MOLECULAR GENETIC VARIATION IN *ECHINOCOCCUS* AND *TAENIA*: AN UPDATE

Donald P McManus

Molecular Parasitology Unit, Tropical Health Program and Australian Centre for International and Tropical Health and Nutrition, The Queensland Institute of Medical Research, 300 Herston Road, Brisbane 4029, Australia

Abstract. An update on our understanding of molecular variation in *Echinococcus* and *Taenia* is provided. Genetic variation within certain species of *Echinococcus* is now a well accepted phenomenon and a number of intraspecific variants or strains of *E. granulosus*, in particular, have been characterized hitherto using a range of procedures. Newly acquired molecular information has now been used in epidemiological studies with *E. granulosus* and in phylogenetic analysis of the genus *Echinococcus*. Similarly, DNA approaches have been applied for taxonomic characterization of the recently recognized Asian *Taenia*, a third form of human *Taenia*, which occurs in Southeast Asia, and which is distinguishable from, but closely related to, *Taenia saginata*.

INTRODUCTION

For effective diagnosis, treatment and control of parasitic diseases it is essential that parasite isolates can be readily and reliably identified. For clinical diagnosis, the aim is to develop simple, reproducible and practical methods of identification. However, with many parasite groups, precise identification is often more problematical than mere recognition of examples of previously characterized groups of organisms. This is because variation in biological features with practical implications is now being detected within a rapidly growing number of parasite species. So that this diversity can be better understood and exploited, these variants need to be thoroughly characterized and categorized. Such parasite variability has profound implications for drug, diagnostics and vaccine development and an overview follows of our current understanding of the extent of molecular variation in the human tapeworms, *Echinococcus* and *Taenia*. The paper focuses on the application of DNA-based procedures in addressing problems of identification, characterization and taxonomic status of these important parasites. A more general overview emphasising the usefulness of applying genetic approaches to parasite identification in general and their value in applied parasitology and systematics is available (McManus and Bowles, in press).

VARIATION IN *ECHINOCOCCUS*

DNA methods for detecting genetic variation

Variation within certain species of *Echinococcus* is now a well accepted phenomenon (Thompson and Lymbery, 1988; Thompson, 1995) and a number of intraspecific variants or strains of *E. granulosus* have been characterized using a range of procedures. These studies have been carried out as genetic diversity may reflect differences in infectivity, especially to humans, with important implications for the epidemiology of hydatid disease. In addition, the phenomenon of strain variation is an important consideration in the future design and development of vaccines, diagnostic reagents and drugs effective against the *Echinococcus* organisms.

A number of traditional approaches including geographical distribution, host range, host specificity, antigenic diversity, metabolism, developmental rate, reproductive biology, growth *in vitro*, infectivity, morphology and protein and, isoenzyme analysis have been usefully employed for characterizing the *Echinococcus* organisms (Thompson and Lymbery, 1988; McManus and Bryant, 1995). However, DNA sequence analysis offers the most direct approach to the characterization of distinct species, subspecies

and strain groups. For the purpose of identification and assessing genetic variability, it is not necessary that genetic characters are directly related to functional or other phenotypic differences between forms. It is sufficient that they consistently act as markers of particular known epidemiological types.

Early DNA studies of genetic variation in *Echinococcus* (McManus and Simpson, 1985; McManus *et al*, 1987; McManus and Rishi, 1989) involved restriction fragment length polymorphism (RFLP) analysis using a Southern blotting approach. This RFLP procedure has recently been greatly simplified, without loss of resolution or accuracy, by linking rDNA RFLP analysis with the polymerase chain reaction (PCR) (Bowles and McManus, 1993a). Using carefully controlled conditions, the random amplified polymorphic DNA-polymerase chain reaction (RAPD-PCR) has also proved useful for distinguishing *Echinococcus* species and genetically distinct strains of *E. granulosus* (Scott and McManus, 1994).

Comparison, between organisms, of the nucleotide sequence of defined DNA segments offers the most direct and sensitive means of detecting genetic variation. Data for portions of the mitochondrial protein-coding genes, cytochrome c oxidase subunit I (COI) and NADH dehydrogenase I (ND1) is available, has revealed twelve different genotypes, including 8 within *E. granulosus* (Bowles, 1993; Bowles and McManus, 1993b; Bowles *et al*, 1992a; Bowles *et al*, 1995) (Table 1).

Table 1

Levels of pairwise nucleotide sequence differences between the cervid strain (G8 genotype) and other genotypes* of *Echinococcus granulosus* for the COI and ND1 genes.

Genotype	G1	G2	G3	G4	G5	G6	G7
COI	5.0	5.0	5.3	10.1	11.0	10.7	11.0
NDI	15.5	15.3	15.3	12.8	7.9	4.3	4.5

* Genotypes defined on the basis of COI sequence (Bowles *et al*, 1992) are: G1, domestic sheep; G2, Tasmanian sheep; G3, buffalo; G4, horse; G5, cattle; G6, camel: G7, pig.

Molecular epidemiological studies

Molecular epidemiological surveys on *E. granulosus* have fully confirmed the distinctiveness between, but uniformity within, the sheep/dog and horse/dog forms of *E. granulosus* in the United Kingdom (McManus and Simpson, 1985; McManus and Rishi, 1989; Bowles *et al*, 1992a; Bowles and McManus, 1993b), corroborated the presence and revealed the host preferences of sheep and camel strains in Kenya (Wachira *et al*, 1993), showed that a single genotypic strain of *E. granulosus* cycles in domestic and sylvatic hosts on mainland Australia (Hope *et al*, 1991, 1992), and proved that the cattle strain is infective to humans (Bowles *et al*, 1992b). In addition, it was shown in a molecular genetic survey that all of 117 *E. granulosus* isolates examined from a range of intermediate hosts, including humans, in north western China were infected with the sheep (G1 genotype) of *E. granulosus* (McManus *et al*, 1994). The findings of each of these studies have considerable implications for public health.

Similar genetic methods have been used to characterize the cervid strain or "northern form" of *E. granulosus* from North America and Eurasia (Bowles *et al*, 1994). The wolf is the principal definitive host and moose and reindeer (Family Cervidae) serve as intermediate hosts; cycles involving dogs and domesticated reindeer also occur (Rausch, 1986). *E. granulosus* of cervid origin differs in a number of biological and clinical respects from domestic strains

of the parasite (Rausch, 1986; Thompson and Lymbery, 1988) and PCR-RFLP analysis of the nuclear ITS1 region of the rDNA repeat was able to readily distinguish the cervid form from other strains of *E. granulosus* (Bowles *et al*, 1994). The complexity of the RFLP patterns obtained with the moose isolates suggested that a number of distinct ITS1 types are present and, indeed, four cervid strain clones of PCR-amplified ITS1 were shown to be different by PCR-RFLP analysis (Bowles *et al*, 1994). In subsequent analysis (Bowles *et al*, 1995), a wide range of ITS1 sequence variation was shown among these clones suggesting that this strain may represent an inter-strain hybrid. The data suggesting also that the cervid strain might be a relatively recently evolved variant and it appears that the process of concerted evolution has not yet homogenized different types of DNA repeat within the genome.

The cervid genotype is of further interest in terms of its mitochondrial ND1 and COI sequence as each of the two regions implied different relationships for the strain. The COI sequence had ambiguities at a number of sites but was most similar to that of the sheep strain (G1 and G2) whereas the ND1 sequence most resembled camel and pig strains (G6, G7) (Table 1). These data may indicate mitochondrial heteroplasmy in this strain. The cervid form does appear to represent a distinct genetic form as evidenced by unique ND1 sequence and ITS1 PCR-RFLP patterns.

Molecular phylogenetic analysis

Phylogenetic relationships among *Echinococcus* species and strains are far from clear based on morphological data (see Lymbery, 1992) and adequate biochemical data are not available. The very large number of characters made available by molecular analysis, and in particular DNA sequence analysis, can prove very useful in the refining of the available morphologically-based phylogenies. Sequence data, involving 3 nucleotide data sets - 2 mitochondrial (partial sequences for COI and ND1) and one nuclear (complete sequence for ribosomal ITS1) - has now been used (Bowles *et al*, 1995) to infer phylogenetic relationships within the *Echinococcus* genus. Phylogenies, based on the principle of parsimony, are shown as phylograms of the single most parsimonious trees produced from the combined mitochondrial data set (Fig 1) and the ITS1 data (Fig 2). The data does not support the hypothesis that *E. granulosus,* as it is currently viewed, is a single valid species. Rather, the strains of *E. granulosus* seem to comprise at least 3 evolutionarily diverse groups (sheep, bovine - including the camel and pig strains - and horse strain groups) with molecular distances between them being comparable to, or greater than, molecular evolutionary distances observed between recognized species. Rausch (1986) considered the cervid strain to be ancestral to the domestic strains of *E. granulosus* which he contended became adapted to synanthropic hosts with the development of animal husbandry. However, although more molecular data is required before firm conclusions can be made on the status of the cervid strain, the phylogenetic analysis does not support the hypothesis that this strain represents the ancestral form of *E. granulosus*. If the cervid strain really was ancestral, one would expect it to appear basally in the phylogenies, rather than in a cluster with other strains. The data also suggest that *E. multilocularis* may not be distinct from *E. granulosus* although *E. vogeli* and *E. oligarthrus* appear distinct and rather distant from the first two. Overall, the molecular data indicates that taxonomic revision of the *Echinococcus* genus is now clearly warranted.

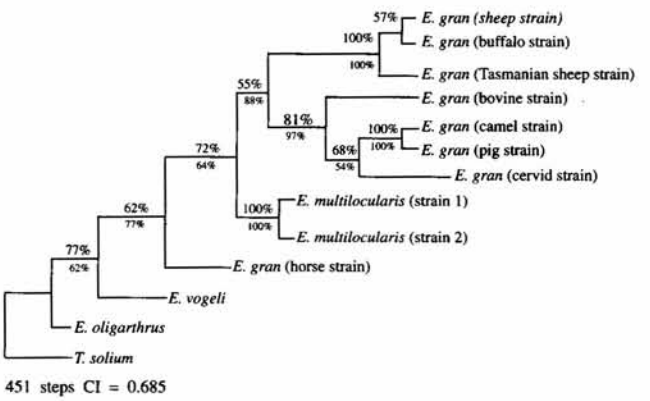

Fig 1 - The single most-parsimonious tree identified using the combined mitochondrial COI and ND1 data sets presented as a phylogram (after Bowles *et al*, 1995).

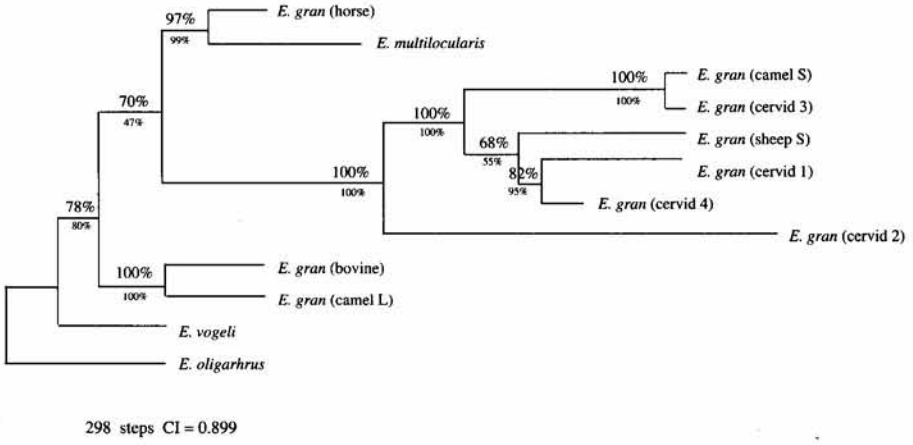

Fig 2 - The single most parsimonious tree identified using the ITS1 data set presented as a phylogram (after Bowles *et al*, 1995).

MOLECULAR VARIATION IN *TAENIA*

The human tapeworms, *Taenia saginata* and *T. solium*, have a cosmopolitan distribution, but in the Asian-Pacific region, *T. saginata* has long been considered the dominant species. This is in spite of the fact that in a number of Asian countries, people frequently eat raw or undercooked pork, but rarely eat beef. Extensive studies on the epidemiological pattern of taeniasis in Southeast Asia, principally by Dr PC Fan and his colleagues (for example, Fan, 1991) have shown that it is another form of human *Taenia* which probably accounts for this paradox. This recently recognized form, originally described in Taiwanese aborigines and hence termed "Taiwan *Taenia*", has now been recorded in Korea, Indonesia, the Philippines and Thailand, and is now generally referred to as Asian *Taenia*.

A number of features distinguish Asian *Taenia* from the other human taeniids. It can infect cattle, goats, monkeys and wild boars, but pigs are the dominant intermediate hosts and the cysticerci are found generally in the liver although experimental infection of pigs has shown that extrahepatic organs (omentum, lungs and serosa of the colon) can account for at least 30% of the infection. These extrahepatic organs may, in fact, be the primary source of natural human transmission of Asian *Taenia* given the eating habits of infected individuals. Early DNA studies (Zarlenga *et al*, 1991) with Asian *Taenia* suggested similarity with *T. saginata*, as does its gross adult morphology, a factor that has clearly confused diagnosis of the parasite in the past. However, a conspicuous rostellum occurs on the scolex of the adult worm and cysticercus of Asian *Taenia* and uterine buds (very short uterine branches without further branching) occur in the gravid proglottides. Furthermore, in contrast to *T. saginata* cysticerci, those of Asian *Taenia* are smaller, they are armed with two rows of (rudimentary) hooklets and have a shorter developmental period (four weeks, compared with 8-18 weeks). In addition to the larval morphology, immunological evidence and the principal intermediate host - the pig - used by the Asian *Taenia* suggest similarity with *T. solium*.

The precise taxonomic status of Asian *Taenia* remained, until recently, unresolved. The new species name of *Taenia asiatica* has recently been proposed, based on morphological criteria, although new molecular evidence (Bowles and McManus, 1994; McManus and Bowles, 1994) does not support the re-classification of Asian *Taenia* as a distinct species. The new study, using similar molecular approaches to those used to categorize *Echinococcus* species and strains referred to above, examined sequence variation in the 5' 28S (large subunit) ribosomal RNA (rRNA) gene, including most of the rapidly evolving D1 region, and the mitochondrial cytochrome c oxidase I (COI) gene for a number of taeniid species so that the genetic distances between *T. saginata, T. solium* and the Asian *Taenia* could be compared with the distances between

accepted species in the family Taeniidae. In general, distinct sequences were obtained with interspecies differences for recognized species within the genus *Taenia* ranging from 5 to 19 nucleotides for the former and 23 to 58 nucleotides for the latter gene. However, the 5ó 28S rRNA sequence for Asian *Taenia* is identical to that of *T. saginata* although the COI sequence differs in nine of the 366 nucleotide positions examined with one of these changes resulting in a coding change. Nucleotide differences between the pairs of taeniid species for the combined COI and 5ó 28S rDNA regions are shown in Table 2. The sequence information indicates *T. saginata* and Asian *Taenia* appear to be more closely related than the other species within the genus.

Table 2

Nucleotide sequence differences between pairs of taxa for the combined mitochondrial cytochrome oxidase I and 5'28 rDNA gene regions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) *T. saginata*									
(2) *Taiwan Taenia*	9								
(3) *T. solium*	55	58							
(4) *T. multiceps*	29	34	43						
(5) *T. ovis*	46	47	46	45					
(6) *T. hydatigena*	56	57	62	55	55				
(7) *T. pisiformis*	58	62	60	56	64	71			
(8) *T. crassiceps*	57	58	67	59	61	53	67		
(9) *T. taeniaefornis*	74	77	75	71	73	74	58	71	
(10) *E. granulosus*	89	92	86	93	78	78	87	76	81

**E.* = *Echinococcus*
Data from Bowles and McManus (1994)

The molecular genetic study fully supports earlier conclusions that the Asian *Taenia* is a genetically distinct entity but is closely related to *T. saginata*. The evidence suggests that classification of the Asian *Taenia* as a subspecies or strain of *T. saginata* is more appropriate than its designation as a new species. Whereas *T. solium* cysticercosis is a serious and chronic disease, *T. saginata* cysticercosis, if it occurs at all, is an extremely rare phenomenon in humans. The molecular data supports the evidence that the Asian *Taenia* is much more closely related to *T. saginata* than to *T. solium,* the public health implication being that the Asian form is unlikely to be an important cause of human cysticercosis. This argument is strengthened by the fact that, despite the high prevalence and intensity of adult Asian *Taenia* reported in Taiwanese aborigines, there appears to be no evidence of significant numbers of cysticercosis or neurocysticercosis cases in these populations.

CONCLUDING COMMENTS

Species and strains within the genera *Echinococcus* and *Taenia* have been defined previously by various identification and characterization methods. Here, the usefulness of some direct DNA approaches in the further characterization of strains and in solving particular systematic problems within these two groups have been assessed. These studies indicate that DNA-based methods are more suitable than conventional approaches for determining the actual genetic identity of a particular organism. One major advantage over alternative methodologies is that DNA sequence comparison allows a direct examination of the genome independently of environmental and ontogenic influences. Another strength of sequence data is that it is absolute rather than relative. Once sequence information is obtained, it is always available for use in comparative studies. Other advantages are that

levels of divergence are relatively easy to quantitate and that the range of nucleotide characters for phylogenetic analysis is practically unlimited. Largely because of the minimal requirements for materials, PCR-based methods, in particular, should prove of especial value in future studies of molecular variation in the taeniid cestodes and other parasitic helminths.

The genetic characterization of the Asian *Taenia* testifies to the power of direct DNA analysis. The identity of the Asian *Taenia* and its relationship with recognized taeniid species was examined quantitatively. Despite a confusing mixture of morphological and epidemiological traits, the Asian *Taenia* was found to be genetically distinct from, but closely related to, one recognized species, *T. saginata*. Since the genetic distance between these two organisms was considerably less than between other pairs of species within the genus, a subspecific designation is recommended (Bowles and McManus, 1994; McManus and Bowles, 1994).

A continuum of genetic variation does not exist among *Echinococcus* isolates as determined by the sequence of evolutionary marker genes. Rather, they can be ordered into distinct, relatively homogeneous, genotypic groups. The sheep, horse and bovine 'strains' of *E. granulosus* were found to be as genetically divergent from one another as recognized species. It is proposed that they represent genuine species rather than intraspecific variants. The Tasmanian sheep and pig strains, although originally defined on the basis of unique epidemiological features or morphology, represent only slight genetic variants of the major groups. Application of genetic approaches in epidemiological studies has revealed the pattern of *E. granulosus* infection in the UK, Australia and Kenya, the fact that the sheep strain is the predominant, if not only, form present in north-west China, and that the bovine strain is infective to humans.

Although genetic studies of organisms within the genus *Echinococcus* have been enlightening, a number of perplexing questions remain unanswered, particularly in regard to the identity and origin of the cervid strain of *E. granulosus*. Resolution of apparent inconsistencies in the cervid strain sequence data may require a more in-depth study of the mitochondrial genome and the rDNA repeats of the nuclear genome. Mitochondrial and ITS1 sequence information from additional *Echinococcus* isolates would undoubtedly contribute to a greater understanding of interrelationships within the genus. As indicated earlier, taxonomic revision of the genus *Echinococcus*, taking into account the new molecular data as well as more traditional information, is now clearly required.

ACKNOWLEDGMENTS

I am indebted to Dr Josephine Bowles who was instrumental in producing much of the molecular data described in this paper. I am also grateful to the Australian National Health and Medical Research Council of Australia and the Tropical Health Program for support of the work described. I would like also to thank the organising committee for kindly inviting me to present this paper at the 2nd Seminar on Food-Borne Parasitic Zoonoses: Current Problems, Epidemiology, Food Safety and Control, December 6-9, 1995, Kosa Hotel, Khon Kaen, Thailand.

REFERENCES

Bowles J. Genetic approaches to the identifi cation and evolutionary study of helminth parasites., University of Queensland 1993. (Unpub lished PhD Thesis)

Bowles J, McManus DP. Molecular variation in *Echino coccus. Acta Tropica* 1993a; 53 291-305.

Bowles J, McManus DP. Rapid discrimination of *Echinococcus* species and strains using a PCR-based RFLP method. *Mole Biochem Parasitol* 1993b; 57, 231-9.

Bowles J, McManus, DP. Genetic characterization of the Asian *Taenia*, a newly described taeniid cestode of humans. *Am J Med Hyg* 1994; 33-44.

Bowles, J, Blair D McManus DP. Genetic variants within the genus *Echinococcus* identified by mitochondrial sequencing. *Molec Biochem Parasitol* 1992a; 54: 165-174.

Bowles J, Blair D, McManus DP. Molecular genetic characterization of the cervid strain (northern form) of *Echinococcus granulosus. Parasitology* 1994; 109: 215-21.

Bowles J, Blair D, McManus DP. A molecular phylogeny of the genus *Echinococcus. Parasitology* 1995; 110: 317-28.

Bowles J, Van Knapen F, McManus DP. Cattle strain of *Echinococcus granulosus* and human infection. *Lancet* 1992b; 339: 1358.

Fan PC. Asian *Taenia saginata*: species or strain. *Southeast Asian J Trop Med Public Health* 1991; 22 (suppl): 245-50.

Hope M, Bowles J, McManus DP. A reconsideration of the *Echinococcus granulosus* strain situation in Australia following RFLP analysis of cystic material. *Int J Parasitol* 1991a; 21: 471-5.

Hope M, Bowles J, Prociv P, McManus DP. A genetic comparison of human and wildlife isolates of *Echinococcus granulosus* in Queensland and the public health implications. *Med J Aust* 1991b; 156: 27-30.

Lymbery AJ. Interbreeding, monophyly and the genetic yardstick: species concepts in parasites. *Parasitol Today* 1992; 8: 208-11.

McManus DP, Bowles J. Asian (Taiwan) *Taenia*: species or strain. Parasitol Today 1994 ; 10: 273-5.

McManus DP, Bowles J. Molecular genetic approaches to parasite identification : their value in applied para sitology and systematics. *Int J Parasitol* (In Press).

McManus DP, Bryant C. Biochemistry, physiology and molecular biology of *Echinococcus*. Thompson RCA, Lymbery AJ, eds. Echinococcus and Hydatid Disease. Wallingford, Oxon, UK: CAB International, 1995: 135-81.

McManus DP, Ding Z, Bowles J. A molecular genetic survey indicates the presence of a single, homogeneous strain of *Echinococcus granulosus* in northwestern China. *Acta Tropica* 1994; 56: 7-14.

McManus DP, Rishi AK. Genetic heterogeneity within *Echinococcus granulosus*: isolates from different hosts and geographical areas characterised with DNA probes. *Parasitology* 1989; 99: 17-29.

McManus DP, Simpso AJG. Identification of the *Echinococcus* (hydatid disease) organisms using cloned DNA markers. *Molec Biochem Parasitol* 1985; 17: 171-8.

McManus DP, Simpson AJG, Rishi AK. Characterization of the hydatid disease organism *Echinococcus granulosus*, from Kenya using cloned DNA markers. In: Geerts S, Kumar V, Brandt J, eds. *Helminth Zoonoses* 1987: pp 29-36.

Rausch RL. Life-cycle patterns and geographic distribution of *Echinococcus* species. In: Thompson RCA, ed. The biology of *Echinococcus* and hydatid disease. London: George Allen and Unwin, 1986: pp 44-80.

Scott JC, McManus DP. The random amplification of polymorphic DNA can discriminate species and strains of *Echinococcus*. *Trop Med Parasitol* 1994; 45: 1-4.

Thompson RCA.Biology and systematics of *Echinococcus*. Thompson RCA, Lymbery AJ, eds. *Echinococcus* and Hydatid Disease. Wallingford, Oxon, UK : CAB International 1995: 1-50.

Thompson RCA, Lymbery AJ. The nature, extent and significance of variation within the genus *Echinococcus*. *Adv Parasitol* 1988; 27: 209-58.

Wachira TM, Bowles J, Zeyhle E, McManus DP. Molecular examination of the sympatry and distribution of sheep and camel strains of *Echinococcus granulosus* in Kenya. *Am J Trop Med Hyg* 1988; 48: 473-9.

Zarlenga DS, McManus DP, Fan PC, Cross JH. Characterization and detection of a newly described Asian taeniid using cloned ribosomal DNA fragments and sequence amplification by the polymerase chain reaction. *Exp Parasitol* 1991; 72: 174-83.