Parasitol Res (1995) 81:388–393

ORIGINAL PAPER

John Ellis · David A. Morrison · Bernd Kalinna

Comparison of the patterns of codon usage and bias between *Brugia, Echinococcus, Onchocerca* and *Schistosoma* species

Received: 4 October 1994 / Accepted: 5 January 1995

Abstract Patterns of codon usage and bias were compared among taxa of the genera *Brugia*, *Echinococcus*, *Onchocerca* and *Schistosoma* by metric multidimensional scaling and three commonly used indices of bias: N_c, GC_{3S} and B. The overall codon usage for each taxon was compared, as was the codon usage for each individual gene within the taxa. Differences in the patterns of codon usage observed between taxa were dependent on the overall base composition of the genes analysed. The codon usage of *Echinococcus* was distinct from that of the other taxa. Furthermore, the pattern of codon usage detected by the average codon usage summed across all genes for each taxon was not shown by all genes from that taxon.

Introduction

Grantham et al. (1981) proposed the genome hypothesis that explained the patterns observed in codon usage as being non-random and species-specific. This hypothesis was subsequently modified to account for observations that codon usage is related specifically to the genome base composition of an organism (D'Onofrio et al. 1991). Furthermore, since genome compartmentalisation is known to occur in eukaryotes, multiple patterns of codon usage are known to occur that reflect the GC content of the different genome compartments. Therefore, in higher eukaryotes the patterns of codon usage shown by genes of an organism often differ. Bias in codon usage may result from not only the influence of the amino acid composition of the gene product but also the action of directional mutation pressure on the base composition of the genome (Andersson and Kurland 1990; Osawa et al. 1992). Although a variety of factors affect the patterns of codon usage shown by an organism, taxonomically related species generally have very similar patterns of codon usage, whereas unrelated species have quite different patterns of codon usage (Ikemura 1985; Long and Gillespie 1991; Ellis et al. 1993).

Recently, Unnasch et al. (1992) reported an analysis of codon usage in genes of *Onchocerca volvulus*, Alvarez et al. (1993) and Kalinna and McManus (1994) reported codon usage for *Echinococcus* species and Hammond (1994) described codon usage in *Brugia*. However, all these analyses were limited to a summary of the total codon usage summed over all the genes of that taxon and did not take into account any heterogeneity within the data set (Sharp and Devine 1989; Lloyd and Sharp 1992a). Since it is now well documented that the extent to which alternative synonymous codons are used by an organism is non-random and that the pattern of codon usage varies not only between species but also between genes of the same species, it is obvious that the approach used in all of these studies will not completely or adequately explain the patterns of codon usage observed in genes of *Brugia*, *Echinococcus* and *Onchocerca*. Consequently, we report herein a thorough comparison of the patterns of codon usage and bias in gene sequences of *Brugia*, *Echinococcus* and *O. volvulus* and compare them with the results reported previously for *Schistosoma mansoni* (Ellis and Morrison 1994).

Materials and methods

The data set (shown in Tables 1 and 2) contains 8 genes derived from 2 species of *Brugia* (*B. malayi* and *B. pahangi*; 4688 codons), 9 genes of *Onchocerca volvulus* (4499 codons) and 10 genes from *Echinococcus* spp. (*E. granulosus* and *E. multilocularis*; 4491 codons). The data set for *Schistosoma mansoni*, which contains 20 genes (11287 codons), has previously been described (Ellis and Morrison 1994); 3 genes of *S. japonicum* (1342 codons) were also included for comparative purposes only. Gene sequences were extracted from the GenBank data base using the accession

J. Ellis (✉) ·D. A. Morrison
Faculty of Science,
University of Technology Sydney, Westbourne St.,
Gore Hill, New South Wales 2065, Australia

B. Kalinna
Molecular Parasitology Unit, Queensland Institute
for Medical Research, Herston Rd.,
Brisbane, Queensland 4006, Australia

Table 1 Codon usage bias in genes of *Onchocerca* and *Brugia* (*L* Length in base pairs; N_c effective number of codons, GC_{3S} G+C content at silent third positions, *GC* gene G+C content as a fraction, *B* standardised measure of bias, *Code* code used in the figures)

Gene[a]	Accession number[b]	L	N_c	GC_{3S}	GC	B	Code	Reference
ONGACT1A	M84916	1131	48.88	0.39	0.45	0.134	21	Zeng and Donelson 1992
ONGCUZNSOD	X57105	477	52.96	0.27	0.44	0.293	22	Henkle et al. 1991
ONGEF1A	M64333	1395	47.59	0.29	0.41	0.180	23	Alarcon and Donelson 1991
ONGMBWMZ	M74066	5874	51.17	0.34	0.41	0.110	24	Werner and Rajan 1992a
ONGRASB	M98812	648	61.00	0.33	0.43	0.159	25	Dissanayake et al. 1992
ONGMSPB	J04663	384	48.43	0.50	0.50	0.255	26	Scott et al. 1989
ONGPARAMYO	M95813	2640	44.28	0.21	0.39	0.269	27	Dahmen et al. 1993
ONGOV7	M37105	489	59.88	0.32	0.39	0.230	28	Lustigman et al. 1992
ONGANTIG	M27807	459	56.88	0.37	0.41	0.144	29	Lobos et al. 1990
BRPCHIT	M73689	1515	52.84	0.30	0.41	0.117	30	Fuhrman et al. 1992
BRPMYOHEAV	M74000	5874	51.47	0.35	0.42	0.097	31	Werner and Rajan 1992b
S77613	S77613	1935	50.00	0.35	0.42	0.118	32	Rothstein and Rajan 1991
BRPANTP	J03971	1647	53.51	0.32	0.41	0.110	33	Perrine et al. 1988
BRPMF22G	X58063	618	50.92	0.20	0.47	0.396	34	Selkirk et al. 1991
BRPRPS13A	M86643	456	51.12	0.39	0.46	0.157	35	Ellenberger et al. 1989
BRPGP29	X63365	672	42.96	0.28	0.37	0.285	36	Cookson et al. 1992
BRPTUBBA	M36380	1347	50.98	0.35	0.45	0.116	37	Guenette et al. 1991

[a] Data-base name
[b] GenBank data-base accession number

Table 2 Codon usage bias in genes of *Echinococcus* and *Schistosoma japonicum*. Definitions of table headings are given in Table 1

Gene[a]	Accession number[b]	L	N_c	GC_{3S}	GC	B	Code	Reference
ECCACTNI	L07773	1128	49.41	0.57	0.52	0.141	38	Da Silva et al. 1993
ECCMDH	L08894	999	51.19	0.62	0.53	0.167	39	Rodrigues et al. 1993
ECCGRPA	M63605	1956	52.16	0.46	0.46	0.098	40	GenBank
ECCEG13A	M95051	1251	58.82	0.56	0.51	0.083	41	GenBank
EGDF1P	X65947	405	57.69	0.64	0.49	0.332	42	Esteves et al. 1993
EGPARAMA	Z21787	2592	57.41	0.59	0.53	0.088	43	Mühlschlegel et al. 1993
ECCUB19A	L23315	231	51.21	0.59	0.50	0.419	44	GenBank
ECCTEGPRO	M61186	1680	55.29	0.57	0.48	0.087	45	Frosch et al. 1991
ECCGRP	M63604	1950	53.40	0.47	0.46	0.083	46	GenBank
ECCANTIGEN	M84807	1281	59.67	0.54	0.50	0.071	47	Hemmings and McManus 1991
SCMENOL	L23324	1305	52.14	0.33	0.42	0.104	48	Waine et al. 1993a
SCMGL3PHDE	L09549	1017	58.16	0.38	0.43	0.104	49	Waine et al. 1993b
SCMFABP	L23322	399	59.50	0.43	0.43	0.337	50	GenBank

[a] Data-base name
[b] GenBank data-base accession number

numbers shown in Tables 1 and 2. Partial gene sequences were excluded from this analysis as they yield anomalous values for measures of bias. Codon-usage tables were compiled using the program CODON FREQUENCY in the GCG computer software package (Devereux et al. 1984) run through the Australian National Genomic Information Service.

The procedures used for these analyses of codon usage and bias have been described in detail elsewhere. Metric multidimensional scaling (MMDS) was performed using the PATN multivariate pattern-analysis program (Belbin 1989) as described by Morrison et al. (1994). This involves deriving a standardised measure of inter-gene divergence in codon usage and displaying these distances as either a dendrogram or an ordination. A two-dimensional ordination was used with the Manhattan distance measure (Faith et al. 1987). Log-likelihood ratio chi-square goodness-of-fit tests were used to identify codons that were over- or under-represented in the data set.

Three independent measures of codon usage bias (N_c, GC_{3S} and B) were determined for genes in the data set. A computer program, CODONS (Lloyd and Sharp 1992b), was used to calculate two of the indicators of codon usage bias for the nucleotide sequences shown in Tables 1 and 2, namely N_c (Wright 1991) and GC_{3S} (Sharp and Devine 1989). N_c (commonly referred to as the "effective number" of codons used by a gene) is a general measure of non-uniformity of codon usage and can take values between 20 (for genes that are highly biased and use only 1 codon for each amino acid) and 61 (for unbiased genes), whereas GC_{3S} is defined as the frequency of G plus C at silent (i.e. synonymously variable) third positions of sense codons (excluding codons for Trp and Met and stop codons). A standardised measure of bias in synonymous codon preference (B) was also calculated, which can take values between 0 (for unbiased genes) and 1 (Long and Gillespie 1991).

Results

An analysis of the total codon usage summed for each data set is shown in Table 3. A chi-square goodness-of-fit test revealed that genes of *Brugia* and *Onchocerca*, like those of *Schistosoma* (Ellis and Morrison 1994), preferred codons containing A or T at the third base position. This contrasts to those of *Echinococcus*, which preferred codons ending in T, C or G. For example, of

Table 3 Synonymous codon usage expressed as a fraction, in *Brugia*, *Echinococcus*, and *O. volvulus* (*B Brugia*, *E Echinococcus*, *O O. volvulus*, *Am acid* amino acid, * over-represented codon [χ^2 significant at $P<0.05$], + under-represented codon [χ^2 significant at $P<0.05$])

Am Acid	Codon	B	E	O	Am acid	Codon	B	E	O
Gly	GGG	0.04+	0.08+	0.05+	Trp	TGG	1.00	1.00	1.00
Gly	GGA	0.32	0.12+	0.26	End	TGA	0.25	0.20	0.11
Gly	GGT	0.48*	0.51*	0.50*	Cys	TGT	0.54	0.38	0.49
Gly	GGC	0.16+	0.29	0.20	Cys	TGC	0.46	0.62	0.51
Glu	GAG	0.26+	0.65*	0.21+	End	TAG	0.25	0.20	0.22
Glu	GAA	0.74*	0.35+	0.79*	End	TAA	0.50	0.60	0.67
Asp	GAT	0.75*	0.57	0.79*	Tyr	TAT	0.57	0.32+	0.58
Asp	GAC	0.25+	0.43	0.21+	Tyr	TAC	0.43	0.68*	0.42
Val	GTG	0.23	0.30	0.18	Leu	TTG	0.28*	0.20	0.32*
Val	GTA	0.21	0.10+	0.26	Leu	TTA	0.13	0.05	0.15
Val	GTT	0.40*	0.34*	0.41*	Phe	TTT	0.54	0.35+	0.48
Val	GTC	0.16	0.26	0.15+	Phe	TTC	0.46	0.65*	0.52
Ala	GCG	0.15+	0.17+	0.13+	Ser	TCG	0.16	0.18	0.16
Ala	GCA	0.34*	0.15+	0.35 *	Ser	TCA	0.30*	0.08+	0.28*
Ala	GCT	0.37*	0.33*	0.37 *	Ser	TCT	0.20	0.18	0.20
Ala	GCC	0.14+	0.35*	0.15+	Ser	TCC	0.09	0.20	0.13
Arg	AGG	0.04+	0.07+	0.04+	Arg	CGG	0.06+	0.07+	0.04+
Arg	AGA	0.14	0.10	0.13	Arg	CGA	0.24	0.13	0.25*
Ser	AGT	0.15	0.16	0.15	Arg	CGT	0.38*	0.34*	0.43*
Ser	AGC	0.10	0.19	0.08	Arg	CGC	0.14	0.29	0.11+
Lys	AAG	0.36+	0.70*	0.34+	Gln	CAG	0.41	0.56	0.35+
Lys	AAA	0.64*	0.30+	0.66*	Gln	CAA	0.59	0.44	0.65*
Asn	AAT	0.72	0.53	0.71*	His	CAT	0.67*	0.36	0.79*
Asn	AAC	0.28+	0.47	0.29+	His	CAC	0.33+	0.64	0.21+
Met	ATG	1.00	1.00	1.00	Leu	CTG	0.16	0.21	0.14
Ile	ATA	0.20+	0.13+	0.14+	Leu	CTA	0.11	0.08+	0.08+
Ile	ATT	0.57	0.43*	0.61*	Leu	CTT	0.24*	0.18	0.23*
Ile	ATC	0.22+	0.44*	0.25	Leu	CTC	0.09+	0.28	0.08+
Thr	ACG	0.22	0.10+	0.16	Pro	CCG	0.24	0.21	0.24
Thr	ACA	0.32	0.18	0.35*	Pro	CCA	0.51*	0.18	0.50*
Thr	ACT	0.32	0.42*	0.35*	Pro	CCT	0.20	0.34	0.20
Thr	ACC	0.14+	0.30	0.15+	Pro	CCC	0.06+	0.27	0.06+

the four possible synonymous codons coding for ALA, GCA and GCT were over-represented in the filarial worms and schistosomes, whereas GCT and GCC were over-represented in *Echinococcus*. Similarly, for GLU GAA was over-represented in the filarial worms and schistosomes, whereas GAG was over-represented in the *Echinococcus* data set. In addition to such obvious differences, all taxa used certain codons infrequently. AGG, AGA and CGG (ARG), CTA (LEU), CAC (HIS), CCG and CCC (PRO) were identified as low-usage codons (i.e. their frequency of usage is less than 10 per 1000 codons; Ellis et al. 1993) in all taxa (data not shown). Codons for CYS and TRP were also infrequently used by all taxa, but this was the result of the low cysteine and tryptophan content of the proteins encoded by the data set. In general, however, these initial observations implied that the patterns of codon usage of *Brugia*, *Onchocerca* and *Schistosoma* were more similar to each other than to that of *Echinococcus*. Consequently, multivariate analysis was performed to investigate this further.

Heterogeneity in codon usage between genes was investigated by MMDS. Since the pattern of codon usage shown by a gene can result from a bias in the amino acid composition of the gene product, codon usage was expressed as a frequency of synonymous codon usage and analysed by MMDS. This method of analysis removes the effect of the amino acid composition of the gene product from the measure of codon usage (Ellis and Morrison 1994; Ellis et al. 1994). The results obtained are shown in Fig. 1, where the data derived for *S. mansoni* are included for comparative purposes. The analysis divides the genes into two non-overlapping groups. The first group contains all of the genes of *Echinococcus* and the second group contains all of the other genes. Two

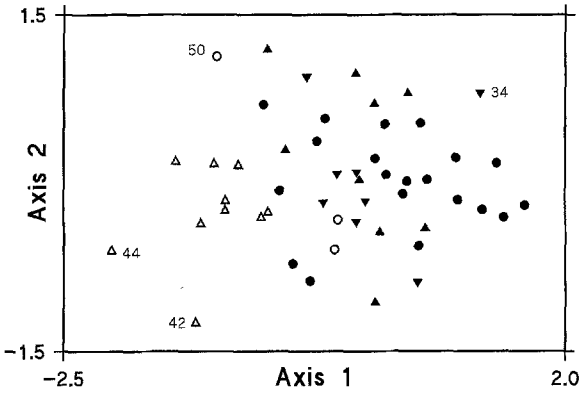

Fig. 1 Metric multidimensional scaling was used to investigate the pattern of codon usage among synonymous codons shown by individual gene sequences in the data set. For each gene sequence, the usage of each codon was expressed as a frequency relative to other synonymous codons. A two-dimensional ordination was used with the Manhattan distance measure (Faith et al. 1987; Belbin 1989). Sequence codes are as given in Tables 1 and 2. The symbols used represent *Brugia* (▼), *Echinococcus* (△), *Onchocerca* (▲), *Schistosoma mansoni* (●) and *S. japonicum* (○)

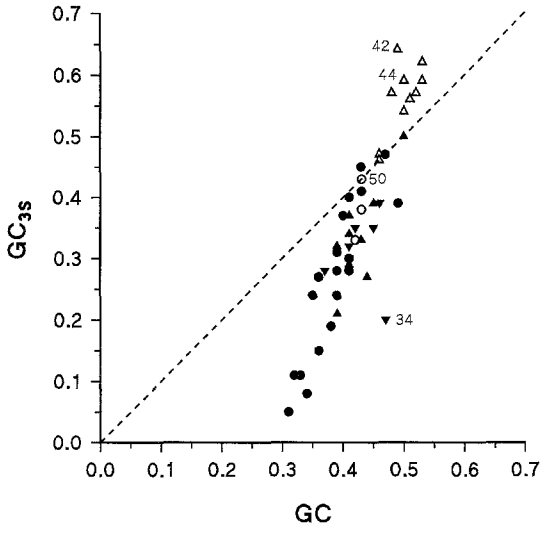

Fig. 2 Frequency of G+C at silent third positions (GC_{3S}) of synonymous codons (excluding codons for Trp and Met and stop codons) relative to the G+C frequency of the whole gene for each of the nucleotide sequences. The *line* represents GC_{3S}=GC. Sequence codes are as given in Tables 1 and 2. Symbols are the same as those shown in Fig. 1

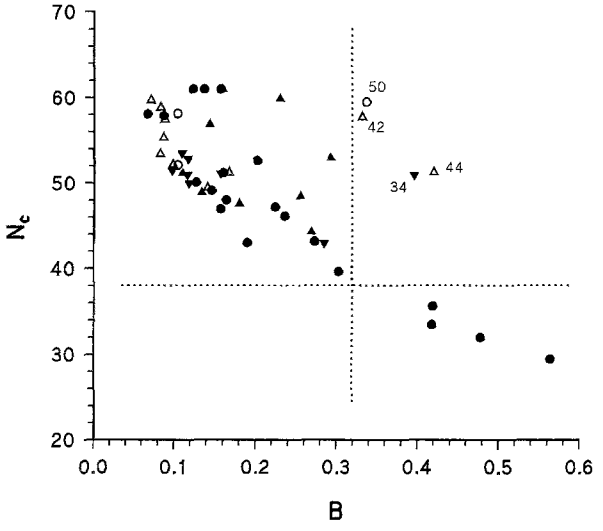

Fig. 3 "Effective" number of codons used by the gene sequence (N_c) relative to the standardised synonymous codon bias (B). Gene sequences *below the horizontal line* are indicated by N_c as having higher codon usage bias than those *above the line*, whereas gene sequences *to the right of the vertical line* are indicated by B as having higher codon bias than those *to the left of the line*. Sequence codes are given as in Tables 1 and 2. Symbols are the same as those shown in Fig. 1

genes of *Echinococcus* (EGDF1P and ECCUB19A) differ in their pattern of codon usage as compared with the rest of the *Echinococcus* genes, which cluster together as a reasonably tight group in the ordination. This analysis, however, shows the patterns of codon usage used by genes of *Brugia*, *Onchocerca* and *Schistosoma* to be more similar to each other than to that shown by genes of *Echinococcus*.

The similarity detected in codon usage between *Brugia*, *Onchocerca* and *Schistosoma* and presented herein is surprising given that current classifications of these organisms group *Echinococcus* and *Schistosoma* into the phylum Platyhelminthes, whereas *Brugia* and *Onchocerca* belong to the phylum Nematoda. However, the data shown in Tables 1 and 2 reveal the reason for the relationships detected. Three independent measures of codon usage bias (N_c, GC_{3S} and B) were calculated for genes in the data set. The variation in N_c, GC_{3S} and B values between genes in the data set confirms the existence of only low levels of bias in the patterns of codon usage shown by these genes. The average GC content of *Echinococcus* genes is significantly higher than that of the other species, explaining the preferences in codon usage observed by the goodness-of-fit tests on the summed data for each taxon. A plot of GC_{3S} against gene GC content expressed as a fraction is shown in Fig. 2. All of the *Echinococcus* genes have GC_{3S} values that are generally higher than those determined for genes from the other taxa. Interestingly, the GC_{3S} values derived from all the genes in the data set fell onto a line that correlated well with that previously identified for *S. mansoni* (Ellis and Morrison 1994).

A plot of N_c against B for the genes is shown in Fig. 3. As noted previously, bias is detected in four genes of *S. mansoni* by both N_c and B (Ellis and Morrison 1994). However, B detected bias in four genes where N_c did not (SCMFABP, EGDF1P, ECCUB19A and BRPMF22G). All four of these genes have high B values and cluster away from the other genes in the ordination of Fig. 1. The first three of these genes contain only a small number of codons, and the inability of N_c to detect bias in them may reflect their small size. It is not clear why N_c does not detect bias in BRPMF22G, since the goodness-of-fit tests show this gene to contain several codons that are either over- or under-represented (not shown).

Discussion

This study investigated the patterns of codon usage and bias in genes of *Brugia*, *Echinococcus*, *Onchocerca* and *Schistosoma* species. There are two main outcomes from this study. The first observation is that the patterns detected by the average codon usage across all genes for each taxon were similar in *Brugia*, *Onchocerca* and *Schistosoma* in that codons over-represented in the data set contained A or T at the third base position, whereas genes of *Echinococcus* preferred codons containing T, C or G. Alvarez et al. (1993) and Kalinna and McManus (1994) concluded that bias in genes of *Echinococcus* was related to an over-representation of codons containing G or C at the third base position. The results presented herein support these claims but also show a significant preference for T at the third base position of some synonymous codons.

The differences detected in codon usage between *Echinococcus* and the other three taxa appear to be relat-

ed directly to a significant difference in the gene base composition between *Echinococcus* as compared with the other taxa. Since directional mutational pressure drives the genome base composition to a higher or lower percentage of GC over time, the genome base composition indirectly determines codon usage (Ellis et al. 1993). Consequently, it would appear that the similarity detected in patterns of codon usage between filarial worms and schistosomes is probably the result of convergent evolution acting on the genome base composition. Such a conclusion is supported by the values assigned to the GC content of genomic DNA from these taxa, which are 20% (*B. malayi*), 36% (*O. volvulus*), 34% (*S. mansoni*) and 44% (*E. multilocularis*; Barrett 1981; Meadows and Simpson 1989; Hammond and Bianco 1992). However, it is noteworthy that the GC content of the genes analysed herein is generally greater than the GC content of the respective genomic DNA from which they are derived.

The second point to be made from this study is that the pattern of codon usage detected by the average codon usage summed across all genes for each taxon is not necessarily that shown by all the genes from that taxon. The analysis presented herein quite clearly shows that all genes in the data set show widely different patterns of codon usage (Fig. 1) and bias (Tables 1, 2). This study therefore significantly extends the observations of Unnasch et al. (1992), Alvarez et al. (1993), Kalinna and McManus (1994) and Hammond (1994), who reported only the average codon usage summed across all genes of *O. volvulus*, *Echinococcus* and *Brugia*, respectively. Recently, Ellis and Morrison (1994) described heterogeneity in codon usage among genes of *S. mansoni*. The results presented herein are in complete agreement with the general trends observed by these authors.

Bias in codon usage has been measured in many ways. Historically, these start with the relatively complicated method of McLachlan et al. (1984), but recently much simpler methods have been proposed. Sharp and Li (1987) devised the codon adaptation index (CAI), whereas Sharp and Devine (1989) suggested GC_{3S}. Shields et al. (1988) used the chi-square coefficient for determining deviation from equal usage of synonymous codons; Wright (1991) proposed N_c, and Long and Gillespie (1991) devised B. Wright (1991) also proposed plotting different measures of bias against each other so as to identify genes containing bias due to the amino acid composition of the gene product or to bias in synonymous codon usage. This approach was therefore adopted in this study. In addition, since there is no reliable information about the level of expression of the gene sequences analysed in this study, it was necessary to restrict our analysis to measures of bias that do not rely on this information. Stenico et al. (1994) presented the results obtained from an analysis of genes of *Caenorhabditis elegans*, which found a correlation between codon usage bias, N_c, GC_{3S} and levels of gene expression. In the present study, analysis of GC_{3S} and N_c values revealed no correlation between base composition and level of codon usage bias within any one taxon, although a strong correlation between GC_{3S} and gene base composition was evident for all taxa. Although N_c and B detected significant heterogeneity in codon usage bias in all genes of the data set studied herein the overall magnitude of the level of bias was generally low and of magnitude similar to that reported for genes of *S. mansoni* (Ellis and Morrison 1994). Consequently, any further investigation into the relationship between bias and levels of gene expression requires the acquisition of the necessary information on a much larger data set than the one analysed herein. Such data may be forthcoming from a variety of "genome sequencing" programs that have recently been initiated.

Acknowledgements We thank Prof. A.M. Johnson of UTS for his encouragement and support of this program.

References

Alarcon CM, Donelson JE (1991) Translational elongation factor 1 alpha (EF-1 alpha) of *Onchocerca volvulus*. Mol Biochem Parasitol 48:105–107

Alvarez F, Garat B, Musto H, Picon M, Ehrlich R (1993) Tendencies in *Echinococcus* sp. codon usage. Mem Inst Oswaldo Cruz 88:345–346

Andersson SGE, Kurland CG (1990) Codon preferences in free-living micro-organisms. Microbiol Rev 54:198–210

Barrett J (1981) Biochemistry of parasitic helminths. Macmillan, London

Belbin L (1989) PATN – pattern analysis package technical reference. CSIRO, Canberra

Cookson E, Blaxter ML, Selkirk ME (1992) Identification of the major soluble cuticular glycoprotein of lymphatic filarial nematode parasites (gp29) as a secretory homolog of glutathione peroxidase. Proc Natl Acad Sci USA 89:5839–5841

Da Silva CM, Ferreira HB, Picon M, Gorfinkel N, Ehrlich R, Zaha A (1993) Molecular cloning and characterisation of actin genes from *Echinococcus granulosus*. Mol Biochem Parasitol 60:209–220

Dahmen A, Gallin M, Schumacher M, Erttman KD (1993) Molecular cloning and pre-mRNA maturation of *Onchocerca volvulus* paramyosin. Mol Biochem Parasitol 57:335–338

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:216–223

Dissanayake S, Xu M, Piessens WF (1992) Filarial parasites contain a *ras* homolog of the *TC4/ran/Spi1* family. Mol Biochem Parasitol 56:259–268

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

Ellenberger DL, Pieniazek NJ, Lammie PJ (1989) Nucleotide sequence of *Brugia pahangi* 17.4 kD protein. Nucl Acids Res 17:10121

Ellis J, Morrison DA (1995) *Schistosoma mansoni*: patterns of codon usage and bias. Parasitology 110:53–60

Ellis J, Griffin H, Morrison DA, Johnson AM (1993) Analysis of dinucleotide frequency and codon usage in the phylum Apicomplexa. Gene 126:163–170

Ellis J, Morrison DA, Avery D, Johnson AM (1994) Comparison of the patterns of codon usage and bias among individual genes of the coccidia and piroplasms. Parasitology 109:265–272

Esteves A, Dallagiovanna B, Ehrlich R (1993) A developmentally regulated gene of *Echinococcus granulosus* codes for a 15.5-kilodalton polypeptide related to fatty acid binding proteins. Mol Biochem Parasitol 58:215–222

Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distances: a theoretical model and computer simulations. Vegetatio 69:57–68

Frosch PM, Frosch M, Pfister T, Schaad V, Bitter-Surmann D (1991) Cloning and characterisation of an immunodominant major surface antigen of *Echinococcus multilocularis*. Mol Biochem Parasitol 48:121–130

Fuhrman JA, Lane WS, Smith RF, Piessens WF, Perler FB (1992) Transmission blocking antibodies recognise microfilarial chitinase in brugian lymphatic filariasis. Proc Natl Acad Sci USA 89:1548–1552

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

Guenette S, Prichard RK, Klein RD, Matlashewski G (1991) Characterisation of a beta-tubulin gene and beta-gene products of *Brugia pahangi*. Mol Biochem Parasitol 44:153–164

Hammond MP (1994) Codon usage and gene organisation in *Brugia*. Parasitol Res 80:173–175

Hammond MP, Bianco AE (1992) Genes and genomes of parasitic nematodes. Parasitol Today 8:299–305

Hemmings L, McManus DP (1991) The diagnostic value and molecular characterisation of an *Echinococcus multilocularis* antigen gene clone. Mol Biochem Parasitol 44:53–62

Henkle KJ, Liebrau E, Muller S, Bergmann B, Walter RD (1991) Characterisation and molecular cloning of a Cu/Zn superoxide dismutase from the human parasite *Onchocerca volvulus*. Infect Immun 59:2063–2069

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2:13–34

Kalinna BH, McManus DP (1994) Codon usage in *Echinococcus*. Exp Parasitol 79:72–76

Lloyd AT, Sharp PM (1992a) Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. Nucleic Acids Res 20:5289–5295

Lloyd AT, Sharp PM (1992b) CODONS: a microcomputer program for codon usage analysis. J Hered 83:239–240

Lobos E, Altmann M, Menrod G, Weiss N, Rudin W, Karam M (1990) Identification of an *Onchocerca volvulus* cDNA encoding a low-molecular-weight antigen uniquely recognised by onchocerciasis patient sera. Mol Biochem Parasitol 39:135–146

Long M, Gillespie JH (1991) Codon usage divergence of homologous vertebrate genes and codon usage clock. J Mol Evol 32:6–15

Lustigman S, Brotman BH, Huima T, Price AM, McKerrow JH (1992) Molecular cloning and characterisation of onchocystatin, a cysteine protease inhibitor of *Onchocerca volvulus*. J Biol Chem 267:17339–17346

McLachlan AD, Staden R, Boswell DR (1984) A method for measuring the non-random bias of a codon usage table. Nucleic Acids Res 12:9567–9575

Meadows HM, Simpson AJG (1989) Codon usage in *Schistosoma*. Mol Biochem Parasitol 36:291–299

Morrison DA, Ellis J, Johnson AM (1994) Comparison of distance matrix techniques for estimating phylogenies from codon usage divergence. J Mol Evol 39:533–536

Mühlschlegel F, Sygulla L, Frosch P, Massetti P, Frosch M (1993) Paramyosin of *Echinococcus granulosus* cDNA sequence and characterisation of a tegumental antigen. Parasitol Res 79:229–264

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–237

Perrine KG, Denker JA, Nilsen TW (1988) A multi-copy gene encodes a potentially protective antigen in *Brugia malayi*. Mol Biochem Parasitol 30:97–104

Rodrigues JJS, Ferreira HB, Zaha A (1993) Molecular cloning and characterisation of an *Echinococcus granulosus* cDNA encoding malate dehydrogenase. Mol Biochem Parasitol 60:157–160

Rothstein N, Rajan TV (1991) Characterisation of an hsp70 gene from the human filarial parasite, *Brugia malayi* (Nematoda). Mol Biochem Parasitol 49:229–237

Scott AL, Dinman J, Sussman DJ, Yentbuhr P, Ward S (1989) Major sperm protein genes from *Onchocerca volvulus*. Mol Biochem Parasitol 36:119–126

Selkirk M, Yazdanbakhsh M, Freedman D, Blaxter M, Cookson E, Jenkins RE, Williams SA (1991) A proline-rich structural protein of the surface sheath of larval *Brugia* filarial nematode parasites. J Biol Chem 266:11002–11008

Sharp PM, Devine KM (1989) Codon usage and gene expression level in *Dictyostelium discoideum*: highly expressed genes do "prefer" optimal codons. Nucleic Acids Res 17:5029–5039

Sharp PM, Li W-H (1987) The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295

Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol 5:704–716

Stenico M, Lloyd AT, Sharp PM (1994) Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res 22:2437–2446

Unnasch T, Katholi CR, Coate LM (1992) *Onchocerca volvulus*: frequency of codon usage. Exp Parasitol 75:457–459

Waine GJ, Becker M, Kalinna BH, Yang W, McManus D (1993a) Cloning and functional expression of a *Schistosoma japonicum* cDNA homologous to the enolase gene family. Biochem Biophys Res Commun 195:1211–1217

Waine GJ, Becker M, Yang W, Kalinna B, McManus DP (1993b) Cloning, molecular characterization, and functional activity of *Schistosoma japonicum* glyceraldehyde-3-phosphate dehydrogenase, a putative vaccine candidate against schistosomiasis. Infect Immun 61:4716–4723

Werner C, Rajan T (1992a) Comparison of the body wall myosin heavy chain sequences from *Onchocerca volvulus* and *Brugia malayi*. Mol Biochem Parasitol 50:255–260

Werner C, Rajan T (1992b) Characterisation of a myosin heavy chain gene from *Brugia malayi*. Mol Biochem Parasitol 50:261–268

Wright F (1990) The effective number of codons used in a gene. Gene 87:23–29

Zeng W, Donelson JE (1992) The actin genes of *Onchocerca volvulus*. Mol Biochem Parasitol 55:207–216