Global phylogeography and genetic diversity of the zoonotic tapeworm Echinococcus granulosus sensu stricto genotype G1

Global phylogeography and genetic diversity of the zoonotic tapeworm *Echinococcus granulosus* sensu stricto genotype G1

Liina Kinkar[a], Teivi Laurimäe[a], Gerardo Acosta-Jamett[b], Vanessa Andresiuk[c], Ibrahim Balkaya[d], Adriano Casulli[e], Robin B. Gasser[f], Joke van der Giessen[g], Luis Miguel González[h], Karen L. Haag[i], Zait Houria[j], Malik Irshadullah[k], Abdul Jabbar[f], David J. Jenkins[l], Eshrat Beigom Kia[m], Maria Teresa Manfredi[n], Hossein Mirhendi[o], Selim M'rad[p], Mohammad Rostami Nejad[q], Myriam Oudni-M'rad[p], Nora Beatriz Pierangeli[r], Francisco Ponce-Gordo[s], Steffen Rehbein[t], Mitra Sharbatkhori[u], Sami Simsek[v], Silvia Viviana Soriano[r], Hein Sprong[g], Viliam Šnábel[w], Gérald Umhang[x], Antonio Varcasia[y], Urmas Saarma[a]*

[a] *Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 50410 Tartu, Estonia*

[b] *Instituto de Medicina Preventiva Veterinaria y Programa de Investigación Aplicada en Fauna Silvestre, Facultad de Ciencias Veterinarias, Universidad Austral de Chile, Valdivia, Chile.*

[c] *Laboratorio de Zoonosis Parasitarias, FCEyN, UNMdP, Funes 3350, CP: 7600 Mar del Plata, Buenos Aires, Argentina*

[d] *Department of Parasitology, Faculty of Veterinary Medicine, University of Atatürk, Erzurum, Turkey*

[e] *World Health Organization Collaborating Centre for the Epidemiology, Detection and Control of Cystic and Alveolar Echinococcosis. European Union Reference Laboratory for Parasites (EURLP). Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy*

[f] *Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia*

[g] *Centre for Infectious Disease Control Netherlands, National Institute for Public Health and Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands*

[h] *Parasitology Department, Centro Nacional de Microbiologia, Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain*

[i] *Departamento de Genética, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, RS, Brazil*

[j] *Parasitology and Mycology department, Mustapha University Hospital, 16000 Algiers, Algeria*

[k] *Section of Parasitology, Department of Zoology, Aligarh Muslim University, Aligarh 202002, India*

[l] *School of Animal and Veterinary Sciences, Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW 2678, Australia*

[m] *Department of Medical Parasitology and Mycology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*

[n] *Department of Veterinary Medicine, Università degli Studi di Milano, via Celoria 10, 20133 Milan, Italy*

[o] *Department of Medical Parasitology and Mycology, School of Medicine, Isfahan University of medical sciences, Isfahan, Iran*

[p] *Laboratory of Medical and Molecular Parasitology-Mycology (LP3M), LR 12ES08. Faculty of Pharmacy, University of Monastir, 5000 Monastir, Tunisia*

[q] *Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[r] *Department of Microbiology and Parasitology, Faculty of Medical Sciences, Comahue National University, Buenos Aires 1400, 8300 Neuquén, Argentina*

[s] *Department of Parasitology, Faculty of Pharmacy, Complutense University, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain*

[t] *Merial GmbH, Kathrinenhof Research Center, Walchenseestr. 8–12, 83101 Rohrdorf, Germany*

[u] *Laboratory Sciences Research Center, Golestan University of Medical Sciences, Gorgan, Iran*

[v] *Department of Parasitology, Faculty of Veterinary Medicine, University of Firat, 23119, Elazig, Turkey*

[w] *Institute of Parasitology, Slovak Academy of Sciences, Košice, Hlinkova 3, 040 01 Košice, Slovakia*

[x] *ANSES, Nancy Laboratory for Rabies and Wildlife, Wildlife surveillance and eco-epidemiology unit, Malzéville 54220, France*

[y] *Laboratory of Parasitology, Veterinary Teaching Hospital, Department of Veterinary Medicine, University of Sassari, Via Vienna 2 - 07100 Sassari, Italy*

*Corresponding author. Tel.: +3727375099. E-mail address: urmas.saarma@ut.ee (U. Saarma).

ABSTRACT

Echinococcus granulosus sensu stricto (s. s.) is the major cause of human cystic echinococcosis worldwide and is listed among the most severe parasitic diseases of humans. To date, numerous studies have investigated the genetic diversity and population structure of *E. granulosus* s. s. in various geographic regions. However, there has been no global study. Recently, using mitochondrial DNA, it was shown that *E. granulosus* s. s. G1 and G3 are distinct genotypes, but a larger dataset is required to confirm the distinction of these genotypes. The objectives of this study were to: (i) investigate the distinction of genotypes G1 and G3 using a large global dataset; (ii) analyse the genetic diversity and phylogeography of genotype G1 on a global scale using near-complete mitogenome sequences. For this study, 222 globally distributed *E. granulosus* s. s. samples were used, of which 212 belonged to genotype G1 and 10 to G3. Using a total sequence length of 11 682 bp, we inferred phylogenetic networks based on the whole *E. granulosus* s. s. dataset (n = 222), G1 dataset (n = 212) and G1 human samples (n = 41). In addition, the Bayesian phylogenetic and phylogeographic analyses were performed. The latter yielded several statistically significant diffusion routes of genotype G1 originating from Turkey, Tunisia and Argentina. We conclude that: (i) using a considerably larger dataset than employed previously, *E. granulosus* s. s. G1 and G3 are indeed distinct mitochondrial genotypes; (ii) the genetic diversity of *E. granulosus* s. s. G1 is high globally, with lower values in South America; (iii) the complex phylogeographic patterns emerging from the phylogenetic and geographic analyses suggest that the current distribution of genotype G1 has been shaped by early livestock diffusion events, along with intensive animal trade in relatively recent history.

86 *Keywords:*

87 Cystic echinococcosis

88 *Echinococcus granulosus*

89 Genetic variability

90 Global phylogeography

91 Mitochondrial genome

92 Livestock domestication

1. Introduction

Echinococcus granulosus sensu lato (s. l.) is the causative agent of cystic echinococcosis (CE), which is one of the most important zoonoses worldwide and a significant global public health concern (e.g., Eckert et al., 2001; Alvarez Rojas et al., 2014; Marcinkute et al., 2015; Budke et al., 2017). CE is listed amongst the most severe parasitic diseases in humans, ranking second in the list of food-borne parasites globally (FAO/WHO, 2014) and representing one of the 17 neglected tropical diseases prioritised by the World Health Organization (WHO, 2015). The life cycle of the parasite involves mainly dogs and wild carnivores as definitive hosts and a wide range of domestic and wild mammals, but also humans, as intermediate or accidental hosts (Eckert et al., 2001; Moks et al., 2006; Deplazes et al., 2011; Laurimaa et al., 2015a).

Echinococcus granulosus s. l. exhibits considerable variation in terms of morphology, host range, infectivity to humans, pathogenicity and other aspects (e.g., Eckert et al., 2001; Thompson, 2008; Gholami et al., 2011; Romig et al., 2015). Molecular studies have identified and characterised a number of genotypes/species within the *E. granulosus* s. l. complex (Bowles et al., 1992, 1994; Thompson and McManus, 2002; Lavikainen et al., 2003; Thompson, 2008; Saarma et al., 2009; Knapp et al., 2011), which are relatively closely related to other species within the genus *Echinococcus* (Knapp et al., 2015). The accurate identification and differentiation of genotypes has important epidemiological implications and informs about the zoonotic potential of particular genotypes. Earlier, the complex was considered to consist of genotypes G1-G8, G10 and *E. felidis* (see Bowles et al., 1992, 1994; Lavikainen et al., 2003; Hüttner et al., 2008), however G2 is no longer considered a valid genotype (Kinkar et al., 2017). Currently, the genotypes regarded as distinct species are *E. granulosus* sensu stricto (s. s.; genotypes G1 and G3; Kinkar et al., 2017), *E. equinus* (G4), *E. ortleppi* (G5) (Thompson and McManus, 2002), whereas the species status of genotypes G6-G10 remains contentious (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 new genotype was discovered in Ethiopia, but its status is not yet clear (Wassermann et al., 2016).

Echinococcus granulosus s. s. (genotypes G1 and G3) is widespread globally, with highly endemic foci in South America, the Mediterranean basin and Central Asia, and particularly affects rural livestock-raising areas (Dakkak et al., 2010; Hajialilo et al., 2012; Rostami et al., 2015; Zhang et al., 2015; Cucher et al., 2016). Some of the main factors contributing to the persistence of CE include the frequent illegal and home slaughtering of animals for food, feeding raw offal to dogs, low public awareness of the disease, large populations of stray dogs and poor hygiene conditions (Eckert et al., 2001; Torgerson and Budke, 2003; Varcasia et al., 2011; Possenti et al., 2016). According to a recent estimate by Alvarez Rojas et al. (2014), *E. granulosus* s. s. is also the most frequently implicated causative agent of CE of humans (88% of cases) worldwide, and thus deserves particular attention.

To date, numerous studies have explored the genetic diversity and population structure of *E. granulosus* s. s. in various geographic regions (Nakao et al., 2010; Casulli et al., 2012; Rostami Nejad et al., 2012; Yanagida et al., 2012; Andresiuk et al., 2013; Yan et al., 2013; Boufana et al., 2014, 2015; Romig et al., 2015; Kinkar et al., 2016; Laurimäe et al., 2016; Hassan et al., 2017). However, there has been no global study. In addition, the analytical power has been low in most studies as the analyses have been based largely on short sequences of mitochondrial DNA (mtDNA), most often on a single gene, e.g., the cytochrome c oxidase subunit 1 gene (*cox1*; 1609 bp; Yanagida et al., 2012; Alvarez Rojas et al., 2016; Alvarez Rojas et al., 2017) or partial sequence of the *cox1* or *nad1* (e.g., Casulli et al., 2012; Andresiuk et al., 2013). Few studies used considerably longer mtDNA sequences (~8270 bp; Kinkar et al., 2016; Laurimäe et al., 2016) and demonstrated significantly better phylogenetic resolution. Due to the variable sequence lengths used thus far (a few hundred bp up to ~8270 bp), the results from different studies and geographic regions are not directly comparable. Therefore, an analysis of near-complete mitogenome sequences

142 in a large geographical scale is required to gain better insight into the global patterns of diversity
143 and phylogeography. Furthermore, the sequences of relatively short mtDNA regions most
144 commonly used to date cannot unequivocally differentiate genotypes G1-G3 due to limited
145 phylogenetic signal (e.g., Casulli et al., 2012; Andresiuk et al., 2013; Romig et al., 2015). Thus,
146 although short mtDNA sequences have been widely used in phylogeographic studies and to
147 develop methods for identifying genotypes (e. g. Boubaker et al., 2013; Laurimaa et al., 2015b), one
148 has to be cautious when interpreting the results based on short mtDNA sequences.

149 By contrast, using near-complete mitogenome sequences (11 443 bp), Kinkar et al. (2017)
150 provided evidence that G1 and G3 are distinct mitochondrial genotypes. As a relatively small
151 number of samples was used in Kinkar et al. (2017), a larger sample size would be preferable to
152 confirm the distinction of the two genotypes (G1 and G3). Therefore, in the present study, we (i)
153 investigated the distinction of the *E. granulosus* s. s. genotypes G1 and G3 using a large global
154 dataset (n = 222), and (ii) analysed the genetic diversity and phylogeography of genotype G1 on a
155 world-wide scale using near-complete mitochondrial genome sequences.

156

157 **2. Materials and methods**

158 *2.1 Parasite material*

159 We sequenced 221 *E. granulosus* s. s. samples and included an additional sequence from
160 Genbank (AB786664; genotype G1 from China; Nakao et al., 2013). Of the 221 samples, 114 were
161 newly sequenced, whereas the rest were from Kinkar et al. (2016 and 2017) and Laurimäe et al.
162 (2016) (Tables S1 and S2). However, additional mtDNA loci were sequenced for these samples in
163 this study. The samples were obtained during routine meat inspections or from hospital cases and
164 were ethanol-preserved at -20°C until further use.

165

166 *2.2 DNA extraction, PCR amplification, sequencing and assembly*

167 Total genomic DNA was extracted from protoscoleces, cyst membranes or adult worms of
168 *E. granulosus* using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim,
169 Germany), following the manufacturer's protocols. For PCR amplification we used 12 primer pairs
170 described in Kinkar et al. (2017). Sequencing was performed using the same primers as for the
171 initial PCR amplification. Cycle parameters for PCR and sequencing were as described in Kinkar et
172 al. (2016). Sequences were assembled using the program CodonCode v6.0.2 and manually curated
173 in BioEdit v7.2.5 (Hall, 1999). All G1 sequences were deposited in the GenBank database under
174 accession nos. XXXX-XXXX.

175

176 *2.3 Phylogenetic analyses*

177 Phylogenetic networks were calculated for three mtDNA sequence datasets: (1) all samples
178 of *E. granulosus* s. s. (n = 222), (2) sequences representing genotype G1 only (n = 212) and (3)
179 sequences representing genotype G1 from humans (n = 41) using Network v4.6.1.5 (Bandelt et al.,
180 1999); http://www.fluxusengineering.com, Fluxus Technology Ltd., 2004. Networks were
181 constructed considering both indels and point mutations.

182 The Bayesian phylogenetic analysis for the whole dataset (n = 222 samples) was performed
183 in the program BEAST 1.8.4 (Drummond et al., 2012) using BEAUti v1.8.4 to generate the initial
184 xml file for BEAST. The general time-reversible nucleotide-substitution model with a proportion of
185 invariable sites and gamma distributed rate variation (GTR+*I*+*G;* Tavaré, 1986; Gu et al., 1995)
186 was determined as the best-fit model of sequence evolution using the program PartitionFinder 2.1.1
187 (Guindon et al., 2010; Lanfear et al., 2012, 2016). Exponential growth coalescent prior (Griffiths
188 and Tavaré, 1994) was chosen for the tree, and a strict molecular clock was assumed owing to the
189 intraspecific nature of the data (Drummond and Bouckaert, 2015). The posterior distribution of

parameters was estimated by Markov Chain Monte Carlo (MCMC) sampling. MCMC chains were run for 10 million states, sampled every 1000 states with 10% burn-in. Log files were analysed using the program Tracer v1.6 (Rambaut et al., 2014). The tree was produced using TreeAnnotator v1.8.4 and displayed in FigTree v.1.4.3 (Rambaut, 2014).

2.4 Population indices

The population diversity indices, such as the number of haplotypes, haplotype diversity and nucleotide diversity, were calculated using the program DnaSP v5.10.01 (Librado and Rozas, 2009). Neutrality indices Tajima's D (Tajima, 1989) and Fu's Fs (Fu, 1997) and the pairwise fixation index (Fst) were calculated using the Arlequin 3.5.2.2 software package (Excoffier et al., 2005). Indices were calculated for four different datasets representing genotype G1: (a) all sequences (n = 212); (b) the three most numerous host species in this study (cattle, sheep and human), (c) five regions (the Americas, Africa, Asia/Australia, Europe and the Middle East), and (d) eight countries for which the sample size exceeded 10: Algeria, Argentina, Brazil, Iran, Italy (comprising continental Italy and Sardinia), Spain, Tunisia and Turkey. In addition, the pairwise fixation index was calculated between genotypes G1 and G3.

2.5 Bayesian phylogeographic analysis

The phylogeographic diffusion patterns of genotype G1 were analysed using a Bayesian discrete phylogeographic approach (Lemey et al., 2009). This approach estimates ancestral locations from the set of sampled locations and annotates the discrete location states to tree nodes (Lemey et al., 2009; Faria et al., 2011). The standard Markov model is extended using a Bayesian Stochastic Search Variable Selection (BSSVS) procedure, which offers a Bayesian Factor (BF) test to identify the most parsimonious description of the phylogeographic diffusion process (Lemey et al., 2009).

Specifically, the intial xml file generated in BEAUti in the Bayesian phylogenetic analysis (see section *2.3*) was edited according to the 'Discrete phylogeographic analysis' tutorial available on the Beast website (http://beast.bio.ed.ac.uk/tutorials). The analysis was performed in BEAST 1.8.4 (Drummond et al., 2012) using the BEAGLE library (Ayres et al., 2011). MCMC chains were run for 50 million states, sampled every 5000 states with 10% burn-in. The effective sampling size (ESS) of estimates was assessed using Tracer v1.6 (Rambaut et al., 2014), and the tree was produced using TreeAnnotator v1.8.4 and displayed in FigTree v.1.4.3 (Rambaut, 2014). The program SpreaD3 v0.9.6 (Bielejec et al., 2016) was used to visualize the output from the Bayesian phylogeographic analysis and to calculate the Bayes Factor supports. Three independent runs were conducted and geographic links that yielded BF > 10 in all three runs were displayed.

3. Results

Near-complete mitogenome sequences representing *E. granulosus* s. s. samples (n = 221) were produced and aligned (length of alignment 11 682 bp). Most sequences were 11 675 bp in length, but some varied from 11 674 bp to 11 678 bp. An additional sequence from GenBank (see section *2.1*) was included, totalling 222 sequences in analysis.

3.1. The phylogenetic network of E. granulosus s. s.

The 222 sequences divided into two haplogroups, separated by 37 mutations (Fig. 3). The largest haplogroup included 212 sequences representing genotype G1, whereas the other haplogroup included 10 samples representing genotype G3. The 212 G1 samples were divided into 171 different haplotypes (Fig. 3). The origin and host species of the G1 samples are shown in Figs. 1 and 2 and Tables 1 and S3. To the best of our knowledge, all human G1 samples used in the analysis were autochthonic cases of CE, except for a Finnish sample, which originated from an

238 Algerian patient who was living in Finland. Therefore the origin of the infection is most likely
239 Algeria.
240
241 *3.2 Bayesian phylogenetic analysis*
242 The Bayesian phylogenetic analysis divided *E. granulosus* s. s. samples into two well-
243 supported clades, corresponding to genotypes G1 and G3 (posterior probability value = 1.00; Fig. 4;
244 Fig. S1). The intraspecific phylogeny of G1 yielded clades with varying support values, of which
245 several clades were well resolved (posterior probability values = 1.00).
246
247 *3.3. The phylogenetic network for genotype G1*
248 The phylogenetic network for genotype G1 was highly divergent (Fig. 5). Among the 171
249 haplotypes, 147 were represented by a single sample, 18 haplotypes included two samples, 5
250 haplotypes (IRA1, BRA1, TUR1, TUR3, TUN5) included 3 samples and one haplotype (ARB1)
251 included 14 samples. The average number of mutational steps between different G1 haplotypes was
252 16 and the maximum 32 (e.g., between TUR12 and ALB2).
253 Multiple haplogroups (monophyletic groups) could be distinguished. Seven such haplogroups
254 (named A-G, respectively) corresponded to the well-supported clusters in the Bayesian
255 phylogenetic tree (posterior probability values = 1.00; see Figs. 4 and 5; see also section *3.2*). Out
256 of the nine haplogroups in grey (Fig. 5), seven were well-supported on the phylogenetic tree
257 (posterior probability values = 1.00; Fig. 5).
258 In some of the monophyletic clusters in the network, haplotypes clustered together according to
259 geographic origin. For example, three monophyletic groups represented haplotypes only from
260 Tunisia (TUN25, TUN11 and TUN1; TUN26 and TUN6; TUN13, TUN3 and TUN18). Another
261 haplogroup (D) was of Middle-East origin, comprising samples from Turkey (TUR8, TUR21,

TUR18, TUR19) and Iran (IRA11). In addition, one group was of African origin and included samples from Tunisia (TUN5, TUN7) and Algeria (ALG9) and another group was from South-America, including haplotypes from Brazil and Argentina (BRA4, ARG2, BRA6). In other monophyletic groups, samples from Eurasia clustered together, some of which comprised haplotypes that were geographically distant from each other, such as an Indian-Iranian group (IND1 and IRA16) and a Turkish-Spanish-Iranian group F (TUR12, TUR24, TUR27, TUR4, TUR9, IRA12 and SPA1). Haplogroup G from Eurasia represented haplotypes from Turkey (TUR32, TUR22, TUR11, TUR36, TUR13, TUR28, TUR26, TUR10, TUR31, TUR33, TUR17, TUR7), Iran (IRA1, IRA13, IRA8, IRA18, IRA7, IRA17, IRA4, IRA9), Albania (ALB1, ALB2), Moldova (MOL2) and Romania (ROM1), and haplogroup C represented haplotypes from Iran (IRA19, IRA6 and IRA5), Moldova (MOL3), Mongolia (MON1) and Romania (ROM2).

The geographically most distant haplotypes that clustered together into haplogroups originated from different continents, including two haplotypes from Australia (AUS1 and AUS2) and a haplotype originating from Algeria (ALG4). However, haplotype AUS3 from Australia clustered together with 12 haplotypes from Africa (TUN8, TUN30, ALG6, TUN12, ALG10, TUN14, TUN23, TUN9, ALG1, TUN10, ALG3 and ALG11) and the haplotypes from Europe (SPA7, SPA4 and FIN1; A). In addition, five haplotypes from Africa (ALG2, TUN15, MOR1, TUN27, ALG8) clustered with haplotype ARG8 from Argentina, and haplotypes ITA7, ITA6, ITA8, and TUN2 from Italy and Tunisia also clustered together.

No host-specific pattern was identified, as the majority of monophyletic clusters included samples from different host species. The most numerous host species in this study, cattle and sheep, were genetically closely related and some haplotypes (TUR17, TUN14 and ARB1) included samples from both hosts. As expected, the haplotypes representing 41 samples from humans did not cluster together and were in different haplogroups, together with samples from other hosts. Haplotype TUN5 from Tunisia represented three samples, one from sheep and two from human and

haplotype TUN15 also from Tunisia represented two samples, one from sheep the other from human.

3.4 The phylogenetic network of human G1 samples

The 41 genotype G1 samples from humans represented 37 distinct haplotypes (Fig. 6). Haplotypes from Tunisia and Algeria were frequently closely related (e.g., TUN22 and ALG12), but some were genetically very distant from one another (e.g. ALG7 and TUN27; separated by 30 mutations). Haplotype ALG1 from Algeria was most closely related to haplotype FIN1; FIN1 was from an Algerian CE patient who was living in Finland. Haplotype MON1 representing two samples from Mongolia was within a monophyletic cluster with haplotype ROM2 from Romania and haplotype IRA3 from Iran with haplotype TUN21 from Tunisia.

3.5 Diversity and neutrality indices

The overall haplotype diversity index for genotype G1 was very high (Hd = 0.994), while the nucleotide diversity was low (π = 0.00133; Table 2). The most numerous host species in this study – cattle, sheep and human – were represented by high haplotype diversity indices (0.987 to 0.995), whereas nucleotide diversities ranged from 0.00128 to 0.00138. The haplotype diversity indices for genotype G1 from the five geographical regions were also high, ranging from 0.926 to 0.994, whereas the nucleotide diversities varied from 0.00083 to 0.00136, with samples from America having the lowest values. Of the countries represented in the present analysis, Argentina had the lowest values of haplotype and nucleotide diversities (Hd = 0.832 and π = 0.00057), whilst the corresponding values for other countries were higher (ranging from 0.956 to 1.000 and π ranging from 0.115 to 0.00143).

310 Neutrality indices Tajima's D and Fu's Fs were negative and statistically highly significant
311 for genotype G1 (D = -2.77, Fs = -23.80; Table 2). Neutrality indices were similar among host
312 species and in the majority of the regions (Africa, the Americas, Europe and the Middle East).
313 However, neutrality indices were lower and insignificant for Asia and Australia. Among the
314 countries included, both neutrality indices were negative and statistically significant for Algeria,
315 Argentina, Tunisia and Turkey, while only Tajima's D (-2.03) was significant for Iran. The
316 neutrality indices calculated for Brazil, Italy and Spain were all negative, and statistically
317 insignificant.

318

319 *3.6. Population differentiation*

320 The Fst value between genotypes G1 and G3 was very high (0.711; $p < 0.00001$). By contrast,
321 low Fst values were observed between cattle, sheep and human samples of G1 (Fst < 0.05; Table 3)
322 and between most of the regions of G1 in this study (Africa, Asia and Australia, Europe and the
323 Middle East), ranging from 0.022 to 0.068 (Table 4). However, higher Fst values (ranging from
324 0.186 to 0.216) were detected between the Americas and the other regions. Among countries, the
325 highest Fst values were seen between Argentina and the Eurasian (Iran, Italy, Spain and Turkey)
326 and African countries (Algeria and Tunisia), ranging from 0.269 to 0.359, while the value was
327 slightly lower between Argentina and Brazil (0.124; Table 5). The Fst values between the remaining
328 countries were mostly less than 0.100. Statistically insignificant values were observed between
329 Europe and Asia-Australia (Table 4) and between Algeria and Tunisia (Table 5).

330

331 *3.7. Bayesian phylogeographic analysis*

332 The Bayesian discrete phylogeographic analysis yielded 18 statistically significant spatial
333 diffusion routes for genotype G1, of which 11 had a BF value of 10 to 100, whereas the BF value

was very high (>100) for seven routes (Fig. 7). A total of seven routes originated from Turkey, two of which had very high statistical support (BF > 100; between Turkey and Iran and Turkey and Greece); six originated from Tunisia, three of which had BF values >100 (between Tunisia and Italy, Tunisia - Algeria and Tunisia - Argentina). Argentina was the ancestral location to Brazil (BF > 100), Mexico and Chile, while Iran was ancestral to India. Algeria was identified as the origin of the sample from a human from Finland.

4. Discussion

The results of this study based on 222 near-complete *E. granulosus* s. s. mitogenome sequences from a worldwide distribution confirmed that genotypes G1 and G3 are indeed distinct genotypes, as reported recently by Kinkar et al. (2017) with a significantly smaller sample size (n = 23). The analysis of the much larger dataset used in the present study also positioned genotypes G1 and G3 into distinct haplogroups, separated by 37 mutations (Fig. 3). This distinction was also well supported by the Bayesian phylogenetic analysis (Fig. 4) and by the high Fst value (0.7nn; $p < 0.00001$) between genotypes G1 and G3. As genotypes G1 and G3 represent distinct mitochondrial lineages and G1 is more widespread with a larger spectrum of hosts, it is possible that there are epidemiological differences between these genotypes. Although this proposal has not yet been explored, the use of up-to-date molecular methods to identify and distinguish these genotypes will be the prerequisite to test this hypothesis. However, sequencing a large portion of the mitochondrial genome is often not feasible in most laboratories, such that establishing a set of diagnostic nucleotides to confidently assign samples to genotypes G1 and G3 is needed (ongoing project).

The results of the present study demonstrated an extremely high global haplotype diversity within genotype G1 (Fig. 5); the 212 samples analysed represented a total of 171 haplotypes (overall haplotype diversity 0.994; Table 2). Haplotype diversities within genotype G1 were high

for different host species, regions and countries (with values being mostly between 0.970 and 1.000; Table 2), whereas Fst values were low (mostly < 0.1; Tables 3-5), pointing to a high genetic diversity and low genetic differentiation between G1 subpopulations globally, possibly due to rapid radiation. However, the South- and Central-American samples (since only one sample was from Mexico, we use henceforth South America) showed slightly lower values of haplotype diversities (particularly Argentina; Hd = 0.832; Table 2) and higher values of Fst (ranging from 0.186 to 0.216 between the Americas and the other regions; Table 4), indicating lower genetic diversity and moderate genetic differentiation of samples from South America compared with those from Africa and Eurasia. This finding is also supported by the phylogenetic network wherein the South-American samples formed a haplogroup (B) with a dominant central haplotype (Fig. 5), suggesting a bottleneck event in the past, while significant negative values of neutrality indices (D = -2.201, Fs = -13.284; Table 2) indicated a population expansion in South America. A possible explanation for this observation is the relatively recent arrival to and sudden expansion of domestic animals (cattle and sheep) in South America during the 15th and 16th Centuries (Rodero et al., 1992) compared with the domestication history in Africa and Eurasia, extending thousands of years BC (Zeder, 2008; Lv et al., 2015). However, as Argentina contributed more to the lower Hd value for South America, another possible reason could be that a relatively large number of the Argentinian samples (24 of 31) originated from the same geographical area (the Buenos Aires province in Argentina). However, the samples from Turkey used in this study also originated from one area in the East (Erzurum and Elazig provinces), but yielded very high haplotype diversity (Hd = 0.991; Table 2). Therefore, the results could reflect a more recent arrival and sudden expansion of *E. granulosus* s. s. genotype G1 in South America.

In addition to the South-American haplogroup B, there were multiple other groups where samples clustered together according to their geographical origin; for example, some of the African samples (Fig. 5). However, the opposite was also observed, and numerous well-supported clusters

383 on the phylogenetic tree comprised samples from various geographic locations (e.g., in haplogroup
384 A, in which African, Australian and European samples clustered together). These observed
385 phylogeographical patterns (along with the low Fst values in Eurasia and Africa) might be the
386 consequence of an extensive livestock trade that has facilitated the dispersal of the parasite over
387 vast geographic areas. Demographic analysis also supported this hypothesis: high haplotype
388 diversity coupled to relatively low nucleotide diversity values observed in this study (Hd = 0.994, π
389 = 0.00133 for the overall population) suggest rapid demographic expansion, supported by
390 significant negative values of neutrality indices Tajima's D (-2.771) and Fu's Fs (-23.802),
391 particularly evident among subpopulations with larger sample sizes (the whole dataset, hosts,
392 African and the Middle Eastern region, Turkey; Table 2). Similar results reflecting populations
393 under expansion have been reported in previous studies in various geographic regions (e.g., Nakao
394 et al., 2010; Casulli et al., 2012; Yanagida et al., 2012; Kinkar et al., 2016; Laurimäe et al., 2016;
395 Hassan et al., 2017).

396 In this study, samples from humans did not cluster together and were frequently positioned
397 with samples from various livestock species (e.g., sheep and goat in group C; sheep and cattle in
398 groups A and F; see Figs. 4 and 5). Furthermore, some of the samples from humans were relatively
399 closely related to samples from wildlife species, such as dingo (group A) and wild boar (group E).
400 Interestingly, the aforementioned human samples were of African origin, whereas the samples from
401 dingo and wild boar were from Australia and Spain, respectively (Fig. 5). The results clearly
402 demonstrate a highly efficient transmission cycle of genotype G1 among different host species
403 (livestock, wildlife and humans) globally. This statement is further supported by the low Fst values
404 among cattle, sheep and human samples (Table 3), suggesting that no particular haplotype is more
405 virulent to humans than any other within genotype G1. However, the Fst values point to a slightly
406 higher genetic similarity between sheep and human samples (Fst = 0.025) compared with cattle and
407 human samples (Fst = 0.046). Interestingly, the majority of the *E. granulosus* s. s. cysts obtained

from cattle are reported as sterile whereas a high fertility rate is characteristic of sheep and human infections (e.g. McManus and Thompson, 2003; Andresiuk et al., 2013; Elmajdoub and Rahman, 2015; Kamelli et al., 2016). The higher genetic similarity between samples of human and sheep origin could indicate better G1 transmission between human and sheep, compared with human and cattle.

As a large portion (29 of 41) of the G1 samples from human studied here originated from Africa, it is not surprising that most of these clustered together in the phylogenetic network (see Fig. 6). The sample from a CE patient in Finland who originated from Algeria, clustered together with another human sample from Algeria and the link between Algeria and Finland was also supported by phylogeographic analysis (Fig. 7), suggesting that the individual was most likely infected in Algeria. The genetic diversity among samples from humans was very high (Hd = 0.995), almost equal to values calculated for cattle and sheep (Hd = 0.992 and 0.987, respectively; Table 2).

The Bayesian phylogeographic analysis revealed a number of statistically significant migration routes which seemed to follow the spread of livestock animals from the centre of domestication during Neolithic times (Zeder, 2008; Lv et al., 2015; Fig. 7). One ancestral location of genotype G1 was Turkey, from which several migration routes originated. The Fertile Crescent of the Middle East is considered as one of the earliest centres of livestock domestication (mainly cattle, sheep, pigs and goats) from where the animals were later distributed east- and westwards during Neolithic times (Bruford et al., 2003; Zeder, 2008; Chessa et al., 2009; Lv et al., 2015; Rannamäe et al., 2016). The phylogeographic results of this study could reflect the early spread of livestock from this region along with *E. granulosus* s. s. genotype G1. Although the possible ancestral location of *E. granulosus* s. s. in the Middle East has been suggested before (e. g. Nakao et al., 2010; Casulli et al., 2012; Yanagida et al., 2012; Kinkar et al., 2016; Hassan et al., 2017), the discrete Bayesian phylogeographic approach used here provided statistical support for this diffusion pattern. In addition, the migration routes from Tunisia to Morocco and Algeria point to a westward movement

433 of genotype G1 in North Africa which is also in accordance with the supposed direction of early
434 dispersal of domesticated animals (cattle, sheep and goat) in this area (Gifford-Gonzalez and
435 Hanotte, 2011).

436 Another location from which several diffusion routes originated was Tunisia: among others,
437 three routes showed a possible migration of genotype G1 from Tunisia to Argentina, Australia and
438 Turkey which could be linked to human/livestock migration in later history. It is possible that
439 during the colonization of Tunisia by the Ottoman Empire (founded by the Turkish) from the 16th to
440 19th Centuries, domestic animals infected with genotype G1 were transported between these
441 regions, and later to other parts of the world, which could also result in Tunisia being one of the
442 centres of radiation, together with Turkey. During the same period (the 15th and 16th Centuries),
443 sheep and other livestock were introduced to the Americas by Spanish and British colonizers.
444 However, some animals that arrived to the Americas could have had an African origin as some of
445 the livestock species (mostly pigs and goats) were taken aboard on the Canary Islands, which were
446 colonized by people from North Africa (Rodero et al., 1992; Rando et al., 1999; also discussed in
447 Alvarez Rojas et al., 2017), possibly explaining the significant diffusion route between Tunisia and
448 Argentina. The ancestral position of Argentina could indicate its possible origin for the other
449 American samples (Brazil, Chile and Mexico). The connection between Tunisia and Australia could
450 also be linked to relatively recent history: it is thought that the sources of Australian sheep could be
451 Spain and/or North Africa, as Merinos raised in North Africa arrived in Australia in the beginning
452 of the 19th Century, as discussed by Jenkins (2005).

453 In conclusion, this is the first study to explore the global patterns of genetic diversity and
454 phylogeography of *E. granulosus* s. s. using near-complete mitogenome sequences. We show that:
455 (i) using a considerably larger dataset than employed previously, *E. granulosus* s. s. genotypes G1
456 and G3 are clearly distinct mitochondrial genotypes; (ii) the genetic diversity within genotype G1 is
457 very high worldwide, with slightly lower values in South America; (iii) the observed complex

458 phylogeographic patterns emerging from the phylogenetic and -geographic analyses suggest that the
459 current distribution of *E. granulosus* s. s. genotype G1 has been shaped by the early livestock
460 diffusion events, along with intensive animal trade in the relatively recent history.
461
462 **Conflict of interest**
463 Authors declare no conflict of interest.
464
465 **Acknowledgements**

466 We would like to thank Ikhlass El Berbri and Allal Dakkak from the Institut Agronomique
467 et Vétérinaire Hassan II, and Oleg Chihai from the Institute of Zoology of the Academy of Sciences
468 of Moldova for their generous help. This work was supported by institutional research funding
469 (IUT20-32) from the Estonian Ministry of Education and Research, and the Estonian Doctoral
470 School of Ecology and Environmental Sciences. RBG's recent research has been supported by the
471 Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC)
472 of Australia, Yourgene Bioscience (Taiwan), Melbourne Water Corporation and The University of
473 Melbourne. This research was also supported by the European Community's Seventh Framework
474 Programme under the grant agreement 602051, Project HERACLES (http://www.Heracles-fp7.eu/).

475 **References**

476

477 Alvarez Rojas, C.A., Ebi, D., Gauci, C.G., Scheerlinck, J.P., Wassermann, M., Jenkins, D.J.,
478 Lightowlers, M.W., Romig, T., 2016. Microdiversity of *Echinococcus granulosus* sensu
479 stricto in Australia. Parasitology 143, 1026–1033. doi:10.1017/S0031182016000445

480 Alvarez Rojas, C.A., Ebi, D., Paredes, R., Acosta-Jamett, G., Urriola, N., Roa, J.C., Manterola, C.,
481 Cortes, S., Romig, T., Scheerlinck, J.-P., Lightowlers, M.W., 2017. High intraspecific
482 variability of *Echinococcus granulosus* sensu stricto in Chile. Parasitol. Int. 66, 112–115.
483 doi:10.1016/j.parint.2016.12.001

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

487 Andresiuk, M.V., Gordo, F.P., Saarma, M., Elissondo, M.C., Taraborelli, A., Casalongue, C.,
488 Denegri, G., Saarma, U., 2013. *Echinococcus granulosus* genotype G1 dominated in cattle
489 and sheep during 2003–2006 in Buenos Aires province, an endemic area for cystic
490 echinococcosis in Argentina. Acta Trop. 127, 136–142.
491 doi:10.1016/j.actatropica.2013.04.008

492 Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P.,
493 Ronquist, F., Swofford, D.L., Cummings, M.P., Rambaut, A., Suchard, M.A., 2011.
494 BEAGLE: an application programming interface and high-performance computing library
495 for statistical phylogenetics. Syst. Biol. 61, 170–173. doi:10.1093/sysbio/syr100

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

498 Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. Spread3:
499 Interactive visualization of spatiotemporal history and trait evolutionary processes. Mol.
500 Biol. Evol. 33, 2167–2169. doi:10.1093/molbev/msw082

501 Boubaker, G., Macchiaroli, N., Prada, L., Cucher, M.A., Rosenzvit, M.C., Ziadinov, I., Deplazes,
502 P., Saarma, U., Babba, H., Gottstein, B., Spiliotis, M., 2013. A multiplex PCR for the
503 simultaneous detection and genotyping of the *Echinococcus granulosus* complex. PLoS
504 Neglect. Trop. D. 7(1): e2017. doi:10.1371/journal.pntd.0002017
505 Boufana, B., Lahmar, S., Rebaï, W., Safta, Z.B., Jebabli, L., Ammar, A., Kachti, M., Aouadi, S.,
506 Craig, P.S., 2014. Genetic variability and haplotypes of *Echinococcus* isolates from Tunisia.
507 Trans. R. Soc. Trop. Med. Hyg. 108, 706–714. doi:10.1093/trstmh/tru138
508 Boufana, B., Lett, W.S., Lahmar, S., Buishi, I., Bodell, A.J., Varcasia, A., Casulli, A., Beeching,
509 N.J., Campbell, F., Terlizzo, M., McManus, D.P., Craig, P.S., 2015. *Echinococcus equinus*
510 and *Echinococcus granulosus* sensu stricto from the United Kingdom: genetic diversity and
511 haplotypic variation. Int. J. Parasitol. 45, 161–166. doi:10.1016/j.ijpara.2014.10.005
512 Bowles, J., Blair, D., McManus, D.P., 1992. Genetic variants within the genus *Echinococcus*
513 identified by mitochondrial DNA sequencing. Mol. Biochem. Parasitol. 54, 165–173.
514 Bowles, J., Blair, D., McManus, D., 1994. Molecular genetic characterization of the cervid strain
515 ('northern form') of *Echinococcus granulosus*. Parasitology 109, 215–221.
516 Bruford, M.W., Bradley, D.G., Luikart, G., 2003. DNA markers reveal the complexity of livestock
517 domestication. Nat. Rev. Genet. 4, 900–910. doi:10.1038/nrg1203
518 Budke, C.M., Casulli, A., Kern, P., Vuitton, D.A., 2017. Cystic and alveolar echinococcosis:
519 Successes and continuing challenges. PLoS Negl. Trop. Dis. 11, e0005477.
520 doi:10.1371/journal.pntd.0005477
521 Casulli, A., Interisano, M., Sreter, T., Chitimia, L., Kirkova, Z., La Rosa, G., Pozio, E., 2012.
522 Genetic variability of *Echinococcus granulosus* sensu stricto in Europe inferred by
523 mitochondrial DNA sequences. Infect. Genet. Evol. 12, 377–383.
524 doi:10.1016/j.meegid.2011.12.014

525 Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R., Pemberton, J.,
526 Beraldi, D., Stear, M., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M., Kazwala, R.,
527 Zhang, Y., Arranz, J.J., Ali, B., Wang, Z., Uzun, M., Dione, M., Olsaker, I., Holm, L.-E.,
528 Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M., Ajmone-Marsan, P.,
529 Bruford, M., Kantanen, J., Spencer, T., Palmarini, M., 2009. Revealing the history of sheep
530 domestication using retrovirus integrations. Science 324, 532–536.
531 doi:10.1126/science.1170587
532 Cucher, M.A., Macchiaroli, N., Baldi, G., Camicia, F., Prada, L., Maldonado, L., Avila, H.G., Fox,
533 A., Gutiérrez, A., Negro, P., Lopez, R., Jensen, O., Rosenzvit, M., Kamenetzky, L., 2016.
534 Cystic echinococcosis in South America: systematic review of species and genotypes of
535 *Echinococcus granulosus* sensu lato in humans and natural domestic hosts. Trop. Med. Int.
536 Health 21, 166–175. doi:10.1111/tmi.12647
537 Dakkak, A., 2010. Echinococcosis/hydatidosis: a severe threat in Mediterranean countries. Vet.
538 Parasitol. 174, 2–11. doi:10.1016/j.vetpar.2010.08.009
539 Deplazes, P., van Knapen, F., Schweiger, A., Overgaauw, P.A., 2011. Role of pet dogs and cats in
540 the transmission of helminthic zoonoses in Europe, with a focus on echinococcosis and
541 toxocarosis. Vet. Parasitol. 182, 41–53. doi:10.1016/j.vetpar.2011.07.014
542 Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge
543 University Press.
544 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti
545 and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. doi:10.1093/molbev/mss075
546 Eckert, J., Deplazes, P., Craig, P., Gemmell, M., Gottstein, B., Heath, D., Jenkins, D., Kamiya, M.,
547 Lightowlers, M., Meslin, F., 2001. Echinococcosis in animals: clinical aspects, diagnosis
548 and treatment. WHO/OIE Manual on echinococcosis in humans and animals: a public health
549 problem of global concern.

550 Elmajdoub, L.O., Rahman, W.A., 2015. Prevalence of Hydatid Cysts in Slaughtered Animals from
551 Different Areas of Libya. Open J. Vet. Med. 5, 1–10. doi:10.4236/ojvm.2015.51001
552 Excoffier, L., Laval, G. and Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package
553 for population genetics data analysis. Evol. Bioinform. Online 1, 47–50.
554 FAO/WHO [Food and Agriculture Organization of the United Nations/World Health Organization],
555 2014. Multicriteria-based ranking for risk management of food-borne parasites.
556 Microbiological Risk Assessment Series No. 23. Rome, 302.
557 Faria, N.R., Suchard, M.A., Rambaut, A., Lemey, P., 2011. Toward a quantitative understanding of
558 viral phylogeography. Curr. Opin. Virol. 1, 423–429. doi:10.1016/j.coviro.2011.10.003
559 Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking
560 and background selection. Genetics 147, 915–925.
561 Gifford-Gonzalez, D., Hanotte, O., 2011. Domesticating Animals in Africa: Implications of Genetic
562 and Archaeological Findings. J. World Prehist. 24, 1–23.
563 Gholami, S., Irshadullah, M., Mobedi, I., 2011. Rostellar hook morphology of larval *Echinococcus*
564 *granulosus* isolates from the Indian buffalo and Iranian sheep, cattle and camel. J.
565 Helminthol. 85, 239–245. doi:10.1017/S0022149X10000520
566 Griffiths, R.C., Tavaré, S., 1994. Sampling theory for neutral alleles in a varying environment.
567 Philos. Trans. R. Soc. Lond. B Biol. Sci. 344, 403–410.
568 Gu, X., Fu, Y.-X., Li, W.-H., 1995. Maximum likelihood estimation of the heterogeneity of
569 substitution rate among nucleotide sites. Mol. Biol. Evol. 12, 546–557.
570 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New
571 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
572 performance of PhyML 3.0. Syst. Biol. 59, 307–321. doi:10.1093/sysbio/syq010
573 Hajialilo, M., Fasihi Harandi, M., Sharbatkhori, M., Mirhendi, H., Rostami, S., 2012. Genetic
574 characterization of *Echinococcus granulosus* in camels, cattle and sheep from the south-east

of Iran indicates the presence of the G3 genotype. J. Helminthol. 86, 263–270. doi:10.1017/S0022149X11000320

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.

Hassan, Z.I., Meerkhan, A.A., Boufana, B., Hama, A.A., Ahmed, B.D., Mero, W.M.S., Orsten, S., Interisano, M., Pozio, E., Casulli, A., 2017. Two haplotype clusters of *Echinococcus granulosus* sensu stricto in northern Iraq (Kurdistan region) support the hypothesis of a parasite cradle in the Middle East. Acta Trop. 172, 201–207. doi:10.1016/j.actatropica.2017.04.028

Hüttner, M., Nakao, M., Wassermann, T., Siefert, L., Boomker, J.D.F., Dinkel, A., Sako, Y., Mackenstedt, U., Romig, T., Ito, A., 2008. Genetic characterization and phylogenetic position of *Echinococcus felidis* Ortlepp, 1937 (Cestoda : Taeniidae) from the African lion. Int. J. Parasitol. 38, 861–868. doi:10.1016/j.ijpara.2007.10.013

Jenkins, D.J., 2005. Hydatid control in Australia: where it began, what we have achieved and where to from here. Int. J. Parasitol. 35, 733–740.

Kamelli, M., Borji, H., Naghibi, A., 2016. Genetic identification of cattle hydatid cyst isolates from northeast and southwest of Iran Mehrab. Ann. Parasitol. 62, 301–305. doi:10.17420/ap6204.65

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. doi:10.1016/j.meegid.2017.04.023

Kinkar, L., Laurimäe, T., Simsek, S., Balkaya, I., Casulli, A., Manfredi, M.T., Ponce-Gordo, F., Varcasia, A., Lavikainen, A., González, L.M., Rehbein, S., van der Giessen, J., Sprong, H.,

600 Saarma, U., 2016. High-resolution phylogeography of zoonotic tapeworm *Echinococcus*
601 *granulosus* sensu stricto genotype G1 with an emphasis on its distribution in Turkey, Italy
602 and Spain. Parasitology 143, 1790–1801. doi:10.1017/S0031182016001530
603 Knapp, J., Gottstein, B., Saarma, U., Millon, L., 2015. Taxonomy, phylogeny and molecular
604 epidemiology of *Echinococcus multilocularis*: from fundamental knowledge to health
605 ecology. Vet. Parasitol. 213, 85–91. doi:10.1016/j.vetpar.2015.07.030
606 Knapp, J., Nakao, M., Yanagida, T., Okamoto, M., Saarma, U., Lavikainen, A., Ito, A., 2011.
607 Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda:
608 Taeniidae): an inference from nuclear protein-coding genes. Mol. Phylogen. Evol. 61, 628–
609 638. doi:10.1016/j.ympev.2011.07.022
610 Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of
611 partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29,
612 1695–1701. doi:10.1093/molbev/mss020
613 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new
614 methods for selecting partitioned models of evolution for molecular and morphological
615 phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. doi:10.1093/molbev/msw260
616 Laurimaa, L., Davison, J., Süld, K., Plumer, L., Oja, R., Moks, E., Keis, M., Hindrikson, M.,
617 Kinkar, L., Laurimäe, T., Abner, J., Remm, J., Anijalg, P., Saarma, U., 2015a. First report of
618 highly pathogenic *Echinococcus granulosus* genotype G1 in European Union urban
619 environment. Parasite Vector 8, 182. doi:10.1186/s13071-015-0796-3
620 Laurimaa, L., Davison, J., Plumer, L., Süld, K., Oja, R., Moks, E., Keis, M., Hindrikson, M.,
621 Kinkar, L., Laurimäe, T., Abner, J., Remm, J., Anijalg, P., Saarma, U., 2015b. Non-invasive
622 molecular diagnostics identifies *Echinococcus multilocularis* spillover to an urban area in
623 Estonia. Emerg. Infect. Dis. 21:163–164. doi:10.3201/eid2101.140136

624 Laurimäe, T., Kinkar, L., Andresiuk, V., Haag, K.L., Ponce-Gordo, F., Acosta-Jamett, G., Garate,
625 T., Gonzàlez, L.M., Saarma, U., 2016. Genetic diversity and phylogeography of highly
626 zoonotic *Echinococcus granulosus* genotype G1 in the Americas (Argentina, Brazil, Chile
627 and Mexico) based on 8279bp of mtDNA. Infect. Genet. Evol. 45, 290–296.
628 doi:10.1016/j.meegid.2016.09.015
629 Lavikainen, A., Lehtinen, M., Meri, T., Hirvelä-Koski, V., Meri, S., 2003. Molecular genetic
630 characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of
631 *Echinococcus granulosus*. Parasitology 127, 207–215.
632 Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its
633 roots. PLoS Comput Biol 5, e1000520. doi:10.1371/journal.pcbi.1000520
634 Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA
635 polymorphism data. Bioinformatics 25, 1451–1452. doi:10.1093/bioinformatics/btp187
636 Lv, F.-H., Peng, W.-F., Yang, J., Zhao, Y.-X., Li, W.-R., Liu, M.-J., Ma, Y.-H., Zhao, Q.-J., Yang,
637 G.-L., Wang, F., Li, J.-Q., Liu, Y.-G., Shen, Z.-Q., Zhao, S.-G., Hehua, Ee., Gorkhali, N.A.,
638 Vahidi, F., M, S., Muladno, M., Naqvi, A.N., Tabell, J., Iso-Touru, T., Bruford, M.W.,
639 Kantanen, J., Han, J.-L., Li, M.-H., 2015. Mitogenomic meta-analysis identifies two phases
640 of migration in the history of eastern Eurasian sheep. Mol. Biol. Evol. 32, 2515–2533.
641 doi:10.1093/molbev/msv139
642 Lymbery, A. J., Jenkins, E. J., Schurer, J. M., Thompson, R. C. A., 2015. *Echinococcus canadensis*,
643 *E. borealis*, and *E. intermedius*. What's in a name? Trends Parasitol. 31, 23–29.
644 Marcinkute, A., Šarkunas, M., Moks, E., Saarma, U., Jokelainen, P., Bagrade, G., Laivacuma, S.,
645 Strupas, K., Sokolovas, V., Deplazes, P., 2015. *Echinococcus* infections in the Baltic region.
646 Vet. Parasitol. 213, 121–131. doi:10.1016/j.vetpar.2015.07.032
647 McManus, D.P., Thompson, R.C., 2003. Molecular epidemiology of cystic echinococcosis.
648 Parasitology 127, S37–S51.

649 Moks, E., Jõgisalu, I., Saarma, U., Talvik, H., Järvis, T., Valdmann, H., 2006. Helminthologic
650 survey of the wolf (*Canis lupus*) in Estonia, with an emphasis on *Echinococcus granulosus*.
651 J. Wildl. Dis. 42, 359–365. doi:10.7589/0090-3558-42.2.359
652 Moks, E., Jõgisalu, I., Valdmann, H., Saarma, U., 2008. First report of Echinococcus granulosus G8
653 in Eurasia and a reappraisal of the phylogenetic relationships of 'genotypes' G5–G10.
654 Parasitology 135, 647–654. doi:10.1017/S0031182008004198
655 Nakao, M., Lavikainen, A., Hoberg, E., 2015. Is *Echinococcus intermedius* a valid species? Trends
656 Parasitol. 31, 342–343.
657 Nakao, M., Li, T., Han, X., Ma, X., Xiao, N., Qiu, J., Wang, H., Yanagida, T., Mamuti, W., Wen,
658 H., Moro, P.L., Giraudoux, P., Craig, P.S., Ito, A., 2010. Genetic polymorphisms of
659 *Echinococcus* tapeworms in China as determined by mitochondrial and nuclear DNA
660 sequences. Int. J. Parasitol. 40, 379–385. doi:10.1016/j.ijpara.2009.09.006
661 Nakao, M., Yanagida, T., Konyaev, S., Lavikainen, A., Odnokurtsev, V.A., Zaikov, V.A., Ito, A.,
662 2013. Mitochondrial phylogeny of the genus *Echinococcus* (Cestoda: Taeniidae) with
663 emphasis on relationships among *Echinococcus canadensis* genotypes. Parasitology 140,
664 1625–1636. doi:10.1017/S0031182013000565
665 Possenti, A., Manzano-Román, R., Sánchez-Ovejero, C., Boufana, B., La Torre, G., Siles-Lucas,
666 M., Casulli, A., 2016. Potential Risk Factors Associated with Human Cystic Echinococcosis:
667 Systematic Review and Meta-analysis. PLoS Negl. Trop. Dis. 10: e0005114.
668 doi:10.1371/journal.pntd.0005114.
669 Rambaut, A., 2014. Figtree, a graphical viewer of phylogenetic trees. Available
670 from: http://tree.bio.ed.ac.uk/software/figtree
671 Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6, Available from
672 http://tree.bio.ed.ac.uk/software/tracer/

673 Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernandez, M., Gonzalez, A.M., Pinto, F., Bandelt, H.J.,
674 1999. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands.
675 Ann. Hum. Genet. 63, 413–428.
676 Rannamäe, E., Lõugas, L., Niemi, M., Kantanen, J., Maldre, L., Kadõrova, N., Saarma, U., 2016.
677 Maternal and paternal genetic diversity of ancient sheep in Estonia from the Late Bronze
678 Age to the Post-Medieval Period, and comparison with other regions in Eurasia. Animal
679 Genetics, doi:10.1111/age.12407
680 Rodero, A., Delgado, J.V., Rodero, E., 1992. Primitive Andalusian livestock and their implications
681 in the discovery of America. Arch. Zootec. 41, 383–400.
682 Romig, T., Ebi, D., Wassermann, M., 2015. Taxonomy and molecular epidemiology of
683 *Echinococcus granulosus* sensu lato. Vet Parasitol. 213, 76–84.
684 doi:10.1016/j.vetpar.2015.07.035
685 Rostami, S., Shariat Torbaghan, S., Dabiri, S., Babaei, Z., Mohammadi, M.A., Sharbatkhori, M.,
686 Fasihi Harandi, M., 2015. Genetic characterization of *Echinococcus granulosus* from a large
687 number of formalin-fixed, paraffin-embedded tissue. Am. J. Trop. Med. Hyg. 92, 588–594.
688 doi:10.4269/ajtmh.14-0585
689 Rostami Nejad, M., Taghipour, N., Nochi, Z., Mojarad, E.N., Mohebbi, S.R., Harandi, M.F., Zali,
690 M.R., 2012. Molecular identification of animal isolates of *Echinococcus granulosus* from
691 Iran using four mitochondrial genes. J. Helminthol. 86, 485–92.
692 doi:10.1017/S0022149X1100071X.
693 Saarma, U., Jogisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., Simsek, S.,
694 Andresiuk, V., Denegri, G., Gonzalez, L.M., Ferrer, E., Garate, T., Rinaldi, L., Maravilla,
695 P., 2009. A novel phylogeny for the genus *Echinococcus*, based on nuclear data, challenges
696 relationships based on mitochondrial evidence. Parasitology 136, 317–328.
697 doi:10.1017/S0031182008005453

698 Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA
699 polymorphism. Genetics 123, 585–595.
700 Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences.
701 Lect. Math. Life Sci. 17, 57–86.
702 Thompson, R., 2008. The taxonomy, phylogeny and transmission of *Echinococcus*. Exp. Parasitol.
703 119, 439–446. doi:10.1016/j.exppara.2008.04.016
704 Thompson, R.A., McManus, D.P., 2002. Towards a taxonomic revision of the genus *Echinococcus*.
705 Trends Parasitol. 18, 452–457. doi:10.1016/S1471-4922(02)02358-9
706 Torgerson, P.R., Budke, C.M., 2003. Echinococcosis – an international public health challenge.
707 Res. Vet. Sci. 74, 191–202.
708 Varcasia, A., Tanda, B., Giobbe, M., Solinas, C., Pipia, A.P., Malgor, R., Carmona, C., Garippa, G.,
709 Scala, A., 2011. Cystic Echinococcosis in Sardinia: Farmers' knowledge and dog infection in
710 sheep farms. Vet. Parasitol. 181, 335-340. doi:10.1016/j.vetpar.2011.05.006
711 Wassermann, M., Woldeyes, D., Gerbi, B.M., Ebi, D., Zeyhle, E., Mackenstedt, U., Petros, B.,
712 Tilahun, G., Kern, P., Romig, T., 2016. A novel zoonotic genotype related to *Echinococcus*
713 *granulosus* sensu stricto from southern Ethiopia. Int. J. Parasitol. 46, 663–668.
714 doi:10.1016/j.ijpara.2016.04.005
715 WHO [World Health Organization], 2015. Investing to overcome the global impact of neglected
716 tropical diseases: Third WHO report on neglected tropical diseases.
717 Yan, N., Nie, H.-M., Jiang, Z.-R., Yang, A.-G., Deng, S.-J., Guo, L., Yu, H., Yan, Y.-B., Tsering,
718 D., Kong, W.-S., 2013. Genetic variability of *Echinococcus granulosus* from the Tibetan
719 plateau inferred by mitochondrial DNA sequences. Vet. Parasitol. 196, 179–183.
720 doi:10.1016/j.vetpar.2013.02.010
721 Yanagida, T., Mohammadzadeh, T., Kamhawi, S., Nakao, M., Sadjjadi, S.M., Hijjawi, N., Abdel-
722 Hafez, S.K., Sako, Y., Okamoto, M., Ito, A., 2012. Genetic polymorphisms of *Echinococcus*

723 *granulosus* sensu stricto in the Middle East. Parasitol. Int. 61, 599–603.
724 doi:10.1016/j.parint.2012.05.014
725 Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins,
726 diffusion, and impact. Proc. Natl. Acad. Sci. 105, 11597–11604.
727 doi:10.1073/pnas.0801317105
728 Zhang, W., Zhang, Z., Wu, W., Shi, B., Li, J., Zhou, X., Wen, H., McManus, D.P., 2015.
729 Epidemiology and control of echinococcosis in central Asia, with particular reference to the
730 People's Republic of China. Acta Trop. 141, 235–243.
731 doi:10.1016/j.actatropica.2014.03.014

732 **Legends to Figures**

733

734 **Fig. 1.** Geographic locations of *Echinococcus granulosus* sensu stricto genotype G1 samples (n =
735 212) analysed in this study.

736

737 **Fig. 2.** Geographic locations of *Echinococcus granulosus* sensu stricto genotype G1 samples from
738 humans (n = 41) used in this study.

739

740 **Fig. 3.** Phylogenetic network of *Echinococcus granulosus* sensu stricto samples based on 11 682 bp
741 of mtDNA. Small black circles are median vectors (i.e. hypothetical haplotypes: haplotypes not
742 sampled or extinct). The larger haplogroup (n = 212) corresponds to the mitochondrial genotype G1
743 and the smaller haplogroup (n = 10) to G3. The small circles and triangles in the haplogroups
744 represent haplotypes. The number on the line connecting the haplogroups indicates the mutational
745 steps between genotypes G1 and G3.

746

747 **Fig. 4.** Bayesian phylogenetic tree inferred from 222 *Echinococcus granulosus* sensu stricto
748 samples. The larger clade (n = 212) corresponds to the mitochondrial genotype G1 and the smaller
749 (n = 10) to G3. Posterior probability values >0.95 are indicated at the nodes. The asterisks indicate
750 haplotypes obtained from humans. Seven clades depicted in blue, yellow, red, green, pink, purple,
751 orange and named A–G, respectively, illustrate clades that received the posterior probability value
752 >0.95 and in which the sample size was equal or higher than 5. Note that the lengths of two
753 branches are reduced (dashed line); for the figure with actual branch lengths, see Fig. S1.

754

755 **Fig. 5.** Phylogenetic network of *Echinococcus granulosus* sensu stricto G1 samples based on 11 682
756 bp of mtDNA. Circles represent haplotypes obtained from livestock and wild animals, triangles
757 represent haplotypes of human origin. Haplotype colours represent different geographical regions:

purple – Africa, green – America, orange – Asia and Australia, blue – Europe, dark red – the Middle East (please note that colours indicated on the right corner of the figure refer to geographic locations of haplotypes, not haplogroups). Haplotype names represent their geographical origin: ALB – Albania, ALG – Algeria, ARG – Argentina, AUS – Australia, BRA – Brazil, CHI – Chile, CHN – China, FIN – Finland (patient from Algeria), FRA – France, GRE – Greece, IND – India, IRA – Iran, ITA – Italy, KAZ – Kazakhstan, MEX – Mexico, MOL – Moldova, MON – Mongolia, MOR – Morocco, ROM – Romania, SPA – Spain, TUN – Tunisia, TUR – Turkey. Host species are indicated with letters inside the haplotypes (C – cattle, S – sheep, H – human, P – pig, G – goat, D – dingo, W – wild boar, B – buffalo). The small number inside haplotypes indicates the frequency of the haplotype. Numbers on the lines represent the number of mutations (single mutations are not marked with a number).

Fig. 6. Phylogenetic network of *Echinococcus granulosus* sensu stricto G1 human samples based on 11 682 bp of mtDNA. Triangles represent haplotypes. Haplotype colours represent different geographical regions: purple – Africa, orange – Asia, blue – Europe and dark red – the Middle East. Haplotype names represent different geographical origins: ALG – Algeria, CHN – China, FIN – Finland (Algerian patient), IRA – Iran, ITA – Italy, KAZ – Kazakhstan, MON – Mongolia, ROM – Romania, SPA – Spain, TUN – Tunisia. The number inside the triangles indicate the frequency of the haplotype. Numbers on the lines represent the number of mutations (single mutations are not marked with a number).

Fig. 7. Statistically significant diffusion routes inferred from the Bayesian phylogeographic analysis based on 212 *Echinococcus granulosus* sensu stricto genotype G1 samples (11 682 bp of mtDNA). Black lines represent significant links (BF > 10), whereas black lines with red outlines represent highly significant links (BF > 100).

Fig. S1. Bayesian phylogenetic tree inferred from 222 *Echinococcus granulosus* sensu stricto samples. The larger clade (n = 212) corresponds to the mitochondrial genotype G1 and the smaller (n = 10) to G3. Posterior probability values >0.95 are indicated at the nodes. The asterisks indicate haplotypes obtained from humans. Seven clades depicted in blue, yellow, red, green, pink, purple, orange and named A-G, respectively, illustrate clades that received the posterior probability value >0.95 and in which the sample size was equal or higher than 5. This is essentially the same as Fig. 4, but with actual branch lengths.

Table 1

Host data for 212 *Echinococcus granulosus* sensu stricto G1 isolates analysed in this study.

Origin	Sheep	Cattle	Human	Goat	Swine	Wild boar	Dingo	Buffalo	Total
1. Turkey	28	14							42
2. Tunisia	17	4	17						38
3. Iran	16	3	2	2					23
4. Argentina	16	14			1				31
5. Brazil		14							14
6. Spain	6		2	3	1	1			13
7. Algeria			12						12
8. Italy	6	2	1	1					10
9. Chile		6							6
10. Australia							3		3
11. Greece	3								3
12. Mongolia			3						3
13. Moldova	2	1							3
14. Romania		1	1						2
15. Albania	2								2
16. Finland (Alg)			1						1
17. France		1							1
18. Kazakhstan			1						1
19. China			1[a]						1
20. India								1	1
21. Mexico					1				1
22. Morocco		1							1
Total	96	61	41	6	3	1	3	1	**212**

[a] Sequence was obtained from GenBank (AB786664; Nakao et al., 2013).

819 **Table 2**

820 Diversity and neutrality indices for *Echinococcus granulosus* sensu stricto G1 samples based on 11
821 682 bp mtDNA sequences.

822

	Diversity				Neutrality	
	n	Hn	Hd ± S.D.	π ± S.D.	*D*	Fs
Total	212	171	0.994±0.002	0.00133±0.00004	-2.77109[a]	-23.80242[b]
Host						
Cattle	61	52	0.992±0.005	0.00138±0.00007	-2.56626[a]	-24.20117[a]
Sheep	96	74	0.987±0.006	0.00128±0.00005	-2.65309[a]	-24.12005[a]
Human	41	37	0.995±0.007	0.00130±0.00008	-2.61502[a]	-18.96890[a]
Region						
Africa	51	43	0.993±0.006	0.00136±0.00007	-2.50107[a]	-20.46636[a]
Asia & Australia	9	8	0.972±0.064	0.00099±0.00014	-1.16779	-0.73526
Europe	35	31	0.993±0.009	0.00136±0.00008	-2.40214[a]	-12.30737[b]
America	52	34	0.926±0.031	0.00083±0.00009	-2.20130[b]	-13.28433[b]
Middle East	65	55	0.994±0.004	0.00132±0.00007	-2.60935[a]	-24.21632[a]
Country						
Algeria	12	12	1.000±0.034	0.00143±0.00014	-1.98613[b]	-3.17349[c]
Argentina	31	19	0.832±0.070	0.00057±0.00014	-2.38545[a]	-5.29367[c]
Brazil	14	12	0.956±0.045	0.00115±0.00012	-1.31585	-1.67741
Iran	23	19	0.980±0.020	0.00120±0.00011	-2.03201[b]	-4.14849
Italy	10	9	0.978±0.054	0.00126±0.00014	-1.32335	-0.77495
Tunisia	38	30	0.987±0.009	0.00132±0.00008	-2.25318[b]	-8.60682[c]
Turkey	42	36	0.991±0.008	0.00137±0.00009	-2.48392[b]	-15.01834[a]
Spain	13	11	0.974±0.039	0.00124±0.00012	-1.61222	-0.92526

823 Abbreviations: number of isolates examined (n), number of haplotypes (Hn), haplotype diversity (Hd), nucleotide
824 diversity (π), Tajima's D (*D*), Fu's Fs (Fs), and standard deviation (S.D.).

825

826 [a] Highly significant p value ($p \leq 0.001$).

827 [b] Highly significant p value ($p < 0.01$).

828 [c] Significant p value ($p < 0.05$).

Table 3

Pairwise fixation index (Fst) values between *Echinococcus granulosus* sensu stricto genotype G1 hosts based on 11 682 bp of mtDNA.

	Cattle	Sheep	Human
Cattle	-		
Sheep	0.01171[a]	-	
Human	0.04620[a]	0.02477[a]	-

[a] Significant p value ($p < 0.05$).

Table 4

Pairwise fixation index (Fst) values between *Echinococcus granulosus* sensu stricto genotype G1 regions based on 11 682 bp of mtDNA.

	Africa	Asia & Aus	Europe	The Americas	Middle East
Africa	-				
Asia & Australia	0.02603[a]	-			
Europe	0.02844[a]	0.02243	-		
America	0.18581[a]	0.21568[a]	0.19073[a]	-	
Middle East	0.06808[a]	0.04671[a]	0.02998[a]	0.20726[a]	-

[a] Significant p value ($p < 0.05$).

Table 5

Pairwise fixation index (Fst) values between *Echinococcus granulosus* sensu stricto genotype G1 countries based on 11 682 bp of mtDNA.

	Algeria	Argentina	Brazil	Iran	Italy	Tunisia	Turkey	Spain
Algeria	-							
Argentina	0.32670[a]	-						
Brazil	0.08251[a]	0.12434[a]	-					
Iran	0.08940[a]	0.33548[a]	0.12860[a]	-				
Italy	0.04580[a]	0.35853[a]	0.10146[a]	0.10366[a]	-			
Tunisia	0.00410	0.26940[a]	0.07992[a]	0.08233[a]	0.05166[a]	-		
Turkey	0.06763[a]	0.27984[a]	0.09946[a]	0.01280[a]	0.07387[a]	0.06480[a]	-	
Spain	0.02989[a]	0.34402[a]	0.10144[a]	0.08996[a]	0.06351[a]	0.04593[a]	0.06133[a]	-

[a] Significant p value ($p < 0.05$).

Table S1
The list of G1 samples from the Americas partially published previously in Laurimäe et al. (2016) and Kinkar et al. (2017).

Lab code in Tartu	Haplotype in this study (11 682 bp)	Haplotype in Laurimäe et al. (2016; 8279 bp)	Haplotype in Kinkar et al. (2017; 11 443 bp)
A1	**ARB1**	ARG1	
A2	**ARG1**	ARG3	
A10	**ARG2**	ARG11	
A13	**ARG3**	ARG13	ARG1
A17	**ARG4**	ARG5	
A19	**ARG5**	ARG16	
A21	**ARG6**	ARG8	
A23	**ARB1**	ARG1	
A29	**ARB1**	ARG1	
A30	**ARB1**	ARG1	
A35	**ARB2**	AB1	
A40	**ARG7**	ARG1	
A41	**ARG8**	ARG14	
A42	**ARG9**	ARG2	
A43	**ARG10**	ARG1	
A47	**ARB1**	ARG1	
A50	**ARB1**	ARG1	
A52	**ARB1**	ARG1	
A53	**ARB1**	ARG1	
A54	**ARB1**	ARG1	
A55	**ARB1**	ARG1	
A57	**ARG11**	ARG12	
TŠ6	**CHI1**	CHI2	
TŠ13	**CHI2**	CHI1	
TŠ14	**CHI1**	CHI2	
TŠ15	**CHI3**	CHI4	
TŠ16	**CHI4**	CHI3	
TŠ18	**CHI2**	CHI1	CHI1
H172	**BRA2**	BRA5	
H408	**BRA1**	BRA3	
H424	**BRA6**	BRA2	
H429	**BRA1**	BRA3	

H442	**BRA9**	BRA6
H567	**ARB2**	AB1
H574	**BRA10**	BRA1
H575	**BRA5**	BRA4
H585	**BRA1**	BRA3
P66	**ARG12**	ARG17
P67	**ARG13**	ARG17
P68	**ARB1**	ARG1
P69	**ARG14**	ARG1
P76	**ARG16**	ARG6
8G	**MEX1**	MEX1

Table S2
The list of G1 samples from Eurasia and Africa partially published previously in Kinkar et al. (2016 and 2017).

Lab code in Tartu	Haplotype in this study (11 682 bp)	Haplotype in Kinkar et al. (2016; 8274 bp)	Haplotype in Kinkar et al. (2017; 11 443 bp)
V8	**GRE1**	GRE1	
HS4	**ROM1**	ROM1	
Fin16	**FIN1**	FIN1	FIN1
IT 10	**ITA2**	ITA6	
AC3	**ITA4**	ITA3	
AC4	**ITA4**	ITA3	
2G	**SPA1**	SPA2	
12G	**SPA2**	SPA3	
ALB3	**ALB1**	ALB1	
ALB4	**ALB2**	ALB2	ALB1
5455	**FRA1**		FRA3
P2	**SPA3**	SPA4	
P15	**SPA5**	SPA5	
P16	**SPA6**	SPA6	
P21	**SPA7**		SPA5
P47	**SPA8**	SPA9	
P51	**SPA9**	SPA10	
P61	**SPA11**	SPA1	
S 2	**TUR2**	TUR1	
S 9	**TUR3**	TUR3	
S 13	**TUR4**	TUR31	
S 14	**TUR5**	TUR4	
S 15	**TUR3**	TUR3	
S 16	**TUR3**	TUR3	
S 19	**TUR6**	TUR5	
S 20	**TUR7**	TUR6	
S30	**TUR8**	TUR11	
S31	**TUR9**	TUR12	
S33	**TUR10**	TUR14	
S53	**TUR11**	TUR20	
S69	**TUR12**	TUR25	
S77	**TUR13**	TUR26	

S78	**TUR14**	TUR27	
S99	**TUR17**	TUR35	
S104	**TUR18**	TUR37	
S107	**TUR19**	TUR39	
S111	**TUR21**	TUR40	
S112	**TUR22**	TUR41	
S117	**TUR23**	TUR44	
S119	**TUR24**	TUR45	
S120	**TUR17**	TUR35	
S121	**TUR25**	TUR46	
S124	**TUR26**	TUR48	
S135	**TUR28**	TUR51	
S136	**TUR29**	TUR52	
S138	**TUR30**	TUR53	
S142	**TUR32**	TUR54	
S144	**TUR33**	TUR55	
S146	**TUR34**	TUR56	
S149	**TUR35**	TUR58	
S154	**TUR36**	TUR62	
U66	**TUN10**		TUN1
MI2	**IND1**		IND2
IR19	**IRA7**		IRA4

Table S3
Data for the 211 *Echinococcus granulosus* sensu stricto G1 isolates sequenced in this study.

Lab code in Tartu	Haplotype	Host	Origin	GenBank accession nr
V8	GRE1	Sheep	Greece	
HS4	ROM1	Cattle	Romania	
Fin16	FIN1	Human	Finland, Algerian patient	
IT3	ITA1	Cattle	Italy, South	
IT10	ITA2	Cattle	Italy, North	
HIP9	ITA3	Human	Italy, Pavia	
AC3	ITA4	Sheep	Italy, Sicily island	
AC4	ITA4	Sheep	Italy, Sicily island	
2G	SPA1	Human	Spain	
7G	ROM2	Human	Romania	
12G	SPA2	Wild boar	Spain	
ALB3	ALB1	Sheep	Albania, Tirana	
ALB4	ALB2	Sheep	Albania, Tirana	
4150	MOR1	Cattle	Morocco, Sidi Kacem	
5455	FRA1	Cattle	France, Oloron-Sainte-Marie	
6200	MOL1	Sheep	Moldova, Centre	
6214	MOL2	Cattle	Moldova, Centre	
6187	MOL3	Sheep	Moldova, South	
P2	SPA3	Sheep	Central Spain	
P3	SPA4	Sheep	Central Spain	
P4	SPA4	Sheep	Central Spain	
P15	SPA5	Sheep	Central Spain	
P16	SPA6	Sheep	Central Spain	
P21	SPA7	Sheep	Central Spain	
P47	SPA8	Pig	Spain, Segovia	
P51	SPA9	Goat	Central Spain	
P52	SPA10	Goat	Central Spain	
P53	SPA9	Goat	Central Spain	
P61	SPA11	Human	Spain, Madrid	
U3	TUN1	Sheep	Tunisia, Sousse	
U8	TUN2	Sheep	Tunisia, Sousse	
U11	TUN3	Sheep	Tunisia, Sousse	
U17	TUN4	Sheep	Tunisia, Sousse	
U30	TUN5	Sheep	Tunisia, Sousse	
U32	TUN6	Sheep	Tunisia, Sousse	
U33	TUN7	Sheep	Tunisia, Sousse	
U44	TUN8	Sheep	Tunisia, Sousse	
U57	TUN8	Sheep	Tunisia, Sousse	
U62	TUN9	Sheep	Tunisia, Sousse	
U66	TUN10	Sheep	Tunisia, Kairouan	
U80	TUN10	Sheep	Tunisia, Kairouan	
U82	TUN11	Sheep	Tunisia, Kairouan	
U110	TUN12	Sheep	Tunisia, Kairouan	
U117	TUN13	Sheep	Tunisia, Kasserine	
U118	TUN14	Sheep	Tunisia, Kasserine	
U120	TUN15	Sheep	Tunisia, Gafsa	

U141	TUN14	Cattle	Tunisia, Sousse
U154	TUN16	Cattle	Tunisia, Monastir
U167	TUN17	Cattle	Tunisia, Kasserine
U183	TUN17	Cattle	Tunisia, Kasserine
S1	TUR1	Sheep	Turkey, Elazig
S2	TUR2	Sheep	Turkey, Elazig
S7	TUR1	Sheep	Turkey, Elazig
S9	TUR3	Sheep	Turkey, Elazig
S12	TUR1	Sheep	Turkey, Elazig
S13	TUR4	Sheep	Turkey, Elazig
S14	TUR5	Sheep	Turkey, Elazig
S15	TUR3	Sheep	Turkey, Elazig
S16	TUR3	Sheep	Turkey, Elazig
S19	TUR6	Cattle	Turkey, Elazig
S20	TUR7	Cattle	Turkey, Elazig
S30	TUR8	Cattle	Turkey, Erzurum
S31	TUR9	Cattle	Turkey, Erzurum
S33	TUR10	Cattle	Turkey, Erzurum
S53	TUR11	Cattle	Turkey, Erzurum
S69	TUR12	Cattle	Turkey, Erzurum
S77	TUR13	Cattle	Turkey, Erzurum
S78	TUR14	Cattle	Turkey, Erzurum
S83	TUR15	Cattle	Turkey, Erzurum
S91	TUR16	Cattle	Turkey, Erzurum
S99	TUR17	Cattle	Turkey, Erzurum
S104	TUR18	Cattle	Turkey, Erzurum
S107	TUR19	Sheep	Turkey, Elazig
S109	TUR20	Sheep	Turkey, Elazig
S111	TUR21	Sheep	Turkey, Elazig
S112	TUR22	Sheep	Turkey, Elazig
S117	TUR23	Sheep	Turkey, Elazig
S119	TUR24	Sheep	Turkey, Elazig
S120	TUR17	Sheep	Turkey, Elazig
S121	TUR25	Sheep	Turkey, Elazig
S124	TUR26	Sheep	Turkey, Elazig
S129	TUR27	Sheep	Turkey, Elazig
S135	TUR28	Sheep	Turkey, Elazig
S136	TUR29	Sheep	Turkey, Elazig
S138	TUR30	Sheep	Turkey, Elazig
S141	TUR31	Sheep	Turkey, Elazig
S142	TUR32	Sheep	Turkey, Elazig
S144	TUR33	Sheep	Turkey, Elazig
S146	TUR34	Sheep	Turkey, Elazig
S148	TUR20	Sheep	Turkey, Elazig
S149	TUR35	Cattle	Turkey, Elazig
S154	TUR36	Sheep	Turkey, Elazig
J1	AUS1	Dingo	Australia
J2	AUS2	Dingo	Australia
J3	AUS3	Dingo	Australia
OU2	TUN18	Human	Tunisia, Kasserine

OU3	TUN19	Human	Tunisia, Sidi bouzid
OU5	TUN20	Human	Tunisia, Sidi bouzid
OU6	TUN20	Human	Tunisia, Sidi bouzid
OU7	TUN21	Human	Tunisia, Sidi bouzid
OU9	TUN22	Human	Tunisia, Sidi bouzid
OU10	TUN15	Human	Tunisia, Kasserine
OU12	TUN23	Human	Tunisia, Sidi bouzid
OU13	TUN24	Human	Tunisia, Kasserine
OU14	TUN25	Human	Tunisia, Gafsa
OU15	TUN5	Human	Tunisia, Kasserine
OU16	TUN26	Human	Tunisia, Mahdia
OU17	TUN27	Human	Tunisia, Kairouan
OU18	TUN5	Human	Tunisia, Mahdia
OU20	TUN28	Human	Tunisia, Kairouan
OU21	TUN29	Human	Tunisia, Kairouan
OU23	TUN30	Human	Tunisia, Mahdia
VA1	ITA5	Goat	Italy, Sardinia
VA3	ITA6	Sheep	Italy, Sardinia
VA6	ITA7	Sheep	Italy, Sardinia
VA7	ITA8	Sheep	Italy, Sardinia
VA14	ITA9	Sheep	Italy, Sardinia
VA16	GRE1	Sheep	Greece
VA17	GRE2	Sheep	Greece
ZA11	ALG1	Human	Algeria
ZA12	ALG2	Human	Algeria, Khenchla
ZA13	ALG3	Human	Algeria, Bouira
ZA20	ALG4	Human	Algeria, Tipaza
ZA23	ALG5	Human	Algeria, Ain Defla
ZA24	ALG6	Human	Algeria, Laghouat
ZA25	ALG7	Human	Algeria, Ouargla
ZA26	ALG8	Human	Algeria, Ain Defla
ZA27	ALG9	Human	Algeria, Blida
ZA31	ALG10	Human	Algeria, Boumerdes
ZA32	ALG11	Human	Algeria, Ain Defla
ZA34	ALG12	Human	Algeria, Ain Defla
A1	ARB1	Cattle	Argentina, 9 de Julio
A2	ARG1	Cattle	Argentina, Castelli
A10	ARG2	Pig	Argentina, Buenos Aires
A13	ARG3	Cattle	Argentina, Balcarce
A17	ARG4	Sheep	Argentina, Tres Arroyos
A19	ARG5	Sheep	Argentina, Tres Arroyos
A21	ARG6	Sheep	Argentina, Tres Arroyos
A23	ARB1	Sheep	Argentina, Mar del Plata
A24	ARB1	Sheep	Argentina, Mar del Plata
A29	ARB1	Sheep	Argentina, Mar del Plata
A30	ARB1	Sheep	Argentina, Mar del Plata
A35	ARB2	Cattle	Argentina, Balcarce
A37	ARB1	Cattle	Argentina, Tres Arroyos
A40	ARG7	Cattle	Argentina, Ayacucho
A41	ARG8	Cattle	Argentina, Balcarce

A42	ARG9	Cattle	Argentina, Balcarce
A43	ARG10	Cattle	Argentina, San Cayetano
A47	ARB1	Sheep	Argentina, Mar del Plata
A50	ARB1	Sheep	Argentina, Mar del Plata
A52	ARB1	Sheep	Argentina, Mar del Plata
A53	ARB1	Sheep	Argentina, Mar del Plata
A54	ARB1	Sheep	Argentina, Tres Arroyos
A55	ARB1	Sheep	Argentina, Tres Arroyos
A57	ARG11	Sheep	Argentina, Tres Arroyos
TŠ6	CHI1	Cattle	Chile, Coquimbo
TŠ13	CHI2	Cattle	Chile, Illapel
TŠ14	CHI1	Cattle	Chile, Illapel
TŠ15	CHI3	Cattle	Chile, Illapel
TŠ16	CHI4	Cattle	Chile, Illapel
TŠ18	CHI2	Cattle	Chile, Illapel
H172	BRA2	Cattle	Brazil, Cachoeira do Sul
H369	BRA3	Cattle	Brazil, Cacapava do Sul
H404	BRA4	Cattle	Brazil, Herval
H408	BRA1	Cattle	Brazil, Arroio Grande
H424	BRA6	Cattle	Brazil
H429	BRA1	Cattle	Brazil
H433	ARB1	Cattle	Brazil, Sao Gabriel
H439	BRA7	Cattle	Brazil, Sao Gabriel
H440	BRA8	Cattle	Brazil, Sao Gabriel
H442	BRA9	Cattle	Brazil
H567	ARB2	Cattle	Brazil, Alegrete
H574	BRA10	Cattle	Brazil, Bagé
H575	BRA5	Cattle	Brazil, Livramento
H585	BRA1	Cattle	Brazil, Alegrete
IR 11	IRA1	Sheep	Iran, Golestan
IR 12	IRA1	Sheep	Iran, Golestan
IR 13	IRA4	Sheep	Iran, Golestan
IR 14	IRA1	Sheep	Iran, Golestan
IR 17	IRA5	Sheep	Iran, Mazandaran
IR 18	IRA6	Sheep	Iran, Mazandaran
IR 19	IRA7	Sheep	Iran, Mazandaran
IR 21	IRA8	Sheep	Iran, Tehran
IR 22	IRA2	Sheep	Iran, Tehran
IR 23	IRA2	Sheep	Iran, Tehran
IR 24	IRA9	Sheep	Iran, Tehran
IR 27	IRA10	Sheep	Iran, Tehran
IR 29	IRA11	Sheep	Iran, Tehran
IR 31	IRA12	Sheep	Iran, Isfahan
IR 32	IRA13	Sheep	Iran, Isfahan
IR 33	IRA14	Sheep	Iran, Isfahan
IR 35	IRA15	Goat	Iran, Isfahan
IR 46	IRA3	Human	Iran, Isfahan
IR 47	IRA3	Human	Iran, Isfahan
IR 49	IRA16	Cattle	Iran, Isfahan
IR 51	IRA17	Cattle	Iran, Isfahan

IR 52	IRA18	Cattle	Iran, Isfahan
P66	ARG12	Cattle	Argentina
P67	ARG13	Cattle	Argentina
P68	ARB1	Cattle	Argentina
P69	ARG14	Cattle	Argentina
P70	ARG15	Cattle	Argentina
P76	ARG16	Sheep	Argentina
J86	MON2	Human	Mongolia
J88	MON1	Human	Mongolia
J91	MON1	Human	Mongolia
SO212	ARG17	Sheep	Argentina, Neuquen
N1	IRA19	Goat	Iran, Lorestan
B20	KAZ1	Human	Kazakhstan
8G	MEX1	Pig	Mexico
MI2	IND1	Buffalo	India

Figure 1

Click here to download high resolution image

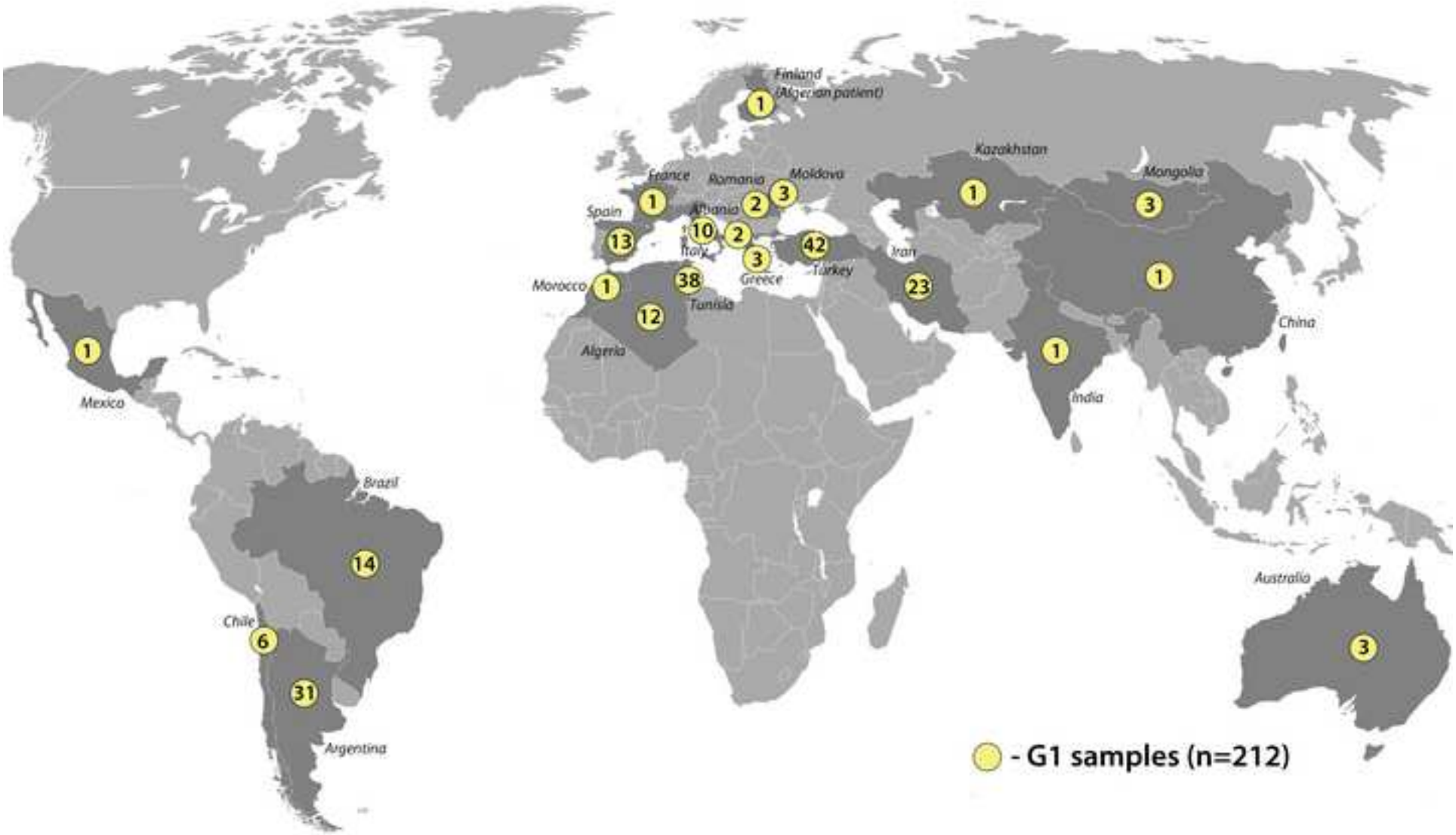


Figure 1 print version

Click here to download high resolution image

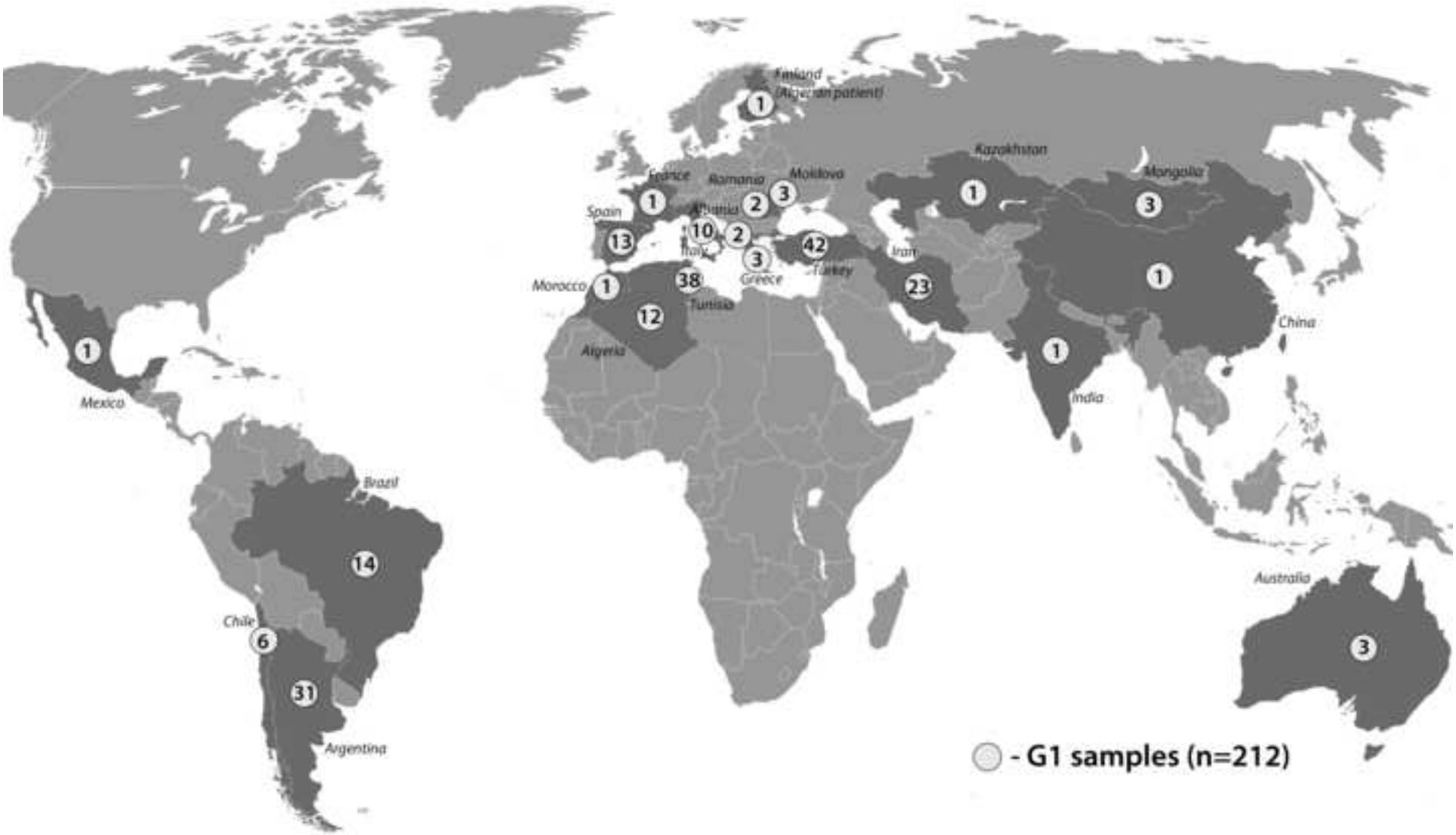


Figure 2

Click here to download high resolution image

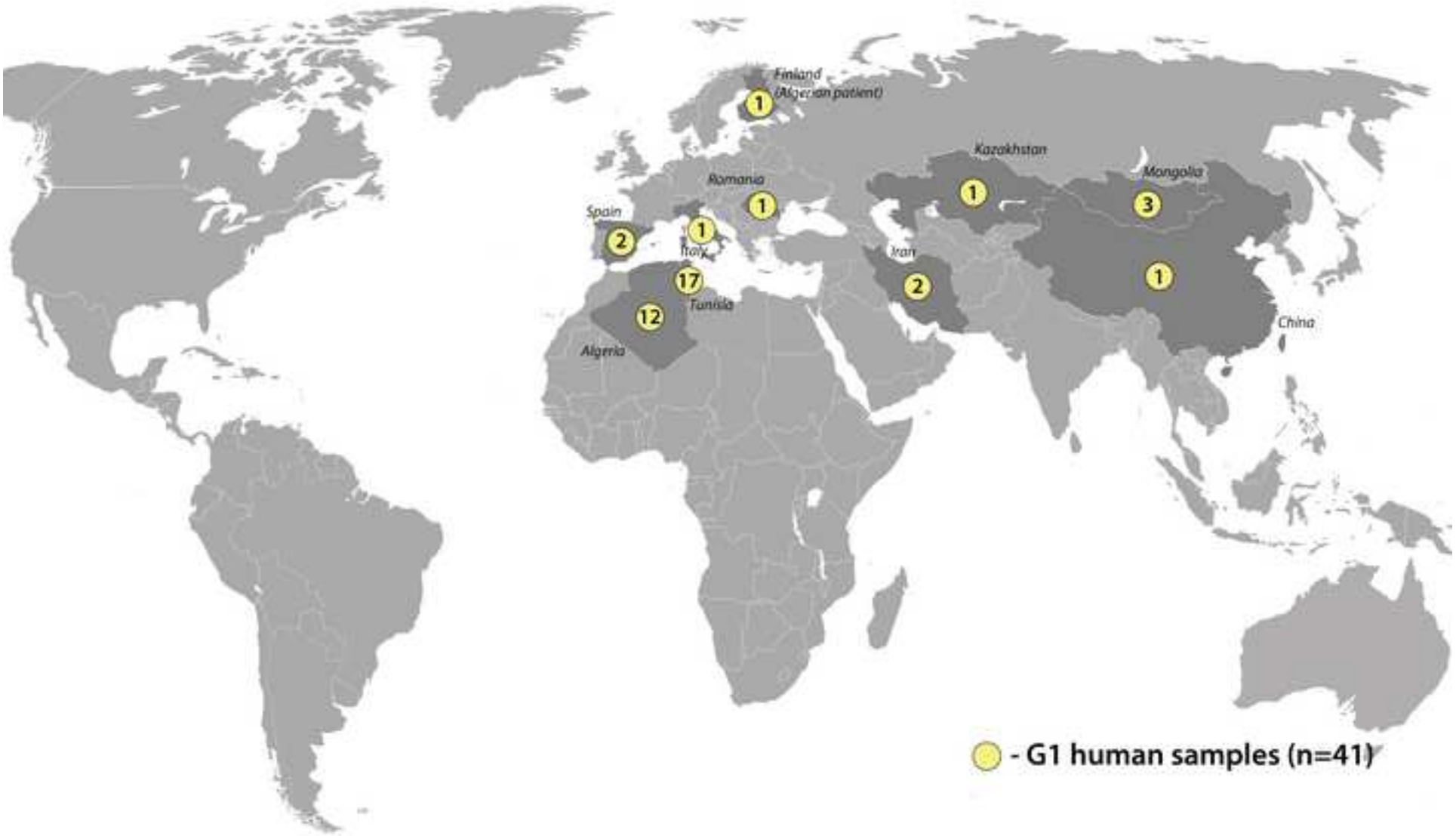


Figure 2 print version

Click here to download high resolution image

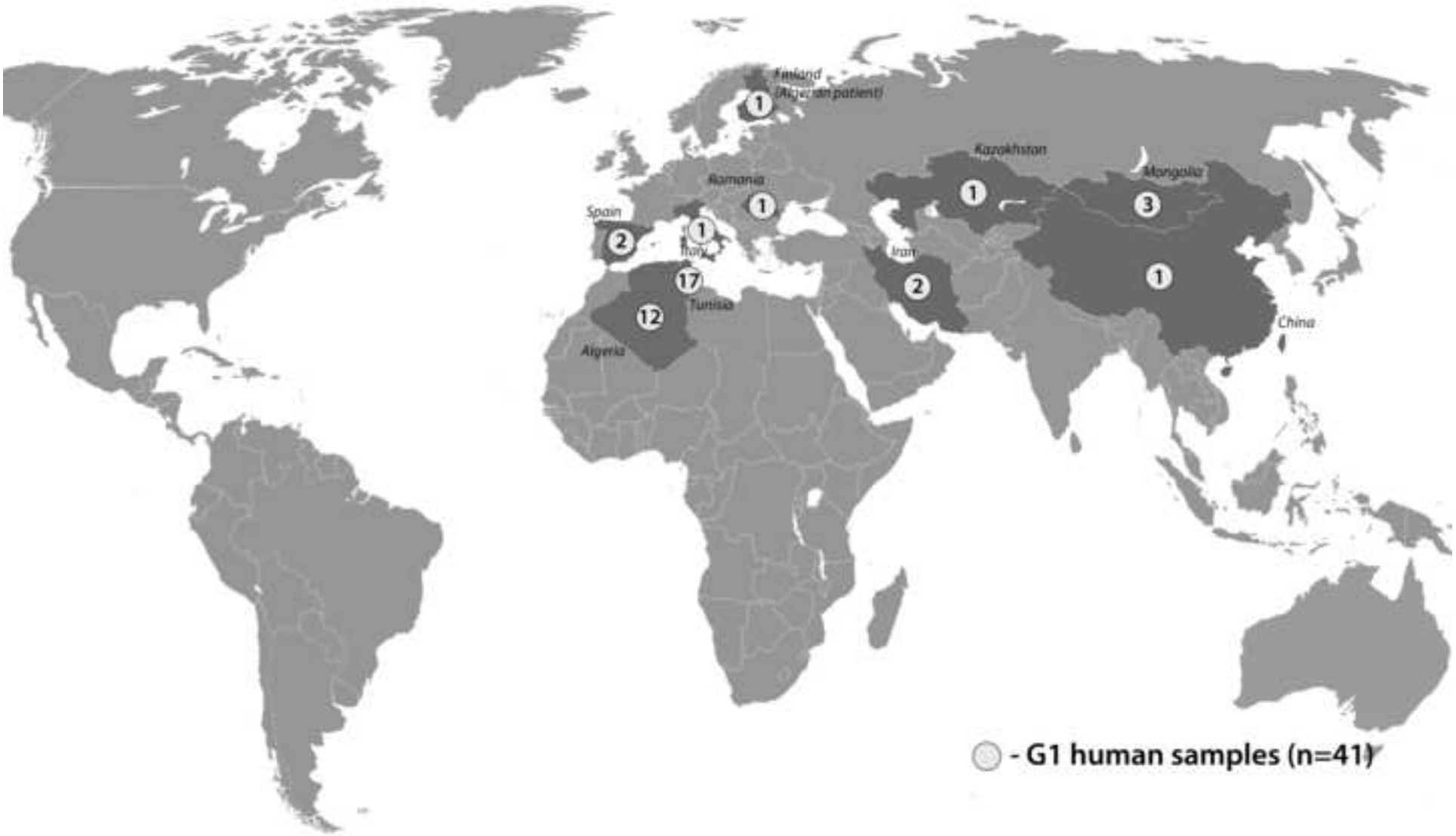


Figure 3

Click here to download high resolution image

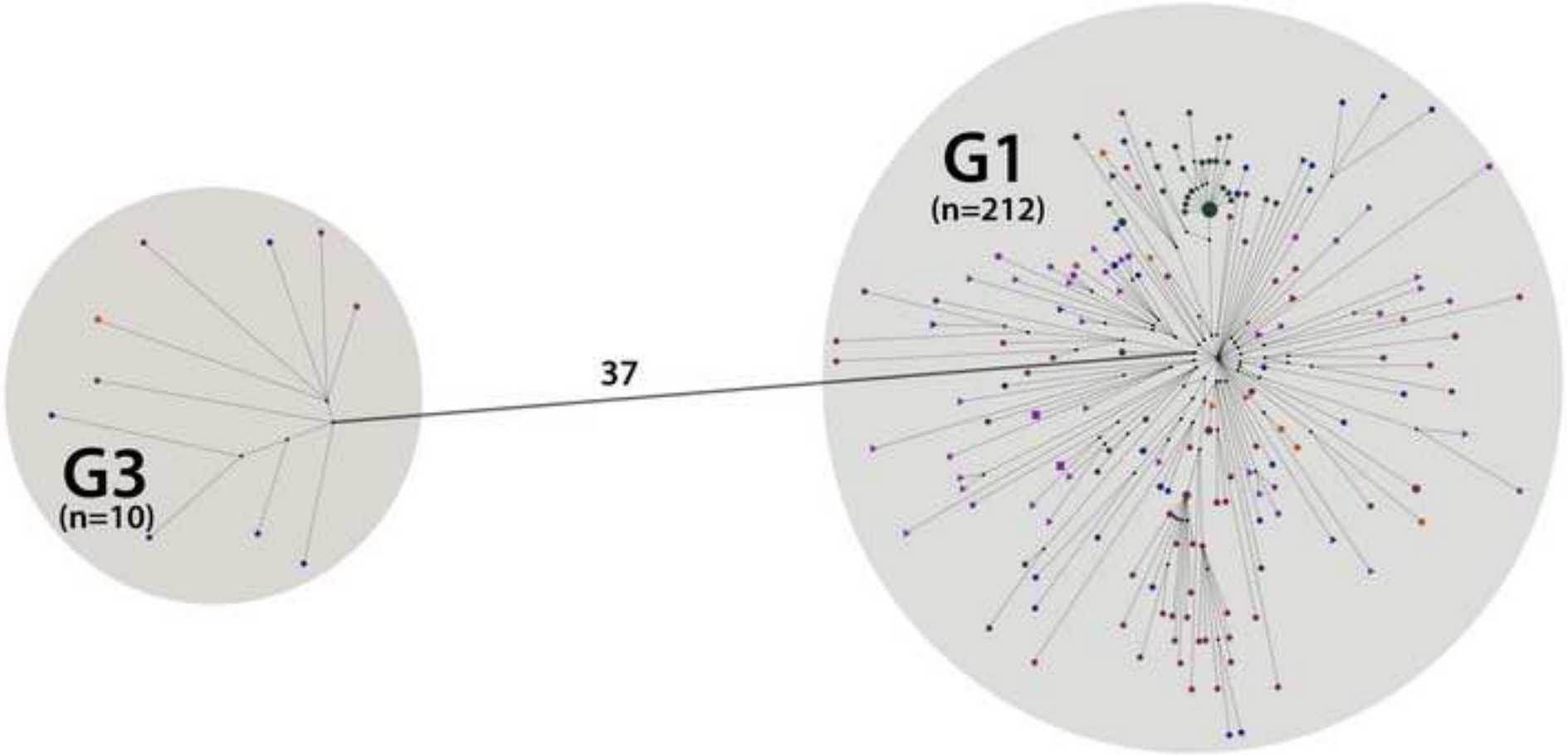


Figure 3 print version
Click here to download high resolution image

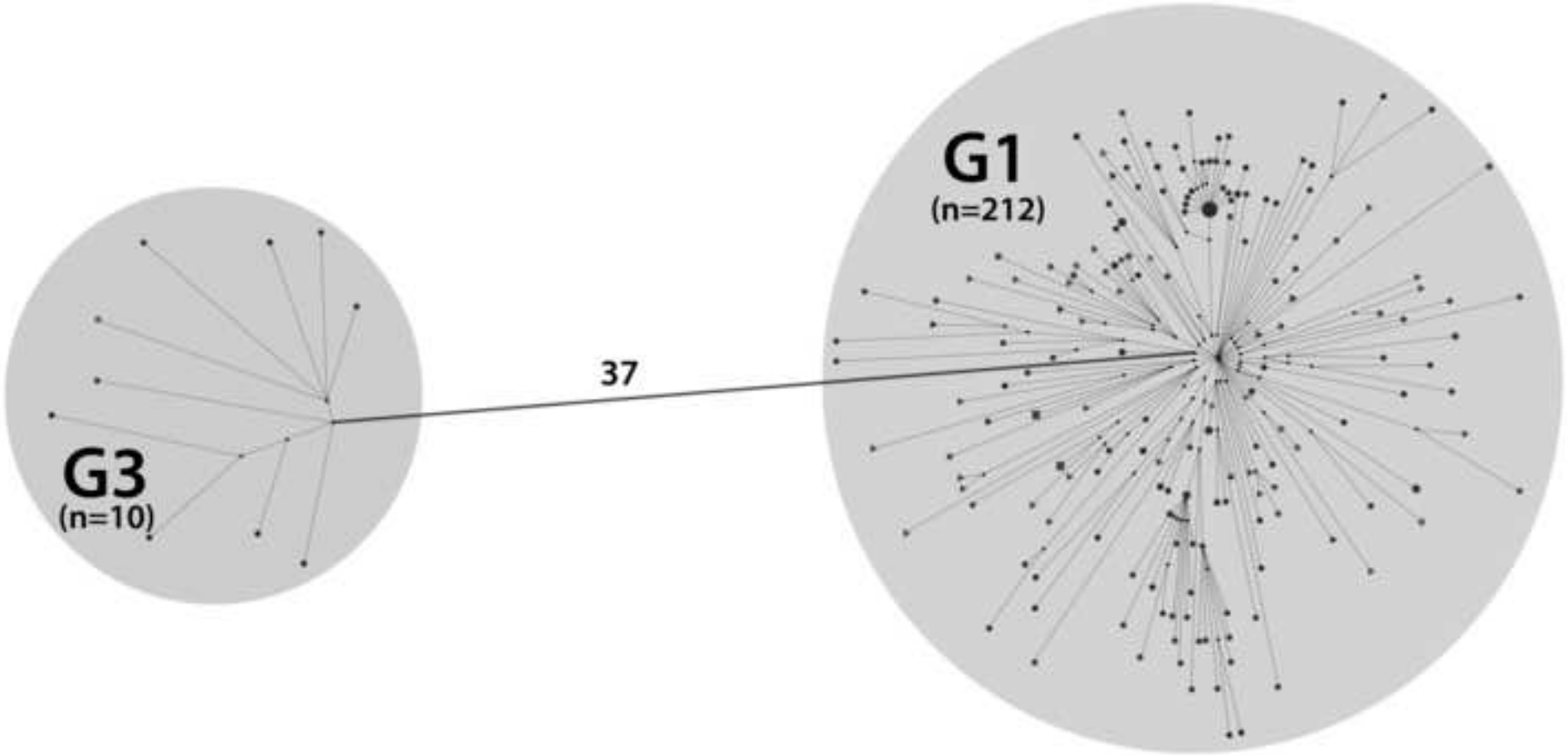


Figure 4

Click here to download high resolution image

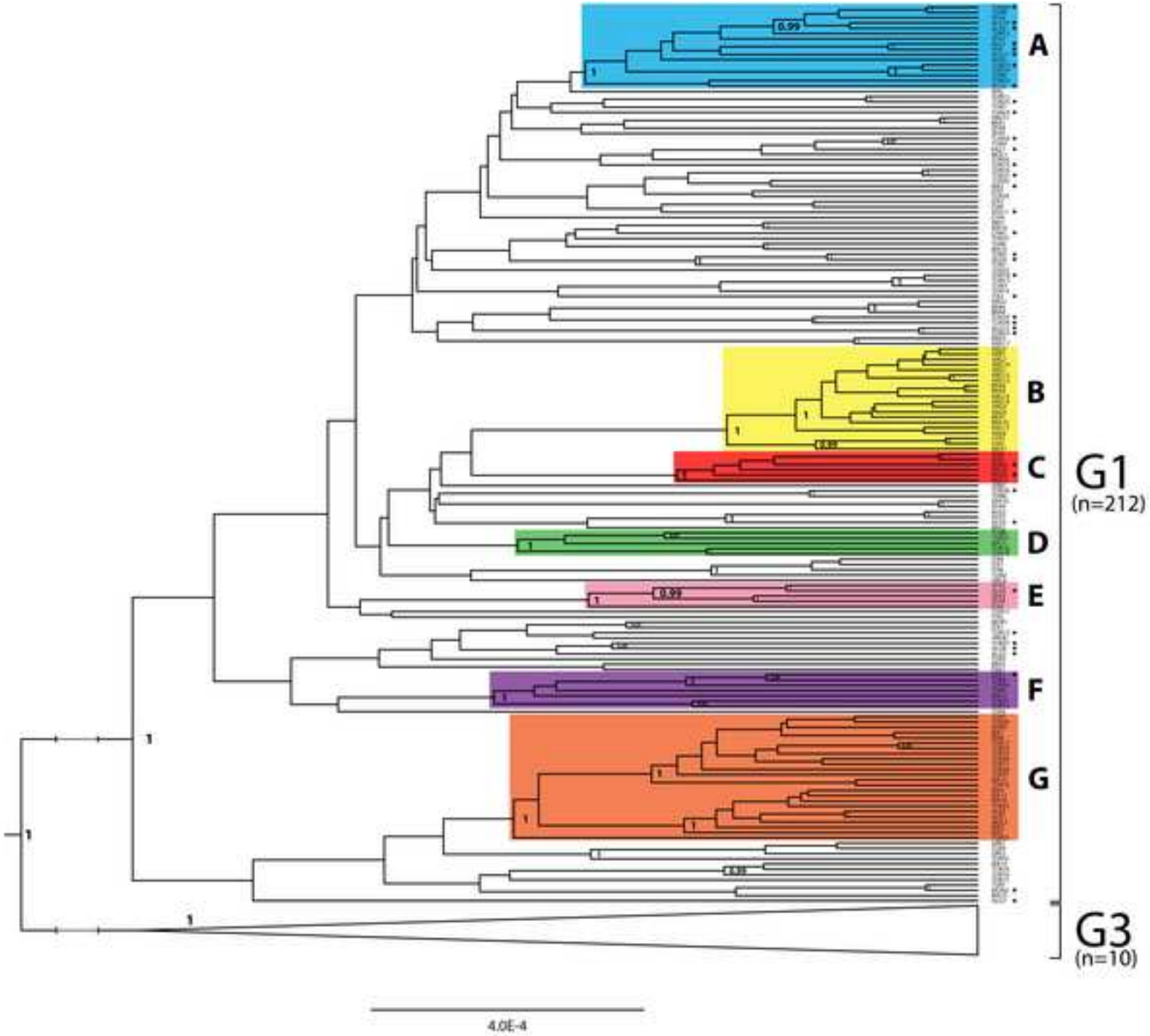


Figure 4 print version

Click here to download high resolution image

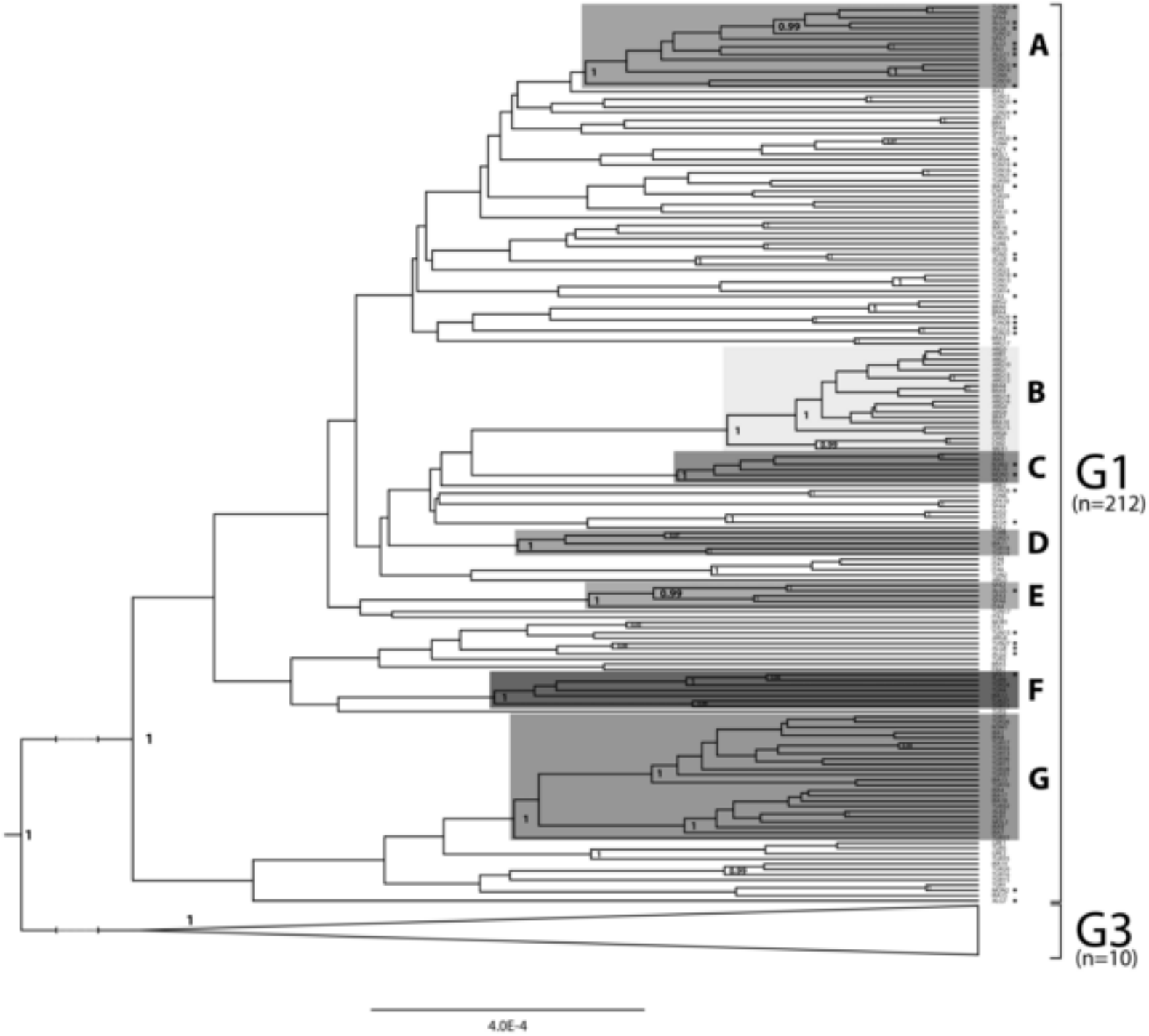


Figure 5

Click here to download high resolution image

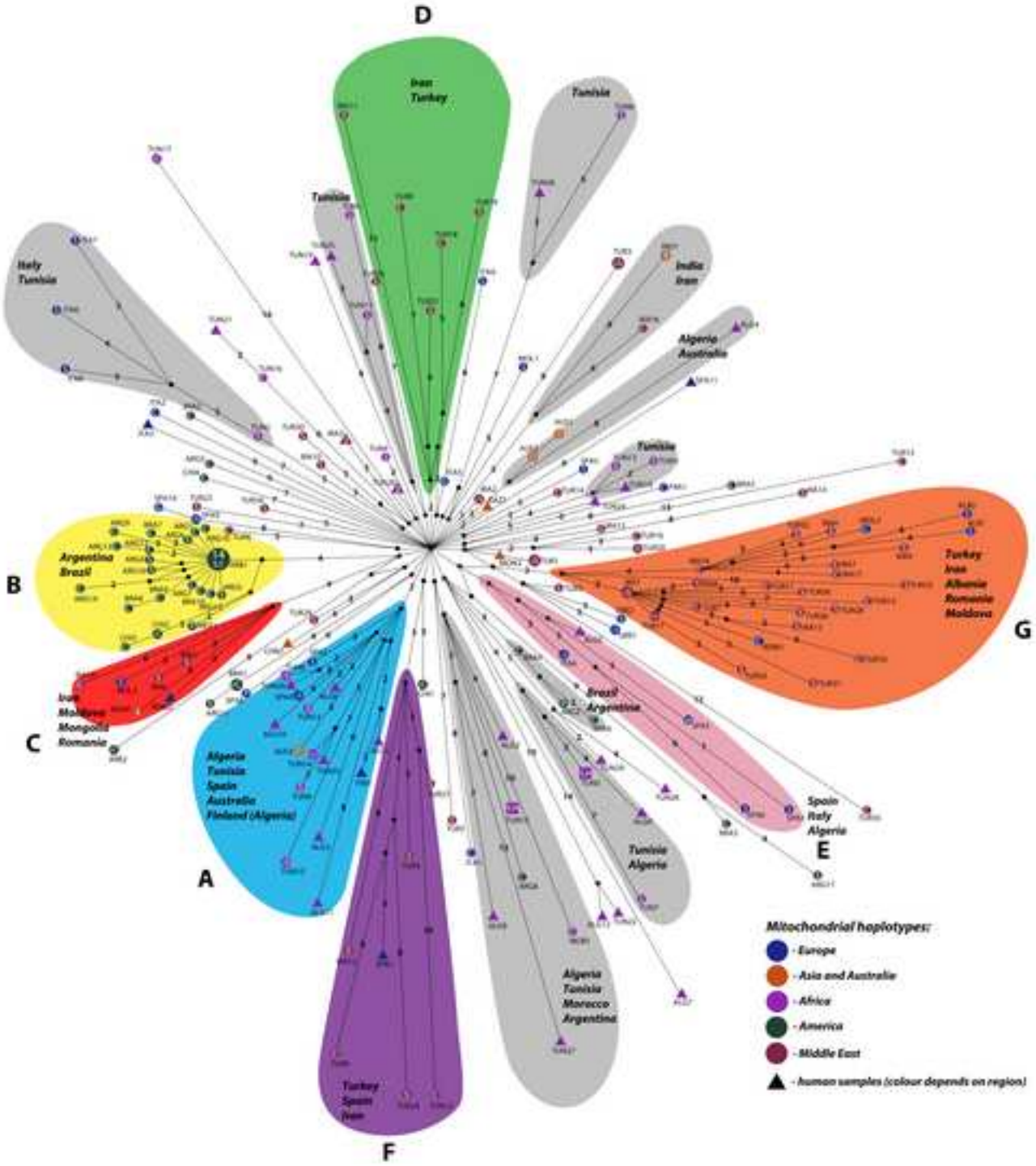


Figure 5 print version

Click here to download high resolution image

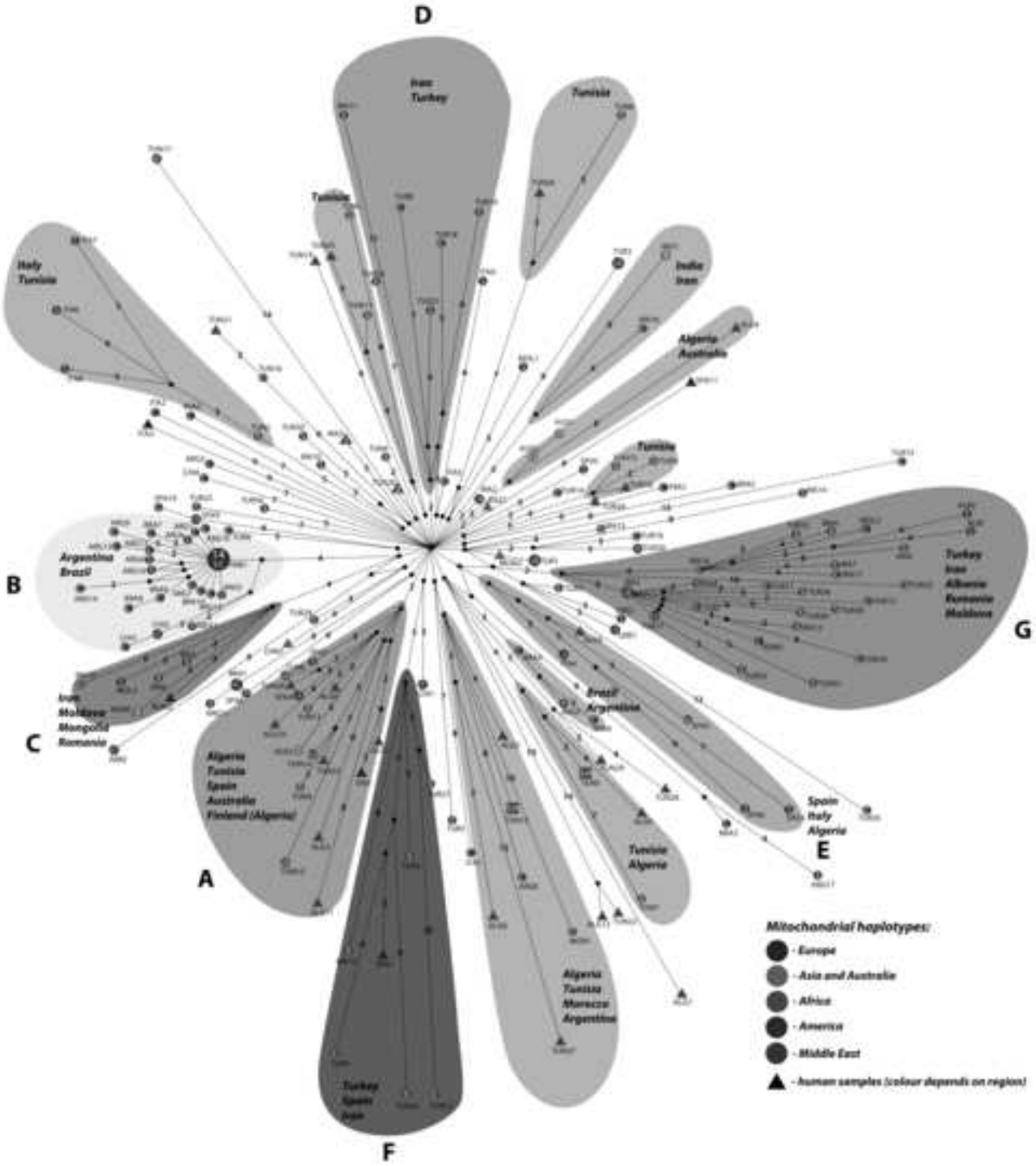


Figure 6

Click here to download high resolution image

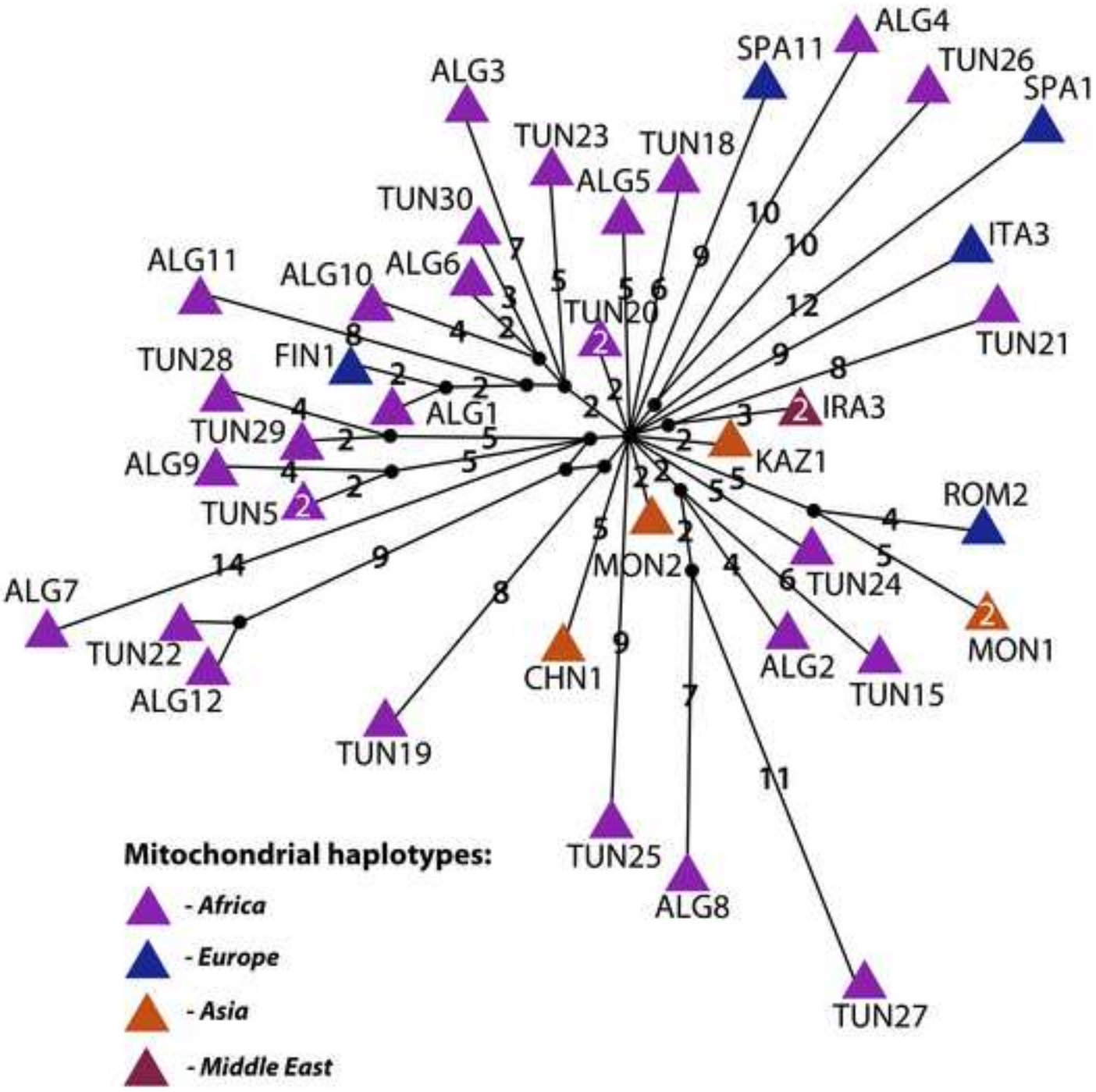


Figure 6 print version

Click here to download high resolution image

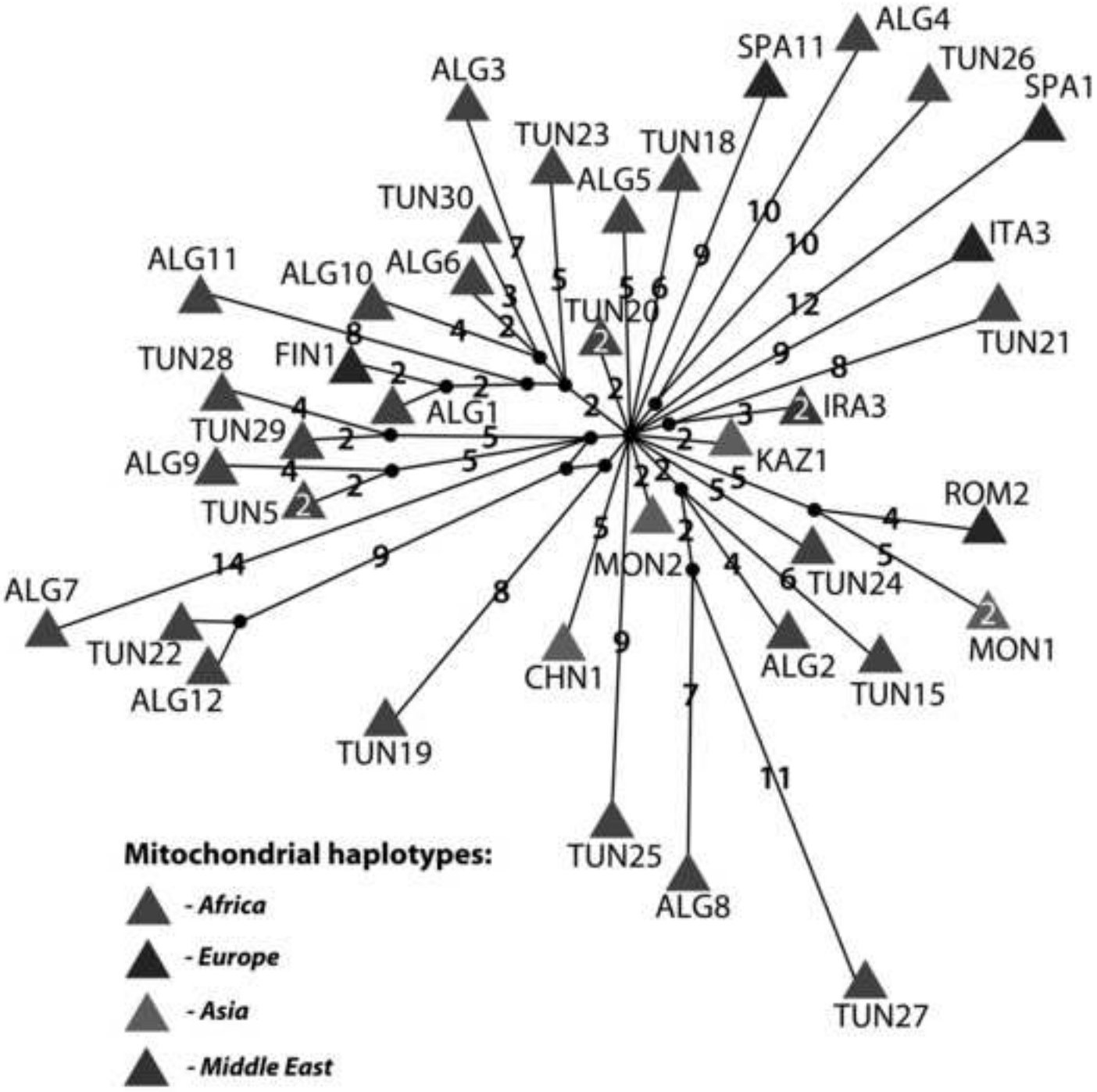


Figure 7

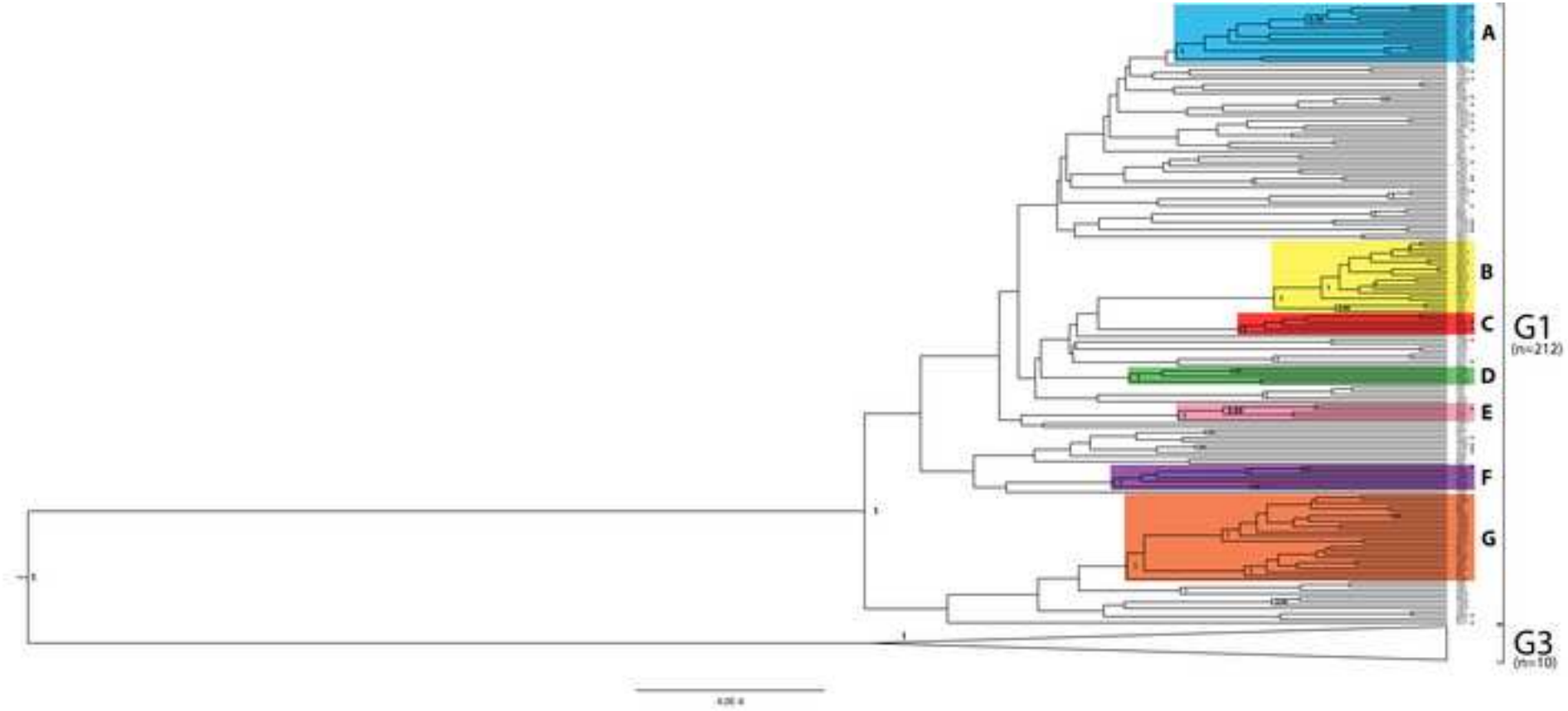
Click here to download high resolution image

Figure 7 print version

Click here to download high resolution image

Figure S1

Click here to download high resolution image



Highlights

- 11 682 bp of mtDNA was analysed for 222 *E. granulosus* s. s. samples globally
- G1 and G3 are distinct mitochondrial genotypes
- The genetic diversity of *E. granulosus* s. s. G1 is extremely high globally
- The main diffusion routes of G1 originated from Turkey, Tunisia and Argentina
- Livestock trade has greatly influenced the present-day diversity of genotype G1

*Graphical Abstract (for review)

Click here to download high resolution image