleotides in the three codon positions

Organism	1st	2th	3th	Preferred consensus codon
Crithidia				
A	24.6	29.3	3.0	
C	23.7	25.4	39.5	GHG
G	39.0	17.7	49.5	
T	12.7	27.6	8.0	
Leishmania				
A	24.4	26.6	4.2	
C	23.2	25.2	40.1	GHG
G	36.3	19.5	44.6	
T	16.1	28.7	11.1	
T. cruzi				
A	28.1	34.4	13.1	
C	20.1	26.7	29.5	GHG
G	39.6	16.9	37.8	
T	12.2	22.0	19.6	
T. brucei (1)				
A	28.3	31.5	19.8	
C	21.0	23.2	28.3	RAB
G	35.5	18.0	26.7	
T	15.2	27.3	25.2	
T. brucei (2)				
A	32.9	35.9	35.3	
C	17.0	26.4	24.2	RAA
G	36.8	16.3	21.3	
T	13.2	21.4	19.1	

H: A-C-T.
R: G-A.
B: C-G-T.
T. brucei (1): protein coding genes.
T. brucei (2): VSG, PARP, ESAG'S coding genes.

regions, but contrary to previous reports (Michels, 1986), A is preferred over T in most codons, and the only termination triplet so far reported is TAA. This difference can be explained by the larger number of gene sequences here analyzed. These genes are probably transcribed by RNA polymerase I (Brown et al., 1992), and are expressed at high levels in the parasite, but since the tRNA composition in this organism is not known, the significance of this observation can not be ascertained. However, Parsons et al. (1991) found that the "housekeeping" genes with high levels of expression, and the highly expressed VSG genes do not coincide in their codon preference, thus demonstrating that in *T. brucei* factors other than aminoacyl-tRNA abundance must be involved in codon choice.

Table V summarizes the frequency of each nucleotide at the the three codon positions, and the consensus sequence for the preferred codon usage.

We can speculate that as a consequence of GC mutational pressure and purifying selection, the nuclear genes of *Leishmania, Crithidia,* and to a less extent *T. cruzi,* have retained features of a primeval genetic system where no ribosomal structures were present (Anderson & Kurland, 1990), and more stable bonds were necessary to give the primitive functions of messengers and adapters. Perhaps the primitive tRNAs of these organisms had GC-rich anticodon regions that determined the bias for synonymous codons ending in G or C. Considering that the most abundant aminoacids in prebiotic times were glycine, alanine, valine, glutamic and proline (Miller, 1986), and assuming that G or C were more frequently used, some authors have proposed a primitive genetic apparatus prefering GNN codons (N = G, C, T, A or U) (Anderson & Kurland, 1990). Then, the trypanosomatidae common ancestor probably favored the stronger GNS or GHG codon (Table V) (H = A, C, T and S = G or C).

ACKNOWLEDGEMENT

To Mr Ian McClure for reviewing the manuscript.

REFERENCES

ANDERSSON, S. G. E. & KURLAND, C. G., 1990. Codon preferences in free-living microorganisms. *Microbiol. Rev., 54:* 198-210.

ARGOS, P.; ROSSMANN, M. G.; ULRICH, M. G.; ZUBER, H.; FRANK, G. & TRASTCHIN, J. D., 1979. Thermal stability and protein structure. *Biochemistry, 18:* 5698-5703.

BERMAN, J. D., 1988. Chemotherapy for leishmaniasis: biochemical mechanisms, clinical efficacy, and future strategies. *Rev. Infec. Diseases, 10:* 560-586.

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

BORST, P., 1986. Discontinuous transcription and antigenic variation in trypanosomas. *Ann. Rev. Biochem., 55:* 701-732.

BROWN, S. D.; HUANG, J. & VAN DER PLOEG, L. H. T., 1992. The promoter for the procyclic acidic repetitive protein (PARP) genes of *Trypanosoma brucei* shares features with RNA polymerase I promoters. *Mol. Cell. Biol., 12:* 2644-2652.

CLAYTON, C., 1988. The molecular biology of the Kinetoplastidae, p. 1-55. In *Genetic Engineering,* Vol 7, AP.

CUROTTO DE LAFAILLE, M. A.; LABAN, A. & WIRTH, D. F., 1992. Gene expression in *Leishmania:* analysis of essential 5' DNA sequences. *Proc. Natl Acad. Sci. U. S. A., 89:* 2703-2727.

GOMEZ, E.; VALDES, A. M.; PIÑERO, D. & HERNANDEZ, R., 1991. What is a genus in the trypanosomatidae family?. Phylogenetic analysis of two small rRNA sequences. *Mol. Biol. Evol., 8:* 254-259.

GRANTHAM, R.; GAUTIER, C.; GOUY, M.; MERCIER, R. & PAVE, A., 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res., 8:* r49-r62.

LANGFORD, C. K.; ULLMAN, B. & LANFEAR, S. M., 1992. *Leishmania:* Codon utilization of nuclear genes. *Exp. Parasitol., 74:* 360-361.

LI, W-H & GRAUR, D., 1991. *Fundamentals of molecular evolution.* p. 219-231. Sinauer Ass., Inc. Publishers, Sunderland, Massachusetts.

MICHELS, P. A. M., 1986. Evolutionary aspects of trypanosomes: analysis of genes, *J. Mol. Evol., 24:* 45-52.

MILLER, S. L., 1986. Current status of the prebiotic synthesis of small molecules. *Chem. Scr., 26B:* 5-11.

MUTO, A. & OSAWA, S., 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl Acad. Sci. U. S. A., 84:* 166-169.

OPPERDOES, F. R., 1983. Glycolysis as target for the development of new trypanocidal drugs. p. 121-131. In Singer, Mansour & Undarza (eds). *Mechanism of drug action.* Academic Press, Orlando.

PARSONS, M.; STUART, K. & SMILEY, B., 1991. *Trypanosoma brucei:* Analysis of codon usage and nucleotide composition of nuclear genes. *Exp. Parasitol., 73:* 101-105.

SAUL, A. & BATTISTUTTA, D., 1988. Codon usage in *Plasmodium falciparum. Mol. Biochem. Parasitol., 27:* 35-42.

SINGER, C. E. & AMES, B. N., 1970. Sunlight ultraviolet and bacterial DNA base ratios. *Science, 170:* 822-826.

SUEOKA, N., 1964. On the evolution of informational macromolecules. p. 479-496. In V. Bryson & H. G. Vogel (eds). *Evolving Genes and Proteins.* Academic Press, New York.

TANNICH, E. & HORSTMANN, R. D., 1992. Codon usage in pathogenic *Entamoeba histolytica. J. Mol. Evol., 34:* 272-273.

TIBAYRENC, M.; KJELLBERG, M. & AYALA, F., 1990. A clonal theory of parasitic protozoa: the population structure of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma. Proc. Natl Acad. Sci. U. S. A., 87:* 2414-2418.

VICKERMAN, K., 1976. The diversity of kinetoplastid flagellates. p. 1-30. In W. H. R. Lumsden & D. A. Evans (eds). *The Biology of the Kinetoplastida.* Academic Press, London, New York, San Francisco.