

TENDENCIES IN *ECHINOCOCCUS* SP. CODON USAGE

FERNANDO ALVAREZ***; BEATRIZ GARAT*; HECTOR MUSTO*/**; MARINA PICON* &
RICARDO EHRLICH*/†

Departamento de Genética* Departamento de Bioquímica** Unidad de Biología Parasitaria,
Facultad de Ciencias, Tristán Narvaja 1674, 11200 Montevideo, Uruguay ***Departamento de Genética,
Facultad de Medicina, Gal. Flores 2125

The analysis of DNA coding sequences of a wide variety of organisms has revealed significant bias in codon usage (K. Wada et al., 1991, *Nucl. Acids Res.*, 19: 1981-1986). Although important variations in the use of alternative synonymous triplets have been found among genes of a given organism (P. M. Sharp et al., 1988, *Nucl. Acids Res.*, 16: 8207-8211), the average codon usage appears to be species-specific. The study of the bias in codon usage is of interest in designing oligonucleotide probes from amino acid sequence data, in translation improvement when expressing a gene in heterologous systems, and in evolutionary analysis.

This report presents the tendencies in codon usage in *Echinococcus* sp. The coding sequences analyzed totaled 2310 codons. The sequences were obtained from already published work (M. V. Lightowers et al., *Mol. Biochem. Parasitol.*, 36: 287-290; H. B. Ferreira & A. Zaha, 1990, in: Basic Research in Helminthiasis, R. Ehrlich, A. Nieto & L. Yarzabal, (eds.), p. 87-104, Logos, Montevideo; J. C. Shepherd, et al., 1991, *Mol. Biochem. Parasitol.*, 44: 81-90; B. Facon, et al., 1991, *Mol. Biochem. Parasitol.*, 45: 233-240). *E. granulosus* homeobox-containing genes, EgHox 1-4, G. Oliver et al. (G. Oliver, et al., 1992, *Gene*, 121: 337-342) Df 1, an *E. granulosus* stage-specific expressed gene, A. Esteves et al. (A. Esteves, et al., 1993, *Mol. Biochem. Parasitol.*, 58: 215-222) and from personal communications: an *E. granulosus* cDNA clone coding for a highly immunogenic protein, A. Zaha (Porto Alegre). Results are summarized in Table I.

We have performed a variation analysis of the data (one-way analysis of variance) in order to discard possible bias in codon usage characteristic of particular groups of genes. The differences between means that appear as significant with this procedure correspond then, to bias in the entire population examined. It is interesting to note that variation analysis exclude some cases presenting apparent significant differences in codon usage (see, for instance, glutamine triplets in Table I). In these cases, the absence of significant differences could be attributed to the differential use of triplets among genes.

Our calculations revealed significant differences in codon usage for glycine, valine, arginine, lysine, isoleucine, leucine and phenylalanine at 1% level, and for cysteine, tyrosine and alanine at 5% level.

Finally, it is interesting to compare the bias in codon usage in *Echinococcus* sp. with that displayed by *Schistosoma mansoni*, the Platyhelminth most largely studied with molecular approaches (H. M. Meadows & A. J. G. Simpson, 1989, *Mol. Biochem. Parasitol.*, 36: 291-293). The comparison reveals that, from the 18 amino acids encoded by more than one codon, 17 display significant differences at 1% level, whereas proline is significant at 5% (Contingency table). Furthermore, while *Echinococcus* sp. exhibits a significant G+C bias in the third codon position, *S. mansoni* displays an A+T bias at the same location (see Table II and K. Wada et al.).

Acknowledgements: to C. Scazzocchio and U. Pettersson for constant encouragement and helpful discussions.

This work was supported by grants of the E.E.C. (C.I. 1.0284-U) and SAREC.

†Corresponding author.

Received 3 November 1992.

Accepted 25 March 1993.

TABLE I
Echinococcus sp. codon frequency table

AA	Codon	Freq.	AA	Codon	Freq.	AA	Codon	Freq.	AA	Codon	Freq.
<i>Lys</i>	AAA	244.50	<i>Glu</i>	GAA	403.42	<i>Gln</i>	CAA	195.59	*	TAA	8.15
	AAG	611.25		GAG	513.25		CAG	289.32	*	TAG	0.00
<i>Asn</i>	AAC	171.15	<i>Asp</i>	GAC	277.10	<i>His</i>	CAC	114.10	<i>Tyr</i>	TAC	191.52
	AAT	211.90		GAT	215.92		CAT	93.72		TAT	89.25
<i>Thr</i>	ACA	118.17	<i>Ala</i>	GCA	134.44	<i>Pro</i>	CCA	81.50	<i>Ser</i>	TCA	73.35
	ACG	77.43		GCG	122.25		CCG	57.05		TCG	114.01
	ACC	191.52		GCC	207.82		CCC	73.35		TCC	134.47
	ACT	187.50		GCT	232.27		CCT	154.84		TCT	122.24
<i>Arg</i>	AGA	85.58	<i>Gly</i>	GGA	52.98	<i>Arg</i>	CGA	130.40	*	TGA	0.00
	AGG	65.20		GGG	32.60		CGG	73.35	<i>Trp</i>	TGG	85.57
<i>Ser</i>	AGC	105.95		GGC	150.78		CGC	187.45	<i>Cys</i>	TGC	105.95
	AGT	110.02		GGT	224.12		CGT	191.52		TGT	52.97
<i>Met</i>	ATG	338.22	<i>Val</i>	GTA	65.20	<i>Leu</i>	CTA	65.20	<i>Leu</i>	TTA	44.82
<i>lle</i>	ATA	57.05		GTG	240.42		CTG	228.12		TTG	154.85
	ATC	207.82		GTC	150.78		CTC	224.12	<i>Phe</i>	TTC	273.02
	ATT	211.90		GTT	142.62		CTT	142.62		TTT	122.25

Frequencies are per 10000 codons. Amino acids for which significant bias in triplet usage were found, are underlined.

TABLE II
Percentages of bases per codon position

	<i>Echinococcus</i> sp.					<i>Schistosoma mansoni</i>				
	A	C	G	T	G+C	A	C	G	T	G+C
I	29.95	23.02	31.66	15.72	54.68	29.36	33.75	16.21	20.70	49.96
II	36.30	20.82	16.54	26.69	37.36	35.00	21.82	17.96	25.24	39.78
III	17.60	27.67	30.03	25.06	57.70	31.65	12.68	15.44	40.25	28.12

Frequencies for *S. mansoni* were calculated from data reported by Wada K. et al.