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 Laurimäe1, Liina Kinkar1, Epp Moks1, Thomas Romig2, Rihab A. Omer3, Adriano Casulli4,5, Gérald Umhang6, Guna Bagrade7, Malik Irshadullah8, Mitra Sharbatkhori9, Hossein Mirhendi10, Francisco Ponce-Gordo11, Silvia V. Soriano12, Antonio Varcasia13, Mohammad Rostami-Nejad14, Vanessa Andresiuk15 and Urmas Saarma1

1Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 51003 Tartu, Estonia; 2Institute of Zoology, Parasitology Unit, University of Hohenheim, 70599 Stuttgart, Germany; 3National University Research Institute, National University Sudan; 4World 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; 5European Union Reference Laboratory for Parasites (EURPL), Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy; 6Arsnes, Wildlife Surveillance and Eco-epidemiology Unit, National Reference Laboratory for *Echinococcus* spp., Nancy Laboratory for Rabies and Wildlife, 54220 Malzéville, France; 7Latvian State Forest Research Institute ‘Silava’, 111 Rīgas str., LV-2169 Salaspils, Latvia; 8Section of Parasitology, Department of Zoology, Aligarh Muslim University, Aligarh 202003, India; 9Laboratory Science Research Center, Golestan University of Medical Sciences, Gorgan, Iran; 10Department of Parasitology and Mycology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran; 11Department of Parasitology, Faculty of Pharmacy, Complutense University, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain; 12Department of Microbiology and Parasitology, Faculty of Medical Sciences, Comahue National University, Buenos Aires 1400, 8300 Neuquén, Argentina; 13Laboratorio de Parasitología e Malattie Parassitarie, Ospedale Didattico Veterinario Dipartimento di Medicina Veterinaria, Università degli Studi di Sassari, Via Winnipeg 2, 07100 Sassari, Italy; 14Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran and 15Laboratorio de Zoonosis Parasitarias, FCEyN, UNMdP, Funes 3350, CP: 7600 Mar del Plata, Buenos Aires, Argentina

**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., 2015a; 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).
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 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, 2018a, 2018b; Laurimäe et al., 2016) have yielded significantly deeper insight into the phylogeny and phylolgeography 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 feldisi* has, to date, remained undisputed (Hütter 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, remains 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 mouse 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; Hütter 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; Laurimäe et al., 2015b) one has to be cautious when interpreting the results. MtDNA represents the evolution of the maternal lineage, 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 view these two distinct species: G6/G7 as *E. intermedium* and G8/G10 as *E. canadensis* (Thompson, 2008; Saarma et al., 2009) or even as three species: G6/G7 as *E. intermedium*, 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...
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 $-20\, ^\circ \text{C}$ 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

---

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

![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.](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 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 northeastern 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.
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; Snabel 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](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 602951, Project HERACLES; http://www.Heracles-ip7.eu/ (to A.C.).

**Declarations of interest.** None.

**References**


Hindrisko 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), e46653.


Kinkar L et al. (2016) High resolution phylogeography of zoonotic tapeworm species. 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. (2016a) 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. (2016b) Genetic diversity and phylogeography of the elusive, but epidemiologically important Echinococcus granulosus sensu stricto genotype G3. Parasitology, https://doi.org/10.1017/S0031182018000549.


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

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

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

Laurimae T et al. (2016) Genetic diversity and phylogography 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.


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.


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.


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