Selection, Recombination and History in a Parasitic Flatworm (*Echinococcus*) Inferred from Nucleotide Sequences

KL Haag, AM Araújo, B Gottstein*, A Zaha**

Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brasil *Institute of Parasitology, University of Berne, Laengass Strasse 122, PO Box 8466, Berne, CH-3001, Switzerland **Departamento de Biotecnologia, Universidade Federal do Rio Grande do Sul, Caixa Postal 15005, 91501-970 Porto Alegre, RS, Brasil

Three species of flatworms from the genus Echinococcus *(*E. granulosus, E. multilocularis *and* E. vogeli*) and four strains of* E. granulosus *(cattle, horse, pig and sheep strains) were analysed by the PCR-SSCP method followed by sequencing, using as targets two non-coding and two coding (one nuclear and one mitochondrial) genomic regions. The sequencing data was used to evaluate hypothesis about the parasite breeding system and the causes of genetic diversification. The calculated recombination parameters suggested that cross-fertilisation was rare in the history of the group. However, the relative rates of substitution in the coding sequences showed that positive selection (instead of purifying selection) drove the evolution of an elastase and neutrophil chemotaxis inhibitor gene (AgB/1). The phylogenetic analyses revealed several ambiguities, indicating that the taxonomic status of the* E. granulosus *horse strain should be revised.*

Key words: *Echinococcus* - parasites - recombination - SSCP - sequencing - phylogeny

Several new insights about the evolution of helminth parasites came out during the last years. *Echinococcus*, a parasite that causes one of the most important and widespread zoonoses, the hydatid disease, is included in this group. The small flatworm uses herbivores as intermediate hosts and carnivores as final hosts. The adult is hermaphrodite and the larval stage (metacestode) is amplified by asexual reproduction.

Four species within the genus are recognised: *E. vogeli* and *E. oligarthrus*, which occur in the neotropical region, *E. multilocularis*, that has an holartic geographic range and *E. granulosus*, that is world-wide distributed. Due to a low intermediate host specificity, *E. granulosus* has been subdivided in several strains, according to the host species used, or to the geographic range of the biological cycle. Some of the evolutionary questions concerning *Echinococcus* are: (1) is the adult mainly self- or cross-fertilising? (2) how do the strains within a species differentiate? (3) what is the true taxonomic status of these strains?

The first question relates to the second one: depending on the breeding system, only one of two modes of strain differentiation can occur. If individual parasites would be mainly selfers (Smyth & Smyth 1964), purifying (negative) selection would quickly eliminate the non-adaptive mutations, due to increased homozygosis. In addition, selfing would lead to a high rate of linkage disequilibrium within parasite populations. In this situation, the genome would be selected as a whole, and not in pieces of recombining DNA. If, on the other hand, populations would undergo outcrossing (Rausch 1967, 1985), free recombination would allow genes to be selected as individual units, and

This work was supported by the Swiss National Science Foundation (project n° 31-45575.95), PADCT/CNPq (Proc. 620081/95-3), EEC (DG XII CI 10284-0), the "Jubiläumsstiftung der Schweitzerischen Lebensversicherungs- und Rentenanstalt für Volksgesundheit und Medizinische Forschung" and the "Sandoz-Stiftung zur Förderung der medizinisch-biologischen Wissenschaften".

This paper reports on research conducted by Karen Luisa Haag as part of her PhD thesis on strain characterisation, genetic variability and breeding systems of *Echinococcus*. It is a result of a collaborative work between the Centro de Biotecnologia (Universidade Federal do Rio Grande do Sul, Brazil) and the Institute of Parasitology (University of Berne, Switzerland). Arnaldo Zaha works primarily with gene organisation and control in *E. granulosus*, Aldo Mellender de Araújo works with evolutionary ecology on a variety of organisms, but mainly insects, and Bruno Gottstein is dealing with molecular aspects of host-parasite interactions in *E. multilocularis.*

+Corresponding author. Fax: +55-51-319.2011. E-mail: haag@dna.cbiot.ufrgs.br

Received 15 June 1998
Accepted 30 July 1998

each genomic sequence would be able to respond singularly to the positive and/or negative selection imposed by the host.

It has also been argued (Thompson et al. 1995, Lymbery & Thompson 1996) that the degree of genetic differentiation of some strains is larger than expected for conspecific groups. Furthermore, if *Echinococcus* is an obligatory selfer, the biological species concept cannot be used to solve the problem (Lymbery 1992, Lymbery & Thompson 1996). In the present study we used the nucleotide sequencing of two coding and two non-coding regions of *Echinococcus* genome to try to elucidate some of the questions above. If parasite populations would have undergone outcrossing during their evolutionary history, we would expect to find recombination among sequences. Additionally, by assessing relative rates of substitution in coding and non-coding regions, it would be possible to evaluate the occurrence of positive and/or negative selection. Finally, genetic distances estimated from those sequences could help to decide whether or not some of the *E. granulosus* strains should be regarded as different species.

MATERIALS AND METHODS

Molecular analyses - Thirty three *E. multilocularis* isolates from different continents (Asia, Europe and North America), hosts (foxes, humans and rodents) and life cycle stages, as well as 110 *E. granulosus* metacestode isolates from different geographic regions (Australia, Europe and Southern Brazil) and strains (bovine, equine, ovine and swine) and one *E. vogeli* isolate were used for genomic DNA extraction and further analyses. DNA extraction was done by standard procedures (McManus & Simpson 1985).

For each isolate, four different targets were amplified by PCR, using primers specific for *Echinococcus* DNA (see procedures in Haag et al. 1997). Two of them were partial intron sequences from an actin gene (ActII - 266 bp) and from an homeobox containing gene (Hbx2 - 331 bp). The other two were coding regions: a partial sequence of a neotrophil chemotaxis inhibitor nuclear gene (AgB/1 - 101 bp) and another partial sequence of the mitochondrial NADH dehydrogenase 1 gene (ND1 - 141 bp).

The nucleotide variation within the PCR products obtained for the four targets was screened by the PCR-SSCP method (see procedures in Haag et al. 1997). Subsequently, two isolates from each SSCP pattern (except in the case of *E. vogeli*) were chosen for direct fluorescence sequencing. For this, the single stranded DNA bands were cut out from the fresh silver-stained SSCP gels, washed and eluted. One ml of the eluted single strands was used for re-amplification with the corresponding primers. These re-amplification products and their respective primers were used for sequencing.

Statistic and phylogenetic analyses - Sequences were aligned by eye (Fig. 1) and the molecular diversity parameters, recombination rates and relative rates of synonymous/non-synonymous substitutions (Ka/Ks) were estimated using DnaSP version 2.0 (Rozas & Rozas 1997). The recombination parameter (C) is calculated based on the average number of nucleotide differences between pairs of sequences (Hudson 1987) and a minimum number of recombination events in the history of the sample (RM) is obtained using a four-gamete test (Hudson & Kaplan 1985).

The genetic distances as well as the neighbour-joining (NJ) trees were estimated with MEGA version 1.0 (Kumar et al. 1993). The parsimony trees were constructed using DNA Penny in Phylip version 3.5c (Felsenstein 1993). For the NJ phylogenetic analysis we used a gamma distance (Kimura 2-parameter model) with gamma parameter a=1. In the parsimony analysis we made a branch-and-bound search to find all most parsimonious trees. Both kinds of trees were constructed using *E. vogeli* as outgroup.

RESULTS

The degree of allele polymorphism found within *E. multilocularis* and within strains of *E. granulosus* was low, as shown in our previous studies (Haag et al. 1997, 1998). Indeed, only one transversion and a single base deletion in the Hbx2 intron occurred among isolates of *E. multilocularis* (Haag et al. 1997). Within the cattle, horse, pig and sheep strains of *E. granulosus* no allele polymorphism was found in the four coding and non-coding loci analysed in the present study.

For this reason, further analyses were done considering the most common variant of *E. multilocularis*, the sequences of the four *E. granulosus* strains and those obtained for the *E. vogeli* isolate {GenBank assession numbers are: AF003748, AF003749, AF003750, AF024661 and AF024662 (Act II); X66818, AF003976, AF003977, AF024663 and AF024664 (Hbx 2); Z26481, Z26482, Z26483, Z26336 and AF024665 (AgB/1); U65748 [ND1 - authors did not provide information about variant sequences published by Bowles and McManus (1993)]}. The molecular diversity parameters estimated from this data set are shown in Table I. The most variable locus was the mitochondrial ND1. Surprisingly, one of the introns (Hbx2) was shown to be very conserved among the referred strains and species, and the AgB/1 nuclear coding region had as much variability as the Act II intron.

A

```
                10                  30                  50
                ,         ,         ,         ,         ,         ,
sheep    TCGTCCAAGACATCAGGTTAGTTGGATAGGTAGGCAGTGTTTCAGCCGCACCGGAACTGG
cattle   ...........................................T..........G...
pig      ...........................................T..........G...
horse    ............................................................
multiloc ............................................................
vogeli   ......................A...G........A......................

                70                  90                  110
                ,         ,         ,         ,         ,         ,
sheep    TACCAACTAGTGGACCAATTTTCTCAAATAAGAGACAGAAATGGTTTGCTTTCATGCACT
cattle   ..........C.........C......A............CA.......C.........
pig      ..........C.........C......A............CA.......C.........
horse    .T.........C.........CTG..T.............CA.......C.........
multiloc .T.........C...T.....CTG..T..........A...CA.......C.........
vogeli   .T.........C.........C....T.............CA.......C.........

                130                 150                 170
                ,         ,         ,         ,         ,         ,
sheep    AAATGTATGGTGAAGAAGTCGGCTTTTCATCTAACTAGATAGGCATGATTAGTGTGGAGA
cattle   ....................................G...................
pig      ......................T.....................................
horse    ..............................G...........A..............
multiloc ..............................................A.............
vogeli   ....A.....................................A.............

                190                 210                 230
                ,         ,         ,         ,         ,         ,
sheep    TCAAGTGCTCTCTTGTAGAGTCGCCATCTGAGGGCAGTCTTTCTATTTTCGCCCTGTGAC
cattle   ......................T...T...............................
pig      ........................T.........GG.C...................
horse    ................A.........T..............................
multiloc ............................................................
vogeli   ...................................................T..GGG.....

                250
                ,         ,
sheep    AACGTACCTATTCCGAAATAATCTTT
cattle   ......................A..
pig      ......................A..
horse    ......................A..
multiloc ........................
vogeli   ................G.........
```

Fig. 1-A: nucleotide sequence alignments of the ActII intron for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

B

```
                    10                  30                  50
                     ,         ,         ,         ,         ,         ,
sheep       CGTCTTAGAAGAGCGATTTGATCGACAAAAGTACCTCAGCAGTGCTGAACGCGCCGAGAT
cattle/pig  ............................................................
horse       ............................................................
multiloc    ............................................................
vogeli      ..................................A....T..............C..

                    70                  90                 110
                     ,         ,         ,         ,         ,         ,
sheep       GTCACGAGACCTGGGGCTCTCTGAAACCCAGGTATGTCACAGCCGATGTCATTAAACATG
cattle/pig  ............................................C.............
horse       ............................................C.............
multiloc    ............................................C.............
vogeli      .....A.A....................................C.............

                   130                 150                 170
                     ,         ,         ,         ,         ,         ,
sheep       GGAAGGGGTGAGAGTAGTTGGAGCGTCACGAAGTGCCAAATTGGGCGCTTGTCAAGCTGC
cattle/pig  ...............T............................................
horse       ............................................................
multiloc    ............................................................
vogeli      ................CA..........................................

                   190                 210                 230
                     ,         ,         ,         ,         ,         ,
sheep       GCCTTTATAACTGTTGAGTGCATCATCACCCATAAAAAATTGGGAGAGAGGGGGGCGGGA
cattle/pig  ............................................................
horse       ..........................................—.................
multiloc    ............................................................
vogeli      ..........................................T.................

                   250                 270                 290
                     ,         ,         ,         ,         ,         ,
sheep       GCGGGTCAAAAGGGTCATCACGGCTCATGCATTAGTAAGATCGTAAAAGGCATGCCTCTA
cattle/pig  ............................................................
horse       ............................................................
multiloc    ................T...........................................
vogeli      ................GGT.......A..............C...................

                   310                 330
                     ,         ,         ,
sheep       ATTATGACCCCCACCACTAGGTGAAAATATG
cattle/pig  ...........................
horse       ...........................
multiloc    .........T.....................
vogeli      ...........................
```

Fig. 1-B: nucleotide sequence alignments of the Hbx2 intron for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

C

```
                 10                  30                  50
                  ,         ,         ,         ,         ,         ,
sheep    CTGGTTGGGGTGGTTACAACAATTATTCATTTTTAAGGTCGGTTCGATGTGCTTTTGGAT
cattle   .A.........A....T..T.G......T..............................
pig      ...........A.......T.G......T..............................
horse    .G.........A....T..T.....C........G.....A...................
multiloc ...........A.......T..A.....G........T..T.................G.
vogeli   .......A...A.......T..C........C..G.....AA....G...........G.

                 70                  90                 110
                  ,         ,         ,         ,         ,         ,
sheep    CTGTTAGGTTTGAGGCTTGTTTTATGTGTGTGGTGATTTTTTGTGCTTTGTGTAGTTGTA
cattle   ................................T.................T......
pig      ................................T........C.....A..CT.....G
horse    ............................C.......................T....C.
multiloc .............A...................A.........T........TAC....
vogeli   ................................A......C..........CT......

                130
                  ,         ,
sheep    GGTATAATTTAATTGATTTTT
cattle   .....................
pig      .....................
horse    ..........GG.........
multiloc ..........G..........
vogeli   ..........GG.........
```

Fig. 1-C: nucleotide sequence alignments of the mitochondrial ND1 for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

D

```
10                  30                  50
                     ,         ,         ,         ,         ,         ,
sheep      AGTGGTTGACCTCTTAAAGGAACTGGAAGAAGTGTTCCAGTTGTTGAGGAAGAAGCTACG
cattle/pig ............................................................
horse      ....................................G.......................
multiloc   .T.........A..........................A....A...............
vogeli     ..........A.G.............................................

                    70                  90
                     ,         ,         ,         ,
sheep      CATGGCACTCAGGTCCCACCTCAGAGGGTTGATTGCTGAAGG
cattle/pig ..C......T.A..............A....G..........
horse      ..C........A..............A....G..........
multiloc   ..C........A..............A....G..........
vogeli     ..C......T.A............A.AA....G..........
```

Fig. 1-D: nucleotide sequence alignments of the AgB/1 for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli.*

TABLE I

Nucleotide diversity (p), theta (q), average number of nucleotide differences (k), number of polymorphic sites (S) and total number of sites (T) of the four non-coding (Act II and Hbx 2) and coding (AgB/1 and ND1) sequences analysed in this study

PARAMETER[a]	Act II	Hbx 2	AgB/1	ND1
p	0.0524	0.0204	0.0559	0.0964
	(0.0001)[b]	(0.0001)	(0.0001)	(0.0002)
q	0.0576	0.0233	0.0618	0.0963
	(0.0008)	(0.0002)	(0.0011)	(0.0023)
k	13.93	6.70	5.70	13.60
S	35	16	13	31
T	266	329	102	141
S/T	0.1316	0.0486	0.1274	0.2198

a: Nei 1987; *b:* Numbers in parentheses are standard deviations.

The recombination parameter (C=4Nc, where c is the recombination rate) among the nuclear sequences was equal to 34.2 (per gene) and 0.0518 (between adjacent sites). The minimum number of recombination events occurring in the history of that sample of sequences was estimated do be Rm=2. Additionally, the relative rates of synonymous and non-synonymous substitutions calculated for the two coding regions showed that, compared to the mitochondrial ND1, the rates of non-synonymous substitutions within AgB/1 were very high (Table II).

As the results of the NJ and parsimony analyses were very similar, we decided to concentrate on the later. A phylogeny obtained by analysing all loci together is shown in Fig. 2. The topology of that tree is in accordance with others, obtained using a larger number of OTUs and other helminths as outgroups (Lymbery 1995). However, the phylogenies constructed for each sequence separately were not congruent. First, most sequences did not provide a single most parsimonious tree: the Hbx2 intron resulted in 15, ND1 in 2 and AgB1 in three equally parsimonious topologies. Second,

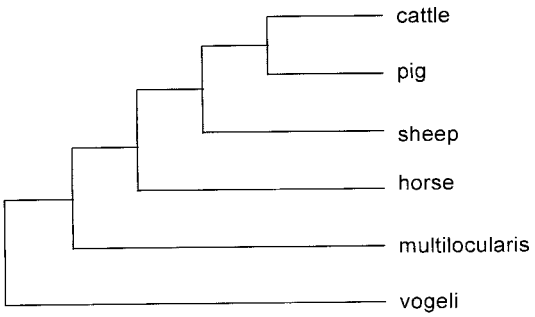

Fig. 2: maximum parsimony phylogenetic tree of *Echinococcus* strains and species obtained using the four coding and non-coding sequences. The tree requires 113 steps (for details, see Materials and Methods).

TABLE II

Relative rates of non-synonymous and synonymous (Ka/Ks) substitutions within ND1 (above diagonal) and AgB1 (below diagonal) coding sequences among the *Echinococcus granulosus* strains, *E. multilocularis* (EM) and *E. volgeli* (EV)

	Sheep	Cattle	Pig	Horse	EM	EV
Sheep		0.20	0.17	0.09	0.13	0.10
Cattle	1.22		0.07	0.06	0.08	0.07
Pig	1.22	0.00		0.05	0.08	0.09
Horse	*	0.31	0.31		0.07	0.04
EM	0.88	0.18	0.18	0.45		0.14
EV	1.48	2.09	2.09	0.90	0.61	

* indeterminacy

ambiguities were found regarding the position of the horse strain: in some instances it is grouped together with the *E. granulosus* strains, and in others it splits before.

A striking result obtained by the genetic distance calculations (Table III) was the high similarity between the cattle and the pig strains. As expected, *E. vogeli* is the most distant group in relation to all other analysed OTUs. The distance values among the other *E. granulosus* strains and between each strain and *E. multilocularis* were quite similar.

DISCUSSION

Previous studies (Lymbery et al. 1997) concluded that cross-fertilisation occurs within *E. granulosus* populations. However, there were also good evidences that outcrossing is not the predominant mating system, since most loci analysed showed monomorphism within strains or large deficiencies of heterozygotes (Lymbery & Thompson 1988, Lymbery et al. 1990, 1997). The results obtained in the present study support those previ-

TABLE III

Jukes-Cantor genetic distances (above diagonal) and their standard deviations (below diagonal) among the *Echinococcus granulosus* strains, *E. multilocularis* (EM) and *E. volgeli* (EV), based on the nuleotide sequences of the four coding and non-coding loci

	Sheep	Cattle	Pig	Horse	EM	EV
Sheep		0.0329	0.0379	0.0392	0.0455	0.0700
Cattle	0.0064		0.0145	0.0317	0.0442	0.0674
Pig	0.0069	0.0042		0.0405	0.0493	0.0700
Horse	0.0070	0.0062	0.0071		0.0392	0.0622
EM	0.0075	0.0074	0.0079	0.0070		0.0635
EV	0.0095	0.0093	0.0095	0.0089	0.0090	

ous findings, suggesting that recombination within the nuclear sequences occurred at least twice during the evolution of the genus. Although the coding and non-coding regions tested here were short, the lack of phylogenetic congruence among the trees constructed for each locus separately could also be due to recombination.

Another explanation for those incongruences is that selection acted independently on each sequence, but this argument could be used only for the coding regions. Indeed, we showed that positive selection did act during the evolution of the AgB/1 gene: most nucleotide replacements found by pairwise comparisons of the sequences were non-synonymous, and the relative rates of non-silent/silent substitutions (Ka/Ks) were greater than one in six out of fifteen comparisons.

Selection was also used to explain the high frequency of heterozygotes found for variant regulatory sequences in populations of *E. granulosus* from the sheep strain. Taken together, all those findings indicate that *Echinococcus* is not an evolutionary dead-end, unable to adapt quickly enough to changing environmental conditions. Nevertheless, it seems that a balance between cross and self-fertilisation was the best solution found by the parasite to keep evolving. It seems that the recombination rates cannot be neither too high, breaking down coadapted gene complexes, nor too low, hindering adaptive changes.

Moreover, the estimated phylogenetic distances and the trees of *Echinococcus* species and strains are in agreement with those reported by Lymbery (1995). The results show that the phylogenetic position of the *E. granulosus* horse strain is ambiguous. For this reason, we agree with the proposal of a taxonomic revision of the genus, based not only on a molecular phylogenetic approach including a larger number of OTUs, but also on other comparative biological data.

REFERENCES

Bowles J, McManus DP 1993. NADH dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus. Internl J Parasitol 23:* 969-972.

Felsenstein J 1993. *Phylip (Phylogeny Inference Package)* version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Haag KL, Zaha A, Araújo AM, Gottstein B 1997. Reduced genetic variability in coding and non-coding regions of *Echinococcus muitilocularis* genome. *Parasitology 115:* 521-530.

Haag KL, Araújo AM, Gottstein B, Siles-Lucas M, Thompson RCA, Zaha A 1998. Breeding system in *Echinococcus granulosus* (Cestoda; Taeniidae); selfing or outcrossing? *Parasitology* (in press).

Hudson RR 1987. Estimating the recombination parameter of a finite population model without selection. *Gen Res 50*: 245-250.

Hudson RR, Kaplan NL 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics 111*: 147-164.

Kumar S, Tamura K, Nei M 1993. *Mega: Molecular Evolutionary Genetics Analysis*, version 1.0. The Pensylvania State University, University Park, PA 16802.

Lymbery AJ 1992. Interbreeding, monophyly and the genetic yardstick: species concepts in parasites. *Parasitol Today 8*: 208-211.

Lymbery AJ 1995. Genetic diversity, genetic differentiation and speciation in the genus *Echinococcus* Rudolphi 1801, p. 51-88. In RCA Thompson, AJ Lymbery (eds), *Echinococcus and Hydatid Disease*, Cab International, Wallingford.

Lymbery AJ, Thompson RCA 1988. Electrophoretic analysis of genetic variation in *Echinococcus granulosus* from domestic hosts in Australia. *Intern J Parasitol 18*: 803-811.

Lymbery AJ, Thompson RCA 1996. Species of *Echinococcus*: pattern and process. *Parasitol Today 12*: 486-491.

Lymbery AJ, Constantine CC, Thompson RCA 1997. Self-fertilization without genomic or population

structuring in a parasitic tapeworm. *Evolution 51*: 289-294.

Lymbery AJ, Thompson RCA, Hobbs RP 1990. Genetic diversity and genetic differentiation in *Echinococcus granulosus* (Batsch, 1786) from domestic and sylvatic hosts on the mainland of Australia. *Parasitology 101*: 283-289.

McManus DP, Simpson AJG 1985. Identification of the *Echinococcus* (hydatid disease) organisms using cloned DNA markers. *Mol Biochem Parasitol 17*: 171-178.

Nei M 1987. *Molecular Evolutionary Genetics*, Columbia University Press, New York, 512 pp.

Rausch RL 1967. A consideration of intraspecific categories in the genus *Echinococcus* Rudolphi,1801 (Cestoda: Taeniidae). *J Parasitol 53*: 484-491.

Rausch RL 1985. Parasitology: retrospect and prospect. *J Parasitol 71*: 139-151.

Rozas J, Rozas R 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Computation Applicated to Biosciences 13*: 307-311.

Smyth JD, Smyth MM 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.

Thompson RCA, Lymbery AJ, Constantine CC 1995. Variation in *Echinococcus*: towards a taxonomic revision of the genus. *Advanc Parasitol 35*: 146-176.