Title
Contingent, non-neutral evolution in a multicellular parasite: natural selection and gene conversion in the Echinococcus granulosus antigen B gene family

Journal
Gene, 333(SUPPL.)

ISSN
0378-1119

Authors
Haag, KL
Alves-Junior, L
Zaha, A
et al.

Publication Date
2004-05-01

DOI
10.1016/j.gene.2004.02.027

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 333 (2004) 157–167
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Contingent, non-neutral evolution in a multicellular parasite: natural selection and gene conversion in the *Echinococcus granulosus* antigen B gene family

K.L. Haag[a,b,*], L. Alves-Junior[c], A. Zaha[d], F.J. Ayala[b]

[a] *Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil*
[b] *Department of Ecology and Evolutionary Biology, University of California, Irvine CA, USA*
[c] *Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil*
[d] *Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil*

Received 14 January 2004; received in revised form 4 February 2004; accepted 12 February 2004

Available online 24 April 2004

Abstract

Recent studies have demonstrated that the *Echinococcus granulosus* antigen B (AgB) interferes with the intermediate hosts' immune response and is encoded by a multigene family. The number of members within the family is still uncertain, but there are several evidences of a large genetic variability. The *E. granulosus AgB* genomic sequences available in nucleotide databases can be grouped into four clades, corresponding to genes *EgAgB1*, *EgAgB2*, *EgAgB3* and *EgAgB4*. In the present study, we use PCR amplifications followed by cloning and sequencing to evaluate the genetic variability for AgB isoforms. Two pairs of primers were independently used for PCR amplification. Both PCR reactions from each of three isolated protoscolex (larvae) were cloned in a plasmid vector and the plasmid inserts of 30 colonies from each cloning experiment were sequenced. Using phylogenetic tools, the 113 *EgAgB* clones are classified as follows: 25 are related to *EgAgB1*, 24 to *EgAgB2*, 9 to *EgAgB3* and 39 to *EgAgB4*. The remaining 16 clones form a separate cluster, which we name *EgAgB5*, more closely related to *EgAgB3* than to any of the other genes. Within each gene group, a number of variant sequences occur, which differ from one another by one or few nucleotides. One *EgAgB3* clone has a premature stop codon (pseudogene) and an *EgAgB2* clone lacks the region corresponding to the intron. The overall variation cannot be explained by differences among the asexual protoscoleces, or by experimental artifacts. Using *Echinococcuss* AgB genes from other species/strains as outgroups, neutrality is rejected for *EgAgB2*, and balancing selection is detected for *EgAgB5*, which also seems to be involved in gene conversion. We suggest that *EgAgB1–EgAgB5* represent a family of contingency genes, that is, genes that are variably expressed, so that some but not others are expressed in each individual parasite. Contingency genes are common in parasitic protozoa and other microparasites, but the *EgAgB* family is the first set identified in a multicellular parasite.

Keywords: Adaptive molecular evolution; Somatic hypermutation; Immune response evasion; Balancing selection

1. Introduction

Parasitic organisms require flexibility to escape from the defenses of their hosts. The antagonistic relationships of immunity and evasion mechanisms between parasites and their hosts generate frequency-dependent selection. The immune system of the host acquires variability through somatic hypermutation, recombination and clonal selection (Kinoshita and Honjo, 2001; Diaz and Casali, 2002). The parasite seeks to evade immunity by various mechanisms, which in parasitic protozoa include switching antigen surface molecules encoded by members of gene families that

Abbreviations: AgB, antigen B; *EgAgB*, *Echinococcus granulosus* antigen B gene; PCR, Polymerase Chain Reaction; RT-PCR, Reverse Transcriptase-Polymerase Chain Reaction; bp, base pairs; kDa, Daltons × 1000; GI, Global Inner; PI, Pairwise Inner; Ka/Ks, average ratio of non-synonymous to synonymous substitutions; EST, Expressed Sequence Tag; ES, Expression Site; VSG, Variant Surface Glycoprotein; V(D)J, Variable Diversity Junction; RAG, Recombinase Activity Gene; APL, Altered Peptide Ligand.

* Corresponding author. Departamento de Genética, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500 Prédio 43323, Caixa Postal 15053, Porto Alegre, RS, CEP 91501-970, Brazil. Tel.: +55-51-3316-6717; fax: +55-51-3316-7311.
E-mail address: karen.haag@ufrgs.br (K.L. Haag).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.02.027

originate by gene duplication and are typically tandemly arranged (Borst, 2002). Hypermutation has not yet been confirmed for parasite surface antigen genes, but recombination and positive selection are known to have important roles in their evolution (Frank, 2002).

The molecular evolutionary process behind the variability of parasite antigens is "contingent"; that is, antigens are variably expressed, so that some but not others are expressed in each individual parasite. The variably expressed genes encoding molecules related to immune response evasion or infectivity have been, accordingly, called "contingency genes" (Moxon et al., 1994). Contingency genes have been described in many species of microparasites (bacteria, fungi and protozoans) and generally comprise large families that frequently contain repeated motifs (including microsatellites), which facilitate recombination. Moreover, contingency gene families from protozoan parasites such as the *var* family from *Plasmodium falciparum* and the expression sites encoding the variant surface glycoprotein (VSG) from African trypanosomes have subtelomeric localization. Located close to telomeres, these genes are prone to ectopic recombination and replication errors (Mefford and Trask, 2002; Barry et al., 2003), which result in genetic variation. Contingency genes have not previously been identified in multicellular parasites, but the recently described *EgAgB* multigene family from the helminth *Echinococcus granulosus* seems to evolve contingently.

E. granulosus adults are very small (around 2 mm) tapeworms living in the intestine of dogs. The parasite also requires an intermediate host, represented by several domestic and wild herbivores, as well as humans, where it causes the cystic hydatid disease. During the larval stage (metacestode), the parasite forms a cyst filled with liquid (hydatid fluid), surrounded by three membrane layers. The innermost, the germinal layer, is capable of asexually originating thousands of larvae (protoscoleces). Each single protoscolex can differentiate into a mature adult worm, if ingested by a dog, or in a secondary cyst, if released into the peritoneal cavity of its intermediate host.

The hydatid fluid is a complex mixture of parasite and host molecules. About 10% of the hydatid fluid protein fraction is composed by a polymeric thermo-stable lipoprotein (Musiani et al., 1978), the antigen B (AgB), whose role in the immune response evasion is suggested by several lines of evidence. First, the antigen inhibits neutrophil chemotaxis and elastases in vitro (Shepherd et al., 1991); second, it elicits a non-protective Th2 immune response (Riganú et al., 2001); and, finally, it induces immune cell apopoptosis in patients with hydatid disease (Riganú et al., 2002). The native antigen, built by subunits of 8 kDa, is highly immunoreactive. Each subunit is encoded by a small (around 400 bp) open reading frame, interrupted by an intron, which usually contains repeated CCT motifs. The first exon encodes a signal peptide, while the second exon encodes the secreted protein. The repeated CCT motifs presumably facilitate intergenic recombination between the signal and encoding exons from different members of the *EgAgB* gene family (see below), which encode the antigenic protein.

Four different AgB-related genes have been previously identified: *EgAgB1* (Frosch et al., 1994), *EgAgB2* (Fernández et al., 1996), *EgAgB3* (Chemale et al., 2001) and *EgAgB4* (Arend et al., 2002). Their expression in the metacestode has been confirmed by isolation of the respective cDNAs. Each class of mRNA represents a different AgB isoform, but further experimental evidence is still required to ascertain whether distinct isoforms actually encode particular subunits of the polymeric native antigen. By comparing sequences obtained from isolates of different strains, we have previously shown an excess of non-synonymous substitutions in the second *EgAgB1* exon (Haag et al., 1998). Since *E. granulosus* strains are intermediate host species-specific, we have suggested that *EgAgB1* is being "shaped" by the distinct immune systems. Furthermore, we have found inside a single metacestode a large number of slightly different *EgAgB2* transcripts (Arend et al., 2002), raising the possibility that selection by the immune system might also act on the level of protoscolex populations within metacestodes from an individual host. The aim of the present study is to evaluate the size (number of gene loci) of the *EgAgB* gene family and to search for *AgB* genetic variability within each gene locus (and the mechanisms generating this variability). We also seek to ascertain whether the *EgAgB* genetic variability observed in large pools of protoscoleces (Arend et al., 2002) can also be found among individual protoscoleces isolated from a single cyst.

2. Materials and methods

2.1. Southern blot

Total genomic DNA of a single hydatid cyst was digested with six restriction enzymes (*Acc*II, *Eco*RI, *Pst*I, *Alu*I, *Rsa*I and *Hae*III), with conditions as recommended by the manufacturer. Ten micrograms of digested DNA were ran in a 0.8% agarose gel, transferred to a positively charged nylon membrane (Amersham Biosciences) and immobilized at 80 °C for 2 h. The fragments were hybridized with PCR products obtained from the same genomic DNA, with the two pairs of primers used in single protoscolex experiments (B1F-B1R and B2F-B2R; see below). The PCR fragments were radioactively labeled with α-^{32} P-dATP by random priming (Gibco). Hybridization was performed at 42 °C for 14 h, the excess of probe was removed as previously described (Chemale et al., 2001) and membranes were exposed to an X-ray-sensitive film (Kodak Omat) using an intensifying screen at −80 °C for 2 days.

2.2. PCR from individual protoscoleces

Two pairs of PCR primers were used to amplify the four classes of AgB genes. Primers B1F (TCGCTCTGGCTC-

TCGTCTCA) and B1R (GTAGATGGTTTATTGAGCAA) usually amplify genes *EgAgB1* and *EgAgB3*, while primers B2F (TTGCTCTCGTGGCTTTCGTG) and B2R (GTGTCCCGACGCATGACTTA) anneal to *EgAgB2* and *EgAgB4* gene sequences. Six protoscoleces from a single metacestode were prepared as follows. First, the whole content of a bovine hydatid cyst was collected by suction and the protoscoleces were washed three times with 1 × PBS (2.7 mM KCl, 2.0 mM KH_2PO_4, 0.137 mM NaCl, 10 mM $NaAPO_4 \cdot 7H_2O$) followed by sedimentation. An aliquot of this sediment was further diluted in sterile water on a Petri dish and individual protoscoleces were pipetted under the microscope with 5 μl of water. As negative controls, two aliquots of 5 μl of water without any protoscolex were handled exactly as the individual protoscolex samples. In a final volume of 20 μl, six single protoscoleces were incubated at 50 °C with proteinase K (0.1 mg/ml) for 2 h. After a further incubation at 95 °C for 30 min, the tube content was briefly centrifuged and a PCR mix was added to the lysate. PCR reactions were done with the ExTaq polymerase kit (Takara) for reduced incorporation errors. Reagents were used at concentrations recommended by the manufacturer, with 20 pmol of each primer in a final volume of 50 μl. Three protoscolex lysates served as templates for B1 and the other three for B2 primer pairs. Amplification was performed with 30 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, followed by a final denaturing step at 94 °C for 30 s and extension at 72 °C for 10 min on a Mastercycle Gradient thermocycler (Eppendorf). PCR reactions were maintained at 4 °C until used.

2.3. Cloning and sequencing

All amplicons were purified from solutions (Qiagen) and checked for DNA concentration and molecular weight by agarose gel electrophoresis (Fig. 1). One microliter of each purified amplification product was cloned in a TOPO TA vector for sequencing (Invitrogen). The recombinant plasmids from 30 colonies per transformation (totalizing 180 colonies) were purified with columns (Qiagen) and sequenced using M13 (− 20) forward and M13 reverse primers (Invitrogen). Both strands were sequenced for every recombinant plasmid. Sequencing was performed with the Big Dye Version 2.0 kit on an ABI 377 sequencer (Applied Biosystems).

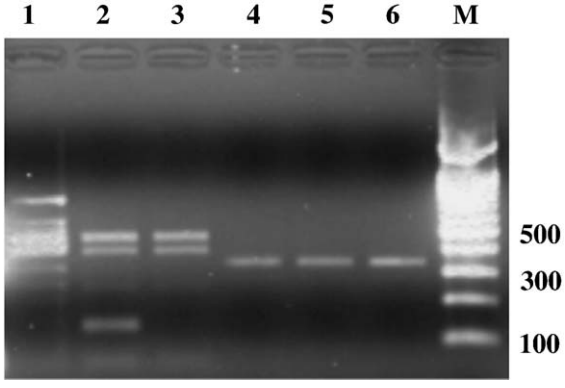

Fig. 1. Agarose gel electrophoresis of single protoscolex PCR products used for cloning. Lanes 1 to 3 correspond to amplifications with primers B1F and B1R, performed on protoscoleces 1 to 3; lanes 4 to 6 were obtained by the amplification of protoscoleces 4 to 6 with primers B2F and B2R; lane 7 is a 100-bp DNA ladder.

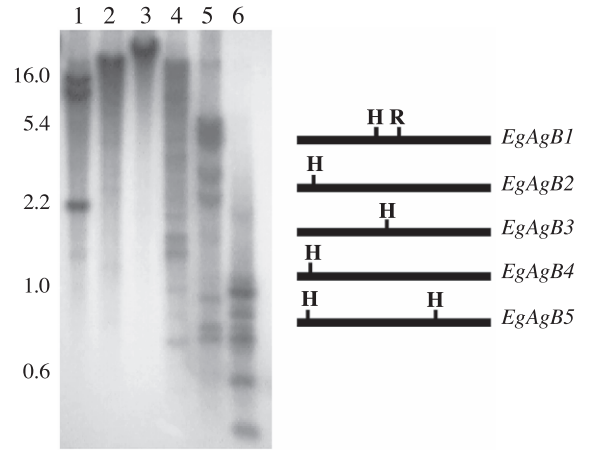

Fig. 2. (Left) Genomic Southern blot of *E. granulosus* DNA digested with *Eco*RI (1), *Pst*I (2), *Alu*I (3), *Rsa*I (4), *Acc*I (5) and *Hae*III (6), and probed with a pool of PCR products obtained with the B1F-B1R and B2F-B2R primer pairs. Numbers on the left indicate the molecular weight (kb) of some major bands. (Right) Restriction map of the five *EgAgB* genes obtained with *Hae*III (H) and *Rsa*I (R); the other enzymes do not have restriction sites within these genes.

2.4. Data analyses

Sequence quality was evaluated by visual inspection of both strands' chromatograms and multiple alignments were generated with the Lasergene software (Dnastar). Phylogenies were constructed by Neighbor Joining, based on the number of different nucleotide sites between pairs of sequences, and tested by bootstrap with 1000 replicates using Mega (Kumar et al., 2001). We used the TCS Version 1.13 software (Clement et al., 2000) to estimate gene genealogies by statistical parsimony and GENECONV Version 1.81 (Sawyer, 1999) to search for gene conversion events. Neutrality tests, and numbers of synonymous and non-synonymous substitutions were calculated with DnaSP Version 3.53 (Rozas and Rozas, 2001). Tests of heterogeneity in the ratio of polymorphic to fixed differences were performed according to McDonald (1998) with DNA Slider (http://udel.edu/~mcdonald/index.html).

3. Results

3.1. Diversity of the EgAgB gene family

Southern experiments indicate that there are several *EgAgB* copies, although the exact number is not immediately apparent (Fig. 2). The digestions with frequent-cutter

enzymes (*Rsa*I, *Acc*II and *Hae*III; lanes 4, 5 and 6 in Fig. 2) indicate the presence of at least seven *EgAgB* gene copies in the *E. granulosus* genome.

From the 180 sequenced PCR-derived clones, 67 were unspecific (from these, 27 were originated from protoscolex 2, which shows a lower than expected molecular weight band in Fig. 1). A phylogeny of the remaining 113 clones (Fig. 3) shows five major clusters, each consisting of nearly identical sequences, which we refer to as "alleles" (but see below) of five different genes. Twenty-five clones are related to *EgAgB1* (alleles M1–M7), 24 to *EgAgB2* (K1–K12) and 39 to *EgAgB4* (A1–A8). The remaining 25 clones are related to *EgAgB3*, but are distributed in two clades (Q1–Q8 and R1–R5), which differ from each other almost as much as the *EgAgB2* (K) differ from the *EgAgB4* (A) sequences (Fig. 3). The differences between the Q and R clades include several nucleotide substitutions in both exons (some of which, in the encoding exon, account for amino acid replacements) and the insertion of a number of CCT repeats inside the intron (Fig. 4). We infer that the Q clade

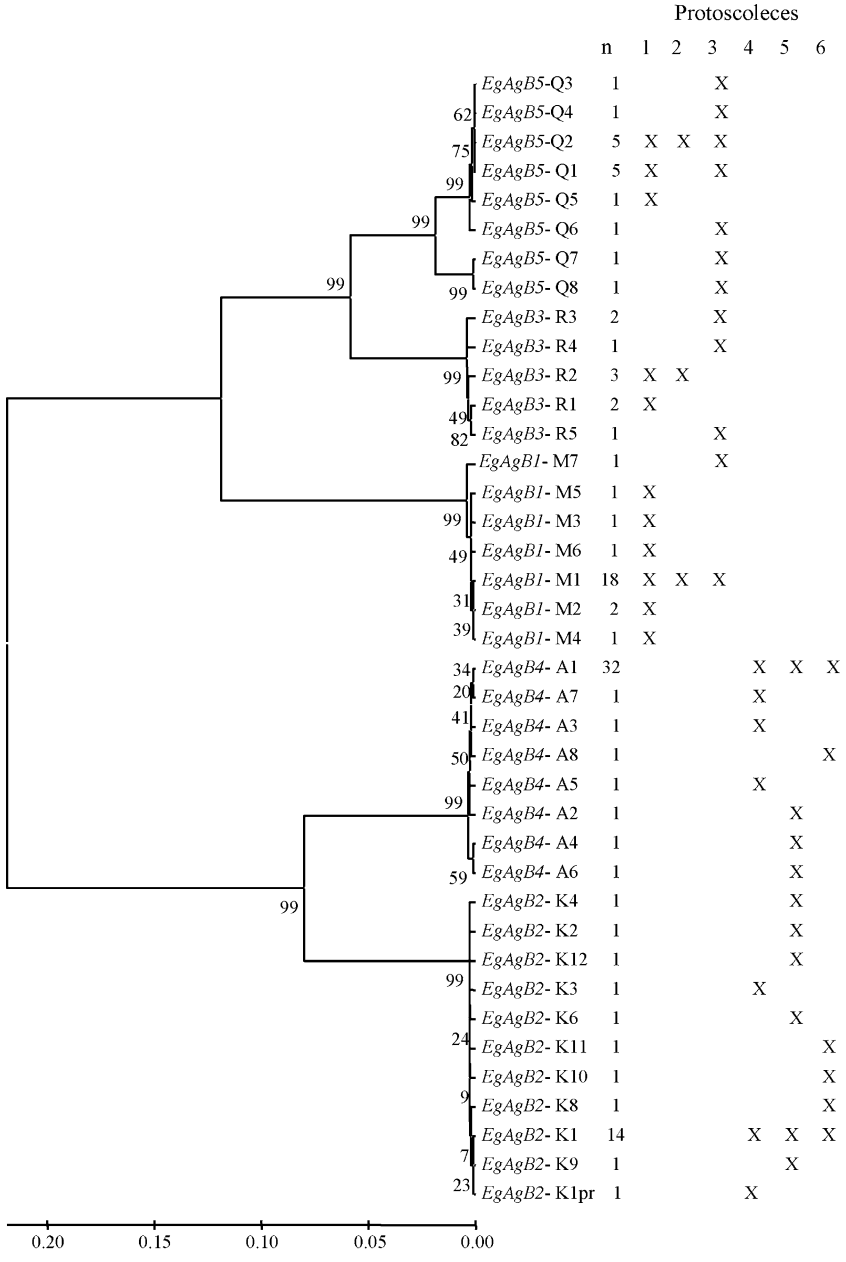

Fig. 3. Neighbor-joining phylogenetic tree of the *EgAgB* sequences cloned from PCR products of single protoscoleces. For each gene, variant sequences are referred as alleles, which are indicated by a letter and a number following the genes' name. The presence and the number of each variant in the six protoscoleces is indicated. Protoscoleces 1–3 were amplified with primers B1F and B1R and 4–6 with primers B2F and B2R. Bootstrap values are based on 1000 replicates; branch lengths correspond to the number of nucleotide differences.

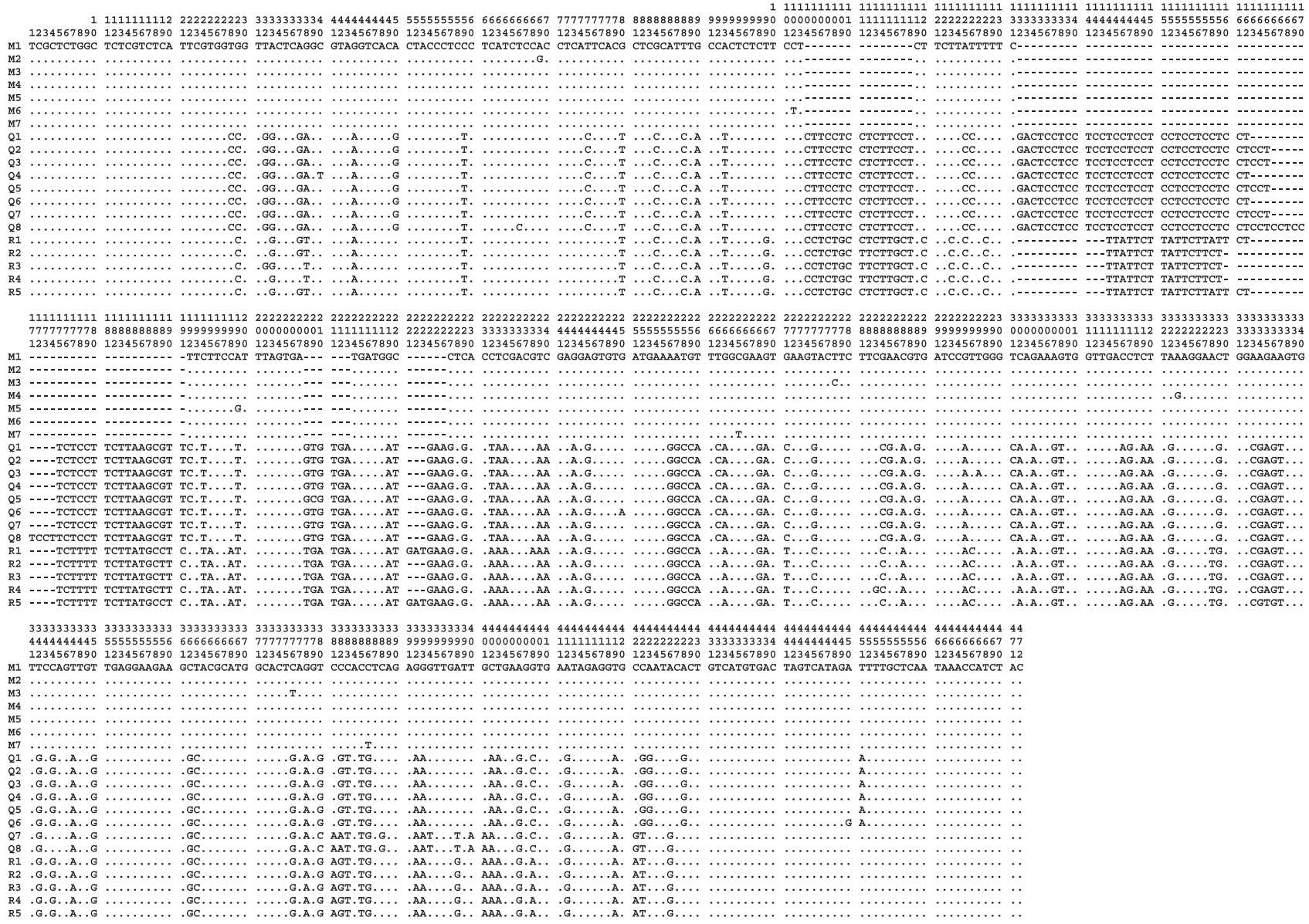

Fig. 4. Alignment of the cloned sequences obtained from single protoscolex PCRs with primers B1F and B1R. M1–M7 correspond to *EgAgB1*, Q1–Q8 to *EgAgB5* and R1–R5 to *EgAgB3*.

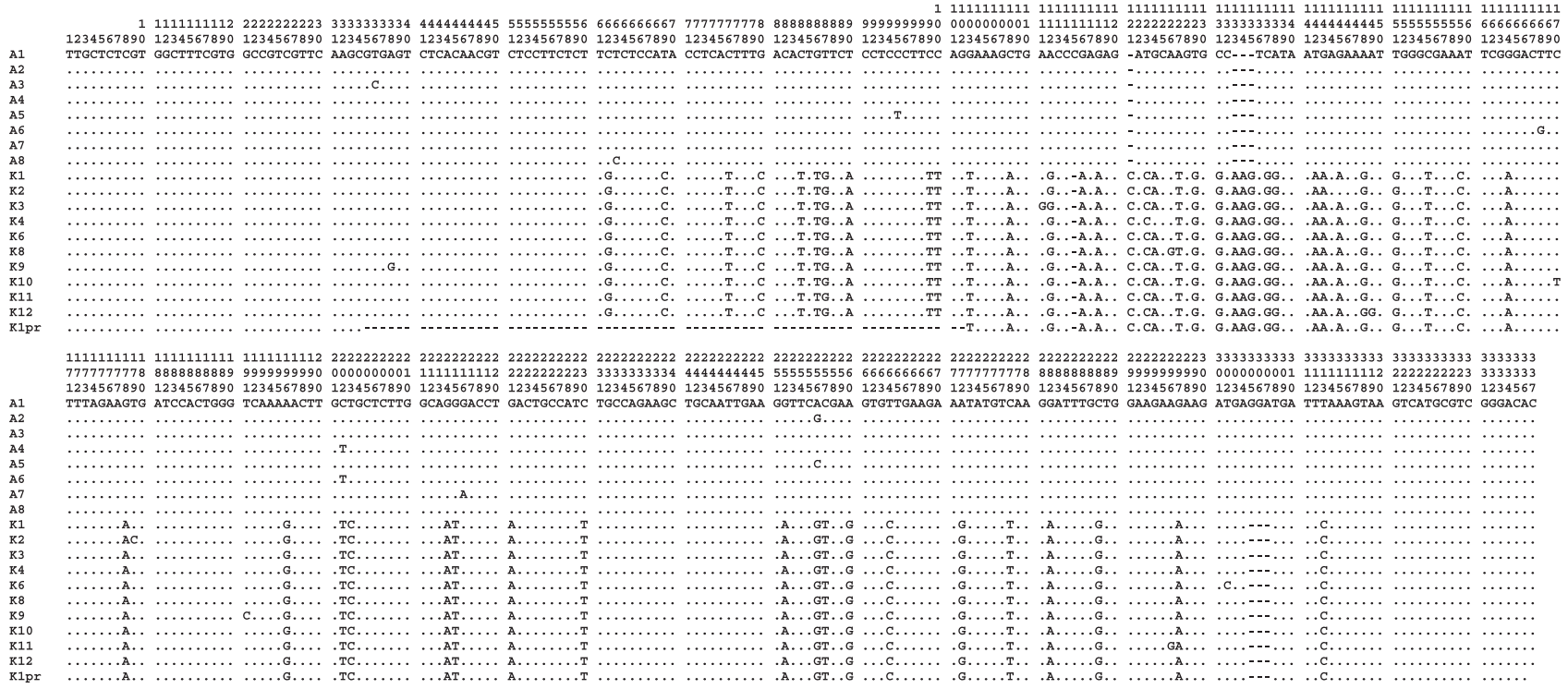

Fig. 5. Alignment of the cloned sequences obtained from single protoscolex PCRs with primers B2F and B2R. A1–A8 correspond to *EgAgB4* and K1–K12 and K1pr to *EgAgB2*.

corresponds to an additional AgB isoform, encoded by a fifth gene locus, *EgAgB5*.

One genomic sequence identical to the most common allele (K1) of *EgAgB2* lacks the intron (K1pr, see Fig. 5). One nonsense mutation occurs in allele Q3 of *EgAgB5* (site 297 in Fig. 4). The ratio of replacements to synonymous substitutions is 3:5 for *EgAgB1* (M alleles), 4:10 for *EgAgB2* (K alleles), 5:8 for *EgAgB3* (R alleles), 6:9 for *EgAgB4* (A alleles) and 10:11 for *EgAgB5* (Q alleles). Transitions exceed the number of transversions in all but *EgAgB3* and *EgAgB5*.

3.2. Parsimony networks

Since *EgAgB* genes each include several variant alleles with minor sequence differences, their phylogenetic relationships might be better represented by networks. Using a 95% confidence probability as a parsimony criterion (statistic parsimony), we have obtained five independent gene networks, representing the five *EgAgB* genes, based on the alignment of all sequences obtained in the present study (Fig. 6). In each gene network, the presumed ancestral allele is the most frequent sequence, which is moreover

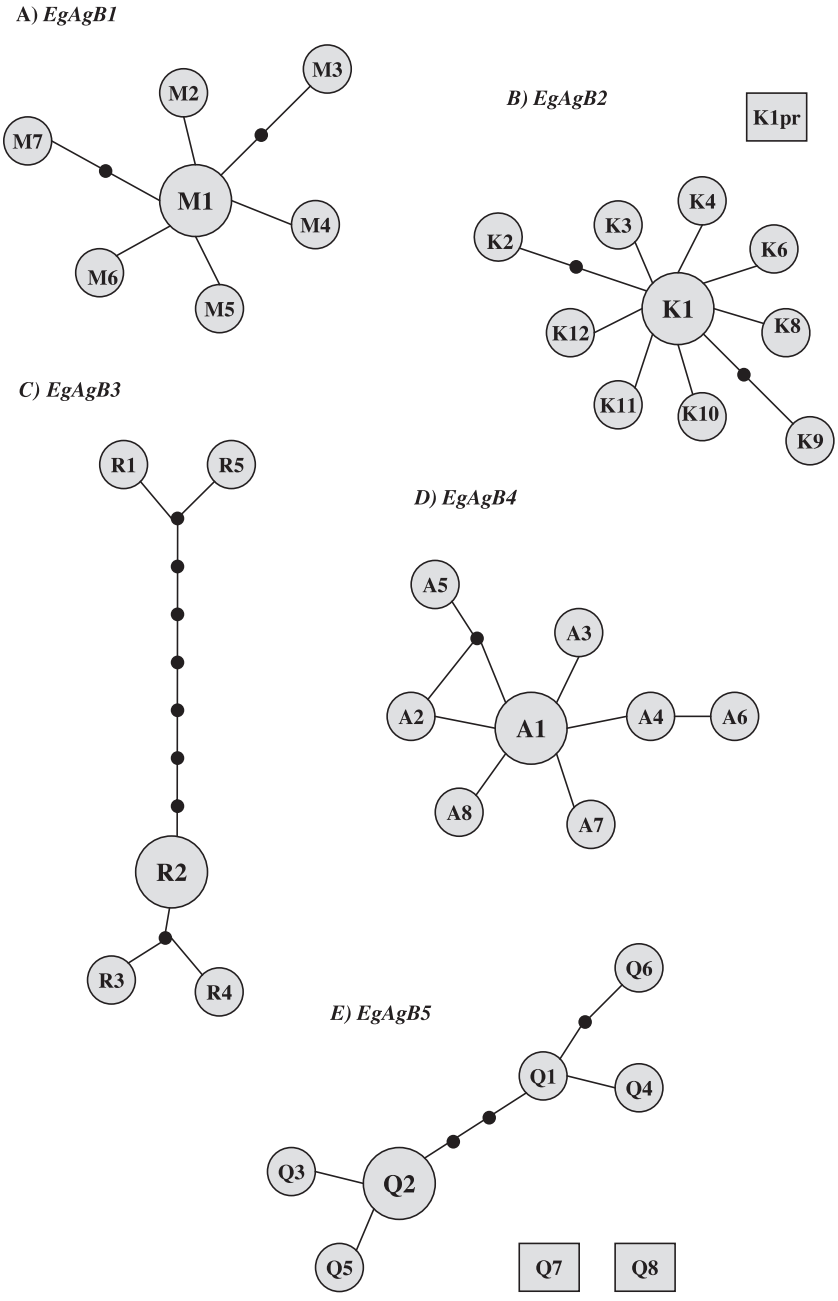

Fig. 6. *EgAgB* allele networks obtained with statistic parsimony (see Section 2). Small solid circles indicate alleles not found in our sample. Rectangles indicate sequences requiring more than nine evolutionary steps in order to be connected to other alleles and, therefore, not supported by a 95% confidence criterion.

Table 1
Eight gene conversion events inferred for pairs of *EgAgB5* alleles, as indicated by their global inner (GI) fragment and pairwise inner (PI) fragment probabilities in the GENECONV analysis

Sequences	GI probabilities		PI probabilities				
	Simulation	BC KA	Simulation	BC KA	Begin	End	Length (bp)
Q1;Q7	0.0082	0.97069	0.0000	0.00131	211	348	138
Q2;Q7	0.0030	0.34405	0.0000	0.00046	1	348	348
Q4;Q7	0.0030	0.33289	0.0000	0.00045	211	348	138
Q5;Q7	0.0121	>1	0.0000	0.00179	215	348	134
Q1;Q8	0.0049	0.57045	0.0000	0.00077	211	348	138
Q2;Q8	0.0049	0.57045	0.0000	0.00077	211	348	138
Q4;Q8	0.0017	0.19296	0.0000	0.00026	211	348	138
Q5;Q8	0.0044	0.47291	0.0000	0.00064	215	348	134

Simulation probabilities are based on 10,000 permutations; BC KA are Bonferroni-corrected probability values. The position of fragments involved in gene conversion is shown by their beginning and ending sites in the alignment; the length (bp) of each fragment is also shown.

found in all analyzed protoscoleces (excepting *EgAgB3*, R2, see Fig. 3). The derived alleles appear each in only one protoscolex (the only exception is Q1, present in two protoscoleces), but single protoscoleces had several distinct copies of the same gene (see Fig. 3). Alleles are separated by a low number of evolutionary steps in "star-like" networks. Indeed, the larger number of steps in the *EgAgB3* and *EgAgB5* genealogies is probably an artifact introduced by sequence gaps. In the parsimony analysis, insertions/deletions of a single nucleotide are considered as a fifth state. Since some of these events are probably not independent, the number of required steps to explain the alleles' differences should be smaller. This kind of artifact also explains why K1pr, which differs from K1 only by the deletion of a stretch corresponding to the intron (sites 35–102 in Fig. 5) did not cluster with the other K alleles. Connections between alleles requiring more than nine steps have a lower than 95% parsimony probability and have therefore been excluded from the total network (alleles K1pr, Q7 and Q8 in Fig. 6).

3.3. Gene conversion

The gene genealogy of *EgAgB4* suggests the occurrence of recombination (see the relationship among A1, A2 and A5 alleles in Fig. 6). Additional evidence of recombination comes from the gene conversion analysis. We have looked for fragments, bounded on each side by discordant sites, for which a pair of sequences is sufficiently similar to be suggestive of gene conversion (Sawyer, 1999). We use an alignment incorporating all nucleotide sequences found in this study, as well as the derived amino acid sequences. With the nucleotide sequences, computations are made using all polymorphic sites (and also, only the silent sites, but we did not find any significantly similar fragments for these). Using all nucleotide polymorphic sites, we have found eight global inner (GI) and 106 pairwise inner (PI) significant fragment pairs. Global *p*-values compare each fragment with all possible fragments for the entire alignment. Only the pairs showing significant GI probabilities are listed in Table 1. The pairwise *p*-values are obtained comparing each fragment with the highest similarity score expected for that sequence pair in the absence of gene conversion. GI and PI probabilities are calculated based on simulations with 10,000 permutations and also on corrected BLAST scores. The values given in Table 1 are Bonferroni-corrected probability values, using Karlin Altschul statistics (Scordis et al., 1999). A 0.0000 PI probability means that no permutation resulted in a longer fragment for the given pair of sequences; a 0.0082 GI probability (pair Q1;Q7, see Table 1) indicates that only 82 out of 10,000 permutations resulted in some fragment this large or longer. Note that all significant fragment pairs include either the Q7 or the Q8 alleles. Amino acid sequences did not lead to significant fragments.

3.4. Neutrality tests and selection

We have used nucleotide sequences from other *Echinococcus* species/strain as outgroups to search for evidence of selection in the *EgAgB* genes. *E. multilocularis* sequences were used as outgroups for *EgAgB1*, *EgAgB2* and *EgAgB4* (*AY265908*, *AY324069* and *AY324068*, respectively), an *E. granulosus* cervid strain sequence (*AY321282*) for *EgAgB3* and an *E. vogeli* outgroup sequence (*AY321292*) for *EgAgB5*. Within a single metacestode, there are departures from neutrality in *EgAgB2*. Natural selection differently influences the number of segregating sites, singletons and

Table 2
Neutrality tests, average ratio of non-synonymous to synonymous nucleotide (Ka/Ks) substitutions per site and nucleotide diversity (π) in the five *EgAgB* genes; *n* represents the number of alleles per gene

Gene	Alleles	*n*	Fu and Li (D)	Fu and Li (S)	Tajima (D)	Ka/Ks	π
EgAgB1	M	7	− 1.666, NS	− 1.780, NS	− 1.576, NS	0.422	0.006
EgAgB2	K	11	− 2.398**	− 2.595**	− 1.990*	0.180	0.007
EgAgB3	R	5	0.083, NS	0.086, NS	0.083, NS	0.467	0.008
EgAgB4	A	8	− 1.444, NS	− 1.606, NS	− 1.477, NS	∞	0.006
EgAgB5	Q	8	0.608, NS	0.498, NS	− 0.079, NS	0.690	0.018

NS = not significant.
* $p < 0.05$.
** $p < 0.001$.

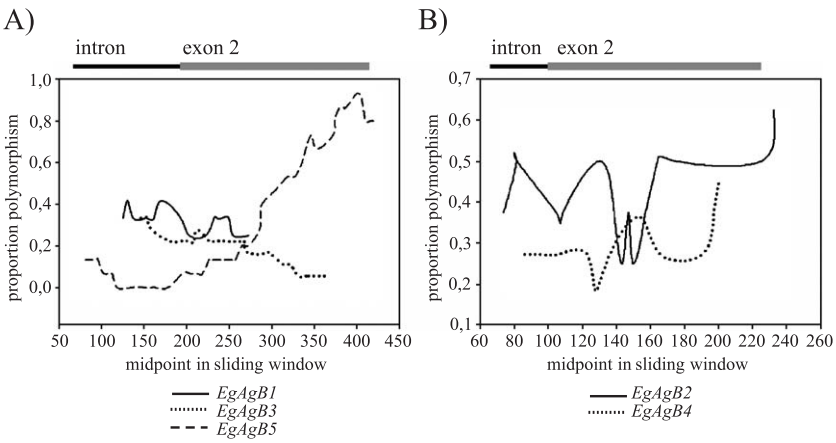

Fig. 7. Polymorphism-to-divergence ratio distribution along the sequences of *EgAgB1*, *EgAgB3* and *EgAgB5* alleles (A), and *EgAgB2* and *EgAgB4* alleles (B).

the mean number of nucleotide differences between sequences. Fu and Li's (1993) and Tajima's (1989) statistics reject neutrality based on the amount of dissimilarity in the estimation of Θ (theta) for these parameters, but demographic events such as bottlenecks can also lead to biased estimates of Θ. Moreover, the alleles analyzed here are obtained from only three protoscoleces (asexual life cycle stages, which are supposed to be clones) from each single cyst. Therefore, more detailed knowledge about the biochemical interaction between parasite and host at this particular life stage, and also about the biological role of AgB, would be required to determine the likely causes that account for rejection of the null hypothesis.

Ka/Ks>1 ratios often reflect positive natural selection. Ka/Ks>1 might be expected for an antigen (and, particularly, one involved in immune evasion), such as AgB. Although, the averaged Ka/Ks ratios between *EgAgB* alleles are relatively high, they are not indicative of positive selection (Table 2). The highest ratio (0.690) is for *EgAgB5*, which also shows the highest nucleotide diversity and, as mentioned before, appears to be involved in gene conversion (Table 1). Furthermore, *EgAgB5* allele polymorphism can be explained by balancing selection. Without selection, fixed and polymorphic substitutions should increase proportionally along lineages. Our data clearly indicate that *EgAgB5* genes have a higher proportion of polymorphisms (opposed to fixed differences) in the 3′ region of exon 2 (Fig. 7), corresponding to the carboxy-terminal end of the inferred protein. The difference is highly significant ($G=30.3785$, $p=0.0010$). The other four *EgAgB* gene loci do not show significant differences between polymorphisms and fixed differences.

4. Discussion

4.1. How many EgAgB genes are there?

With the data currently available, it is not possible to know exactly how many copies per genome occur in the *E. granulosus AgB* gene family. The results shown in the present study, however, indicate redundancy and variability. There seem to be no fewer than five gene loci, each present in several, slightly different copies ("alleles").

The PCR-based cloning experiments and the Southern blot experiments indicate the presence of a number of *EgAgB* genes. Phylogenetic analyses separate our sequences in five major clusters, which we call genes (gene loci *EgAgB1* to *EgAgB5*). Further investigation is required to explain the presence of a number of slightly variant sequences (which we call alleles) within each gene locus in single protoscoleces, derived from the same hydatid cyst. Since protoscoleces are produced asexually within the cyst, three scenarios could explain this observation: (1) PCR or cloning artifacts; (2) variation induced during clonal multiplication, leading to genetic differences among protoscoleces from the same cyst and/or cells from the same protoscolex; (3) variation among tandemly repeated copies of each *EgAgB* gene, with low degree of concerted evolution.

We do not believe that experimental artifacts are the sole or main explanation for our results, because similar genetic variability has been found in previous studies. Experiments of RT-PCR followed by cloning (Arend et al., 2002) and the construction of *Echinococcus* EST libraries (Fernández et al., 2002) have indicated the presence of sequences with nucleotide substitutions, some of them identical and in the same sites as the alleles characterized here. Our previous attempts to screen total or subtracted genomic libraries failed to isolate any *EgAgB*-related clone. However, the cloning of *EgAgB* 5′ flanking regions (data not shown) indicates that the same gene is located in different genomic regions. Moreover, very similar alleles (belonging to the same gene locus) seem to be transcribed from distinct promoter sequences (results not shown), suggesting that they are differentially expressed. Therefore, we believe that individual *EgAgB* gene loci are repeated within each cell (third scenario).

The second and third scenarios are not mutually exclusive. Indeed, the networks in Fig. 6 suggest the occurrence of somatic hypermutation (clonal variation), with several

rare variants frequently originating from ancestral alleles. We found only one network where alleles are ordered in a single branch, namely, the *EgAgB4* alleles A1, A4 and A6 (Fig. 6). These phylogenetically ordered variants occur only in protoscolex 5 (Fig. 2), suggesting that mutations on A1 leading to A6 might be protoscolex-specific. It is therefore possible that duplications of the same *EgAgB* gene within one genome, as well as clonal variation among alleles from different protoscoleces, account for the diversity of sequences found in the present study.

4.2. Gene conversion

Although parasite contingency genes are always repetitive, their expression is usually restricted to a particular genomic location, called expression site (ES). In the clonal pathogenic bacterium *Borrelia hermsii*, serotype switching occurs when one serotype-specific variant replaces another by duplicative transposition into the ES (review in Rich et al., 2001). In the African protozoan *Trypanosoma brucei*, switching of the VSG genes to the ES occurs by gene conversion (review in Borst, 2002). Our gene conversion analyses have shown fragments extending from sites 1, 211 or 215 to 348 (Table 1) in alleles Q7 and Q8 (*EgAgB5*), with a higher than expected similarity to the other Q alleles, which is suggestive of gene conversion.

Vertebrate antibodies acquire diversity by gene conversion, recombination and hypermutation. For example, exchanging V(D)J gene regions through recombination, immune cells are able to produce different classes of antibodies. The process seems to have evolved from an ancestral transposon; proteins RAG1 and RAG2, which mediate the recombination process, can catalyze a full transposition reaction (Agrawal et al., 1998). It is tempting to imagine a similar mechanism generating variation in the *EgAgB* genes, since we isolated a genomic clone lacking the intron from one of the protoscoleces, which could be explained by retrotransposition. However, we are still far from understanding the implication of the similarities of *Echinococcus* AgB variability with respect to the immune system, nor are we yet certain about the exact organization and causes of the parasite's antigenic variation.

4.3. Natural selection

Based on our data, we can reject neutrality for *EgAgB2* (see Table 2). The action of selection on AgB isoforms from a single cyst might result from the direct interaction between AgB epitopes and their targets, so as to evade the host's immune response. In malaria, epitope variation interferes with priming of T lymphocytes through antagonism with the T-cell receptor (Plebanski et al., 1999). Coexisting parasite strains facilitate each other's survival by down-regulating cellular immune responses using altered peptide ligand (APL) antagonism (Gilbert et al., 1998). A similar antagonistic interaction could explain the fact that the polymorphisms of *EgAgB2* and *EgAgB4* alleles have complimentary (mirror) distributions (Fig. 7). Furthermore, the excess polymorphism in the 3′ coding region of *EgAgB5* suggests balancing selection. These indications of selection provide insights for future studies seeking to decipher the effects of different AgB epitopes in the evasion of immune response.

APL antagonism has been found in other pathogens, such as *Leishmania*, where altered ligand peptides differing in a single amino acid failed to activate interleukin 4, which directs the subsequent Th2 cell response (Pingel et al., 1999). It is possible that the presence of slightly different AgB isoforms within a metacestode facilitates the modulation of the host's immune system. Further studies about the effects of simultaneously occurring variant AgB molecules on T-cell responses might resolve this issue. Whether *EgAgB* alleles are repeated in the genome of a single parasite cell (third scenario), or in different cells from one metacestode (second scenario), we suggest that variation is selected for perhaps by its contingent role in the immune evasion process.

4.4. Concluding remarks

Antigenic switching in helminth parasites has not been previously described. The putative protein isoforms encoded by the five *EgAgB* genes differ 44–81% in amino acid sequence. A switch from one isoform to another could, therefore, be sufficient to evade immune response. Studies with polyclonal antisera indicate that nondenatured globular proteins differing by at least 30–40% in sequence will not cross-react (Wilson et al., 1977). Although mRNAs for *EgAgB1–4* isoforms have been previously identified in protoscoleces (*EgAgB5* has been uncovered by the present study for the first time), there are still no studies about control of gene expression in this gene family. It would be of great importance to determine whether the *EgAgB* genes are in fact differentially expressed, and also at which point in the life cycle is each gene switched on.

It will also be important to determine the location of the *EgAgB* genes in the *E. granulosus* genome and how they are organized. Some genes responsible for antigenic variation in microparasites are known to have subtelomeric localization, where they experience high rates of mutation, recombination and silencing. If the *EgAgB* genes are subtelomeric, this could explain why previous attempts to isolate sequences from genomic libraries did not succeed. Subtelomeric genes are usually underrepresented in genomic libraries, due to the paucity of restriction sites in these particular regions. New efforts to build genomic libraries taking into account this possibility should be made in order to isolate large genomic regions containing more than one *EgAgB* gene.

In conclusion, our evidence suggests that the *EgAgB* genes are contingency genes, which would be the first contingency genes described in helminths. Additional studies will be necessary to confirm this hypothesis.

Acknowledgements

Research supported by NIH grant GM42397 to F.J.A. and grants from Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT/CNPq-620538/98-8) to K.L.H. The work was part of K.L.H. post-doctoral activities at UCI, with a fellowship from Coordenaç ão de Aperfeiç oamento de Pessoal de Nível Superior (CAPES).

References

Agrawal, A., Eastman, Q.M., Schatz, D.G., 1998. Implications of transposition mediated by V(D)J-recombination proteins RAG1nd RAG2 for origins of antigen-specific immunity. Nature 394, 744–751.

Arend, A.C., Alves-Junior, L., Chemale, G., Zaha, A., Haag, K.L., 2002. Cloning and characterization of sequences encoding an *Echinococcus granulosus* antigen (AgB) of diagnostic value. The American Journal of Tropical Medicine and Hygiene 67, 299–300 (Supplement).

Barry, J.D., Ginger, M.L., Burton, P., McCulloch, R., 2003. Why are parasite contingency genes often associated with telomeres? International Journal for Parasitology 33, 29–45.

Borst, P., 2002. Antigenic variation and allelic exclusion. Cell 109, 5–8.

Chemale, G., Haag, K.L., Ferreira, H.B., Zaha, A., 2001. *Echinococcus granulosus* antigen B is encoded by a gene family. Molecular and Biochemical Parasitology 116, 233–237.

Clement, M., Derington, J., Posada, D., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9, 1657–1660.

Diaz, M., Casali, P., 2002. Somatic immunoglobulin hypermutation. Current Opinion in Immunology 14, 235–240.

Fernández, V., Ferreira, H.B., Fernández, C., Zaha, A., 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., 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.

Frank, S.A., 2002. Immunology and Evolution of Infectious Disease. Princeton Univ. Press, Princeton.

Frosch, P., Hartmann, M., Mühlschlegel, F., Frosch, M., 1994. Sequence heterogeneity of the echinococcal antigen B. Molecular and Biochemical Parasitology 64, 171–175.

Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Gilbert, S.C., Plebanski, M., Gupta, S., Morris, J., Cox, M., Aidoo, M., Kwiatkowski, D., Greenwood, B.M., Whittle, H.C., Hill, A.V.S., 1998. Association of malaria parasite, population structure, HLA, and immunological antagonism. Science 279, 1173–1177.

Haag, K.L., Gottstein, B., Araújo, A.M., Zaha, A., 1998. Selection, recombination and history in a parasitic flatworm (*Echinococcus*) inferred from nucleotide sequences. Memórias do Instituto Oswaldo Cruz 93, 695–702.

Kinoshita, K., Honjo, T., 2001. Linking class-switch recombination with somatic hypermutation. Nature Reviews. Molecular Cell Biology 2, 493–503.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. Mega2: Molecular Evolutionary Genetics Analysis software. Distributed by the authors, Arizona State University, Tempe, AZ, USA, available at http://www.megasoftware.net/.

McDonald, J.H., 1998. Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Molecular Biology and Evolution 15, 77–384.

Mefford, H.C., Trask, B.J., 2002. The complex structure and dynamic evolution of human subtelomeres. Nature Reviews. Genetics 3, 91–102.

Moxon, E.R., Rainey, P.B., Novak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current Biology 4, 24–33.

Musiani, P., Piantelli, M., Lauriola, L., Arru, E., Pozuuoli, R., 1978. *Echinococcus granulosus*: specific quantification of the two most immunoreactive antigens in hydatid fluids. Journal of Clinical Pathology 31, 475–478.

Pingel, S., Launois, P., Fowell, D.J., Turck, C.W., Southwood, S., Sette, A., Glaichenhaus, N., Louis, J.A., Locksley, R.M., 1999. Altered ligands reveal limited plasticity in the T cell response to a pathogenic epitope. Journal of Experimental Medicine 189, 1111–1120.

Plebanski, M., Lee, E.A.M., Hannan, C.M., Flanagan, K.L., Gilbert, S.C., Gravenor, M.B., Hill, A.V.S., 1999. Altered peptide ligands narrow the repertoire of cellular immune responses by interfering with T-cell priming. Nature Medicine 5, 565–570.

Rich, S.M., Sawyer, S.A., Barbour, A.G., 2001. Antigen polymorphism in *Borrelia hermsii*, a clonal pathogenic bacterium. Proceedings of the National Academy of Sciences of the United States of America 18, 15038–15043.

Riganú, R., Profumo, E., Bruschi, F., Carulli, G., Azzarà, A., Ioppolo, S., Buttari, B., Ortona, E., Margutii, P., Teggi, A., 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.

Riganú, R., Buttari, B., Profumo, E., Ortona, E., Margutti, P., Teggi, A., Siracusano, A., 2002. Apopoptosis modulation in PMBC from patients with cystic echinococcosis. The American Journal of Tropical Medicine and Hygiene 67, 298.

Rozas, J., 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.

Sawyer, S.A., 1999. GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at http://www.math.wustl.edu/ ~ sawyer.

Scordis, P., Flower, D.R., Attwood, T.K., 1999. FingerPRINTScan: intelligent searching of the PRINTS motif database. Bioinformatics 15, 799–806.

Shepherd, J.C., Aitken, A., McManus, D., 1991. A protein secreted in vivo by *Echinococcus granulosus* inhibits elastase activity and neutrophil chemotaxis. Molecular and Biochemical Parasitology 44, 81–90.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Wilson, A.C., Carlson, S.S., White, T.J., 1977. Biochemical evolution. Annual Review of Biochemistry 46, 573–639.