Infection, Genetics and Evolution 74 (2019) 103941

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Analysis of *nad2* and *nad5* enables reliable identification of genotypes G6 and G7 within the species complex *Echinococcus granulosus* sensu *lato*

Teivi Laurimäe[a], Liina Kinkar[a], Thomas Romig[b], Gérald Umhang[c], Adriano Casulli[d,e], Rihab A. Omer[f], Mitra Sharbatkhori[g], Hossein Mirhendi[h], Francisco Ponce-Gordo[i], Lorena E. Lazzarini[j], Silvia V. Soriano[j], Antonio Varcasia[k], Mohammad Rostami-Nejad[l], Vanesa Andresiuk[m], Pablo Maravilla[n], Luis Miguel González[o], Monika Dybicz[p], Jakub Gawor[q], Mindaugas Šarkūnas[r], Viliam Šnábel[s], Tetiana Kuzmina[t], Eshrat Beigom Kia[u], Urmas Saarma[a,*]

[a] *Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 51003 Tartu, Estonia*
[b] *Institute of Zoology, Parasitology Unit, University of Hohenheim, 70599 Stuttgart, Germany*
[c] *Anses, Wildlife Surveillance and Eco-epidemiology Unit, National Reference Laboratory for Echinococcus spp., Nancy Laboratory for Rabies and Wildlife, 54220 Malzéville, France*
[d] *World Health Organization Collaborating Centre for the Epidemiology, Detection and Control of Cystic and Alveolar Echinococcosis (in humans and animals), Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy*
[e] *European Union Reference Laboratory for Parasites (EURLP), Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy*
[f] *National University Research Institute, National University Sudan, Khartoum, Sudan*
[g] *Infectious Diseases Research Center, Golestan University of Medical Sciences, Gorgan, Iran*
[h] *Department of Parasitology and Mycology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran*
[i] *Department of Parasitology, Faculty of Pharmacy, Complutense University, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain*
[j] *Department of Microbiology and Parasitology, Faculty of Medical Sciences, Comahue National University, Buenos Aires, 1400, 8300, Neuquén, Argentina*
[k] *Laboratorio di Parassitologia e Malattie Parassitarie, Ospedale Didattico Veterinario Dipartimento di Medicina Veterinaria, Università degli Studi di Sassari, Via Vienna 2, 07100 Sassari, Italy*
[l] *Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[m] *Laboratorio de Zoonosis Parasitarias, FCEyN, UNMdP, Funes 3350, CP: 7600 Mar del Plata, Buenos Aires, Argentina*
[n] *Hospital General "Dr. Manuel Gea Gonzalez", Departamento de Ecologia de Agentes Patogenos, DF 14080, Mexico*
[o] *Parasitology Department, Centro Nacional de Microbiologia, Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain*
[p] *Department of General Biology and Parasitology, 5 Chałubińskiego Str., 02-004 Warsaw, Medical University of Warsaw, Poland*
[q] *W. Stefański Institute of Parasitology, Polish Academy of Science, Twarda51/55, Warsaw 00-818, Poland*
[r] *Department of Veterinary Pathobiology, Veterinary Academy, Lithuanian University of Health Sciences, Tilžes Street 18, 47181 Kaunas, Lithuania*
[s] *Institute of Parasitology, Slovak Academy of Sciences, Košice, Hlinkova 3, 040 01 Košice, Slovakia*
[t] *I.I. Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, 01030 Kyiv, Ukraine*
[u] *Department of Medical Parasitology and Mycology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*

ARTICLE INFO

Keywords:
Cystic echinococcosis
Echinococcus granulosus sensu *lato*
Genotype G6
Genotype G7
Genotype identification
Hydatid disease

ABSTRACT

The larval stages of tapeworms in the species complex *Echinococcus granulosus* sensu *lato* cause a zoonotic disease known as cystic echinococcosis (CE). Within this species complex, genotypes G6 and G7 are among the most common genotypes associated with human CE cases worldwide. However, our understanding of ecology, biology and epidemiology of G6 and G7 is still limited. An essential first step towards this goal is correct genotype identification, but distinguishing genotypes G6 and G7 has been challenging. A recent analysis based on complete mitogenome data revealed that the conventional sequencing of the *cox1* (366 bp) gene fragment mistakenly classified a subset of G7 samples as G6. On the other hand, sequencing complete mitogenomes is not practical if only genotype or haplogroup identification is needed. Therefore, a simpler and less costly method is required to distinguish genotypes G6 and G7. We compared 93 complete mitogenomes of G6 and G7 from a wide geographical range and demonstrate that a combination of *nad2* (714 bp) and *nad5* (680 bp) gene fragments would be the best option to distinguish G6 and G7. Moreover, this method allows assignment of G7 samples into haplogroups G7a and G7b. However, due to very high genetic variability of G6 and G7, we suggest to construct a phylogenetic network based on the *nad2* and *nad5* sequences in order to be absolutely sure in genotype

* Corresponding author.
E-mail address: Urmas.Saarma@ut.ee (U. Saarma).

https://doi.org/10.1016/j.meegid.2019.103941
Received 12 March 2019; Received in revised form 13 June 2019
Available online 25 June 2019

assignment. For this we provide a reference dataset of 93 concatenated *nad2* and *nad5* sequences (1394 bp in total) containing representatives of G6 and G7 (and haplogroups G7a and G7b), which can be used for the reconstruction of phylogenetic networks.

1. Introduction

Cystic echinococcosis (CE) is a severe zoonotic disease with cosmopolitan distribution that is caused by the larval stage of tapeworms belonging to the species complex *Echinococcus granulosus* sensu *lato* (*s.l.*) (Deplazes et al., 2017). Various canids (dogs, wolves, jackals, dingoes) may act as definitive hosts for the adult stage of this parasite (e.g., Moks et al., 2006; Laurimaa et al., 2015; Schurer et al., 2016; Thompson, 2017). Intermediate hosts for the larval stage (fluid filled cyst) are various wild and domestic herbivores and omnivores (e.g., sheep, goats, camels, pigs, cattle, yaks, moose, wild boars), but also humans (Romig et al., 2017). While CE is asymptomatic for the definitive host, it can cause severe health problems and can be fatal to the intermediate host, including humans, if not adequately treated (Alvarez Rojas et al., 2014; Thompson, 2017).

Initially two mitochondrial DNA (mtDNA) gene fragments *cox1* (366 bp) and *nad1* (471 bp) were used to define ten *E. granulosus s.l.* genotypes, designated as G1 to G10 (Bowles et al., 1992, 1994; Bowles and McManus, 1993; Scott et al., 1997; Lavikainen et al., 2003). Later studies have, however, revealed that genotypes G2 and G9 are not valid, since G2 is a microvariant of G3 (Kinkar et al., 2017) and G9 a microvariant of G7 (Kedra et al., 1999). Significant genetic, morphological, life cycle and host range differences between *E. granulosus s.l.* genotypes have provided grounds to consider a number of these genotypes as separate species (e.g., Lymbery, 2017; Romig et al., 2017; Thompson, 2017). Genotypes G1 and G3 are now regarded as *E. granulosus* sensu stricto *(s.s.),* G4 as *Echinococcus equinus*, and G5 as *Echinococcus ortleppi* (Thompson and McManus, 2002; Kinkar et al., 2017). The phylogeny and species status of genotypes G6, G7, G8 and G10 has, however, remained controversial (Moks et al., 2008; Thompson, 2008; Saarma et al., 2009; Knapp et al., 2011, 2015; Lymbery et al. 2015; Nakao et al., 2015). Recently, a study by Yanagida et al. (2017), which was based on two nuclear genes, suggested that gene flow may exist between the genotypic groups G6/G7 and G8/G10. However, another study based on six nuclear genes suggested that G6-G10 consists of two species, one comprising genotypes G6/G7 and the other species G8/G10 (Laurimäe et al., 2018a). While the suitable species name for the G8/G10 genotypic group is *Echinococcus canadensis*, the species name for G6/G7 is under dispute (Nakao et al., 2015; Saarma et al., unpublished).

Within *E. granulosus s.l.*, genotype G6 is considered to be the second most common genotype associated with human CE cases worldwide (Alvarez Rojas et al., 2014; Cucher et al., 2016). However, human CE infections with genotype G7 are also more frequent than previously thought (Jabbar et al., 2011; Dybicz et al., 2013; Alvarez Rojas et al., 2014). These two genotypes are prevalent or highly prevalent in the Mediterranean area, on the Tibetan Plateau, and in the southern regions of South America. Genotype G6 is more prevalent than G7 in Africa, Asia and Middle East, whereas G7 is more common in Europe (Deplazes et al., 2017). The typical intermediate hosts for G6 are goats and camels, whereas for G7 the most common is the pig (Romig et al., 2017). Differentiating between genotypes G6 and G7, but also between haplogroups G7a and G7b, of which G7b seems to be restricted to the islands of Corsica and Sardinia (Laurimäe et al., 2018b) and to the adjacent regions in the mainland of Italy (Laurimäe et al., 2019), is relevant to determine whether there are significant ecological, biological or epidemiological differences between these genetic groups (e.g., infectivity to humans, host assemblages, distribution ranges). However, it has remained unclear to what extent the differences in distribution ranges of these groups reflect differences in host affinity (especially relevant to G6 and G7), and to what extent it is the result of historical events and livestock trading. Additionally, it also remains to be studied whether the pathogenicity of these genetic groups is different in terms of human infections, which can be important for planning effective cystic echinococcosis control strategies.

The two markers used for the original description of genotypes G6 and G7 (fragments of *cox1*, 366 bp; *nad1,* 471 bp; Bowles et al., 1992; Bowles and McManus, 1993) have been widely applied to distinguish genotypes within *E. granulosus s.l.*, including G6 and G7 (e.g., Beato et al., 2013; Rodriguez-Prado et al., 2014; Zhang et al., 2014; Rostami et al., 2015). Recently a number of studies have employed the complete *cox1* gene (1608 bp), providing a more detailed molecular characterization of these genotypes (e.g., Konyaev et al., 2013; Addy et al., 2017). The data obtained through employing these markers have played an important role in distinguishing various *E. granulosus s.l.* species and genotypes, and in understanding their evolution and transmission. However, it has become evident that these markers do not allow for a consistent and reliable assignment of samples to genotypes G6 and G7. It is especially relevant for samples that do not cluster clearly neither to G6 nor G7, but are positioned between these genotypes. Even the data of complete *cox1* (1608 bp) has not allowed reliable assignment of such intermediate samples into genotypes (e.g., Nakao et al., 2013; Addy et al., 2017; Laurimäe et al., 2018b). Other genetic markers (e.g., *16S rRNA, atp6*) have occasionally also been used to distinguish genotypes (e.g., Šnábel et al., 2009; Boubaker et al., 2013, 2016; Nikmanesh et al., 2017), but their efficacy has remained contentious. In order to develop a method that is reliable and simple, a global dataset is required that is based on a large number of samples originating from different geographical regions, and for which mitogenome sequences are determined.

Recently, studies on the genetic diversity of *E. granulosus s.l.* have started to employ near-complete or complete mitogenomes for a large set of samples, covering a wide geographical distribution range of genotypes G1/G3 and G6/G7 (Kinkar et al., 2016, 2018a, 2018c; Laurimäe et al., 2016, 2018b). These studies have opened possibilities to find mtDNA sequences that are sufficient for the reliable distinction of closely related genotypes. As demonstrated in Kinkar et al. (2018b) for *E. granulosus s.s.* genotypes G1 and G3, the traditionally used *cox1* and *nad1* genes proved unreliable, and it was found that a *nad5* gene fragment (680 bp) is optimal to identify G1 and G3. A relatively large dataset has recently also been established for genotypes G6 and G7 (Laurimäe et al., 2018b) and it was discovered, based on complete mitogenomes, that genotype G7 is represented by two haplogroups G7a and G7b. Notably, it was found that based on the *cox1* (366 bp) gene fragment, some of the G7b samples would have been mistakenly classified as genotype G6, and the complete *cox1* (1608 bp) gene sequences positioned these samples closer to genotype G6 than to G7. However, the phylogenetic network and Bayesian phylogenetic analysis based on complete mitogenomes clearly showed this cluster to belong to G7 (Laurimäe et al., 2018b).

Although the studies using near-complete or complete mitogenomes have demonstrated their significant advantages, sequencing complete mitogenomes of *Echinococcus* can be challenging (Kinkar et al., 2019) and often not practical when only genotype or haplogroup assignment is required. A much simpler, but still highly reliable method is therefore required that is based on the analysis of a minimum set of genetic markers that allow confident distinction of genotypes G6 and G7, and haplogroups G7a and G7b.

The main aim of the current study was to determine mtDNA gene regions that would be optimal for reliable and easy genotype distinction

of *E. granulosus s.l.* G6 and G7 samples, as well as haplogroups G7a and G7b. We also aimed to develop an affordable assay for the amplification of such regions.

2. Material and methods

2.1. Complete mitochondrial genome sequences

In order to determine new suitable markers, which would allow for reliable assignment of samples into *E. granulosus s.l.* genotypes G6 and G7, as well as into haplogroups G7a and G7b, a total of 93 complete mitogenome sequences of G6 and G7 from GenBank were used for the current analysis (Fig. 1; accession numbers: MH300929-MH300970; MH300972-MH301022; Laurimäe et al., 2018b; see Table S1 for additional data). Of these, 26 represented genotype G6, and 67 genotype G7, whereas genotype G7 was represented by sequences belonging to haplogroups G7a (n = 55) and G7b (n = 12). The highly divergent isolate Gmon from Mongolia (Laurimäe et al., 2018b) was excluded from the current analysis since there was only one sample and its genotypic identity remained unclear (it can be a new genotypic cluster within the *E. granulosus s.L. complex*). More samples that are genetically similar to Gmon are needed to clarify its status.

The complete mitogenome sequences were 13,549–13,554 bp in length, which were aligned using Clustal W multiple sequence alignment in BioEdit v.7.2.5 (Thompson et al., 1994; Hall, 1999). The total length of the alignment was 13,556 bp (alignment available at Mendeley Data; DOI: 10.17632/d2w6xnmz9y.1). Using the invertebrate mitochondrial genetic code, protein-coding genes of all 93 sequences were translated into amino acids in Geneious v.11.0.5 (Biomatters, Auckland, New Zealand; Kearse et al., 2012). Subsequently, open reading frames (ORFs), codon positions and amino acid changes (synonymous/nonsynonymous) at polymorphic positions were determined. Defining positions were according to the reference mitogenome sequence MH300955 in GenBank (Laurimäe et al., 2018b).

2.2. Sequencing of nad2 and nad5 genes

Since we found that the *nad2/nad5* combination provides reliable means to distinguish genotypes G6 and G7, as well as haplogroups G7a and G7b, two new primers for *nad2* were designed for PCR amplification: G7for (GTGTTGTGTTGTTGATAGATTG) and G7rev (GTAAAAAT AATCACCACCCAAC). These primers yield a PCR product of 781 bp in length. The primers for the *nad5* gene region were from in Kinkar et al. (2018b), yielding a PCR product of 759 bp.

Fifteen sequences of *E. granulosus s.l.* genotypes G6 (n = 5) and G7 (n = 10), including representatives of G7a (n = 5) and G7b (n = 5), were used to test the *nad2* and *nad5* primers. PCR was carried out in separate reactions for the *nad2* and *nad5* primer pairs, both in a volume

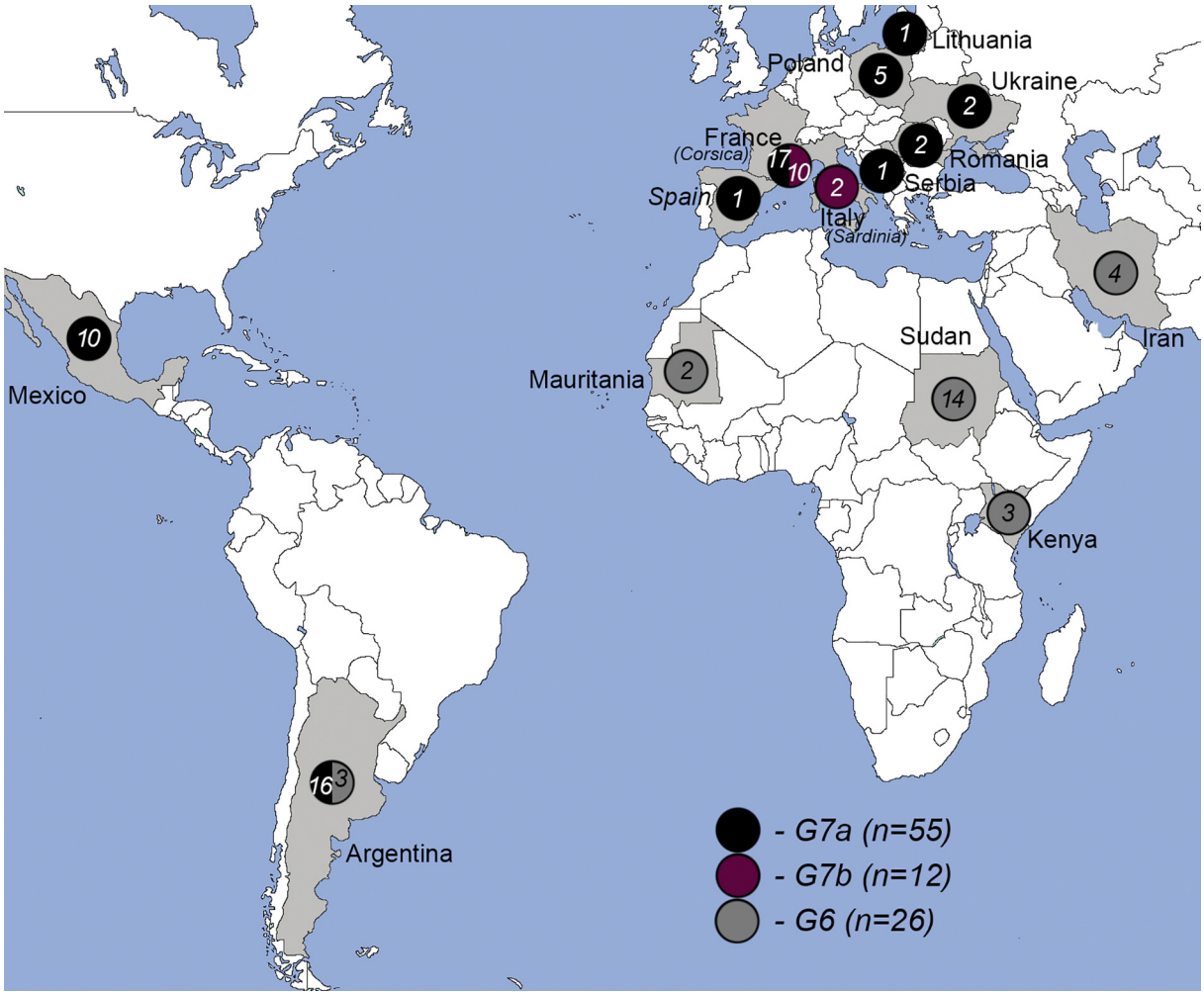

Fig. 1. Geographic locations of the 93 complete mitogenome sequences of *Echinococcus granulosus* sensu *lato* genotypes G6 and G7 (including haplogroups G7a and G7b) obtained from GenBank (MH300929-MH300970; MH300972-MH301022; Laurimäe et al., 2018b). Black dots depict haplogroup G7a (n = 55) samples, purple dots G7b (n = 12), and grey dots genotype G6 (n = 26) samples. Numbers inside the dots represent the number of samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of 20 µl, using 1 U x 5× FIREPol® Master Mix (Solis BioDyne, Tartu, Estonia), 0.25 µM of each primer, and 10 ng of purified genomic DNA. A touchdown protocol was used for PCR: initial denaturation at 95 °C for 15 min, followed by 10 cycles of 95 °C for 20 s, 55 °C for 45 s (annealing temperature progressively reduced by −0.5 °C in each cycle), and 72 °C for 1 min; followed by 25 cycles of 95 °C for 20 s, 50 °C for 45 s, 72 °C for 1 min; and finishing with a final elongation step at 72 °C for 5 min. In order to validate the size of the PCR products against a reference molecular weight standard O'GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific), 10 µl of each PCR product was examined on 1.4% agarose gel electrophoresis in 1xTAE buffer. Purification reactions were carried out for the remaining 10 µl of PCR products by adding 1 U of FastAP Thermosensitive Alkaline Phosphatase and 1 U Exonuclease I (Thermo Fisher Scientific) to the PCR products, which were then incubated at 37 °C for 30 min, and subsequently heated at 80 °C for 15 min in order to inactivate the enzymes.

Sequencing was performed at the Estonian Biocentre Core Laboratory (Tartu, Estonia). Both DNA strands were sequenced by using the same primers (one primer per reaction) as for the initial PCR reactions. BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, USA) was utilised for sequencing, following the manufacturer's protocols. Cycling parameters were 96 °C 1 min, followed by 25 cycles of 96 °C 10 s, 50 °C 15 s and 60 °C 4 min. Sequences were resolved on the ABI 3130xl sequencer (Applied Biosystems). Sequences were assessed for quality and consensus sequences were assembled using the program Codon Code Aligner v.6.0.2. Sequences were aligned using Clustal W multiple sequence alignment in the program BioEdit v.7.2.5 (Thompson et al., 1994; Hall, 1999). Subsequently, chromatograms of sequences were checked for double peaks and reading errors. The sequences were then compared to a high quality reference sequence, and observed mutations were determined as valid if a clear, high quality peak of the corresponding nucleotide was present on the chromatogram. After sequencing and trimming, the final length of the sequences for the *nad2* gene region was 714 bp, and for the *nad5* region 680 bp.

2.3. Phylogenetic networks

Phylogenetic networks were constructed separately for five datasets using Network v4.612 (Bandelt et al., 1999; http://www.fluxus-engineering.com, Fluxus Technology Ltd., 2004), with both indels and point mutations considered. The datasets were as follows: i) *nad2* (714 bp) + *nad5* (680 bp); ii) *cox1* (366 bp) sensu Bowles et al. (1992); iii) full *cox1* (1608 bp) gene; iv) *nad2* (714 bp) + *nad5* (680 bp) + *cox1* (366 bp); v) *nad2* (714 bp) + *nad5* (680 bp) + full *cox1* (1608 bp).

Table 1
Defining positions in the complete mitochondrial genomes for differentiating *Echinococcus granulosus* sensu *lato* genotypes G6 and G7, and haplogroups G7a and G7b. Alignment length is 13,556 bp. Positions are according to GenBank reference MH300955 (Laurimäe et al., 2018b). Positions marked in bold are within the new markers *nad2* and *nad5*. Abbreviations for amino acids: Ala – Alanine, Asn – Asparagine, Cys – Cysteine, Glu - Glutamic acid, Gly – Glycine, Ile – Isoleucine, Met – Methionine, Thr – Threonine, Leu – Leucine, Phe – Phenylalanine, Pro – Proline, Ser – Serine, Tyr – Tyrosine, Val – Valine.

Gene	Defining position	G6	G7a	G7b	Codon position	Amino acid change	Notes
tRNA-Arg	412	C	T	T	–	–	–
nad5	656	G	A	A	1st	Gly (G6)/Ser(G7)	–
nad5	**714**	**C**	**T**[a]	**T**	**2nd**	**Ala (G6)/Val (G7)**	[a]2 G7a samples had C (Ala)
nad5	**804**	**A**	**G**	**G**	**2nd**	**Asn (G6)/Ser (G7)**	**–**
nad5[b]	**1060**	**T**	**C**	**T**	**3rd**	**Synonymous (Phe)**	**–**
nad5[b]	1595	T	A	T	1st	Cys (G6, G7b)/Ser (G7a)	–
nad5	1854	C	T	T	2nd	Ser (G6)/Phe (G7)	–
cox3	2222	C	T	T	2nd	Ala (G6)/Val (G7)	–
cob	3413	T	C	C	3rd	Synonymous (Ile)	–
cob	3587	G	A	A	3rd	Synonymous (Val)	–
nad4	4446	C[a]	T	T	2nd	Ala (G6)/Val (G7)	[a]1 G6 sample had T (Val)
nad4	5029	C[a]	T	T	3rd	Synonymous (Val)	[a]2 G6 samples had T (synonymous)
atp6	5928	T	C[a]	C	3rd	Synonymous (Val)	[a]2 G7a samples had A (synonymous)
atp6	5959	G	A	A	1st	Val (G6)/Met (G7)	–
atp6	6160	A	T	T	1st	Thr (G6)/Ser (G7)	–
nad2	**6353**	**T**	**G**[a]	**G**	**3rd**	**Phe (G6)/Leu (G7)**	[a]1 G7a sample had T (Phe), 1 G7a sample had A (synonymous)
nad2[b]	**6491**	**A**	**G**	**A**	**3rd**	**Synonymous (Val)**	**–**
nad2	**6524**	**A**[a]	**G**	**G**	**3rd**	**Synonymous (Val)**	**–**
nad2[b]	**6620**	**C**	**A**	**C**	**3rd**	**Ile (G6, G7b)/Met (G7a)**	**–**
Intergenic region	Between 7085 and 7086	Inserted G	–	–	–	–	–
nad1	7788	T	C	C	3rd	Synonymous (Cys)	–
nad1	7884	G	A	A	3rd	Synonymous (Leu)	–
nad1	8005	A	G	G	1st	Met (G6)/ Val (G7)	–
nad3	8593	C	T	T	3rd	Synonymous (Phe)	–
cox1	9656	G	A	A	3rd	Synonymous (Leu)	–
cox1[b]	9734	T	C	T	3rd	Synonymous (Pro)	–
cox1[b]	10,550	T	C	T	3rd	Synonymous (Tyr)	–
L-rRNA	11,375	C[a]	T	T	–	–	[a]3 G6 samples had T
L-rRNA	11,491	T	G[a]	G	–	–	[a]1 G7a sample had T
S-rRNA	11,762	G	A	A	–	–	–
S-rRNA[b]	11,766	A	G	A	–	–	–
cox2	12,759	A	G	G	3rd	Synonymous (Met)	–
cox2	12,819	T	C	C	3rd	Synonymous (Ser)	–
nad6	13,390	A	G	G	3rd	Synonymous (Leu)	–
nad6[b]	13,447	A	G	A	3rd	Synonymous (Glu)	–
nad6	13,499	T	C	C	1st	Synonymous (Leu)	–
nad6	13,513	C	T	T	3rd	Synonymous (Ala)	–

[a] Positions where one or more samples had a different nucleotide than the rest of the samples of the same genotype in that position.
[b] Positions where G7b shared the same nucleotide with genotype G6.

3. Results

3.1. Defining positions in the mitochondrial genome for distinguishing G6, G7a and G7b

In the complete mitogenome dataset, 37 informative positions were found that allowed the distinction of G6 and G7, and G7a and G7b (Table 1). Of these, 31 positions were in protein-coding genes, of which 12 represented non-synonymous substitutions (resulting in amino acid change), and 19 were synonymous (amino acid remained the same). A total of 29 positions were of diagnostic value for differentiating genotypes G6 and G7, i.e. positions where genotype G6 samples were represented by one genotype-specific nucleotide, whereas genotype G7 samples had a different genotype-specific nucleotide (Table 1). However, of these 29, seven represented positions where one or more samples were represented by a different nucleotide, other than what was characteristic to the rest of the samples of the same genotype. For example, at position 4446 (*nad4* gene region) G6 samples had C in that position, and G7 (both G7a and G7b) samples had T, but one G6 sample had the same nucleotide as genotype G7 samples (nucleotide T). Similar situation was observed at position 11,491 (l-rRNA) – genotype G6 samples were characterised by nucleotide T, and genotype G7 (both G7a and G7b) by nucleotide G, whereas one G7a sample had nucleotide T (same as genotype G6 samples). Therefore, these seven positions were not considered as consistently genotype defining, since they did not allow for unequivocal distinguishing of G6 and G7.

In the mitogenome there were altogether eight informative positions where haplogroup G7b samples had the same nucleotide as genotype G6 samples, whereas genotype G7 haplogroup G7a samples had another nucleotide in that position. For example, in the position 9734 of the *cox1* gene region: G6 samples had nucleotide T and G7a had nucleotide C, while G7b shared the same nucleotide T in that position as genotype G6 samples (Table 1). However, in the whole mitogenome there were 11 additional positions specific to haplogroup G7b, whereas G6 and G7a had a different nucleotide in these positions. For example, position 6992 in the *nad2* gene region: G6 and G7a samples had nucleotide A, whereas G7b samples had G (Table S2).

3.2. New mtDNA markers for distinguishing G6 and G7, and also G7a and G7b

Three gene regions were found in the mitogenome that had the highest number of positions for distinguishing G6, G7, G7a and G7b: *nad2* (four positions), *nad5* (six positions) and *nad6* (four positions). Within the *nad2* gene region there were two diagnostic positions for differentiating genotypes G6 and G7, and two positions that allowed for the distinguishing of G7 haplogroups G7a and G7b. Additionally, in the *nad5* gene there were four diagnostic positions for differentiating G6 and G7, and two positions for allocating samples into G7a and G7b. In addition, there was one position within the *nad2* gene and one position on the *nad5* gene that was not consistently genotype defining: i) at position 6353 (*nad2* gene) one G7a sample had nucleotide T, which was characteristic to genotype G6 samples; and one G7a sample had nucleotide A, while the rest of genotype G7 (both G7a and G7b) samples had nucleotide G; ii) at position 714 (*nad5* gene) two G7a samples shared the same nucleotide C as genotype G6 samples, while the rest of G7 (both G7a and G7b) samples were characterised by nucleotide T. Within the *nad6* gene there were three positions with diagnostic value for distinguishing G6 and G7, and only one position diagnostic for differentiating G7a and G7b. However, for a more reliable assigment of samples into G6, G7a and G7b, it was determined that the combination of *nad2* and *nad5* gene regions would be optimal (Fig. 2).

3.3. Defining positions within the cox1 gene

We also explored whether the full *cox1* gene (1608 bp) is suitable for allocating samples into G6-G7 and G7a-G7b, and found that there were altogether three informative positions (Table 1). Of these three, one position (9656) allowed for a clear discrimination of genotypes G6 and G7, whereas at the remaining two positions (9734 and 10,550) haplogroup G7b samples shared the same nucleotide with genotype G6 samples (both were characterised by nucleotide T), while G7a had C.

3.4. Validation of new primers by PCR and sequencing of nad2 and nad5 genes

Fifteen samples of G6 (n = 5), G7a (n = 5) and G7b (n = 5) were successfully amplified with the newly designed primers for *nad2* and with primers for *nad5* previously described in Kinkar et al. (2018b), resulting in PCR products of 781 bp and 759 bp, respectively (data not shown). After sequencing and trimming, the final length of the sequences was 714 bp (*nad2*) and 680 bp (*nad5*), corresponding to positions 6280–6993 (*nad2*) and 691–1370 (*nad5*) in the mitogenome sequence MH300955 in GenBank (Laurimäe et al., 2018b). All four informative positions in the *nad2* gene located in this 714 bp gene fragment. Of the six informative positions in the full *nad5* gene, three were positioned within the 680 bp gene fragment (Table 1).

3.5. Phylogenetic networks

The phylogenetic network of combined *nad2* (714 bp) and *nad5* (680 bp), 1394 bp in total, allowed for a clear distinction of G6 and G7 genotypes, as well as of haplogroups G7a and G7b (Fig. 2). However, the phylogenetic network of the *cox1* (366 bp) gene fragment did not

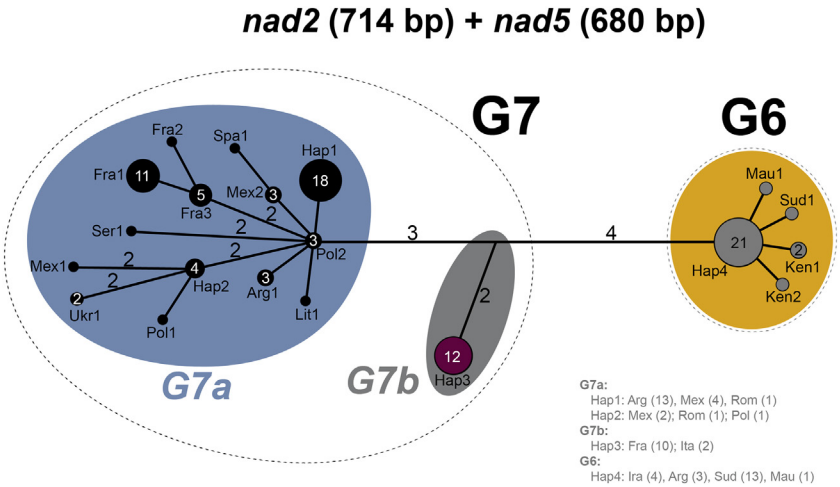

Fig. 2. Phylogenetic network of concatenated *nad2* (714 bp) and *nad5* (680 bp) gene sequences (1394 bp in total). Black dots depict haplogroup G7a (n = 55) samples, purple dots G7b (n = 12), and grey dots genotype G6 (n = 26) samples. Numbers inside the dots represent the number of samples. Numbers above the lines indicate the number of mutations. Lines without numbers denote one-mutational steps. Haplotypes comprising samples from a single country are marked with country three letter abbreviations. Haplotypes comprising samples from multiple countries are named "Hap", and their sample origins are listed in the bottom right corner (the numbers in brackets beside the country name represent the number of samples). Genotype G6 is marked with an orange background. Genotype G7 haplogroup G7a is defined with a blue background, and genotype G7 haplogroup G7b is marked with a grey background. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

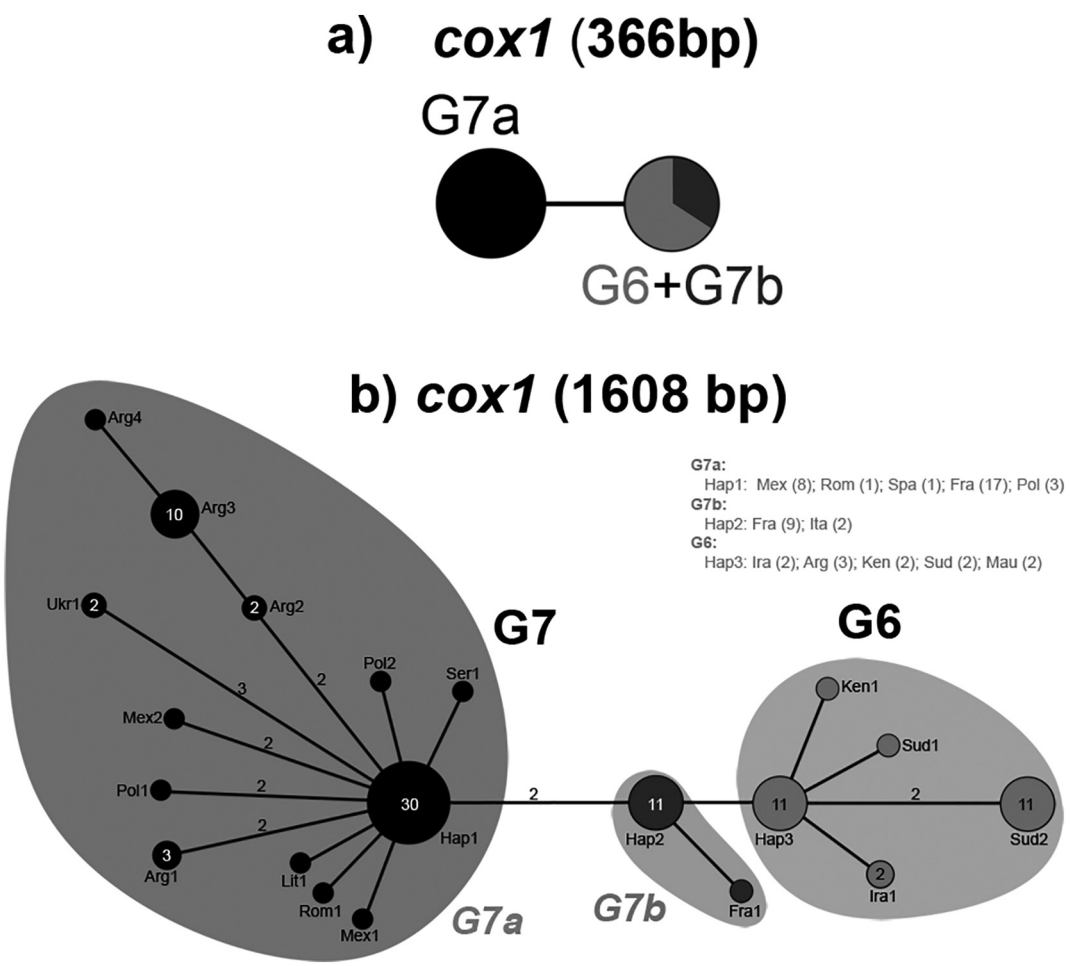

Fig. 3. Phylogenetic networks constructed for two different genes using the same set of samples as for constructing the network depicted in Fig. 2: a) *cox1* (366 bp) gene fragment; b) complete *cox1* (1608 bp) gene region. Black circles represent genotype G7a samples, dark grey circles mark G7b samples, while light grey depicts samples belonging to genotype G6. Numbers inside the dots represent the number of samples. Numbers above the lines indicate the number of mutations. Lines without numbers denote one-mutational steps. Haplotypes in Fig. 3b comprising samples from a single country are marked with country three letter abbreviations. Haplotypes comprising samples from multiple countries are named "Hap", and their sample origins are listed in the top right corner of Fig. 3b (the numbers inside brackets beside the country name represent the number of samples).

have sufficient phylogenetic power to accurately allocate samples into the correct genotype (Fig. 3a). It is important to note that G7b samples were grouped together into one haplotype with G6 samples, and the rest of the G7a samples were just one mutation apart from the G6 + G7b haplotype (Fig. 3a). Even on the basis of the complete *cox1* gene (1608 bp), the accurate allocation of samples remained challenging – G7b positioned closer to G6 than to G7a samples (Fig. 3b). However, if higher resolution is required (e.g., for phylogeographical analysis), using the *nad2* + *nad5* gene fragments (1394 bp in total) together with either 366 bp of the *cox1* gene fragment (Fig. S1) or with the complete *cox1* (1608 bp) gene (Fig. S2) can increase the phylogenetic resolution of the analysis.

4. Discussion

In terms of human and animal health it is of paramount importance to determine whether there are significant epidemiological, ecologial or biological differences between *E. granulosus s.l.* genotypes/species (for example, in infectivity and clinical course in humans). The basis for such research relies on the correct identification of genotypes. However, consistent allocation of samples into genotypes G6 and G7 based on the widely used, but relatively short sequences of *cox1* and *nad1*, has remained ambiguous, as a number of samples have been observed to be in intermediate positions between G6 and G7. Therefore a number of these in-between samples have been classified as G6/G7, without specifying the genotype. While for the analyses of genetic diversity, population structure and phylogeography (e.g., migration patterns) it is recommended to use complete or near-complete mitogenome sequence data (Kinkar et al., 2016, 2017, 2018a, 2018c; Laurimäe et al., 2016, 2018b), sequencing the whole mitogenome for genotype assignment is often not feasible, nor practical. Therefore we aimed to determine a new set of mtDNA markers for reliable (but also easy and cost-effective) assignment of samples to genotypes G6 and G7, and also to haplogroups G7a and G7b.

In the current study the *nad2* (714 bp) and *nad5* (680 bp) mitochondrial gene fragments (1394 bp in total) were determined as the best combination for reliable allocation of samples into genotypes G6 and G7, but also to haplogroups G7a and G7b (Table 1; Fig. 2). The *nad2* and *nad5* combination provides altogether four diagnostically relevant positions for differentianting between genotypes G6 and G7. Additionally, within *nad2* and *nad5* there are three informative positions for allocating samples into haplogroups G7a and G7b. Thus, we strongly recommend sequencing both the *nad2* and *nad5* gene fragments for reliable G6-G7 genotype discrimination, as well as for allocating samples into haplogroups G7a and G7b. We encourage to use this approach for accurate genotype determination, as correctly assigned samples lay a solid basis for future research to determine any epidemiologically significant differences between the different genetic

variants. Although the new assay is slightly more expensive and time consuming compared to sequencing just one gene fragment (e.g., 366 bp of *cox1*), it is still cost-effective and easy to perform. However, note that for the initial screening or if only species determination (and not of genotypes) within the *E. granulosus s.L. complex* is required, the conventional *cox1* (366 bp) gene fragment is still a good option.

Out of the 29 positions with diagnostic value for distinguishing G6 and G7, 22 allowed for consistent allocation of samples into the correct genotype cluster. The remaining seven represented positions where one or more samples had a different nucleotide than what was characteristic to the rest of the samples of the same genotype (e.g., positions 4446, 5029, and 5928; Table 1). When the sample size of G6 and G7 becomes significantly larger in time, there will likely be fewer genotype defining positions than reported in the current study, as the genetic variability of G6 and G7 is very high. Therefore, for reliable distinction of genotypes G6 and G7, we suggest constructing a phylogenetic network based on the *nad2* and *nad5* sequences (1394 bp in total). The sequences can be aligned with the reference sequences provided in this study (supplementary file "Reference Sequences for G6-G7.fas"; the same alignment can be accessed also via Mendeley; DOI: 10.17632/d2w6xnmz9y.1). The reference dataset includes sequences of genotypes G6 and G7, but also of haplogroups G7a and G7b from geographically wide locations.

Intrestingly, there were a total of eight positions within the whole mitogenome where G7 haplogroup G7b shared the same nucleotide with genotype G6, while haplogroup G7a samples were defined by a different nucleotide (e.g., position 13,447: G6 and G7b had nucleotide A; G7a had nucleotide G). One such position was within the widely used *cox1* (366 bp) gene fragment (Table 1; pos. 9734). As a result, haplogroup G7b samples clustered into one haplotype together with genotype G6 samples, and therefore would have been mistakenly classified as genotype G6 samples (Fig. 3a). This would in turn imply that genotype assignments based on the *cox1* (366 bp) gene fragment could be erroneous, at least in some cases. Furthermore, even the complete *cox1* (1608 bp) gene did not improve the phylogenetic power enough to reliably allocate samples into genotypes. Within the complete *cox1* gene there were two positions where both G6 and G7b were characterised by the same nucleotide, and only one genotype defining position. As also discussed in Laurimäe et al. (2018b), this means that based on the complete *cox1* gene the G7b samples would be positioned closer to the genotype G6 cluster (Fig. 3b).

Whether there are any significant ecological, biological or epidemiological differences between G6 and G7, but also between G7a and G7b, remains to be studied. Therefore, in order to evaluate these potential differences both in animals and in humans, it is important to accumulate data about the exact genotypes and haplogroups without limiting it to the species within *E. granulosus s.l.*. So far it has been especially difficult to allocate the haplotypes that position between G6 and G7 samples of G7b into the correct genotype cluster, and as a result the prevalence, distribution range and host assemblages of this haplogroup have remained unknown. Although G7b was first defined in our recent study (Laurimäe et al., 2018b), there has also been an indication of similar genetic structuring of genotype G7 from earlier studies based on sequences of *cox1* and *nad1* genes (Umhang et al., 2014; Addy et al., 2017). However, the exact genetic relationship of the samples with their respective genotypes remained unclear since these were placed in an intermediate position between the genotypes G6 and G7. The combination of gene regions of *nad2* (714 bp) and *nad5* (680 bp) now allow confident assignment of samples into their respective genotypes and haplogroups.

While other mtDNA markers and assays were developed for genotype distinction (e.g., Rostami Nejad et al., 2008; Šnábel et al., 2009; Boubaker et al., 2013, 2016; Mogoye et al., 2013), these were based on a relatively small number of samples and short sequences. Moreover, at that time only few mitogenomes were available to confirm and validate the reliability of such markers, especially for genotypes G6 and G7. Although the widely used *cox1* gene will remain relevant for species discrimination within *E. granulosus s.l.*, it lacks power to distinguish genotypes. Therefore we suggest to use the *nad5* marker for the correct differentiation of genotypes G1 and G3 (Kinkar et al., 2018b), and the combination of *nad2/nad5* for the differentiation of G6 and G7, or haplogroups G7a and G7b. However, if a more comprehensive analysis is required (e.g., phylogeographical), sequencing near-complete or complete mitogenomes is highly recommended, but if this is not possible due to limited funding, combining the *nad2* and *nad5* sequences with those of *cox1* and/or *nad1*, would enhance the resolution of phylogeographical analysis considerably compared to the commonly used short *cox1* gene fragment of 366 bp.

5. Conclusions

The accurate identification of genotypes within the species complex *E. granulosus* sensu *lato* has important epidemiological implications, as it can inform about the zoonotic potential of different genotypes and facilitates communication among scientists, veterinarians and medical doctors. For correct determination of *E. granulosus s.l.* genotypes G6 and G7, and also haplogroups G7a and G7b, we suggest to use an assay developed in this study that is based on the sequencing of *nad2* (714 bp) and *nad5* (680 bp) gene fragments (Fig. 4). We also suggest to align

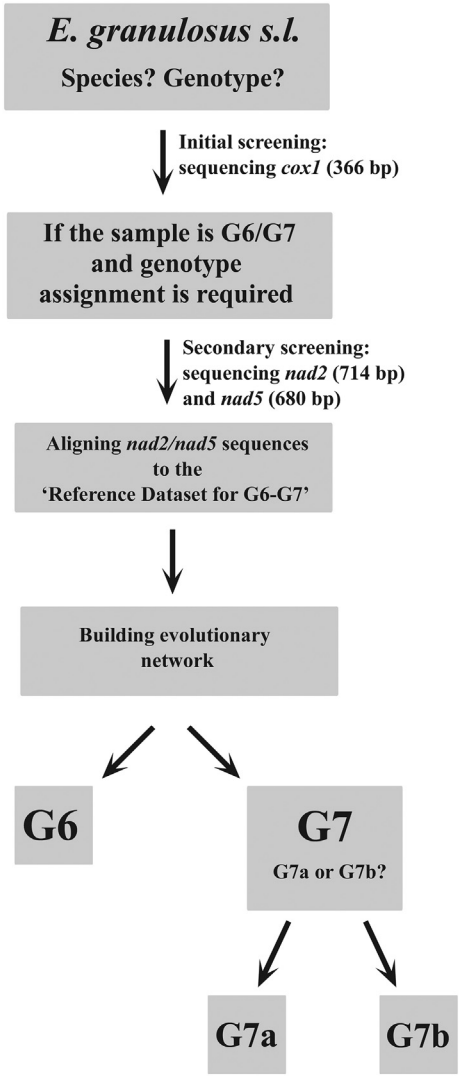

Fig. 4. A flowchart demonstrating the assignment of *E. granulosus* sensu *lato* samples into genotypes G6 and G7. For genotype G7, it is possible to assign samples also to haplogroups G7a and G7b (see Laurimäe et al., 2018b).

sequences with the reference dataset of *nad2* and *nad5* provided here as a supplementary file "Reference Sequences for G6-G7.fas". However, we recommend constructing a phylogenetic network based on the concatenated *nad2* and *nad5* sequences. This enables to ascertain the correct genotype (and haplogroup) with confidence.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.103941.

Acknowledgements

This work was supported by institutional research funding (IUT20-32) from the Estonian Ministry of Education and Research, and the Estonian Doctoral School of Ecology and Environmental Sciences. This work was partially supported by ERANET-LAC 2nd Joint Call and the Italian Ministry of Health - NDTND project (http://www.eranet-lac.eu).

Declarations of interest

None.

References

Addy, F., Wassermann, M., Kagendo, D., Ebi, D., Zeyhle, E., Elmahdi, I.E., Umhang, G., Casulli, A., Harandi, M.F., Aschenborn, O., Kern, P., Mackenstedt, U., Romig, T., 2017. Genetic differentiation of the G6/7 cluster of *Echinococcus canadensis* based on mitochondrial marker genes. Int. J. Parasitol. 47, 923–931. https://doi.org/10.1016/j.ijpara.2017.06.003.

Alvarez Rojas, C.A., Romig, T., Lightowlers, M.W., 2014. *Echinococcus granulosus* sensu lato genotypes infecting humans – review of current knowledge. Int. J. Parasitol. 44, 9–18. https://doi.org/10.1016/j.ijpara.2013.08.008.

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Beato, S., Parreira, R., Roque, C., Gonçalves, M., Silva, L., Maurelli, M.P., Cringoli, G., Grácio, M.A., 2013. *Echinococcus granulosus* in Portugal: the first report of the G7 genotype in cattle. Vet. Parasitol. 198, 235–239. https://doi.org/10.1016/j.vetpar.2013.08.021.

Boubaker, G., Macchiaroli, N., Prada, L., Cucher, M.A., Rosenzvit, M.C., Ziadinov, I., Deplazes, P., Saarma, U., Babba, H., Gottstein, B., Spiliotis, M., 2013. A multiplex PCR for the simultaneous detection and genotyping of the *Echinococcus granulosus* complex. PLoS Negl. Trop. Dis. 7 (1), e2017. https://doi.org/10.1371/journal.pntd.0002017.

Boubaker, G., Marinova, I., Gori, F., Hizem, A., Müller, N., Casulli, A., Jerez Puebla, L.E., Babba, H., Gottstein, B., Spiliotis, M., 2016. A dual PCR-based sequencing approach for the identification and discrimination of *Echinococcus* and *Taenia* taxa. Mol. Cell. Probes 30, 211–217.

Bowles, J., McManus, D.P., 1993. NADH dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus*. Int. J. Parasitol. 23, 969–972.

Bowles, J., Blair, D., McManus, D.P., 1992. Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. Mol. Biochem. Parasitol. 54, 165–173.

Bowles, J., Blair, D., McManus, D., 1994. Molecular genetic characterization of the cervid strain ('northern form') of *Echinococcus granulosus*. Parasitology 109, 215–221.

Cucher, M.A., Macchiaroli, N., Baldi, G., Camicia, F., Prada, L., Maldonado, L., Avila, H.G., Fox, A., Gutiérrez, A., Negro, P., López, R., Jensen, O., Rosenzvit, M., Kamenetzky, L., 2016. Cystic echinococcosis in South America: systematic review of species and genotypes of *Echinococcus granulosus sensu lato* in humans and natural domestic hosts. Tropical Med. Int. Health 21, 166–175. https://doi.org/10.1111/tmi.12647.

Deplazes, P., Rinaldi, L., Alvarez Rojas, C.A., Torgerson, P.R., Harandi, M.F., Romig, T., Antolova, D., Schurer, J.M., Lahmar, S., Cringoli, G., Magambo, J., Thompson, R.C., Jenkins, E.J., 2017. Global distribution of alveolar and cystic echinococcosis. Adv. Parasitol. 95, 315–493. https://doi.org/10.1016/bs.apar.2016.11.001.

Dybicz, M., Gierczak, A., Dabrowska, J., Rdzanek, L., Michalowicz, B., 2013. Molecular diagnosis of cystic echinococcosis in humans from Central Poland. Parasitol. Int. 62, 364–367. https://doi.org/10.1016/j.parint.2013.03.005.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Jabbar, A., Narankhajid, M., Nolan, M.J., Jex, A.R., Campbell, B.E., Gasser, R.B., 2011. A first insight into the genotypes of *Echinococcus granulosus* from humans in Mongolia. Mol. Cell. Probes 25, 49–54. https://doi.org/10.1016/j.mcp.2010.11.001.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kedra, A.H., Swiderski, Z., Tkach, V.V., Dubinsky, P., Pawlowski, Z., Stefaniak, J., Pawlowski, J., 1999. Genetic analysis of *Echinococcus granulosus* from humans and pigs in Poland, Slovakia and Ukraine. A multicenter study. Acta Parasitol. 44, 248–254.

Kinkar, L., Laurimäe, T., Simsek, S., Balkaya, I., Casulli, A., Manfredi, M.T., Ponce-Gordo, F., Varcasia, A., Lavikainen, A., Gonzalez, L.M., Rehbein, S., van der Giessen, J., Sprong, H., Saarma, U., 2016. High-resolution phylogeography of zoonotic tapeworm *Echinococcus granulosus* sensu stricto genotype G1 with an emphasis on its distribution in Turkey, Italy and Spain. Parasitology 143, 1790–1801. https://doi.org/10.1017/S0031182016001530.

Kinkar, L., Laurimäe, T., Sharbatkhori, M., Mirhendi, H., Kia, E.B., Ponce-Gordo, F., Andresiuk, V., Simsek, S., Lavikainen, A., Irshadullah, M., Umhang, G., Oudni-M'rad, M., Acosta-Jamett, G., Rehbein, S., Saarma, U., 2017. New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm *Echinococcus granulosus* sensu stricto. Infect. Genet. Evol. 52, 52–58. https://doi.org/10.1016/j.meegid.2017.04.023.

Kinkar, L., Laurimäe, T., Acosta-Jamett, G., Andresiuk, V., Balkaya, I., Casulli, A., Gasser, R.B., van der Giessen, J., Gonzalez, L.M., Haag, K.L., Zait, H., Irshadullah, M., Jabbar, A., Jenkins, D.J., Kia, E.B., Manfredi, M.T., Mirhendi, H., M'rad, S., Rostami Nejad, M., Oudni-M'rad, M., Pierangeli, N.B., Ponce-Gordo, F., Rehbein, S., Sharbatkhori, M., Simsek, S., Soriano, S.V., Sprong, H., Šnabel, V., Umhang, G., Varcasia, A., Saarma, U., 2018a. Global phylogeography and genetic diversity of the zoonotic tapeworm *Echinococcus granulosus* sensu stricto genotype G1. Int. J. Parasitol. 48, 729–742. https://doi.org/10.1016/j.ijpara.2018.03.006.

Kinkar, L., Laurimäe, T., Acosta-Jamett, G., Andresiuk, V., Balkaya, I., Casulli, A.,Gasser, R., Gonzalez, L.M., Haag, K.L., Houria, Z., Irshadullah, M., Jabbar, A., Jenkins, D.D., Manfredi, M.T., Mirhendi, H., M'rad, S., Rostami-Nejad, M., Oudni-M'rad, M., Pierangeli, N.B., Ponce-Gordo, F., Rehbein, S., Sharbatkhori, M., Kia, E.B., Simsek, S., Soriano, S.V., Sprong, H., Snabel, V., Umhang, G., Varcasia, A., Saarma, U. (2018b) Distinguishing *Echinococcus granulosus* sensu stricto genotypes G1 and G3 with confidence: a practical guide. Infect. Genet. Evol. 64, 178–184. Doi: https://doi.org/10.1016/j.meegid.2018.06.026.

Kinkar, L., Laurimäe, T., Balkaya, I., Casulli, A., Zait, H., Irshadullah, M., Sharbatkhori, M., Mirhendi, H., Rostami Nejad, M., Ponce-Gordo, F., Rehbein, S., Kia, E.B., Simsek, S., Šnabel, V., Umhang, G., Varcasia, A., Saarma, U., 2018c. Genetic diversity and phylogeography of the elusive, but epidemiologically important *Echinococcus granulosus* sensu stricto genotype G3. Parasitology 145, 1613–1622. https://doi.org/10.1017/S0031182018000549.

Kinkar, L., Korhonen, P.K., Cai, H., Gauci, C.G., Lightowlers, M.W., Saarma, U., Jenkins, D.J., Li, J., Li, J., Young, N.D., Gasser, R.B., 2019. Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (sensu stricto) genotype G1. Parasit. Vectors 12, 238. https://doi.org/10.1186/s13071-019-3492-x.

Knapp, J., Nakao, M., Yanagida, T., Okamoto, M., Saarma, U., Lavikainen, A., Ito, A., 2011. Phylogenetic relationships within Echinococcus and Taenia tapeworms (Cestoda: Taeniidae): an inference from nuclear protein-coding genes. Mol. Phylogenet. Evol. 61, 628–638. https://doi.org/10.1016/j.ympev.2011.07.022.

Knapp, J., Gottstein, B., Saarma, U., Millon, L., 2015. Taxonomy, phylogeny and molecular epidemiology of Echinococcus multilocularis: from fundamental knowledge to health ecology. Vet. Parasitol. 213, 85–91. https://doi.org/10.1016/j.vetpar.2015.07.030.

Konyaev, S.V., Yanagida, T., Nakao, M., Ingovatova, G.M., Shoykhet, Y.N., Bondarev, A.Y., Odnokurtsev, V.A., Loskutova, K.S., Lukmanova, G.I., Dokuchaev, N.E., Spiridonov, S., Alshinecky, M.V., Sivkova, T.N., Andreyanov, O.N., Abramov, S.A., Krivopalov, A.V., Karpenko, S.V., Lopatina, N.V., Dupal, T.A., Sako, Y., Ito, A., 2013. Genetic diversity of *Echinococcus* spp. in Russia. Parasitology 140, 1637–1647. https://doi.org/10.1017/S0031182013001340.

Laurimaa, L., Davison, J., Süld, K., Plumer, L., Oja, R., Moks, E., Keis, M., Hindrikson, M., Kinkar, L., Laurimäe, T., Abner, J., Remm, J., Anijalg, P., Saarma, U., 2015. First report of highly pathogenic *Echinococcus granulosus* genotype G1 in European Union urban environment. Parasit. Vectors 8 (182). https://doi.org/10.1186/s13071-015-0796-3.

Laurimäe, T., Kinkar, L., Andresiuk, V., Haag, K.L., Ponce-Gordo, F., Acosta-Jamett, G., Garate, T., Gonzalez, L.M., Saarma, U., 2016. Genetic diversity and phylogeography of highly zoonotic *Echinococcus granulosus* genotype G1 in the Americas (Argentina, Brazil, Chile and Mexico) based on 8279 bp of mtDNA. Infect. Genet. Evol. 45, 290–296. https://doi.org/10.1016/j.meegid.2016.09.015.

Laurimäe, T., Kinkar, L., Moks, E., Romig, T., Omer, R.A., Casulli, A., Umhang, G., Bagrade, G., Irshadullah, M., Sharbatkhori, M., Mirhendi, H., Ponce-Gordo, F., Soriano, S.V., Varcasia, A., Rostami-Nejad, M., Andresiuk, V., Saarma, U., 2018a. Molecular phylogeny based on six nuclear genes suggests that *Echinococcus granulosus* sensu lato genotypes G6/G7 and G8/G10 can be regarded as two distinct species. Parasitology 145, 1929–1937. https://doi.org/10.1017/S0031182018000719.

Laurimäe, T., Kinkar, L., Romig, T., Omer, R.A., Casulli, A., Umhang, G., Gasser, R., Jabbar, A., Sharbatkori, M., Mirhendi, H., Ponce-Gordo, F., Lazzarini, L., Soriano, S.V., Varcasia, A., Rostami-Nejad, M., Andresiuk, V., Maravilla, P., Gonzalez, L., Dybicz, M., Gawor, J., Šarkunas, M., Snabel, V., Kuzmina, T., Saarma, U., 2018b. The benefits of analysing complete mitochondrial genomes: deep insights into the phylogeny and population structure of *Echinococcus granulosus* sensu lato genotypes G6 and G7. Infect. Genet. Evol. 64, 85–94. https://doi.org/10.1016/j.meegid.2018.06.016.

Laurimäe, T., Kinkar, L., Varcasia, A., Dessi, G., Sgroi, G., D'Alessio, N., Veneziano, V., Saarma, U., 2019. First detection of zoonotic tapeworm *Echinococcus granulosus* sensu lato genotype G7 in continental Italy. Parasitol. Res. https://doi.org/10.1007/s00436-019-06346-2.

Lavikainen, A., Lehtinen, M.J., Meri, T., Hirvelä-Koski, V., Meri, S., 2003. Molecular genetic characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of *Echinococcus granulosus*. Parasitology 127, 207–215.

Lymbery, A. J. (2017). Phylogenetic pattern, evolutionary processes and species delimitation in the genus *Echinococcus*. Adv. Parasitol. 95, 111–145. https://doi.org/10.1016/bs.apar.2016.07.002

Mogoye, B.K., Menezes, C.N., Wong, M.L., Stacey, S., von Delft, D., Wahlers, K.,

Wassermann, M., Romig, T., Kern, P., Grobusch, M.P., Frean, J., 2013. First insights into species and genotypes of *Echinococcus* in South Africa. Vet. Parasitol. 196, 427–432. https://doi.org/10.1016/j.vetpar.2013.03.033.

Moks, E., Jõgisalu, I., Saarma, U., Talvik, H., Järvis, T., Valdmann, H., 2006. Helminthologic survey of the wolf (*Canis lupus*) in Estonia, with an emphasis on *Echinococcus granulosus*. J. Wildl. Dis. 42, 359–365. https://doi.org/10.7589/0090-3558-42.2.359.

Moks, E., Jõgisalu, I., Valdmann, H., Saarma, U., 2008. First report of *Echinococcus granulosus* G8 in Eurasia and a reappraisal of the phylogenetic relationships of 'genotypes' G5-G10. Parasitology 135, 647–654.

Nakao, M., Yanagida, T., Konyaev, S., Lavikainen, A., Odnokurtsev, V.A., Zaikov, V.A., Ito, A., 2013. Mitochondrial phylogeny of the genus *Echinococcus* (Cestoda:Taeniidae) with emphasis on relationships among *Echinococcus canadensis* genotypes. Parasitology 140, 1625–1636. https://doi.org/10.1017/S0031182013000565.

Nakao, M., Lavikainen, A., Hoberg, E., 2015. Is *Echinococcus intermedius* a valid species? Trends Parasitol. 31, 342–343.

Nikmanesh, B., Mirhendi, H., Mahmoudi, S., Rokni, M.B., 2017. Multilocus sequence analysis of *Echinococcus granulosus* strains isolated from humans and animals in Iran. Exp. Parasitol. 183, 50–55. https://doi.org/10.1016/j.exppara.2017.10.002.

Rodriguez-Prado, U., Jimenez-Gonzalez, D.E., Avila, G., Gonzalez, A.E., Martinez-Flores, W.A., Mondragon de la Peña, C., Hernandez-Castro, R., Romero-Valdovinos, M., Flisser, A., Martinez-Hernandez, F., Maravilla, P., Martinez-Maya, J.J., 2014. Short report: genetic variation of *Echinococcus canadensis* (G7) in Mexico. The Am. J. Trop. Med. Hygiene 91 (6), 1149–1153. https://doi.org/10.4269/ajtmh.14-0317.

Romig, T., Deplazes, P., Jenkins, D., Giraudoux, P., Massolo, A., Craig, P.S., Wassermann, M., Takahashi, K., de la Rue, M., 2017. Ecology and life cycle patterns of *Echinococcus* species. Adv. Parasitol. 95, 213–314. https://doi.org/10.1016/bs.apar.2016.11.002.

Rostami Nejad, M., Nazemalhosseini Mojarad, E., Nochi, Z., Fasihi Harandi, M., Cheraghipour, K., Mowlavi, G.R., Zali, M.R., 2008. *Echinococcus granulosus* strain differentiation in Iran based on sequence heterogeneity in the mitochondrial 12S rRNA gene. J. Helminthol. 82, 343–347. https://doi.org/10.1017/S0022149X0804594X.

Rostami, S., Shariat Torbaghan, S., Dabiri, S., Babaei, Z., Ali Mohammadi, M., Sharbatkhori, M., Fasihi Harandi, M., 2015. Genetic characterization of *Echinococcus granulosus* from a large number of formalin-fixed, paraffin-embedded tissue samples of human isolates in Iran. The Am. J. Trop. Med. Hygiene 92 (3), 588–594. https://doi.org/10.4269/ajtmh.14-0585.

Saarma, U., Jõgisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., Simsek, S., Andresiuk, V., Denegri, G., Gonzalez, L.M., Ferrer, E., Garate, T., Rinaldi, L., Maravilla, P., 2009. A novel phylogeny for the genus *Echinococcus* based on nuclear data challenges relationships based on mitochondrial evidence. Parasitology 136, 317–328. https://doi.org/10.1017/S0031182008005453.

Schurer, J.M., Pawlik, M., Huber, A., Elkin, B., Cluff, H.D., Pongracz, J.D., Gesy, K., Wagner, B., Dixon, B., Merks, H., Bal, M.S., Jenkins, E.J., 2016. Intestinal parasites of gray wolves (Canis lupus) in northern and western Canada. Can. J. Zool. 94, 643–650.

Scott, J.C., Stefaniak, J., Pawlowski, Z.S., McManus, D.P., 1997. Molecular genetic analysis of human cystic hydatid cases from Poland: identification of a new genotypic group (G9) of *Echinococcus granulosus*. Parasitology 114, 37–43.

Šnábel, V., Altintas, N., D'Amelio, S., Nakao, M., Romig, T., Yolasigmaz, A., Gunes, K., Turk, M., Busi, M., Hüttner, M., Sevcová, D., Ito, A., Altintas, N., Dubinský, P., 2009. Cystic echinococcosis in Turkey: genetic variability and first record of the pig strain (G7) in the country. Parasitol. Res. 105, 145–154. https://doi.org/10.1007/s00436-009-1376-2.

Thompson, R., 2008. The taxonomy, phylogeny and transmission of *Echinococcus*. Exp. Parasitol. 119, 439–446. https://doi.org/10.1016/j.exppara.2008.04.016.

Thompson, R.C.A., 2017. Biology and systematics of *Echinococcus*. Adv. Parasitol. 95, 65–109. https://doi.org/10.1016/bs.apar.2016.07.001.

Thompson, R.A., McManus, D.P., 2002. Towards a taxonomic revision of the genus *Echinococcus*. Trends Parasitol. 18, 452–457. https://doi.org/10.1016/S1471-4922(02)02358-9.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Umhang, G., Richomme, C., Hormaz, V., Boucher, J.M., Boué, F., 2014. Pigs and wild boar in Corsica harbor *Echinococcus canadensis* G6/7 at levels of concern for public health and local economy. Acta Trop. 13, 64–68. https://doi.org/10.1016/j.actatropica.2014.02.005.

Yanagida, T., Lavikainen, A., Hoberg, E.P., Konyaev, S., Ito, A., Sato, M.O., Zaikov, V.A., Beckmen, K., Nakao, M., 2017. Specific status of *Echinococcus canadensis* (Cestoda: Taeniidae) inferred from nuclear and mitochondrial gene sequences. Int. J. Parasitol. 47, 971–979. https://doi.org/10.1016/j.ijpara.2017.07.001.

Zhang, T., Yang, D., Zeng, Z., Zhao, W., Liu, A., Piao, D., Jiang, T., Cao, J., Shen, Y., Liu, H., Zhang, W., 2014. Genetic characterization of human-derived hydatid cysts of *Echinococcus granulosus* Sensu Lato in Heilongjiang Province and the first report of G7 genotype of *E. canadensis* in humans in China. PLoS One 9 (10), e109059. https://doi.org/10.1371/journal.pone.0109059.