© Indian Academy of Sciences

RESEARCH ARTICLE

Variability of the *EG95* antigen-coding gene of *Echinococcus granulosus* in animal and human origin: implications for vaccine development

V. SREEVATSAVA[1], SUMANTA DE[1], S. BANDYOPADHYAY[1], P. CHAUDHURY[2], A. K. BERA[1], RAMACHANDRAN MUTHIYAN[3], ARUN KUMAR DE[3], P. PERUMAL[3], JAI SUNDER[3], GAYATRI CHAKRABORTY[3] and D. BHATTACHARYA[3]*

[1]*Eastern Regional Station, Indian Veterinary Research Institute, Kolkata 700 037, India*
[2]*Indian Veterinary Research Institute, Bareilly 243 122, India*
[3]*ICAR-Central Island Agricultural Research Institute, Port Blair 744 101, India*
*For correspondence. E-mail: debasis63@rediffmail.com.

Abstract. In the present study, the genetic variability of the *EG95* protein-coding gene in several animal and human isolates of *Echinococcus granulosus* was investigated. A total of 24 isolates collected from cattle, buffalo, sheep, goat, dog and man were amplified by *Eg95*-coding gene-specific primers. From the generated sequence information, a conceptual amino acid sequence was deduced. Phylogenetically, the *Eg95* coding gene belongs to the Eg95-1/Eg95-2/Eg95-3/Eg95-4 cluster. Further confirmation on the maximum composite likelihood analysis revealed that the overall transition/transversion bias was 2.913. This finding indicated that there is bias towards transitional and transversional substitution. Using artificial neural networks, a B-cell epitope was predicted on primary sequence information. Stretches of amino acid residues varied between animal and human isolates when hydrophobicity was considered. Flexibility also varied between larval and adult stages of the organism. This observation is important to develop vaccines. However, cytotoxic T-lymphocyte epitopes on primary sequence data remained constant in all isolates. In this study, agretope identification started with hydrophobic amino acids. Amino acids with the same physico-chemical properties were present in the middle. The conformational propensity of the *Eg95*-coding gene of 156 amino acid residues had α-turns and β-turns, and α-amphipathic regions up to 129, 138–156 and 151–155 residues, respectively. The results indicated potential T-cell antigenic sites. The overall Tajima's *D* value was negative (-2.404165), indicative of negative selection pressure.

Introduction

Hydatidosis or cystic echinococcosis (CE) caused by the larval stage of the tapeworm *Echinococcus granulosus* is a classic example of cyclozoonosis (Eckert and Thompson 1997). The parasite is distributed worldwide (Nunnari *et al.* 2012) and causes significant economic losses to the meat industry, mainly due to condemnation of edible offal namely liver and lungs of food animals (Torgerson and Heath 2003). Besides, the disease has a potential public health impact that is recognized by the World Health Organization (WHO 2003). Echinococcosis in man is often expensive and difficult to treat. Sometimes treatment of the disease requires either extensive surgery or prolonged chemotherapy or the simultaneous use of both (WHO 2018). In Asia and Africa, around 50 million people are at risk of developing the disease (Hemphill and Kern 2008). Office International des Epizooties has recognized CE as a multispecies disease since this species has low intermediate host specificity and infects domestic and wild ungulates, and humans (Thompson and McManus 2001; Gauci *et al.* 2002). Intermediate hosts get the infection after the ingestion of parasite eggs excreted by the canid definitive host (Alvarez-Rojas *et al.* 2012). Within the intermediate host, the unilocular cyst develops mainly in the lung and liver (Torgerson and Heath 2003). The cyst contains an inner germinal membrane that is externally supported by an acellular laminated layer. By clonal expansion of the germinal membrane, protoscoleces develop (Thompson *et al.* 1995).

1

Within the host, the helminth parasites survive for a long time. This is possible due to the developed and altered strategies to feed and reproduce after evading host immune attack. The evasion mechanism is mediated by proteins that are collectively known as secreted and membrane bound (S/M) proteins. S/M proteins induce apoptosis of the dendritic cells (Nono *et al.* 2012) and produce an immunosuppressive environment in the infection phase (Huang *et al.* 2016). Since these proteins are expressed before the host immune system, S/M components are candidates for ameliorated diagnosis tests as well as a target for new drugs and immunoprophylactic agents (Rosenzvit *et al.* 2006). A total of 12 S/M proteins have been reported as immunodiagnostic and immunoprophylactic agents (Rosenzvit *et al.* 2006). Of these, Eg95 is a candidate antigen used for immunization of animals. This biomolecule is expressed in the oncosphere and is used for immunization against the larval stage of the disease. Earlier reports suggest that there was a 62% decrease in the prevalence of infection in sheep, which occurs due to the development of the larval stage of the parasite (Lightowlers and Heath 2004; Larrieu *et al.* 2015). This antigen is expressed by oncospheres, protoscoleces and immature and mature adult worms (Zhang *et al.* 2003). A vaccine has been developed based on a recombinant antigen of EG95 and it has been shown to be effective in the prevention of the disease in different intermediate hosts (Lightowlers *et al.* 1996, 1999). A perusal of the available literature suggests that *EG95* is a multigene family and comprises seven isoforms (EG95-1 to EG95-7). All the isoforms except EG95-7 comprise three exons that are separated by two introns (Chow *et al.* 2001). Exon 1 encodes a signal peptide and exons 2 and 3 encode mature peptides (Haag *et al.* 2009).

There is accumulated evidence to show that antigens encoded by a multigene family show a high degree of polymorphism (Arend *et al.* 2004; Kamenetzky *et al.* 2005). This occurs due to natural selection pressure which in turn leads to the diversification of antigen-coding genes. There are two models of natural selection: purifying selection (negative selection) and directional selection (positive selection) (Haag *et al.* 2009). The selection pressure of an antigen-coding gene is exclusively dependent on synonymous and nonsynonymous substitution which occurs due to a change in the nucleotide sequence (Pan *et al.* 2010a). An antigen-coding gene undergoing positive selection pressure exhibits an excess of nonsynonymous substitution. This type of substitution indicates the heterogeneous selection pressure exerted by the host (Haag *et al.* 2009). On the other hand, large deletions are mainly found within introns 1 and 2 of this antigen-coding gene which is a footprint of gene conservation (Haag *et al.* 2009). Therefore, the use of a vaccine based on single recombinant antigens may select for antigenically variant parasites that are ineffective as vaccines (Boubaker

et al. 2014). Hence, the practical use of the developed EG95 vaccine may be affected by the antigenic variability of the EG95 protein expressed in the different intermediate hosts. This problem can be overcome by monitoring antigenic variants among parasite populations and incur necessary modification for appropriate vaccine development (Lightowlers *et al.* 2003). Moreover, extensive isolate variations have been observed in the parasite with 10 designated genotypes (G1–G10) that may impact the epidemiology, pathology and control of CE (Spotin *et al.* 2017).

Allelic polymorphisms of the *Eg95* gene may impact the efficacy of the vaccine based on *Eg95* (Gauci *et al.* 2018). Therefore, understanding the polymorphism of the *EG95* genes in different *E. granulosus* isolates is very important. Moreover, it has been observed that EG95 epitopes are conformational (Woollard *et al.* 2000), indicating that a few amino acid substitutions may affect protection dramatically (Haag *et al.* 2009). A crucial step to design a peptide vaccine involves the identification of T cell, B cell and cytotoxic T-lymphocyte (CTL) epitopes. The experimental scanning of the epitopes requires the synthesis of overlapping peptides that span the entire sequence of a protein antigen. But this approach is costly and labour intensive. The alternative approach is to find out the epitopes by *in silico* analysis (Bhasin and Raghava 2004). In the present study, we investigated the genetic variability of the *EG95* protein-coding gene in several animal and human isolates of *E. granulosus*. The antigenic variability of the deduced EG95 proteins was also assessed, and B and T cell epitopes were predicted to have huge implications for stain-specific vaccine development.

Materials and methods

Chemicals, reagents, media and buffers

Molecular biology and analytical grade chemicals were procured from Fermentas, Hi-Media, Amresco, Progema, SRL, CISCO, Merck and Sigma. Media and buffers used in this study were purchased from Hi-Media, SRL and Difco.

Collection of parasite samples

Cyst samples from animals were collected from local abattoirs situated at Kolkata, West Bengal. Adult parasites were collected from experimentally infected pups. Human cyst samples were collected from RG Kar Medical College Hospital, Kolkata, West Bengal, India. The collected samples were transported to the laboratory by maintaining an appropriate cold chain. Parasitic samples from 21 animals and three human patients were collected (table 1).

Table 1. *E. granulosus* isolates used for the genetic diversity of EG95.

| Isolate designation	Genomic DNA source	Host	Genotype
B1	Px-sc	Buffalo	G1
B2	Px-sc	Buffalo	G1
B3	Px-sc	Buffalo	G1
B11	Px-sc	Buffalo	G1
B25	Px-sc	Buffalo	G1
B27	Px-sc	Buffalo	G1
B30	Px-sc	Buffalo	G1
B69	Px-sc	Buffalo	G5
C2	Px-sc	Cattle	G1
C8	Px-sc	Cattle	G1
S1	Px-sc	Sheep	G1
S2	Px-sc	Sheep	G1
S3	Px-sc	Sheep	G1
S4	Px-sc	Sheep	G1
S20	Px-sc	Sheep	G1
G1	Px-sc	Goat	G1
G3	Px-sc	Goat	G1
G4	Px-sc	Goat	G1
G8	Px-sc	Goat	G1
G12	Px-sc	Goat	G1
H1	Px-sc	Human	G1
H2	Px-sc	Human	G1
H3	Px-sc	Human	G1
D1	Aw	Dog	G1

Px-sc, protoscoleces of the metacestode stage of *E. granulosus*; Aw, adult worms; G1, common sheep strain; G5, cattle strain. Parasite isolates were genotyped based on a PCR-RFLP method (Pan *et al.* 2010a).

Isolation of genomic DNA

Genomic DNA from fresh protoscoleces of individual hydatid cysts and from adult worms was extracted using a commercial kit (Q-BIOgene, Carlsbad, USA) as per the protocol of the manufacture, as previously described (Pan *et al.* 2010b). Following the phenol–chloroform extraction, genomic DNA was precipitated with isopropanol and dissolved in Tris-EDTA buffer (pH 8.0) and kept at −80°C until further use.

Identification of the genotype of isolates of E. granulosus

The identification of the genotype was performed using a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method as described previously (Pan *et al.* 2009). In brief, a 770-bp fragment was amplified by PCR and the amplified fragment was digested with *Alu*1 and *Cas*p61 restriction enzymes. Genotypes were identified by restriction profiles of the fragment (Pan *et al.* 2009).
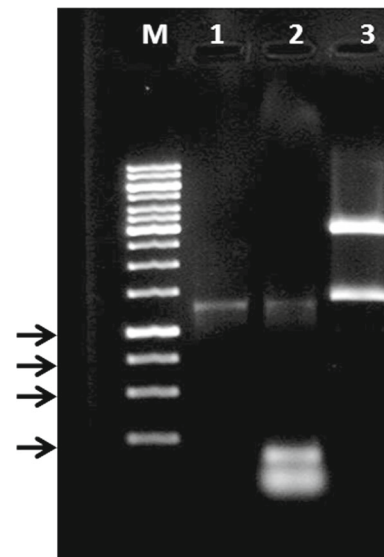
Figure 1. Clone confirmation of the *Eg95* coding gene. M = 1-kb DNA ladder (arrows from the bottom indicating 250, 500, 750 and 1000 bp). Lane 1, PCR product; lane 2, colony PCR product; lane 3; released product after RE digestion.

Amplification of the EG95 antigen-coding region

The EG95-coding region was amplified by PCR using primers described by Haag *et al.* (2009). The PCR was carried out in a total volume of 25 μL containing DNA (100 ng), 200 μM of deoxyribonucleotide triphosphates, 10× PCR buffer, 2.5 mM $MgCl_2$, 10 pmol each of forward and reverse primers and 1 unit of *Taq* DNA polymerase. The PCR was carried out in an initial denaturation (94°C) for 5 min followed by 37 cycles. For the first two cycles, the annealing temperature was 60°C, followed by 55°C for the next 10 cycles with decrement of 1°C after each cycle. For the last 25 cycles, the annealing temperature was 45°C. For all the cycles, 30 s each was given for denaturation and annealing, and 1 min was given for extension. The confirmation of the amplicon was performed by agarose gel electrophoresis (figure 1).

Cloning of the PCR product and sequencing

Amplicons were cloned into the pTZ57R/T vector using the InsTAclone PCR cloning kit (Fermentas, Burlington, USA) following the manufacturer's protocol. The ligation product was transformed into *E. coli* DH5α competent cells. The recombinant clones were identified by blue/white selection and the positive white colonies were analysed for the confirmation of correct insert and orientation in the purified plasmids by the colony PCR and by double digestion with *Eco*R1 and *Bam*H1 (figure 1). Then, the purified plasmids were sequenced with standard primers (M13F and M13R) using a chain termination method.

The prediction of the promiscuous MHC-II binding site
was evaluated as described earlier by Singh and Raghava
(2001). The B-cell epitope was predicted using online soft-
ware developed by Saha and Raghava (2006). The CTL
epitope was predicted as described by Bhasin and Raghava
(2004). An antigenic index and conformational propensity
were deduced by Lasergene 8.0 (DNA Star, USA) and a
PSIPRED secondary structure prediction method (Jones
1999). Tajima's neutrality test and transition/transversion
bias were calculated using MEGA v5.1 (Tamura *et al.*
2007).

Phylogenetic analysis

Multiple sequence alignment of the deduced amino acid
sequences was performed using Clustal W. Phylogenetic
analysis was performed by the distance-based neighbour-
joining method using MEGA 5.1 (CEMI, Tempe, USA).
Tree evaluation was performed by bootstrapping for 1000
replicates to test the accuracy of the phylogenetic tree.
Sequence information was compared with the available
sequence information of EG 95-1 (AAL35393), Eg95-2
(AAG40127), Eg95-3 (AAG40128), Eg95-4 (AAG40124),
Eg95-5 (AAG40125) and Eg95-6 (AAG40123).

Results

Genotype of isolates of E. granulosus

The genotype of eight buffalo, two cattle, five sheep, five
goat, three human isolates and one adult worm was per-
formed through PCR-RFLP. It was found that all isolates
belonged to the common sheep strain (G1) except one buf-
falo isolate that was a cattle strain (G5) (table 1).

Pairwise amino acid sequence variation

The nucleotide sequences of the EG95-coding region of 24
isolates used in the study were generated and the sequences
were deposited in a public database (NCBI with acces-
sion numbers HM345583–HM345607). The conceptual
amino acid sequences were deduced using universal genetic
code. The multiple sequence alignment based on the
deduced-amino acid sequences along with their conserved
and variable regions of the 24 isolates used in the study and
the EG95 of six isoforms is presented in figure 2. On the
basis of the amino acid sequence variation, one goat (G8)
and two sheep (S3 and S4) isolates were indistinguishable
from each other. Similarly, both goat isolates (G4 and G12)
and three human isolates (H1, H2 and H3) were indistin-
guishable from each other. Similarity of the amino acid

sequence was also observed between two goat isolates (G1
and G3) and one buffalo isolate (B1). Sequence variation
between 24 isolates varied from 1.21 to 11.56% (table 2).

Phylogenetic analysis

From the phylogenetic analysis, it was revealed that four
isoforms of the *EG95* gene described in the centralized
repositories of the public database (Eg95-1, EG95-2, Eg95-
3 and Eg95-4) belonged to one cluster. On the other hand,
EG95-5 and EG95-6 belonged to another cluster. The iso-
lates characterized in the present study were homologous
to four isoforms (Eg95-1, Eg95-2, Eg95-3 and Eg95-4) of
the *Eg95*-coding gene (figure 3).

The probability of substitution was estimated by using
the maximum composite likelihood estimate. Transitional
and transversional substitution varied from 13.08 to
27.26 and 2.68 to 3.33, respectively. The overall tran-
sition/transversion bias (R) was 2.913. The patterns of
nucleotide substitution for both purine and pyrimidine
bases are shown in table 3.

Variation with respect to stretches of α-helix, β-sheet,
α-amphipathic and β-amphipathic was not observed
between animal and human isolates. Positive values of
the antigenic index were seen up to 131–137 amino acid
residues in both animal and human isolates (table 4).

Physico-chemical properties of the amino acids

Physico-chemical properties of the deduced-amino acid
sequence of the 24 isolates are presented in table 5. The
positive GRAVY value indicates that the protein is polar
in nature.

Prediction of B-cell and CTL epitopes

The prediction of the B-cell epitope was done on the basis
of seven physico-chemical properties. Stretches of amino
acid sequences varied between animal and human isolates
when hydrophobicity was considered. Stretches of amino
acid sequences varied between the larval stage and adult
stage of the parasite when flexibility was considered.

Further, stretches of amino acid sequences varied
between the larval stage of animal and human isolates, as
well as the adult stage of canine isolates when accessibility
was taken into consideration. Only one stretch (LRN-
HFNLT) of the amino acid sequence was seen within the
sequence information of the amino acid sequence of the
adult stage of *E. granulosus* when turns within the sequence
were assessed. Stretches of amino acid sequences varied
between animal and human isolates when the exposed
surface was considered. However, stretches of amino acid
sequences did not vary between all isolates when polarity
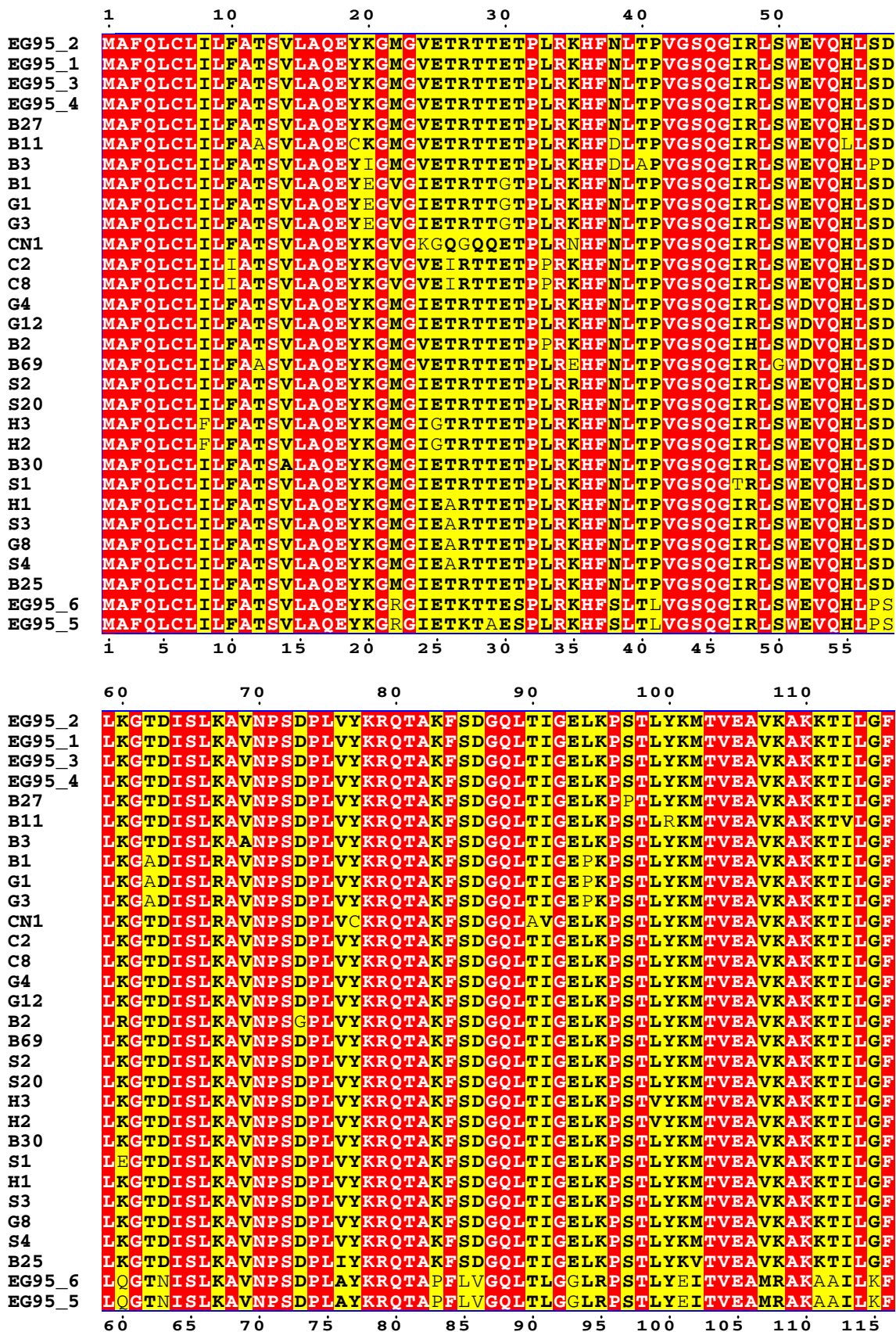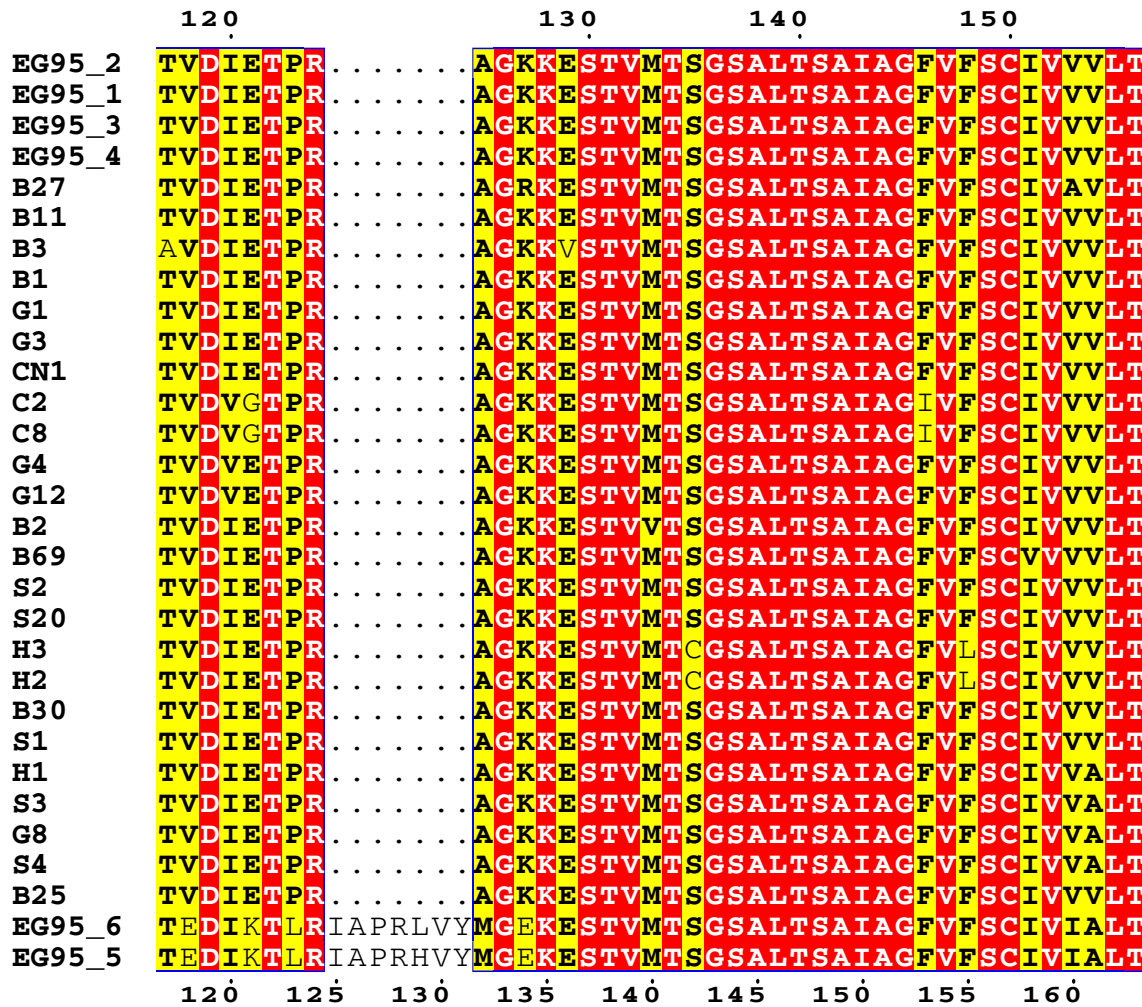and antigenic propensity were considered (table 6).

```
            1        10        20        30        40        50
EG95_2   MAFQLCLILFATSVLAQEYKGMGVETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
EG95_1   MAFQLCLILFATSVLAQEYKGMGVETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
EG95_3   MAFQLCLILFATSVLAQEYKGMGVETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
EG95_4   MAFQLCLILFATSVLAQEYKGMGVETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
B27      MAFQLCLILFATSVLAQEYKGMGVETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
B11      MAFQLCLILFAASVLAQECKGMGVETRTTETPLRKHFDLTPVGSQGIRLSWEVQLLSD
B3       MAFQLCLILFATSVLAQEYIGMGVETRTTETPLRKHFDLAPVGSQGIRLSWEVQHLPD
B1       MAFQLCLILFATSVLAQEYEGVGIETRTTGTPLRKHFNLTPVGSQGIRLSWEVQHLSD
G1       MAFQLCLILFATSVLAQEYEGVGIETRTTGTPLRKHFNLTPVGSQGIRLSWEVQHLSD
G3       MAFQLCLILFATSVLAQEYEGVGIETRTTGTPLRKHFNLTPVGSQGIRLSWEVQHLSD
CN1      MAFQLCLILFATSVLAQEYKGVGKGQGQQETPLRNHFNLTPVGSQGIRLSWEVQHLSD
C2       MAFQLCLILIATSVLAQEYKGVGVEIRTTETPPRKHFNLTPVGSQGIRLSWEVQHLSD
C8       MAFQLCLILIATSVLAQEYKGVGVEIRTTETPPRKHFNLTPVGSQGIRLSWEVQHLSD
G4       MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRKHFNLTPVGSQGIRLSWDVQHLSD
G12      MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRKHFNLTPVGSQGIRLSWDVQHLSD
B2       MAFQLCLILFATSVLAQEYKGMGVETRTTETPPRKHFNLTPVGSQGIHLSWDVQHLSD
B69      MAFQLCLILFAASVLAQEYKGMGVETRTTETPLREHFNLTPVGSQGIRLGWDVQHLSD
S2       MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRRHFNLTPVGSQGIRLSWEVQHLSD
S20      MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
H3       MAFQLCLFLFATSVLAQEYKGMGIGTRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
H2       MAFQLCLFLFATSVLAQEYKGMGIGTRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
B30      MAFQLCLILFATSALAQEYKGMGIETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
S1       MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRKHFNLTPVGSQGTRLSWEVQHLSD
H1       MAFQLCLILFATSVLAQEYKGMGIEARTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
S3       MAFQLCLILFATSVLAQEYKGMGIEARTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
G8       MAFQLCLILFATSVLAQEYKGMGIEARTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
S4       MAFQLCLILFATSVLAQEYKGMGIEARTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
B25      MAFQLCLILFATSVLAQEYKGMGIETRTTETPLRKHFNLTPVGSQGIRLSWEVQHLSD
EG95_6   MAFQLCLILFATSVLAQEYKGRGIETKTTESPLRKHFSLTLVGSQGIRLSWEVQHLPS
EG95_5   MAFQLCLILFATSVLAQEYKGRGIETKTAESPLRKHFSLTLVGSQGIRLSWEVQHLPS
            1    5   10   15   20   25   30   35   40   45   50   55
```

```
            60        70        80        90        100       110
EG95_2   LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
EG95_1   LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
EG95_3   LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
EG95_4   LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
B27      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPPTLYKMTVEAVKAKKTILGF
B11      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLRKMTVEAVKAKKTVLGF
B3       LKGTDISLKAANPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
B1       LKGADISLRAVNPSDPLVYKRQTAKFSDGQLTIGEPKPSTLYKMTVEAVKAKKTILGF
G1       LKGADISLRAVNPSDPLVYKRQTAKFSDGQLTIGEPKPSTLYKMTVEAVKAKKTILGF
G3       LKGADISLRAVNPSDPLVYKRQTAKFSDGQLTIGEPKPSTLYKMTVEAVKAKKTILGF
CN1      LKGTDISLRAVNPSDPLVCKRQTAKFSDGQLAVGELKPSTLYKMTVEAVKAKKTILGF
C2       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
C8       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
G4       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
G12      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
B2       LRGTDISLKAVNPSGPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
B69      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
S2       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
S20      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
H3       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTVYKMTVEAVKAKKTILGF
H2       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTVYKMTVEAVKAKKTILGF
B30      LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
S1       LEGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
H1       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
S3       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
G8       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
S4       LKGTDISLKAVNPSDPLVYKRQTAKFSDGQLTIGELKPSTLYKMTVEAVKAKKTILGF
B25      LKGTDISLKAVNPSDPLIYKRQTAKFSDGQLTIGELKPSTLYKVTVEAVKAKKTILGF
EG95_6   LQGTNISLKAVNPSDPLAYKRQTAPFLVGQLTLGGLRPSTLYEITVEAMRAKAAILKF
EG95_5   LQGTNISLKAVNPSDPLAYKRQTAPFLVGQLTLGGLRPSTLYEITVEAMRAKAAILKF
           60   65   70   75   80   85   90   95  100  105  110  115
```

Figure 2. *(contd)*

```
                    120                130            140              150
EG95_2    TVDIETPR........AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
EG95_1    TVDIETPR........AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
EG95_3    TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
EG95_4    TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
B27       TVDIETPR.......AGRKESTVMTSGSALTSAIAGFVFSCIVAVLT
B11       TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
B3        AVDIETPR.......AGKKVSTVMTSGSALTSAIAGFVFSCIVVVLT
B1        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
G1        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
G3        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
CN1       TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
C2        TVDVGTPR.......AGKKESTVMTSGSALTSAIAGIVFSCIVVVLT
C8        TVDVGTPR.......AGKKESTVMTSGSALTSAIAGIVFSCIVVVLT
G4        TVDVETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
G12       TVDVETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
B2        TVDIETPR.......AGKKESTVVTSGSALTSAIAGFVFSCIVVVLT
B69       TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCVVVVLT
S2        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
S20       TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
H3        TVDIETPR.......AGKKESTVMTCGSALTSAIAGFVLSCIVVVLT
H2        TVDIETPR.......AGKKESTVMTCGSALTSAIAGFVLSCIVVVLT
B30       TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
S1        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
H1        TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVALT
S3        TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVALT
G8        TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVALT
S4        TVDIETPR.......AGKKESTVMTSGSALTSAIAGFVFSCIVVALT
B25       TVDIETPR......AGKKESTVMTSGSALTSAIAGFVFSCIVVVLT
EG95_6    TEDIKTLRIAPRLVYMGEKESTVMTSGSALTSAIAGFVFSCIVIALT
EG95_5    TEDIKTLRIAPRHVYMGEKESTVMTSGSALTSAIAGFVFSCIVIALT
            120   125   130   135   140   145   150   155   160
```

Figure 2. Amino acid sequences of the *Eg95*-coding gene of *E. granulosus* isolated from animals and humans. Red residues indicate an amino acid residue that is conserved in between different pieces of sequence information. Yellow residues differ from the consensus.

Table 2. Pairwise amino acid sequence variation of the *EG95*-coding gene of animal and human isolates of *E. granulosus*.

	B30	S20	G8/S3/S4	G4/G12	S1	B25	B27	B69	H1/H2/H3	B2	B1/G1/G2	B11	S2/C2/C8	B3	D1
B30	–	1.21	1.82	1.82	1.82	1.82	3.65	4.26	3.65	4.87	4.26	4.87	5.48	5.48	7.92
S20	1.21	–	1.82	1.82	1.82	1.82	3.04	3.65	3.65	4.87	4.26	4.87	5.48	5.48	7.31
G8/S3/S4	1.82	1.82	–	2.43	2.43	2.43	3.65	4.87	4.26	5.48	4.87	5.48	5.48	6.09	7.92
G4/G12	1.82	1.82	2.43	–	2.43	2.43	3.65	3.65	4.26	4.26	4.87	5.48	4.87	6.09	8.53
S1	1.82	1.82	2.43	2.43	–	2.43	3.65	4.87	4.26	4.87	4.87	5.48	6.09	6.09	8.53
B25	1.82	1.82	2.43	2.43	2.43	–	3.65	4.87	4.26	5.48	4.87	5.48	6.09	6.09	8.53
B27	3.04	3.04	3.65	3.65	3.65	3.65	–	4.87	5.48	5.48	6.09	5.48	6.09	6.09	9.14
B69	4.26	3.65	4.87	3.65	4.87	4.87	4.87	–	6.70	5.48	7.31	5.48	7.31	7.31	9.75
H1/H2/H3	3.65	3.65	4.26	4.26	4.26	4.26	5.48	6.70	–	7.31	6.70	7.31	7.92	7.92	9.14
B2	4.87	4.87	5.48	4.26	4.87	5.48	5.48	5.48	7.31	–	7.92	7.31	6.70	7.29	10.97
B1/G1/G2	4.26	4.26	4.87	4.87	4.87	4.87	6.09	7.31	6.70	7.92	–	7.92	7.31	7.92	8.53
B11	4.87	4.87	5.48	5.48	5.48	5.48	5.48	5.48	7.31	7.31	7.92	–	7.92	6.70	10.97
S2/C2/C8	5.48	5.48	5.48	4.87	6.09	6.09	6.09	7.31	7.92	6.70	7.31	7.92	–	8.53	9.75
B3	5.48	5.48	6.09	6.09	6.09	6.09	6.09	7.31	7.92	7.92	7.92	6.70	8.53	–	11.58
D1	7.92	7.31	7.92	8.53	8.53	8.53	9.14	9.75	9.14	10.97	8.53	10.97	9.75	11.58	–

B, buffalo; C, cattle; G, goat; D, adult worm from dog; H, human isolate.
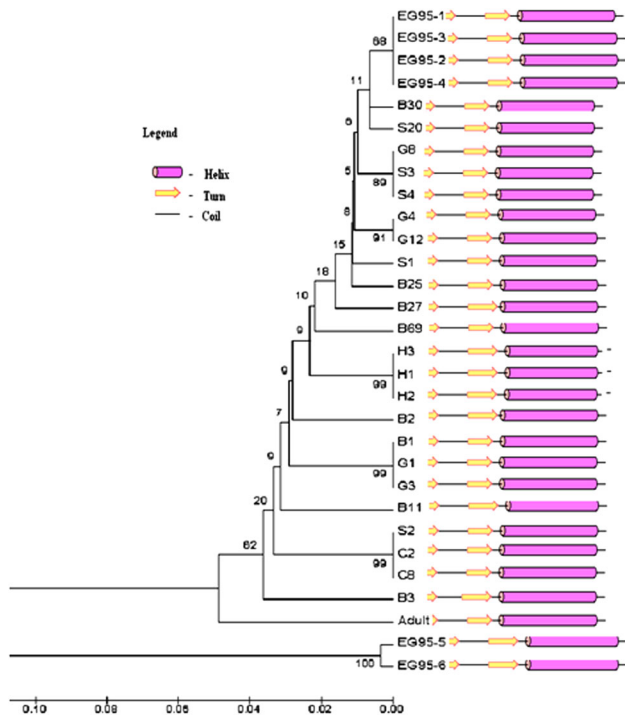
Figure 3. Phylogenetic analysis on the amino acid sequence of the *Eg95* coding gene of *E. granulosus* by the unweighted pair group method. Numbers at nodes represent percentage occurrence of clades in 1000 bootstrap replications (B, buffalo isolate; C, cattle isolate; S, sheep isolate; G, goat isolate; Adult, adult worm; H, human isolate) and conformational propensity of the individual isolate.

Table 3. Maximum composite likelihood estimate of the pattern of nucleotide substitution.

	A	T	C	G
A	–	*2.95*	*2.68*	**21.96**
T	*3.33*	–	**13.08**	*2.68*
C	*3.33*	**14.41**	–	*2.68*
G	**27.26**	*2.95*	*2.68*	–

A, adenine; T, thymine; C, cytosine; G, guanine. Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of different transitional substitutions have been shown in bold and those of transversional substitutions are shown in italics. The nucleotide frequencies are 0.286 (A), 0.253 (T/U), 0.23 (C) and 0.23 (G). The transition/transversion rate ratios are $k_1 = 8.182$ (purines) and $k_2 = 4.88$ (pyrimidines). The overall transition/transversion bias was $R = 2.913$, where $R = [A^*G^*k_1 + T^*C^*k_2]/[(A + G)^*(T + C)]$. Codon positions included were 1st + 2nd + 3rd. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 470 positions in the final dataset.

The prediction of the CTL epitope revealed three stretches of CTL epitopes within 164 amino acid residues which started at 48, 92 and 142 positions. Detailed information on CTL epitopes is depicted in table 7.

Prediction of the promiscuous MHC-II binding site

The larval stage isolated from animals exhibited more than 50% binding propensity with 17 MHC-II alleles. The larval stage of human isolates showed more than 50% binding propensity with 13 MHC-II alleles. The adult stage of *E. granulosus* presented more than 50% binding propensity with 18 MHC-II alleles (table 8).

Selection pressure of the EG95-coding gene

Twenty-four isolates considered in the study had 326 segregating sites. By sliding window analysis, the sites of nonsynonymous and synonymous valley were analysed. The highest Tajima's *D* value (−1.8923) was observed within 1–101 nucleotide bases. The lowest Tajima's *D* value was observed within 302–401 nucleotide stretches (table 9, figure 4). The overall *D* value of 471 bases was − 2.404165. The value was statistically significant (*P* < 0.001).

Discussion

E. granulosus has a worldwide distribution and is an important zoonotic parasite. Although domestic livestock are its most important host, numerous herbivorous wild life species may also act as intermediate hosts (Jenkins and Macpherson 2003). Of the 10 reported genotypes (G1–G10), common sheep strain (G1) is promiscuous in its range of suitable intermediate host species (Barnes *et al.* 2009). This strain is distributed in parts of South America, southern and eastern Europe, northern and eastern Africa and parts of Asia and the Australian region (Eckert and Thompson 1997). Most organisms isolated from human patients have been confirmed as belonging to the common sheep strain (Arbabi *et al.* 2017). In the present study, of the 24 isolates of *E. granulosus* (both larvae and adults), 23 belonged to the common sheep strain (table 1). Therefore, considering the zoonotic significance and importance of the strain, control of the transmission of this organism is warranted.

Control of the disease is possible by preventing the infection of the definitive host or by immunization of the intermediate hosts. Immunization against the parasite is possible after understanding the protective immune response that allows for the identification of antigens eliciting the protective immune response (Craig *et al.* 2007). Immunization against the parasite may be performed by using S/M proteins that are expressed by the parasite (Rosenzvit *et al.* 2006). Previous vaccination trials indicated that EG95 induced 95% protection against hydatid infection in an experimental trial (Lightowlers *et al.* 1996). Later on, the vaccine was proved to be effective in Australia, New Zealand, Argentina and China (Lightowlers *et al.* 1999; Larrieu *et al.* 2015). The vaccine has also been

Table 4. Antigenic index and conformational propensity of the *Eg95* coding gene of 156 amino acid residues.

| Name of the isolate | Stretches of amino acid residues | | | | |
	α-Region (Chou–Fasman)	β-Region (Chou–Fasman)	α-Amphipathic regions (Eisenberg)	β-Amphipathic regions (Eisenberg)	Antigenic index (Jameson–Wolf)
Larval stage of animal isolates	1–18, 47–60, 65–69, 77–83, 97–114, 124–129	87–90, 115–120, 128–132, 137–155	16–18, 25, 28–39, 41–54, 63–68, 70–73, 77–84, 86–88, 98–103, 104–106, 107–112, 113–115, 120–130, 140–142, 155	1–2, 3–6, 7–8, 21–29, 32–39, 40–41, 44–58, 61–64, 66–69, 77–80, 81–83, 85–86, 89–93, 95–100, 104–105, 106–107, 110–120, 122, 110–120, 122	14–15, 16–35, 36–37, 40–43, 45–47, 54–55, 57–66, 68–76, 79–88, 90–97, 102–111, 119–131, 134–137
Larval stage of human isolates	1–18, 47–60, 65–69, 77–83, 105–114, 121–129	23–30, 88–90, 98–104, 109–120, 128–131, 136–138	16–18, 20–39, 42–54, 64–68, 71–72, 79–83, 87–88, 99–103, 105, 108–111, 114–115, 121–128, 136–137, 140–141, 151	1–6, 8, 23–24, 27, 29, 32–39, 41, 43–58, 62–69, 78–80, 82–83, 85–86, 89–92, 95–101, 104–105, 107, 109–120, 122	17–24, 27–37, 41–47, 55, 56–66, 67–77, 79–88, 91–98, 102–112, 119–131
Adult worm	1–18, 47–60, 75–83, 88–93, 98–114, 125–129	19–22, 65–69, 115–120, 128–132, 137–156	15–25, 29, 33–39, 42–54, 62–73, 78–83, 88, 99–103, 105, 108–112, 114–115, 121–130, 140–141, 151	2,4,6,8,20–39, 41, 44–58, 60–70, 78–80, 82–83, 85, 89–91, 95–101, 104, 107, 110–120, 122	17–37, 40–47, 55, 57–66, 68–76, 79–88, 91–97, 102–110, 119–131, 135–136

Table 5. Physico-chemical parameters of the deduced amino acid sequence of the 24 isolates.

Isolate name	Length (aa)	MW	PI	Extinction coefficients	Instability index	Aliphatic index	Gravy
Buffalo isolates (*n* = 8)							
B1	156	16835.56	9.17	10, 095	27.37	96.22	0.122
B2	156	16831.57	9.33	10, 095	26.61	94.94	0.110
B3	156	16820.69	9.14	10, 095	26.43	100.00	0.221
B11	156	16808.69	9.24	7115	28.49	98.08	0.169
B25	156	16938.76	9.30	10, 095	24.65	98.72	0.124
B27	156	16952.77	9.33	10, 095	25.43	94.36	0.081
B30	156	16928.74	9.30	10, 095	27.16	95.00	0.092
B69	156	16855.61	8.91	10, 095	25.25	95.58	0.125
Cattle isolates (*n* = 2)							
C2	156	16752.60	9.45	10, 095	20.10	101.79	0.159
C8	156	16752.60	9.45	10, 095	20.10	101.79	0.159
Goat isolates (*n* = 5)							
G1	156	16835.56	9.17	10, 095	27.37	96.22	0.122
G3	156	16835.56	9.17	10, 095	27.37	96.22	0.122
G4	156	16928.74	9.30	10, 095	24.34	95.58	0.106
G8	156	16898.72	9.30	10, 095	27.16	95.64	0.108
G12	156	16928.74	9.30	10, 095	24.34	95.58	0.106
Sheep isolates (*n* = 5)							
S1	156	16945.68	8.91	10, 095	28.25	93.72	0.077
S2	156	16984.81	9.33	10, 095	32.06	96.22	0.104
S3	156	16898.72	9.30	10, 095	27.16	95.64	0.108
S4	156	16898.72	9.30	10, 095	27.16	95.64	0.108
S20	156	16984.81	9.33	10, 095	32.06	96.22	0.104
Human isolates (*n* = 3)							
H1	156	16898.72	9.30	10, 095	27.16	95.64	0.108
H2	156	16886.77	9.35	10, 095	22.08	95.58	0.147
H3	156	16886.77	9.35	10, 095	22.08	95.58	0.147
Canine isolate (*n* = 1)							
CN1	156	16759.48	9.24	8605	27.00	95.58	0.098

Table 6. Predicted B cell epitopes of the Eg95 antigen of animal and human origin.

Physico-chemical properties	Larval stage of animal isolates	Larval stage of human isolates	Adult stage of canine isolates
Hydrophilicity	ETRTTETP (25), SDLKGTD (57), ETPRAGKKESTV (121)	GTRTTETP (25), SDLKGTD (57), ETPRAGKKESTV (121)	GKGQGQQETP (23), SDLKGTD (57), ETPRAGKKESTV (121)
Flexibility	MGVETRTT (22), IETPRAGKKES (120)	MGIGTRTT (22), IETPRAGKKES (120)	QEYKGVGKGQGQ (17), IETPRAGKKES (120)
Accessibility	LAQEYKGMGVETR (15), TTETPLRKHFNLT (28), KAVNPSDPLVYKRQT (67), AKFSD (82), GELKPSTLYK (92), EAVKAKKTIL (105), DIETPRAGKKESTVM (119)	LAQEYKGMG (15), GTRTTET-PLRKHFNLT (25), KANPSDPLVYKQ-TAKFSD (67), GELKPSTVYK (92), EAVKAKKTIL (105), DIET-PRAGKKESTVM (119)	LAQEYKGVGKGQGQQE TPLRNHFN (15), RAVNPSD (67), KQTAKFSD (78), GELKPSTLYK (92), EAVKAKKTIL (105), DIET-PRAGKKESTVM (119)
Turns	–	–	LRNHFNLT (33)
Exposed surface	KPSTLYK (95), EAVKAKK (105), TPRAGKKES (122)	EAVKAKK (105), TPRAGKKES (122)	KPSTLYK (95), EAVKAKK (105), TPRAGKKES (122)
Polarity	ETRTTETPLRKHFNLT (25), EAVKAKK (105), ETPRAGKKESTVM (121)	TETPLRKHFNLY (29), EAVKAKK (105), ETPRAGKKESTVM (121)	ETPLRMH (30), EAVKAKK (105), ETPRAGKKESTVM (121)
Antigenic propensity	FQLCLILF (3), VQHLSDL (53), PSDPLVYKQT (71), GFVFSCIVVVLT (147)	FQLCLFLF (3), VQHLSDL (53), PSDPLVYKQT (71), LKPSTVY (94), GFVLSCIVVVLT (147)	FQLCLILF (3), VQHLDSL (53), PSDPLVCKQT (71), GFVFSCIVVVLT (147)

Start position of the sequence is indicated in parentheses.

Table 7. CTL epitopes of the EG95 antigen of animal and human origin.

Sequence (start position)		
Larval stage isolated from animals	Larval stage isolated from humans	Adult stage isolated from canines
RLSWEVQHL (48)	RLSWEVQHL (48)	RLSWEVQHL (48)
IAGFVFSCI (142)	FVLSCIVVV (145)	IAGFVFSCI (142)
ELKPSTLYK (92)	ELKPSTYK (92)	ELKPSTYK (92)

attempted to be used in wild animals with 96–100% protective efficacy (Barnes *et al.* 2009).

In the past, based on preliminary reports, it was postulated that Eg95 has broad applicability as a vaccine (Lightowlers *et al.* 1999). This conclusion was arrived since there was no available information on the genetic variability of the gene encoding Eg95 antigen, but with the increase noesis in molecular biology, it has been worked out that the *Eg95*-encoding gene is a member of a multigene family (Zhang *et al.* 2003). DNA sequence analysis of cloned genomic DNA indicated that the *Eg95* gene family consists of at least seven members, one of which

(Eg95-7) is a pseudogene (Chow *et al.* 2001). Therefore, the use of a vaccine based on a single recombinant antigen (Eg95) may select for antigenitically variant parasites that are insusceptible to the vaccine. This problem may be ducked by monitoring the parasite population for antigenic variants and modifying the vaccine appropriately (Lightowlers *et al.* 2003).

As far as the phylogenetic relationship is concerned, Eg95-1 and Eg95-2, and Eg95-3 and Eg95-4, belong to one cluster. In contrast, Eg-95-5 and Eg-95-6 belong to another cluster and have 76–77% similarity with Eg95-1 (Chow *et al.* 2001). In the present study, from phylogenetic

Table 8. Stretches of the agretope within the *Eg95* coding antigen along with the more than 50% binding propensity with MHC-II alleles.

Type of isolate	Name of the allele	Amino acid sequence of stretches of agretope	Per cent binding propensity
Larval stage of animal origin	DRB1_0309	FVFSCIVVV	50.53
	DRB1_0404	LILFATSAL	51.14
	DRB1_0410	LILFATSAL	51.51
	DRB1_0423	LILFATSAL	51.14
	DRB1_0701	FVFSCIVVV	64.66
	DRB1_0703	FVFSCIVVV	64.66
	DRB1_0801	LVYKQTAKF	56.98
	DRB1_0804	LVYKQTAKF	62.50
	DRB1_0806	LVYKQTAKF	68.60
	DRB1_0813	ISLKAVNPS	52.87
	DRB1_1101	FQLCILFA	53.01
	DRB1_1128	FQLCLILFA	50.57
	DRB1_1301	LVYKQTAKF	61.36
	DRB1_1305	FQLCLILFA	50.57
	DRB1_1327	LVYKQTAKF	61.36
	DRB1_1328	LVYKQTAKF	61.36
	DRB1_1501	LILFATSAL	53.67
	DRB1_1506	LILFATSAL	53.67
Larval stage of human origin	DRB1_0309	FVLSCIVVV	52.63
	DRB1_0410	LFLFATSVL	55.32
	DRB1_0701	FVLSCIVVV	66.38
	DRB1_0703	FVLSCIVVV	66.38
	DRB1_0801	LVYKQTAKF	56.98
	DRB1_0804	LVYKQTAKF	62.50
	DRB1_0806	LVYKQTAKF	68.60
	DRB1_0813	ISLKAVNPS	52.87
	DRB1_1301	LVYKQTAKF	61.36
	DRB1_1327	LVYKQTAKF	61.36
	DRB1_1328	LVYKQTAKF	61.36
	DRB1_1501	LFLFATSVL	50.61
	DRB1_1506	LFLFATSVL	50.61
Adult isolate of canine origin	DRB1_0309	FVFSCIVVV	50.53
	DRB1_0402	LRNHFLTP	51.04
	DRB1_0404	LILFATSVL	51.14
	DRB1_0410	LILFATSVL	58.51
	DRB1_0423	LILFATSVL	51.14
	DRB1_0701	FVFSCIVVV	64.66
	DRB1_0703	FVFSCIVVV	64.66
	DRB1_0804	LVCKQTAKF	52.50
	DRB1_0806	LVCKQTAKF	59.30
	DRB1_0813	SLRAVNPS	52.87
	DRB1_1101	FQLCLILFA	53.01
	DRB1_1128	FQLCLILFA	50.57
	DRB1_1301	LVCKQTAKF	52.27
	DRB1_1305	FQLCLILFA	50.57
	DRB1_1327	LVCKQTAKF	52.27
	DRB1_1328	VCKQTAKF	52.27
	DRB1_1501	LRNHFNLTP	54.08
	DRB1_1506	LRNHFNLTP	54.08

analysis (figure 3) and pairwise amino acid sequence variation (figure 2), it can be concluded that all isolates belonged to the Eg95-1/Eg95-2/Eg-95-3/Eg95-4 cluster. Therefore, from the present findings, it may be concluded that both the strains considered during the study (G1 and G5) can be used without any chance of vaccine resistance due to the absence of antigenic flexibility which obviously excludes the chances of common serotype replacement

(Kennedy and Read 2017). This finding is helpful because proteins expressed by Eg95-1 and other members of the cluster are host protective while the protective efficacy of Eg95-5 and Eg95-6 is yet to be assessed (Lightowlers *et al.* 2003).

Due to nucleotide substitution, the antigenicity and efficacy of the EG95 vaccine may vary in G6/G7 genotypes.

Table 9. Sliding window analysis of a selection of the *Eg95* antigen coding gene.

Window	Midpoint	D	Sign
1–101	50	− 1.8923	*
26–126	75	− 2.2661	**
51–151	101	− 2.4132	**
76–176	126	− 2.5673	***
102–201	151	− 2.5735	***
127–226	176	− 2.5254	***
152–251	201	− 2.5109	***
177–276	226	− 2.5683	***
202–301	251	− 2.5841	***
227–326	276	− 2.6160	***
252–351	301	− 2.6072	***
277–376	326	− 2.5632	***
302–401	351	− 2.6211	***
327–426	376	− 2.5773	***
352–451	401	− 2.4859	***
377–471	424	2.5027	***

*$P < 0.1$; **$P < 0.01$; and ***$P < 0.001$.
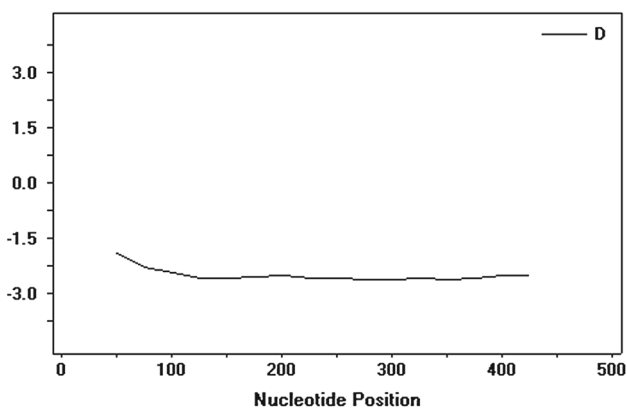
Figure 4. Sliding window analysis of selection pressure of the *Eg95* antigen coding gene of *E. granulosus*. D = Tajima's *D* statistic; *X*-axis depicts nucleotide position and *Y*-axis depicts Tajima's *D* value.

The identification of the B-cell epitope has three different implications, namely development of antibody therapeutics (Regenmortel 2006), peptide-based vaccines (Dudek *et al.* 2010) and immunodiagnostics (Leinikki *et al.* 1993). It has been clearly described in the past that over 70% of the discontinuous B-cell epitopes are composed of 1–6 amino acids. In the present investigation, short stretches of discontinuous B-cell epitopes have been predicted (table 6). The correlations between B-cell epitope localization to its physico-chemical properties (table 6) are important observations for the development of the EG95 vaccine, which has been perceived earlier as well (Pellequer *et al.* 1991). For B-cell epitope formation, hydrophilicity and antigenic propensity are primary factors although other inter-related factors like flexibility cannot be ignored

(Li *et al.* 2013). Therefore, the difference in the predicted B-cell epitope among larval stages of animal and human isolates and canine isolate (table 6) is a very important observation for designing the EG95 vaccine when the vaccine has to be developed from two different stages of the parasite. This information seems to be the pioneer observation on the development of the EG95 vaccine and has not been reported earlier.

The identification of the CTL epitope is crucial to understanding the roles of T-cell activation and designing of synthetic vaccines (Brunak and Buus 2000). In the present study, the predicted CTL epitopes of the *Eg95*-coding gene have remained constant in terms of distribution in all the isolates (table 7).

The T cell is activated after the formation of the trimolecular complex between its antigenic binding receptor, an MHC molecule and an antigenic peptide. Therefore, antigens recognized by the T cell have two distinct interaction sites. One is the epitope which interacts with the T-cell receptor and other is the agretope. The agretope interacts with the MHC molecule. Peptides that bind to class II MHC molecules contain a sequence of 7–10 amino acids. This sequence provides a major contact point with the MHC-II molecule (Goldsby *et al.* 2000). In the present study, stretches of the agretope varied from 8–10 amino acid residues (table 8). Stretches of the agretope contained aromatic or hydrophobic residues at the amino terminus and three additional hydrophobic residues in the middle (Goldsby *et al.* 2000). In the present study, the agretope identified on the *Eg95*-coding gene started with hydrophobic amino acids (either with phenylalanine or isoleucine). The amino acids with the same physico-chemical properties (hydrophobic) were present in the middle of the agretope. The agretope with MHC-II binding propensity (more than 50%) indicated a potential interaction site for the formation of the trimolecular complex.

Two antigenic properties contribute to antigen processing, presentation and recognition. The properties are amphipathicity and α-helicity. Most helper T-cell antigenic sites are amphipathic and have α-helices. Under certain circumstances, T-cell antigenic sites exhibit a β-sheet (Spouge *et al.* 1987). The conformational propensity of the *Eg95*-coding gene of 156 amino acid residues revealed α- turns and β-turns and α-amphipathic regions up to 129, 138–156 and 151–155 residues, respectively, in both animal and human isolates. This result indicated the potential T-cell antigenic site on the *Eg95*-coding gene. Positive values of the antigenic index were seen up to 137 residues. This finding indicated the presence of the immunodominant helper T-lymphocyte antigenic site.

Tajima's *D* test value (− 2.404165) has indicated negative/purifying selection pressure that varied within stretches of the EG95 antigen. This result indicated that there was an excess of low-frequency polymorphisms within 1–101 nucleotide bases compared to other stretches of nucleotide sequences (table 9). Earlier literature suggests

that there are two types of natural selection pressures in universal biological evolution: positive (Darwinian) and negative/purifying. The latter selection pressure is responsible for detrimental pseudogenization (Zhang 2008). Therefore, negative selection pressure of the EG95 antigen depicts that the biomolecule can be expressed and used for the development of vaccines without pseudogenization.

In summary, this study dealt with the genetic variability of the *EG95* protein-coding gene in 24 animal and human isolates of *E. granulosus*. All isolates belonged to one cluster and were closely similar to four isoforms of the EG95 antigen. Epitope mapping analysis revealed that physico-chemical properties of the amino acids present on the B-cell linear epitope varied between human and animal isolates as well as between larval and adult stages of the parasites. The agretope present on the *Eg95*-coding gene showed the presence of hydrophilic amino acids and conformational propensity deduced on the basis of the α-helix and β-sheet indicated the potential T-cell antigenic receptor site which helps in the presentation, processing and recognition of antigens.

References

Alvarez-Rojas C. A., Gauci C. G., Nolan M. J., Harandi M. F. and Lightowlers M. W. 2012 Characterization of the *Eg95* gene family in the G6 genotype of *Echinococcus granulosus*. *Mol. Biochem. Parasitol.* **183**, 115–121.

Arbabi M., Pirestani M., Delavari M., Hooshyar H., Abdoli A. and Sarvi S. 2017 Molecular and morphological characterizations of *Echinococcus granulosus* from human and animal isolates in Kashan, Markazi Province, Iran. *Iran J. Parasitol.* **12**, 177–187.

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

Barnes T. S., Hinds L. A., Jenkins D. J., Coleman G. T., Colebrook A. L., Kyngdon C. T. *et al.* 2009 Efficacy of the EG95 hydatid vaccine in a macropodid host, the tammar wallaby. *Parasitology* **136**, 461–468.

Bhasin M. and Raghava G. P. 2004 Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* **22**, 3195–3204.

Boubaker G., Gottstein B., Hemphill A., Babba H. and Spiliotis M. 2014 *Echinococcus* P29 antigen: molecular characterization and implication on post-surgery follow-up of CE patients infected with different species of the *Echinococcus granulosus* complex. *PLoS One* **9**, e98357.

Brunak S. and Buus S. 2000 Identifying cytotoxic T cell epitopes from genomic and proteomic informat ion: 'The human MHC project'. *Rev. Immunogenet.* **2**, 477.

Chow C., Gauci C. G., Cowman A. F. and Lightowlers M. W. 2001 A gene family expressing a host-protective antigen of *Echinococcus granulosus*. *Mol. Biochem. Parasitol.* **118**, 83–88.

Craig P. S., McManus D. P., Lightowlers M. W., Chabalgoity J. A., Garcia H. H., Gavidia C. M. *et al.* 2007 Prevention and control of cystic echinococcosis. *Lancet Infect. Dis.* **7**, 385–394.

Dudek N. L., Perlmutter P., Aguilar M. I., Croft N. P. and Purcell A. W. 2010 Epitope discovery and their use in peptide based vaccines. *Curr. Pharm. Des.* **16**, 3149–3157.

Eckert J. and Thompson R. C. A. 1997 Intraspecific variation of *Echinococcus granulosus* and related species with emphasis on their infectivity to humans. *Acta Trop.* **64**, 19–34.

Gauci C., Merli M., Muller V., Chow C., Yagi K., Mackenstedt U. *et al.* 2002 Molecular cloning of a vaccine antigen against infection with the larval stage of *Echinococcus multilocularis*. *Infect. Immun.* **70**, 3969–3972.

Gauci C. G., Alvarez-Rojas C. A., Chow C. and Lightowlers M. W. 2018 Limitations of the *Echinococcus granulosus* genome sequence assemblies for analysis of the gene family encoding the EG95 vaccine antigen. *Parasitology* **145**, 807–813.

Goldsby R. A., Kindt T. J. and Osborne B. A. 2000 *Kuby immunology*, 4th edition, pp. 63–81, 173–198. W. H. Freeman, New York.

Haag K. L., Gottstein B. and Ayala F. J. 2009 The EG95 antigen of *Echinococcus* spp. contains positively selected amino acids, which may influence host specificity and vaccine efficiency. *PLoS One* **4**, e5362.

Hemphill A. and Kern P. 2008 Special issue: experimental studies in echinococcosis. *Exp. Parasitol.* **119**, 437–438.

Huang F., Dang Z., Suzuki Y., Horiuchi T., Yagi K., Kouguchi H. *et al.* 2016 Analysis on gene expression profile in oncospheres and early stage metacestodes from *Echinococcus multilocularis*. *PLoS Negl. Trop. Dis.* **10**, e0004634.

Jenkins D. J. and Macpherson C. N. 2003 Transmission ecology of *Echinococcus* in wild-life in Australia and Africa. *Parasitology* **127**, S63–S72.

Jones D. T. 1999 Protein secondary structure prediction based on position specific scoring matrices. *J. Mol. Biol.* **292**, 195–202.

Kamenetzky L., Muzulin P. M., Guttierrez A. M., Angel S. O., Zaha A., Guarnera E. A. *et al.* 2005 High polymorphism in genes encoding antigen B from human infections strains of *Echinococcus granulosus*. *Parasitology* **131**, 805–815.

Kennedy D. A. and Read A. F. 2017 Why does drug resistance readily evolve but vaccine resistance does not? *Proc. Biol. Sci.* **284**, 20162562.

Larrieu E., Mujica G., Gauci C. G., Vizcaychipi K., Seleiman M., Herrero E. *et al.* 2015 Pilot field trial of the EG95 vaccine against ovine cystic echinococcosis in Rio Negro, Argentina: second study of impact. *PLoS Negl. Trop. Dis.* **9**, e0004134.

Leinikki P., Lehtinen M., Hyoty H., Parkkonen P., Kantanen M. L. and Hakulinen J. 1993 Synthetic peptides as diagnostic tools in virology. *Adv. Virus Res.* **42**, 149–186.

Li Y., Liu X., Zhu Y., Zhou X., Cao C., Hu X. *et al.* 2013 Bioinformatic prediction of epitopes in the Emy162 antigen of *Echinococcus multilocularis*. *Exp. Ther. Med.* **6**, 335–340.

Lightowlers M. W. and Heath D. D. 2004 Immunity and vaccine control of *Echinnococuss granulosus* infection in animal intermediate hosts. *Parassitologia* **46**, 27–31.

Lightowlers M. W., Jensen O. and Fernandez E. 1999 Vaccination trails in Australia and Argentina confirm the effectiveness of the EG95 hydatid vaccine in sheep. *Int. J. Parasitol.* **29**, 531–534.

Lightowlers M. W., Lawrence S. B., Gauci C. G., Young J., Ralston M. J., Maas D. *et al.* 1996 Vaccination against hydatidosis using a defined recombinant antigen. *Parasite Immunol.* **18**, 457–462.

Lightowlers M. W., Gauci C. G., Chow C., Drew D. R., Gauci S. M., Jenkins D. J. *et al.* 2003 Molecular and genetic characterization of the host-protective oncospheres antigens of taeniid cestode parasites. *Int. J. Parasitol.* **33**, 1207–1217.

Nono J. K., Pletinckx K., Lutz M. B. and Brehm K. 2012 Excretory/secretory-products of *Echinococcus multilocularis* Larvae induce apoptosis and tolerogenic properties in dendritic cells *in vitro*. *PLoS Negl Trop Dis.* **6**, e1516.

Nunnari G., Pinzone M. R., Gruttadauria S., Celesia B. M., Madeddu G., Malaguarnera G. *et al.* 2012 Hepatic echinococcosis: clinical and therapeutic aspects. *World J. Gastroenterol.* **18**, 1448–1458.

Pan D., Bera A. K., Bandyopadhyay S., Das S. K., Bandyopadhyay S., Bhattacharyya S. *et al.* 2009 Relative expression of antigen B coding gene of bubaline isolates of *Echinococcus granulosus* in fertile and sterile cysts. *J. Helminthol.* **28**, 1–4.

Pan D., Bera A. K., Das S. K., Bandyopadhyay S., Manna B. and Bhattacharya D. 2010a Polymorphism and natural selection of antigen B1 of *Echinococcus granulosus* isolated from different host assemblages in India. *Mol. Biol. Rep.* **37**, 1477–1482.

Pan D., Bera A. K, De S., Bandyopadhyay S., Das S. K., Manna B. *et al.* 2010b Relative expression of the 14-3-3 gene in different morphotypes of cysts of *Echinococcus granulosus* isolated from the India buffalo. *J. Helminthol.* **15**, 1–4.

Pellequer J. L., Westhof E. and Van Regenmortel M. H. 1991 Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol.* **203**, 176–201.

Regenmortel M. H. V. V. 2006 Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines. *J. Mol. Recognit.* **19**, 183–187.

Rosenzvit M. C., Camica F., Kamenentzky L., Muzulin P. M. and Gutierrez A. M 2006 Identification and intra-specific variability analysis of secreted membrane-bound proteins from *Echinococcus granulosus*. *Parasitol Int.* **55**, 63–67.

Saha S. and Raghava G. P. 2006 Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**, 40–48.

Singh H. and Raghava G. P. 2001 Propred: prediction of HLA-DR binding sites. *Bioinformatics* **7**, 1236–1237.

Spotin A., Mahami-Oskouei M., Harandi M. F., Baratchian M., Bordbar A., Ahmadpour E. *et al.* 2017 Genetic variability of *Echinococcus granulosus* complex in various geographical populations of Iran inferred by mitochondrial DNA sequences. *Acta Trop.* **165**, 10–16.

Spouge J. L., Guy H. R., Cornette J. L., Margalit H., Cease K., Berzofsky J. A. *et al.* 1987 Strong conformational propensities enhance T cell antigenicity. *J. Immunol.* **138**, 204–212.

Tamura K., Dudley J., Nei M. and Kumar S. 2007 MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

Thompson R. C. A., Lymbery A. J. and Constantine C. C. 1995 Variation in *Echinococcus*: towards a taxonomic revision of the genus. *Adv. Parasitol.* **35**, 145–175.

Thompson R. C. A. and McManus D. P. 2001 Aetiology: parasites and life-cycles. In *WHO/OIE manual on echinococcosis in humans and animals: A public health problem of global concern* (ed. J. Eckert, M. A. Gemmel, F. X. Meslin, Z. S. Pawlowski), pp. 1–9. World Organization for Animal Health, Paris.

Torgerson P. R. and Heath D. D. 2003 Transmission dynamics and control options for *Echinococcus granulosus*. *Parasitology* **127**, 43–58.

WHO (2018) Echinococcosis. Who.int./news-rooms (accessed on 14-01-2019).

WHO-Informal Working Group on Echinococcosis 2003 *PAIR: Puncture, aspiration, injection, Re-aspiration. An option for the treatment of cystic echinococcosis*. WHO, Geneva.

Woollard D. J., Heath D. D. and Lightowlers M. W. 2000 Assessment of protective immune responses against hydatid disease in sheep by immunization with synthetic peptide antigens. *Parasitology* **121**, 145–153.

Zhang J. 2008 Positive selection, not negative selection, in the pseudogenization of *rcsA* in *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **105**, E69; author reply E70.

Zhang W., Lij Y. H., Zhang Z., Turson G., Aili H., Wang J. *et al.* 2003 *Echinococcus granulosus* from Xinjiang, PR China cDNAs encoding the EG95 vaccine antigen are expressed in different life cycle stages and are conserved in the oncosphere. *Am. J. Trop. Med. Hyg.* **68**, 40–43.

Corresponding editor: H. A. RANGANATH