Title
Searching for antigen B genes and their adaptive sites in distinct strains and species of the helminth Echinococcus

Journal
Infection Genetics and Evolution, 6(4)

ISSN
1567-1348

Authors
Haag, KL
Zanotto, PMA
Alves-Junior, L
et al.

Publication Date
2006-07-01

DOI
10.1016/j.meegid.2005.07.003

Available online at www.sciencedirect.com
SCIENCE DIRECT
Infection, Genetics and Evolution 6 (2006) 251–261
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

Searching for antigen B genes and their adaptive sites in distinct strains and species of the helminth *Echinococcus*

K.L. Haag [a,c,*], P.M.A. Zanotto [b], L. Alves-Junior [c], R.B. Gasser [d,e], A. Zaha [f], F.J. Ayala [g]

[a] *Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500 Prédio 43323, Caixa Postal 15053, Porto Alegre, RS, CEP 91501-970, Brazil*
[b] *Laboratório de Evolução Molecular e Bioinformática, Instituto de Ciências Biomédicas, Universidade de São Paulo, Brazil*
[c] *Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Brazil*
[d] *Department of Veterinary Science, The University of Melbourne, Australia*
[e] *Biotechnology Research Institute, Macquarie University, Australia*
[f] *Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Brazil*
[g] *Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California at Irvine, USA*

Received 5 May 2005; received in revised form 20 July 2005; accepted 26 July 2005
Available online 3 October 2005

Abstract

Twenty-seven PCR-derived antigen B (AgB) nucleotide sequences from four *Echinococcus* species (*Echinococcus granulosus*, *Echinococcus multilocularis*, *Echinococcus oligarthrus* and *Echinococcus vogeli*) were aligned with 78 already published sequences, to generate a maximum likelihood phylogeny of the AgB multigene family. The phylogenetic analysis confirms that the family is constituted by four groups of genes present in each one of the four species (*AgB1*, *AgB2*, *AgB3* and *AgB4*), and suggests that it originated by ancient duplication events preceding speciation within the genus. *AgB5* sequences, which had been formerly suggested to correspond to a putatively new AgB subunit, cluster with *AgB3*. Likelihood tests suggest that AgB gene evolution may have been driven by heterogeneous selection pressures acting on particular *AgB1*, *AgB3* and *AgB4* codons. No selection is detected in *AgB2*. We discuss implications of our findings in terms of AgB biology and its use as a diagnostic tool.

Keywords: *Echinococcus*; Hydatid disease; Antigen B; Positive selection; Phylogeny; Molecular evolution

1. Introduction

Antigen B (AgB), the major antigen of *Echinococcus granulosus*, which causes the cystic hydatid disease, is a 120–160 kDa lipoprotein composed of heterogeneous multimers of 8-kDa subunits encoded by a multigene family. Four AgB genes have been characterized, *AgB1*, *AgB2*, *AgB3* and *AgB4* (Shepherd et al., 1991; Fernández et al., 1996; Chemale et al., 2001; Arend et al., 2004), each encoding a related, but distinct AgB subunit. However, genomic Southern blots suggest that the gene family should include at least seven genes (Haag et al., 2004). The native lipoprotein amounts to a large fraction of the metacestode cyst fluid, and its capacity of modulating the host immune response suggests that it may play a relevant role in the host/parasite interaction. Furthermore, since parasites are unable to synthesize most of their own lipids, these molecules must be taken from the host.

Helminth lipid-binding proteins (LBPs) are frequently small and rich in alpha helices, and are essential for the uptake and transport of fatty acids, which are nutrients and structural components of the cell membrane. Parasite LBPs may also be involved in host invasion (McKerrow and Salter, 2002) and parasite evasion from host immune responses (Furlong and Caulfield, 1989) and pathogenicity (Bradley et al., 2001). The *Echinococcus* AgB is predicted to be alpha helix-rich (González-Sapienza and Cachau, 2003), but its binding properties to hydrophobic compounds and its polymeric nature suggest a distinct biological role in relation to the parasite LBPs

* Corresponding author. Tel.: +55 513 316 6717; fax: +55 513 316 6727.
E-mail address: karen.haag@ufrgs.br (K.L. Haag).

1567-1348/$ – see front matter
doi:10.1016/j.meegid.2005.07.003

characterized so far. Chemale et al. (2005) suggest that AgB could be involved in detoxifying mechanisms, sequestering and buffering xenobiotics, since it does not release lipids in competition assays, and therefore, does not seem to direct fatty acid metabolic pathways.

AgB seems to be involved in the evasion of host immune responses, as indicated by its ability to inhibit polymorphonuclear cell recruitment (Shepherd et al., 1991; Riganò et al., 2001). Studying the immunological profile of human patients with different stages of hydatid disease, it has been shown that parasite survival is always associated with a Th2 lymphocyte polarization in its host, which is induced by AgB (Riganò et al., 2001; Riganò et al., 2004). It is possible that the protein has a critical role in the parasite establishment during the early stages of infection, since AgB starts to be recognized by specific antibodies as early as 5 weeks post infection in experimental infected mice (Zhang et al., 2003). In human infections, AgB is highly immunogenic, leading to its use as a diagnostic tool. Indeed, AgB has the highest diagnostic potential among the major *E. granulosus* antigens, and synthetic peptides representing the N-terminal extension of AgB1 exhibit higher diagnostic sensitivity and specificity than the native AgB (González-Sapienza et al., 2000). Recent immunological studies with peptides derived from this region have determined that the stretch EVKYFFER contains the contact residues of epitopes defined by monoclonal antibodies directed against the N-terminal portion of AgB1 (González-Sapienza and Cachau, 2003). The authors predict that the stretch corresponds to an alpha helix, where the mapped contact residues are brought together in the protein polar side.

At the nucleotide level, a huge amount of AgB diversity has been found. Our earlier studies of *E. granulosus* population genetic variability using *AgB1* as a marker indicated an excess of nonsynonymous over synonymous substitutions, when comparing sequences derived from different strains (Haag et al., 1998). Because *E. granulosus* strains are frequently adapted to a particular intermediate host species, it was reasoned that the *AgB1* polymorphisms should have been positively selected. However, even within a single *E. granulosus* metacestode, several slightly different sequences for a single gene have been found (Arend et al., 2004). Considering that during development the metacestode should be faced with heterogeneous selective pressures exerted by the host, and that some mutations could originate during the process of metacestode growing and protoscolex formation, we have searched for positive selection based on the nucleotide diversity found among three individual *E. granulosus* protoscoleces isolated from the same metacestode (Haag et al., 2004). We have shown that the C-terminal region of a putatively new gene (*AgB5*) has a significantly high ratio of polymorphisms to fixed differences, which could be indicative of balancing selection. We also showed that neutrality is rejected for AgB2 variants taken from the same cyst, but there was no evidence of adaptive evolution for the other three already characterized AgB genes. In the present study we apply codon-based evolutionary models to search for positively selected amino acids using a large set of AgB nucleotide sequences derived from distinct *Echinococcus* species and strains.

2. Materials and methods

2.1. PCR, cloning and sequencing

We amplified AgB genes from *E. granulosus* (cervid strain), *Echinococcus multilocularis*, *E. oligrarthrus* and *Echinococcus vogeli* by PCR using the primer sequences and procedures previously described (Haag et al., 2004). All amplicons were purified from solutions (Qiagen) and 1 μl of each purified amplification product was cloned in a TOPO TA vector for sequencing (Invitrogen). The recombinant plasmids from 10 colonies per transformation were purified with columns (Qiagen) and automatically sequenced using M13 (-20) forward and M13 reverse primers (Invitrogen). Both strands were sequenced for every recombinant plasmid. We used the program Electropherogram Quality Analysis (available at http://bioinformatica.ucb.br/electro.html) to read and evaluate the quality of our sequences. This information was used as a measure of confidence for correcting sequencing errors detected by comparing the electropherograms from both strands.

2.2. Phylogeny reconstruction

In addition to the sequences obtained by ourselves, we made an exhaustive GenBank database search, ending up with a total of 105 *AgB* sequences. These sequences were selected based on their distinctiveness in exon II, which encodes the secreted protein. The exon II sequences were aligned with the Clustal X program (Thompson et al., 1997). The alignment was adjusted by visual inspection and used for phylogenetic analysis with PAUP 4.0. We made a heuristic search of the best phylogeny using the SPR branch-swapping algorithm and likelihood as an optimality criterion, based on the HKY 85+G model, with parameters estimated from the data. Substitution rate variation was assumed to be gamma distributed, with $\alpha = 0.5585$.

3. Searching for positive selection

For positive selection analyses we excluded the GenBank sequences derived from ESTs, which are single-reads, sequences differing only by deletions/insertions and sequences containing premature stop codons or gaps which could lead to non-functional protein products. The analyses were performed for each AgB subunit separately, using an underlying phylogenetic tree specific for each group of sequences (with n sequences equal to 16 for *AgB1*, 23 for *AgB2*, 16 for *AgB3* and 21 for *AgB4*).

Positively selected codons are inferred when ω, the ratio of dN (rate of nonsynonymous substitution) to dS (rate of synonymous substitution) in individual sites, is greater than one. We used three distinct approaches to infer ω: the empirical Bayes, the fixed effects likelihood (FEL) and the random effects

likelihood (REL), relying only on those positively selected sites inferred by all three methods.

3.1. Empirical Bayes

The PAML (Yang, 1997) package was used for fitting distinct models of codon substitution on phylogenetic trees by maximum likelihood (Nielsen and Yang, 1998; Yang et al., 2000), with the incorporation of heterogeneous ω ratios among amino acid sites, and then using a Bayes approach to infer which sites are under selection.

Presence of positive selection, i.e., d*N*/d*S* greater than one ($\omega > 1$) was tested by fitting the following models on each of four B antigen genealogies: single-rate (M0), neutral (M1), selection (M2), discrete (M3), beta (M7) and beta&W (M8). Briefly, M0 assumes only one ω value for all sites, M1 assumes 0 = 0 and 1 = 1 in different proportions (p_0 and p_1), M2 adds a new category of codons (ω_2, with p_2) to the previous model which can assume values greater than 1, M3 classifies sites in K discrete classes ($K = 8$, in our case), with both the ω ratios and frequencies estimated from the data. M7 assumes 10 codon categories with ω values following a beta distribution, within an interval from 0 to 1, not allowing for positive selection. Model M8 adds a single discrete class to the beta distribution, to account for sites under positive selection ($\omega > 1$).

A likelihood ratio test (LRT) was used to compare the log likelihoods of M0 versus M2 (with 2 d.f.), M0 versus M3 (with 4 d.f.), M1 versus M2 (with 2 d.f.), M1 versus M3 (with 4 d.f.), M2 versus M3 (with 3 d.f.) and nd M7 versus M8 (with 2 d.f.), where the number of degrees of freedom is determined by the number of free parameters estimated by each model. The LRT can be applied to nested models. In this regard, M2 and M3 nest M0 and M1, M2 is nested in M3 and M7 is nested in M8. The rejection of M0 and M1 by M2 (with $\omega > 1$) is the most conservative indication of positive selection. However, testing for M3 and M8 (i.e., when M3 rejects M1 or M2 and M8 rejects M7) provides further support for the detection of individual sites. Individual positively selected sites are identified using the Bayes formula, when the posterior probability of $\omega > 1$ is higher than 0.95.

3.2. FEL and REL

Initial phylogeny estimation is performed with the neighbor-joining method using the Tamura-Nei distance, and an iterative process is used to find the best fitting nucleotide evolutionary model in parallel with the best tree reconstruction for each gene. The final tree is used to infer the ancestral states, based on the codon model developed by Muse and Gaut (1994).

Searching for positively selected sites using FEL involves fitting substitution rates on a site-by-site basis (Pond and Frost, 2005). It has an advantage over the empirical Bayes approach, in that it does allow variable synonymous substitution rates as well as variable nonsynonymous substitution rates, but it requires the computation of a much higher number of parameters. FEL is based in a model proposed by Suzuki (2004a), in which the ratio d*N*/d*S* is estimated for each codon using maximum likelihood, and a LRT is used to test whether this ratio is significantly different from 1. Amino acid sites are considered as positively selected when p in the LRT is smaller than 0.05

REL relies on a more general underlying nucleotide substitution model, which is similar to the M3 model of Yang et al. (2000). However, since it allows rate variation in both synonymous and nonsynonymous sites, it seems to lead to a reduced number of type I (false positive) errors, when compared to M3 (Pond and Frost, 2005). Amino acid sites are considered as positively selected when the posterior probability of $\omega > 1$ is higher than 0.95. We use the FEL and REL analysis available at http://www.datamonkey.org.

4. Results

4.1. Phylogeny of AgB genes

The phylogenetic relationship among the 105 *AgB* exon II sequences is shown in Fig. 1.

We use the GenBank accession number to identify sequences obtained via database search. The nucleotide sequences identified with a letter followed by a number are the alleles found in our previous study of *E. granulosus AgB* variation (Haag et al., 2004), which were obtained from only three protoscoleces from a single metacestode; those shown in bold letters are sequences obtained in the present study. A list of GenBank accession numbers and designations for all sequences and their corresponding species/strains is shown in Appendix A. Sequences are separated into four main clades, connected by long branches, which correspond to the already described AgB1, AgB2, AgB3 and AgB4 subunits. The AgB3 clade also contains alleles from an AgB isoform which was previously suggested to correspond to a fifth AgB subunit, AgB5 (Haag et al., 2004), but there is not yet evidence that the *AgB5* genes are functional. Interestingly, all *E. granulosus* cervid strain and *Echinococcus oligarthrus* sequences from the AgB3 clade are clustered with the AgB5 (Q) alleles, and not with those corresponding to the already known *E. granulosus* AgB3 subunit (R alleles). The four major clades contain sequences from each of the four recognized *Echinococcus* species *E. granulosus*, *E. multilocularis*, *E. oligarthrus* and *E. vogeli*, with the exception of the *AgB2* clade, which does not include any *E. vogeli* sequence.

4.2. Diversifying evolution in the AgB gene family

Our analyses indicate that three of the four AgB genes are evolving by positive selection. Except for *AgB2*, where the highest log likelihood is obtained with the single-rate model (M0), the AgB sequence data fit better the discrete model (M3), which allows positive selection (Table 1). Indeed, evolutionary models accounting for positive selection also perform better than the neutral models in the LRT for *AgB1*, *AgB3* and *AgB4*. The average d*N*/d*S* ratios calculated with the empirical Bayes, FEL and REL methods for *AgB* genes are shown in Tables 2 and 3. The highest mean d*N*/d*S* is obtained for *AgB4* with all

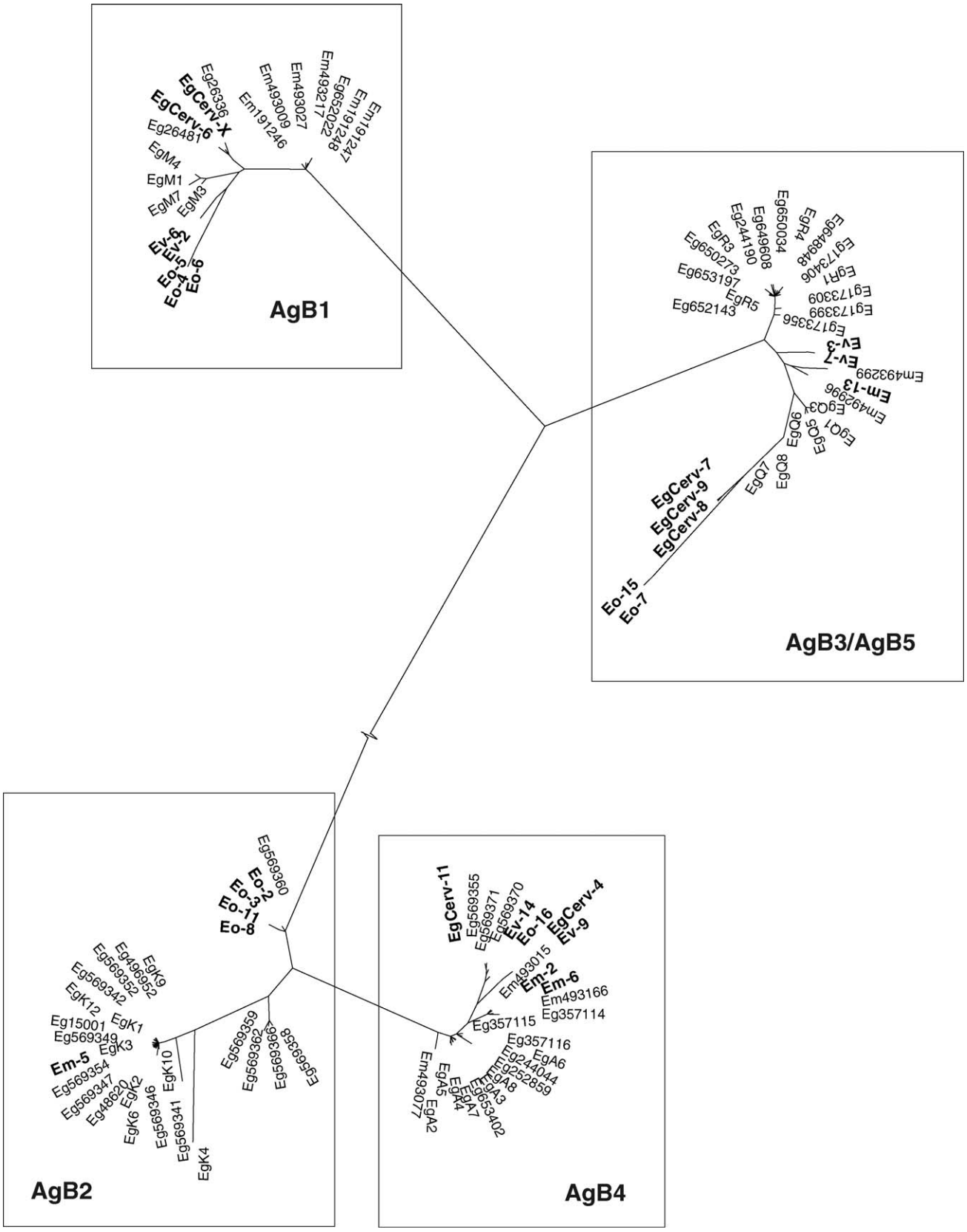

Fig. 1. Maximum likelihood phylogenetic tree inferred from 27 AgB nucleotide sequences obtained in the present study (shown in bold letters), 29 sequences derived from three protoscoleces from a single *E. granulosus* metacestode (shown with a letter followed by a number) and 49 sequences sampled from GenBank shown by their respective accession numbers (see Appendix A for sequence details). Sequences are separated in four main groups, corresponding to subunits AgB1, AgB2, AgB3 (which does not clearly separate from AgB5) and AgB4. The long internal branch connecting AgB1–AgB3/AgB5 to AgB2–AgB4 clades has been shortened for convenience.

Table 1
Log likelihood of *AgB* sequence data under M0–M3 and M7–M8 evolutionary models

Gene	Model						LRT	
	M0	M1	M2	M3	M7	M8	Comparison	*p*
AgB1	−554.099	−551.481	−550.259	−550.248	−551.536	−550.410	M0 × M2	0.0215
AgB2	−644.810	−646.754	−645.837	−645.306	−646.124	−645.310	–	–
AgB3	−650.993	−643.024	−642.023	−639.052	−642.403	−642.403	M0 × M2	0.0001
							M0 × M3	0.0001
AgB4	−523.694	−525.460	−520.462	−519.574	−525.502	−519.574	M0 × M2	0.0394
							M1 × M2	0.0067
							M1 × M3	0.0191
							M7 × M8	0.0027

The comparisons showing statistically significant results in the likelihood ratio test (LRT) are displayed.

methods. *AgB1* and *AgB3* sequences lead to similar results with regard to positive selection. The LRT between M0 and M2 suggests diversifying evolution ($p < 0.05$ for *AgB1* and $p < 0.01$ for *AgB3*), but the difference between M7 and M8 is not statistically significant. *AgB2* has the lowest d*N*/d*S* ratio, and does not exhibit positive selection according to any of the three methods.

Since the methods differ considerably in detecting codons under selection pressure, only those sites inferred to be selected by all three methods are indicated in the protein alignments of Figs. 2 and 3. We plotted the site-by-site differences between d*N* and d*S* (d*N* − d*S*, estimated using the REL approach in the same figures) to manifest how the positive and/or negative selection pressures are distributed along the protein sites. The N-terminal region from the AgB1 and AgB3 subunits includes de highest d*N* − d*S*. The site-by-site d*N* − d*S* differences of AgB2 are generally negative and close to zero, in contrast with AgB4, which has quite variable differences, with higher values in the carboxy-terminal end and the lowest values in the N-terminal end. Indeed, the basic amino acid arginine at site 13 of AgB4 corresponds to a d*N* − d*S* "valley" (see Fig. 3) and appears as a negative selected site by both FEL and REL ($p < 0.01$, in the LRT of FEL, data not shown).

Codon 18 of *AgB1* is under positive selection (Fig. 2A). In the sequences included in our sample it always encodes a hydrophobic amino acid, valine, alanine or metionine, but one *E. granulosus* sequence (Z263366, see Appendix A for details) encodes arginine, a basic amino acid, in this particular site. In the N-terminal end of AgB3 a repeated motif encoding aspartic and glutamic acids is positively selected (Fig. 2B). In site 3, the repeated motif is interrupted either by a cystein or an arginine in the *E. granulosus* Q alleles, or by a glycine, in *E. multilocularis*. Both AgB2 and AgB4 contain repeated stretches of the two acidic amino acids, but in their COH-terminal ends (see Fig. 3). The number of these repeated aspartic and glutamic amino acids is highly polymorphic in AgB4, which could be due to positive selection, although this conclusion is not statistically supported by all three methods.

Table 2
Positively selected codons detected with the empirical Bayes approach

Gene	Positively selected sites[a]	d*N*/d*S*	Proportion
AgB1	10, 11, 15, 18, 21, 26, 31, 32, 33, 34, 36, 43, 44, 46, 49, 51, 53, 54, 56, 57, 59, 60, 61, 64, 65	$\omega_{0-2} = 0.000$	0.350
		$\omega_{3-6} = 1.156$	0.543
		$\omega_7 = 3.5838$	0.107
AgB2	70	$\omega_{0-5} = 0.361$	0.803
		$\omega_{6-7} = 2.056$	0.197
AgB3	3	$\omega_{0-3} = 0.083$	0.500
		$\omega_{4-6} = 0.992$	0.470
		$\omega_7 = 7.659$	0.030
AgB4	9, 10, 12, 16, 18, 21, 24, 29, 30, 32, 34, 40, 46, 49, 50, 61, 62, 64, 65, 70, 71, 72	$\omega_{0-2} = 0.0000$	0.489
		$\omega_3 = 5.483$	0.511
		$\omega_4 = 8.492$	0.000
		$\omega_5 = 8.964$	0.000
		$\omega_6 = 11.328$	0.000
		$\omega_7 = 13.268$	0.000

The d*N*/d*S* ratios and their frequency along the four *AgB* genes correspond to model M3.

[a] Models M3 and M8 detect the same set of positively selected sites.

5. Discussion

5.1. History of AgB genes

Our results confirm that the *Echinococcus* AgB gene family includes four major groups of genes, originated from duplication events which occurred in the early evolution of the genus. The long internal branch separating the AgB1/AgB3/AgB5 clade from AgB2/AgB4 clade clearly suggests that the first duplication event occurred before *Echinococcus* speciation. In our experiments, we did not find an AgB2-homolog clone for *E. vogeli*, but our failure in isolating clones belonging to a particular AgB clade could be due to experimental bias. It is reasonable to assume that the other two duplication events, leading to the four AgB subunits, also preceded *Echinococcus* speciation.

Within each major AgB clade, the sequences do not cluster following a distinct taxonomic pattern. For example, the *E. oligarthrus* AgB2 sequences are closely related to an *E.*

Table 3
Selection results obtained with fixed effects likelihood (FEL) and random effects likelihood (REL) for the four *AgB* genes

Gene	FEL			REL		
	Average d*N*/d*S*	Positively selected codons	Negatively selected codons	Average d*N*/d*S*	Positively selected codons	Negatively selected codons
AgB1	1.041	18	2, 3, 22, 53	1.173	18	–
AgB2	0.544	–	17, 22, 42, 58, 67, 68	0.577	–	–
AgB3	0.798	3	6, 11, 19, 21, 22, 30, 31	0.842	3, 28, 66	–
AgB4	1.348	71	13, 42	1.435	9, 10, 16, 21, 24, 29, 30, 32, 34, 40, 46, 50, 62, 64, 65, 70, 71, 72	13

granulosus camel strain variant, in a clade well separated from all other *E. granulosus* AgB2 sequences. One explanation for this finding is that *E. granulosus* strains are quasi-species. Indeed, the genus has been a theme of an endless taxonomical debate, and there is evidence that a revision is called for (Thompson and McManus, 2002). Another possible explanation, which we consider in the present paper, is that AgB genes are subject to strong natural selection. Heterogeneous selective pressures in different directions on distinct protein sites, probably exerted by the host, could lead to a complex pattern of

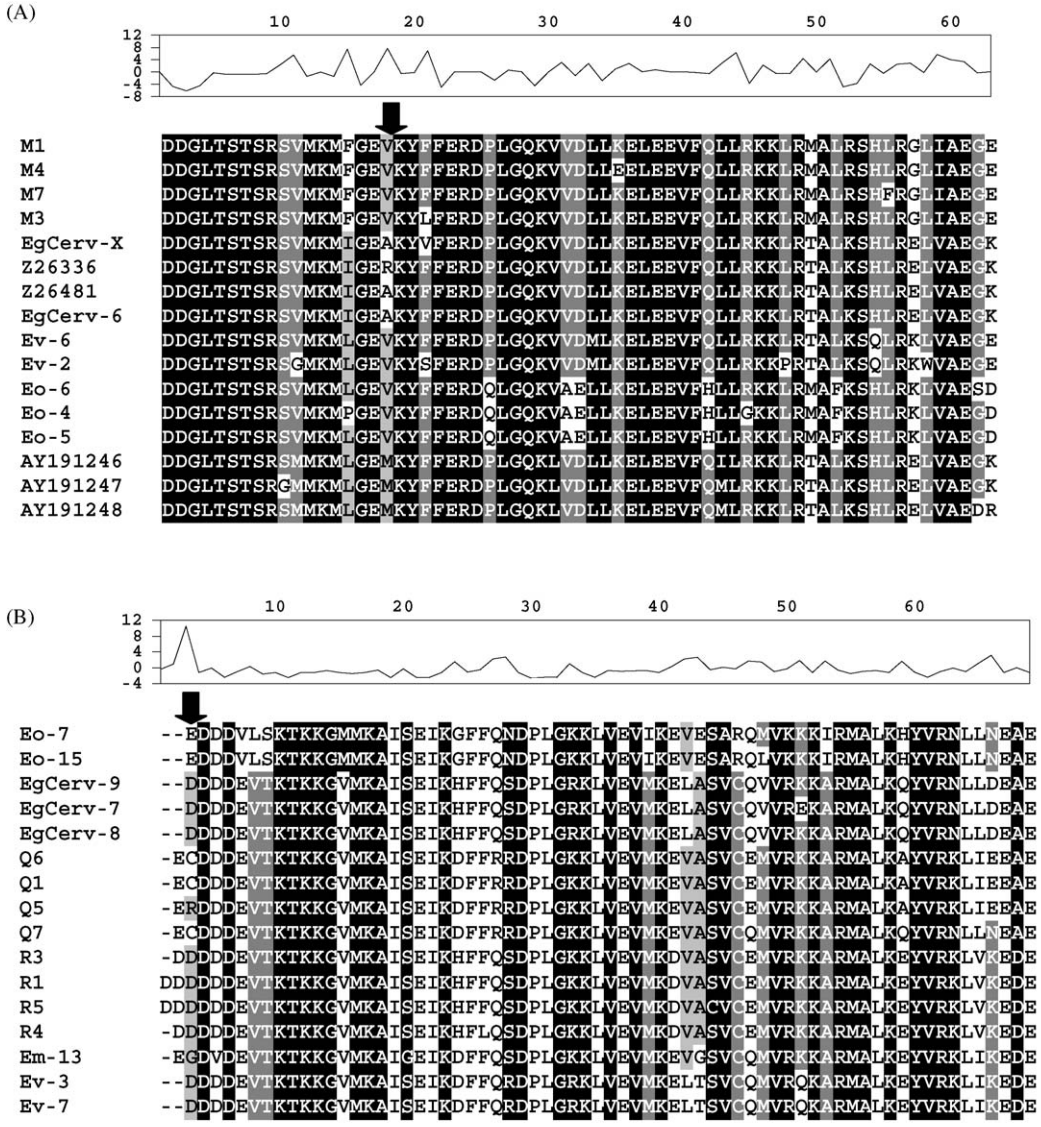

Fig. 2. Alignment of the deduced amino acid sequences of *AgB1* (A) and *AgB3* (B) used for positive selection analyses (see Appendix A for sequence details). Positive selected sites identified by the three methods of analyses are indicated with a black arrow and the site-by-site d*N* − d*S* differences calculated by REL are plotted on top of the alignment.

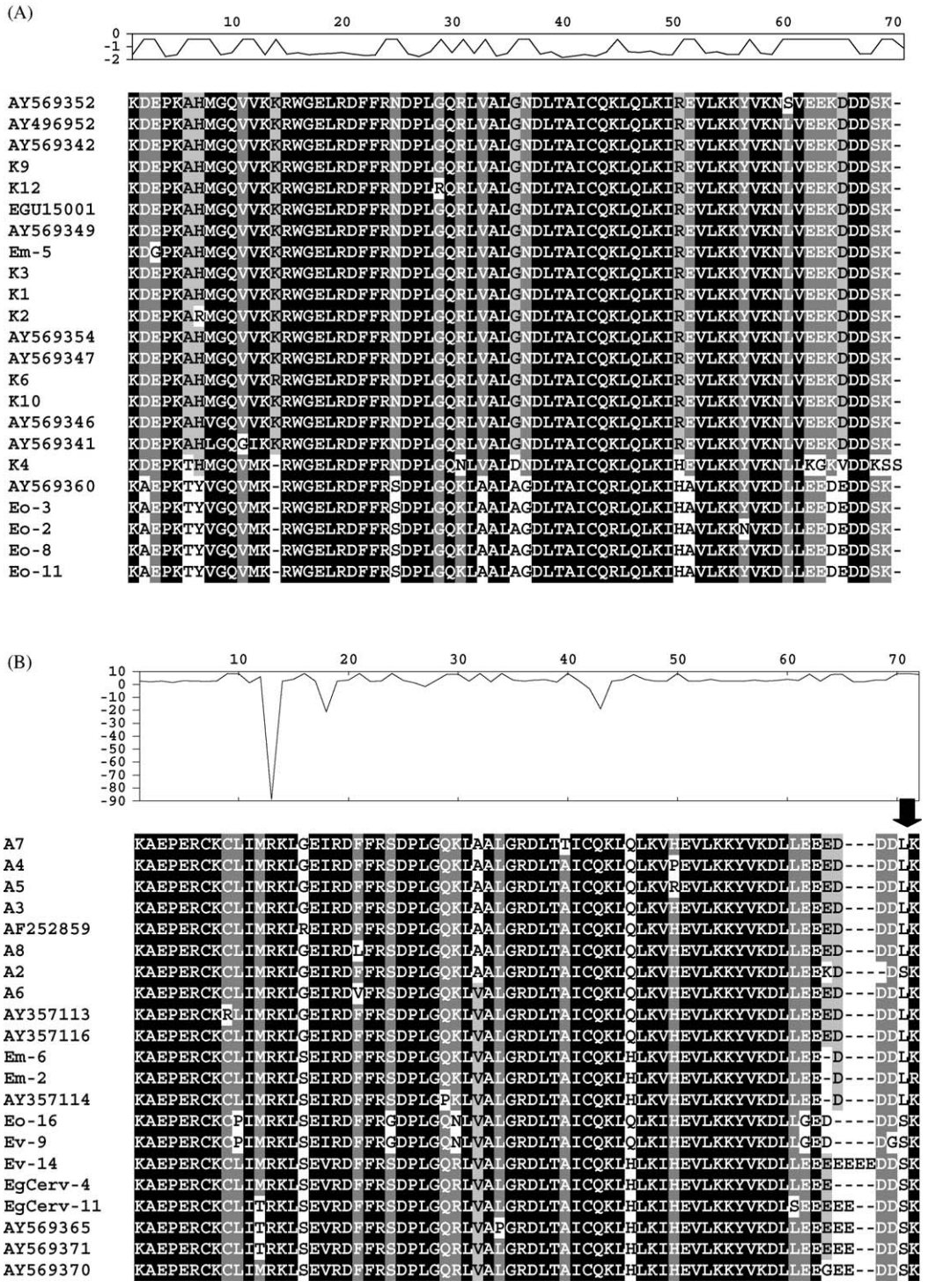

Fig. 3. Alignment of the deduced amino acid sequences of *AgB2* (A) and *AgB4* (B) used for positive selection analyses (see Appendix A for sequence details). The AgB4 positive selected site identified by the three methods of analyses is indicated with a black arrow and the site-by-site dN − dS differences calculated by REL are plotted on top of the alignment.

evolution. Moreover, it seems that these genes evolve quite rapidly. Our primers do not amplify any AgB homolog in other Cestode species, and the low-molecular weight antigens with similar structures characterized for these groups do not show enough similarity with AgB for obtaining reliable sequence alignments.

In our previous work, we suggested that *AgB5* encodes a distinct AgB subunit (Haag et al., 2004), but the present phylogenetic analyses do not discriminate between isoforms *AgB3* and *AgB5*. We already know that *AgB3* is tandemly repeated in the *E. granulosus* genome (unpublished data), and we have previously suggested that *AgB5* could be involved in

gene conversion mechanisms (Haag et al., 2004). We, therefore, suggest that non-homologous recombination may have homogenized the nucleotide sequences of the redundant copies of those genes, yielding a poor phylogenetic resolution in this particular branch.

5.2. Is the effect of natural selection on AgB statistically supported?

Likelihood methods have been widely used to study positive selection over the past few years, although their accuracy in detecting sites under selection on a coding sequence has been questioned (Suzuki and Nei, 2004; Zhang, 2004). It has been argued that the likelihood approach is too sensitive to assumption violations, and false positives (type I errors) can easily arise in these circumstances. One potential source of error is recombination, given that no unique tree topology can satisfactorily account for the evolution of the sequences. By simulation, Anisimova et al. (2003) concluded that with more than three recombination events in a set of 10 sequences, the LRT can yield false positives of positive selection, although recombination affects less the identification of individual sites under positive selection by the empirical Bayes approach. We do not know to which extent recombination may have impacted our data. The minimum number of recombination events (Hudson and Kaplan, 1985) estimated from our datasets is 3 for *AgB1* ($n = 16$), 2 for *AgB2* ($n = 23$), 7 for *AgB3* ($n = 16$) and 4 for *AgB4* ($n = 21$), indicating that, at least for *AgB3*, our conclusion of positive selection might have been confounded by recombination. We have followed the recommendations of Anisimova et al. (2002) for greater accuracy and power in the Bayes approach for detecting positive selected sites, trying to include sequences from several different evolutionary lineages and analyzing different models of evolution. Furthermore, we have used two other likelihood-based methods to detect codons under selective pressure (FEL and REL); our conclusions are conservative in relying only on positively selected codons identified by all three methods. We are aware that the evolutionary models tested by maximum likelihood, or any other model, are always simplistic, and therefore, unrealistic, but this does not invalidate our results.

5.3. Implications and future prospects

Finding positively selected amino acids in a parasite antigen is not, in itself, a novelty. Parasite versus host relationships include well-known examples of diversifying selection. But our findings have two main implications. First, discovering how the selective pressures are distributed among the distinct AgB subunits, and distinct sites within the protein, helps to understand the functional roles of AgB. Although site 70 in AgB2 appears to be positively selected under the M3 and M8 models (Table 2), our analyses indicate that the gene evolves neutrally, because these models do not perform better than the neutral ones in the LRT. On the other hand, AgB4, which is phylogenetically related to AgB2, and highly polymorphic, has a complex pattern of evolution, with both positive and negative selection pressures occurring along the same molecule. Indeed, some preliminary studies of quantification of *AgB* gene expression in *E. multilocularis* indicate that *AgB2* is down regulated, while *AgB4* is highly expressed in metacestodes (unpublished data). Moreover, while searching for AgB sequences in the *Echinococcus* EST data libraries, we found only one AgB2 isoform and several AgB4 sequences. Thus, the distinct evolutionary patterns of these genes might be a consequence of their regulation and, finally, implies distinct functional roles for AgB2 and AgB4, which deserve further investigation. As an initial guess, it could be inferred that AgB subunits or protein sites involved with the binding of hydrophobic compounds should underlie purifying selection, while those communicating with the host immune system should be positively selected. As a multifunctional protein, AgB offers a rich field for biochemical investigation.

The present results are not in contradiction with our previous rejection of neutrality for *AgB2* (Haag et al., 2004, where we used sequences derived from a single cyst) for two reasons. First, neutrality rejection by the tests does not always imply positive selection, because other factors can yield the same result. For example, adaptive sweeps as well as population expansions can lead to significant and negative values of Tajima's D (Escalante et al., 2004). Second, since in the previous work all AgB2 sequences came from protoscoleces originated inside the same metacestode, they told us a different evolutionary story. Those sequences are only informative about processes occurring during the development of a single cyst and do not contain any information about adaptive processes related to the interaction with distinct host individuals/species.

A second implication of our findings relates to the fact that AgB is widely used as a tool for diagnostic tests. Recombinant proteins/peptides derived from distinct AgB subunits are being evaluated for serology, due to their high immune reactivity. Our finding of a positively selected amino acid in AgB1 (site 18 in the N-terminal region) is consistent with empirical studies, since it corresponds to one of the monoclonal antibody-contact residues mapped by González-Sapienza and Cachau (2003). Furthermore, our results show that the N-terminal AgB1 region includes sites with the largest dN/dS differences (see Fig. 2) and corresponds to a peptide with higher diagnostic potential than the protein from which it is derived (González-Sapienza et al., 2000). Positive selection is also inferred for sites in the N-terminal end of AgB3 and the carboxy-terminal end of AgB4, but since both contain variably repeated amino acids, it is possible that recombination in the respective genes has been mistakenly interpreted as positive selection. Repeated motifs generate conditions for mispairing, and polymorphism in repeated motifs can be generated by unequal crossing over. Further studies should be performed to better understand whether these sites indeed correspond to epitopes.

The recombinant AgB2 is one of the antigens with best performance in immunodiagnostics (Rott et al., 2000; Virginio et al., 2003), showing high sensitivity and specificity, which is somehow surprising in face of our results. If patients produce a strong immune response against AgB2, we would expect it to be positively selected, but the indications of positive selection

acting on this subunit are weak. However, since AgB2 shows an overall 70% amino acid identity with AgB4, it is possible that the signal detected in ELISA experiments is derived from an epitope shared between both proteins. Following a recent theoretical analysis made by Suzuki (2004b), which shows that the neutralization epitopes of poliovirus surface proteins always contain negative selected sites, we suggest that a combination of empirical studies of AgB epitope mapping and serology, and theoretical analyses of positive/negative selection on these epitopes could maximize the sensitivity and specificity of immunodiagnostic tests.

Acknowledgements

Thanks to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) for supporting our work.

Appendix A

Subunit	Accession	Designation	Species (strain)	Reference
AgB1	AY265912	Eo-6	*E. oligarthrus*	Present study
	AY265910	Eo-4	"	"
	AY265911	Eo-5	"	"
	AY265913	Ev-2	*E. vogeli*	"
	AY265914	Ev-6	"	"
	AY871011	EgM3	*E. granulosus* (G1)	Haag et al. (2004)
	AY871019	EgM7	"	"
	AY773091	EgM1	"	"
	AY871013	EgM4	"	"
	Z26481	Eg26481	*E. granulosus*	Frosch et al. (1994)
	AY265907	EgCerv-6	*E. granulosus* (G8)	Present study
	AY264402	EgCerv-X	"	"
	Z26336	Eg26336	*E. granulosus*	Frosch et al. (1994)
	AY191246	Em191246	*E. multilocularis*	Mamuti et al. (2004)
	BU493009	Em493009	"	Fernandez et al. (2002)
	BU493027	Em493027	"	"
	BU493217	Em493217	"	"
	CN652022	Eg652022	*E. granulosus*	Fernandez and Maizels (2001)
	AY191248	Em191248	*E. multilocularis*	Mamuti et al. (2004)
	AY191247	Em191247	"	"
AgB2	AY569358	Eg569358	*E. granulosus* (G5)	Kamenetzky et al. (2004)
	AY569366	Eg569366	*E. granulosus* (G7)	"
	AY569362	Eg569362	*E. granulosus* (G6)	"
	AY569359	Eg569359	"	"
	AY871034	EgK4	*E. granulosus* (G1)	Haag et al. (2004)
	AY569341	Eg569341	"	Kamenetzky et al. (2004)
	AY871029	EgK10	"	Haag et al. (2004)
	AY569346	Eg569346	"	Kamenetzky et al. (2004)
	AY871036	EgK6	"	Haag et al. (2004)
	AY871040	EgK2	"	"
	L48620	Eg48620	*E. granulosus*	Fernández et al. (1996)
	AY569347	Eg569347	*E. granulosus* (G1)	Kamenetzky et al. (2004)
	AY569354	Eg569354	*E. granulosus* (G2)	"
	AY324071	Em-5	*E. multilocularis*	Present study
	AY871036	EgK3	*E. granulosus* (G1)	Haag et al. (2004)
	AY569349	Eg569349	"	Kamenetzky et al. (2004)
	AY496952	EgK1	*E. granulosus* (G1)	Haag et al. (2004)
	EGU15001	Eg15001	*E. granulosus*	Fernández et al. (1996)
	AY871037	EgK12	*E. granulosus* (G1)	Haag et al. (2004)
	AY569342	Eg569342	"	Kamenetzky et al. (2004)
	AY569352	Eg569352	*E. granulosus* (G2)	"
	AY496952	Eg496952	*E. granulosus* (G1)	Arend et al. (2004)
	AY871039	EgK9	*E. granulosus* (G1)	Haag et al. (2004)
	AY324075	Eo-8	*E. oligarthrus*	Present study
	AY324076	Eo-11	"	"
	AY324074	Eo-3	"	"
	AY324073	Eo-2	"	"
	AY569360	Eg569360	*E. granulosus* (G6)	Kamenetzky et al. (2004)
AgB3	CN652143	Eg652143	*E. granulosus*	Fernandez and Maizels (2001)
	AY871026	EgR5	*E. granulosus* (G1)	Haag et al. (2004)
	CN653197	Eg653197	*E. granulosus*	Fernandez and Maizels (2001)

Appendix A *(Continued)*

Subunit	Accession	Designation	Species (strain)	Reference
	CN650273	Eg653197	"	"
	AY871015	EgR3	*E. granulosus* (G1)	Haag et al. (2004)
	BI244190	Eg244190	*E. granulosus*	Fernandez and Maizels (2001)
	CN649608	Eg649608	"	"
	CN650034	Eg650034	"	"
	AY871016	EgR4	*E. granulosus* (G1)	Haag et al. (2004)
	CN648948	Eg648948	*E. granulosus*	Fernandez and Maizels (2001)
	BQ173406	Eg173406	"	"
	AY871010	EgR1	*E. granulosus* (G1)	Haag et al. (2004)
	BQ173309	Eg173309	*E. granulosus*	Fernandez and Maizels (2001)
	BQ173399	Eg173399	"	"
	BQ173356	Eg173356	"	"
	AY321291	Ev-3	*E. vogeli*	Present study
	AY321292	Ev-7	"	"
	BU493299	Em493299	*E. multilocularis*	Fernandez et al. (2002)
	AY321287	Em-13	"	Present study
	BU492996	Em492996	"	Fernandez et al. (2002)
	AY871020	EgQ3	*E. granulosus* (G1)	Haag et al. (2004)
	AY871009	EgQ1	"	"
	AY871023	EgQ5	"	"
	AY871021	EgQ6	"	"
	AY871025	EgQ8	"	"
	AY871024	EgQ7	"	"
	AY321283	EgCerv-7	*E. granulosus* (G8)	Present study
	AY321285	EgCerv-9	"	"
	AY321284	EgCerv-8	"	"
	AY321289	Eo-15	*E. oligarthrus*	"
	AY321288	Eo-7	"	"
AgB4	AY324067	EgCerv-11	*E. granulosus* (G8)	Present study
	AY569355	Eg569355	*E. granulosus* (G2)	Kamenetzky et al. (2004)
	AY569371	Eg569371	*E. granulosus* (G7)	"
	AY569370	Eg569370	*E. granulosus* (G6)	"
	AY324085	Ev-14	*E. vogeli*	Present study
	AY324077	Eo-16	*E. oligarthrus*	"
	AY324065	EgCerv-4	*E. granulosus* (G8)	Present study
	AY324080	Ev-9	*E. vogeli*	Present study
	BU493015	Em493015	*E. multilocularis*	Fernandez et al. (2002)
	AY34068	Em-2	"	Present study
	AY324072	Em-6	"	"
	BU493166	Em493166	"	Fernandez et al. (2002)
	AY357114	Eg357114	*E. granulosus* (G1)	Arend et al. (2004)
	AY357115	Eg357115	"	"
	AY357116	Eg357116	"	"
	AY357113	EgA6	"	Haag et al. (2004)
	BI244044	Eg244044	*E. granulosus*	Fernandez and Maizels (2001)
	AF252859	Eg252859	"	Lu et al. (2000)
	AY871032	EgA8	*E. granulosus* (G1)	Haag et al. (2004)
	AY871028	EgA3	"	"
	CN653402	Eg653402	*E. granulosus*	Fernandez and Maizels (2001)
	AY871031	EgA7	*E. granulosus* (G1)	Haag et al. (2004)
	AY357112	EgA4	"	"
	AY871030	EgA5	"	"
	AY871027	EgA2	"	"
	BU493077	Em493077	*E. multilocularis*	Fernandez et al. (2002)

References

Anisimova, M., Bielawski, J.P., Yang, Z., 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. 19, 950–958.

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229–1236.

Arend, A.C., Zaha, A., Ayala, F.J., Haag, K.L., 2004. The *Echinococcus granulosus* antigen B shows a high degree of genetic variability. Exp. Parasitol. 108, 76–80.

Bradley, J.E., Nirmalan, N., Klager, S.L., Faulkner, H., Kennedy, M.W., 2001. River blindness: a role for parasite retinoid-binding proteins in the generation of pathology? Trends Parasitol. 17, 471–475.

Chemale, G., Ferreira, H.B., Barrett, J., Brophy, P.M., Zaha, A., 2005. *Echinococcus granulosus* antigen B hydrophobic ligand binding properties. Biochim. Biophys. Acta 1747, 189–194.

Chemale, G., Haag, K.L., Ferreira, H.B., Zaha, A., 2001. *Echinococcus granulosus* antigen B is encoded by a gene family. Mol. Biochem. Parasitol. 116, 233–237.

Escalante, A.A., Cornejo, O.E., Rojas, A., Udhayakumar, V., Lal, A.A., 2004. Assessing the effect of natural selection in malaria parasites. Trends Parasitol. 20, 388–395.

Fernández, V., Ferreira, H.B., Fernández, C., Zaha, A., Nieto, A., 1996. Molecular charaterisation of a novel 8-kDa subunit of *Echinococcus granulosus* antigen B. Mol. Biochem. Parasitol. 77, 247–250.

Fernandez, C., Brehm, K., Maizels, R.M., 2002. Analysis of genes expressed in *Echinococcus multilocularis* metacestodes. GenBank direct submission, unpublished.

Fernandez, C., Maizels, R.M., 2001. A survey of genes expressed in *Echinococcus granulosus* metacestode. GenBank direct submission, unpublished.

Frosch, P., Hartman, M., Muhlschlegel, F., Frosch, M., 1994. Sequence heterogeneity of the echinococcal antigen B. Mol. Biochem. Parasitol. 64, 171–175.

Furlong, S.T., Caulfield, J.P., 1989. Schistosoma mansoni: synthesis and release of phospholipids, lysophospholipids, and neutral lipids by schistosomula. Exp. Parasitol. 69, 65–77.

González-Sapienza, G., Cachau, R.E., 2003. Identification of critical residues of an immunodominant region of *Echinococcus granulosus* antigen B. J. Biol. Chem. 278, 20179–20184.

González-Sapienza, G., Lorenzo, G., Nieto, A., 2000. Improved immunodiagnosis of cystic hydatid disease by using a synthetic peptide with higher diagnostic value than that of its parent protein *Echinococcus granulosus* antigen B. J. Clin. Microbiol. 11, 3979–3983.

Haag, K.L., Alves-Junior, L., Zaha, A., Ayala, F.J., 2004. Contingent, non-neutral evolution in a multicelullar parasite: natural selection and gene conversion in the *Echinococcus granulosus* antingen B gene family. Gene 333, 157–167.

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

Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.

Kamenetzky, L., Muzulin, P., Gutierrez, A.M., Angel, S.O., Zaha, A., Roserzvit, M.C., 2004. Genetic variation of one of the major antigenic components of *Echinococcus granulosus* cyst fluid, antigen B. GenBank direct submission, unpublished.

Lu, J.H., Gao, J.S., Pan, X.H., Shan, Z.X., Yu, X.B., 2000. Cloning and identification of *Echinococcus* antigen B gene. GenBank direct submission, unpublished.

Mamuti, W., Yamasaki, H., Sako, Y., Nakao, M., Xiao, N., Nakaya, K., Sato, N., Vuitton, D.A., Piarroux, R., Lightowlers, M.W., Craig, P.S., Ito, A., 2004. Molecular cloning, expression and serologic evaluation of an 8-kilodalton subunit of antigen B from *Echinococcus multilocularis*. J. Clin. Microbiol. 42, 1082–1088.

McKerrow, J.H., Salter, J., 2002. Invasion of skin by *Schistosoma cercariae*. Trends Parasitol. 18, 193–195.

Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715–724.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino-acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Pond, S.L.K., Frost, S.D.W., 2005. Not so different at all: a comparison of methods for detecting amino-acid sites under selection. Available at http://www.datamonkey.org/.

Riganò, R., Buttari, B., de Falco, E., Profumo, E., Ortona, E., Marguti, P., Scottà, C., Teggi, A., Siracusano, A., 2004. *Echinococcus granulosus*-specific T-cell lines derived from patients various clinical stages of cystic echinococcosis. Parasite Immunol. 26, 45–52.

Riganò, R., Profumo, E., Bruschi, F., Carulli, G., Azarrà, A., Ioppolo, S., Buttari, B., Ortona, E., Margutti, P., Teggi, A., Siracusano, A., 2001. Modulation of human immune response by *Echinococcus granulosus* antigen B and its possible role in evading host defenses. Infec. Immun. 69, 288–296.

Rott, M.B., Fernández, V., Farias, S., Ceni, J., Ferreira, H.B., Haag, K.L., 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 Trop. 75, 331–340.

Shepherd, J.C., Aitken, A., McManus, D.P., 1991. A protein secreted in vivo by *Echinococcus granulosus* inhibits elastase activity and neutrophil chaemotaxis. Mol. Biochem. Parasitol. 44, 81–90.

Suzuki, Y., 2004a. New methods for detecting positive selection at single amino acid sites. J. Mol. Evol. 59, 11–19.

Suzuki, Y., 2004b. Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene 328, 127–133.

Suzuki, Y., Nei, M., 2004. False-positive selection identified by ML-based methods: examples from the Sig1 gene of the diatom Thalassiosira weissflogii and the tax gene of a human T-cell lymphotrophic virus. Mol. Biol. Evol. 21, 914–921.

Thompson, R.C.A., McManus, D.P., 2002. Towards a taxonomic revision of the genus Echinococcus. Trends Parasitol. 18, 452–457.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Virginio, V., Hernandez, A., Rott, M.B., Monteiro, K.M., Zandonai, A.F., Nieto, A., Zaha, A., Ferreira, H.B., 2003. A set of recombinant antigens from *Echinococcus granulosus* with potential for use in the immunodiagnosis of human cystic hydatid disease. Clin. Exp. Immunol. 132, 309–315.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comp. Appl. Biosci. 13, 555–556.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.

Zhang, J., 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. Mol. Biol. Evol. 21, 1332–1339.

Zhang, W., You, H., Li, J., Zhang, Z., Turson, G., Aili, H., Wang, J., McManus, D.P., 2003. Immunoglobulin profiles in a murine intermediate host model of resistance for *Echinococcus granulosus* infection. Parasite Immunol. 25, 161–168.