Relative merits of nuclear ribosomal internal transcribed spacers and mitochondrial CO1 and ND1 genes for distinguishing among *Echinostoma* species (Trematoda)

J. A. T. MORGAN* *and* D. BLAIR

Department of Zoology and Tropical Ecology, James Cook University, Townsville, Queensland, Australia 4811

(*Received 23 July 1997; revised 26 September 1997; accepted 26 September 1997*)

SUMMARY

Cryptic species, belonging to the 37 collar-spine *Echinostoma* group, were distinguished using nuclear rDNA ITS (884 bases) and mtDNA CO1 (257 bases) and ND1 (530 bases) sequences. Sequences were obtained from five 37 collar-spine species, *Echinostoma trivolvis*, *E. paraensei*, *E. caproni*, *E. revolutum* and *E.* sp.I, a parthenogenetic isolate from Africa. Three geographic isolates of *E. caproni* were compared. Average sequence divergence among the 37 collar-spine species range from 2·2% in the rDNA ITS through 8% for the CO1 and 14% for the ND1. In addition, genes were sequenced from 2 non 37 collar-spine species, *E. hortense* and an undescribed Australian species, *E.* sp. (Aus). For each gene, distances of terminals from a predicted ancestral sequence were calculated. These indicated that ND1 is diverging significantly faster than the other 2 regions. In the CO1 gene most substitutions are synonymous and saturation has been reached for the majority of pairwise comparisons. The ND1 gene exhibits greater pairwise divergence but less evidence of saturation due to weaker conservation of first and second codon positions. The ITS has no amino acid coding constraints and displays no evidence of saturation. Although all 3 regions successfully distinguished the nominal species, ND1 appears to be the most informative region for investigating relationships within the 37 collar-spine group.

Key words: *Echinostoma*, ribosomal DNA, internal transcribed spacer, mitochondrial DNA, cytochrome *c* oxidase subunit 1 (CO1), nicotinamide adenine dinucleotide dehydrogenase subunit 1 (ND1).

INTRODUCTION

Molecular genetic characterization is being used increasingly to distinguish among morphologically similar parasites. PCR technology and DNA sequencing techniques permit the identification of species, strains and populations from a small quantity of tissue from any stage in their life-history. Very little information is available comparing the relative divergence rates of nuclear versus mitochondrial DNA and what is known may not be applicable to trematodes.

Trematodes of the family Echinostomatidae are abundant and cosmopolitan parasites that generally cycle through snails and aquatic vertebrates, usually birds. The genus *Echinostoma* includes nominal species and species groups characterized by the number of collar spines surrounding the oral sucker. The 37 collar-spine group contains a number of cryptic species which have proved difficult to identify in the past (Kanev, 1985). In a previous study, 6 species of *Echinostoma*, originally described using non-molecular criteria, were distinguished using sequence data from the nuclear rDNA internal transcribed spacers (ITS1 and 2) (Morgan & Blair, 1995). The spacers were found to be highly conserved within species and displayed between 1·1 and 19·2% sequence divergence between species.

Here we report the use of mitochondrial DNA to distinguish among the same nominal *Echinostoma* species belonging to the 37 collar-spine group, *E. trivolvis* (Cort, 1914), *E. revolutum* (Frölich, 1802), *E. caproni* (Richard, 1964), *E. paraensei* (Lie & Basch, 1967), and an undescribed African isolate, *E.* sp.I, currently maintained by J. Jourdane, Perpignan, France. Four specimens, representing 3 geographic isolates of *E. caproni*, were sequenced, *E. caproni* (c) (Madagascar), *E. caproni* (l) (Egypt) and *E. caproni* (k) (Cameroon). In the past these strains were treated as species and labelled *E. caproni*, *E. liei* and *E.* sp.II respectively by Morgan & Blair (1995). Despite their ability to cross-fertilize, Trouvé *et al.* (1996) distinguished representatives of each of these isolates of *E. caproni* electrophoretically using diagnostic loci. Sequence data from both mitochondrial cytochrome *c* oxidase subunit 1 (CO1) and nicotinamide adenine dinucleotide dehydrogenase subunit 1 (ND1) genes were used to investigate phylogenetic relationships for comparison with those defined previously using nuclear ribosomal DNA sequences (Morgan & Blair, 1995). The relative merits of each DNA region for investigating phylogenetic relationships is assessed by comparing nu-

* Corresponding author: Tel: +077 814242. Fax: +077 251570. E-mail jessica.morgan@jcu.edu.au.

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

cleotide composition, transition bias, saturation levels and divergence rates.

MATERIALS AND METHODS

Sample collection, DNA extraction and amplification

Echinostomes were obtained from either naturally infected hosts or laboratory strains originating from various countries (Table 1). In addition 2 non 37 collar-spined species were included, *E. hortense* (Asada, 1926), which has 28 collar spines, and *E.* sp. (Aus), which has more than 40 collar spines. These 2 species acted as outgroups for phylogenetic analyses and for estimates of DNA divergence levels. DNA was extracted from the flukes using standard, previously described, techniques (Morgan & Blair, 1995).

Mitochondrial genes were amplified by PCR using *Taq* polymerase and primers designed from *Fasciola hepatica* sequence (Garey & Wolstenholme, 1989) as reported by Bowles (1993). Primers used to generate the partial ND1 fragment were forward, JB11 (5′AGATTCGTAAGGGGCCTAATA3′) and reverse, JB12 (5′ACCACTAACTAATTCACTTTC3′). The CO1 region utilized the forward primer, JB3 (5′TTTTTTGGGCATCCTGAGGTTTAT3′) and the reverse primer, JB13 (5′TCATGAAAACACCTTAATACC3′).

Cloning and sequencing

PCR products were ligated into a TA cloning vector and clones were generated by transforming INVαF′ *E. coli* cells provided in Invitrogen's TA cloning kit. Positive recombinants were purified using Wizard Minipreps (Promega) and double stranded sequencing was completed using Sequenase Version 2.0 (USB) and [α-^{32}P]-dATP (Bresatec).

Sequencing primers included the universal primers USP and RSP in addition to the 4 PCR primers. At least 2 clones were sequenced per isolate with additional clones sequenced as necessary to resolve ambiguous sites. Mitochondrial sequences were aligned by eye against published *F. hepatica* sequences (Garey & Wolstenholme, 1989). Translation of sequences into amino acids used a modified universal code (Table 2). Levels of intraspecific variation were assessed by sequencing 2 isolates of *E. trivolvis*, and 4 isolates of *E. caproni*, see Table 1.

The ITS2 sequences used were those published previously by Morgan & Blair (1995).

Phylogenetic analysis

Trees were generated for each coding region using all sites and again with bases at the third codon position removed to assess the effect of silent mutations. Additional trees were produced from inferred amino acid sequences. The amount of phylogenetic signal within each data set was tested by measuring the degree of skewness, measured with the g1 statistic, from length frequency distributions of 1000 random trees (PAUP – Swofford, 1993).

Maximum parsimony (PAUP – Swofford, 1993), maximum likelihood (PHYLIP – Felsenstein, 1989) and distance matrix (TREECON – Van de Peer & De Wachter, 1993) analyses were applied to the data. *F. hepatica* mitochondrial sequence was used to root trees for all methods. Maximum parsimony trees were found using branch-and-bound searches.

Use of different distance estimators (Kimura 2-parameter, Tamura and Tamura-Nei) had no effect on tree topology. For simplicity the Kimura 2-parameter algorithm was adopted for all analyses. Distance matrices were converted to trees using Saitou & Nei's (1987) neighbour joining technique in TREECON. Manipulating the transition:transversion ratio (TS:TV) settings from 2 through to 10 had no effect on tree topology, so for maximum likelihood predictions this value was fixed at 2 and empirical base frequencies were used to calculate probabilities. Maximum likelihood predictions assuming a molecular clock were calculated to enable a likelihood ratio test of the molecular clock hypothesis (DNAML and DNAMLK in PHYLIP – Felsenstein, 1989). Branches were tested for all inferred trees using bootstrap analysis on 1000 random trees.

DNA saturation and genetic distances

Nucleotide frequencies at each codon position were compared for both mitochondrial genes using a χ^2 homogeneity test to determine if particular bases were favoured at specific positions.

Saturation occurs when there are multiple substitutions at the same site. When a random sequence (base frequency 1:1:1:1) reaches saturation the observed number of mutations approaches a stable ratio of 2 transversions/transition. Nucleotide composition influences this ratio which can be estimated using the following equation modified from Holmquist (1983).

Equation 1.

$$\text{Saturation asymptote} = \frac{(AxG)+(CxT)}{(A+G)(C+T)}.$$

Levels of saturation were assessed by plotting the TS:TV ratio for each pairwise comparison against its corresponding Kimura 2-parameter distance (Taylor, Hebert & Colbourne, 1996). Saturation levels were observed over an extended range of genetic distances by including published sequences for *F. hepatica* (Garey & Wolstenholme, 1989), *Schistosoma mansoni* CO1 (Bowles, Blair & McManus, 1995) and both CO1 and ND1 from *Echinococcus granulosus* (Bowles, Blair & McManus, 1992; Bowles & McManus, 1993) in the analysis. In addition the pairwise number of transitions and transversions among all taxa were plotted against

Table 1. Species list indicating country of origin and contributor for sequenced specimens and GENBANK numbers for species used in comparisons

Species	Originated from	Maintained and provided by
Echinostoma trivolvis	North America	J. E. Huffman/D. W. Hosier – USA
E. trivolvis	North America	P. M. Nollen – USA
E. caproni (c)	Madagascar	B. Fried/J. E. Huffman – USA
E. caproni (c)	Madagascar	N. O. Christensen – Denmark
E. caproni (l)	Egypt	J. Jourdane – France = *E. liei* in Morgan & Blair (1995)
E. caproni (k)	Cameroon	J. Jourdane – France = *E.* sp.II in Morgan & Blair (1995)
E. paraensei	Brazil	L. A. Hertel – USA
E. revolutum	Germany	I. Kanev – Bulgaria
E. sp.I.	Niger	J. Jourdane – France
E. hortense	East Asia	S. H. Lee – Korea
E. sp. (Aus)	Northern Australia	J. A. T. Morgan – Australia
Fasciola hepatica	Garey & Wolstenholme (1989)	GENBANK Accession #X15613
Schistosoma mansoni	Bowles *et al.* (1995)	GENBANK Accession #U22162
Echinococcus granulosus	Bowles *et al.* (1992)	GENBANK Accession #M84661
Echinococcus granulosus	Bowles & McManus (1993)	GENBANK Accession #S70921

Table 2. Modifications to the mitochondrial code for amino acid translation in flatworms

Codon	Mitochondrial code	Flatworm code	Reference
TGA	STOP	Tryptophan	Garey & Wolstenholme (1989)
AGA	Arginine	Serine	Garey & Wolstenholme (1989)
AGG	Arginine	Serine	Garey & Wolstenholme (1989)
AAA	Lysine	Asparagine	Ohama *et al.* (1990)
ATA	Methionine	Isoleucine	Bowles *et al.* (1992)

their Kimura 2-parameter distances for each gene region (Nunn & Cracraft, 1996). The number of transitions and transversions, TS:TV ratios and Kimura 2-parameter pairwise distances were calculated using MEGA 1.02 (Kumar, Tamura & Nei, 1993).

The effects of gene (term loosely applied to include ITS) and species on sequence divergence from an ancestral sequence were measured and tested using a factorial ANOVA model in SPSS 6.1 for Windows (1989). The ancestral sequence was inferred using parsimony (PAUP – Swofford, 1993) for each gene. Pairwise distances between each species and the ancestral sequence were calculated using the Kimura 2-parameter algorithm. A natural log (ln) transformation was required for the multiplicative ANOVA model. The interaction value was used as an error term. ANOVA assumptions of normality, homoscedasticity and additivity were considered.

RESULTS

Primary sequence alignments from 257 bases of CO1 (Fig. 1) and 530 bases of ND1 (Fig. 2) include *F. hepatica*. The 7 nominal species and 2 of the 3 *E. caproni* isolates were distinguished using sequences from the mtDNA genes. *Echinostoma caproni* (c) and *E. caproni* (l) could not be distinguished from each other using either region and are regarded as identical for the analysis. Sequences align to positions 311–840 (ND1) and 2575–2831 (CO1) in the *F. hepatica* mtDNA molecule sequence, published by Garey & Wolstenholme (1989). The combined mitochondrial data set consists of 787 bases with 295 variable sites, 154 being informative for parsimony. Within the 37 collar-spined group there were 183 variable sites, of which 94 sites were informative.

Base composition was heavily A–T biased, 64 % across both ND1 and CO1. The most common base was T, 45 %, and the least common was C, 10 %. No deletions or insertions were required to align either gene. Intra-individual differences between clones were not detected. No intraspecific variability was observed between the 2 *E. trivolvis* or *E. caproni* (c) isolates.

Sequence and amino acid comparisons (Table 3) indicate that greatest differentiation occurs in the ND1 gene. The *E. caproni* strains displayed the lowest level of variation observed. At most they differed by 2 nucleotides in the CO1 (0·8 %) and 13 nucleotides in the ND1 (2·5 %). *Echinostoma* species ranged from 6·3 to 14·1 % sequence divergence in the

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

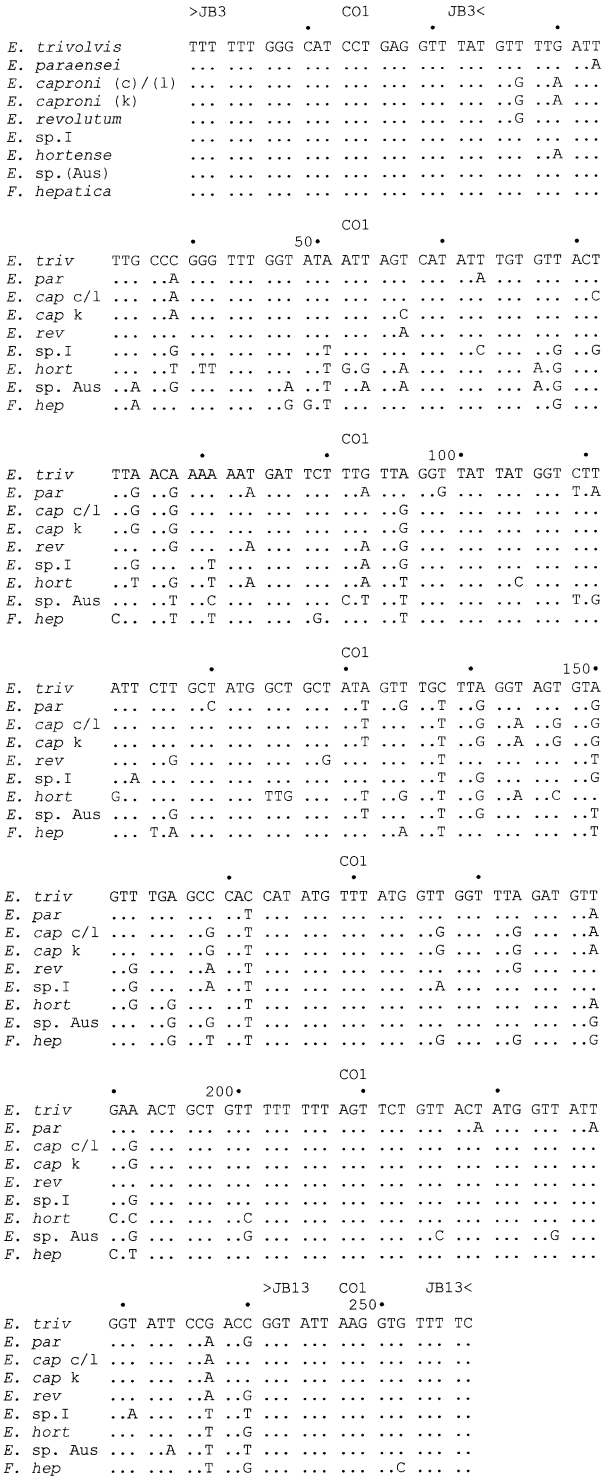

Fig. 1. Partial mtDNA cytochrome subunit 1 (CO1) sequences for 7 echinostome species showing alignment with *Fasciola hepatica*. A small dot (·) indicates identity with *Echinostoma trivolvis*. A large dot (●) marks every tenth base. Primer positions are marked on the alignment.

CO1 and 12·3 to 30·8% in the ND1. *Echinostoma hortense* displayed the greatest divergence across both genes.

Despite differences in their primary DNA sequences, all of the 37 collar-spine species translated identical CO1 amino acid sequences. The remaining species differed by between 2 and 8 amino acids of the 85 translated for the CO1. Changes in amino acid composition occurred among all species for the ND1. *Echinostoma caproni* (c)/*E. caproni* (l) diverged from *E. caproni* (k) at 4 amino acids out of the 176 translated. *Echinostoma hortense* displayed greater amino acid divergence from the remaining echinostomes than did *F. hepatica*.

Phylogenetic analysis

Significant cladistic structure was detected in both data sets with the ND1 gene showing greater left skew (CO1 $g1 = -0{\cdot}58$, $P = 0{\cdot}000$, ND1 $g1 = -0{\cdot}88$, $P = 0{\cdot}000$). Some of this difference can be attributed to the difference in the number of sites between the 2 data sets.

Maximum parsimony analyses of the CO1 data produced 3 trees, each 124 steps long, with a consistency index of 0·734. The strict consensus parsimony tree was congruent with both distance matrix and maximum likelihood trees. The CO1 consensus tree is not shown as it displayed the same broad groupings as the ND1 tree (see below). In the CO1 tree, the 37 collar-spine species formed a group with *E.* sp.I at its base. There was a close relationship among the *E. caproni* isolates. No other branches were resolved among the 37 collar-spine species. Differing topology between CO1 trees generated by maximum likelihood (DNAML) and maximum likelihood with a molecular clock (DNAMLK) prevented testing for a molecular clock. Trees generated by removing third codon positions, using transversions only, or amino acid sequences, did not resolve relationships any further due to the conservative nature of CO1.

The ND1 tree was able to resolve more branches than the rDNA ITS (Morgan & Blair, 1995) and CO1 trees. ND1 sequences produced 2 trees using maximum parsimony each 491 steps long with consistency index 0·719. The trees differed only in their positioning of *E. revolutum*. The tree in Fig. 3 shows the parsimony topology identical to that obtained using maximum likelihood. Distance matrix methods also support this position for *E. revolutum*, however, they fail to cluster *E. hortense* with *E.* sp. (Aus). Maximum likelihood found all branches were significantly positive. The molecular clock hypothesis was rejected using the likelihood ratio test ($\chi^2 = 24{\cdot}2$, $P < 0{\cdot}001$). Identical tree topology was obtained when only transversions were analysed. Removing the third codon position and comparing amino acid sequences produced less informative trees with a collapsed internal branch marked with an asterisk in Fig. 3.

DNA saturation and genetic distances

Strong nucleotide bias in the mtDNA, favouring T against C, had a marked effect on saturation levels.

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

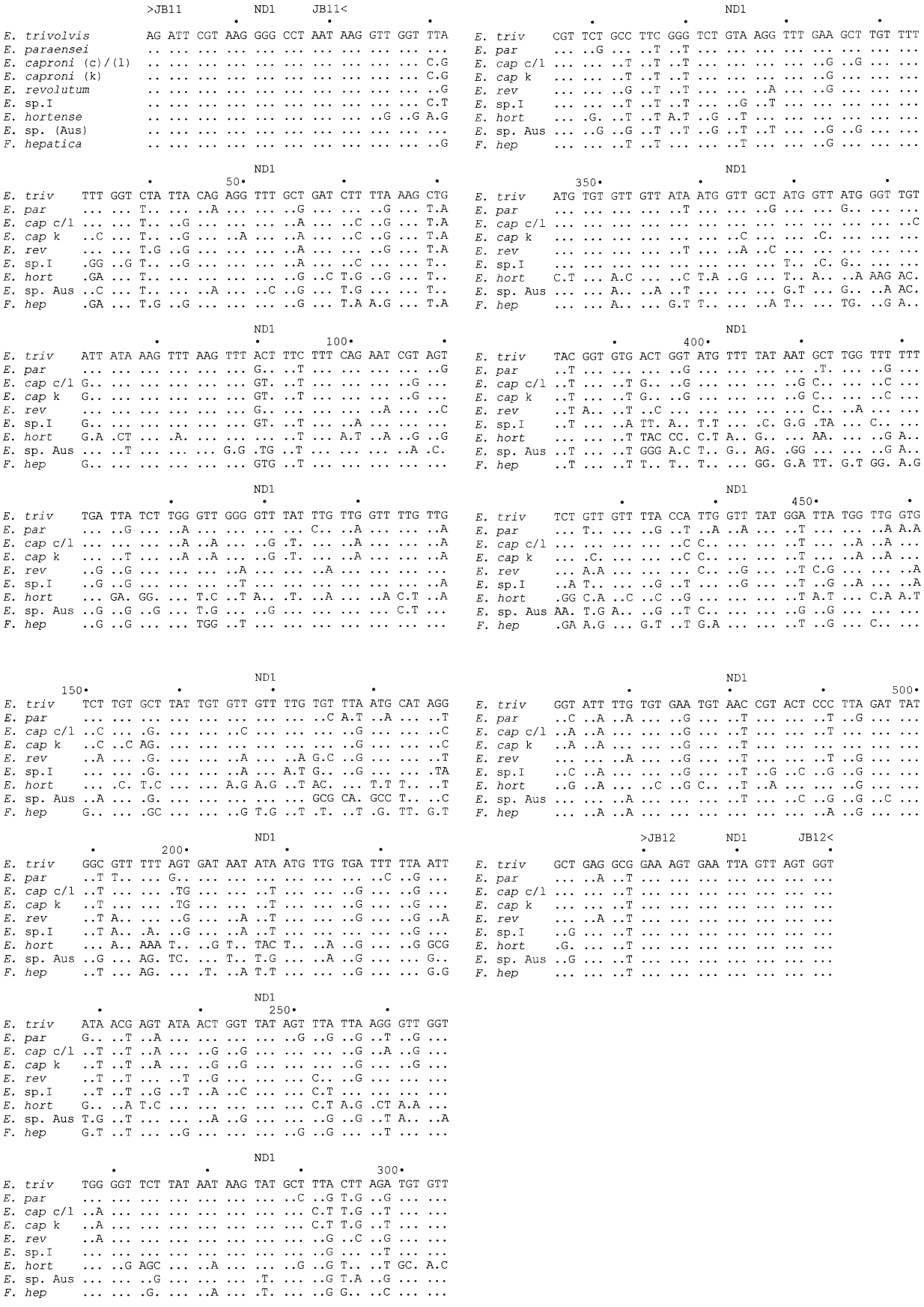

Fig. 2. Partial mtDNA nicotinamide adenine dinucleotide dehydrogenase subunit 1 (ND1) sequences for 7 echinostome species showing alignment with *Fasciola hepatica*. A small dot (·) indicates identity with *Echinostoma trivolvis*. A large dot (●) marks every tenth base. Primer positions are marked on the alignment.

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

Table 3. Percentage primary sequence/percentage amino acid pairwise comparisons for CO1 above the diagonal and ND1 below the diagonal

Species	*E. triv*	*E. par*	*E. cap* c/l	*E. cap* k	*E. rev*	*E.* sp.I.	*E.* sp. (Aus)	*E. hort*	*F. hep*
E. trivolvis	—	8·6/0·0	7·8/0·0	7·8/0·0	6·3/0·0	8·2/0·0	12·2/2·4	14·1/9·4	10·2/4·7
E. paraensei	12·3/6·3	—	8·6/0·0	8·6/0·0	9·0/0·0	10·6/0·0	13·7/2·4	13·7/9·4	14·5/4·7
E. caproni (c)/(l)	13·1/4·6	13·0/9·1	—	0·8/0·0	7·8/0·0	8·6/0·0	12·9/2·4	13·3/9·4	12·2/4·7
E. caproni (k)	13·6/6·3	14·0/10·2	2·5/2·3	—	7·5/0·0	9·0/0·0	12·6/2·4	12·9/9·4	12·2/4·7
E. revolutum	13·3/4·0	13·8/7·4	12·9/5·7	13·6/7·4	—	8·6/0·0	12·2/2·4	13·3/9·4	10·6/4·7
E. sp.I.	15·9/11·4	16·7/12·0	15·0/10·8	15·2/11·4	14·6/10·2	—	11·4/2·4	13·3/9·4	11·4/4·7
E. sp. (Aus)	21·5.17·6	21·5/18·8	22·1/19·3	22·8/21·0	22·6/18·8	22·8/18·8	—	14·1/7·0	11·4/4·7
E. hortense	29·4/34·1	29·2/33·5	30·1/34·1	30·5/33·5	30·5/34·7	30·1/33·0	30·8/33·0	—	14·8/9·4
F. hepatica	20·6/23·9	21·0/25·0	21·6/23·3	22·4/25·0	19·7/24·4	19·7/21·6	22·3/23·9	28·6/35·8	—

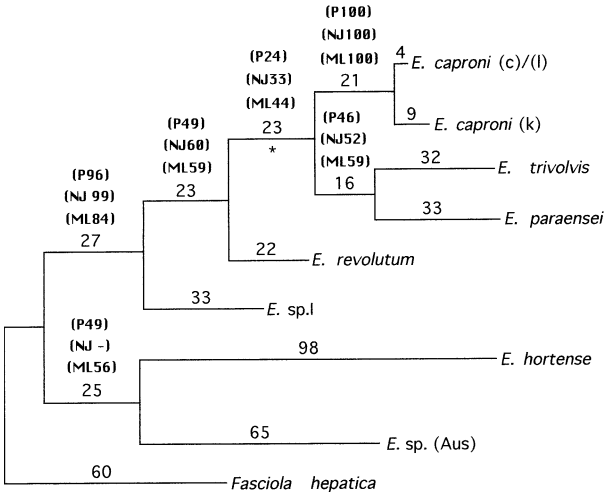

Fig. 3. One of the 2 equally parsimonious trees inferred from partial ND1 sequence data for nominal echinostome species using *Fasciola hepatica* as an outgroup. Identical topology was obtained using maximum likelihood analyses. Bootstrap values for each method of analysis (Parsimony, Neighbour Joining and Maximum Likelihood) appear in brackets above branch lengths. An asterisk indicates the branch that collapses when trees are inferred from either amino acid data or from sequences with third codon position removed. Assigned branch lengths were those calculated in PAUP (Swofford, 1993).

The saturation asymptotes calculated from observed nucleotide frequencies (see Eqn 1) are 0·35 for ND1 and 0·38 for CO1. Codon position had a significant effect on base frequency (CO1 $\chi^2 = 349$, D.F. = 6, $P = 0{\cdot}00$; ND1 $\chi^2 = 128$, D.F. = 6, $P = 0{\cdot}00$) with T favoured over C at the third position. Saturation levels were further influenced as the majority of variable sites were in this position. Recalculating saturation asymptotes using third codon positions only, lowers them to 0·26 for ND1 and 0·21 for CO1.

The level of saturation observed for each DNA region was assessed by plotting the TS:TV ratio (Fig. 4A, B, C) and the actual number of transitions or transversions (Fig. 4D, E, F) against pairwise genetic distances. The TS:TV ratio shows a general decrease as genetic distances increase. Compared to the other gene regions, most of the TS:TV ratio values for CO1 are positioned near the saturation asymptote (Fig. 4A). Transversions outnumber transitions for the majority of comparisons (Fig. 4D). ND1 spans the widest range of genetic distances with distant comparisons approaching saturation (Fig. 4B and E). None of the ITS comparisons reached saturation (Fig. 4C) with transitions consistently more common than transversions (Fig. 4F).

Gene region has a significant effect on the average Kimura 2-parameter distances from their respective ancestral sequence ($F = 32{\cdot}976$, $P \ll 0{\cdot}05$). These distance measures support the increasing rate of divergence from ITS through CO1 to ND1 (Fig. 5). An apparent anomaly is that the mean values for CO1 and ITS are similar despite their percentage sequence divergence differing by 4-fold. This effect might be due to saturation in the CO1 gene leading to incorrect inference of variable sites in the ancestral sequence. Individual species did not affect the average distance from the ancestral sequence ($F = 0{\cdot}723$, $P = 0{\cdot}656$). Although the data did not conform to a normal distribution, the plot of studentized residuals displayed roughly equal variances. ANOVAs are known to be robust to considerable deviations from their underlying assumptions of normality and heteroscedasticity (Zar, 1984).

DISCUSSION

Mitochondrial DNA sequencing has successfully distinguished all of the echinostome species identified *a priori* using non-molecular criteria. Phylogenetic analyses largely agree with previous results using nuclear rDNA internal transcribed spacers. No intraspecific variation was observed between either of the 2 *E. caproni* (c) or *E. trivolvis* isolates. These isolates had been maintained in distinct laboratories, however, their specific collection locations are not known. *Echinostoma caproni* (c) and *E. caproni* (l) were completely identical across both mitochondrial genes suggesting that these strains share a closer relationship to each other than either does to *E.*

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

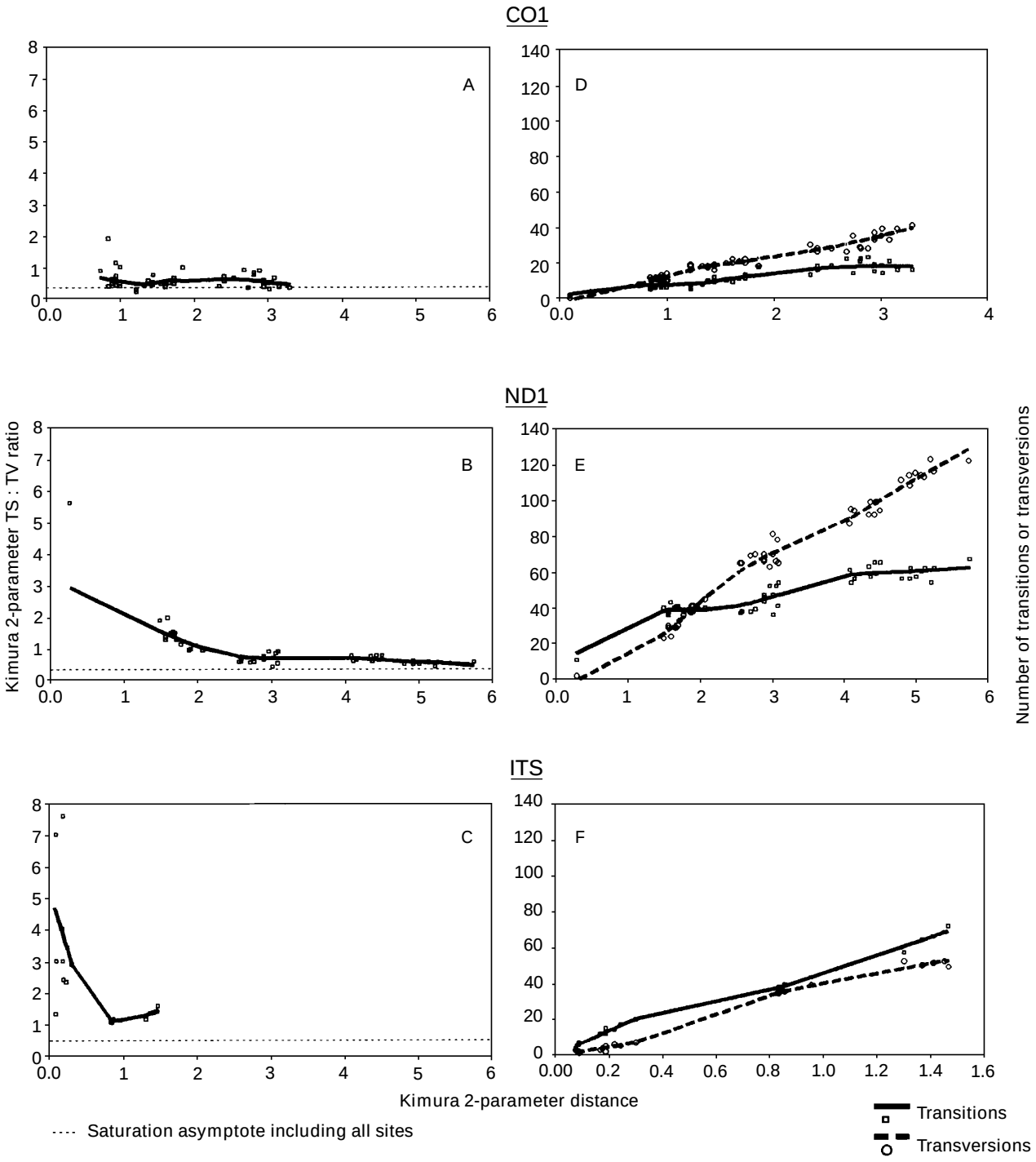

Fig. 4. Saturation plots for each gene region (CO1, ND1 and ITS) displaying the TS:TV ratio against Kimura 2-parameter distance in A, B and C and the number of transitions or transversions against Kimura 2-parameter distance in D, E and F. The saturation asymptote for each gene is indicated with a dotted line in A, B and C. Lines of best fit have been drawn to display the pattern of increase in transitions versus transversions in D, E and F. Distances were calculated using MEGA. In addition to the echinostomes, CO1 plots (A, D) include pairwise comparisons among *Fasciola hepatica*, *Schistosoma mansoni* and *Echinococcus granulosus*. ND1 plots (B, E) include *F. hepatica* and *E. granulosus*.

caproni (k) which diverged by 0·8 % in the CO1 and 2·5 % in the ND1. The lowest divergence detected between recognized species was 5-fold greater than these values.

E. revolutum was the first 37 collar-spine echinostome to be re-assessed using detailed morphological descriptors. Identification of this species relies on a single character, the number of paraoesophageal

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

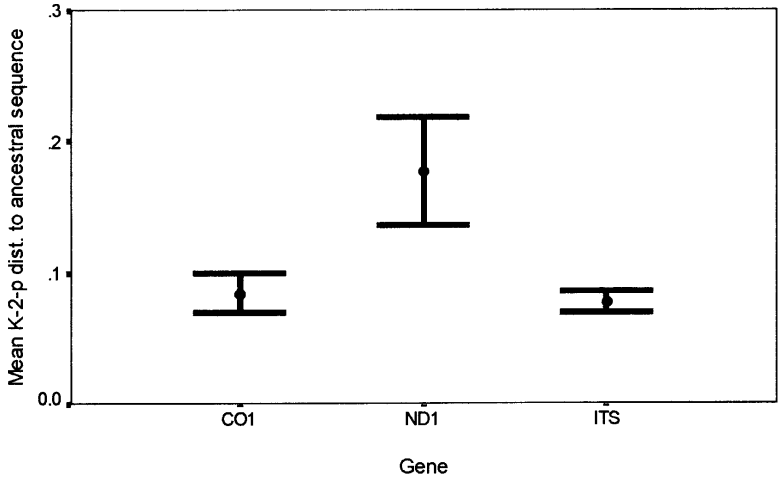

Fig. 5. Plot illustrating the significant effect of gene on the mean Kimura 2-parameter distance from the inferred ancestral sequence to the terminal taxa. The ancestral sequence was inferred using parsimony. No species effect was detected. The 95% confidence interval around each mean, based on a sample size of 8 for each gene, is shown.

glands in the oral region of the cercaria (Kanev, 1994). Visualization of this feature involves staining large numbers of cercariae and often produces ambiguous results. The mtDNA ND1 sequence of *E. revolutum* differs from its nearest relative, *E. caproni*, at 68 sites. Many of these variable positions are shared with other species. When all of the ND1 sequences are aligned, *E. revolutum* can be distinguished from every other species using any 1 of 15 unique sites. Using molecular techniques a single specimen at any stage in its life-cycle can now be accurately identified using small amounts of tissue.

Among the 37 collar-spine species the average CO1 sequence divergence was 8% compared to 14% in the ND1. These means fall well within the ranges distinguishing other parasites including species in the genera *Taenia* (CO1 between 6·3 and 18·2% (Bowles & McManus, 1994; Okamoto *et al.* 1995)), *Echinococcus* (CO1 6 and 11·5% and ND1 12·7 and 16·3% (Bowles & McManus, 1993)) and *Schistosoma* (CO1 8 and 21% (Bowles *et al.* 1995)).

Amino acid conservation was maintained through the CO1 with all nucleotide changes being silent among the 37 collar-spine echinostomes. The ND1 gene was less conserved with up to 12% amino acid divergence among the 37 collar-spine group.

Phylogenetic results indicate that the 37 collar-spine group form a monophyletic sister group to the 2 non 37 collar-spine species. Although bootstrap values are weak they support the notion that the American species (*E. trivolvis* and *E. paraensei*) are sister taxa relative to the *E. caproni* strains from Madagascar, Egypt and Cameroon. Within the 37 collar-spine group *E.* sp.I, remains the most diverged. This result is supported by rDNA ITS data (Morgan & Blair, 1995).

The observed mtDNA thymine bias (45%) is similar to that found in other trematodes (Garey & Wolstenholme, 1989) Codons rarely end with cytosine and if present they are never conserved across all species. The echinostomes differed from *F. hepatica* by displaying an A/T bias rather than G/T. An A/T bias is also seen in insects (Sperling & Hickey, 1994).

Despite the low divergence rate of CO1, it has reached saturation before either of the other regions. Saturation has been reached even within the 37 collar-spine group. High gene conservation is restricting variable sites to third codon positions, where transversions are quickly masking transitions. This conservation is reflected in the small pairwise distances and low TS:TV ratio. Divergence levels may be underestimated for CO1 due to saturation at these variable sites.

ND1 is diverging at a rapid rate and displaying a wider range of genetic distances that asymptote at saturation for distant comparisons outside of the 37-collar spine group. Compared to the CO1 a greater proportion of variable sites occurs at first and second codon positions resulting in saturation not being reached as quickly. The TS:TV ratio is low for all pairwise species comparisons but remains high among the *E. caproni* strains. These closely related strains have only recently diverged and transitions still outnumber transversions.

The ITS is evolving relatively slowly. With no amino acid coding constraints a larger proportion of sites is presumably available to mutate. This explains the observed transition bias with no evidence of saturation.

CO1 sequences were useful for distinguishing the nominal species despite their limited phylogenetic value. Of the 3 gene regions sequenced, ND1 appears to be the most informative for investigating relationships within the 37 collar-spine group. Although strain variability was detected in both the CO1 and ND1 genes, neither was capable of distinguishing *E. caproni* (c) from *E. caproni* (l). Future population studies on this species should concentrate on faster evolving regions of DNA.

High levels of saturation in the mitochondrial

https://doi.org/10.1017/S0031182097002217 Published online by Cambridge University Press

genes indicate that they should not be used to estimate relationships among distantly related species of trematodes. The rDNA ITS is better suited to these comparisons. However, accurate sequence alignments for distant species are more difficult to obtain for non-coding DNA.

We are indebted to N. O. Christensen, B. Fried, L. A. Hertel, D. W. Hosier, J. E. Huffman, J. Jourdane, I. Kanev, S. H. Lee and P. M. Nollen for providing specimens for this study. Special thanks go to John Morgan for his help with the statistical analyses used in this article.

REFERENCES

ASADA, J. (1926). Über eine neue Species von *Echinochasmus* und ihren Entwicklungsgang. [Abstract]. *Transactions of the Japanese Pathological Society* **16**, 113.

BOWLES, J. (1993). Genetic approaches to the identification and evolutionary study of helminth parasites. Ph.D. dissertation, University of Queensland, Australia.

BOWLES, J., BLAIR, D. & MCMANUS, D. P. (1992). Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology* **54**, 165–174.

BOWLES, J., BLAIR, D. & MCMANUS, D. P. (1995). A molecular phylogeny of the human schistosomes. *Molecular Phylogenetics and Evolution* **4**, 103–109.

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

BOWLES, J. & MCMANUS, D. P. (1994). Genetic characterization of the Asian *Taenia*, a newly described Taeniid cestode in humans. *American Journal of Tropical Medicine and Hygiene* **50**, 33–44.

CORT, W. W. (1914). Larval trematodes from North American fresh water snails. Preliminary report. *Journal of Parasitology* **1**, 65–84.

FELSENSTEIN, J. (1989). PHYLIP – Phylogeny interference package (Version 3.2). *Cladistics* **5**, 164–166.

FRÖLICH, J. A. (1802). Beyträge zur Naturgeschichte der Eingeweidewürmer. *Naturforscher Halle* **29**, 5–96.

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

HOLMQUIST, R. (1983). Transitions and transversions in evolutionary decent: An approach to understanding. *Journal of Molecular Evolution* **19**, 134–144.

KANEV, I. (1985). On the morphology, biology, ecology and taxonomy of *E. revolutum* group (Trematoda: Echinostomatidae: *Echinostoma*). Ph.D. dissertation, University of Sofia, Bulgaria. (English abstract)

KANEV, I. (1994). Life cycle, delimitation and redescription of *Echinostoma revolutum* (Frölich, 1802) (Trematoda: Echinostomatidae). *Systematic Parasitology* **28**, 125–144.

KUMAR, S., TAMURA, K. & NEI, M. (1993). *MEGA: Molecular Evolutionary Genetics Analysis*. Version 1.02. The Pennsylvanian State University, University Park, PA.

LIE, K. J. & BASCH, P. F. (1967). The life history of *Echinostoma paraensei* sp.n. (Trematoda: Echinostomatidae). *Journal of Parasitology* **53**, 1192–1199.

MORGAN, J. A. T. & BLAIR, D. (1995). Nuclear rDNA ITS sequence variation in the trematode genus *Echinostoma*: an aid to establishing relationships within the 37 collar-spine group. *Parasitology* **111**, 609–615.

NUNN, G. B. & CRACRAFT, J. (1996). Phylogenetic relationships among the major lineages of the birds-of-paradise (Paradisaeidae) using mitochondrial DNA gene sequences. *Molecular Phylogenetics and Evolution* **5**, 445–459.

OHAMA, T., OWASA, S., WATANABE, K. & JUKES, T. H. (1990). Evolution of the mitochondrial genetic code IV. AAA as an asparagine codon in some animal mitochondria. *Journal of Molecular Evolution* **30**, 329–332.

OKAMOTO, M., BESSHO, Y., KAMIYA, M., KUROSAWA, T. & HORII, T. (1995). Phylogenetic relationships within *Taenia taeniaeformis* variants and other taeniid cestodes inferred from the nucleotide sequence of the cytochrome c oxidase subunit I gene. *Parasitology Research* **81**, 451–458.

RICHARD, J. (1964). Trematodes d'oiseaux de Madagascar (Note III). Espèces de la famille Echinostomatidae Poche 1926. *Annales de Parasitologie Humaine et Comparée* **39**, 607–620.

SAITOU, N. & NEI, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

SPERLING, F. A. H. & HICKEY, D. A. (1994). Mitochondrial DNA sequence variation in the spruce budworm species complex (Choristoneura: Lepidoptera). *Molecular Biology and Evolution* **11**, 656–665.

SWOFFORD, D. L. (1993). *PAUP: Phylogenetic Analysis Using Parsimony*. Version 3.1, Computer program distributed by the Illinois Natural History Survey Champaign, Illinois.

TAYLOR, D. J., HEBERT, P. D. N. & COLBOURNE, J. K. (1996). Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Molecular Phylogenetics and Evolution* **5**, 495–510.

TROUVÉ, S., RENAUD, F., DURAND, P. & JOURDANE, J. (1996). Selfing and outcrossing in a parasitic hermaphrodite helminth (Trematoda, Echinostomatidae). *Heredity* **77**, 1–8.

VAN DE PEER, Y. & DE WACHTER, R. (1993). TREECON: A software package for the construction and drawing of evolutionary trees. *CABIOS* **9**, 177–182.

ZAR, J. H. (1984). *Biostatistical Analysis*. Prentice-Hall International, London.