Apparent dominance of the G1-G3 genetic cluster of *Echinococcus granulosus* strains in the central inland region of Portugal.

Sílvia Beato[1,3], Ricardo Parreira[2], Manuela Calado[1] and Maria Amélia A. Grácio[1]

[1] Unidade de Helmintologia e Malacologia Médicas (UHMM)/Unidade de Parasitologia e Microbiologia Médicas (UPMM), Instituto de Higiene e Medicina Tropical (IHMT)/Universidade Nova de Lisboa (UNL), Rua da Junqueira 100, 1349-008 Lisboa, Portugal.

[2] Unidade de Virologia/Unidade de Parasitologia e Microbiologia Médicas (UPMM), Instituto de Higiene e Medicina Tropical (IHMT)/Universidade Nova de Lisboa (UNL), Rua da Junqueira 100, 1349-008 Lisboa, Portugal.

[3] Escola Superior de Saúde Dr. Lopes Dias, Instituto Politécnico de Castelo Branco, Avenida empresário Campos Talagueira, 6000-767, Castelo Branco, Portugal

Corresponding author: Sílvia Beato

Email: silvia.beato@ihmt.unl.pt

Phone: +351 21 365 26 00, ext. 507; Fax: +351 21 36321 05

Abstract

Infection by the larval stage of the cestode *Echinococcus granulosus* causes a disease known as cystic echinococcosis or hydatidosis, which is one of the most widespread zoonotic infections of veterinary and medical importance. Numerous studies have shown that *E. granulosus* exists as a complex of strains differing in a wide variety of criteria. Ten distinct genotypes (G1-G10) have been identified with potential impact on the pathology, epidemiology and the effect of the measures implemented for the control of hydatidosis. Our main objective was to carry out a preliminary analysis of the genotypes of *E. granulosus* circulating in the central inland region of Portugal.

Parasite samples (hydatid cysts, n=27) were isolated from the liver and lung of sheep and cattle. The DNA extracted from protoscoleces isolated from the fertile cysts served a template for the PCR amplification of part of the mitochondrial cytochrome c oxidase subunit 1 (*cox1*), ATP synthase F0 subunit 6 (*atp6*) as well as the large (*rrnL*/16S) and small (*rrnS*/12S) ribosomal RNA genes. Similarity searches with homologous sequences in the databanks indicated very high similarity with references assigned to the G1, G3 and/or G1-G3 complex of *Echinococcus* strains. Phylogenetic analysis (Bayesian approach) supported these observations, and confirmed the assignment of all the analyzed sequences to the G1-G3 genetic cluster.

Keywords: *Echinococcus granulosus*; hydatic cyst; G1-G3 genotypes; Portugal; mitochondrial DNA.

Cystic echinococcosis (also known as hydatidosis or hydatid disease) is one of the most important parasitic infections of livestock, and a considerable cause of morbidity and mortality in the world. This long known disease remains, still today, one of the most important helminthic zoonoses and is regarded as a significant worldwide public health problem [1]. Its etiological agent is a parasite known as *Echinococcus granulosus*. In its natural cycle this cestode has dogs and other canids as definitive hosts, whereas its larval stage (the metacestode) can be found in a number of ungulates including sheep, goats, horses and pigs. Accidentally, it can also be transmitted to a series of other mammals such as rodents, marsupials, non-human primates and humans [2]. Transmission to humans frequently results from close contacts with infected dogs carrying the parasite's eggs on their fur or, indirectly, as a result of ingestion of contaminated water or food [3].

The taxonomy and phylogeny of the genus *Echinococcus* has remained a controversial issue for several years [1]. A number of *E. granulosus* strains, designated G1 to G10 have been recognized [3, 4], all of which appear to be adapted to particular life cycle patterns and host assemblages [5]. A high degree of genetic diversity between E. *granulosus* strains is one of this parasite's features. In recent years a number of molecular approaches have allowed a more thorough genetic characterization of the different *E. granulosus* strain types so far identified, and supported the elevation of two of them, formerly known as the G4 and G5 strains, to the species status (*E. equinus* and *E. ortleppi*, respectively) [4].

Recent epidemiological data regarding the frequency, geographic distribution, and host range of the *E. granulosus* genetic variants in Europe is lacking. Apart from its impact on the development of control strategies, this information also provides insights on the

putative differential pathogenicity and growth characteristics of the parasite's genetic variants in humans, or their potential differences in response to therapeutics. All these reasons have prompted us to conduct this survey of *Echinococcus* genetic variants circulating in the central inland region of Portugal. Human and animal cystic echinococcosis cases have been previously reported in this region, previously defined as hyper endemic for *E. granulosus* infection [6] and an important public health problem.

A total of 58 hydatic cysts were collected from the lung (n=26) or liver (n=32) of sheep and cattle in a slaughterhouse servicing 5 different localities in central Portugal (Fig. 1). Thirty-one of these cysts, classified as infertile, calcified or contaminated (bacteria), were discarded. The remainder 27 (26 from sheep, 1 from cattle) fertile cysts were further processed. The genomic DNA from each fertile cyst was extracted from protoscoleces preserved in 70% ethanol using the High Pure PCR Template Preparation kit (Roche, Mannhein, Germany), as indicated by the supplier.

Amplification, by PCR, of part of the mitochondrial cytochrome c oxidase subunit 1 gene (*cox1*) was carried out using the previously described JB3 and JB4.5 primers and reaction conditions [7]. A multiple sequence alignment of complete mitochondrial DNA sequences from a total of 17 different *Echinococcus* strains (listed in Fig. 3), and assigned to 8 different species, was constructed with MAFFT vs. 6 [8] using sequence data obtained from the public databases (GenBank/EMBL/DDBJ). It served as a starting point for the design of pairs of oligonucleotides allowing the amplification of parts of the ATP synthase F0 subunit 6 (*atp6*) as well as the large (*rrnL*/16S) and small (*rrnS*/12S) ribosomal RNA genes. The primers used were as follows: *atp6* (ATP6F: 5'-AAACTGTRGGGTTCATGTCYC-3' and ATP6R: 5'-CACAACATAAAHGGAAAYAAACCAAAC-3'), *rrnS* (12SrF: 5'-GGTTTATTTGCCTTTTGCATCATGC-3' and 12SrR: 5'-

CCTAAGTCAACATCGAGGTGGCAAAC-3', and *rrnL* (16SrF: 5'-AGCCAGGTCGGTTCTTATCTATTG-3' and 16SrR: 5'-CGAGGGTGACGGGCGGTGTGTAC-3'). For these 3 genes, PCR conditions included an initial denaturation step at 95ºC for 5 min, followed by 35 cycles of 95ºC for 45 sec, 61ºC for 1 min and 72ºC for 45 sec, followed by a final extension for 7 min. In all cases 0.6μM was the final concentration of primers used per reaction carried out with the Illustra™ puReTaq Ready-to-go PCR beads' system (GE Healthcare, Buckinghamshire, UK). The obtained PCR amplicons were purified from the reaction mixtures using the QIAquick PCR Purification kit (Qiagen, Valencia, USA), and directly sequenced. Nucleotide sequence similarity searches were carried out using BLASTn (available at http://www.ncbi.nlm.nih.gov).

Phylogenetic inference was based on a Bayesian Markov chain Monte Carlo approach, run for $6x10^6$ generations under a GTR model, using MrBayes v3.0b4 [9], with nucleotide rate heterogeneity estimated using a γ distribution for the variable sites. The nucleotide sequences reported in this study were deposited at the EMBL/GenBank/DDBJ sequence databases under accession numbers FN646353-FN646362, FN646364-FN646378, FN666904 and FN666905 (*coxI*), FR668537-FR668555 (*atp6*), FR666874-FR666903 (*rrnL*) and FR667922-FN667949 (*rrnS*).

Since it evolves more rapidly than nuclear DNA, mitochondrial genes have, a priori, the potential to resolve phylogenetic and taxonomic problems regarding the analysis of closely related taxa. Furthermore, the large ensemble of data already available in the databases have lead us to initiate this study with the analysis of *cox1*, one of the most extensively studied mitochondrial genes. A specific DNA segment amplified, and sequenced from the 27 fertile cysts, revealed almost total nucleotide sequence conservation, as polymorphisms were only found at two of the positions analyzed.

Similarity searches with sequences deposited in the public databases (n>380) revealed over 99% identity (10 best matches) with partial *cox1* sequences. Most of these were referred to as having been amplified from either sheep or cattle, and only one was referred to as originating from a water buffalo (DQ104331). The overwhelming majority of them was classified as G1 (the common sheep strain), or included in a G1-G3 complex. Similar results were obtained for the *atp6*, *rrnS* and *rrnL* sequences (data not shown).

The relationships between the sequences here described, and several other references deposited in the databases, was also carried out through phylogenetic reconstruction using a Bayesian approach. In a preliminary analysis involving only the *cox1* sequences (due to their wide representation in the sequence databases), and contrary to what had been previously reported [4], *E. vogeli* and *E. equinus* (not *E. oligarthrus*) occupied basal positions in the obtained phylogenetic tree (Fig. 2). However, these differences may be explained by the non-overlapping data sets used (considerably shorter in the analysis presented here). The G6 to G10 strains formed a monophyletic cluster supported by maximum posterior probability. The tight segregation of these strains in phylogenetic trees has previously prompted the assignment of all these variants to a single species, designated *E. canadensis* [10]. Nevertheless, the analysis here presented revealed a clear separation between the G10 strain and a very tight G6-G7-G8 cluster of reference sequences, which warrants further investigation.

The study of *Echinococcus* strains obtained from Portuguese animals was further extended with the analysis of partial *atp6*, *rrnL* and *rrnS* sequences. While the amplification of *atp6* was only possible for a total number of 19 samples due to exhaustion of the available material and/or its degradation, we were able to amplify ribosomal DNA segments from the 27 samples from which *cox1* sequences had been

previously obtained. For that reason, the assessment of phylogenetic relationships between *Echinococcus* strains, graphically depicted in Fig. 3, was based on the construction of Bayesian trees involving the analysis of separate *atp6* (Fig. 3A, 575 aligned nucleotides) and *cox1*/*rrnS*/*rrnL* (Fig. 3B, 1703 aligned nucleotides) concatenated sequence datasets. Both phylogenetic trees disclosed a congruent association between *E. ortleppi* (G5) and a cluster including *E. canadensis* (G6-G8), as well as the inclusion of all the Portuguese sequences analyzed in a statistically consistent cluster with low genetic variability, and containing the two *E. granulosus* references used. The larger size of the *cox1*/*rrnS*/*rrnL* concatenated dataset also allowed a better segregation of the major clusters of sequences, while in the *atp6* tree a polytomy excluded only the *E. shiquicus* and *E. oligarthrus* references. Curiously, in the *cox1*/*rrnS*/*rrnL* tree (Fig. 3B) two sequences (indicated by *), clustering together with statistical support, segregate prematurely from all the others included in the *E. granulosus* group. Although the topology of the *atp6* tree is not exactly congruent, these two sequences still cluster within the *E. granulosus* radiation with significant statistical support (Fig. 3A). Curiously, the *cox1* fragment of both sequences had high similarity (BLAST analysis) with the water buffalo (G3) *Echinococcus* strain DQ104331 mentioned above.

The G1 variant, also known as the common sheep strain, is the most important *E. granulosus* strain in Europe. In the Mediterranean region in particular, where sheep farming is extensive, its presence coincides with the highest levels of human cystic echinococcosis [3, 11]. Nevertheless, the G3 strain, which is considered a poorly characterized genetic variant that infects buffaloes and cattle, has already been described in Greece and Italy [13, 14]. The available genetic data has been disclosing a high degree of similarity between these two strains (G1/G3), which can also be

extended to G2, or the Tasmanian sheep strain. This has lead several authors to suggest the inclusion of the G1, G2 and G3 strains into a single species designated *E. granulosus sensu stricto* [4, 11, 12]. The analysis here presented also supports this suggestion. Indeed, all the *cox1* sequences analyzed clustered with a posterior probability of 1.00 in a cluster, which contained all the G1 to G3 references used in this study (n=88). The short sequence analyzed most certainly impacts the low genetic variability observed and consequent uncertain resolution of this cluster. However, even the analysis of a larger sequence resulting from the concatenation of several mitochondrial genes has not unambiguously improved the resolution of the G1 to G3 strains [4] which is clearly restricted by the paucity of sequence data for mitochondrial markers from G2-G3 *Echinococcus* strains.

Although the analysis here presented involved a small number of *Echinococcus* samples, it is the first genetic characterization of the parasite carried out in Portugal. The assessment of the genetic diversity of the *Echinococcus* strains circulating in the central inland part of the country disclosed an apparent dominance of the G1-G2-G3 cluster (well defined in the *cox1* tree, Fig. 2). An extended study of the parasite's genetic makeup, involving the examination larger set of *Echinococcus* strains and additional mitochondrial (*nad1* and *cytB*) and nuclear markers (*cal*¸ *mdh, actII* is currently being undertaken.

Acknowledgements

We would thank the animal health officials Dr. Barreira Junior, Dra. Ana Maria Menezes, Dra. Maria Otília Reis and Dr. Henrique Domingues, for their cooperation in

this study. This study was supported by Fundação para a Ciência e a Tecnologia (Portugal) through UPMM funding.

References

[1] Thompson RC. The taxonomy, phylogeny and transmission of *Echinococcus*. Exp Parasitol 2008; 119:439-446.

[2] Eckert J, Deplazes P. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. Clin Microbiol Rev 2004;17:107-135.

[3] Romig T. Epidemiology of echinococcosis. Langenbecks Arch Surg 2003;388:209-217.

[4] Nakao M, McManus DP, Schantz PM, Craig PS, Ito A. A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. Parasitology 2007;134:713-722.

[5] McManus D, Thompson RCA. Molecular epidemiology of cystic echinococcosis, Parasitology 2003;127:S37-51.

[6] David De Morais JA A Hidatidologia Em Portugal. Ed. Fundação Calouste Gulbenkian: Lisboa, 1998. (In Portuguese).

[7] Bowles J, Blair D, McManus DP. Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. Mol Biochem Parasitol 1992;54:165-173.

[8] Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 2008;9:286-298.

[9] Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 2001;17:754-755.

[10] Lavikainen A, Lehtinen MJ, Laaksonen S, Agren E, Oksanen A, Meri S. Molecular characterization of *Echinococcus* isolates of cervid origin from Finland and Sweden. Parasitology 2006;133:565-70.

[11] Romig T, Dinkel A, Mackenstedt U. The present situation of echinococcosis in Europe. Parasitol Int 2006;55:187-191.

[12] Thompson RCA, Lymbery AJ, Constantine CC. Variation in *Echinococcus*: towards a taxonomic revision of the genus. Adv Parasitol 1995;35:145-176.

[13] Capuano F, Rinaldi L, Maurelli MP, Perugini AG, Veneziano V, Garippa G,Genchi C, Musella V, Cringoli G. Cystic echinococcosis in water buffaloes: epidemiological survey and molecular evidence of ovine (G1) and buffalo (G3) strains. Vet Parasitol 2006;137:262-268.

[14] Varcasia A, Canu S, Kogkos A, Pipia AP, Scala A, Garippa G, Seimenis A. Molecular characterization of *Echinococcus granulosus* in sheep and goats of Peloponnesus, Greece. Parasitol Res 2007;101:1135-1139.

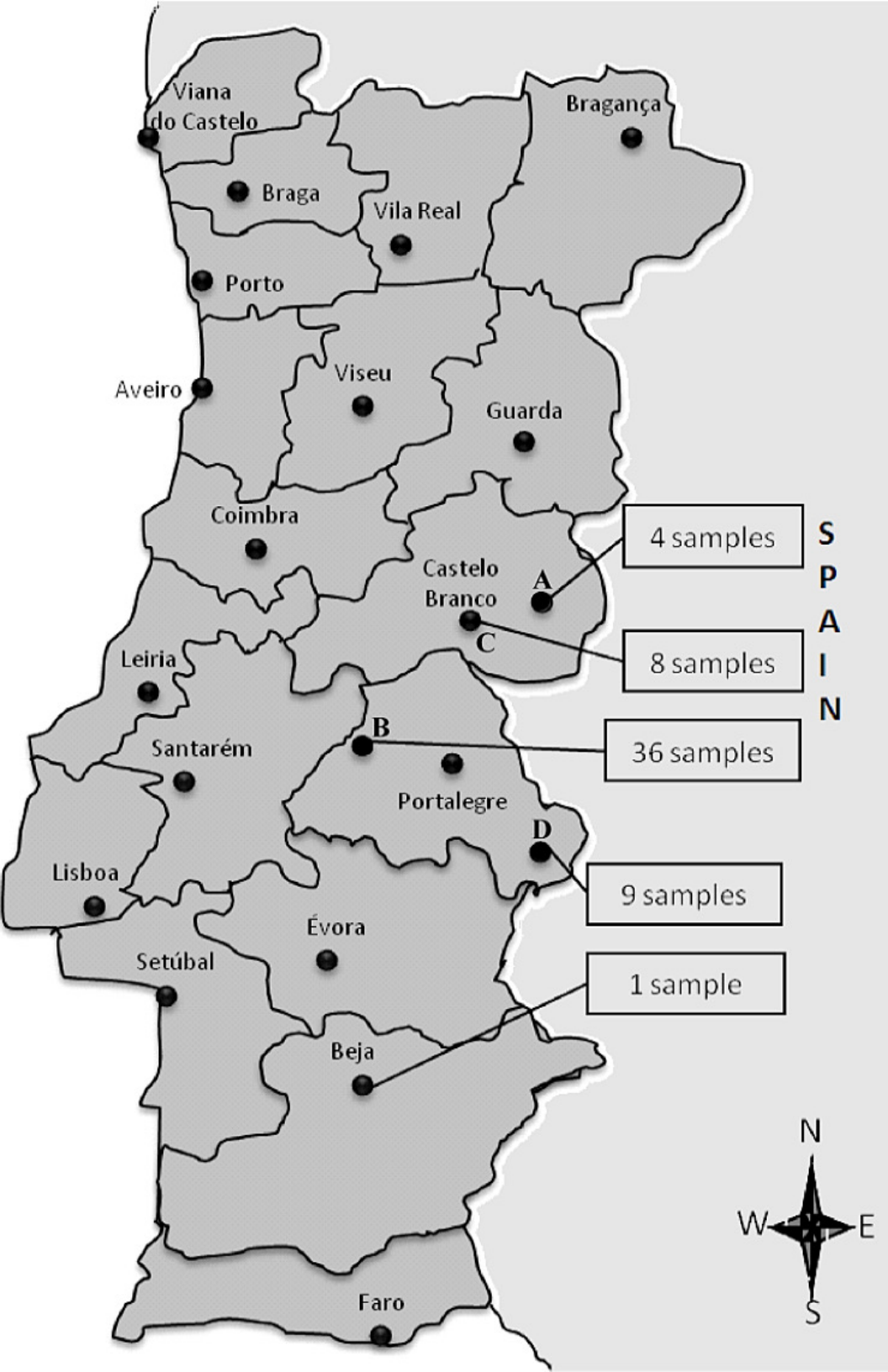
Figure Captions

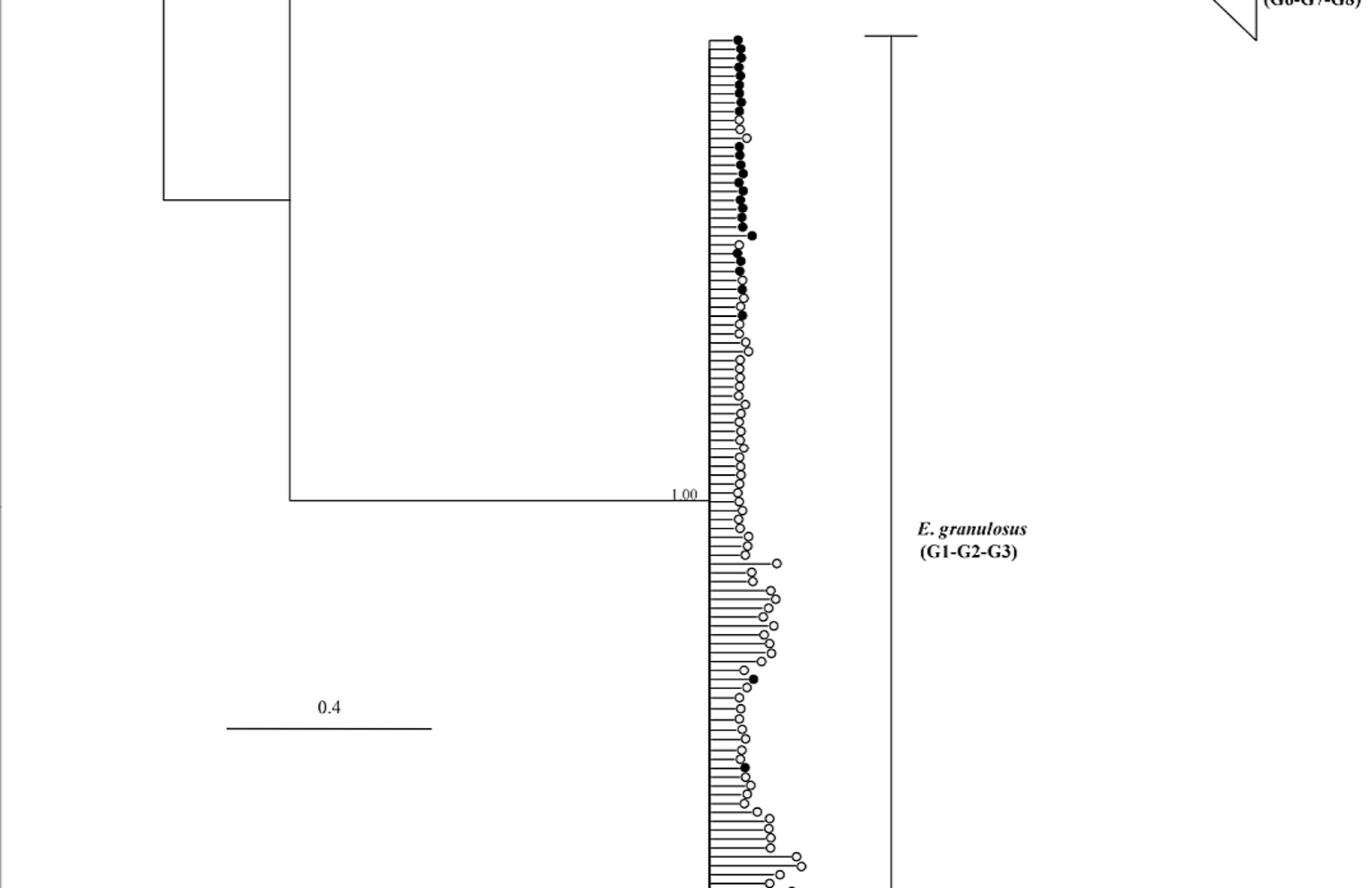
Figure 1 – Map of Portugal showing the geographic origin and number (indicated by the rectangles) of *the Echinococcus granulosus* strains analyzed in this study. The localities indicated by the letters A-D (A-Idanha-a-Nova, B-Ponte de Sor, C-Castelo Branco, D-Elvas) represent the origin of the infected animals from which *Echinococcus* mitochondrial sequences were obtained.

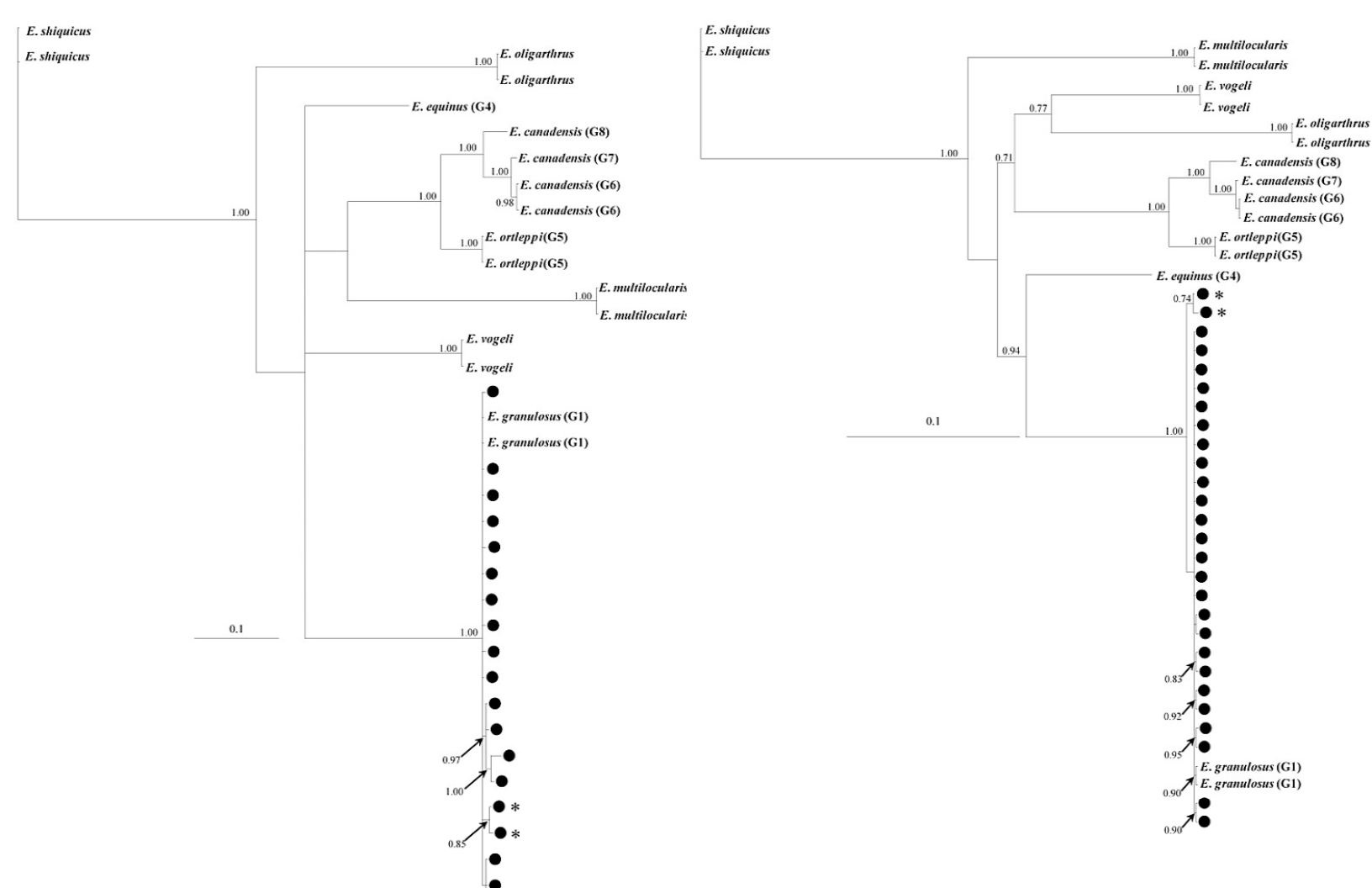
Figure 2 - Phylogenetic tree (Bayesian analysis) generated from the analysis of partial *Echinococcus cox1* sequences. Branch lengths are proportional to the number of nucleotide changes per site. At selected branch nodes, the numbers indicate values of Bayesian probability. The sequences obtained from Portuguese strains of *E. granulosus* are indicated by the black circles. The tree was rooted using a *Taenia taeniaeformis* (AB221484) as the outgroup sequence. The different *Echinoccoccus* references (species and accession numbers) used in this analysis, indicated by the grey circles, were as follows: *E. multilocularis* - M84669, M84668; *E. equinus* - AJ508035, AJ508036, EF143834, M84664; *E. oligarthrus* - M84671; *E. ortleppi* - M84665; *E. granulosus* G10 - AF525457; *E granulosus* G6-G7-G8 cluster - AB271910, AB271911, AB271912, AB271236, AB274020, DQ062858, DQ341580, DQ341582, DQ341584, DQ856468, EU151431, M84666, M84667; *E. granulosus* G1-G2-G3 cluster - AB033407, AB458672, AB458673, AB458674, AB458675, AB470527, AJ508013, AJ508019, AY278068, AY679144-AY679146, AY686559, AY850565, DQ062857, DQ109036, DQ131582, DQ269943, DQ269947, DQ333185, DQ341564, DQ341566, DQ341568, DQ341579, DQ356881, DQ356882, DQ356883, DQ856466, DQ856467, EF367241-EF367266, EF367269, EF367270, EF367271, EF367273-EF367276, EF367292,

EF367294, EF393619, EF545563, EF595654, EU006775, EU006776, EU006781, EU006784, EU072107, EU072108, EU072110, EU178103, EU178105, EU503084, EU929083, FJ608720, FJ608726, FJ608749, FJ608759, FJ608760, M84661, M84662, M84663, U50464, U50464.

Figure 3 - Bayesian phylogenetic tree of partial mitochondrial *Echinococcus atp6* (A) and *cox1*/*rrnS*/*rrnL* concatenated sequences (B). Branch lengths are proportional to the number of nucleotide changes per site. At selected branch nodes, the numbers indicate values of Bayesian posterior probability. In both trees a similar set of reference sequences was used, which included *E. shiquicus* (NC_009460 and AB208064), *E. multilocularis* (NC_000928 and AB018440), *E. vogeli* (NC_009462 and AB208546), *E. oligarthrus* (NC_009461 and AB208545), *E. canadensis* (NC_011121, AB208063, AB235847, and AB235848), *E. ortleppi* (NC_011122 and AB235846), *E. equinus* (AF346403) and *E. granulosus* (NC_008075 and AF297617).


Viana do Castelo
Braga
Bragança
Vila Real
Porto
Aveiro
Viseu
Guarda
Coimbra
Castelo Branco
A
C
Leiria
Santarém
B
Portalegre
D
Lisboa
Évora
Setúbal
Beja
Faro
4 samples
8 samples
36 samples
9 samples
1 sample
SPAIN
N
W
E
S


1.00
0.4
E. granulosus
(G1-G2-G3)


E. shiquicus
E. shiquicus
E. oligarthrus
E. oligarthrus
E. equinus (G4)
E. canadensis (G8)
E. canadensis (G7)
E. canadensis (G6)
E. canadensis (G6)
E. ortleppi(G5)
E. ortleppi(G5)
E. multilocularis
E. multilocularis
E. vogeli
E. vogeli
E. granulosus (G1)
E. granulosus (G1)
0.1
1.00
0.98
0.97
0.85
0.71
0.74
0.94
0.83
0.92
0.95
0.90