A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes

M. NAKAO[1]*, D. P. McMANUS[2], P. M. SCHANTZ[3], P. S. CRAIG[4] *and* A. ITO[1]

[1] *Department of Parasitology, Asahikawa Medical College, Asahikawa, Hokkaido 078-8510, Japan*
[2] *Molecular Parasitology Laboratory, The Queensland Institute of Medical Research and the University of Queensland, Brisbane, Queensland 4029, Australia*
[3] *Division of Parasitic Diseases, National Centers for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30341, USA*
[4] *Cestode Zoonoses Research Group, Bioscience Research Institute and School of Environment and Life Sciences, University of Salford, Great Manchester M5 4WT, UK*

SUMMARY

Taxonomic revision by molecular phylogeny is needed to categorize members of the genus *Echinococcus* (Cestoda: Taeniidae). We have reconstructed the phylogenetic relationships of *E. oligarthrus*, *E. vogeli*, *E. multilocularis*, *E. shiquicus*, *E. equinus*, *E. ortleppi*, *E. granulosus* sensu stricto and 3 genotypes of *E. granulosus* sensu lato (G6, G7 and G8) from their complete mitochondrial genomes. Maximum likelihood and partitioned Bayesian analyses using concatenated data sets of nucleotide and amino acid sequences depicted phylogenetic trees with the same topology. The 3 *E. granulosus* genotypes corresponding to the camel, pig, and cervid strains were monophyletic, and their high level of genetic similarity supported taxonomic species unification of these genotypes into *E. canadensis*. Sister species relationships were confirmed between *E. ortleppi* and *E. canadensis*, and between *E. multilocularis* and *E. shiquicus*, regardless of the analytical approach employed. The basal positions of the phylogenetic tree were occupied by the neotropical endemic species, *E. oligarthrus* and *E. vogeli*, whose definitive hosts are derived from carnivores that immigrated from North America after the formation of the Panamanian land bridge. Host-parasite co-evolution comparisons suggest that the ancestral homeland of *Echinococcus* was North America or Asia, depending on whether the ancestral definitive hosts were canids or felids.

Key words: *Echinococcus*, mitochondrial genome, phylogeny, taxonomic revision.

INTRODUCTION

Cestodes of the family Taeniidae are parasitic in mammals and are placed within either of 2 genera, *Taenia* and *Echinococcus* (Hoberg *et al*. 2000; Kumaratilake and Thompson, 1982). The adult tapeworms live in the digestive tract of their definitive hosts, while the bladder-like larvae occur in the internal organs of intermediate hosts. The predator-prey relationships of the mammalian hosts perpetuate the life-cycle of the taeniid cestodes. The tiny adult tapeworms of the genus *Echinococcus* utilize canids and/or felids as definitive hosts, and their larvae develop into hydatid cysts in intermediate hosts including ungulates, rodents and humans. The 2 main species, *E. granulosus* and *E. multilocularis*, cause cystic and alveolar echinococcosis in humans, respectively. The organisms of *Echinococcus* have a biphasic propagation manner comprising a sexual reproductive phase in the hermaphroditic adults and an asexual proliferation of protoscoleces in the larval stage. The self-fertilization of the adults and the clonal amplification of the larvae are responsible for genetic monomorphism within isolated *Echinococcus* populations (Lymbery and Thompson, 1996; Haag *et al*. 1999), and may lead, eventually, to the speciation of *Echinococcus* (Smyth and Smyth, 1964).

The classification of *Echinococcus* has been controversial because of the paucity of phenotypic characters and inadequate taxonomic descriptions. In taxonomic terms, many species and subspecies of *Echinococcus* were originally described, based mainly on host-parasite specificity characteristics (Ortlepp, 1937; Lopez-Neyra and Soler Planas, 1943; Williams and Sweatman, 1963; Verster, 1965). However, most of these taxa were regarded as synonyms for *E. granulosus* (Rausch, 1967), and the subsequent reclassification permitted only 4 morphospecies, namely *E. granulosus*, *E. multilocularis*, *E. oligarthrus* and *E. vogeli* (Rausch and Bernstein, 1972). The biological variants of *E. granulosus*, which utilize particular ungulates as intermediate hosts, have been designated as strains (Thompson *et al*. 1995). Molecular genetic studies, using mainly mitochondrial DNA (mtDNA) sequences, supported the strain definition and identified 10 genotypes (G1 to G10) within *E. granulosus* (Bowles *et al*. 1994,

* Corresponding author: Department of Parasitology, Asahikawa Medical College, Asahikawa, Hokkaido 078-8510, Japan. Tel: +81 166 68 2422. Fax: +81 166 68 2429. E-mail: nakao@asahikawa-med.ac.jp

Table 1. Mitochondrial genomes of *Echinococcus* and *Taenia* tapeworms analysed in this study

Species (genotype)	bp in total length	Nucleotide compositions of sense strand				Database accession no.	References
		% A	% G	% T	% C		
E. oligarthrus	13,791	20·8	23·3	48·5	7·4	AB208545	This study
E. vogeli	13,750	18·9	25·3	48·2	7·6	AB208546	This study
E. granulosus (G1)	13,588	19·2	25·0	47·9	8·0	AF297617	Le *et al.* (2002)
E. equinus (G4)	13,598	19·9	24·3	48·0	7·7	AF346403	Le *et al.* (2002)
E. ortleppi (G5)	13,717	19·6	23·9	48·2	8·3	AB235846	This study
E. sp. (G6)	13,721	19·5	24·1	48·1	8·3	AB208063	This study
E. sp. (G7)	13,719	19·5	24·1	48·2	8·2	AB235847	This study
E. sp. (G8)	13,717	19·5	24·2	48·2	8·2	AB235848	This study
E. multilocularis	13,738	20·6	23·4	48·5	7·6	AB018440	Nakao *et al.* (2002*b*)
E. shiquicus	13,807	20·1	24·1	48·3	7·5	AB208064	This study
T. solium	13,709	25·4	19·7	46·7	8·3	AB086256	Nakao *et al.* (2003)

1995; Scott *et al.* 1997; Lavikainen *et al.* 2003). The taxonomic revision of *Echinococcus* is now required to determine the status of sibling or cryptic species. To date, reviving the names of synonyms, the oversimplified *E. granulosus* has been split into *E. granulosus* sensu stricto (genotypes G1 to G3), *E. equinus* (G4) and *E. ortleppi* (G5), but the species status of genotypes G6 to G10 is still ambiguous (Thompson and McManus, 2003). Most recently, a new sibling species, *E. shiquicus*, whose adult is morphologically similar to *E. multilocularis*, has been found in Tibet (Xiao *et al.* 2005).

Mitochondrial DNA (mtDNA) has more power than nuclear DNA in reconstructing phylogenetic relationships among closely related species because of the rapid sequence evolution (Brown *et al.* 1979), and large data sets derived from mitochondrial genomes have the potential for resolving problematic issues in taxonomy (Mueller *et al.* 2004). In this study, the phylogeny of the genus *Echinococcus* was clarified by using the complete mitochondrial genomes of representative taxa in order to improve the consistency between species status and traditional nomenclatures. The concatenated data sets of nucleotide and amino acid sequences derived from all mitochondrial genes were analysed with an array of phylogenetic methods including maximum likelihood and partitioned Bayesian analyses. In particular, the latter is the only method that can be used to analyse a large data set by incorporating complex substitution models into the statistical process (Ronquist and Huelsenbeck, 2003; Nylander *et al.* 2004). Also, the ancestral origin of *Echinococcus* was inferred, based on the phylogeny of *Echinococcus* and the evolutionary history of mammalian host animals.

MATERIALS AND METHODS

Taxon sampling

Ten taxa of *Echinococcus* and 1 outgroup (*Taenia solium*) were used in this study (Table 1). Complete published sequences were available for mitochondrial genomes of *E. granulosus* sensu stricto, *E. equinus*, *E. multilocularis* and *T. solium* (Le *et al.* 2002; Nakao *et al.* 2002*b*, 2003). In addition, complete mtDNA sequences were obtained for larval isolates of *E. oligarthrus* (Panama), *E. vogeli* (Colombia), *E. shiquicus* (Tibet), *E. ortleppi* (Argentina), *E. granulosus* genotype G6 (Kazakhstan), *E. granulosus* genotype G7 (Poland) and *E. granulosus* genotype G8 (USA). Previous studies (Bowles *et al.* 1995; Lavikainen *et al.* 2003) indicate that the taxa of *Echinococcus* examined in this study cover a range of major sequence variation in mtDNA within the genus.

DNA sequencing

Genomic DNA purified from each of the *Echinococcus* isolates was used as a template for the polymerase chain reaction (PCR). As reported previously (Nakao *et al.* 2003), DNA fragments (0·7–2·5 kb) covering complete mitochondrial genomes were amplified by PCR, and their DNA sequences were compiled into a total sequence. In brief, the initial amplification of mtDNA was carried out using primers designed from the conserved regions of *E. multilocularis* mtDNA (Nakao *et al.* 2002*b*). The sequences of these amplicons allowed us to design additional primers, whereby the remaining unknown regions were amplified. The cycle sequencing kits, DYEnamic ET terminator (Amersham Biosciences) and BigDye terminator (Applied Biosystems), were used for the direct sequencing of PCR products. After each sequencing reaction, synthesized ladders were run on an ABI PRISM 377 genetic analyser (Applied Biosystems). Large DNA templates were sequenced by primer walking.

Data processing

Mitochondrial genomes were annotated as reported previously (Nakao *et al.* 2002*b*). Open reading

frames (ORFs) of 12 protein-coding genes for ATPase subunit 6 (*atp6*), cytochrome *b* (*cob*), cytochrome *c* oxidase subunits 1 to 3 (*cox1* to *3*) and NADH dehydrogenase subunits 1 to 6 and 4L (*nad1* to *6* and *nad4L*) were determined using the echinoderm mitochondrial genetic code (Nakao *et al.* 2000; Telford *et al.* 2000). Large and small subunit rRNA genes (*rrnL* and *rrnS*) and 22 tRNA genes (*trn*) were identified by their sequence similarities and secondary structures. Multiple alignments and basic statistics of nucleotide and amino acid sequences were achieved by ClustalX 1.82 (Thompson *et al.* 1997) and MacClade 4.08 (Maddison and Maddison, 2000). PAUP4.0b10 (Swofford, 2002) was applied to make pairwise distance matrices. The conservative regions of mitochondrial genomes were displayed by the program plotcon of EMBOSS 3.00 (Rice *et al.* 2000) using the alignment data of complete nucleotide sequences from all *Echinococcus* taxa. Gene-by-gene alignment was carried out for the 11 taxa (10 *Echinococcus* and 1 *Taenia*) to prepare data sets for phylogenetic analyses. Based on the alignments of deduced amino acid sequences, the nucleotide sequences of each protein-coding gene were aligned manually. The alignments of rRNA and tRNA genes were corrected by deleting ambiguous positions, such as loops and indels. The resultant alignments were concatenated into 3 data sets as follows: data set #1 comprising nucleotide sequences for all genes (total 13,130 sites); data set #2 comprising nucleotide sequences for 12 protein-coding genes excluding the third codon position (total 6746 sites); and data set #3 of amino acid sequences for the 12 proteins (total 3361 sites). Non-coding regions were considered unusable for phylogenetic analyses because they contained highly repetitive sequences.

PAUP4.0b10 (Swofford, 2002) was employed for maximum likelihood (ML) analysis using the nucleotide data sets #1 and #2. A suitable nucleotide substitution model and its parameter values were determined by Akaike Information Criterion (AIC) implemented in MODELTEST 3.7 (Posada and Crandall, 1998). A full heuristic search algorithm was utilized to estimate the ML trees. To conduct ML analysis for proteins, the program proml of PHYLIP 3.6 (Felsenstein, 2005) was run using the amino acid data set #3. The Jones-Taylor-Thornton (JTT) model for the substitution of amino acids was used with a gamma-shape parameter estimated from the data. In both analyses, the robustness of inferred trees was tested by non-parametric bootstrapping with 1000 pseudoreplicates. Consensus trees were constructed by the 50% majority rule and rooted with an outgroup taxon.

Partitioned Bayesian analyses were implemented using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). The nucleotide data set #1 comprising all genes was divided into 15 partitions. Twelve protein-coding genes and 2 rRNA genes were treated as individual data partitions and all tRNA genes were assigned as a single data partition. The best-fit nucleotide substitution models of each partition were selected by AIC implemented in MrModeltest 2.2 (modified from MODELTEST by J. A. A. Nylander, Uppsala University, Sweden). The amino acid data set #3 comprising all inferred proteins was divided into 12 partitions and their substitution models were selected by AIC in PROTTEST 1.26 (Abascal *et al.* 2005). Using each of the partitioned data sets, a Metropolis-coupled Markov chain Monte Carlo analysis was run for 1 million generations to estimate the posterior probabilities of trees. The tree sampling was taken every 1000 generations and all tree samples before the chain reached a stationarity were discarded as burn-in. Two independent analyses were conducted to verify the consistency of results.

RESULTS

Characterization of mitochondrial genomes

All mtDNAs examined in this study were in the form of a covalently closed-circular molecule of ∼14 kilo base pairs (Table 1). The genomes comprised 12 protein-coding genes, 2 rRNA genes, 22 tRNA genes (1 for each of 18 amino acids and 2 each for serine and leucine) and 2 short non-coding regions involved in replication and translation. The gene for ATPase subunit 8 was absent. Most of the genes were separated by a few bases, and the protein-coding genes contained no introns. The gene orders were identical among all of the genomes examined. All genes were located on the same strand, which was strongly biased towards thymine and against cytosine (Table 1). The nucleotide A+T contents of the *Echinococcus* mtDNAs ranged between 67·1 and 69·3%. The protein-coding genes were initiated by either an ATG or GTG codon and terminated by a TAG or TAA codon. The abbreviated stop codons T and TA, which are modified into a complete TAA codon by polyadenylation, were detected in the *Taenia* mtDNA (Nakao *et al.* 2003) but not in the *Echinococcus* mtDNAs.

The similarity scores of complete nucleotide sequences among 10 *Echinococcus* mtDNAs were plotted for every nucleotide position (Fig. 1). Conservative sequences were demonstrated in *cob*, *nad3*, *cox1*, *cox2* and *rrnS*, while the non-coding region, located upstream from *nad5*, was highly polymorphic due to the deletion and insertion of repetitive sequences. An extreme imbalance of the similarity scores was observed within the sequences for *nad5* and *rrnL*. Concerning the tRNA genes, the sequences of the anticodon arm (17 bases for each tRNA) were well conserved. However, the loops
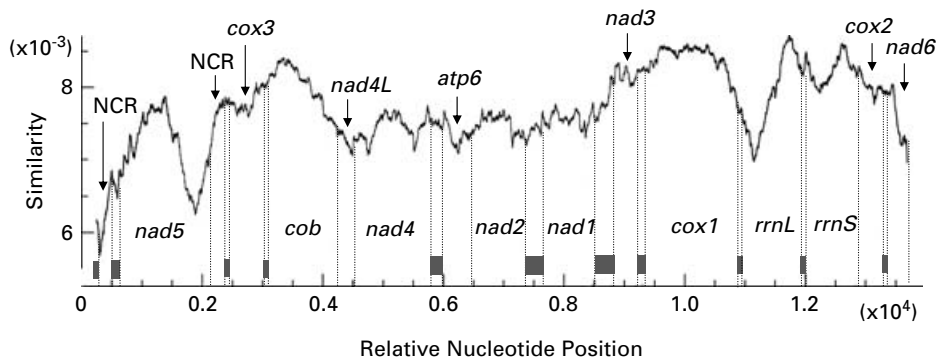
Fig. 1. Similarity of nucleotide sequences among mitochondrial genomes of 10 *Echinococcus* taxa. The positions of genes and non-coding regions are marked with their abbreviations. Shaded quadrates indicate genes for transfer RNA. Abbreviations: *atp6*, ATPase subunit 6; *cox1* to *3*, cytochrome *c* oxidase subunits 1 to 3; *nad1* to *6* and *nad4L*, NADH dehydrogenase subunits 1 to 6 and 4L; NCR, non-coding region; *rrnL* and *rrnS*, large and small subunit ribosomal RNAs.

of the D and TΨC arms were variable (data not shown).

For all *Echinococcus* taxa, the number of nucleotide substitutions in all protein-coding genes was counted at every codon position after the alignments of each gene were concatenated. Of a total of 3426 variable sites, the second codon position had the fewest sites ($n=426$, 12·4%), followed by the first codon position ($n=904$, 26·4%), whereas the third codon position had the most sites ($n=2096$, 61·2%). The transition/transversion ratios for the first, second and third codon positions were 2·55, 1·68 and 1·69, respectively. The average frequencies of amino acid replacements are illustrated by MacClade's bubble charts using the amino acid data set #3 of all proteins (Fig. 2). The most frequent amino acid changes among the mitochondrial proteins predicted for the *Echinococcus* taxa were Val to Ile, Ile to Val, Leu to Phe and Phe to Leu. Most of these changes corresponded directly to the transitional substitutions of the first codon position.

Phylogenetic analyses

The phylogenetic relationships among 10 *Echinococcus* taxa were estimated mainly by using the nucleotide data set #1 representing all of the mitochondrial genes. To reduce the influence of saturation at the third codon position, the nucleotide data set #2 of 12 protein-coding genes (without third codon positions) and the amino acid data set #3 of 12 proteins were also used. The substitution models of these data sets for ML and partitioned Bayesian analyses are shown in Table 2. The Tamura-Nei (TrN)+G+I model was selected for ML analysis as the best fit model to describe the data set #1. Parameters for this model included base frequencies (A=0·2114, C=0·0653, G=0·2340 and T=0·4893), substitution rate matrix (AC=1·0000, AG=11·6176, AT=1·0000, CG=1·0000, CT=8·3353 and GT=1·0000), gamma distribution shape
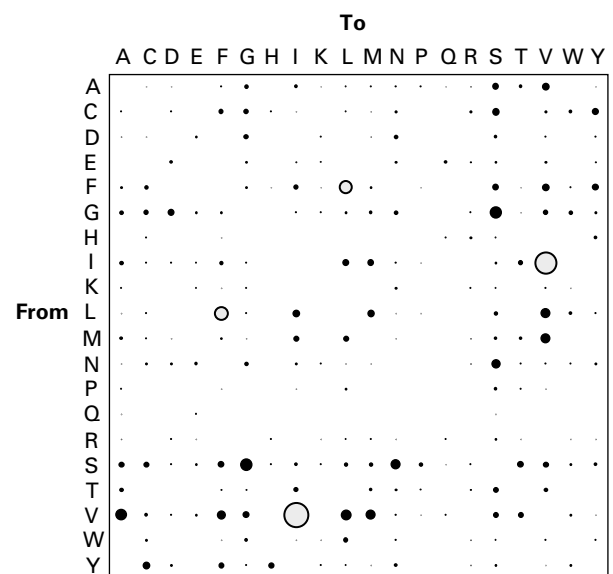

Fig. 2. The average frequencies of amino acid replacements in all mitochondrial proteins among 10 *Echinococcus* taxa. The diameter of the largest circle corresponds to 278 changes and other circles are scaled in diameter proportionally. Open circles represent the frequent changes more than 70. The values were determined from the alignment of amino acid sequences by MacClade 4.08 using the cladogram obtained by ML analysis.

(0·9340) and proportion of invariable sites (0·4179). The tree topology inferred from the ML analysis is displayed as a cladogram (Fig. 3A) and as a phylogram (Fig. 3B). Two clades (α and β) were supported with high bootstrap values (98–100%). Clade α included *E. ortleppi* and genotypes G6, G7 and G8, whereas clade β consisted of *E. multilocularis* and *E. shiquicus*. The phylogram depicted a close relationship between the members of clade α. The other taxa, *E. oligarthrus*, *E. vogeli*, *E. granulosus* sensu stricto and *E. equinus*, were divided deeply into independent branches with moderate bootstrap

Table 2. Nucleotide and amino acid substitution models used for maximum likelihood and partitioned Bayesian analyses

(The models of individual genes were applied to partitioned Bayesian analysis, while the models of total sequences were used for maximum likelihood analysis.)

Genes	Nucleotide		Amino acid	
	No. sites examined	Substitution models	No. sites examined	Substitution models
atp6	519	HKY+G+I	172	JTT+G
cob	1,068	HKY+G+I	355	CpREV+G
cox1	1,611	GTR+G+I	536	MtMam+G+I
cox2	582	GTR+G+I	193	JTT+G+I
cox3	648	GTR+G+I	215	JTT+G
nad1	897	GTR+G+I	298	JTT+G
nad2	885	GTR+G+I	294	JTT+G
nad3	348	HKY+G	115	CpREV+G
nad4	1,263	GTR+G+I	420	JTT+G
nad4L	261	HKY+G	86	MtMam+G
nad5	1,581	GTR+G+I	526	JTT+G+I
nad6	456	GTR+G+I	151	JTT+G+I
rrnL	949	GTR+G		
rrnS	703	GTR+G		
trn	1,359	GTR+G+I		
Total (#1)	13,130	TrN+G+I (#3)	3,361	JTT+G
(#2)	6,746	GTR+G+I		
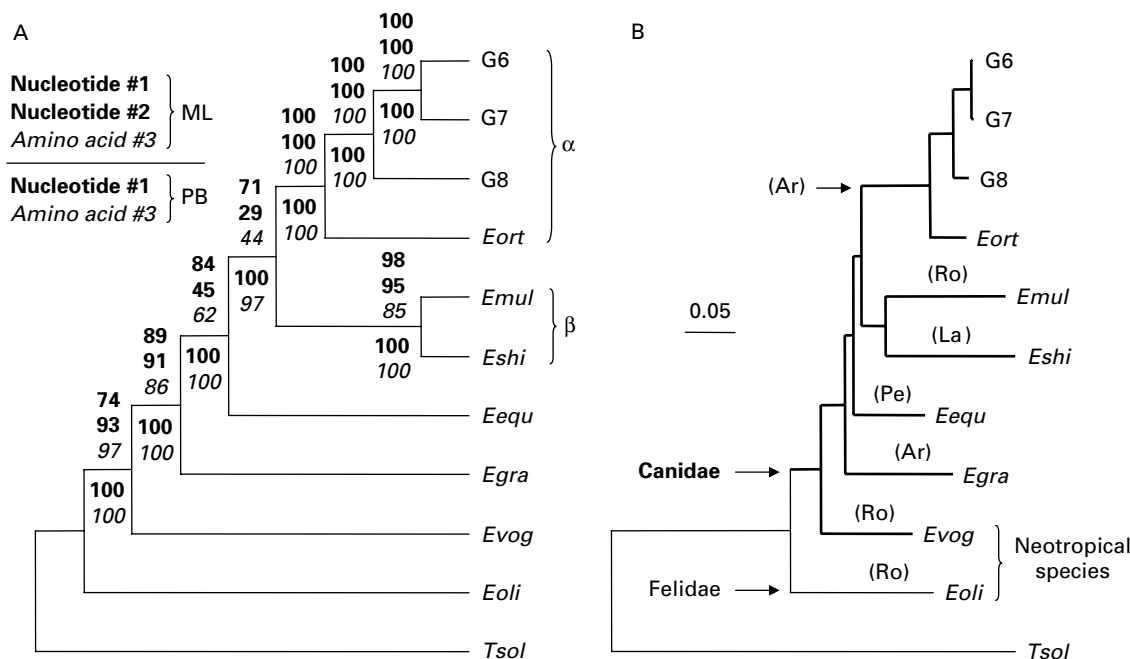
Fig. 3. Phylogenetic trees for members of the genus *Echinococcus* inferred from mitochondrial genome data. Abbreviations: *Eequ*, *E. equinus*; *Egra*, *E. granulosus* sensu stricto; *Emul*, *E. multilocularis*; *Eoli*, *E. oligarthrus*; *Eort*, *E. ortleppi*; *Eshi*, *E. shiquicus*; *Evog*, *E. vogeli*; G6 to G8, *Echinococcus* genotypes; *Tsol*, *Taenia solium* (outgroup). (A) Cladogram obtained through maximum likelihood (ML) and partitioned Bayesian (PB) analyses. ML bootstrap proportions (%) are shown above branches, while Bayesian posterior probabilities (%) are shown below. The confidence values of both analyses using the nucleotide data sets #1 and #2 are shown in boldface. Numbers in italics represent the values obtained from the amino acid data set #3. (B) Phylogram obtained by ML analysis using the data set #1. The scale bar represents the estimated number of nucleotide substitutions per nucleotide site. Bold lines indicate the topology of *Echinococcus* using Canidae as definitive hosts. The orders of intermediate hosts (Ar, Artiodactyla; La, Lagomorpha; Pe, Perissodactyla; Ro, Rodentia) are mapped to branches.

values (71–89%). Further ML analyses using the data sets #2 and #3 produced phylogenies with unresolved nodes (Fig. 3A). In trees derived from both data sets, the basal positions of *E. oligarthrus* and *E. vogeli* were strongly supported, whereas the phylogenetic relationships among *E. granulosus* sensu stricto, *E. equinus* and clades α and β were uncertain because their nodes showed low bootstrap values. Partitioned Bayesian analyses using data sets #1 and #3 resulted in a robust phylogeny (Fig. 3A). The posterior probabilities of each node mostly reached 100% for both data sets. The tree topology of the Bayesian analyses was consistent with those of the ML analyses. The Bayesian analyses supported a sister group relationship between clades α and β. Furthermore, the resultant trees clarified the phylogenetic positions of *E. granulosus* sensu stricto and *E. equinus*, which remained as uncertain in the ML analyses.

Host-parasite relationships

Mammalian hosts for the members of *Echinococcus* were mapped to the parasite phylogram to examine host-parasite relationships (Fig. 3B). The definitive hosts were represented by the families Felidae and Canidae. The orders Rodentia, Artiodactyla, Perissodactyla and Lagomorpha were used to simplify an abundance of the intermediate hosts. A switch in definitive hosts might occur in the basal node of the phylogram because only *E. oligarthrus* utilizes Felidae, but the other monophyletic taxa colonize Canidae including dogs and foxes. Clade β consists of *E. multilocularis* and *E. shiquicus* which especially colonize foxes. Ancestral definitive hosts, whether Felidae or Canidae, were uncertain. Concerning intermediate hosts, a coevolutionary relationship was evident in clade α consisting of *E. ortleppi* and genotypes G6, G7 and G8. The members of the clade colonize Artiodactyla including cattle (*E. ortleppi*), camels (G6), pigs (G7) and cervids (G8). The basal species of *E. oligarthrus* and *E. vogeli* are associated typically with Rodentia, suggesting that rodents might be their ancestral intermediate hosts.

Divergence of closely related taxa

Pairwise divergence values for mtDNA sequences were compared to evaluate the species status of *E. ortleppi* and genotypes G6, G7 and G8. Within the family Taeniidae, *Taenia saginata* and *Taenia asiatica* are recognized as a pair of recently evolved species (Hoberg *et al.* 2001) and are useful as a reference for closely related species. The complete nucleotide sequences of the *cob* and *cox1* genes have been determined for *T. saginata* and *T. asiatica* (Nakao *et al.* 2002 *a*). Therefore, these genes were selected for comparisons of pairwise divergence. The

Table 3. Pairwise divergence values of mtDNA sequences among closely related taxa of *Echinococcus* (*E. ortleppi* and genotypes G6, G7, and G8) and between sister species of *Taenia* (*T. saginata* and *T. asiatica*)

(Kimura's 2-parameter distances were computed from concatenated sequences of *cob* and *cox1* (total 2676 sites) using a gamma-shape parameter (a=1).)

	G6	G7	G8
G7	0·004		
G8	0·037	0·038	
E. ortleppi	0·059	0·060	0·058
T. saginata versus *T. asiatica*		0·046	

matrix of Kimura's 2-parameter distance was calculated from the concatenated data set for the *cob* and *cox1* sequences (total 2676 sites) using a gamma shape parameter (a=1). The pairwise divergence value between *T. saginata* and *T. asiatica* was used as a standard for incipient speciation. As shown in Table 3, the remarkably low value of pairwise divergence between G6 and G7 suggests that these genotypes belong to a single species. On the other hand, the species status of G8 is problematic, because the divergence values between G8 and the other genotypes were close to the threshold value between *T. saginata* and *T. asiatica*. The divergence values between *E. ortleppi* and the 3 genotypes were at interspecific levels.

DISCUSSION

Concurrent descriptions of species, strains and genotypes have been used to record the biodiversity of *Echinococcus* in the context of medical and veterinary criteria for diagnosing and controlling echinococcosis in various settings (McManus and Thompson, 2003; Thompson and McManus, 2003). As a general rule, strains and genotypes are employed to categorize intraspecific variation. However, previous molecular phylogenetic studies on *Echinococcus* (Bowles *et al.* 1995; Thompson *et al.* 1995) demonstrated that all strains or genotypes of *E. granulosus* could not be assembled into a monophyletic group, indicating that *E. granulosus* sensu lato is a cryptic species complex. Accordingly, the arbitrary use of strains and genotypes without a correction of the old nomenclature causes unnecessary confusion. The taxonomy of the cryptic species should be based entirely on their phylogenetic relationships, and appropriate scientific names should be used instead of strains and genotypes. The present study provides useful information for dividing *E. granulosus* sensu lato into distinct species.

The phylogeny of *Echinococcus* presented in this study is the first to be reconstructed from whole mitochondrial genome sequences. Confidence in the
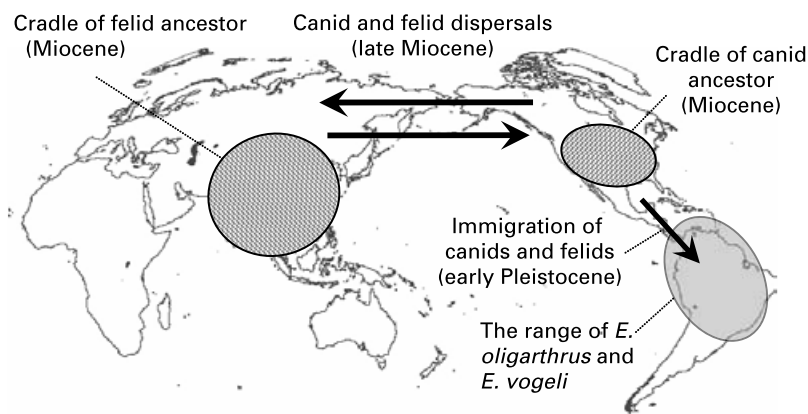
Fig. 4. The proposed movements of ancient definitive hosts of *Echinococcus*. The double arrows on the Bering Strait show the dispersals of primitive canids and felids during the late Miocene. The single arrow on Central America represents the immigration of carnivores into South America during the early Pleistocene.

proposed topology was increased by the overall support of the nodes with ML and partitioned Bayesian analyses. However, divergence dates cannot be estimated due to the absence of fossil records. The resultant phylogeny included representative genotypes of *E. granulosus* sensu lato. Genotypes G2 (Tasmanian sheep strain) and G3 (buffalo strain) were not used in this study because they have very similar mtDNA sequences to that of genotype G1 (sheep strain) (Bowles *et al.* 1995). Thompson *et al.* (1995) regarded the genotypes G1, G2 and G3 as a single species, and these constitute the species *E. granulosus* sensu stricto. The present phylogeny confirms the validity of *E. granulosus* sensu stricto and supports the recent revisions of genotypes G4 (horse strain) and G5 (cattle strain) into *E. equinus* and *E. ortleppi*, respectively (Thompson and McManus, 2003). Moreover, the tree depicted a monophyly of *E. ortleppi* and genotypes G6 (camel strain), G7 (pig strain) and G8 (cervid strain). The other genotypes G9 (Polish human strain) and G10 (Fennoscandian cervid strain) were not analysed in this study. However, close relationships between G7 and G9 and between G8 and G10 have been reported using mtDNA sequences (Scott *et al.* 1997; Lavikainen *et al.* 2003; Xiao *et al.* 2005; Thompson *et al.* 2006). In this study, the values of pairwise divergence among G6, G7 and G8 were lower than a threshold value for speciation estimated from *T. saginata* and *T. asiatica*. Using multiple loci of mtDNA (*nad1*, *atp6* and *cox1*), Lavikainen *et al.* (2006) showed that the camel, pig and cervid strains clustered into a monophyletic group and that the cervid genotypes (G8 and G10) could not be separated out from this group. This result and our data suggest that the camel, pig and cervid strains may belong to a single species. The same suggestion has been made by Xiao *et al.* (2005) and Thompson *et al.* (2006). The camel and pig strains have not previously been named (Thompson *et al.* 1995), but the cervid strain was

formerly named *E. granulosus canadensis* (Cameron, 1960; Webster and Cameron, 1961) or *E. granulosus borealis* (Sweatman and Williams, 1963). In terms of priority for nomenclature, *E. canadensis* is the most suitable for genotypes G6 to G10. Based on the resultant phylogeny, we recommend a taxonomic revision to unify these genotypes into *E. canadensis*. Lavikainen *et al.* (2005) have already made the same proposal. In our recommendation, *E. canadensis* and *E. ortleppi* are regarded as 'sister species'. An alternative revision to this unified classification is to describe a new species for the camel and pig strains (genotypes G6, G7 and G9). In this case, the use of *E. canadensis* would be limited to the cervid strain (genotypes G8 and G10). The morphological similarities between the camel and pig strains (Eckert *et al.* 1993) provide a significant piece of evidence to support the new species. Further, comprehensive population genetic, morphological and ecological studies are required for a final decision about the taxonomic status of the camel, pig and cervid strains.

A sister species relationship between *E. multilocularis* and *E. shiquicus* was demonstrated in this study. The morphological similarities between both adult tapeworms (Xiao *et al.* 2005) and the common usage of foxes as definitive hosts support this relationship. The extreme exogenous proliferation of larval cysts is characteristic of *E. multilocularis* and is responsible for the serious pathological effects associated with alveolar echinococcosis. In contrast, the larval cyst of *E. shiquicus* is basically unilocular (Xiao *et al.* 2005). The relationship between *E. multilocularis* and *E. shiquicus* strongly suggests that *E. multilocularis* only acquired the ability of the exogenous proliferation during the bifurcation of species. Although the larvae of *E. oligarthrus* and *E. vogeli* are also polycystic, the present phylogeny indicates that both species are not related to *E. multilocularis* in the evolutionary process of acquiring exogenous proliferation.

The most striking feature of the present phylogeny is that basal positions are occupied by *E. oligarthrus* and *E. vogeli*. A previous study using partial nucleotide sequences of mitochondrial genes (*cox1* and *nad1*) also ranked these species as basal lineages, but the result was not fully supported due to the short sequences examined (Bowles *et al.* 1995). The basal positions of *E. oligarthrus* and *E. vogeli* are important in considering phylogeographic patterns because both species are endemic to the Neotropical region and colonize wildlife as definitive and intermediate hosts. As compared with these endemic species, *E. granulosus* sensu stricto, *E. equinus*, *E. ortleppi* and *E. canadensis* are inadequate for phylogeographical considerations because worldwide livestock movement in historic and prehistoric times seriously disturbed their original distributions. Nevertheless, all mammals, which can serve as definitive and intermediate hosts for *Echinococcus*, are restricted to members of the superorders Laurasiatheria and Euarchontoglires of the Northern Hemisphere origin (Murphy *et al.* 2001), suggesting that the adaptive radiation of the mammals is linked with the evolution of *Echinococcus*.

We propose plausible hypotheses for the history of *Echinococcus* based on geological events and the evolution of felids and canids. Several neotropical felids such as jaguarondi, ocelot and jaguar are definitive hosts for *E. oligarthrus*, whereas only the bush dog is known to be a definitive host for *E. vogeli* (D'Alessandro *et al.* 1981; Rausch and Bernstein, 1972). As shown in Fig. 4, these carnivores are derived from immigrants from North America after the formation of the Panamanian land bridge. The immigration of North American felids and canids occurred during the early Pleistocene and these evolved into endemic species in South America (Marshall *et al.* 1982). Accordingly, it seems likely that the ancestral species of *Echinococcus* invaded South America together with the immigrants. The present members of *Echinococcus* usually utilize canids as definitive hosts except for *E. oligarthrus*. In this study, it was uncertain whether the ancestral definitive hosts were felids or canids. If they were canids, one could speculate that North America was the cradle of *Echinococcus*, as canids originated in North America (Wayne *et al.* 1997; Wang *et al.* 2004). Another plausible scenario should be considered in case that the ancestral hosts were felids. Recently, a large-scale molecular phylogenetic study on felids suggested that modern felids arose in Asia (Johnson *et al.* 2006), and their lineages migrated repeatedly between Eurasia and North America and invaded Africa. Considering felid evolution, Asia should also be considered as a candidate for the cradle of *Echinococcus*.

In Africa, *E. granulosus* sensu stricto, *E. ortleppi* and *E. canadensis* (genotype G6) prevail in domesticated animals (McManus and Thompson, 2003), whereas *Echinococcus* infections of wildlife including lions, zebras, wildebeest and warthogs suggest a distinst species or strain. This parasite was formerly named as *E. felidis* (Ortlepp, 1937; Verster, 1965) and now categorized as the lion strain of *E. granulosus* sensu lato (Thompson and McManus, 2003). The genetic makeup of this strain is completely unknown because wildlife preservation currently precludes its sampling and molecular analysis. Further studies are required to clarify the taxonomic status of the lion strain of *E. granulosus* sensu lato. The molecular phylogeny of *Echinococcus* should be tested further by adding this taxon.

REFERENCES

Abascal, F., Zardoya, R. and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.

Bowles, J., Blair, D. and McManus, D. P. (1994). Molecular genetic characterization of the cervid strain ('northern form') of *Echinococcus granulosus*. *Parasitology* **109**, 215–221.

Bowles, J., Blair, D. and McManus, D. P. (1995). A molecular phylogeny of the genus *Echinococcus*. *Parasitology* **110**, 317–328.

Brown, W. M., George, M. Jr. and Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA* **76**, 1967–1971.

Cameron, T. W. M. (1960). The incidence and diagnosis of hydatid cyst in Canada. *Echinococcus granulosus* var. *canadensis*. *Parassitologia* **2**, 381–390.

D'Alessandro, A., Rausch, R. L., Morales, G. A., Collet, S. and Angel, D. (1981). *Echinococcus* infections in Colombian animals. *The American Journal of Tropical Medicine and Hygiene* **30**, 1263–1276.

Eckert, J., Thompson, R. C. A., Lymbery, A. J., Pawlowski, Z. S., Gottstein, B. and Morgan, U. M. (1993). Further evidence for the occurrence of a distinct strain of *Echinococcus granulosus* in European pigs. *Parasitology Research* **79**, 42–48.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author via the internet. Department of Genome Sciences, University of Washington, Seattle, Washington, USA.

Haag, K. L., Araujo, A. M., Gottstein, B., Siles-Lucas, M., Thompson, R. C. A. and Zaha, A. (1999). Breeding systems in *Echinococcus granulosus* (Cestoda; Taeniidae): selfing or outcrossing? *Parasitology* **118**, 63–71.

Hoberg, E. P., Jones, A., Rausch, R. L., Eom, K. S. and Gardner, S. L. (2000). A phylogenetic hypothesis for species of the genus *Taenia* (Eucestoda: Taeniidae). *The Journal of Parasitology* **86**, 89–98.

Hoberg, E. P., Alkire, N. L., de Queiroz, A. and Jones, A. (2001). Out of Africa: origins of the *Taenia* tapeworms in humans. *Proceedings of the Royal Society of London, B* **268**, 781–787.

Johnson, W. E., Eizirik, E., Pecon-Slattery, J., Murphy, W. J., Antunes, A., Teeling, E. and O'Brien, S. J. (2006). The late Miocene radiation of modern Felidae: a genetic assessment. *Science* **311**, 73–77.

Kumaratilake, L. M. and Thompson, R. C. A. (1982). A review of the taxonomy and speciation of the genus *Echinococcus* Rudolphi 1801. *Zeitschrift für Parasitenkunde* **68**, 121–146.

Lavikainen, A., Lehtinen, M. J., Ågren, E. and Meri, S. (2005). Phylogeny of the Fennoscandian cervid strain of *Echinococcus granulosus*. *Bulletin of the Scandinavian-Baltic Society for Parasitology* **14**, 92.

Lavikainen, A., Lehtinen, M. J., Meri, T., Hirvela-Koski, V. and Meri, S. (2003). Molecular genetic characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of *Echinococcus granulosus*. *Parasitology* **127**, 207–215.

Lavikainen, A., Lehtinen, M. J., Laaksonen, S., Ågren, E., Oksanen, A. and Meri, S. (2006). Molecular characterization of *Echinococcus* isolates of cervid origin from Finland and Sweden. *Parasitology* **133**, 565–570.

Le, T. H., Pearson, M. S., Blair, D., Dai, N., Zhang, L. H. and McManus, D. P. (2002). Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*. *Parasitology* **124**, 97–112.

Lopez-Neyra, C. R. and Soler Planas, M. A. (1943). Revision del genero *Echinococcus* Rud y description de una especie nuéva Parasita intestinal del perro en Almeria. *Revista ibérica de parasitología* **3**, 169–194.

Lymbery, A. J. and Thompson, R. C. A. (1996). Species of *Echinococcus*: pattern and process. *Parasitology Today* **12**, 486–491.

Maddison, D. R. and Maddison, W. P. (2000). MacClade 4: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts, USA.

Marshall, L. G., Webb, S. D., Sepkoski, J. J. and Raup, D. M. (1982). Mammalian evolution and the Great American Interchange. *Science* **215**, 1351–1357.

McManus, D. P. and Thompson, R. C. A. (2003). Molecular epidemiology of cystic echinococcosis. *Parasitology* **127** (Suppl.), S37–S51.

Mueller, R. L., Macey, J. R., Jaekel, M., Wake, D. B. and Boore, J. L. (2004). Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. *Proceedings of the National Academy of Sciences, USA* **101**, 13820–13825.

Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., Teeling, E., Ryder, O. A., Stanhope, M. J., de Jong, W. W. and Springer, M. S. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351.

Nakao, M., Okamoto, M., Sako, Y., Yamasaki, H., Nakaya, K. and Ito, A. (2002*a*). A phylogenetic hypothesis for the distribution of two genotypes of the pig tapeworm *Taenia solium* worldwide. *Parasitology* **124**, 657–662.

Nakao, M., Sako, Y. and Ito, A. (2003). The mitochondrial genome of the tapeworm *Taenia solium*: a finding of the abbreviated stop codon U. *The Journal of Parasitology* **89**, 633–635.

Nakao, M., Sako, Y., Yokoyama, N., Fukunaga, M. and Ito, A. (2000). Mitochondrial genetic code in cestodes. *Molecular and Biochemical Parasitology* **111**, 415–424.

Nakao, M., Yokoyama, N., Sako, Y., Fukunaga, M. and Ito, A. (2002*b*). The complete mitochondrial DNA sequence of the cestode *Echinococcus multilocularis* (Cyclophyllidea: Taeniidae). *Mitochondrion* **1**, 497–509.

Nylander, J. A., Ronquist, F., Huelsenbeck, J. P. and Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53**, 47–67.

Ortlepp, R. J. (1937). South African Helminths. Part I. *The Onderstepoort Journal of Veterinary Science and Animal Industry* **9**, 311–336.

Posada, D. and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

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

Rausch, R. L. and Bernstein, J. J. (1972). *Echinococcus vogeli* sp. n. (Cestoda: Taeniidae) from the bush dog, *Speothos venaticus* (Lund). *Zeitschrift für Tropenmedizin und Parasitologie* **23**, 25–34.

Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Scott, J. C., Stefaniak, J., Pawlowski, Z. S. and 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.

Smyth, J. D. and Smyth, M. M. (1964). Natural and experimental hosts of *Echinococcus granulosus* and *E. multilocularis*, with comments on the genetics of speciation in the genus *Echinococcus*. *Parasitology* **54**, 493–514.

Sweatman, G. K. and Williams, R. J. (1963). Comparative studies on the biology and morphology of *Echinococcus granulosus* from domestic livestock, moose and reindeer. *Parasitology* **53**, 339–390.

Swofford, D. L. (2002). PAUP*: phylogenetic analysis using parsimony (and other methods) 4.0 beta. Sinauer Associates, Sunderland, Massachusetts, USA.

Telford, M. J., Herniou, E. A, Russell, R. B. and Littlewood, D. T. (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, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Thompson, R. C. A., Boxell, A. C., Ralston, B. J., Constantine, C. C., Hobbs, R. P., Shury, T. and Olson, M. E. (2006). Molecular and morphological characterization of *Echinococcus* in cervids from North America. *Parasitology* **132**, 439–447.

Thompson, R. C. A., Lymbery, A. J. and Constantine, C. C. (1995). Variation in *Echinococcus*: towards a taxonomic revision of the genus. *Advances in Parasitology* **35**, 145–176.

Thompson, R. C. A. and McManus, D. P. (2003). Towards a taxonomic revision of the genus *Echinococcus*. *Trends in Parasitology* **18**, 452–457.

Verster, A. J. M. (1965). Review of *Echinococcus* species in South Africa. *The Onderstepoort Journal of Veterinary Research* **32**, 7–118.

Wang, X., Tedford, R. H., van Valkenburgh, B. and Wayne, R. K. (2004). Ancestry: evolutionary history, molecular systematics, and evolutionary ecology of Canidae. In *Biology and Conservation of Wild Canids* (ed. MacDonald, D. W. and Sillero-Zubiri, C.), pp. 39–54. Oxford University Press, Oxford.

Wayne, R. K., Geffen, E., Girman, D. J., Koepfli, K. P., Lau, L. M. and Marshall, C. R. (1997). Molecular systematics of the Canidae. *Systematic Biology* **46**, 622–653.

Webster, G. A. and Cameron, T. W. M. (1961). Observations on experimental infections with *Echinococcus* in rodents. *Canadian Journal of Zoology* **39**, 877–891.

Williams, R. J. and 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.

Xiao, N., Qiu, J., Nakao, M., Li, T., Yang, W., Chen, X., Schantz, P. M., Craig, P. S. and Ito, A. (2005). *Echinococcus shiquicus* n. sp., a taeniid cestode from Tibetan fox and plateau pika in China. *International Journal for Parasitology* **35**, 693–701.