Parasitology

cambridge.org/par

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

Teivi Laurimäe[1], Liina Kinkar[1], Epp Moks[1], Thomas Romig[2], Rihab A. Omer[3], Adriano Casulli[4,5], Gérald Umhang[6], Guna Bagrade[7], Malik Irshadullah[8], Mitra Sharbatkhori[9], Hossein Mirhendi[10], Francisco Ponce-Gordo[11], Silvia V. Soriano[12], Antonio Varcasia[13], Mohammad Rostami-Nejad[14], Vanessa Andresiuk[15] and Urmas Saarma[1]

[1]Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 51003 Tartu, Estonia; [2]Institute of Zoology, Parasitology Unit, University of Hohenheim, 70599 Stuttgart, Germany; [3]National University Research Institute, National University Sudan; [4]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; [5]European Union Reference Laboratory for Parasites (EURLP), Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy; [6]Anses, Wildlife Surveillance and Eco-epidemiology Unit, National Reference Laboratory for *Echinococcus* spp., Nancy Laboratory for Rabies and Wildlife, 54220 Malzéville, France; [7]Latvian State Forest Research Institute 'Silava', 111 Rigas str., LV-2169 Salaspils, Latvia; [8]Section of Parasitology, Department of Zoology, Aligarh Muslim University, Aligarh 202002, India; [9]Laboratory Science Research Center, Golestan University of Medical Sciences, Gorgan, Iran; [10]Department of Parasitology and Mycology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran; [11]Department of Parasitology, Faculty of Pharmacy, Complutense University, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain; [12]Department of Microbiology and Parasitology, Faculty of Medical Sciences, Comahue National University, Buenos Aires 1400, 8300 Neuquén, Argentina; [13]Laboratorio di Parassitologia e Malattie Parassitarie, Ospedale Didattico Veterinario Dipartimento di Medicina Veterinaria, Università degli Studi di Sassari, *Via* Vienna 2, 07100 Sassari, Italy; [14]Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran and [15]Laboratorio de Zoonosis Parasitarias, FCEyN, UNMdP, Funes 3350, CP: 7600 Mar del Plata, Buenos Aires, Argentina

Research Article

Cite this article: Laurimäe T *et al* (2018). 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

Received: 22 January 2018
Revised: 12 March 2018
Accepted: 5 April 2018
First published online: 21 May 2018

Key words:
Cystic echinococcosis; hydatid disease; nuclear genes; nuclear phylogeny; zoonosis; zoonotic pathogens

Author for correspondence:
Urmas Saarma, E-mail: Urmas.Saarma@ut.ee

Abstract

Tapeworms of the species complex of *Echinococcus granulosus sensu lato* (*s. l.*) are the cause of a severe zoonotic disease – cystic echinococcosis, which is listed among the most severe parasitic diseases in humans and is prioritized by the World Health Organization. A stable taxonomy of *E. granulosus s. l.* is essential to the medical and veterinary communities for accurate and effective communication of the role of different species in this complex on human and animal health. *E. granulosus s. l.* displays high genetic diversity and has been divided into different species and genotypes. Despite several decades of research, the taxonomy of *E. granulosus s. l.* has remained controversial, especially the species status of genotypes G6–G10. Here the Bayesian phylogeny based on six nuclear loci (7387 bp in total) demonstrated, with very high support, the clustering of G6/G7 and G8/G10 into two separate clades. According to the evolutionary species concept, G6/G7 and G8/G10 can be regarded as two distinct species. Species differentiation can be attributed to the association with distinct host species, largely separate geographical distribution and low level of cross-fertilization. These factors have limited the gene flow between genotypic groups G6/G7 and G8/G10, resulting in the formation of distinct species. We discuss ecological and epidemiological differences that support the validity of these species.

Introduction

Tapeworms belonging to the species complex of *Echinococcus granulosus sensu lato* (*s. l.*) have been identified as the cause of a chronic zoonotic disease known as cystic echinococcosis (CE), a disease that has considerable impact on both livestock and human health worldwide (Craig *et al.*, 2017; WHO, 2017). The general life cycle of *E. granulosus s. l.* involves various carnivores as definitive hosts for the adult stage, including mostly dogs in both rural and urban areas, and wolves (Moks *et al.*, 2006; Schurer *et al.*, 2014; Laurimaa *et al.*, 2015*a*; Thompson, 2017). Both domesticated and wild large mammalian herbivores act as intermediate hosts for the larval stage. The larval stage is in the form of hydatid cysts that are predominantly located in the liver and/or lungs of the intermediate hosts. Cysts can be fertile or sterile, depending on presence or absence of protoscoleces, respectively. While these cysts can cause significant health problems for the infected intermediate hosts, the infection in the definitive host is usually asymptomatic (Thompson, 2017).

CAMBRIDGE UNIVERSITY PRESS

The taxonomy of *E. granulosus s. l.* has been a challenging issue for decades. It is well established that this parasite complex displays high genetic diversity and on the basis of mitochondrial DNA (mtDNA) it has been divided into eight different genotypes (G1, G3, G4–G8; and G10; Bowles *et al.*, 1992, 1994; Lavikainen *et al.*, 2003; Kinkar *et al.*, 2017). Several of these recognized mitochondrial genotypes have differences in their lifecycles, hosts ranges and morphology (Thompson and McManus, 2002; Romig *et al.*, 2017; Thompson, 2017). These differences have provided grounds to consider some of these genotypes as distinct species: G1 and G3 as *E. granulosus sensu stricto* (*s. s.*; Kinkar *et al.*, 2017), G4 as *Echinococcus equinus* and G5 as *Echinococcus ortleppi* (Thompson and McManus, 2002; Lymbery, 2017). The analytical power has been low in most studies as the analyses have been based largely on short sequences of mtDNA, most often on a fragment of a single gene (e.g. Casulli *et al.*, 2012; Andresiuk *et al.*, 2013). Recent studies based on considerably longer mtDNA sequences (Kinkar *et al.*, 2016, 2018*a*, 2018*b*; Laurimäe *et al.*, 2016) have yielded significantly deeper insight into the phylogeny and phylogeography of different genotypes. For example, using sequences of nearly complete mitochondrial genomes and three nuclear genes, Kinkar *et al.* (2017) have revised the status of *E. granulosus s. s.* and demonstrated that genotypes G1 and G3 are distinct mitochondrial genotypes, whereas G2 is not a separate genotype or even a monophyletic cluster, but belongs to G3. On the other hand, nuclear data revealed no genetic separation of G1 and G3, suggesting that these genotypes form a single species due to ongoing gene flow. The authors concluded that in the taxonomic sense, genotypes G1 and G3 can be treated as a single species *E. granulosus s. s.*, and that G1 and G3 should be regarded as distinct genotypes only in the context of mitochondrial data, whereas G2 was recommended to be excluded from the genotype list (Kinkar *et al.*, 2017). A recently discovered isolate from Ethiopia is tentatively retained in *E. granulosus s. s.* as a genotype distant from G1/G3 awaiting taxonomic positioning (Wassermann *et al.*, 2016). In contrast, the species status of *E. equinus* (G4), *E. ortleppi* (G5) and *Echinococcus felidis* has, to date, remained undisputed (Hüttner *et al.*, 2008; Thompson, 2008; Saarma *et al.*, 2009; Knapp *et al.*, 2011; Romig *et al.*, 2015; Lymbery, 2017).

The species status of the four *E. granulosus s. l.* genotypes G6, G7, G8 and G10, however, has remained uncertain. In addition to genetic differences, there are also various ecological and epidemiological differences between these mitochondrial genotypes. Genotypes G6 and G7 are known to be perpetuated predominantly in a domestic life cycle involving goats, camels or pigs as intermediate hosts, and dogs as definitive hosts; however, a recent study found members of the G6/G7 cluster to be widespread in wild mammals of southern Africa (Romig *et al.*, 2017). Genotypes G8 and G10 are mostly circulating in a sylvatic cycle with moose and reindeer acting as intermediate hosts, and wolves as definitive hosts. Moreover, these four genotypic groups are largely allopatric. The distribution range of G6 and G7 covers more southern areas such as Western Europe, the Mediterranean area, Africa, South and Central America, and the Middle East (e.g. Varcasia *et al.*, 2006, 2007; Lymbery *et al.*, 2015). Genotypes G8 and G10 have been found to coexist in the northern hemisphere – mostly in northern part of Europe (e.g. Estonia, Finland, Sweden and Latvia), Northern Asia and Canada (Moks *et al.*, 2006, 2008; Konyaev *et al.*, 2013; Schurer *et al.*, 2014; Marcinkute *et al.*, 2015; Oksanen and Lavikainen, 2015).

Previous studies have mostly focused on mtDNA to resolve the phylogeny and taxonomic status of genotypes G6–G10 (e.g. Moks *et al.*, 2008; Nakao *et al.*, 2013; Addy *et al.*, 2017). These studies demonstrated that the cervid genotype G10 is a sister taxon to the camel–pig genotypes G6/G7, rather than assuming a sister position with the other cervid genotype G8. It was therefore suggested to combine G6–G10 into a single species which, in terms of priority, should be *E. canadensis* (Nakao *et al.*, 2007; Hüttner *et al.*, 2008). Moreover, the mitochondrial studies placed *E. multilocularis* in the midst of the *E. granulosus* genotypes, rendering the *E. granulosus* complex paraphyletic and contradicting the classical taxonomy of the genus (reviewed in Knapp *et al.*, 2015). Although mtDNA sequences are widely used and various parasite identification methods have been developed based on these (e.g. Boubaker *et al.*, 2013; Laurimaa *et al.*, 2015*b*) one has to be cautious when interpreting the results. MtDNA represents the evolution of the maternal linage, which can have different trajectories than that of the species. As argued in Saarma *et al.* (2009), once a new mtDNA mutation becomes fixed in a population, the new mitochondrial lineage separates from the ancestral one; from this point onwards, mutations continue to fix progressively in an independent manner in both the new and ancestral mitochondrial lineages, and mitochondrial lineages continue to diverge. However, this does not mean that these separate mitochondrial lineages have necessarily become separate biological entities – genetic exchange between different taxa can only be assessed with nuclear markers. Thus, it was clear that nuclear data are needed to clarify the taxonomy of the genus *Echinococcus*. Indeed, a phylogeny radically different from that of mtDNA data was inferred by using sequences of five nuclear genes (5086 bp in total); this analysis placed G8 and G10 as sister taxa, and *E. multilocularis* clearly separate from the *E. granulosus s. l.* complex (Saarma *et al.*, 2009). However, in this work G6/G7 were represented by isolates from cattle and pig, and it was not evident which of these two genotypes these isolates belonged to (probably G7). Since the clear distinction between G6 and G7 was not made in this study, the exact phylogenetic relations between G6 and G7 remained obscure. On the other hand, the analysis performed by Knapp *et al.* (2011) based on a different set of nuclear loci suggested (in line with the mtDNA data), that *E. granulosus s. l.* complex could be paraphyletic. Unfortunately, this study did not include G10 and therefore the exact phylogenetic relations in the G6–G10 group remained unresolved. Thus, despite numerous attempts to revise the phylogeny and taxonomy of genotypes G6–G10, no consensus has been reached. Some authors have proposed to treat G6–G10 provisionally as a single species *E. canadensis* awaiting further data from the nuclear genome (e.g. Nakao *et al.*, 2007; Moks *et al.*, 2008; Nakao *et al.*, 2013; Romig *et al.*, 2015; Addy *et al.*, 2017), while others as two distinct species: G6/G7 as *E. intermedius* and G8/G10 as *E. canadensis* (Thompson, 2008; Saarma *et al.*, 2009) or even as three species: G6/G7 as *E. intermedius*, G8 as *E. borealis* and G10 as *E. canadensis* (Lymbery *et al.*, 2015).

The main aim of this study was to use a more extensive range of nuclear loci and include all four genotypes (G6–G8 and G10) in a phylogenetic analysis to resolve their taxonomic status.

Materials and methods

Parasite material

Samples of *E. granulosus* genotypes G5–G10 used in this study (41 in total) originated from various regions and intermediate or final hosts (Fig. 1, Table 1). Genotype G5 was represented by two samples from India. Samples of genotype G6 ($n = 8$) were from three continents: South America (Argentina, $n = 3$), Africa (Kenya, $n = 1$; Sudan, $n = 2$) and Eurasia (Iran, $n = 2$). Samples of genotype G7 ($n = 14$) were from South America (Argentina, $n = 3$), and Eurasia (Spain, $n = 1$; France, $n = 7$; Italy, $n = 2$; Romania, $n = 1$). Samples that belonged to genotype G8 ($n = 8$) originated from Eurasia (Estonia, $n = 3$; Latvia, $n = 5$), whereas

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press

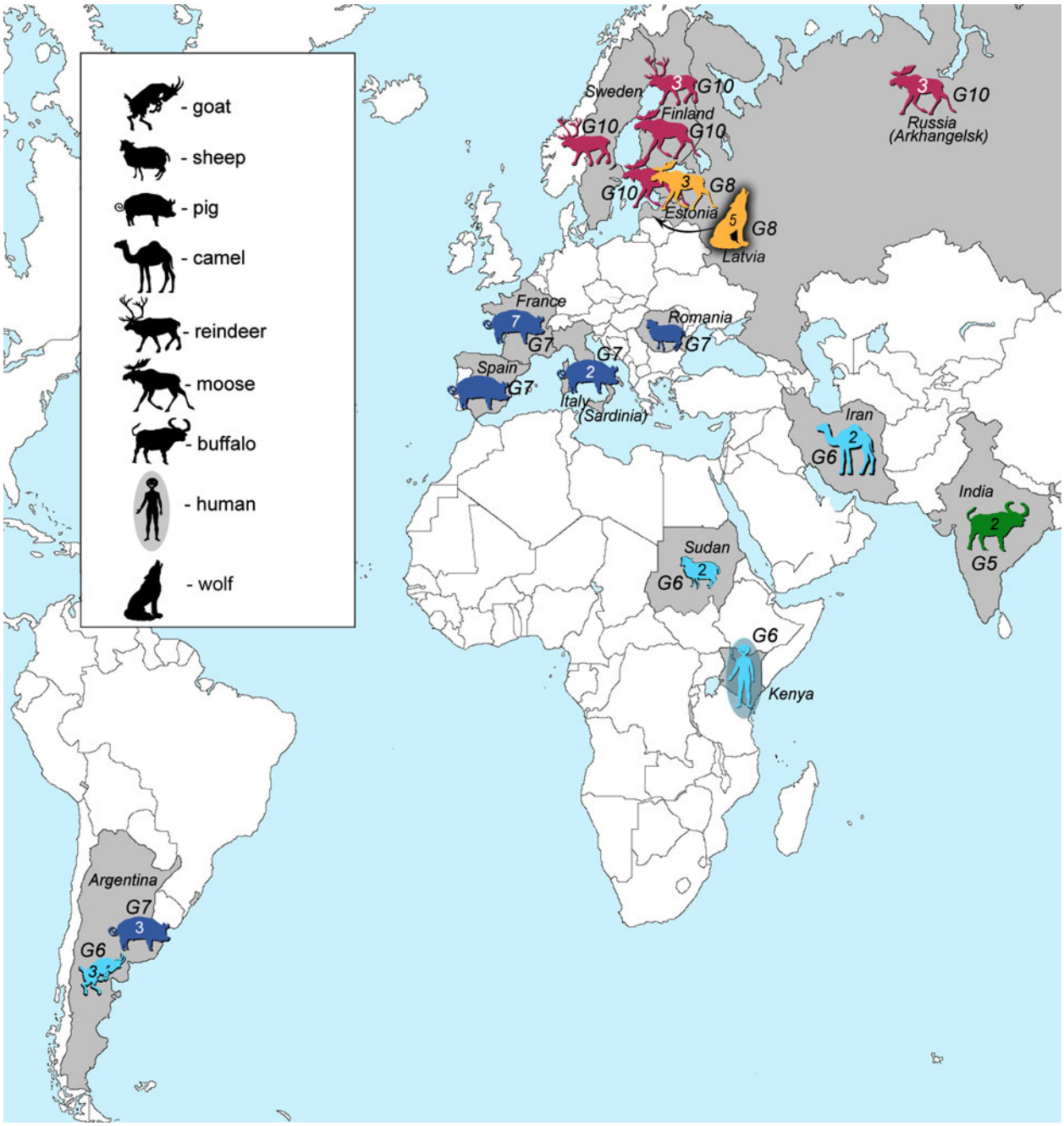

Fig. 1. Geographic locations and host species (intermediate or final) of all of the analysed samples in this study. Numbers inside the animal figures stand for the number of samples collected. Green colour represents *E. ortleppi* (G5) samples, cyan genotype G6 samples, dark blue genotype G7 samples, orange G8 and pink G10 samples.

genotype G10 was represented by nine samples from Eurasia (Sweden, $n = 1$; Finland, $n = 4$; Russia, $n = 3$; Estonia, $n = 1$). DNA of the Finnish, Swedish and Russian specimens were provided by Antti Lavikainen. These specimens were already used and their genotypic identities defined in previously published studies (Lavikainen *et al.*, 2006; Nakao *et al.*, 2013). Samples were ethanol-preserved at −20 °C until further use.

PCR amplification and sequencing of six nuclear loci

High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) was used to extract DNA from either protoscoleces or cyst membranes, following the manufacturer's protocols. Six nuclear genes were chosen for PCR amplification and sequencing: transforming growth factor beta receptor kinase (*tgf*; 1137 bp), calreticulin (*cal*; 1138 bp), elongation factor 1 alpha (*ef1*; 1055 bp), ezrin–radixin–moesin-like protein (*elp*; 780 bp), phosphoenolpyruvate carboxykinase (*pepck*; 1506 bp) and DNA polymerase delta (*pold*; 1771 bp). For further details on cycle parameters for PCR and sequencing see Saarma *et al.* (2009; *tgf*, *cal*, *ef1*, *elp*) and Knapp *et al.* (2011; *pepck*, *pold*). Nuclear sequences were deposited in GenBank (MG766944–MG767169). Consensus sequences were assembled using Codon Code Aligner v6.0.2. BioEdit v7.2.5 (Hall, 1999) was used for multiple alignments with Clustal W and for manual correction of sequences.

As polymorphic sites where the same mutations are shared between genotypes G6 and G10 have been found in *pepck* and *pold* genes (Yanagida *et al.*, 2017), we also checked our aligned sequences for polymorphic sites that discriminate between

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press

Table 1. Data for samples of *Echinococcus granulosus sensu lato* genotypes G5–G10

Sample no.	nDNA	Genotype	Host	Origin	Accession no *ef1*	Accession no *cal*	Accession no *tgf*	Accession no *elp*	Accession no *pepck*	Accession no *pold*
1	G5	G5	Buffalo	India	MG766944	MG766984	MG767024	MG767064	MG767098	MG767135
2	G5	G5	Buffalo	India	MG766945	MG766985	MG767025	MG767065	MG767099	MG767136
3	G6	G6	Camel	Iran, Isfahan	MG766946	MG766986	MG767026	–	–	MG767137
4	G6	G6	Camel	Iran, Isfahan	MG766947	MG766987	MG767027	MG767066	MG767100	MG767138
5	G6	G6	Goat	Argentina, Neuquen	MG766948	MG766988	MG767028	MG767067	MG767101	MG767139
6	G6	G6	Goat	Argentina, Neuquen	MG766949	MG766989	MG767029	MG767068	MG767102	MG767140
7	G6	G6	Goat	Argentina, Neuquen	MG766950	MG766990	MG767030	MG767069	MG767103	MG767141
8	G6	G6	Human	Kenya, Turkana	MG766951	MG766991	MG767031	–	MG767104	–
9	G6	G6	Sheep	Sudan, Nyala	MG766952	MG766992	MG767032	–	MG767105	MG767142
10	G6	G6	Sheep	Sudan, Nyala	MG766953	MG766993	MG767033	MG767070	MG767106	MG767143
11	G7e	G7	Pig	Argentina, Buenos Aires	MG766954	MG766994	MG767034	MG767071	MG767107	MG767144
12	G7e	G7	Pig	Argentina	MG766955	MG766995	MG767035	MG767072	MG767108	MG767145
13	G7c	G7	Pig	Argentina, Neuquen	MG766956	MG766996	MG767036	–	MG767109	MG767146
14	G7c	G7	Pig	Central Spain, Segovia	MG766957	MG766997	MG767037	MG767073	MG767110	MG767147
15	G7d	G7	Sheep	Romania	MG766958	MG766998	MG767038	–	–	–
16	G7d	G7	Pig	Italy, Sardinia	MG766959	MG766999	MG767039	MG767074	MG767111	MG767148
17	G7d	G7	Pig	Italy, Sardinia	MG766960	MG767000	MG767040	MG767075	MG767112	MG767149
18	G7d	G7	Pig	France, Corsica	MG766961	MG767001	MG767041	MG767076	MG767113	MG767150
19	G7d	G7	Pig	France, Corsica	MG766962	MG767002	MG767042	MG767077	MG767114	–
20	G7d	G7	Pig	France, Corsica	MG766963	MG767003	MG767043	MG767078	MG767115	MG767151
21	G7d	G7	Pig	France, Corsica	MG766964	MG767004	MG767044	MG767079	MG767116	MG767152
22	G7d	G7	Pig	France, Corsica	MG766965	MG767005	MG767045	MG767080	MG767117	MG767153
23	G7d	G7	Pig	France, Corsica	MG766966	MG767006	MG767046	MG767081	MG767118	MG767154
24	G7d	G7	Pig	France, Corsica	MG766967	MG767007	MG767047	MG767082	MG767119	MG767155
25	G8a	G8	Moose	Estonia	MG766968	MG767008	MG767048	MG767083	MG767120	MG767156
26	G8a	G8	Moose	Estonia	MG766969	MG767009	MG767049	MG767084	MG767121	MG767157
27	G8a	G8	Moose	Estonia	MG766970	MG767010	MG767050	MG767085	MG767122	MG767158
28	G8a	G8	Wolf	Latvia	MG766971	MG767011	MG767051	MG767086	MG767123	MG767159
29	G8a	G8	Wolf	Latvia	MG766972	MG767012	MG767052	MG767087	MG767124	MG767160
30	G8a	G8	Wolf	Latvia	MG766973	MG767013	MG767053	MG767088	MG767125	MG767161
31	G8a	G8	Wolf	Latvia	MG766974	MG767014	MG767054	MG767089	MG767126	MG767162

32	G8a	G8	Wolf	Latvia	MG766975	MG767015	MG767055	MG767090	MG767127	MG767163	
33	G8b[a]	G8	Moose	USA	–	–	–	–	FN567995	FN568366	
34	G10d	G10	Moose	Estonia	MG766976	MG767016	MG767056	–	–	–	
35	G10c[b]	G10	Moose	Finland	EU834907	EU834940	EU834916	EU834895	–	–	
36	G10c	G10	Reindeer	Finland	MG766977	MG767017	MG767057	MG767091	MG767128	MG767164	
37	G10c	G10	Reindeer	Finland	MG766978	MG767018	MG767058	MG767092	MG767129	MG767165	
38	G10c	G10	Reindeer	Finland	MG766979	MG767019	MG767059	MG767093	MG767130	MG767166	
39	G10c	G10	Reindeer	Sweden	MG766980	MG767020	MG767060	MG767094	MG767131	MG767167	
40	G10c	G10	Moose	Russia, Arkhangelsk	MG766981	MG767021	MG767061	MG767095	MG767132	–	
41	G10c	G10	Moose	Russia, Arkhangelsk	MG766982	MG767022	MG767062	MG767096	MG767133	MG767168	
42	G10c	G10	Moose	Russia, Arkhangelsk	MG766983	MG767023	MG767063	MG767097	MG767134	MG767169	

[a]Sample sequences of *pepck* and *pold* obtained from the GenBank database (Knapp *et al.*, 2011).
[b]Sample sequences of *ef1*, *cal*, *tgf* and *elp* obtained from the GenBank database (Saarma *et al.*, 2009).

genotypes G6 and G10, as well as for positions where mutations were shared between these genotypes.

Bayesian phylogeny

Bayesian phylogenies were constructed for two datasets, both based on six nuclear genes (7387 bp in total): (1) Dataset 1 (a total of 40 sequences): 39 samples of G6–G8 and G10 analysed in this study, and one additional G8 sample from GenBank, originating from the USA (accession numbers for *pepck* and *pold* were FN567995 and FN568366, respectively; Knapp *et al.*, 2011); (2) Dataset 2 (a total of 42 sequences): the same set of samples as in Dataset 1 and two additional sequences of genotype G5.

The best-fit nucleotide substitution model was selected on the basis of BIC (Bayesian Information Criterion) scores using jModelTest 2 (Guindon and Gascuel, 2003; Darriba *et al.*, 2012). Bayesian phylogenetic analysis was performed in BEAST 1.8.4 (Drummond *et al.*, 2012) using StarBeast (Heled and Drummond, 2010). Posterior distributions of parameters were estimated by using the MCMC (Markov Chain Monte Carlo) sampling. Total length of the chain was 10 000 000 and the parameters were logged every 1000 generations. The resulting phylogenetic trees were summarized and annotated using TreeAnnotator 1.8.4 and visualized with FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

Results

Total length of the alignment based on six nuclear loci was 7387 bp: *ef1* 1055 bp, *cal* 1138 bp, *tgf* 1137 bp, *elp* 780 bp, *pepck* 1506 bp and *pold* 1771 bp. However, a few of the samples did not yield positive results for all analysed nuclear loci, but as BEAST allows analysis with some missing data, these samples were also included in the analysis (Table 1). All of the samples were homozygotes at all six nuclear loci.

Across the six nuclear loci, 12 polymorphic positions were found that discriminated between G6/G7 and G8/G10. However, in the *pepck*, mutations in three positions were shared between two G7 isolates (samples 13 and 14) and G8/G10 isolates. According to GenBank reference FN567995 (Knapp *et al.*, 2011) these positions were: 236; 1435–1436; 1513.

The best-fit nucleotide substitution model used for the nuclear DNA (nDNA) data was GTR + I + G. The Bayesian phylogeny for the Dataset 1 revealed that genotypes G6 and G7 formed one clade, whereas G8 and G10 another (Fig. 2). Posterior probability values for both nodes assigning G6/G7 and G8/G10 into two different clades were very high (1.00). According to the evolutionary (general lineage) species concept, they can be regarded as two distinct species.

Internal nodes for the clade G8/G10 also received high posterior probability values (0.98 and 1.00). It was shown that G8b (the GenBank sample from the USA) was a sister taxon to G10d (Estonia) and that G10c was a sister taxon to the G8b/G10d clade. Additionally, the tree topology indicated that G8a was positioned as a basal taxon in relation to the G10c/G10d/G8b clade. Similarly to G8/G10 clade, the internal nodes for G6/G7 also received high posterior probability values (0.96 and 1.00). The resultant tree topology shows that G6 is a sister taxon to G7e and that G7d is sister to G6/G7e. G7c occupied a basal position inside the G6/G7 clade.

We also performed a phylogenetic analysis for the Dataset 2 (included G5), as well as with only the samples for which all six nuclear loci were sequenced (Table 1). These analyses yielded essentially the same phylogenetic relations between G6 and G10 as with the larger dataset (Fig. 2; Supplementary Fig. S1–S3).

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press

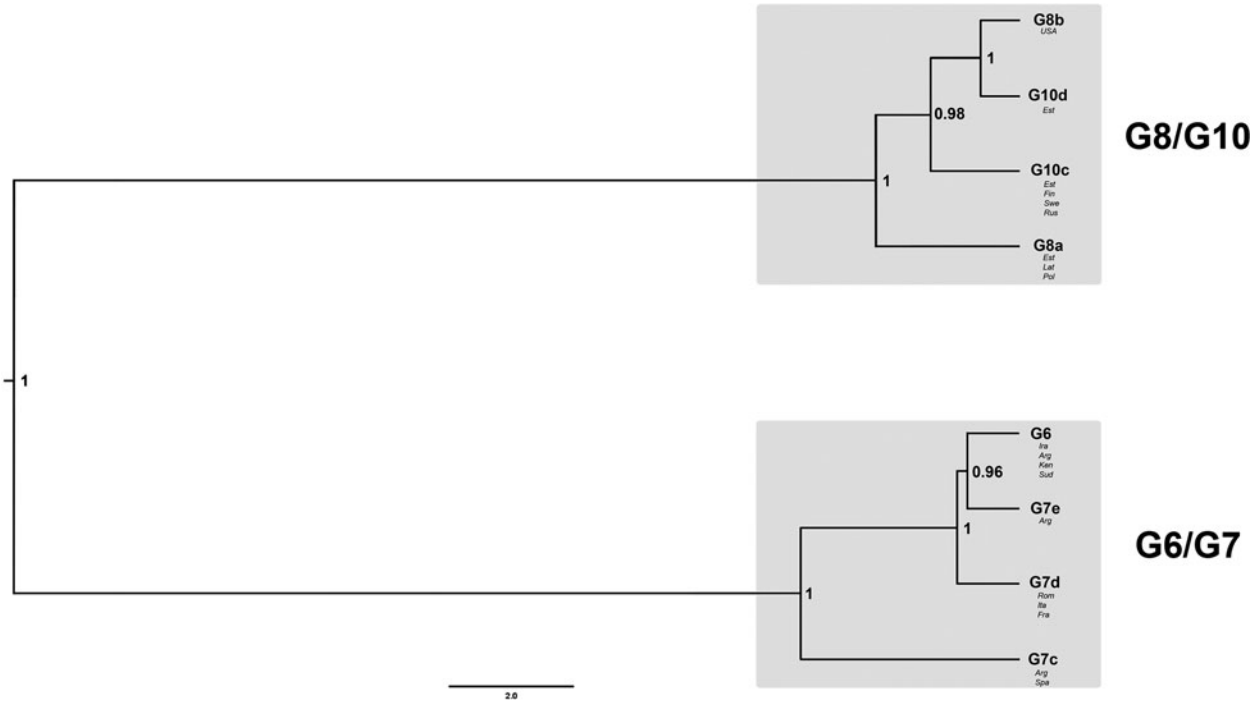

Fig. 2. Bayesian phylogeny of genotypes G6–G8 and G10, based on sequences of six nuclear loci (Dataset 1; 40 samples). The numbers on nodes represent posterior probability values. For further details on the included samples see Table 1.

Discussion

A stable taxonomy of *E. granulosus s. l.* is essential to the medical and veterinary communities for accurate and effective communication of the role of different species in this complex on human and animal health. Despite several decades of research, the taxonomy of *E. granulosus s. l.* has remained controversial and a subject of intense discussion (Saarma *et al.*, 2009; Knapp *et al.*, 2011; Lymbery *et al.*, 2015; Nakao *et al.*, 2015). Most of the studies aiming to resolve the taxonomic status of genotypes G6–G10 have relied on mtDNA (e.g. Lavikainen *et al.*, 2003; Nakao *et al.*, 2007; Hüttner *et al.*, 2008; Moks *et al.*, 2008; Nakao *et al.*, 2013; Addy *et al.*, 2017). However, the mitochondrial genome can only reveal the evolutionary history of the maternal lineage, which can be different to that of the species. For species delimitation, a key component is the analysis of genetic exchangeability, which can be effectively studied only by using various loci from the nuclear genome (Saarma *et al.*, 2009). Until recently, only two studies have analysed multiple nuclear loci to infer the phylogeny of *E. granulosus s. l.* (Saarma *et al.*, 2009; Knapp *et al.*, 2011), yielding contradictory results. Moreover, both studies did not include all genotypes of the G6–G10 complex.

Nuclear data and taxonomy of G6–G10

The Bayesian phylogeny based on six nuclear loci clustered the camel–pig genotypes G6/G7 into one clade and the cervid genotypes G8/G10 into another clade (Fig. 2). This result provides strong support for the hypothesis according to which genotypes G6–G10 are divided into two species (Thompson, 2008; Saarma *et al.*, 2009). The internal division of the G6/G7 and G8/G10 clades provides evidence for gene flow between G6 and G7, as well as between G8 and G10, but non-existent or very limited gene flow between genotypic groups G6/G7 and G8/G10. The latter seems to be supported by a recent study based on two nuclear loci, which suggested some degree of gene flow between genotypic groups G6/G7 and G8/G10 (Yanagida *et al.*, 2017); however, their result could be also due to incomplete lineage sorting (see below). Since G6 and G7 are not distinct taxa based on nuclear data (notice in Fig. 2 that G7e forms a subclade with G6, while other isolates of G7 are sister to this), it demonstrates that the gene flow between G6 and G7 has been sufficient to guarantee that G6 and G7 have not diverged from each other. Exactly the same is valid for G8 and G10 (notice in Fig. 2 that G8 and G10 do not form separate subclades, but the isolates of both genotypes are not monophyletic).

Gene flow can occur under conditions of sympatry between both genotypic groups (i.e. G6/G7 and G8/G10). To date, none of the studies have demonstrated sympatry of all these four genotypes. Nevertheless, there are regions where at least some genotypes of these two genotypic groups are potentially sympatric. One such region is in north-eastern Europe, where G8 (this study) and G10 have been recorded from Latvia, and G7 in neighbouring Lithuania (Marcinkute *et al.*, 2015). Considering that wolves (as a main definitive host species for G8/G10) can cover very long distances and their populations in Europe are connected over the distance of more than 800 km (Hindrikson *et al.*, 2017), the possibility for gene flow between G7 and G8/G10 is potentially there, and yet the genotypic groups G6/G7 and G8/G10 are clearly separate on the nuclear phylogeny. Another region of potential sympatry is in eastern Russia, where G6 has been found in relative geographical proximity with G10 (>500 km between the reported cases) (Konyaev *et al.*, 2013; Yanagida *et al.*, 2017). Nevertheless, the phylogeny based on six nuclear loci (current study) shows also that gene flow between genotypic groups G6/G7 and G8/G10 has not been sufficient to merge all four genotypes into a single clade (species). A recent study by Yanagida *et al.* (2017) based on two nuclear loci (*pepck* and *pold*) indicated that some degree of gene flow might occur between G6/G7 and G8/G10 as they found few polymorphic sites where mutations were shared among G6/G7 and G8/G10. Based on this, they suggested that G6–G10 could be considered as one species. However, there were only a limited number of polymorphic characters in the two analysed loci, which may likely

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press

be the reason why their conclusion is not supported by the results of our study. One possible explanation for the shared characters reported in Yanagida *et al.* (2017) could be due to incomplete lineage sorting, which means that due to the relatively recent evolutionary divergence of G6/G7 and G8/G10, some loci have not had enough time to diverge and as a result there are still shared characters between different genotypes. This is actually evident also from our results. When we examined all six nuclear loci of our study (that include also *pepck* and *pold* used by Yanagida *et al.*, 2017), there are several positions in the alignment where the same nucleotide is shared between different genotypic groups. For example, in *pepck* all isolates of G8 and G10 have A in the shown position in Fig. 3, but remarkably A is also found in two isolates of G7. And yet, despite of some shared mutations between different genotypes, there are a large number of characters specific only either to the genotypic group G6/G7 or to G8/G10, and as a result in the phylogenetic tree the camel–pig genotypes G6/G7 firmly form one clade and the cervid genotypes G8/G10 another (Fig. 2). While we cannot rule out the possibility that to some extent gene flow (hybridization) between these two genotypic groups can occur, as suggested in Yanagida *et al.* (2017), the nDNA evidence in our study that is based on a larger number of nuclear loci compared with Yanagida and co-authors, clearly shows the phylogenetic division of G6–G10 into two clades, G6/G7 and G8/G10. According to the evolutionary (general lineage) species concept, these two clades can be regarded as distinct species as they represent two distinct evolutionary lineages and other data also support this (see below).

Limited gene flow between species, i.e. hybridization, is in fact relatively common in nature. Possibly the most popular example is the hybridization between wolves and dogs (e.g. Hindrikson *et al.*, 2012; Leonard *et al.*, 2014). In general, it has been estimated that 10–30% of multicellular animal and plant species hybridize regularly (Abbott *et al.*, 2013). Hybridization is also well-known among parasites, it is known for example between different species of helminths (*Taenia*, *Trichinella*, *Schistosoma*, *Fasciola*, *Ascaris*) and protozoa (*Plasmodium*, *Leishmania*, *Toxoplasma* and *Trypanosoma*) (Arnold, 2004; Detwiler and Criscione, 2010; King *et al.*, 2015). Hybridization between closely related tapeworm species in Taeniidae has been demonstrated between *T. saginata* and *T. asiatica* (e.g. Okamoto *et al.*, 2010; Yamane *et al.*, 2013). The occurrence of hybridization does not mean that two hybridizing species, if clearly separate on the phylogeny, should therefore be regarded as a single species, it just means that reproductive barrier between species has not yet fully developed.

Although our study did not include samples from the whole geographical range of the genotypes, we argue that this is not a major limitation, since our result shows that even in close geographical proximity these genotypic groups maintain their genetic differences. Moreover, our data included samples from north-eastern Europe where genotypes G7, G8 and G10 have been recorded in relatively close geographic areas. A need for including samples from the whole geographical range of the species would have been critical if the genetic data showed no differentiation on a smaller scale, but this was not the case here. Our results demonstrated that gene flow between G6/G7 and G8/G10 genotypic groups in relatively close geographical areas has been insufficient to merge them into a single clade, and instead they form two statistically well supported separate clades (species). One of the possible contributing factors for the limited gene flow between G6/G7 and G8/G10 could also be the reproduction mode of *E. granulosus s. l.* Although cross-fertilization can occur (e.g. Haag *et al.*, 2011), the main mode of reproduction appears to be self-fertilization (Lymbery, 2017; Thompson, 2017). As the potential for outcrossing between genotypic groups is rare, the evolutionary potential for genetic differentiation and species divergence is high (Lymbery, 2017).

Ecological, epidemiological and morphological differences of the two species

The division of G6–G10 into two separate species is also supported by other data that can be found in detail in Thompson (2008) and Saarma *et al.* (2009). Briefly, while G6/G7 is known to be typically circulating in the domestic cycle (camels, goats, pigs and dogs), G8/G10 cycles primarily in the sylvatic cycle, between cervids (moose, elk and reindeer) and wolves (Thompson and McManus, 2002; Lymbery, 2017). Although G6 is commonly involved in a cycle between goats/camels–dogs and G7 mainly pigs–dogs, these two also share some overlap in

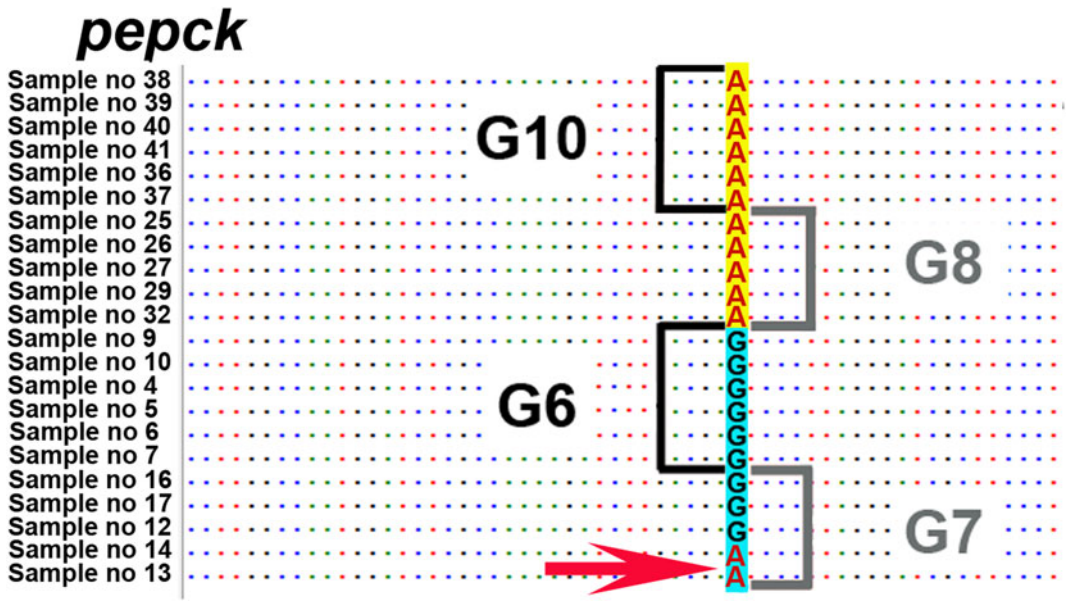

Fig. 3. Nucleotide position on *pepck* locus, where the same nucleotide A is shared between two samples of genotype G7 and the G8/G10 genotypic group. Depicted position according to FN567995 from the GenBank database is 236 (Knapp *et al.*, 2011). Sample numbers correspond to sample numbers in Table 1.

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press

their life cycles as both can infect the same intermediate hosts, such as goats and humans (Cardona and Carmena, 2013; Alvarez Rojas *et al.*, 2014; Addy *et al.*, 2017), and definitive host – dogs. Most likely dogs act as vectors for both genotypes, providing opportunities for outcrossing. Moreover, the geographical distribution of G6/G7 is largely different from G8/G10. G6 and G7 are sympatric in Turkey, Argentina and Peru (e.g. Moro *et al.*, 2009; Šnabel *et al.*, 2009; Soriano *et al.*, 2010; Simsek *et al.*, 2011; Lymbery *et al.*, 2015). The recent discovery of the G6/G7 cluster in African wildlife is highly interesting on phylogeographical grounds and is currently further explored (Romig *et al.*, 2017). The cervid strains G8 and G10 are, in contrast, distributed in the northern part of Eurasia and North America (Lavikainen *et al.*, 2003; Thompson *et al.*, 2006; Moks *et al.*, 2008) and so far there are only a few recorded occurrences of G6 in northern latitudes (Konyaev *et al.*, 2013; Yanagida *et al.*, 2017). As G6/G7 circulate primarily in the domestic cycle and G8/G10 in the sylvatic cycle, the probability that parasites from different genotypic groups co-occur in the same definitive host and cross-fertilize is very low. On the other hand, since G6/G7 share the same final host (dog) cross-fertilization has apparently been frequent enough to guarantee that G6/G7 have not diverged. The same is valid for G8/G10, which utilize wild canids (mostly wolves) as definitive hosts. Thus, the association with distinct host species, largely separate geographical distribution and limited rate of cross-fertilization are the main factors that have limited the gene flow between genotypic groups G6/G7 and G8/G10. As a result, these genotypic groups can be regarded as distinct species.

Morphological comparisons of adult worms of G6/G7 and G8/G10 are scarce. Recently, it has been found that genotypes G6 and G7 share similar morphological characteristics, e.g. long terminal segment when compared with the total adult worm length, genital pore is generally anterior in the mature segment and rostellar hook morphometric data have also given similar results for both of these genotypes (Soriano *et al.*, 2016). Based on the limited data, it has been suggested that there are some morphological differences in the reproductive anatomy between G6/G7 and G8/G10, and between rostellar hook morphology (Thompson *et al.*, 2006; Lymbery *et al.*, 2015; Soriano *et al.*, 2016). However, these differences need to be further confirmed as neither direct comprehensive morphological nor extensive ecological comparisons between G6/G7 and G8/G10 have been made so far. Such studies would provide additional valuable information for species delimitation.

Based on priority, the species name for G8/G10 should be *E. canadensis*; however, the species name for G6/G7 warrants further discussions. It has been proposed to use *E. intermedius* for G6/G7 (Thompson, 2008; Saarma *et al.*, 2009); however, this name is highly problematic since the original description by Lopez-Neyra and Soler Planas (1943) did not describe intermediate host and no original type specimen for *E. intermedius* can be found (Nakao *et al.*, 2015).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0031182018000719

Acknowledgements. We would like to thank Antti Lavikainen and Tetsuya Yanagida for helpful comments on the manuscript.

Financial support. This work was supported by institutional research funding (IUT20-32) from the Estonian Ministry of Education and Research (to U.S.); the Estonian Doctoral School of Ecology and Environmental Sciences (to T.L. and L.K.); Deutsche Forschungsgemeinschaft grant Ro 3753/3-1 (to T.R.); the European Community's Seventh Framework Programme under the grant agreement 602051, Project HERACLES; http://www.Heracles-fp7.eu/ (to A.C.).

Declarations of interest. None.

References

Abbott R *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229–246.

Addy F *et al.* (2017) Genetic differentiation of the G6/7 cluster of *Echinococcus canadensis* based on mitochondrial marker genes. *International Journal for Parasitology* **47**, 923–931.

Alvarez Rojas CA, Romig T and Lightowlers MW (2014) *Echinococcus granulosus* sensu lato genotypes infecting humans – review of current knowledge. *International Journal for Parasitology* **44**, 9–18.

Andresiuk MV *et al.* (2013) *Echinococcus granulosus* genotype G1 dominated in cattle and sheep during 2003–2006 in Buenos Aires province, an endemic area for cystic echinococcosis in Argentina. *Acta Tropica* **127**, 136–142.

Arnold M (2004) Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology* **13**, 997–1007.

Boubaker G *et al.* (2013) A multiplex PCR for the simultaneous detection and genotyping of the *Echinococcus granulosus* complex. *PLoS Neglected Tropical Diseases* **7**(1), e2017.

Bowles J, Blair D and McManus DP (1992) Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology* **54**, 165–173.

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

Cardona GA and Carmena D (2013) A review of the global prevalence, molecular epidemiology and economics of cystic echinococcosis in production animals. *Veterinary Parasitology* **192**, 10–32.

Casulli A *et al.* (2012) Genetic variability of *Echinococcus granulosus* sensu stricto in Europe inferred by mitochondrial DNA sequences. *Infection, Genetics and Evolution* **12**, 377–383.

Craig PS *et al.* (2017) Echinococcosis: control and prevention. *Advances in Parasitology* **96**, 55–158.

Darriba D *et al.* (2012) Jmodeltest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**(8), 772.

Detwiler JT and Criscione CD (2010) An infectious topic in reticulate evolution: introgression and hybridization in animal parasites. *Genes* **210**, 102–123.

Drummond AJ *et al.* (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969–1973.

Guindon S and Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52**, 696–704.

Haag KL *et al.* (2011) Reappraising the theme of breeding systems in *Echinococcus*: is outcrossing a rare phenomenon? *Parasitology* **138**, 298–302.

Hall TA (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Heled J and Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**(3), 570–580.

Hindrikson M *et al.* (2012) Bucking the trend in wolf-dog hybridization: first evidence from Europe of hybridization between female dogs and male wolves. *PLoS ONE* **7**(10), e46465.

Hindrikson M *et al.* (2017) Wolf population genetics in Europe: a systematic review, meta-analysis and suggestions for conservation and management. *Biological Reviews* **92**, 1601–1629.

Hüttner M *et al.* (2008) Genetic characterization and phylogenetic position of *Echinococcus felidis* Ortlepp, 1937 (Cestoda: Taeniidae) from the African lion. *International Journal for Parasitology* **38**, 861–868.

King KC *et al.* (2015) Hybridization in parasites: consequences for adaptive evolution, pathogenesis, and public health in a changing world. *PLoS Pathogens* **11**(9), e1005098.

Kinkar L *et al.* (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.

Kinkar L *et al.* (2017) New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm *Echinococcus granulosus* sensu stricto. *Infection, Genetics and Evolution* **52**, 52–58.

Kinkar L *et al.* (2018*a*) Global phylogeography and genetic diversity of the zoonotic tapeworm *Echinococcus granulosus* sensu stricto genotype G1. *International Journal for Parasitology*, in press.

Kinkar L *et al.* (2018*b*) Genetic diversity and phylogeography of the elusive, but epidemiologically important *Echinococcus granulosus* sensu stricto genotype G3. *Parasitology*. https://doi.org/10.1017/S0031182018000549.

Knapp J *et al.* (2011) Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda: Taeniidae): an inference from nuclear protein-coding genes. *Molecular Phylogenetics and Evolution* **61**, 628–638.

Knapp J *et al.* (2015) Taxonomy, phylogeny and molecular epidemiology of *Echinococcus multilocularis*: from fundamental knowledge to health ecology. *Veterinary Parasitology* **213**, 85–91.

Konyaev SV *et al.* (2013) Genetic diversity of *Echinococcus* spp. *in Russia*. *Parasitology* **140**, 1637–1647.

Laurimaa L *et al.* (2015*a*) First report of highly pathogenic *Echinococcus granulosus* genotype G1 in European Union urban environment. *Parasites and Vectors* **8**, 182.

Laurimaa L *et al.* (2015*b*) Noninvasive detection of *Echinococcus multilocularis* tapeworm in urban area, Estonia. *Emerging Infectious Diseases* **21**, 163–164.

Laurimäe T *et al.* (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. *Infection, Genetics and Evolution* **45**, 290–296.

Lavikainen A *et al.* (2003) Molecular genetic characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of *Echinococcus granulosus*. *Parasitology* **127**, 207–215.

Lavikainen A *et al.* (2006) Molecular characterization of *echinococcus* isolates of cervid origin from Finland and Sweden. *Parasitology* **133**, 565–570.

Leonard JA *et al.* (2014) Impact of hybridization on the conservation of wild canids. In Gompper ME (ed.), *Free Ranging Dogs and Wildlife Conservation*. Oxford: Oxford University Press, pp. 170–184.

Lopez-Neyra CR and Soler Planas MA (1943) Revision del genero *Echinococcus* Rud y descripcion de una especie nueva pararita intestinal del perro en Almeria. *Revista ibérica de parasitología* **3**, 169–194, (in Spanish).

Lymbery AJ (2017) Phylogenetic pattern, evolutionary processes and species delimitation in the genus *Echinococcus*. *Advances in Parasitology* **95**, 111–145.

Lymbery AJ *et al.* (2015) *Echinococcus canadensis*, *E. borealis*, and *E. intermedius*. What's in a name? *Trends in Parasitology* **31**, 23–29.

Marcinkute A *et al.* (2015) *Echinococcus* infections in the Baltic region. *Veterinary Parasitology* **213**, 121–131.

Moks E *et al.* (2006) Helminthologic survey of the wolf (*Canis lupus*) in Estonia, with an emphasis on *Echinococcus granulosus*. *Journal of Wildlife Diseases* **42**, 359–365.

Moks E *et al.* (2008) First report of *Echinococcus granulosus* G8 in Eurasia and a reappraisal of the phylogenetic relationships of 'genotypes' G5–G10. *Parasitology* **135**, 647–654.

Moro PL *et al.* (2009) Molecular identification of *Echinococcus* isolates from Peru. *Parasitology International* **58**, 184–186.

Nakao M, Lavikainen A and Hoberg E (2015) Is *Echinococcus intermedius* a valid species? *Trends in Parasitology* **31**, 342–343.

Nakao M *et al.* (2007) A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. *Parasitology* **134**, 713–722.

Nakao M *et al.* (2013) Mitochondrial phylogeny of the genus *Echinococcus* (Cestoda: Taeniidae) with emphasis on relationships among *Echinococcus canadensis* genotypes. *Parasitology* **140**, 1625–1636.

Okamoto M *et al.* (2010) Evidence of hybridization between *Taenia saginata* and *Taenia asiatica*. *Parasitology International* **59**, 70–74.

Oksanen A and Lavikainen A (2015) *Echinococcus canadensis* transmission in the North. *Veterinary Parasitology* **213**, 182–186.

Romig T, Ebi D and Wassermann M (2015) Taxonomy and molecular epidemiology of *Echinococcus granulosus* sensu lato. *Veterinary Parasitology* **231**, 76–84.

Romig T *et al.* (2017) Ecology and life cycle patterns of *Echinococcus* species. *Advances in Parasitology* **95**, 213–314.

Saarma U *et al.* (2009) A novel phylogeny for the genus *Echinococcus*, based on nuclear data, challenges relationships based on mitochondrial evidence. *Parasitology* **136**, 317–328.

Schurer JM *et al.* (2014) *Echinococcus multilocularis* and *Echinococcus canadensis* in wolves from western Canada. *Parasitology* **141**, 159–163.

Simsek S, Kaplan M and Ozercan IH (2011) A comprehensive molecular survey of *Echinococcus granulosus* in formalin-fixed paraffin-embedded tissues in human isolates in Turkey. *Parasitology Research* **109**, 411–416.

Šnabel V *et al.* (2009) Cystic echinococcosis in Turkey: genetic variability and first record of the pig strain (G7) in the country. *Parasitology Research* **105**, 145–154.

Soriano SV *et al.* (2010) Molecular characterization of *Echinoccocus* isolates indicates goats as reservoir for *Echinococcus canadensis* G6 genotype in Neuquen, Patagonia Argentina. *Parasitology International* **59**, 626–628.

Soriano SV *et al.* (2016) First study about the development of adult *Echinococcus canadensis* G6 genotype of goat origin in experimentally infected dogs. *Veterinary Parasitology* **228**, 6–12.

Thompson RCA (2008) The taxonomy, phylogeny and transmission of *Echinococcus*. *Experimental Parasitology* **119**, 439–446.

Thompson RCA (2017) Biology and systematics of *Echinococcus*. *Advances in Parasitology* **95**, 65–109.

Thompson RCA and McManus DP (2002) Towards a taxonomic revision of the genus *Echinococcus*. *Trends in Parasitology* **18**, 452–457.

Thompson RCA *et al.* (2006) Molecular and morphological characterization of *Echinococcus* in cervids from North America. *Parasitology* **132**, 439–447.

Varcasia A *et al.* (2006) Molecular characterization of *Echinococcus granulosus* strains in Sardinia. *Parasitology Research* **98**, 273–277.

Varcasia A *et al.* (2007) Molecular characterization of *Echinococcus granulosus* in sheep and goats of Peloponnesus, Greece. *Parasitology Research* **101**, 1135–1139.

Wassermann M *et al.* (2016) A novel zoonotic genotype related to *Echinococcus granulosus* sensu stricto from southern Ethiopia. *International Journal for Parasitology* **46**, 663–668.

WHO (2017). Echinococcosis. Fact Sheet (Updated March 2017). World Health Organization. Available at http://www.who.int/mediacentre/factsheets/fs377/en/ (last accessed 12.12.2017).

Yamane K *et al.* (2013) Genotypic relationships between *Taenia saginata*, *Taenia asiatica* and their hybrids. *Parasitology* **140**, 1595–1601.

Yanagida T *et al.* (2017) Specific status of *Echinococcus canadensis* (Cestoda: Taeniidae) inferred from nuclear and mitochondrial gene sequences. *International Journal for Parasitology* **47**, 971–979.

https://doi.org/10.1017/S0031182018000719 Published online by Cambridge University Press