Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*

T. H. LE[1], M. S. PEARSON[1], D. BLAIR[2], N. DAI[1], L. H. ZHANG[1] *and* D. P. McMANUS[1]*

[1] *Molecular Parasitology Laboratory, Australian Centre for International and Tropical Health and Nutrition, The Queensland Institute of Medical Research and The University of Queensland, Brisbane, Queensland 4029, Australia*
[2] *School of Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia*

(*Received 23 June 2001; revised 9 August 2001; accepted 9 August 2001*)

SUMMARY

Unlike other members of the genus, *Echinococcus granulosus* is known to exhibit considerable levels of variation in biology, physiology and molecular genetics. Indeed, some of the taxa regarded as 'genotypes' within *E. granulosus* might be sufficiently distinct as to merit specific status. Here, complete mitochondrial genomes are presented of 2 genotypes of *E. granulosus* (G1–sheep-dog strain: G4–horse-dog strain) and of another taeniid cestode, *Taenia crassiceps*. These genomes are characterized and compared with those of *Echinococcus multilocularis* and *Hymenolepis diminuta*. Genomes of all the species are very similar in structure, length and base-composition. Pairwise comparisons of concatenated protein-coding genes indicate that the G1 and G4 genotypes of *E. granulosus* are almost as distant from each other as each is from a distinct species, *E. multilocularis*. Sequences for the variable genes *atp6* and *nad3* were obtained from additional genotypes of *E. granulosus*, from *E. vogeli* and *E. oligarthrus*. Again, pairwise comparisons showed the distinctiveness of the G1 and G4 genotypes. Phylogenetic analyses of concatenated *atp6*, *nad1* (partial) and *cox1* (partial) genes from *E. multilocularis*, *E. vogeli*, *E. oligarthrus*, 5 genotypes of *E. granulosus*, and using *T. crassiceps* as an outgroup, yielded the same results. We conclude that the sheep-dog and horse-dog strains of *E. granulosus* should be regarded as distinct at the specific level.

Key words: *Echinococcus granulosus*, mitochondrial genome, mitochondrial DNA, strain, genotype; horse-dog strain, sheep-dog strain, phylogeny.

INTRODUCTION

Only 4 of the 16 nominal *Echinococcus* species are generally accepted as being taxonomically valid– *Echinococcus granulosus*, *E. multilocularis*, *E. vogeli* and *E. oligarthrus*. All the other taxa are regarded *as* subspecific variants or strains of *E. granulosus* (Thompson & McManus, 2001). These conclusions were based on differences in morphologies of adult worms, host ranges, life-cycle patterns, the nature and location of the hydatid cyst, and biochemical and molecular characteristics (Thompson & McManus, 2001). It is now increasingly clear that the near-cosmopolitan *E. granulosus* exhibits considerable variation at the genetic level and that a re-evaluation of its taxonomy is merited. Indeed, based on a range of different biological, epidemiological, biochemical and molecular-genetic criteria, separate species status for the horse-dog (G4 genotype) and sheep-dog (G1 genotype) strains has been advocated (Bowles, Blair & McManus, 1995; Thompson, 1995; Thompson, Lymbery & Constantine, 1995). The extensive intra-specific variation in nominal *E. granulosus* must impact on the epidemiology, pathology and control of hydatid disease (Thompson & Lymbery, 1988; Thompson, 1995), with important implications also for the design and development of vaccines, diagnostic reagents and drugs. By contrast, there appears to be very limited genetic variation within *E. multilocularis* (McManus & Bryant, 1995; Haag *et al*. 1997; Rinder *et al*. 1997), and there are no available data to indicate that either *E. vogeli* or *E. oligarthrus* is variable.

Mitochondrial (mt) sequences provide rich sources of data for research in evolutionary biology, population genetics and phylogenetics and are increasingly being used in studies of the genus *Echinococcus* (see Le, Blair & McManus, 2000*a*). To date, molecular studies, using mainly mtDNA sequences, have identified 9 distinct genotypes within *E. granulosus* (Bowles, Blair & McManus, 1992, 1994; Bowles & McManus, 1993*a*, *b*; Scott & McManus, 1994; Scott *et al*. 1997). Nonetheless, there is still a paucity of information regarding the structure and characteristics of the mt genomes of this and other

* Corresponding author: Molecular Parasitology Laboratory, Australian Centre for International and Tropical Health and Nutrition, The Queensland Institute of Medical Research and The University of Queensland, Brisbane, Queensland 4029, Australia. Tel: +61 7 3362 0401. Fax: +61 7 3362 0104. E-mail: donM@qimr.edu.au

DOI: 10.1017/S0031182001008976 Printed in the United Kingdom

Table 1. Position and characteristics of mitochondrial genes and non-coding sequences in *Echinococcus granulosus* (genotype G1), *E. multilocularis* and *Taenia crassiceps*

(Egr: *Echinococcus granulosus* (genotype 1); Emu: *E. multilocularis*; Tcr: *Taenia crassiceps*. NR1: first non-coding region; NR2: second non-coding region. *See text concerning start and stop codons for *cox1* and the length of *trnT*. The sequence tract indicated here for *trnT* forms a tRNA lacking a paired DHU arm (see text).)

	Length of genes and sequences						Codon used for						Position		
	Nucleotide			Amino acid			Initiation			Termination			(5′→3′)		
Gene and sequence	Egr	Emu	Tcr	Egr	Emu	Tcr	Egr	Emu	Tcr	Egr	Emu	Tcr	Egr	Emu	Tcr
cox3	648	648	645	215	215	214	ATG	ATG	GTG	TAG	TAG	TAG	1–648	1–648	1–645
trnH	65	68	71										649–713	650–717	646–716
cob	1068	1068	1074	355	355	357	ATG	ATG	ATG	TAA	TAA	TAA	717–1784	720–1787	720–1793
nad4L	261	261	261	86	86	86	GTG	GTG	ATG	TAA	TAG	TAG	1798–2058	1798–2058	1793–2053
nad4	1260	1260	1260	419	419	419	ATG	ATG	GTG	TAG	TAG	TAG	2019–3278	2019–3278	2014–3273
trnQ	62	61	63										3282–3343	3285–3345	3279–3345
trnF	63	63	64										3343–3405	3345–3407	3342–3405
trnM	66	65	67										3402–3467	3404–3468	3402–3468
atp6	513	516	513	170	171	170	ATG	ATG	GTG	TAG	TAG	TAA	3473–3985	3747–3989	3469–3981
nad2	882	882	879	293	293	292	ATG	ATG	ATG	TAG	TAG	TAG	3994–4875	3998–4879	3981–4859
trnV	63	63	64										4900–4962	4902–4964	4867–4930
trnA	64	64	64										4968–5031	4970–5033	4935–4998
trnD	65	63	66										5032–5096	5034–5096	5002–5067
nad1	894	894	894	297	297	297	GTG	ATG	ATG	TAA	TAG	TAG	5100–5993	5100–5993	5075–5968
trnN	66	66	67										6010–6075	6011–6076	5971–6037
trnP	63	63	66										6082–6144	6083–6145	6053–6118
trnI	62	62	64										6145–6206	6146–6207	6119–6182
trnK	62	66	65										6213–6274	6214–6279	6188–6252
nad3	348	348	348	115	115	115	ATG	ATG	GTG	TAG	TAA	TAG	6277–6624	6280–6627	6255–6602
$trnS_{1(AGN)}$	59	59	59										6623–6681	6639–6697	6601–6659
trnW	67	65	64										6690–6756	6706–6770	6661–6724
*cox1**	1581	1581	1596	526	526	531	GTG	GTA	ATG	TAG	TAG	TAG	6787–8367	6801–8381	6746–8341
*trnT**	55	55	58										8367–8421	8385–8439	8337–8394
rrnL (*16S*)	978	983	963										8422–9399	8441–9423	8395–9357
trnC	63	64	59										9400–9462	9424–9487	9358–9416
rrnS (*12S*)	719	723	722										9463–10181	9488–10210	9417–10138
cox2	582	582	585	193	193	194	GTG	GTG	ATG	TAG	TAG	TAG	10182–10763	10211–10792	10139–10723
trnE	67	68	69										10779–10845	10810–10877	10725–10793
nad6	456	456	453	151	151	150	ATG	ATG	ATG	TAG	TAA	TAA	10849–11304	10880–11335	10796–11248
trnY	66	66	69										11316–11381	11348–11413	11256–11324
NR1	66	183	65										11382–11447	11414–11596	11325–11389
$trnL_{1(CUN)}$	73	73	65										11448–11520	11597–11669	11390–11454
$trnS_{2(UCN)}$	58	58	59										11559–11616	11707–11764	11480–11538
$trnL_{3(UUN)}$	64	64	65										11630–11693	11778–11841	11545–11609
trnR	58	58	56										11703–11760	11857–11914	11617–11672
nad5	1572	1575	1569	523	524	522	ATG	ATG	ATG	TAG	TAA	TAA	11763–13334	11916–13490	11674–13242
NR2	184	177	194										13335–13518	13491–13667	13243–13436
trnG	67	68	64										13519–13585	13668–13735	13437–13500

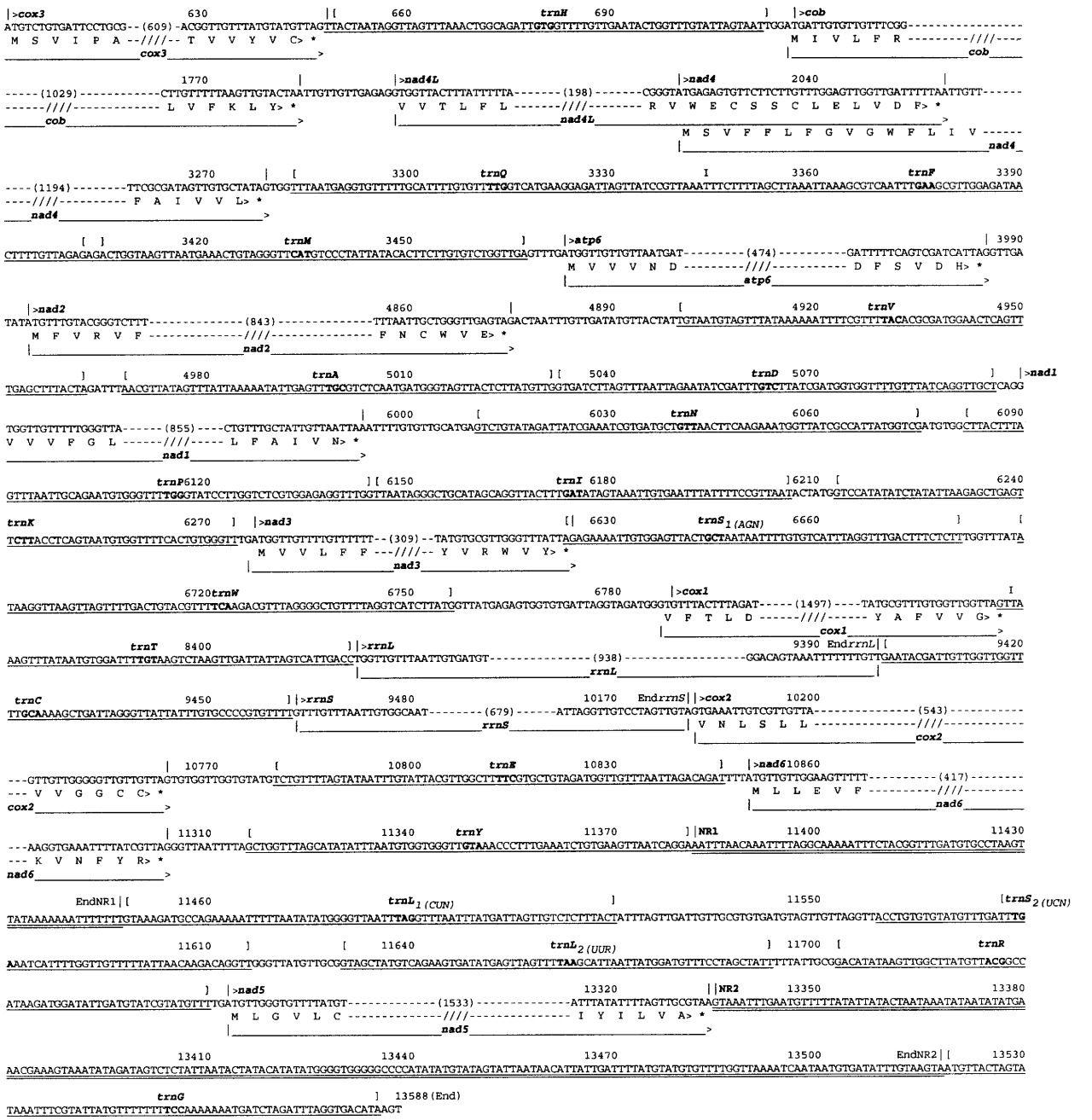

Fig. 1. Abbreviated mitochondrial sequence of *Echinococcus granulosus* (G1 genotype) showing gene arrangement (see text for details).

cestodes which has hindered efforts to further advance epidemiological and phylogenetic studies and to address taxonomic questions (McManus & Bryant, 1995; Bowles *et al*. 1995; Le *et al*. 2000*a*).

Here, we describe, characterize and compare the complete mt sequences for the common sheep (G1 genotype) and horse (G4 genotype) strains of *E. granulosus* relative to the mt sequence of *E. multilocularis* (M. Nakao, GeneBank Accession number AF018440) and another taeniid species, *Taenia crassiceps* (Le *et al*. 2000). These comparisons show that the G1 and G4 genotypes of *E. granulosus* are almost as distinct from each other as either is from *E. multilocularis*. Some comparisons have also been made with the recently published complete sequence of the mt genome of the more distantly related cyclophyllidean cestode, *Hymenolepis diminuta* (see von Nickisch-Rosenegk, Brown & Boore, 2001). Additionally, a comparison of protein-encoding genes (*atp6*, *nad1*, *nad3* and *cox1*) for a number of *Echinococcus* genotypes and species (*E. granulosus* genotypes G1, G4, G6, G7, G8; *E. multilocularis*; *E. vogeli*; *E. oligathrus*) is made for consideration of their genetic variation and phylogeny.

Table 2. Nucleotide codon usage for mitochondrial protein-encoding genes of *Echinococcus* and *Taenia crassiceps*

(EgrG1: *Echinococcus granulosus* (G1 genotype), 3355 codons used for 3343 amino acids and 12 stop codons. EgrG4: *Echinococcus granulosus* (G4 genotype), 3355 codons used for 3343 amino acid and 12 stop codons. Emu: *Echinococcus multilocularis*, 3357 codons used for 3345 amino acids and 12 stop codons. Tcr: *Taenia crassiceps*, 3359 codons used for 3347 amino acids and 12 stop codons. NC, nucleotide codons; Ab, amino acid abbreviation; No., number of codons. Putative initiation (ATG, GTA and GTG) and termination codons (TAA and TAG) are underlined.)

		EgrG1		EgrG4		Emu		Tcr				EgrG1		EgrG4		Emu		Tcr	
NC	Ab	No.	%	No.	%	No.	%	No.	%	NC	Ab	No.	%	No.	%	No.	%	No.	%
TTT	Phe	378	11·2	393	11·7	404	12·0	412	12·3	TAT	Tyr	205	6·1	201	6·0	193	5·7	199	5·9
TTC	Phe	20	0·6	14	0·4	14	0·4	20	0·6	TAC	Tyr	11	0·3	14	0·4	19	0·6	18	0·5
TTA	Leu	145	4·3	158	4·7	181	5·4	314	9·3	TAA	*	4	0·1	6	0·2	4	0·1	4	0·1
TTG	Leu	304	9·0	292	8·7	272	8·1	154	4·6	TAG	*	8	0·2	6	0·2	8	0·2	8	0·2
CTT	Leu	32	1·0	24	0·7	24	0·7	24	0·7	CAT	His	46	1·4	49	1·5	46	1·4	51	1·5
CTC	Leu	2	0·1	1	< 0·1	—	0·0	—	0·0	CAC	His	4	0·1	3	0·1	3	0·1	1	< 0·1
CTA	Leu	8	0·2	12	0·4	7	0·2	15	0·4	CAA	Gln	8	0·2	7	0·2	10	0·3	13	0·4
CTG	Leu	15	0·4	12	0·4	14	0·4	7	0·2	CAG	Gln	17	0·5	18	0·5	14	0·4	8	0·2
ATT	Ile	140	4·2	136	4·1	149	4·4	174	5·2	AAT	Asn	77	2·3	83	2·5	87	2·6	89	2·6
ATC	Ile	9	0·3	10	0·3	5	0·1	2	0·1	AAC	Asn	4	0·1	4	0·1	1	< 0·1	5	0·1
ATA	Ile	50	1·5	61	1·8	67	2·0	126	3·7	AAA	Asn	16	0·5	13	0·4	18	0·5	59	1·8
ATG	Met	90	2·7	96	2·9	91	2·7	99	2·9	AAG	Lys	42	1·2	44	1·3	43	1·3	49	1·5
GTT	Val	267	7·9	255	7·6	240	7·1	183	5·4	GAT	Asp	78	2·3	79	2·4	74	2·2	81	2·4
GTC	Val	12	0·4	7	0·2	9	0·3	3	0·1	GAC	Asp	2	0·1	3	0·1	2	0·1	3	0·1
GTA	Val	47	1·4	65	1·9	75	2·2	75	2·2	GAA	Glu	17	0·5	20	0·6	17	0·5	31	0·9
GTG	Val	139	4·1	131	3·9	112	3·3	67	2·0	GAG	Glu	48	1·4	46	1·4	46	1·4	26	0·8
TCT	Ser	98	2·9	96	2·9	99	2·9	123	3·7	TGT	Cys	140	4·2	129	3·8	144	4·3	126	3·7
TCC	Ser	6	0·2	3	0·1	2	0·1	2	0·1	TGC	Cys	9	0·3	11	0·3	4	0·1	9	0·3
TCA	Ser	20	0·6	23	0·7	36	1·1	40	1·2	TGA	Trp	29	0·9	37	1·1	33	1·0	51	1·5
TCG	Ser	36	1·1	30	0·9	21	0·6	15	0·4	TGG	Trp	66	2·0	60	1·8	58	1·7	35	1·0
CCT	Pro	36	1·1	45	1·3	45	1·3	48	1·4	CGT	Arg	35	1·0	34	1·0	34	1·0	42	1·2
CCC	Pro	2	0·1	2	0·1	—	0·0	1	< 0·1	CGC	Arg	2	0·1	4	0·1	1	< 0·1	1	< 0·1
CCA	Pro	14	0·4	9	0·3	13	0·4	16	0·5	CGA	Arg	1	< 0·1	3	0·1	7	0·2	2	0·1
CCG	Pro	17	0·5	15	0·4	13	0·4	7	0·2	CGG	Arg	12	0·4	12	0·4	9	0·3	3	0·1
ACT	Thr	60	1·8	68	2·0	65	1·9	70	2·1	AGT	Ser	93	2·8	107	3·2	113	3·4	104	3·1
ACC	Thr	2	0·1	1	< 0·1	—	0·0	—	0·0	AGC	Ser	9	0·3	6	0·2	7	0·2	1	< 0·1
ACA	Thr	8	0·2	2	0·1	9	0·3	14	0·4	AGA	Ser	25	0·7	30	0·9	31	0·9	43	1·3
ACG	Thr	19	0·6	20	0·6	14	0·4	8	0·2	AGG	Ser	49	1·5	38	1·1	33	1·0	29	0·9
GCT	Ala	48	1·4	52	1·5	59	1·8	50	1·5	GGT	Gly	146	4·4	140	4·2	155	4·6	111	3·3
GCC	Ala	9	0·3	4	0·1	4	0·1	—	0·0	GGC	Gly	12	0·4	10	0·3	5	0·1	7	0·2
GCA	Ala	7	0·2	8	0·2	8	0·2	6	0·2	GGA	Gly	22	0·7	22	0·7	21	0·6	33	1·0
GCG	Ala	16	0·5	15	0·4	11	0·3	6	0·2	GGG	Gly	62	1·8	56	1·7	54	1·6	35	1·0

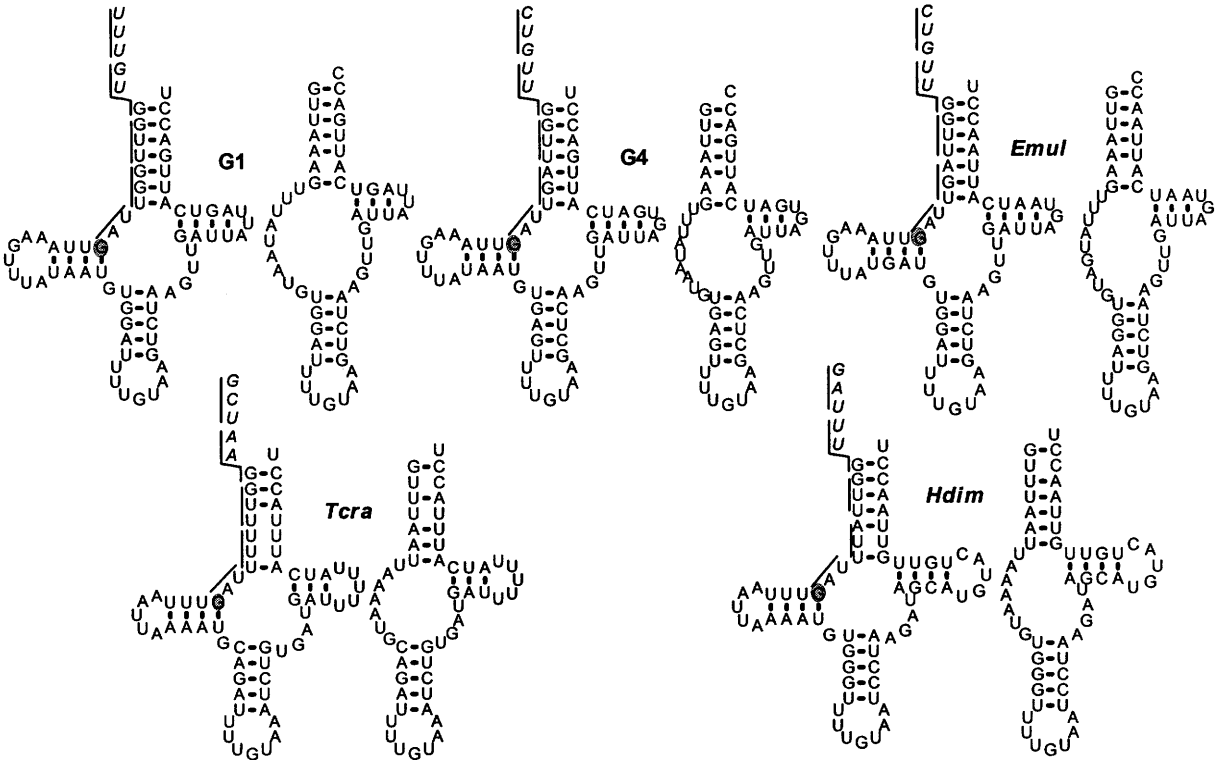

Fig. 2. Alternative structures for tRNA(T) in *Echinococcus granulosus* genotypes G1 and G4 (indicated as G1 and G4 in figure), *E. multilocularis* (*Emul*), *Taenia crassiceps* (*Tcra*) and *Hymenolepis diminuta* (*Hdim*). See text for details. The left-hand drawing of each pair shows the tRNA(T) structure with a paired DHU arm. If *cox1* is terminated with the codon TAG, then there is a 10 nt overlap between *cox1* and *trnT*. The reading frame of the overlapping sequence tract is indicated by vertical (or diagonal) lines and the 5 nt in *cox1* preceding tRNA(T) are shown in italics (with T shown as U for consistency). The right-hand drawing of each pair shows the alternative structure for tRNA(T) lacking a paired DHU arm. In each case, this structure starts with the nucleotide (G) at the end of the putative TAG stop codon for *cox1*. Thus, there needs be no overlap between *cox1* and *trnT* if the TAG stop codon is abbreviated to T or TA, or at most a 1 nt overlap if the full stop codon is used.

MATERIALS AND METHODS

Parasite materials and determination of mtDNA sequence

Echinococcus granulosus G1 (sheep strain) and G4 (horse strain) genotypes were of United Kingdom origin, G6 (camel strain) was obtained from Kenya, G7 (pig strain) was obtained from Poland and the G8 (cervid strain) was of Alaskan origin. *E. vogeli* was obtained from South America and *E. oligarthrus* was from Panama. Techniques for genomic DNA extraction from starting materials (protoscoleces in all cases) and PCR application for obtaining the mt fragments have been described (Le, Blair & McManus 2001*a*). The *Taenia crassiceps* (American strain: Zarlenga & George, 1995) mtDNA molecule was sequenced from available mt clones in combination with PCR (see Le *et al*. 2000). The complete mtDNA sequences for genotypes 1 and 4 of *E. granulosus* were also obtained using PCR strategies (Le *et al*. 2001). In brief, a combination of 'long PCR' and conventional PCR amplified overlapping fragments spanning the mt genome. Some PCR products were sequenced directly while others were cloned. Primer-walking was used to obtain overlapping sequences on both strands. Sequencing of PCR fragments and/or recombinant plasmid DNA was performed on an automated sequencer (ABI 377, Applied Biosystems) using specific or M13 universal sequencing primers. Both strands were completely sequenced and at least 6 sequences (3 from each strand) were aligned to obtain the final sequence for characterization. PCR was also used to amplify and subsequently sequence the *atp6* gene from the *E. granulosus* genotypes G1, G4, G6, G7, G8, *E. vogeli* and *E. oligarthrus* and the *nad3* gene from all these taxa except *E. oligarthrus*.

Sequence analysis

Sequences were aligned using AssemblyLIGN v 1.9c and analysed using the MacVector 6.5.3 package (Oxford Molecular Group). Preliminary identity of a sequence or a region was assigned by comparison with corresponding platyhelminth sequences obtained by us (Le *et al*. 2000; Le, Blair & McManus, 2000*b*) or available in the GenBank database (http://www.ncbi.nlm.nih.gov/Web/Genbank) using BLAST searches. Protein-encoding genes were

Table 3. Base composition in the complete mtDNA, protein-encoding and ribosomal RNA (rRNA) sequences
(EgrG1: *Echinococcus granulosus* (G1 genotype); EgrG4: *E. granulosus* (G4 genotype), Emu: *E. multilocularis*, Tcr: *Taenia crassiceps*.)

	Complete mtDNA sequence					Protein-encoding sequence					rRNA-encoding sequence				
spp.	Length (bp)	T %	C %	A %	G %	T+A %	Length (bp)	T %	C %	A %	G %	T+A %	Length (bp)	T %	C %	A %	G %	T+A %
EgrG1	13588	47·9	8·0	19·1	25·0	67·0	10065	49·8	7·6	16·9	25·7	66·7	3095	42·4	9·6	25·0	23·0	67·4
EgrG4	13598	48·0	7·7	19·9	24·3	67·9	10065	50·0	7·4	17·8	24·9	67·8	3106	42·2	9·4	25·7	22·6	67·9
Emu	13738	48·5	7·6	20·6	23·4	69·1	10071	50·5	7·1	18·4	24·0	68·9	3098	42·5	9·7	25·7	22·1	68·2
Tcr	13503	48·6	7·6	25·4	18·3	74·0	10077	50·5	7·1	23·4	18·9	72·9	3097	43·4	9·1	30·3	17·2	73·7

identified by sequence similarity of translated open reading frames to mt gene sequences available in the GenBank database. The platyhelminth mt genetic code (Garey & Wolstenholme, 1989; Telford *et al*. 2000; Nakao *et al*. 2000) was used for translation as done previously for a number of platyhelminth species (Le *et al*. 2000). The possibility of unusual initiation and termination codons (Wolstenholme, 1992) was considered when characterizing protein-encoding genes. In the case of the small, poorly conserved genes (*nad3*, *nad4L*, *nad6*), hydrophilicity profiles, drawn in MacVector 6.5.3, were additionally used to confirm identity.

The identities of the ribosomal RNA sequences were established based on their similarity with those found in other parasitic platyhelminthes (Le *et al*. 2000*a*, *b*, 2001) and by their potential to form rRNA-like secondary structures. Ends of rRNA genes were not determined experimentally: consequently, these genes were assumed to consist of the entire sequence tract lying between flanking genes. Most of the transfer RNAs were identified by preliminary screening with tRNAscan-SE (Lowe & Eddy, 1997) with parameters specified for mitochondrial/chloroplast DNA using the invertebrate mt genetic code for tRNA prediction (available at http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Remaining tRNA genes were identified by inspection of the sequences, taking into account both sequence similarity to homologues from other species and ability to form the appropriate secondary structure. All secondary structures were drawn using RNAViz (De Rijk & De Wachter, 1997). Throughout, we have used the convention for abbreviating names of mt genes and their products as used by von Nickisch-Rosenegk *et al*. (2001).

The extent of genetic divergence among the detected mt genotypes was estimated by pairwise comparisons of nucleotide and inferred amino acid sequences. These were aligned by eye and submitted to MEGA2 (Kumar *et al*. 2001) for phylogenetic analysis. Pairwise distances among nucleotide sequences were calculated using the Kimura 2-parameter method to compensate for multiple substitutions. Distances among inferred amino acid sequences were calculated using a Poisson correction for multiple hits. Trees were constructed using the minimum evolution approach. *Taenia crassiceps* was used as the outgroup for rooting trees. Bootstrap resampling was used to gain an indication of the level of support for internal branches.

RESULTS AND DISCUSSION

Gene organization and content

The complete mt sequences for *Echinococcus granulosus* G1 genotype (EgrG1) (13 588 bp, GenBank Accession number AF297617), *E. granulosus* G4 genotype (EgrG4) (13 598 bp, GenBank Accession

Table 4. Amino acid codon usage of the mitochondrial protein-encoding genes

(AA: abbreviation of amino acid codons as 3 letters; Ab: as 1 letter. No.; number of codons. Tcr; *Taenia crassiceps*; Emu; *Echinococcus multilocularis*; Egr; *E. granulosus* (G1; genotype 1; G4; genotype 4).)

		Tcr		Emu		EgrGl		EgrG4				Tcr		Emu		EgrGl		EgrG4	
AA	Ab	No.	%	No.	%	No.	%	No.	%	AA	Ab	No.	%	No.	%	No.	%	No.	%
Ala	A	62	1·9	82	2·4	80	2·4	79	2·4	Met	M	100	3·0	91	2·7	90	2·7	96	2·9
Cys	C	135	4·0	148	4·4	149	4·4	140	4·2	Asn	N	153	4·6	106	3·2	97	2·9	100	3·0
Asp	D	84	2·5	76	2·3	80	2·4	82	2·5	Pro	P	72	2·1	71	2·1	69	2·1	71	2·1
Glu	E	57	1·7	63	1·9	65	1·9	66	2·0	Gln	Q	21	0·6	24	0·7	25	0·7	25	0·7
Phe	F	431	12·9	418	12·5	398	11·9	407	12·2	Arg	R	48	1·4	51	1·5	50	1·5	53	1·6
Gly	G	186	5·5	235	7·0	242	7·2	228	6·8	Ser	S	357	10·7	342	10·3	336	10·1	333	10·0
His	H	52	1·6	49	1·5	50	1·5	52	1·6	Thr	T	92	2·8	88	2·6	89	2·7	91	2·7
Ile	I	302	9·0	221	6·6	199	5·9	207	6·2	Val	V	329	9·8	436	13·1	465	13·9	458	13·7
Lys	K	49	1·5	43	1·3	42	1·3	44	1·3	Trp	W	86	2·6	91	2·8	95	2·9	97	2·9
Leu	L	514	15·3	498	14·9	506	15·1	499	14·9	Tyr	Y	217	6·5	212	6·3	216	6·4	215	6·4

number AF346403), and *T. crassiceps* (13 503 bp, GenBank Accession number AF216699) were determined. The genomes are relatively small with that of *T. crassiceps* being the smallest known among metazoans (Wolstenholme, 1992; Boore, 1999; Le *et al*. 2000). The coding portions (97·4–98·6 % of the total mt genome) and the protein-encoding portions (around 74 %) are similar in length in all species and genotypes. Individual genes are very similar in length among the cestode species. The positions, lengths, and other features of genes and non-coding sequences for *E. granulosus* G1 genotype and *T. crassiceps* are compared with *E. multilocularis* (Nakao *et al*. 2001) in Table 1. The complete sequence for the *E. granulosus* G1 genotype is presented semi-schematically in Fig. 1.

All the 36 genes typically found in helminth mt genomes (12 protein-, 22 tRNA- and 2 rRNA-encoding genes) have been identified and are transcribed in the same direction (Fig. 1). As is the case with other helminths (Okimoto *et al*. 1992; Keddie, Higazi & Unnasch, 1998; Le *et al*. 2000; Le, Blair & McManus 2000*a*, *b*, 2001; von Nickisch-Rosenegk *et al*. 2001) *atp8* is absent. The gene arrangement is basically identical in all cestode species (although in *H. diminuta*, the adjacent $trnS_{2(UCN)}$ and $trnL_{1(CUN)}$ have exchanged places relative to the situation in taeniids – see von Nickisch-Rosenegk *et al*. 2001) and is similar to that found in trematodes (except *S. mansoni*, see Le *et al*. 2000, 2001*b*). Genes abut one another or are separated by short intergenic sequences. However, each genome has 2 somewhat longer non-coding regions: one (designated NR1) sited between *trnY* and $trnL_{1(CUN)}$, and the other (designated NR2) located downstream of *nad5*. The lengths of NR2 are similar among all the cestodes. In the case of NR1, however, that of *E. multilocularis* is 3 times the length seen in any of the other cestodes and accounts for the overall larger mt genome size of this species.

Some pairs of adjacent genes overlap in the mt genomes of the cestodes reported here: (i) there is an overlap of 40 nt (including the stop codon of *nad4L*) between *nad4L* and *nad4* in a different reading frame, a phenomenon seen in all sequenced parasitic platyhelminths, with the exceptions of *S. mansoni* (overlap only 28 nt: Le *et al*. 2000, 2001) and *H. diminuta* (overlap of 16 nt: von Nickisch-Rosenegk *et al*. 2001); (ii) a 1 nt overlap (T) occurs between *trnQ* and *trnF* in all *Echinococcus* species and genotypes, but not in *T. crassiceps*; (iii) a 4 nt overlap is present in all cestode species between *trnF* and *trnM*; (iv) depending on interpretation, an overlap of up to 10 nt occurs between the 3′ end of *cox1* and *trnT* in all cestodes (discussed further below); (v) 2 nt (AG) at the 5′ end of $trnS_{1(AGN)}$ are shared with the termination TAG codon of *nad3* in the G1 and G4 genotypes of *E. granulosus* and *T. crassiceps*, but not in *E. multilocularis*; (vi) in *T. crassiceps*, the stop codon of *cob* overlaps by 1 nt with *nad4L* and a similar situation occurs for *atp6* and *nad2*. In the cases listed in (v) and (vi) it is possible that the stop codon of the upstream gene is in fact abbreviated (to TA or T), as has been noted in a number of mitochondrial genomes (see Wolstenholme, 1992; Le *et al*. 2001; Le, Blair & McManus, 2001).

Initiation and termination codons

In almost all cases, ATG or GTG initiate, and TAA or TAG terminate translation of protein-encoding genes among the taeniid cestodes (Table 1). However, the same start and stop codons are not always used in all homologous genes among the different species (Table 1). For example, GTG acts as an initiation codon in *nad1* of *E. granulosus* G1 genotype whereas ATG performs the same function in the G4 genotype, *T. crassiceps* and *E. multilocularis*. The *E. granulosus* G4 genotype utilizes the stop codons TAA and TAG equally (Table 2), unlike the

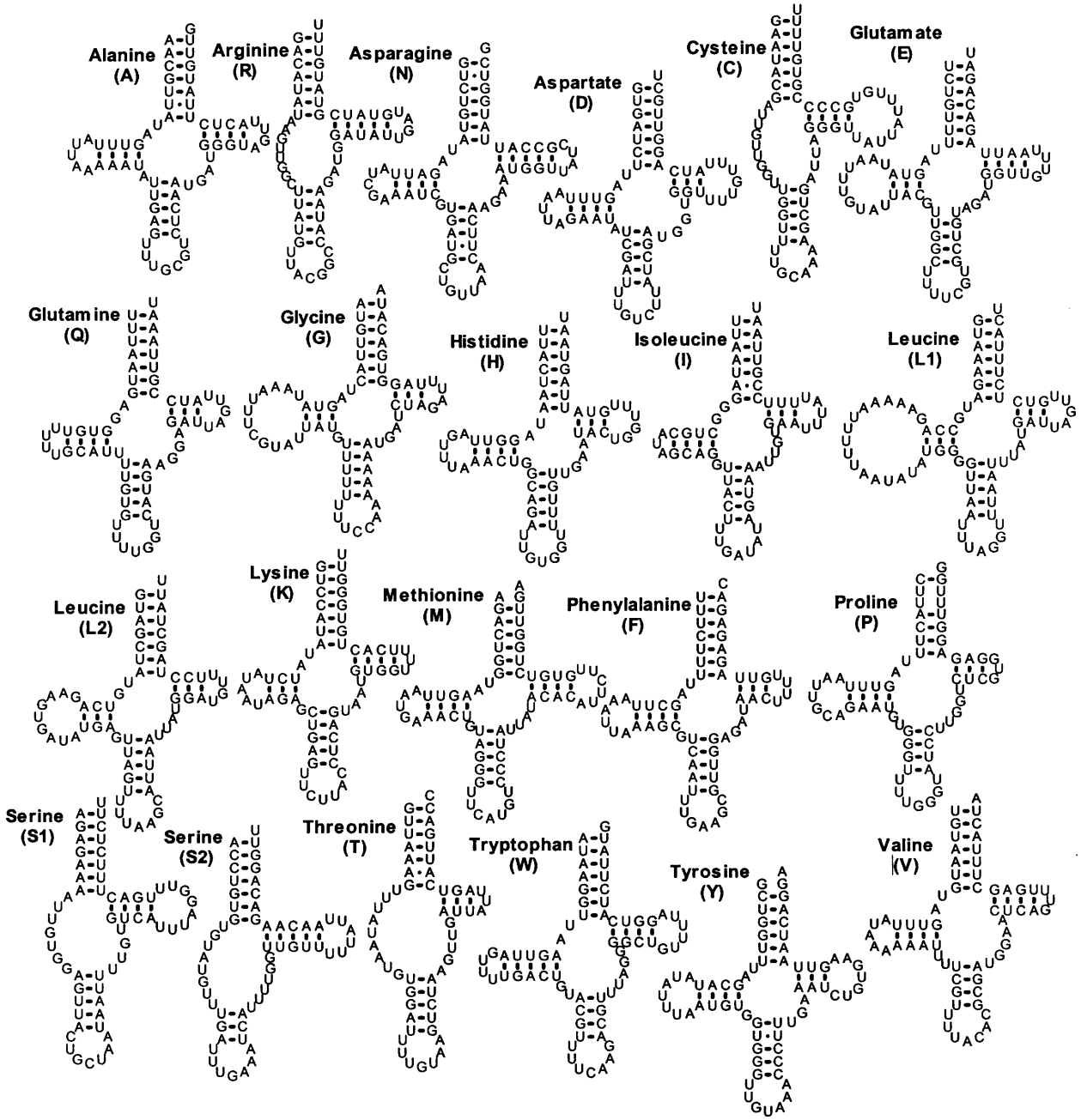

Fig. 3. Secondary structure models for the 22 tRNAs of *Echinococcus granulosus* G1 genotype. See text for details. The structure shown for tRNA(T) is the form lacking the DHU arm.

situation in other taxa in which TAA is less common.

Resolution of initiation and termination codons in *cox1* has proved to be difficult. In all 4 taeniid species or genotypes, a typical initiation codon (ATG) is found near the start of *cox1*. We would regard this as the true start codon, except for the fact that there is a 2-nt deletion just downstream of it in the G4 genotype of *E. granulosus*, thus changing the reading frame for this taxon. A more likely start codon for the *Echinococcus* species/genotypes is therefore GTG/GTA located 9 codons downstream from the ATG codon. This position aligns with the codon (GTT) chosen by von Nickisch-Rosenegk *et al.* (2001) as the initiator of transcription in *cox1* of *H. diminuta*. The codon TTG found at this position in *T. crassiceps* could be a start codon. However, there is an in-frame ATG located 3 codons upstream of this which might be the true start codon in that species.

In their analysis of the mt genome of *H. diminuta*, von Nickisch-Rosenegk *et al.* (2001) inferred that *cox1* terminated with an abbreviated stop codon (T) and thus did not overlap the downstream *trnT*. They also pointed out that an in-frame stop codon (TAG) occurred downstream of the abbreviated codon, implying a 10 nt overlap with *trnT* if this were the true stop codon. We have examined this region in our sequences from taeniids. For each of the cestode species, it is possible to construct alternative structures for tRNA(T), one with a paired DHU arm and

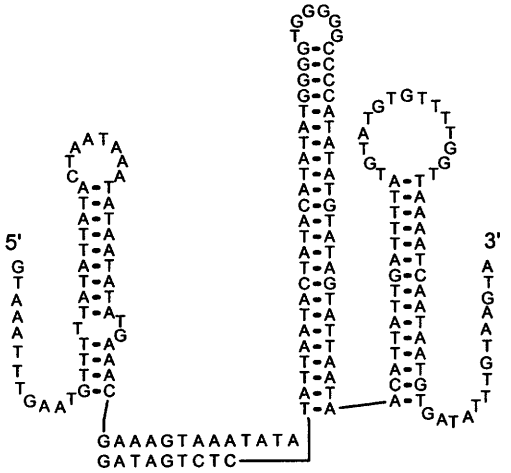

Fig. 4. Putative secondary structure for the NR2 (non-coding region 2) of *Echinococcus granulosus* G1 genotype.

one lacking this arm (Fig. 2). Structures lacking the DHU arm overlap by only 1 nt (G) with the TAG codon mentioned above. Abbreviation of such a codon to T or TA remains a possibility. Structures possessing the DHU arm, such as the structure figured by von Nickisch-Rosenegk *et al*. (2001), overlap *cox1* by 10 nt, assuming that TAG is the stop codon. Von Nickisch-Rosenegk *et al*. (2001) suggested that the codon TTG, the last nt of which overlaps the structure possessing the DHU arm, might be abbreviated to T and act as a stop codon, thus eliminating any overlap with the downstream *trnT*. Codons in the same position as this TTG in other cestode taxa do not always start with T (Fig. 2 – only *E. multilocularis* and *E. granulosus* G4 genotype have such codons) and therefore could not act as abbreviated stop codons. We think it likely that the true stop codon for *cox1* is the TAG (or TA) mentioned above. However, we are undecided as to which structure for tRNA(T) is to be preferred and consequently we are uncertain as to the extent of overlap between *cox1* and *trnT*.

Nucleotide and amino acid composition

The A+T content of the complete mt genomes differ slightly among the *Echinococcus* species and genotypes (67·0–69·1 %) on the one hand and *T. crassiceps* (74 %) on the other (Table 3). There is very low use of C (~ 8 %) in all species, but the use of A and G differs; lower A (~ 20 %) and higher G (24–25 %) occur in *Echinococcus* compared with *T. crassiceps* (Table 3). These values are consistent throughout the protein and ribosomal coding sequences (Table 3). The amino acid compositions of the protein-encoding mtDNA sequences are shown in Table 4. Phe (11·9–12·9 %), Leu (14·9–15·3 %), Ser (10·0–10·7 %), and Val (9·8–13·7 %) are the most used, making up ~ 50 % of the total number, and Gln (0·6–0·7 %), His (1·5–1·6 %), Lys (1·3–1·5 %) and Arg (~ 1·5) are the least frequently used. This correlates with the high T and low C composition of the genes and the correspondingly frequent use of T in codons.

Transfer RNAs

The complement of 22 tRNA-encoding genes in each of the cestode mt genomes presented here is typical of that found in other metazoans. As an example, predicted structures of tRNAs in the mtDNA of *E. granulosus* (G1 genotype) are presented in Fig. 3. Uncertainties about the structure for tRNA(T) have been discussed above and alternatives shown in Fig. 2. Lengths of tRNAs (ranging from 54–73 nt) are similar between genes of the 4 cestodes, but these sizes are less conserved between the genera *Echinococcus* and *Taenia*. The most different in length is $trnL_{1(CUN)}$ (73 nt in both *Echinococcus* species but only 65 nt in *T. crassiceps*) (Table 1). As in other parasitic platyhelminths and a number of other metazoans, both the tRNAs specific for serine lack a DHU arm (Fig. 3). The *trnR* and *trnC* genes, in all species, have a DHU replacement loop, a feature never (*trnR*) or sometimes (*trnC*) observed in trematodes (see Le *et al*. 2001). There is an unusually large loop closing the DHU arm in the $trnL_{1(CUN)}$ gene structure of *E. granulosus* (Fig. 3) and *E. multilocularis*.

Non-coding sequences

Apart from short intergenic sequences ranging from 1 to 39 nt (the longest in all 4 taeniid taxa being that between $trnL_{1(CUN)}$ and $trnS_{2(UCN)}$,), there are 2 other intergenic or non-coding regions (NR) which are functionally unassigned (Table 1). One, designated NR1, lies between *trnY* and $trnL_{1(CUN)}$ and is much shorter in the *E. granulosus* genotypes (66 nt) than in *E. multilocularis* (183 nt) but very similar in length and sequence to NRl in *T. crassiceps* (65 nt). NRl in *T. crassiceps*, with the inclusion of a few bases at the 3′ end of the *trnY*, forms a stem of 23 bp with a capping loop of 7 nt (von Nickisch-Rosenegk *et al*. 2001). Despite a degree of sequence similarity with *T. crassiceps*, the NR1 of *E. granulosus* G1 and G4 genotypes can form only a much shorter stem of 7 bp or fewer, and the inclusion of the 3′ end of *trnY* does not lead to formation of a longer stem. Similarly, the initial 65 nt of the *E. multilocularis* NR1, which has some sequence similarity with the NR1 in other taeniids, cannot fold on itself to form long stem-loop structures. However, the complete NR1 in *E. multilocularis* (183 nt) has the potential to form long stems (von Nickisch-Rosenegk *et al*. 2001). It is noteworthy that among the 4 taeniid species/genotypes discussed here, *E. multilocularis* stands out in possessing a long NR1 with a strong secondary structure.

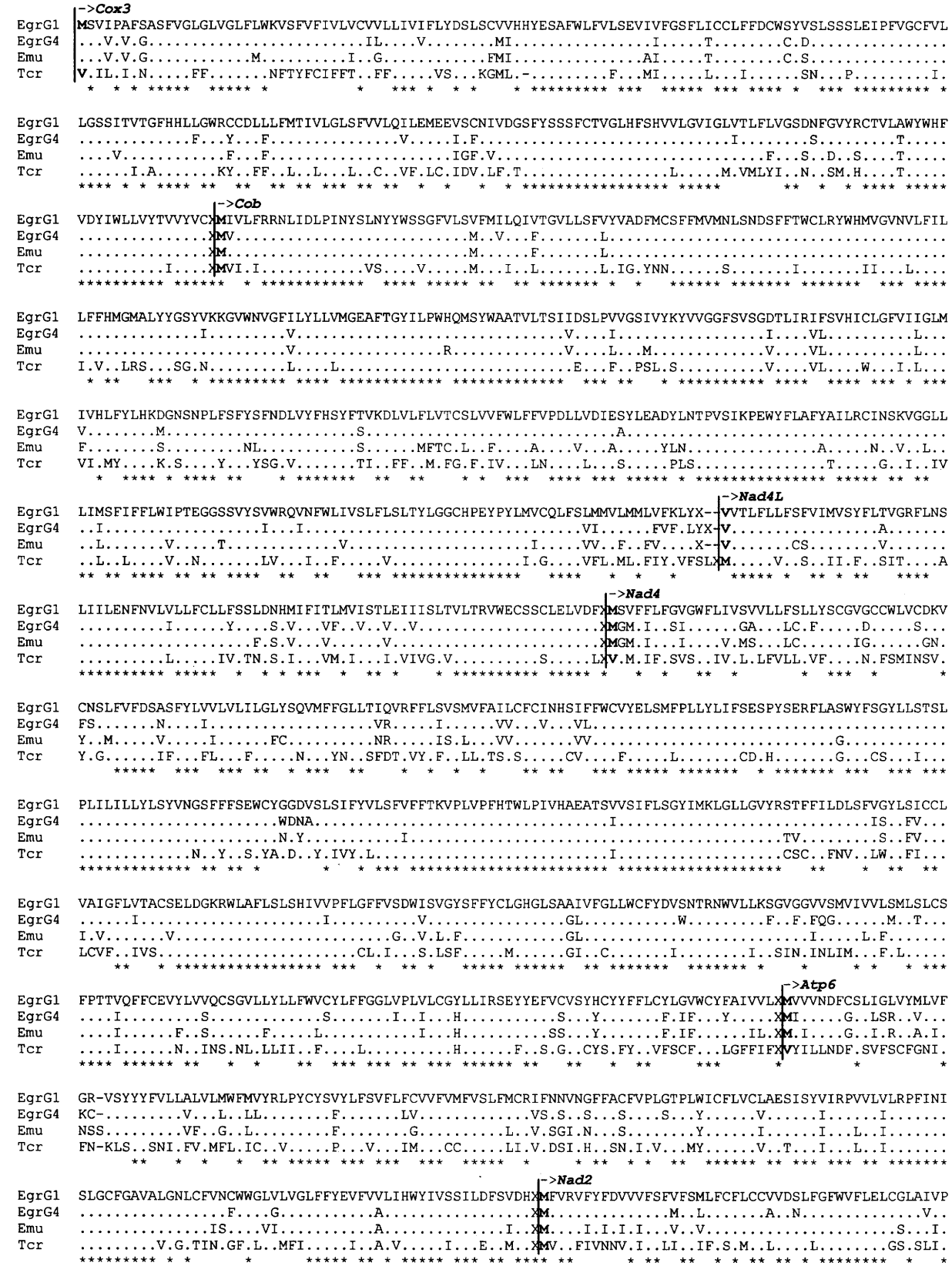

Fig. 5. For legend see p. 108.

The NR2 is more uniform in length among *Echinococcus* species/genotypes (184 in G1, 182 in G4 and 177 in *E. multilocularis*) and the sequences are similar in all cases. Von Nickisch-Rosenegk *et al.* (2001) have proposed a secondary structure for this region in *E. multilocularis*. Comparisons among the *Echinococcus* species/genotypes allow us to refine this slightly, and our interpretation for *E. granulosus* G1 genotype (there are only minor differences in the other 2 forms of *Echinococcus*) is shown in Fig 4. In

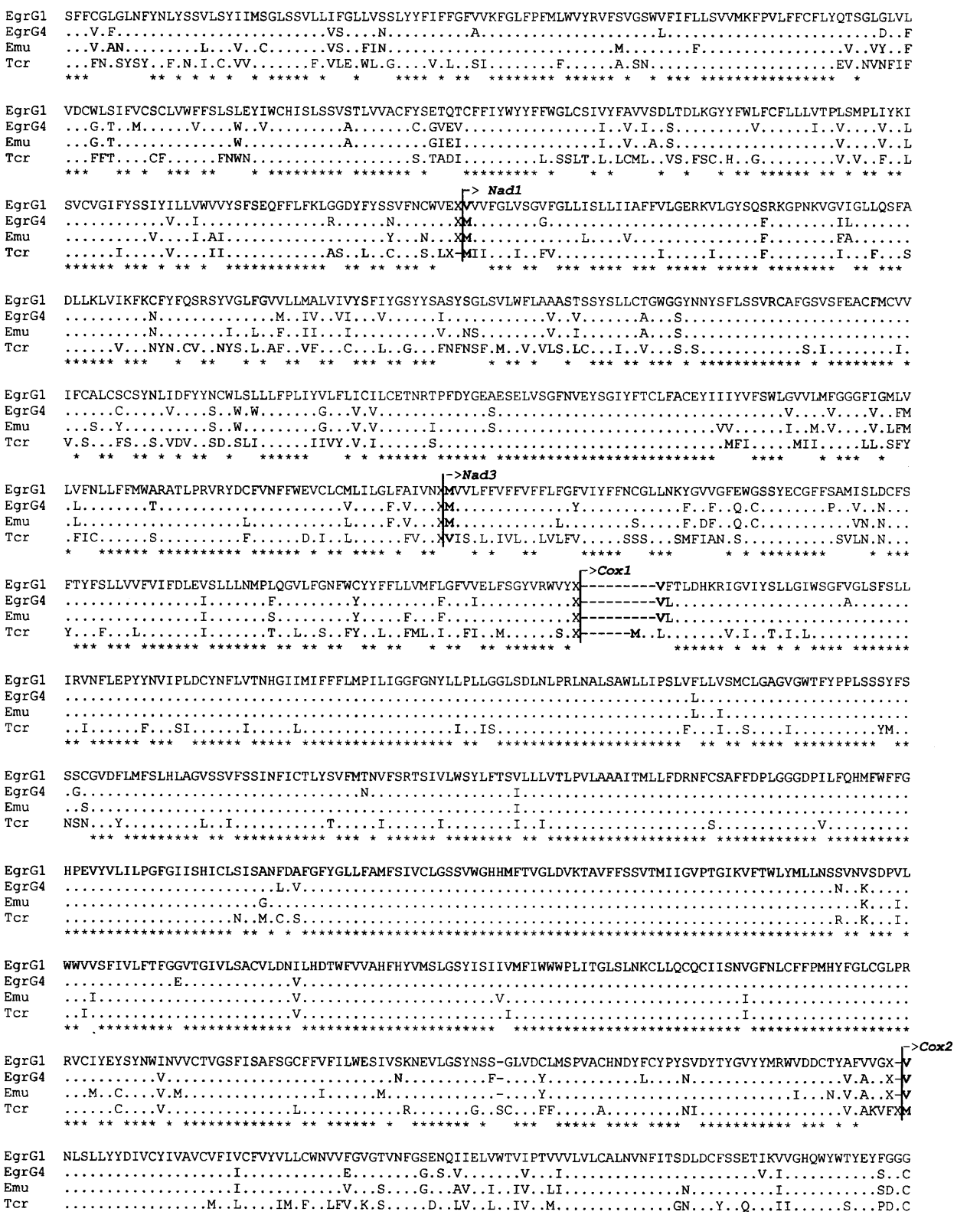

Fig. 5. For legend see p. 108.

contrast with NR1, the NR2 of *T. crassiceps* is very different from those of the *Echinococcus* species in sequence and in secondary structure (as proposed by von Nickisch-Rosenegk *et al*. 2001) as well as being slightly longer (194 nt).

Mitochondrial sequence variation in Echinococcus *and* E. granulosus *genotypes*

Now that complete mt genomes are available for 2 genotypes of *E. granulosus*, for *E. multilocularis* and

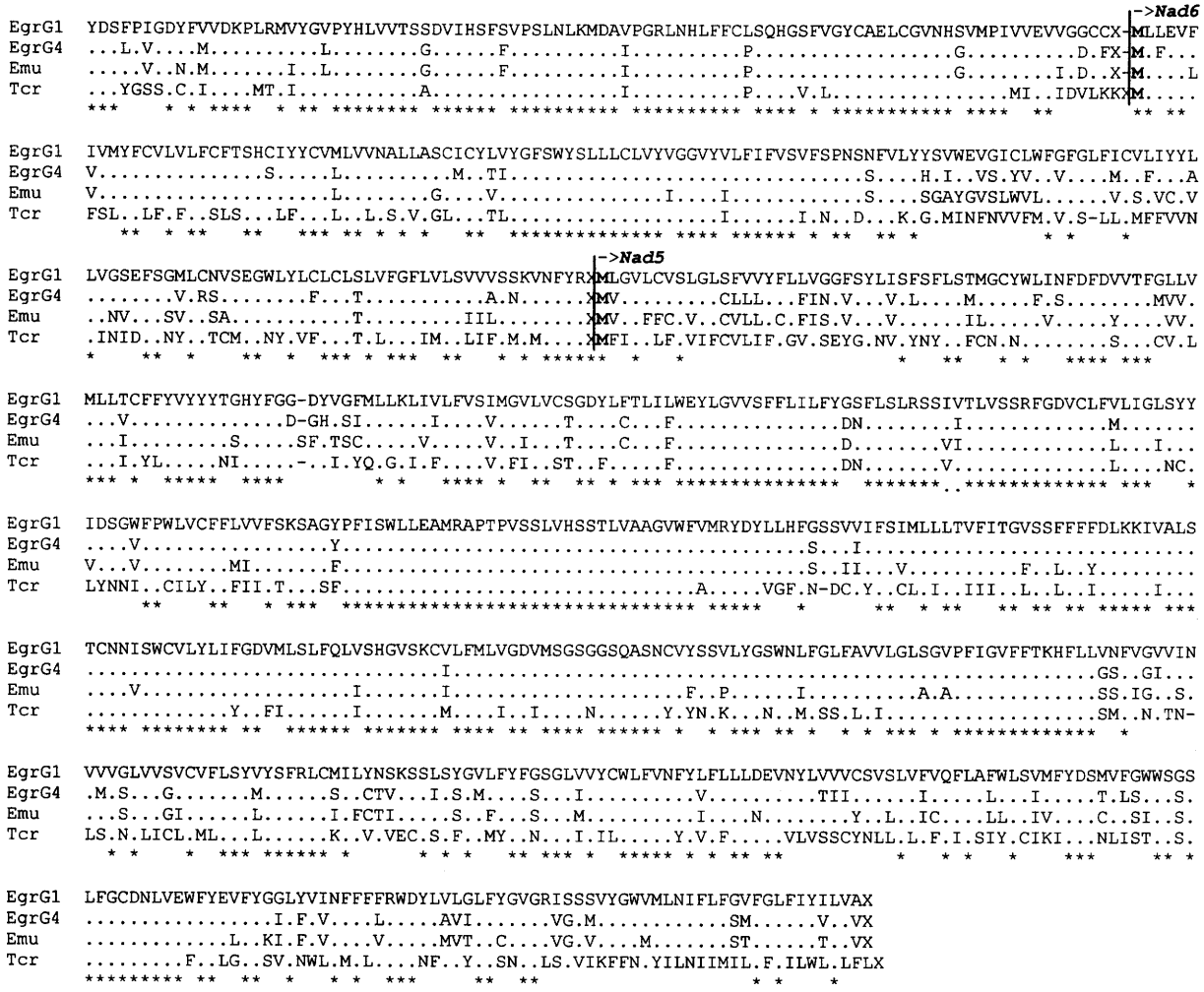

Fig. 5. An alignment of amino acid sequences of the 12 nt protein-encoding genes of *Echinococcus granulosus* genotypes 1 (EgrG1) and 4 (EgrG4), *E. multilocularis* (Emu) and *Taenia crassiceps* (Tcr). Termination codons are marked with the letter X. Dots (.) indicate residues identical with those in EgrG1. Sites conserved in all taxa are indicated by an asterisk (*) under the alignment. Amino acids for the initiation codons (either M or V) are shown in bold to mark the start position of the proteins. See text concerning the start codon for *cox1*.

for an additional taeniid (*T. crassiceps*), we are in a position to use these data to (i) make a preliminary statement as to which mt genes are the most variable and therefore likely to be useful at shallow phylogenetic depths (e.g. at the level of species or genotype) and (ii) measure the divergence between genotypes of *E. granulosus* relative to other taeniids.

A useful first step in assessing variability of genes is to inspect alignments of different genes. Fig. 5 shows an alignment of all 12 protein sequences from the 4 taxa. Differences are most noticeable among proteins such as Cox3, Nad4L, Atp6, Nad3, Cox2 and Nad6 that are generally less conserved in mt genomes. Some proteins, such as Nad5, have tracts that are highly conserved and tracts that are very variable. Cox1 is the most conserved protein among these species, as has been observed in other parasitic platyhelminths (Le *et al.* 2001; Le, Blair & McManus, 2001). The assumption that *cox1* is therefore a good candidate gene for the study of deep phylogenies needs to be tested. Morgan & Blair (1998) found that, despite its apparent conservatism, the *cox1* gene in trematodes had only a relatively few sites free to vary and consequently became saturated with substitutions even at shallow phylogenetic depths.

For 2 of the variable genes, *atp6* and *nad3*, we obtained sequences from additional taxa: *E. granulosus* genotypes 1, 4, 6, 7, 8 (EgrG1, EgrG4, EgrG6, EgrG7 and EgrG8), *E. multilocularis*, *E. vogeli*, *E. oligarthrus* (not *nad3*) and *T. crassiceps*. The percentage pairwise comparison of nucleotide and amino acid composition for *nad3* is shown in Table 5A and for *atp6* is presented in Table 5B. The nucleotide divergence is less than amino acid divergence in all cases, implying that there are few synonymous substitutions. Of the 348 nucleotide positions in the *nad3* alignment, 40 (11·5%) were variable among the *Echinococcus* species and genotypes and 103 (29·6%) were variable when com-

Table 5. Percentage pairwise divergences of nucleotides (above diagonal) and amino acids (below diagonal) of the *nad3* gene (A) and *atp6* gene (B) for genotypes G1, G4, G6, G7, G8 of *Echinococcus granulosus*, *E. multilocularis*, *E. vogeli*, *E. oligarthrus* (B only) and *Taenia crassiceps*

(Egr; *Echinococcus granulosus* (genotypes 1, 4, 6, 7, 8 designated as G1, G4, G6, G7, and G8, respectively), Emu; *E. multilocularis*, Evo; *E. vogeli*, Eol; *E. oligarthrus* and Tcr; *Taenia crassiceps*.)

A

	EgrG1	EgrG4	EgrG6	EgrG7	EgrG8	Emu	Evo	Tcr
EgrG1	—	7·5	7·8	7·8	8·1	10·6	10·9	28·5
EgrG4	11·3	—	8·3	8·3	8·6	11·2	11·5	29·0
EgrG6	11·3	12·2	—	0·0	2·0	7·8	8·6	29·6
EgrG7	11·3	12·2	0·0	—	2·0	7·8	8·6	29·6
EgrG8	12·2	13·0	6·1	6·1	—	8·6	9·5	29·0
Emu	13·0	13·9	11·3	11·3	12·2	—	10·6	29·0
Evo	15·7	16·5	13·9	13·9	13·9	13·9	—	29·3
Tcr	39·1	40·0	40·0	40·0	40·0	33·9	34·5	—

B

	EgrG1	EgrG4	EgrG6	EgrG7	EgrG8	Emu	Evo	Eol	Tcr
EgrG1	—	13·8	16·2	15·8	16·6	19·8	16·4	19·3	36·0
EgrG4	16·5	—	13·8	13·8	15·2	17·0	14·2	17·9	34·1
EgrG6	19·4	15·9	—	0·6	4·7	17·0	16·0	16·0	32·9
EgrG7	18·8	15·3	1·2	—	4·5	17·1	16·0	16·0	33·1
EgrG8	19·4	16·5	5·9	4·7	—	16·7	15·6	17·2	33·3
Emu	18·7	17·0	18·1	17·5	17·5	—	15·1	18·2	32·4
Evo	18·8	13·5	17·6	17·1	16·5	14·0	—	16·0	33·1
Eol	21·2	19·4	20·6	19·4	20·0	18·1	16·5	—	32·8
Tcr	42·4	38·2	41·3	41·3	41·3	38·2	38·4	40·7	—

Table 6. Divergence (%) in mitochondrial protein-coding (nucleotide; above diagonal, and amino acid; below diagonal) and in nucleotide sequences of *rrnL* (above diagonal) and *rrnS* (below diagonal) of the cestodes reported in this study

(For length of individual protein-encoding and ribosomal-encoding sequences, see Table 1.)

	EgrG1	EgrG4	Emu	Tcr	EgrG1	EgrG4	Emu	Tcr
	Protein-coding sequences				*rrnL* and *rrnS* sequences			
EgrG1		12·37	14·97	27·01		8·76	11·05	23·73
EgrG4	11·57		13·01	26·37	8·18		11·24	24·47
Emu	13·67	11·53		25·73	11·20	10·24		25·41
Tcr	30·60	30·78	29·58		22·45	22·56	22·25	

parisons with *T. crassiceps* were included (Table 5A). Levels of nucleotide variation were greater in *atp6* (516 positions) than in *nad3*: 69 (19·8%) variant sites among *Echinococcus* species and genotypes and 125 (36%) variant sites when comparisons with *T. crassiceps* were included (Table 5B). Alignments of the predicted amino acid sequences revealed 18 (15·7%) and 36 (21·2%) differences in the *nad3* and *atp6* proteins, respectively, among *Echinococcus* genotypes, and 46 (40%) and 72 (42·4%) respectively between *Echinococcus* and *T. crassiceps* (Table 5A,B). The variation in *nad3* was similar to that reported previously (Bowles *et al.* 1992, 1994; Bowles & McManus, 1993*b*) for fragments of the *nad1* and *cox1* genes among *E. granulosus* genotypes and *E. multilocularis*. However, *atp6* exhibits greater levels of variation and should be useful for discriminating taxa at shallow phylogenetic levels.

Pairwise differences among genes can give a measure of relative levels of divergence among taxa. Such a comparison, of the complete nucleotide sequences of the protein-encoding genes and of the 2 subunits of ribosomal RNA (small; *rrnS* and large; *rrnL*), is shown in Table 6. The *E. granulosus* G1

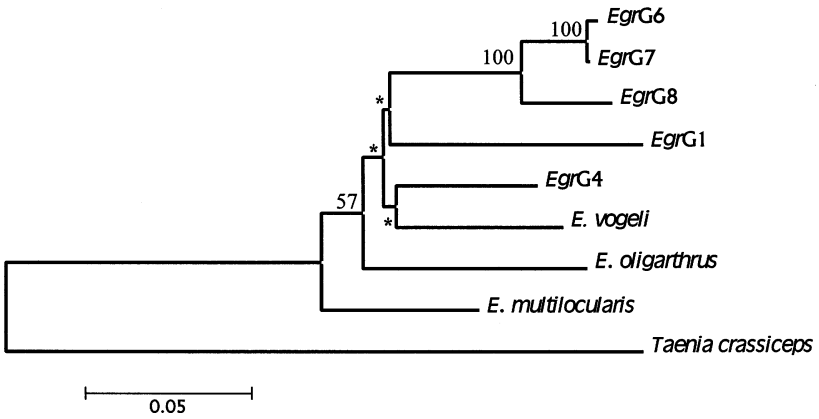

Fig. 6. Inferred relationships among species and genotypes of *Echinococcus*, using *Taenia crassiceps* as an outgroup. Concatenated sequences of *atp6*, *nad1* (partial) and *cox1* (partial) were analysed. A distance matrix was constructed from the inferred amino acid sequences using a Poisson correction for multiple hits and the tree constructed using the minimum evolution approach. Five hundred bootstrap resamplings were carried out. Branches with bootstrap support values less than 50 % are indicated with an asterisk. EgrG1, EgrG4, EgrG6-EgrG8 are the different genotypes of *E. granulosus*. Units on scale bar: changes per site.

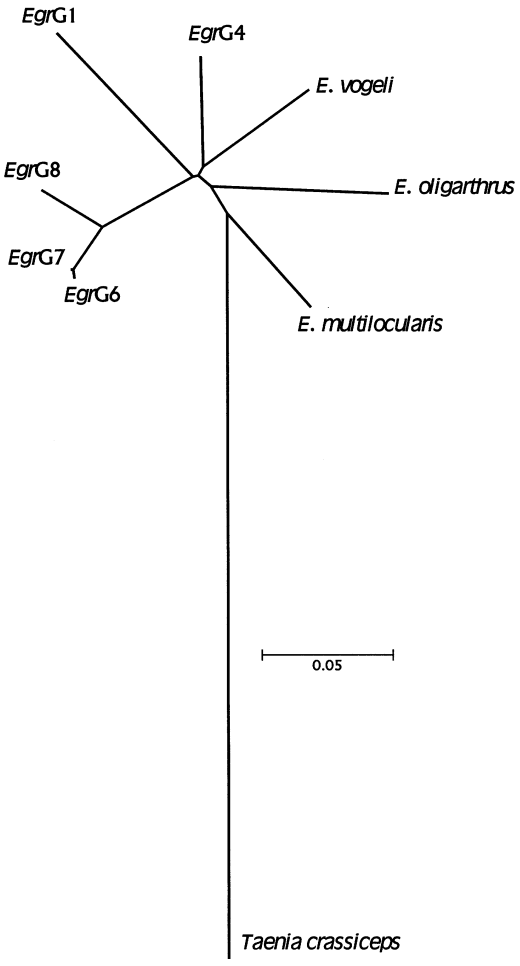

Fig. 7. Inferred relationships among species and genotypes of *Echinococcus* shown as an unrooted tree. Concatenated sequences of *atp6*, *nad1* (partial) and *cox1* (partial) were analysed. A distance matrix was constructed from the nucleotide sequences using the Kimura 2-parameter correction for multiple hits and the tree constructed using the minimum evolution approach. Taxon labels as for Fig. 6. Units on scale bar: changes per site.

genotype differs from the G4 genotype by 12·4 % (nucleotides (nt)), and 11·6 % (amino acids (aa)), a level similar to differences between these two genotypes and *E. multilocularis* (13–15 % nt; and 11·5–13·5 % aa) (Table 6). As expected, divergence is considerably higher when any member of the genus *Echinococcus* is compared with *T. crassiceps* (26–30 % nt and aa differences), suggesting that saturation has not been reached within *Echinococcus*. In both the *rrnL* and *rrnS* genes, the G1 and G4 genotypes of *E. granulosus* differ by 11 % from *E. multilocularis* and differ from each other by 8 % (Table 6). As rRNAs are known to be conserved among related taxa, the differences between *E. granulosus* genotypes is noteworthy. The comparisons reported here suggest that EgrG1 and EgrG4 are as distinct from each other as either is from *E. multilocularis*.

Another approach to investigating levels of divergence is by means of phylogenetic trees. For this, we used nt sequences (complete *atp6*, partial *nad1* (Bowles & McManus, 1993*a*) and partial *cox1* (Bowles *et al*. 1992)) for genotypes 1, 4, 6, 7, 8 (EgrG1, EgrG4, EgrG6, EgrG7 and EgrG8) of *E. granulosus*, *E. multilocularis*, *E. vogeli*, *E. oligarthrus* and *T. crassiceps*. The alignment was 1353 nt (451 aa) long with 543 variable sites (168 for aa) and 262 parsimony-informative sites (67 for aa). The tree in Fig. 6 was constructed from inferred amino acid sequences. Five hundred bootstrap resamplings were conducted. *T. crassiceps* was chosen as the outgroup for rooting the tree. The branches indicated by an asterisk were supported by fewer than 50 % of the resampled data sets and therefore should be regarded as poorly supported. The tree in Fig. 7 was constructed from nucleotide sequences and is presented without an explicit root simply to show more clearly the shortness of the internal branches separating the *Echinococcus* taxa.

It is clear that EgrG4, EgrG1, *E. vogeli* and *E. oligarthrus* are almost equidistant from each other in terms of mt sequences. Furthermore, the *E. granulosus* G1 and G4 genotypes are also almost equidistant from the G6-8 genotype cluster, although there is some structure in this latter group. *E. multilocularis* appears as basal within the genus, but again the branch placing it there is rather poorly supported. Given this, recognition of the sheep-dog (G1 genotype) and the horse-dog (G4 genotype) strains (and possibly also the G6-8 genotypes) as separate species is appropriate. In the case of the sheep and horse strains, a wealth of other strongly supporting information (based on differences in morphological, biological, epidemiological, *in vitro* and *in vivo* developmental and biochemical features) is available (Thompson & Lymbery, 1988; McManus & Bryant, 1995; Thompson, 1995; Thompson & McManus, 2001).

The horse-dog form of *E. granulosus* was recognized as distinct from the common sheep strain and originally promoted as a distinct subspecies, *E. granulosus equinus*, by Williams & Sweatman (1963) based on morphological and host specificity criteria. This classification was rejected by Rausch (1967) because the horse and sheep strains exist sympatrically. However, although the two may be sympatric, their epidemiological patterns and host ranges vary and the form adapted to horses, unlike the sheep form, appears poorly or non-infective to humans (Thompson & Smyth 1975). Despite the opinion of Rausch (1967), therefore, the discrete nature of the 2 forms is quite clear and the molecular and phylogenetic evidence from this and previous studies suggests the case for reinstatement of their formal taxonomic status as subspecies/species is now overwhelming.

This work was supported by grants from the National Health and Medical Research Council of Australia, the Australian Research Council, The Queensland Institute of Medical Research and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR).

REFERENCES

BOORE, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research* **27**, 1767–1780.

BOWLES, J., BLAIR, D. & MCMANUS, D. P. (1992). Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology* **54**, 165–173.

BOWLES, J., BLAIR, D. & MCMANUS, D. P. (1994). Molecular genetic characterization of the cervid strain ('northern form') of *Echinococcus granulosus*. *Parasitology* **109**, 215–221.

BOWLES, J., BLAIR, D. & MCMANUS, D. P. (1995). A molecular phylogeny of the genus *Echinococcus*. *Parasitology* **110**, 317–328.

BOWLES, J. & MCMANUS, D. P. (1993 *a*). Rapid discrimination of *Echinococcus* species and strains using a polymerase chain reaction-based RFLP method. *Molecular and Biochemical Parasitology* **57**, 231–239.

BOWLES, J. & MCMANUS, D. P. (1993 *b*). NADH dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus*. *International Journal for Parasitology* **23**, 969–972.

DE RIJK, P. & DE WACHTER, R. (1997). RnaViz, a program for the visualisation of RNA secondary structure. *Nucleic Acids Research* **25**, 4679–4684.

GAREY, J. R. & WOLSTENHOLME, D. R. (1989). Platyhelminth mitochondrial DNA, evidence for early evolutionary origin of a tRNA(SerAGN) that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. *Journal of Molecular Evolution* **28**, 374–387.

HAAG, K. L., ZAHA, A., ARAUJIA, A. M. & GOTTSTEIN, B. (1997). Reduced genetic variability within coding and non-coding regions of the *Echinococcus multilocularis* genome. *Parasitology* **115**, 521–529.

KEDDIE, E. M., HIGAZI, T. & UNNASCH, T. R. (1998). The mitochondrial genome of *Onchocerca volvulus*, sequence, structure and phylogenetic analysis. *Molecular and Biochemical Parasitology* **95**, 111–127.

KUMAR, S., TAMURA, K., JAKOBSEN, I. B. & NEI, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* (in the Press).

LE, T. H., BLAIR, D., AGATSUMA, T., HUMAIR, P. F., CAMPBELL, N. J. H., IWAGAMI, M., LITTLEWOOD, D. T. J., PEACOCK, B., JOHNSTON, D. A., BARTLEY, J., ROLLINSON, D., HERNIOU, E. A., ZARLENGA, D. S. & MCMANUS, D. P. (2000). Phylogenies inferred from mitochondrial gene orders – a cautionary tale from the parasitic flatworms. *Molecular Biology and Evolution* **17**, 1123–1125.

LE, T. H., BLAIR, D. & MCMANUS, D. P. (2000 *a*). Mitochondrial genomes of human helminths and their use as markers in population genetics and phylogeny. *Acta Tropica* **77**, 243–256.

LE, T. H., BLAIR, D. & MCMANUS, D. P. (2000 *b*). Mitochondrial DNA sequences of human schistosomes, the current status. *International Journal for Parasitology* **30**, 283–290.

LE, T. H., BLAIR, D. & MCMANUS, D. P. (2001). Complete DNA sequence and gene organization of the mitochondrial genome of the liver fluke, *Fasciola hepatica* L. (Platyhelminthes; Trematoda) *Parasitology* **123**, 609–621.

LE, T. H., HUMAIR, P. F., BLAIR, D., AGATASUMA, T. & MCMANUS, D. P. (2001). Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. *Molecular and Biochemical Parasitology* **117**, 61–71.

LOWE, T. & EDDY, S. R. (1997). tRNAscan-SE: a program improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955–964.

MCMANUS, D. P. & BRYANT, C. A. (1995). Biochemistry, physiology and molecular biology of *Echinococcus*. In *The Biology of Echinococcus and Hydatid Disease* (ed. Thompson, R. C. A. & Lymbery, A. J.), pp. 135–181. CAB International, Wallingford, UK.

MORGAN, J. A. T. & BLAIR, D. (1998). Relative merits of nuclear ribosomal internal transcribed spacers and

mitochondrial COI and NDI genes for distinguishing among *Echinostoma* species (Trematoda). *Parasitology* **116**, 289–297.

NAKAO, M., SAKO, Y., YOKOYAMA, N., FUKUNAGA, M. & ITO, A. (2000). Mitochondrial genetic code in cestodes. *Molecular and Biochemical Parasitology* **11**, 415–424.

OKIMOTO, R., MACFARLANE, J. L., CLARY, D. O. & WOLSTENHOLME, D. R. (1992). The mitochondrial genome of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130**, 471–498.

RAUSCH, R. L. (1967). A consideration of intraspecific categories in the genus *Echinococcus* Rudolphi, 1801 (Cestoda: Taeniidae). *Journal of Parasitology* **53**, 484–491.

RINDER, H., RAUSCH, R. L., TAKAHASHI, K., KOPP, H., THOMSCHE, A. & LOSCHER, T. (1997). Limited range of genetic variation in *Echinococcus multilocularis*. *Journal of Parasitology* **83**, 1045–1050.

SCOTT, J. C. & MCMANUS, D. P. (1994). The random amplification of polymorphic DNA can discriminate species and strains of *Echinococcus*. *Tropical Medicine and Parasitology* **45**, 1–4.

SCOTT, J. C., STEFANIAK, J., PAWLOWSKI, Z. S. & MCMANUS, D. P. (1997). Molecular genetic analysis of human cystic hydatid cases from Poland, identification of a new genotypic group (G9) of *Echinococcus granulosus*. *Parasitology* **114**, 37–43.

TELFORD, M. J., HERNIOU, E. A., RUSSELL, R. B. & LITTLEWOOD, D. T. J. (2000). Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proceedings of the National Academy of Sciences, USA* **97**, 11359–11364.

THOMPSON, R. C. A. (1995). Biology and systematics of *Echinococcus*. In *The Biology of Echinococcus and Hydatid Disease* (ed. Thompson, R. C. A. & Lymbery, A. J.), pp. 1–50. CAB International, Wallingford, Oxon, UK.

THOMPSON, R. C. A. & LYMBERY, A. J. (1988). The nature, extent and significance of variation within the genus *Echinococcus*. *Advances in Parasitology* **27**, 209–258.

THOMPSON, R. C. A., LYMBERY, A. J. & CONSTANTINE, C. C. (1995). Variation in *Echinococcus*, towards a taxonomic revision of the genus. *Advances in Parasitology* **35**, 145–176.

THOMPSON, R. C. A. & MCMANUS, D. P. (2001). Aetiology: parasites and life cycles. WHO/OIE *Manual on Echinococcosis in Humans and Animals* (ed. Eckert, J., Gemmell, M. A., Meslin, F.-X. & Pawlowski, Z. S.), pp. 1–19. CAB International, Wallingford, Oxon, UK.

THOMPSON, R. C. A. & SMYTH, J. D. (1975). Equine hydatidosis: a review of the current status in Great Britain and the results of an epidemiological survey. *Veterinary Parasitology* **1**, 107–127.

VON NICKISCH-ROSENEGK, M., BROWN, W. M. & BOORE, J. L. (2001). Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: gene arrangements indicate that platyhelminths are eutrochozoans. *Molecular Biology and Evolution* **18**, 721–730.

WILLIAMS, R. J. & SWEATMAN, G. K. (1963). On the transmission, biology and morphology of *Echinococcus granulosus equinus*, a new subspecies of hydatid tapeworm in horses in Great Britain. *Parasitology* **53**, 391–407.

WOLSTENHOLME, D. R. (1992). Animal mitochondrial DNA, structure and evolution. *International Reviews of Cytology* **141**, 173–216.

ZARLENGA, D. S. & GEORGE, M. (1995). *Taenia crassiceps*: cloning and mapping of mitochondrial DNA and its application to the phenetic analysis of a new species of *Taenia* from Southeast Asia. *Experimental Parasitology* **81**, 604–607.