

# High polymorphism in genes encoding antigen B from human infecting strains of *Echinococcus granulosus*

L. KAMENETZKY<sup>1,2</sup>, P. M. MUZULIN<sup>1</sup>, A. M. GUTIERREZ<sup>1</sup>, S. O. ANGEL<sup>2</sup>, A. ZAHA<sup>3</sup>, E. A. GUARNERA<sup>1</sup> and M. C. ROSENZVIT<sup>1\*</sup>

<sup>1</sup>Departamento de Parasitología, Instituto Nacional de Enfermedades Infecciosas ANLIS "Dr. Carlos G. Malbrán" Av. Velez Sarsfield 563, Buenos Aires (1281), Argentina

<sup>2</sup>Instituto Tecnológico de Chascomús (IIB-INTECH)-Conicet/UNSAM, Chascomús, Argentina

<sup>3</sup>Centro de Biotecnología e Departamento de Biología Molecular e Biotecnología, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

(Received 22 April 2005; revised 26 May 2005; accepted 28 May 2005)

## SUMMARY

*Echinococcus granulosus* antigen B (AgB) is encoded by a gene family and is involved in the evasion of the host immune response. *E. granulosus* exists as a number of strains (G1–G10) that differ in biological characteristics. We used PCR-SSCP followed by DNA sequencing to evaluate sequence variation and transcription profile of AgB in 5 *E. granulosus* strains†. Twenty-four genomic sequences were isolated and clustered in 3 groups related to 2 of the 5 reported AgB genes. AgB4 genes were present in almost all strains, whereas AgB2 were present as functional genes exclusively in G1/G2 cluster, and as non-functional genes in G5 and the G6/G7 cluster, suggesting inter-strain variation. The AgB transcription patterns, analysed by RT-PCR, showed that AgB2 and AgB4 genes were transcribed in G1, while only the AgB4 gene was transcribed in G7 strain. Cysts from the same strain or cluster shared more genomic and cDNA variants than cysts from different strain or cluster. The level of nucleotide and deduced amino acid sequence variation observed is higher than that reported so far for coding genes of other helminths. Neutrality was rejected for AgB2 genes. These data show the genetic polymorphism of antigen-coding genes among genetically characterized strains of *E. granulosus*.

Key words: genetic polymorphism, antigen B, *Echinococcus granulosus*, strains, RT-PCR, SSCP.

## INTRODUCTION

The larval stage of *Echinococcus granulosus* is the causative agent of cystic hydatid disease, a major health and economic problem in many countries around the world. This stage is characterized by the formation and growth, in the host internal organs, of a unilocular cyst filled with a hydatid cyst fluid (HCF). HCF contains a complex mixture of proteins of host and parasite origin. One of the most abundant parasite antigens of HCF is the oligomeric lipoprotein antigen B (AgB), which is commonly used in immunodiagnosis of hydatid disease (Lightowlers and Gottstein, 1995; Virginio *et al.* 2003). AgB is involved in the evasion of the immune response due to its ability to inhibit elastase activity and neutrophil chemotaxis (Shepherd, Aitken and McManus, 1991); and to elicit a non-protective Th2 cell response (Rigano *et al.* 2001). It has been shown that

AgB is encoded by a gene family consisting of AgB1 (Shepherd *et al.* 1991), AgB2 (Fernández *et al.* 1996) and AgB3 (Chemale *et al.* 2001). Recently, 2 additional gene loci related to AgB2 and AgB3 genes named respectively AgB4 and AgB5 have been reported (Haag *et al.* 2004). The comparative analysis of the diagnostic potential of antigens encoded by some of these genes showed that the recombinant antigen AgB2 had the best diagnostic performance (Rott *et al.* 2000; Virginio *et al.* 2003).

The species *E. granulosus* comprises a number of intraspecific variants or strains that differ in biological features such as intermediate host specificity, developmental rate and infectivity to humans (Thompson and McManus, 2001; Lavikainen *et al.* 2003). The strains (G1–G10) were named according to their most commonly identified intermediate host and classified by mitochondrial gene sequencing and restriction fragment length polymorphisms. Among others, 5 of these strains were found in humans: sheep (G1), Tasmanian sheep (G2), cattle (G5), camel (G6) and pig (G7) strains (Eckert and Thompson, 1997; Rosenzvit *et al.* 1999; Kamenetzky *et al.* 2002; Turceková *et al.* 2003; Bart *et al.* 2004). Analysis by independent molecular markers allowed the differentiation of 2 groups of strains: G1/G2 cluster and G6/G7 cluster, leaving G5 strain outside but closer to the G6/G7 group

\* Corresponding author: Departamento de Parasitología, Instituto Nacional de Enfermedades Infecciosas ANLIS "Dr. Carlos G. Malbrán" Av. Velez Sarsfield 563, Buenos Aires (1281), Argentina. Fax: +5411 43032382. E-mail: marar@anlis.gov.ar

† Nucleotide sequence data reported in this paper are available in the GenBank™ database under the Accession numbers AY569341–AY569371, AY159484, AY159485, AY608606, AY608607, AY614001.

(Bowles, Blair and McManus, 1995; Lymbery and Thompson, 1996; Rosenzvit *et al.* 2001; Kamenetzky *et al.* 2002; Bartholomei-Santos *et al.* 2003). The extensive sequence differences found among mitochondrial (Bowles, Blair and McManus, 1992; Bowles and McManus, 1993*a*), ribosomal (Bowles and McManus, 1993*b*), housekeeping (Haag *et al.* 1998*a*) and non-coding repetitive genes (Rosenzvit *et al.* 2001) from *E. granulosus* strains suggest that genetic variation could also exist in antigen coding genes. Comparison of excretory/secretory proteins from *E. granulosus* hydatid fluids from several hosts showed some differential patterns (Siles-Lucas and Cuesta-Bandera, 1996). The analysis of sequences related to AgB1 showed variation between parasites with different hosts and geographical origin (Frosch *et al.* 1994). Also, Haag *et al.* (1998*b*) found genetic variability in a partial sequence of AgB1 gene among 4 *E. granulosus* strains. In addition, a high degree of polymorphism of AgB genes was shown in the protoscoleces of a single hydatid cyst (Haag *et al.* 2004). However, the genetic variation and transcription profile of antigen B coding genes in *E. granulosus* strains has not been systematically analysed so far. This could have important implications for both AgB function and its utility in the development of future diagnostic tools. The aim of this study was to determine the extent of variation and transcription profile of AgB encoding genes in several human infecting strains of *E. granulosus*.

#### MATERIALS AND METHODS

##### *DNA and RNA extraction*

Total *E. granulosus* genomic DNA was prepared from fresh, frozen in liquid nitrogen or 70% ethanol preserved protoscoleces of *E. granulosus* by conventional techniques (Maniatis, Fritsch and Sambrook, 1989). Total RNA was prepared from fresh protoscoleces using TRIzol reagent (Gibco, BRL). Each sample was incubated for 30 min at 37 °C with 1 unit of RQ1 RNase-Free DNase (Promega). An *E. granulosus* isolate refers to protoscoleces obtained from a single hydatid cyst.

##### *Analysis of E. granulosus strain*

*E. granulosus* strain was determined as previously described (Rosenzvit *et al.* 2001, Kamenetzky *et al.* 2002). Seven parasite isolates were used for AgB sequence analysis: 2 isolates from G1 (sheep strain), 1 from G2 (Tasmanian sheep strain), 1 from G5 (cattle strain), 1 from G6 (camel strain) and 2 from G7 (pig strain) (Table 1).

##### *Amplification of Antigen B genes*

The set of primers used was the same as those used by Fernández *et al.* (1996), which are expected to

amplify AgB2 and AgB4 sequences. Polymerase chain reaction (PCR) was performed in a final 20 µl volume containing sample DNA (10–50 ng), 200 µM of each dNTP (Amersham Biosciences, UK), 2.5 mM MgCl<sub>2</sub>, 10 pmol of each AgB8/2 specific primers (Fernández *et al.* 1996), and 1 unit of *Thermus aquaticus* DNA polymerase in reaction buffer (Promega, Madison, WI). These primers anneal on 5' and 3'UTR regions close to the ATG and TAA codons respectively. The PCR conditions were as follows: an initial denaturing step (95 °C for 180 s) followed by 35 cycles, 95 °C for 60 s (denaturation), 55 °C for 60 s (annealing), 72 °C for 90 s (extension), and a final extension step (72 °C for 10 min). The specificity and size of the amplification products were assessed by electrophoresis in 1% (w/v) Tris-acetate-EDTA (TAE) agarose gels and stained with ethidium bromide.

##### *Cloning and analysis by single-strand conformation polymorphism (SSCP)*

Amplification products were extracted from the agarose gel (QIAEX II Gel extraction kit, Qiagen) and cloned in T vector (pGem-T Easy Vector System 1 Promega, Madison, WI). Then 20–40 colonies from each isolate were screened for AgB inserts by colony-PCR. Each recombinant colony was grown in a 96-well plate containing 100 µl of LB super broth medium and incubated at 37 °C overnight. The PCR master mix was inoculated with 0.5 µl of each overnight culture. The PCR conditions were as above. Five µl of each PCR product were denatured for 5 min at 94 °C in a 95% formamide, 0.025% xylene cyanole and 0.025% bromophenol blue containing buffer and chilled immediately on ice. The electrophoresis was carried out in 10% acrylamide: bis-acrylamide (49:1) non-denaturing polyacrylamide gels containing 10% of glycerol, at 200 V for 3 h at 4 °C in Tris-borate-EDTA (TBE) buffer. The pattern of bands were visualized by silver staining.

##### *Sequencing and data analysis*

The plasmid inserts were sequenced using the Big Dye Terminator Kit on an ABI 377 sequencer (Applied Biosystems, Foster City, CA) and an Eppendorf Mastercycler gradient 5331 version 1.2 DNA Thermal Cycler. Both strands were sequenced for every clone analysed. DNA sequences were aligned using the program CLUSTAL X 1.81 version. The nucleotide diversity ( $\pi_N$ ), which estimates the average number of substitutions between any two sequences, was determined using the program DNAsp version 3.51 (Rozas and Rozas, 2001). Clustering of AgB variants was done by the parsimony method using the program PAUP\* version 4.0b4a and by UPGMA and Neighbour-joining methods using MEGA 2.1 program (Kumar *et al.*

Table 1. Genomic analysis of *Echinococcus granulosus* antigen B

(Strain, sample source, number of clones analysed per cyst, genomic variants and number of clones with each variant from a total of 249 analysed clones.)

Strain (genotype)	Sample source	Total no. of clones analysed	No. of clones	Variant
Sheep (G1)	DNA cyst 1	40	12	EgB2G1v9*
			12	EgB2G1v8
			10	EgB2G1v4*
			4	EgB2G1v6*
			1	EgB2G1v7*
	DNA cyst 2	30	1	EgB4G1v20 <sup>+</sup>
			17	EgB2G1v4*
			7	EgB2G1v1*
			2	EgB2G1v2
			2	EgB2G1v3*
Tasmanian Sheep (G2)	DNA cyst 3	30	14	EgB2G2v4*
			7	EgB2G2v12
			7	EgB2G2v13*
			1	EgB2G2v10
			1	EgB2G2v11
Pig (G7)	DNA cyst 4	30	18	EgB4G7v20 <sup>+</sup>
			11	EgB2G7v15
			1	EgB4G7v22
	DNA cyst 5	40	29	EgB4G7v20 <sup>+</sup>
			4	EgB4G7v24
			3	EgB2G7v18
			2	EgB2G7v15
			1	EgB4G7v23
Camel (G6)	DNA cyst 6	35	1	EgB2G7v19
			19	EgB4G6v20 <sup>+</sup>
			14	EgB2G6v15
			1	EgB2G6v16
			1	EgB2G6v17
Cattle (G5)	DNA cyst 7	44	32	EgB2G5v14
			12	EgB4G5v21
Total: 249				

\* Genomic variants that code for the same AgB2 protein. + Genomic variants that code for the same AgB4 protein.

2001). Trees were obtained with bootstrap replication of 1000, addition of sequences was at random with 5 replications. Evidence for selection was determined by comparing the rate of synonymous and non-synonymous substitutions using Nei and Gojobori's method (Nei and Gojobori, 1986) with the Jukes-Cantor correction, calculated using the MEGA 2.1 program (Kumar *et al.* 2001). Standard errors were determined by 1000 bootstrap replications. Codon based tests of selection (Fisher's exact test and Z-test) were performed using the same program. Tajima's test of neutrality was performed using the DNAsp Version 3.51.

#### Transcription analysis

For the design of PCR primers for cDNA amplification, all different variants of AgB from G1 and G7 strains obtained were aligned using CLUSTAL X 1.81 version. The sequences of the primers used were 5' GGATCCTTCGTGGCCGTCGTTCAAGC

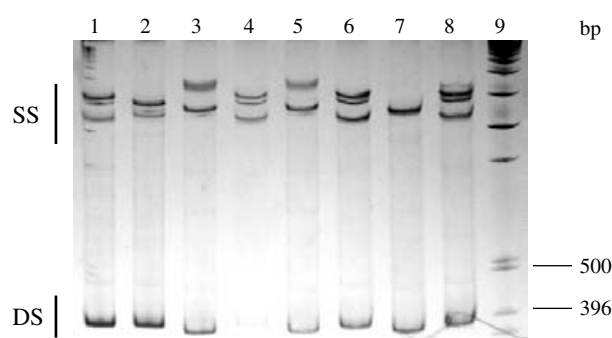


Fig. 1. Representative PCR-SSCP assay of AgB. AgB genes were PCR amplified and cloned. Twenty to forty colonies from each isolate were screened for AgB inserts by colony-PCR. Each PCR product was denatured and loaded in a different lane of a polyacrylamide gel. The observed single-strand patterns are: A (lanes 1, 4, 6 and 8); B (lane 2); C (lanes 3, 5) and D (lane 7). Molecular marker lane 9. Two to five clones of each pattern were sequenced. SS: Single-strand DNA, DS: Double-strand DNA.



3' (primer forward) and 5' GTCGACAAATCAT-GTGTCCCGACGCA 3' (primer reverse). First-strand synthesis was carried out with total RNA of each isolate, 100 pmol of primer reverse, 350 units of M-MLV Reverse Transcriptase (Promega), 2.5 mM of each dNTP (Amersham Biosciences, UK) and 1 unit of RNasin Ribonuclease Inhibitor (Promega) in reaction buffer (Promega). The mix was incubated for 1 h at 35 °C and then inactivated at 95 °C during 5 min. To distinguish between cDNA amplification products from those that would arise from DNA contamination we designed RT-PCR primers that span the intron of the genomic sequence. Genomic DNA contamination would produce a PCR fragment of larger size than the product generated from the cDNA. Furthermore, a control reaction without including reverse transcriptase was performed in each RT-PCR assay. We did not observe bands in this control (data not shown). The PCR reaction, product cloning, PCR-SSCP analysis and sequencing were done as above.

## RESULTS

### Isolation of genomic AgB variants

The primer set derived from the sequence of AgB8/2 (Fernández *et al.* 1996) amplified the expected 390 bp band from the DNA of all isolates analysed. After cloning the PCR products, the inserts of 30 to 40 recombinant clones derived from each isolate were amplified by PCR and analysed by SSCP. Several different patterns were observed (Fig. 1) for all the strains analysed. In order to validate the PCR-SSCP technique, 2–5 independent clones representing each of the observed patterns were analysed by DNA sequencing. No differences in nucleotide sequences were observed in clones sharing the same SSCP pattern. A total of 24 different AgB-related genomic variants was isolated (Table 1). Since these sequences cluster either with AgB2 or AgB4, we named them EgB2 or EgB4 to adopt consistent nomenclature with Haag *et al.* (2004), followed by the strain (G1–7) and a number representing the variant after a 'v'. Hence, all variants with the same number after the 'v' have identical nucleotide sequence. Several AgB-related variants were found in all the strains analysed. Most variants found in G1 strain were related to the AgB2 gene, and a variant related to the AgB4 gene was detected in only 1 out of 70 clones analysed for this strain. In G5, G6 and G7 strains both types of genes were present. However, AgB2-related sequences present in these 3 strains probably

represent pseudogenes (see below). G2 strain showed only variants related to AgB2 gene. The variant v4 found in 41 clones and shared between G1 and G2 strains was identical to the already published AgB2-related sequence (Fernández *et al.* 1996). Isolates of the same strain or cluster shared more variants than those from different strains or clusters. For example, variant v4 was present in all isolates from G1/G2 and v15 in all isolates from G6/G7. The nucleotide differences between variants occurred at the same position in sequences amplified in independent PCR reactions and many of them were shared between different isolates. Most of the mutations were detected in only 2 regions of the gene (i.e. intron and exon 2). Our results are in agreement with data from 2 previous studies (Fernández *et al.* 2002; Haag *et al.* 2004) that reported the presence of AgB sequences with nucleotide substitutions, some of them identical and in the same sites as those reported here. The set of primers used in our work was specific for AgB2 and AgB4. No amplification products were obtained when plasmids containing AgB1, AgB3 or AgB5 genes were used as templates (data not shown).

### Nucleotide sequence analysis of AgB variants

The 3 regions of AgB genes showed different levels of variation. In exon 1, only 5 polymorphic sites were observed, while the intron and exon 2 differed in several nucleotides. The distribution of polymorphisms across the gene is not random (Fig. 2). Phylogenetic analysis was done by the maximum parsimony method, using all the nucleotide sequences found in this work, and related nucleotide sequences from other species of the genus *Echinococcus* and *E. granulosus* cervid strain (G8) (GenBank Accession nos. AY324065 to AY324085). Two related antigenic proteins from *Taenia crassiceps* and *Taenia solium* (Zarlenga, Rhoads and al-Yaman, 1994; Chung *et al.* 1999; Saghir *et al.* 2000) were used as outgroups. As can be seen in Fig. 3, the isolated AgB-genomic sequences could be clustered in 3 groups: one with AgB4-related sequences, another with AgB2-related sequences and the last one containing sequences related to AgB2 that probably represent pseudogenes, named AgB2p (see below). Sequences of the AgB4 group were present in almost all strains, while variants of AgB2 group were present in G1/G2 cluster and those of AgB2p group in G5 strain and G6/G7 cluster. Analysis using UPGMA and Neighbour-joining methods yielded trees which were similar in topology

Fig. 2. Alignment of all AgB genomic variants obtained from *Echinococcus granulosus* G1, G2, G5, G6 and G7 strains. The codes are the same as in Table 1. Dots indicate identical nucleotide with respect to EgB2G1v8. Open box indicates intron sequence, grey box indicates a substitution in splicing site present in AgB2 pseudogenes. Exon 1: positions 1 to 59. Exon 2: positions 128 to 347. Alignment was done with Program Clustal X 1.81.

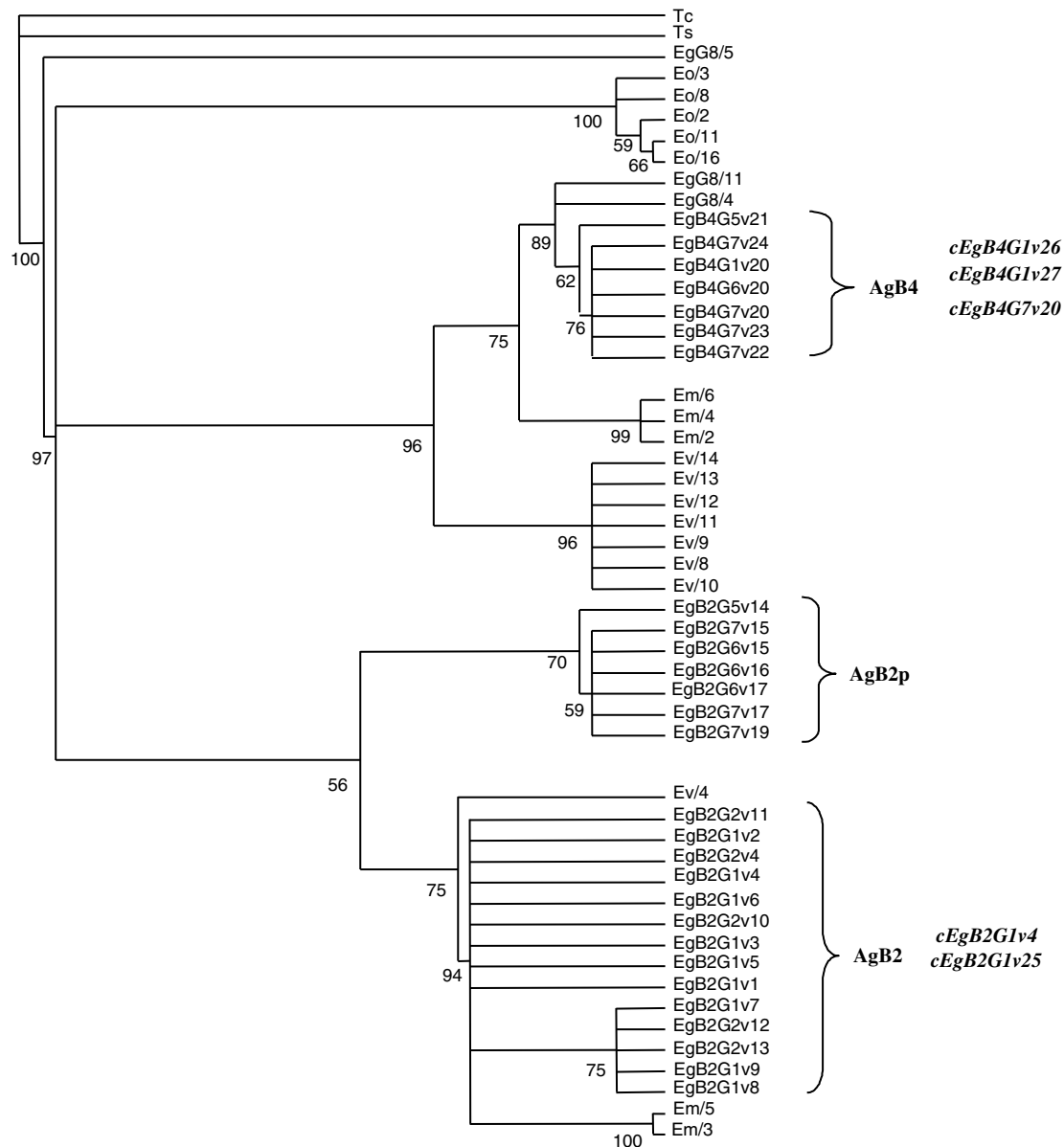


Fig. 3. Phylogenetic tree obtained using maximum parsimony method on data from AgB nucleotide sequences found in *Echinococcus granulosus* G1, G2, G5, G6 and G7 strains and *E. vogeli*, *E. multilocularis*, *E. oligarthrus* and *E. granulosus* G8 strain (GenBank Accession nos. AY324065 to AY324085). Numbers at nodes represent percentage occurrence of clades in 1000 bootstrap replications of the data. The represented taxa are: Ts, *T. solium*; Tc, *T. crassiceps*; Ev, *E. vogeli*; Em, *E. multilocularis*; Eo, *E. oligarthrus*; EgG8, *E. granulosus* cervid strain; followed by the clone number; Eg, *E. granulosus* (the codes are the same as in Table 1). Three groups of sequences (AgB2, AgB2p and AgB4) and cDNA variants related to each group are indicated on the right.

and clustered AgB sequences in the same groups (data not shown).

#### Nucleotide diversity and amino acid identity analysis

Due to the high number of variants found, we were interested in quantifying the variability of the 3 AgB groups: AgB2, AgB2p and AgB4. When each group was separately considered, a high sequence conservation was observed (Table 2), resulting in low nucleotide diversity and high amino acid identity values. However, when 2 groups were considered; e.g., AgB2 and AgB2p groups, reflecting the differences of AgB2-related sequences between

G1/G2 and the other strains; the nucleotide diversity values in the intron and the exon 2 regions were around 10 times higher than before. Also, when AgB2 and AgB4 groups were considered together the amino acid identity observed was only 69% in exon 2, the secreted region of the protein. The presence of these two groups of proteins may confer a wider antigenic repertoire to G1 strain.

#### Amino acid sequence analysis

Only 11 different amino acid sequences could be deduced from the 24 AgB-related nucleotide sequences. This can be explained by 2 facts: (1) many of the

Table 2. Nucleotide diversity ( $\pi_N$ ) with standard deviation (S.D.) and amino acid identity (AI) of AgB subunits from *Echinococcus granulosus* strains

		Exon 1	Intron	Exon 2
AgB2	$\pi_N$ (S.D.)	0·0073 (0·0032)	0·0188 (0·0030)	0·0053 (0·0011)
	AI	90%	—	96%
AgB2p	$\pi_N$ (S.D.)	0·0097 (0·0043)	0·0168 (0·0058)	0·0091 (0·0028)
	AI	—*	—	—*
AgB4	$\pi_N$ (S.D.)	0·0097 (0·0043)	0	0·0053 (0·0019)
	AI	100%	—	94%
AgB2 + AgB2p	$\pi_N$ (S.D.)	0·0135 (0·0028)	0·0416 (0·0042)	0·0499 (0·0074)
	AI	—*	—	—*
AgB2 + AgB4	$\pi_N$ (S.D.)	0·0137 (0·0029)	0·0708 (0·0079)	0·1024 (0·0156)
	AI	90%	—	60%
AgB2p + AgB4	$\pi_N$ (S.D.)	0·0093 (0·0032)	0·0649 (0·0067)	0·0752 (0·0068)
	AI	—*	—	—*

\* AgB2p are pseudogenes.

nucleotide changes were on the third nucleotide of the codon, with no change in the deduced amino acid sequence, and (2) no amino acid sequences were deduced from AgB2p sequences, present in G5, G6 and G7, because of a substitution in position 126, a transversion A/T, that probably generates a non-functional splicing site. This substitution was found in products of independent PCR reactions performed with template DNA from 1 G5, 2 G7 and 1 G6 cysts. If the AgB2p gene is not spliced a premature stop codon is generated (see Fig. 2). This suggests that AgB2p sequences correspond to pseudogenes, which have the same general structure of the functional genes (AgB2) but cannot be translated into a functional protein. This last observation was corroborated by the transcription analysis (see below).

One of the 11 protein variants, corresponding to the AgB2 group was the most represented in the 3 cysts from the G1/G2 cluster while another protein variant, belonging to AgB4 group (see footnotes in Table 1) was the most represented in the 3 cysts from the G6/G7 cluster.

#### Synonymous and non-synonymous substitution rates in AgB groups

In order to test for selection pressures on the AgB genes, we compared the rate of synonymous and non-synonymous substitutions within each group of sequences. The rate of non-synonymous substitutions,  $d_N$ , was lower in AgB2 and AgB4 groups of sequences (Table 3). Also, only for these groups the ratio of  $d_N/d_S$  was less than 1. Particularly in the AgB2 group of sequences, significant departure from neutrality was observed ( $P < 0·05$ , Z-test) suggesting that there is purifying selection against non-synonymous substitutions in these AgB genes. The  $d_N$  value was higher in AgB2p and the ratio of  $d_N/d_S$  was greater than 1 but the neutral hypothesis could

Table 3. Comparison between extents of non-synonymous substitutions ( $d_N$ ) and synonymous substitutions ( $d_S$ ) in AgB sequences

	N <sup>a</sup>	$d_N$ (S.E.) <sup>b</sup>	$d_S$ (S.E.) <sup>c</sup>	$d_N/d_S$
AgB2	13	0·0030 (0·0014)	0·015 (0·006)	0·18*
AgB2p	6	0·013 (0·004)	0·005 (0·005)	2·6
AgB4	5	0·006 (0·003)	0·011 (0·008)	0·54

\* Neutrality hypothesis rejection ( $P < 0·05$ , Z-test, purifying selection).

<sup>a</sup> Number of sequences per group.

<sup>b</sup> Number of non-synonymous mutations per non-synonymous site, calculated using the Nei and Gojobori method, with standard error estimated by 1000 bootstrap replications in parenthesis.

<sup>c</sup> Number of synonymous mutations per synonymous site, calculated as in footnote b.

Codes of AgB groups are the same as figure 3.

not be rejected with Fisher's exact test and Tajima's statistic test.

#### Transcription analysis of AgB variants

To study the transcription pattern of AgB genes in the larval stage of *E. granulosus*, cysts from G1 and G7 strains belonging to the mentioned clusters, were subject to RT-PCR followed by cloning and PCR-SSCP analysis.

Six different cDNA variants were obtained (Table 4), most of them corresponding to the genomic variants isolated in this work. The cDNA variants represented in a higher number of clones, were also the most abundant in the genomic analysis. Transcription of AgB2 variants was observed in both G1 cysts, most of them corresponding to the clone cEgB2G1v4 with the same deduced amino acid sequence as the AgB2 sequence previously reported by Fernández *et al.* (1996). AgB4-related variants were

Table 4. Expression analysis of *Echinococcus granulosus* antigen B

(Strain, genotype, sample source, total number of clones, number of clones with each cDNA variant and variants obtained from a total of 90 clones analysed.)

Strain (genotype)	Sample source	Total no. of clones analysed	No. of clones	Variant (frequency)
Sheep (G1)	RNA cyst 1	20	19	cEgB2G1v4*
			1	cEgB2G1v25
	RNA cyst 8	20	18	cEgB2G1v4*
			1	cEgB4G1v26
			1	cEgB4G1v27
Pig (G7)	RNA cyst 4	20	19	cEgB4G7v20 <sup>+</sup>
			1	cEgB4G7v28 <sup>+</sup>
	RNA cyst 9	30	30	cEgB4G7v20 <sup>+</sup>
		Total: 90		

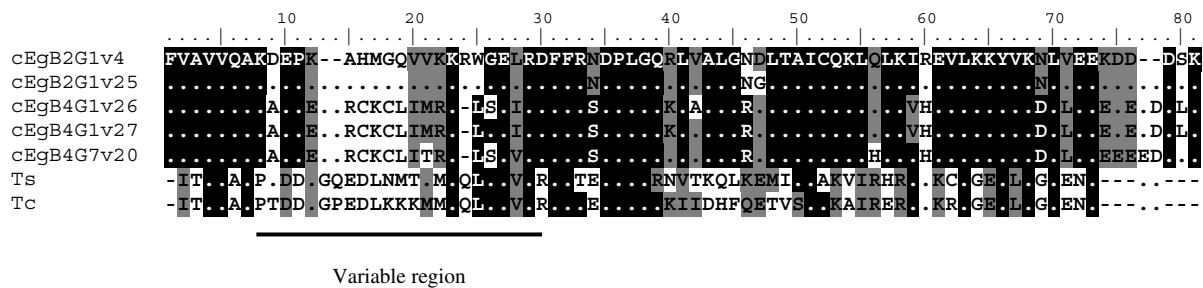
\* cDNA variants that code for the same AgB2 protein. <sup>+</sup>cDNA variants that code for the same AgB4 protein.

Fig. 4. Alignment of all AgB2 cDNA sequences obtained from protoscoleces of *Echinococcus granulosus* G1 and G7 strains. Two immunogenic proteins from the same family are also aligned: Ts, *Taenia solium*; Tc, *Taenia crassiceps*. The N-terminal variable region was underlined (positions 8 to 30). Dots indicate identical amino acids with respect to cEgB2G1v4 variant. Identical residues are presented in black, and conserved substitutions are presented in grey. The codes are the same as in Table 4. Alignment was done with Clustal X 1.81.

found in only 2 clones of 1 G1 cyst. G7 strain showed only AgB4-related cDNA variants, with an amino acid identity of 88% with respect to G1 AgB4-related cDNA sequences and 68% with G1 cDNA AgB2-related variants. According to what was observed in the genomic analysis, no AgB2-related cDNA variants were detected in G7 strain. In agreement with the genomic analysis, only cysts of the same strain shared cDNA variants.

When all cDNA variants were aligned, it was evident that positions 9 to 30 of the AgB N-terminal region are more variable (only 30% identity) than the AgB central region (85% identity, positions 31–68) between cEgB2 and cEgB4 protein variants (Fig. 4). Interestingly, the N-terminal region of exon 2 of AgB subunits was shown to concentrate the epitopes for human antibody recognition (González-Sapienza, Lorenzo and Nieto, 2000). The amino acid identity between these AgB proteins of *E. granulosus* is lower than that observed in related antigens from *T. crassiceps* and *T. solium* (74% identical within the corresponding variable region). Variation in the carboxyl-region from AgB was also observed, AgB4 protein variants from G7 strain have a stretch of 6

glutamate residues between positions 72 and 80, whereas AgB2 proteins variants have only 2.

#### DISCUSSION

The combined use of PCR-SSCP and DNA sequencing allowed the analysis of polymorphism of AgB-related genes in 5 genetically characterized strains of *E. granulosus*. Although the existence of intraspecific variation in *E. granulosus* is well accepted and it has been suggested that the strains differ in their antigenicity, this is the first systematic analysis of genetic variability in an antigen-coding gene among strains of the parasite. Most studies on helminth molecular polymorphisms involved non-coding sequences. Only a few studies aiming to determine intraspecific variation between potential alleles of coding genes were undertaken, mainly in nematodes (reviewed by Maizels and Kurniawan-Atmadja, 2002). Interestingly, the level of amino acid and nucleotide variation observed here was higher than that detected in those studies.

A high degree of polymorphism in AgB family coding genes was found in each *E. granulosus* hydatid



cyst. Our results also showed a substantial level of inter-strain variation in AgB-related genes. This is supported by the finding that AgB2 and AgB4 related genomic sequences were present and expressed at the RNA level in the G1/G2 cluster, while only AgB4-related genomic and cDNA sequences were detected as potentially functional genes in the G5 and G6/G7 clusters. Also, AgB2-related sequences present in the G5 and G6/G7 clusters showed a high degree of nucleotide divergence with respect to AgB2 sequences present in the G1/G2 cluster. Furthermore, cysts from the same strain or same cluster shared more genomic and cDNA variants than cysts from different strains or cluster. Although we can not discard differences in the amplification efficiency or cloning, the strains also seem to differ in the relative proportion of some of the AgB genes: only 1.4% of the clones from G1 strain were related to AgB4 while in the G5, G6 and G7 clusters the percentages were 27.3, 54.2 and 74.3% respectively. The distribution of AgB nucleotide sequences among strains is in agreement with previous studies with mitochondrial or nuclear markers used for strain identification (Bowles *et al.* 1995; Lymbery and Thompson, 1996; Rosenzvit *et al.* 2001; Kamenetzky *et al.* 2002; Bartholomei-Santos *et al.* 2003) which also grouped G1/G2 strains and G6/G7 strains as 2 separate clusters, leaving G5 strain outside but closer to the G6/G7 group.

AgB2 may be subject to purifying selection pressure. The statistical test employed rejected the null hypothesis that AgB2-related sequences are neutral to selection in concordance with data presented by Haag *et al.* (2004). These sequences, present only in sheep strains (G1/G2), may therefore play a role in adaptation to a specific intermediate host. Although several species can act as intermediate hosts for *E. granulosus* sheep strain, sheep are the hosts in which the greater proportion of fertile cysts are formed (Schantz *et al.* 1995; Thompson and McManus, 2002). By contrast, in the AgB2p group of sequences there is not significant selective pressure, which is consistent with the fact that these sequences were not found in the transcription analysis. AgB4-related sequences were observed in all the strains analysed, contrasting with the situation observed for AgB2, which so far was found in only 2 *E. granulosus* strains. It would be interesting to determine whether this group of sequences plays a particular role in infection or maintenance in the host. In the AgB2 and AgB4 groups, many nucleotide variants coding for the same or nearly the same protein were found. However, the transcription analysis showed a low level or absence of transcripts for some of these genomic sequences in the larval stage of *E. granulosus*. It may be possible that during other stages of development or under other conditions, such as different host environments, these AgB variants are transcribed.

Only in the G1 strain AgB2 and AgB4 subunits, sharing only 68% amino acid identity, were found in the genomic and transcription analysis. This is in agreement with studies showing that this strain had relatively higher variability in mitochondrial and nuclear sequences (Haag *et al.* 1998a; Rosenzvit *et al.* 2001; Kamenetzky *et al.* 2002; Bartholomei-Santos *et al.* 2003). Coincidentally, G1 strain has the widest host and geographical range. It would be interesting to compare the AgB expression profiles from G1 strain infecting different individual intermediate hosts as humans or different species of ungulates. This would allow the question of whether the expression of different AgB isoforms is related to the host species to be answered.

The AgB sequence variation was more concentrated in the N-terminal region of the molecule. In fact, the N-terminal variability in subunits from *E. granulosus* strains was higher than the variability observed in the taeniid AgB-related antigens. The deduced amino acid sequences found in the present study were also detected in the oligomeric native AgB (González *et al.* 1996). This suggests that these subunits are effectively translated and may form part of the oligomeric native protein. The native AgB is commonly used in immunodiagnosis of hydatid disease and its N-terminal extension concentrates the immunoreactive B cell epitopes of the native molecule (González-Sapienza *et al.* 2000), therefore it would be important to analyse the performance of each AgB subunit in immunodiagnosis of human hydatid disease.

In conclusion, our results suggest that one of the main antigens of cyst hydatid fluid is highly polymorphic and variable in its transcription profile in human infecting strains of *E. granulosus*.

We would like to thank Drs Wenbao Zhang and Alberto Parra for providing parasite material for analysis and Dr Henrique B. Ferreira for the critical reading of the manuscript. Research was supported by CABBIO, CONICET, Instituto Nacional de Enfermedades Infecciosas, ANLIS 'Dr. Carlos G. Malbrán' and 'Fundación Alberto J. Roemmers'.

#### REFERENCES

- Bart, J. M., Bardonnnet, K., Elfegoun, M. C., Dumon, H., Dia, L., Vuitton, D. A. and Piarroux, R.** (2004). *Echinococcus granulosus* strain typing in North Africa: comparison of eight nuclear and mitochondrial DNA fragments. *Parasitology* **128**, 229–239.
- Bartholomei-Santos, M. L., Heinzelmann, L. S., Oliveira, R. P., Chemale, G., Gutierrez, A. M., Kamenetzky, L., Haag, K. L. and Zaha, A.** (2003). Isolation and characterization of microsatellites from the tapeworm *Echinococcus granulosus*. *Parasitology* **126**, 599–605.
- Bowles, J., Blair, D. and McManus, D. P.** (1992). Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology*. **54**, 165–173.

- Bowles, J., Blair, D. and McManus, D. P.** (1995). A molecular phylogeny of the genus *Echinococcus*. *Parasitology* **110**, 317–328.
- Bowles, J. and McManus, D. P.** (1993a). NADH hydrogenase I gene sequences compared for species and strains of the genus *Echinococcus*. *International Journal for Parasitology* **23**, 969–972.
- Bowles, J. and McManus, D. P.** (1993b). Rapid discrimination of *Echinococcus* species and strains using a polymerase chain reaction-based RFLP method. *Molecular and Biochemical Parasitology* **57**, 231–240.
- Chemale, G., Haag, K. L., Ferreira, H. B. and Zaha, A.** (2001). *Echinococcus granulosus* antigen B is encoded by a gene family. *Molecular and Biochemical Parasitology* **116**, 233–237.
- Chung, J. Y., Bahk, Y. Y., Huh, S., Kang, S. Y., Kong, Y. and Cho, S. Y.** (1999). A recombinant 10-kDa protein of *Taenia solium* metacestodes specific to active neurocysticercosis. *Journal of Infectious Diseases* **180**, 1307–1315.
- Eckert, J. and Thompson, R. C.** (1997). Intraspecific variation of *Echinococcus granulosus* and related species with emphasis on their infectivity to humans. *Acta Tropica* **64**, 19–34.
- Fernández, V., Ferreira, H. B., Fernandez, C., Zaha, A. and Nieto, A.** (1996). Molecular characterization of a novel 8 kDa subunit of *Echinococcus granulosus* antigen B. *Molecular and Biochemical Parasitology* **77**, 247–250.
- Fernández, C., Gregory, W. F., Loke, P. and Maizels, R. M.** (2002). Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Molecular and Biochemical Parasitology* **122**, 171–180.
- Frosch, P., Hartmann, M., Muhlschlegel, F. and Frosch, M.** (1994). Sequence heterogeneity of the echinococcal antigen B. *Molecular and Biochemical Parasitology* **64**, 171–175.
- González, G., Nieto, A., Fernandez, C., Orn, A., Wernstedt, C. and Hellman, U.** (1996). Two different 8 kDa monomers are involved in the oligomeric organization of the native *Echinococcus granulosus* antigen B. *Parasite Immunology* **18**, 587–596.
- González-Sapienza, G., Lorenzo, C. and Nieto, A.** (2000). Improved immunodiagnosis of cystic hydatid disease by using synthetic peptide with higher diagnostic value than that of its parent protein, *Echinococcus granulosus* Antigen B. *Journal of Clinical Microbiology* **38**, 3979–3983.
- Haag, K. L., Araujo, A. M., Gottstein, B., Siles-Lucas, M., Thompson, R. C. and Zaha, A.** (1998a). Breeding systems in *Echinococcus granulosus* (Cestoda, Taeniidae): selfing or outcrossing? *Parasitology* **118**, 63–71.
- Haag, K. L., Araujo, A. M., Gottstein, B. and Zaha, A.** (1998b). Selection, recombination and history in a parasitic flatworm (*Echinococcus*) inferred from nucleotide sequences. *Memorias do Instituto Oswaldo Cruz* **93**, 695–702.
- Haag, K. L., Alves-Junior, L., Zaha, A. and Ayala, F. J.** (2004). Contingent, non-neutral evolution in a multicellular parasite: natural selection and gene conversion in the *Echinococcus granulosus* antigen B gene family. *Gene* **333**, 157–167.
- Kamenetzky, L., Gutierrez, A. M., Canova, S. G., Haag, K. L., Guarnera, E. A., Parra, A., García, G. E. and Rosenzvit, M. C.** (2002). Several strains of *Echinococcus granulosus* infect livestock and humans in Argentina. *Infection Genetics and Evolution* **2**, 129–136.
- Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M.** (2001). MEGA2: Molecular Evolutionary Analysis software. *Bioinformatics* **17**, 1244–1245.
- 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.
- Lightowlers, M. W. and Gottstein, B.** (1995). Echinococcosis/hydatidosis: antigens, immunological and molecular diagnosis. In *Echinococcus and Hydatid Disease* (ed. Thompson, R. C. A. and Lymbery, A. J.), pp. 355–410. CAB International, Wallingford, Oxon, UK.
- Lymbery, A. J. and Thompson, R. C.** (1996). Species of *Echinococcus*: pattern and process. *Parasitology Today* **12**, 486–491.
- Maizels, R. M. and Kurniawan-Atmadja, A.** (2002). Variation and polymorphism in helminth parasites. *Parasitology* **125**, S25–S37.
- Maniatis, T., Fritsch, E. F. and Sambrook, J.** (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
- Nei, M. and Gojobori, T.** (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.
- Rigano, R., Profumo, E., Bruschi, F., Carulli, G., Azzara, A., Ioppolo, S., Buttari, B., Ortona, E., Margutti, P., Teggi, A. and Siracusano, A.** (2001). Modulation of human immune response by *Echinococcus granulosus* antigen B and its possible role in evading host defenses. *Infection and Immunity* **69**, 288–296.
- Rosenzvit, M. C., Zhang, L.-H., Kamenetzky, L., Canova, S. G., Guarnera, E. A. and McManus, D. P.** (1999). Genetic variation and epidemiology of *Echinococcus granulosus* in Argentina. *Parasitology* **118**, 523–530.
- Rosenzvit, M. C., Canova, S. G., Kamenetzky, L. and Guarnera, E. A.** (2001). *Echinococcus granulosus*: intraspecific genetic variation assessed by a DNA repetitive element. *Parasitology* **123**, 381–388.
- Rott, M. B., Fernandez, V., Farias, S., Ceni, J., Ferreira, H. B., Haag, K. L. and Zaha, A.** (2000). Comparative analysis of two different subunits of antigen B from *Echinococcus granulosus*: gene sequences, expression in *Escherichia coli* and serological evaluation. *Acta Tropica* **75**, 331–340.
- Rozas, J. and Rozas, R.** (2001). DnaSP: DNA Sequence Polymorphism. Distributed by the authors, Department of Genetics, University of Barcelona, available at <http://www.bio.ub.es/~julio/DnaSP.html>.
- Saghir, N., Conde, P. J., Brophy, P. M. and Barret, J.** (2000). A new diagnostic tool for neurocysticercosis is a member of a cestode specific hydrophobic ligand binding protein family. *FEBS Letters* **487**, 181–184.

- Schantz, P. M., Chai, J., Craig, P. S., Eckert, J., Jenkins, D. J., Macpherson, C. N. L. and Thakur, A.** (1995). Epidemiology and control of hydatid disease. In *Echinococcus and Hydatid Disease* (ed. Thompson, R. C. A. and Lymbery, A. J.), pp. 233–331. CAB International, Wallingford, Oxon, UK.
- Shepherd, J. C., Aitken, A. and McManus, D.** (1991). A protein secreted *in vivo* by *Echinococcus granulosus* inhibits elastase activity and neutrophil chemotaxis. *Molecular and Biochemical Parasitology* **44**, 81–90.
- Siles-Lucas, M. and Cuesta-Bandera, C.** (1996). *Echinococcus granulosus* in Spain: strain differentiation by SDS-PAGE of somatic and excretory/secretory proteins. *Journal of Helminthology* **70**, 253–257.
- Thompson, R. C. A. and McManus, D. P.** (2001). Aetiology: parasites and life-cycles. In *World Health Organization/OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern* (ed. Eckert, J., Gemmell, M. A., Meslin, F.-X. and Pawlowski, Z. S.), pp. 1–19. World Organization for Animal Health, Paris, France.
- Thompson, R. C. and McManus, D. P.** (2002). Towards a taxonomic revision of the genus *Echinococcus*. *Trends in Parasitology* **18**, 452–457.
- Turcekova, L., Snabel, V., D'Amelio, S., Busi, M. and Dubinsky, P.** (2003). Morphological and genetic characterization of *Echinococcus granulosus* in the Slovak Republic. *Acta Tropica* **85**, 223–229.
- Virginio, V. G., Hernandez, A., Rott, M. B., Monteiro, K. M., Zandonai, A. F., Nieto, A., Zaha, A. and Ferreira, H. B.** (2003). A set of recombinant antigens from *Echinococcus granulosus* with potential for use in the immunodiagnosis of human cystic hydatid disease. *Clinical and Experimental Immunology* **132**, 309–315.
- Zarlenga, D. S., Rhoads, M. L. and al-Yaman, F. M.** (1994). A *Taenia crassiceps* cDNA sequence encoding a putative immunodiagnostic antigen for bovine cysticercosis. *Molecular and Biochemical Parasitology* **67**, 215–223.