Pakistan Veterinary Journal

ISSN: 0253-8318 (PRINT), 2074-7764 (ONLINE)
DOI: 10.29261/pakvetj/2024.143

RESEARCH ARTICLE

High Incidence of G1 Genotype Found in the Levant Revealed by Sequence-based Association Analysis of *Echinococcus granulosus* (sensu stricto)

Enas Al-khlifeh[1]*, Ayed Alshammari[2] and Hussein Alnasarat[3]

[1]Department of Medical Laboratory Science, Al-Balqa Applied University, Al-Salt (19117), Jordan; [2]Department of Biology, College of Science, University of Hafr Al-Batin, Hafr Al-Batin, 39511 Saudi Arabia; [3]Department of Biology, Al-Hussain Bin Talal University, Ma'an, (71111), Jordan
*Corresponding author: Al-khlifeh.en@bau.edu.jo

ARTICLE HISTORY (23-507)

Received: September 11, 2023
Revised: February 23, 2024
Accepted: February 26, 2024
Published online: March 20, 2024

Key words:
Echinococcus granulosus
Cystic echinococcosis
Haplotype diversity
Cox1
Nad2
Jordan

A B S T R A C T

Hydatid cyst disease, caused by *Echinococcus granulosus* sensu lato (*s.l.*), is linked to endemic areas in the Levant region, such as Jordan. There is a scarcity of epidemiological data about this disease. Our goal was to pinpoint the particular genotypes accountable for sheep and goat infection in unexplored regions of Southern Jordan. Partial sequences of cytochrome C oxidase (*Cox1*, 450 bp) and NADH dehydrogenase subunit 2 (*Nad2*, 820 bp) of the *E. granulosus* from Jordan were analysed and compared with *E. granulosus s.l.* reference genomes, to infer the phylogeny of the species. This analysis demonstrated a significant degree of intraspecific similarity and identified the presence of *E. granulosus* sensu stricto (G1, G3). The G1 genotype was the most prevalent (*Cox1*, n = 17/17); (*Nad2*, n = 13/18). The study examined partial *Cox1* sequences from different countries (Jordan, Lebanon, Iran, India) and regions (South America) to analyze the genetic architecture and phylogeography of the G1 genotype at regional and global levels. One primary haplotype (Hap_1) was present in the research locations. We hypothesized that a certain branch of G1, known as Hap_1, was prevalent among regions that were not geographically connected. Future research on the biology and distribution of *E. granulosus s.s.* will be enhanced by this information.

To Cite This Article: Al-khlifeh E, Alshammari A and Alnasarat H, 2024. High incidence of G1 genotype found in the levant revealed by sequence-based association analysis of *Echinococcus granulosus* (sensu stricto). Pak Vet J. http://dx.doi.org/10.29261/pakvetj/2024.143

INTRODUCTION

Parasitic helminthiases are one of the most severe infections consistently recorded worldwide. *Echinococcus granulosus* is a tapeworm belonging to the Taeniidae family that leads to zoonotic illnesses, including cystic echinococcosis (CE), which can be infectious and harmful to both people and animals. Canids, including domestic and wild dogs, serve as definitive hosts for *E. granulosus* during their heteroxenous life cycle. Various omnivorous and herbivorous animals serve as intermediary hosts instead (Romig *et al*., 2017). *E. granulosus s.l.* is currently recognized as a group of cryptic species with different levels of selectivity towards intermediate hosts. The term "cryptic" arises from differences in physical and developmental patterns, along with genetic variability within a species (Casulli *et al*., 2019). The current terminologies for the *E. granulosus s.l.* species complex are *E. granulosus* sensu stricto (*s.s.*) (G1-G3), *E. equinus* (G4), *E. ortleppi* (G5), *E. felidis* (lion strain), and the genotype cluster *E. canadensis* (G6-G10), with unresolved taxonomic issues (Vuitton *et al*., 2020). Three genotypes (G1, G2, and G3) were initially identified within *E.granulosus s.s.*, with G1 referred to as the "cosmopolitan strain of sheep and goats" and G3 as the "buffalo strain." It has been proposed that G2 is a subvariant of G3 that also causes CE in sheep (Kinkar *et al*., 2017). Research on several genotypes of *E. granulosus s.s.* is needed as it is the primary pathogenic agent of human cystic echinococcosis (Alvarez Rojas *et al*., 2014). A substantial number of human cystic echinococcosis cases worldwide are attributed to the G1 genotype (Casulli *et al*., 2022).

The tapeworm's larval stage, referred to as the metacestode stage, primarily develops in the liver and lungs in the form of fluid-filled sacs known as hydatid cysts. Dogs get parasites by consuming internal organs containing fully developed larval cysts, which also include several immature scoleces. Every protoscolex matures into an adult tapeworm within the intestines of dogs (Craig *et al*., 2007). Infectious eggs are released into the environment during maturation, along with the feces of sick dogs. Both humans

and other herbivores contract CE by consuming grass or drinking water contaminated with eggs. Both demographic and behavioral factors contribute to the transmission of infection (Craig *et al*., 2007). Monitoring studies have shown that CE is more prevalent in rural areas, with rates of 7.4% in the Middle East and North Africa (Galeh *et al.*, 2018), 50% in South China (Gao *et al*., 2021), and over 11% in certain rural portions of South America (Uchiumi *et al.*, 2021). One of the main factors that increases the likelihood of infection recycling is human activity, namely feeding dogs raw offal with *E. granulosus* cysts after slaughtering animals at home, in addition to direct interaction between dogs and farm animals (Larrieu *et al*., 2019).

The phylogenetic approach established the Middle Eastern origin hypothesis for *E. granulosus s.s.* colonization of intermediate hosts. This hypothesis suggests that the colonization began where domesticated animals and plants were first brought under human control and then spread to the Far East and Europe (Yanagida *et al*., 2012; Kinkar *et al*., 2018a). Sheep may have been domesticated in the northern Levant, an area in the Middle East that includes Jordan, based on zooarcheological data (Meadows, 2014). Studying the genetic differentiation of *E. granulosus s.s.* populations in different regions and analyzing haplotypes has provided valuable insights into regional variation and evolution (Eckert and Deplazes, 2004; Nakao *et al*., 2010; Yanagida *et al*., 2012; Mehmood *et al*., 2020). Yanagida *et al*. (2012) found a shared G1 haplotype in Jordanian and Iranian populations, as well as in the Fareast population of *E. granulosus s.s*., using the entire mt-*Cox1* gene. They observed a high level of haplotype diversity but little nucleotide diversity. They determined that there was a significant amount of gene flow among Middle Eastern isolates. Nakao *et al*. (2010) found a founder effect in Chinese populations *of E. granulosus s.s.* from a distant region. They suggested that bottleneck events happened in the original population of *E. granulosus s.s.* during the introduction of domestic sheep in China as an intermediate host. A recent hypothesis was suggested after studying the genetic variability of the mtDNA of *E. granulosus s.s.* G3 genotype in Pakistan compared to other global populations using the partial *Nad5* gene. The findings indicate a higher presence of G3 in the South Asian region (Mehmood *et al*., 2022). The accuracy of these findings in representing the genetic composition of *E. granulosus s.s.* needs validation through more samples, especially from the Levant region, known as the primary location for intermediary host domestication. Comparing the population genetic architecture of *E. granulosus s.s.* in various endemic regions is crucial for understanding its ancestral origin and worldwide dispersal processes.

CE infections are predominantly prevalent in the Middle East, particularly in the Levant countries (Monge-Maillo *et al*., 2019). It is essential to compare *E. granulosus s.s.* populations in the Middle Eastern regions with those in other global locations. Jordan is a reference point due to its historical role as a center connecting Asia, Africa, and Europe in the Middle East. Jordan has a high prevalence of CE due to the continuous occurrence of surgically treated cases. Al-Qaoud *et al*. (2003a) and Himsawi *et al*. (2019) findings align with Kamhawi (1995), indicating that cystic echinococcosis is prevalent in Jordan (Al-Qaoud *et al*., 2003a; Himsawi *et al.*, 2019; Kamhawi, 1995). Abundant clinical data, including retrospective investigations and case reports, are accessible regarding the prevalence of the condition in Jordan (Hijjawi *et al*., 2018). The infection was initially documented in 1966 (Kamhawi, 1995). The infection rate has since risen. Past research indicates that the G1 genotype-cosmopolitan common strain is prevalent in Jordan (Al-Qaoud *et al*., 2003b; Yanagida *et al*., 2012; Issa *et al*., 2018).

For genetic structure and phylogenetic research, the most reliable markers for *E. granulosus* species complex is mitochondrial DNA (mtDNA). They are commonly utilized in the examination of genetic alterations within a species due to their tiny size in comparison to chromosomes, preserved genetic components, and quick rate of evolution (Sun *et al*. 2021). The mitochondrial *Cox1* marker is commonly utilized (Han *et al*. 2019). Yet, in the *Cox1* gene, G1 and G3 showed variation at two specific locations according to the initial short 366 bp fragment sequences provided by (Bowles *et al*., 1994). *Cox1* markers are now considered inadequate for accurately classifying specimens as G1 or G3 genotypes, as shown by (Casulli *et al.*, 2012) (Romig *et al.*, 2015; Kinkar *et al*., 2018b). Recent studies have shown that the *Nad2* gene is reliable for identifying genotypes within the *E. granulosus s.l.* species complex and distinguishing between G1 and G3 genotypes (Wang *et al*., 2014; Samari *et al*., 2022).

Using the *Cox1* and *Nad2* genes, this work aimed at determining and examining the genetic variation of *E. granulosus s.s.* populations from the Levant and worldwide, to infer evolutionary relationships with populations. We initially investigated genetic variations in *Cox1* and *Nad2* of *E. granulosus s.s.* sequences in the Levant and other geographical regions sourced from GenBank; examined the prevalent genotype in Jordan and its phylogeny to evaluate the genetic structures in Jordan and compare them with other geographical populations. We studied *E. granulosus s.s.* haplotype diversity to evaluate the present evolutionary theory of the spread of *E. granulosus s.s.* The study also examined nucleotide-level changes and the features of mitochondrial mutations in the partial *Cox1* and *Nad2* genes.

MATERIALS AND METHODS

Jordanian *E. granulosus s.s.*: A search for nucleotide sequence data conducted utilizing the resource choice "nucleotide," in the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). A total of thirty-five gene sequences were obtained for Jordanian *E. granulosus s.s. Cox1* (n = 17) and *Nad2* (n = 18) gene sequences. The Cox1 accession numbers were OQ300479–OQ300495, while the *Nad2* gene accession numbers were OQ204092-OQ204109. The remaining sequences matched global Cox1 and Nad2 (Supplementary Table S1 and S2).

Sequence alignment and phylogeny: MEGA 11 software was used for the phylogenetic analysis of the

sequences (Tamura *et al.,* 2021). Multiple alignments were performed using the usual program settings. For partial *Cox1* target fragments, phylogeny was inferred using the neighbor-joining method (Saitou and Nei, 1987) at 1625 locations included in the final dataset. For *Nad2*, the evolutionary distance was determined using the Tamura-Nei model and Maximum Likelihood approach (Yang, 1994). The tree with the highest log probability was selected. For phylogenetic inference, the nucleotide sequences of the isolates examined in this study were aligned with GenBank-retrieved sequences for each genotype of *E. granulosus s.l.* (G1-G10). The reference sequences for *E. granulosus s.l.* are listed with their accession IDs in Supplementary Table (S1 and S2).

Nucleotide sequence variation and haplotype analysis: We calculated the rate of evolution in substitutions (i.e., inferred nucleotide changes) per site, where 'site' refers to a single nucleotide position in both partial *Cox1* and *Nad2* sequences among Jordanian *E. granulosus s.s.* and other *Cox1* and *Nad2* deposits available in the NCBI databank. This method was described by (Moses *et al.,* 2003). MEGA version 11 was used to identify the polymorphisms, transitions, and transversions. Only mutations that occurred several times at the same location were considered. Transitions and transversions were independently recorded as single-base mutations in the first, second, and third codon positions. Areas with gaps or incomplete data were excluded using pairwise deletion. The total transition or transversion bias (R) was calculated as $R = [A \times G \times k1 + T \times C \times k2]/ [(A + G) \times (T + C)]$. AG, GA, TC, and CT are well-known transition patterns. There were eight forms of transversions: A-T, T-A, A-C, C-A, G-T, T-G, G-C, and C-G. All the patterns in the *Cox1* and *Nad2* scoring regions were considered in this study. MEGA 11 was used to record the A, T, G, C, AT, and GC contents of all *E. granulosus s.s.* sequences obtained in this study.

Using the population analysis program DnaSP 6 (Rozas *et al.,* 2017), haplotype network analysis was performed using sequences examined in the FASTA format for both targeted mitochondrial genes. However, we focused on the partial *Cox1* gene-based phylogeography of the G1-genotype. The additional *E. granulosus s.s.* -G1 sequences that were downloaded from NCBI for comparison are included in the supplementary Table (S1) along with pertinent details about the places of origin of the sequences. We included 17 sequences from Iran, four from South America, three from India, one from Peru, one from Iraq, and one from Mongolia. The samples were chosen randomly. G1 isolates from Algeria (accession numbers KC897682, 84, 85, 86, 88, and AF 297617) were included in the *Nad2*-based haplotype study. The lack of sufficient *E. granulosus s.s.* -G1 sequences of the *Nad2* gene in the target countries have prevented researchers from concluding the phylogeography based on this mitogene. The haplotype count (h), nucleotide diversity (π), and haplotype diversity values (Hd) were computed as diversity index parameters for all sequences, irrespective of their place of origin. Using the genome in the Nexus format, we created a reticulate tree of haplotypes using the University of Otago Population 2023 (http://popart.otago.ac.nz).

Supplementary Table S1: The reference sequences of *Cox1* gene for *E. granulosus s.l.* obtained from the Genebank, listed with their accession IDs and country of origin.

	Accession Number	Country		Accession Number	Country
1	G1 JX854034.1	India	16	G8 KX685896.1	China
2	G5 KX685893.1	China	17	G9 KX685897.1	China
3	G10 KX685898.1	China	18	G5 JX854035.1	India
4	G3 KJ162562.1	Iran	19	G5 M84665.1	Holland
5	G1 MT796487.1	Argentina	20	G7 M84667.1	Poland
6	G1 MW732663.1	Peru	21	G7 KJ556997.1	China
7	G1 HM563013.1	Iran	22	G6 AB921084.1	Egypt
8	G1 HM563012.1	Iran	23	G6 AB921058.1	Egypt
9	G3 MG745736.1	Iran	24	G6 HM563018.1	Iran
10	G3 JN604103.1	Iran	25	G10 AF525457.1	Finland
11	G3 M84662.1	India	26	G10 AB777911.1	Russia
12	G3 M84663.1	India	27	G4 KP101616.1	United Kingdom
13	G3 KP339048.1	Iran	28	G4 M84664.1	Unknown
14	G3 HM563016.1	Iran	29	AB465246.1	South Korea
15	G3 DQ856466.1	Greece			

Supplementary Table S2: The reference sequences of *Nad2* gene for *E. granulosus s.l.* obtained from the Genebank, listed with their accession IDs and country of origin.

	Accession Number	
1	G1 MG672248	Iran
2	G1 KC897682	China
3	G3 MG682542.1	India
4	G3 MG682537.1	Iran
5	G3 MG682523.1	Greece
6	G3 MG682518.1	Italy
7	G1 MG672292.1	Algeria
8	G3 MG682515.1	Iran
9	G3 MG682511.1	Finland
10	G7 MH301016.1	France
11	G7 MH301003.1	Poland
12	G7 MH300974.1	Mexico
13	G6 MH300947.1	Sudan
14	G6 MH300952.1	Sudan
15	G6 MH300940.1	Sudan
16	NC 009938.1	Unknown
17	AB732959.1	Unknown
18	AF216697.1	Australia

RESULTS

Molecular identification of hydatid cysts

Confirmation of the *E. granulosus s.s.* (genotypes G1, G3): The consensus sequences of the incomplete *Cox1* fragment (n = 17) were 400-424 bp long and exhibited sequence homology of ≥92.91% to 100% with GenBank entries of G1 (Fig. 1A). Phylogenetic analysis showed a significant genetic difference between the studied isolates and the chosen outgroup (*Taenia saginata*), with an average length of 750 base pairs determined by similarity comparisons. Thirteen Jordanian *E. granulosus s.s.* sequances were grouped together with the reference G1-genotype sequence from India (accession number: JX854034.1) in the Bayesian phylogenetic analysis. Isolates OQ300489, OQ300480, and OQ300479 showed variations from the standard G1 reference sequence due to substitutions at different locations, resulting in the formation of a unique haplotype (Fig. 2A). The sequence OQ300482 was located between genotypes G1 and G3.

The homology between the incomplete *Nad2* consensus sequences (n= 18) and the G1 entry in GenBank was around 98.5%. Neighbor-joining cluster analysis indicated an estimated 88% commonality within the same species. To compare the *E.granulosus s.s.* isolates in the study, *T. saginata*, *Dipylidium caninum*, and *Fasciola hepatica* were employed as outgroups to

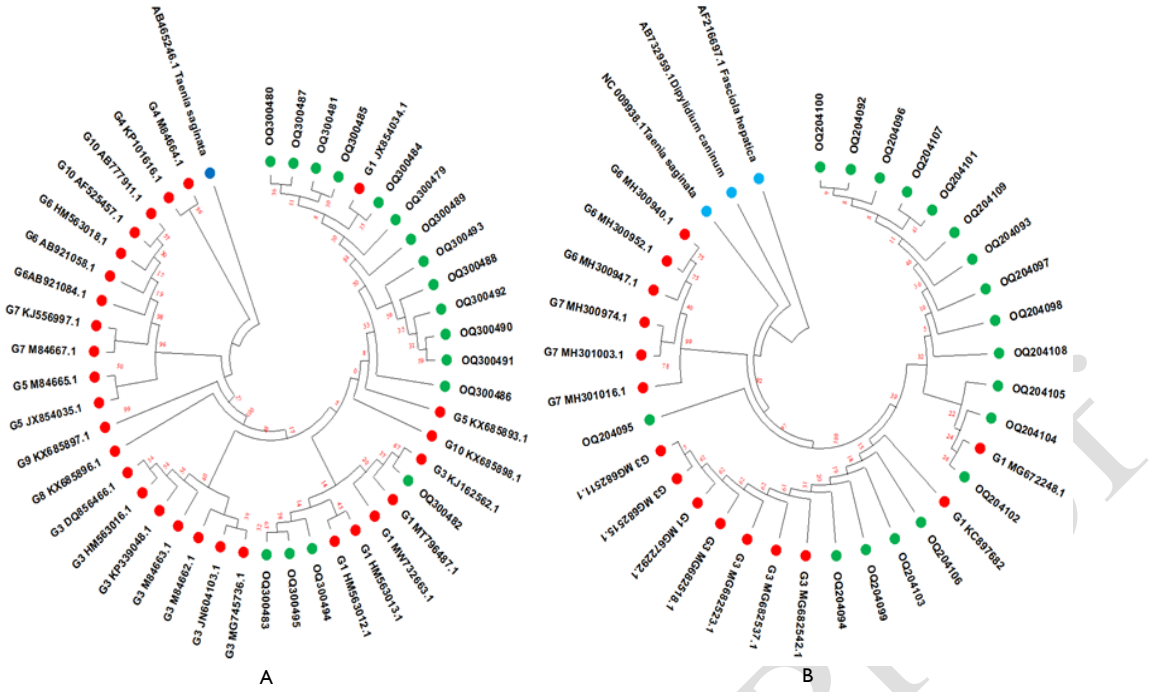

Fig. 1: Phylogenetic view of the *E. granulosus s.s.* genotype from Jordan (Green) and reference sequences of known genotypes G1-G10 (Red), employing partial *Cox1* (424 bp) (A) and partial *Nad2* (B) The analysis was based on ClustalW alignment and Neighbor-Joining method of tree construction with 1000 bootstrap iterations is shown, the tree's in-dependability was assessed. Each sequence has a name and an accession number. *Taenia saginata*, *Dipylidium caninum,* and *Fasciola hepatica* were used as an outgroup.

show the level of link with other species of similar or different lineage (Fig. 1B). The phylogenetic tree showed a significant clade of 13 Jordanian isolates with 89% mean intraspecific similarity, clustering with a cattle G1 genotype from Iran (accession number: MG672248). Isolate OQ204095 was classified into a paraphyletic clade near the G3 genotype cluster. Four more sequences (accession number: OQ204094, OQ204099, OQ204103, and OQ204106) were found between G1 and G3 reference sequences. Isolate OQ204103 exhibited the lowest average similarity comparison score (75.39%) among all sequences analyzed. The Jordanian sequences were categorized as *E. granulosus s.s.* (genotypes G1-G3).

Fig. (1) Phylogenetic view of the *E. granulosus s.s.* genotype from Jordan (Green) and reference sequences of known genotypes G1-G10 (Red), employing partial *Cox1* (424 bp) (A) and partial *Nad2* (B) The analysis was based on ClustalW alignment and Neighbor-Joining method of tree construction with 1000 bootstrap iterations is shown, the tree's in-dependability was assessed. Each sequence has a name and an accession number. *Taenia saginata*, *Dipylidium caninum* and *Fasciola hepatica* were used as an outgroup.

Polymorphisms and haplotype analysis: A total of 50 sequences were analyzed, comprising 470 alignment positions of Jordanian sequences, along with representative sequences from Lebanon, Iran, South America, India, Mongolia, Peru, and Iraq. The analysis focused on *Cox1* gene analog sequences spanning 469 base pairs. There were 12 unique locations with variant sites in the *Cox1* sequence (Fig. 2A). 750 alignment locations revealed 38 distinct *Nad2* nucleotide alterations. 10 haplotypes were discovered for *Cox1* and 24 haplotypes for *Nad2* among the *E. granulosus s.s.* strains.

Fig. 2B displays the haplotype network for the dominant genotype (G1) of *E. granulosus s.s.* based on *Cox1* sequences. Hap_1 was the most common haplotype found in isolates from all geographic regions. It held a central position in the network and represented 80% (40/50) of the total sequences analyzed. The mutational variations between Hap_1 and the other haplotypes ranged from one to six. The two Indian isolates were merged into a single haplotype (Hap_2) of a single mutation step from the central haplotype. Eight sequences produced distinct haplotypes, with six originating from Jordanian isolates displaying a star-shaped extension from the original haplotype (Hap_1). Hap_3 and Hap_4 are unique Iranian samples.

The haplotype diversity of *E. granulosus s.s.* based on *Cox1* was minimal (Hd = 0.3624 < 0.5%), and the nucleotide diversity was also low (π = 0.02 < 5%). Analysis of the partial *Nad2* gene identified 24 unique haplotypes as shown in Table 1. Each sample displayed a distinct haplotype, resulting in a high haplotype diversity value (Hd=1.0000), whereas the nucleotide diversity value was low (π = 0.01 < 5%).

Nucleotide sequence variation: Two sequence variation patches were observed in the multiple alignments of the partial *Cox1* and *Nad2* fragments (Fig. 3). The variance was primarily caused by nucleotide transversions and transitions. Table 2 outlines nucleotide substitution patterns in both mitogenes. 6 G/A and 19 T/C transitions were identified in the *Cox1* sequences, while 4 G/A and 26 T/C transitions were observed in *Nad2* sequences. Seven sites in *Cox1* exhibited a G/T transversion, while two sites showed an A/T transversion. The most common kind of mutation in *Nad2* was G/T transversions, with 23 occurrences, followed by 17 A/T, 4 G/C and just 1 A/C. Transition-to-transversion ratios (R) of *Cox1* sequences

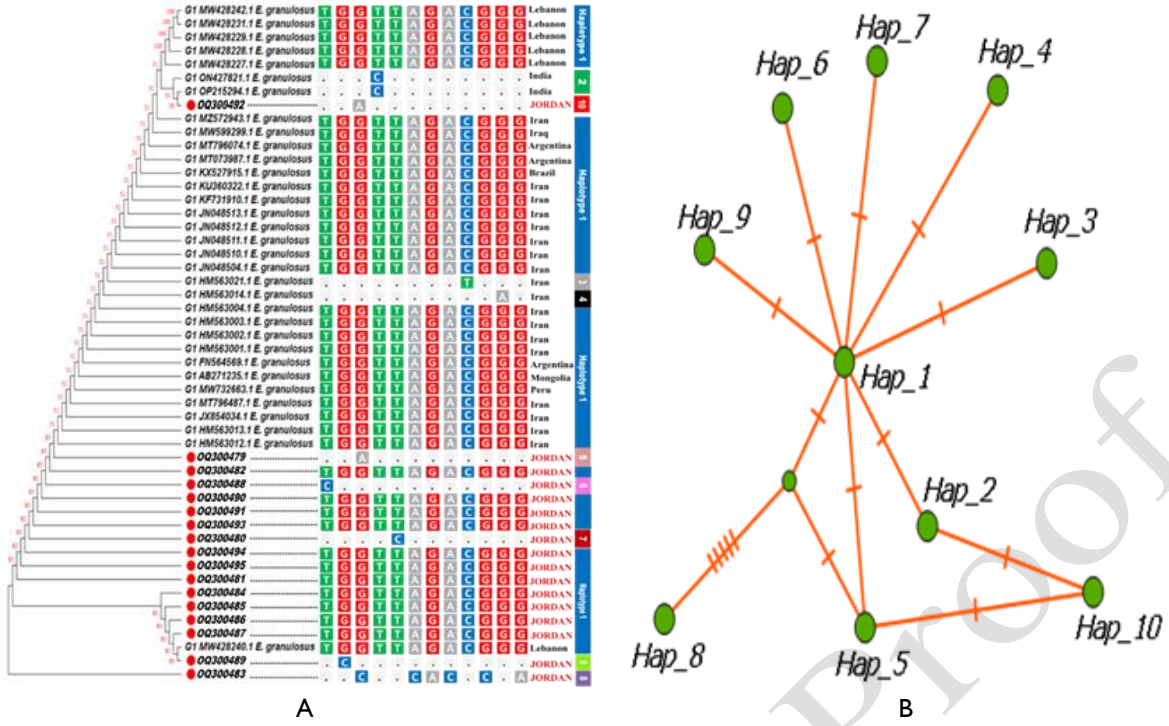

Fig. 2: *Echinococcus granulosus s.s.* phylogeography based on partial *Cox1* (469 bp). (A)The left panel shows the Maximum Likelihood tree of *E. granulosus* (G1) genotypes from Jordan and regional sequences from GenBank. On the right are the nucleotide base changes that set each variant haplotype apart in addition to their country of origin. (B) Haplotype network for the dominant genotype (G1) of *E. granulosus s.s.* using incomplete *Cox1* gene sequences. Forty isolates are included in Hap_1 (Jordan, 10; Lebanon, 6; Iran, 15; South America, 4; one each from India, Peru, Iraq, and Mongolia). Two isolates from India are included in Hap_2, Hap_3, Hap_4 are distinct Iranian isolates, and Hap5–Hap10 are Jordanian isolates. Vetting marks symbolize the number of mutations that distinguish haplotypes.

Table 1: Haplotype of the G1 based on *Nad2* partial sequences and the accession numbers of the isolates comprising the groups.

Haplotype	Isolate-ID	Accession Number
Hap1	GIKC897688	GIKC897688
Hap2	GIKC897686	GIKC897686
Hap3	GIKC897685	GIKC897685
Hap4	GIKC897684	GIKC897684
Hap5	GIKC897682	GIKC897682
Hap6	AF297617.1	AF297617.1
Hap7	NSNEG52	OQ204109
Hap8	NSNEG50	OQ204108
Hap9	NSNEG49	OQ204107
Hap10	NSNEG48	OQ204106
Hap11	NSNEG47	OQ204105
Hap12	NSNEG46	OQ204104
Hap13	NSNEG45	OQ204103
Hap14	NSNEG43	OQ204102
Hap15	NSNEG42	OQ204101
Hap16	NSNEG41	OQ204100
Hap17	NSNEG39	OQ204099
Hap18	NSNEG38	OQ204098
Hap19	NSNEG35	OQ204097
Hap20	NSNEG34	OQ204098
Hap21	NSNEG33	OQ204097
Hap22	NSNEG32	OQ204096
Hap23	NSNEG02	OQ204095
Hap24	NSNEG01	OQ204094

Table 2: highest composite likelihood was used to determine the pattern of nucleotide substitution for partial mitochondrial *Cox1* and *Nad2* genes in *Echinococcosis granulosus s.s.* The probability (r) of switching from one base (row) to another indicated in each entry (column). The rates of different transitional replacements are shown in bold and the rates of transversion substitutions are shown in *italics*.

	Cox1				*Nad2*			
	A	T	C	G	A	T	C	G
A	-	*3.3*	*0.59*	4.34	-	*5.75*	*0.8*	14.53
T	*1.21*	-	11.98	*2*	*1.85*	-	6.49	*2.64*
C	*1.21*	66.84	-	*2*	*1.85*	46.69	-	*2.64*
G	2.62	*3.3*	*0.59*	-	10.22	*5.75*	*0.8*	-

(*n*=50) was 3.58. The (R) value of *Nad2's* was 2.186. Calculations were made for single base changes at every codon position in the scoring region. The number of base substitutions at three positions in each codon was equal for *Cox1* and *Nad2* (Fig. 3). The sequences were mostly dinucleotide repeats. TT was detected in 46% and 50% of *Cox1* and *Nad2* gene sequences, respectively. Ten percent of the AA repeats occurred, whereas about 25% of the GG repeats did.

DISCUSSION

Ruminants in Jordan were found to carry the genotypes G1 and G3 of *E. granulosus*. According to Issa *et al.* (2018), G1 was found to be the most prevalent with "sheep–dog" and "sheep–camel" cycles (Issa *et al.*, 2018). According to this study, the G1 is most common in Jordan's southern regions, which is in line with findings from livestock studies carried out in Jordan's northern regions (Al-Qaoud *et al.*, 2003b). As a result, the G1 genotype is associated with areas in Jordan where there is significant sheep herding and has a variable local distribution. According to epidemiological research, echinococcosis is currently considered an endemic zoonotic disease in the Levant and other Middle Eastern countries where G1 is the main genotype (Joanny *et al.*, 2021).

Actually, it seems that the G1 is more prevalent in several countries, including Iran (Siyadatpanah *et al.*, 2019), Sardinia (Mehmood *et al.*, 2021); Bangladesh (Faruk *et al.*, 2017); China (Guo *et al.*, 2019) (Wang *et al.*, 2014); Nepal (Donadeu *et al.*, 2020); South America (i.e. Argentina, Brazil, Chile, Peru, Uruguay) (Cucher *et al.*, 2016). This revelation has significant public health

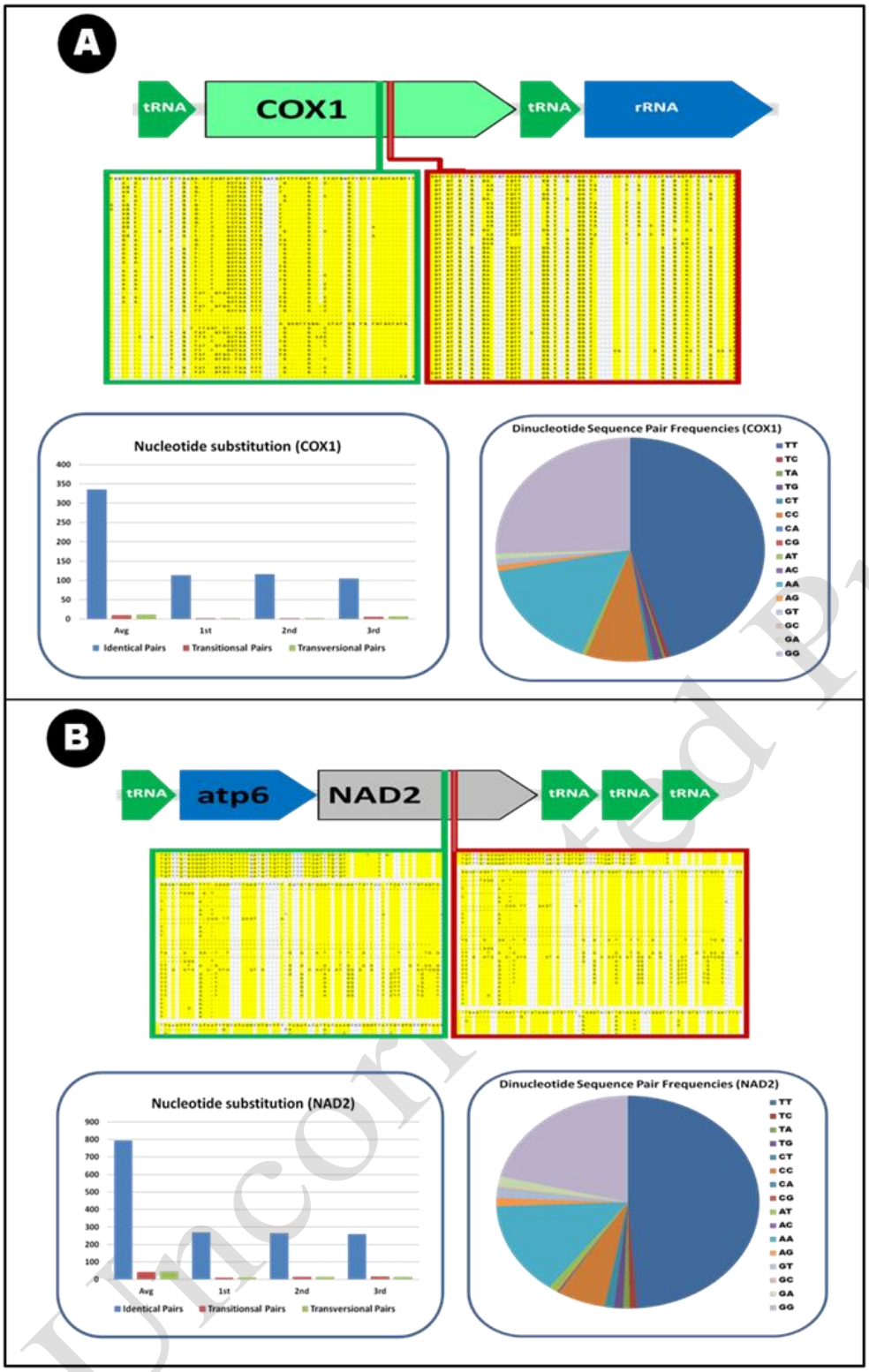

Fig. 3: Nucleotide sequence variation and multiple alignments of the *Echinococcus granulosus s.s.* -*Cox1* and *Nad2* fragments. (A) Location of variable regions in the targeted segment of the *Cox1* in the region of the nucleotide position between 9654 and 10078. (B) Location of variable regions in the targeted segment of the *Nad2* gene in the nucleotide position between 6172 and 7053. The overall frequency of base pair transitions in comparison to the frequency of base pair transversion at the three-codon sites is displayed for each mitogene along with the dinucleotide frequency. counting

implications because the G1 is considered the main source of human infection worldwide (Casulli *et al.*, 2022). Livestock husbandry poses a serious threat to human health in rural areas.

Applying two mitochondrial genes to this study allowed for successful analysis. The justification for the proposed methodology is that phylogenetic gene analysis has become more reliable with the use of phylogenetic markers in combination as opposed to a single marker. The phylogeny of the two genes sequenced in this study reached a consensus for classifying the Jordanian isolates studied as *E. granulosus s.s.* (genotypes G1–G3). However, *Nad2* showed higher resolution and differentiated between the G1 and G3 genotypes. Similar results have been reported for *E. granulosus s.s.* and other genotypes (Wang *et al.*, 2014; Laurimäe *et al.*, 2019; Samari *et al.*, 2022).

Haplotype analysis results for the partial *Cox1* and *Nad2* genes were obviously different. Nonetheless, research elucidating haplotype diversity makes clear that haplotypes defined by a particular gene may differ from those based on other mitochondrial genes (Busi *et al.*,

2007; Morgan-Richards *et al.*, 2017) thus our results should not be inconsistent with that information. In some instances, when the entire mitogenome is sequenced, the same haplotype reflects separate parasite variations (Busi *et al.*, 2007; Ohiolei *et al.*, 2019; Selcuk *et al.*, 2022a). Therefore, the length of the gene regions under investigation was a major factor in the haplotype disparity between the *Cox1* and *Nad2* in this study. Previous *E. granulosus* genotyping studies have focused on analyzing 469 bp segments of the *Cox1* gene. Sequencing wider gene regions of *Nad2* resulted in the identification of a greater number of haplotypes.

Based on *Cox1*, a major haplotype (Hap_1) can be identified, which is significant because *Cox1* is the primary phylogenetic barcode of the tapeworm *E. granulosus.* Less genetic variation was found between isolates from the several geographic areas that this study was focused on. G1 sequences from Argentina, Peru, Mongolia, Iran, Iraq, Jordan, and Lebanon were included in Hap_1. Understanding this pattern is essential to comprehending how humans become infected. If this be the most common haplotype in the sheep cycle, it would be beneficial to carry out a comparable investigation on *E. granulosus s.s.* from different regions of the country. It would be also interesting to investigate human samples in subsequent research to see if this haplotype is most common in human infections.

Currently, data on the haplotype diversity of *E. granulosus s.s.* in Jordan are limited. A previous study by Yanagida *et al.* (2012) supported the occurrence of the common haplotype in Jordan and Iran based on the complete *Cox 1* gene sequence (1609 bp) (Yanagida *et al.*, 2012). They discovered that isolates from Iran and Peru belonged to the same haplotype group as those from Jordan. These findings are in agreement with our results because the major haplotype group in this study also contained one Peruvian sequence and 16 sequences from Iran.

The increased frequency of the Hap_1 haplotype across the broad geographic region supports the idea that gene flow and a lack of genetic differentiation have historically been characteristic of *E. granulosus s.s.* in many Mediterranean and European locations (Kinkar *et al.*, 2016; Tamarozzi *et al.*, 2020; Bonelli *et al.*, 2021). An array of evolutionary processes explains this phenomenon. These haplotypes, for instance, may have started in one location before spreading to others (e.g., by human transport of dogs and other host animals). It is probable that this tapeworm has recently spread given the pattern of closely related haplotypes and very low nucleotide diversity values.

It has previously been proposed that the occurrence of a common haplotype could be explained by selective pressure (Selcuk *et al.*, 2022b), which could affect the spread of the common haplotype. The haplotypes we observed are not likely to have undergone a profitable mutation. Given the known phylogenetic relationships between the sequences, it can be inferred that the founder's amino acid sequence corresponds to that of the other minor haplotypes, indicating a low degree of genetic variation among the isolates from the examined regions. According to the parsimony network analysis, the *Cox1* haplotypes (except for Hap_10) displayed a star-like expansion from the main founder haplotype, suggesting that the populations in Jordan are not entirely distinct from one another. Furthermore, the mutational analysis of *Cox1* revealed a relatively high transition/transversion bias ($R = 3.581 > 2.0$; $n = 50$), indicating a lack of genetic heterogeneity amongst the local populations, whereas the average ratio in coding locations typically does not surpass three.

Understanding the divergence thresholds at which different transitions saturate nucleotide sites can help us understand how parasite evolve. When sequences start to diverge, the ratio of transition to transversion substitutions is at its highest; as divergence time develops, it subsequently falls. We found that the transition/ transversion ratio for the *Nad2* gene was much lower than for *Cox1* suggesting that the mutations in this gene are not saturated and appropriate for studying the genetic variation of *E. granulosus s.s.* However, because of its greater R-value, *Cox1* seems to be able to infer haplotype diversity better than *Nad2*. This bias is partly due to the increased frequency of methylation cytosine mutations in thymine. Our study revealed this trend (see Table 2) since the potential (r) for the shift from methylation cytosine to thymine was substantially larger. The predilection for nonsynonymous transversions may contribute to coding region bias. Both transitions and transversions can alter the amino acid composition of the corresponding protein; however, biochemical differences in the protein products are usually greater for transversions. The likelihood of purifying selection against transversions increases, particularly if the transversion occurs at the third codon position. In our refinement analysis of the first, second, and third codon positions, we found that the number of base substitutions in *Nad2* and *Cox1* was identical at the three positions. In contrast, it has been discovered that the underlying chemistry of this mutation favors transitions over transversions in Asian Taenia tapeworms (Eom *et al*., 2002).

In this work, we sought to ascertain if the locations of one or more mutations in *Nad2* and *Cox1* fragments tended to be more frequently found in specific regions of the gene or if they were dispersed randomly along the span of the chosen fragments. Rather than mutations spreading across the fragment, both genes had alterations in a few specific areas. Given that genes may naturally be predisposed to mutation, these areas could be hotspots for mutations (Galtier *et al.*, 2006). Moreover, these mutational sections co-occurred between various genotypes used for comparison and between isolates of the same genotype (G1 in this study). Hence, it was postulated that these mutational sites contribute to haplotype divergence and polymorphisms. Multiple studies have reported differences in mitogenomic nucleotide sequences among parasitic helminths (Jeon *et al.*, 2011; Metwally *et al.*, 2019). Most of the sequences were composed of dinucleotide repeats. TT repeats were the most common in both *Cox1* and *Nad2*, followed by slightly fewer GC repeats. This low GC content compared to that of other repeats is in accordance with insect and helminth mitochondrial DNA (Clare *et al.*, 2008; Abe *et al.*, 2021; Cantara *et al.*, 2022).

Conclusion: The most prevalent genotype of the tapeworm *E. granulosus s.s.* discovered in isolates from

southern Jordan was G1. The partial *Cox1* and *Nad2* genes showed varying degrees of resolution between the G1 and G3 genotypes. In our mitogenome study, we deduced the evolutionary path of the G1 genotype. Along with South America and several Asian countries, the Levant region encompasses Jordan and Lebanon appears to be prominent in the genetic subgroup of G1 seen in the (Hap_1) haplotype, which is related to the *Cox1* gene. Considering its prominent position in the haplotype network and its co-occurrence in the Levant, where domestication originated, it is conceivable that this haplotype represents an ancient variation in the parasite. This discovery has significant epidemiological implications because it coincides with some of the main hotspots of hydatidosis worldwide. Mutations in the mitogenomes of *Cox1* and *Nad2* have been linked to two small regions rather than widely distributed and are believed to account for genetic and haplotype diversity. We recommend the use of the full-length *Nad2* gene in future studies to better discriminate *E. granulosus s.s.* G1 and G3 genotypes.

Acknowledgements: We gratefully acknowledge Dr. Saqer Abu-Shattal for his assistance in Figures productions.

Ethics approval and consent to participate: Not applicable

Consent for publication: All of the authors have approved the contents of this paper and have agreed to the submission policies.

Competing Interests: We declare that we have no competing interest

Authors' contributions

Enas Al-khlifeh: Provided concepts and ideas, designed the study, defined the intellectual content, collected and analyzed the data, search the literature, produced figures and wrote the manuscript.

Ayed Alshammari: Search the literature and revising the manuscript.

Hussein Al Nasrat: Analyzed the data, produced figures and revising the manuscript.

All authors have reviewed and approved the manuscript.

Funding: Funding provided by Al-Balqa Applied University- Deanship of scientific research (DSR) (grant number DSR-2022#511).

Availability of data and materials: Data available on request.

REFERENCES

Abe N, Baba T, Nakamura Y, *et al.*, 2021. Global analysis of cytochrome c oxidase subunit I (cox1) gene variation in Dibothriocephalus nihonkaiensis (Cestoda: Diphyllobothriidae). Curr Res Parasitol Vector Borne Dis 1:100042.

Al-Qaoud KM, Craig PS and Abdel-Hafez SK, 2003a. Retrospective surgical incidence and case distribution of cystic echinococcosis in Jordan between 1994 and 2000. Acta Trop 87:207–214.

Al-Qaoud KM, Abdel-Hafez SK and Craig PS, 2003b. Canine echinococcosis in northern Jordan: increased prevalence and dominance of sheep/dog strain. Parasitol Res 90:187–191.

Alvarez Rojas CA, Romig T and Lightowlers MW, 2014. Echinococcus granulosus sensu lato genotypes infecting humans--review of current knowledge. Int J Parasitol 44:9–18.

Bonelli P, Dei Giudici S, Peruzzu A, *et al.*, 2021. Identification of Echinococcus granulosus Genotypes G1 and G3 by SNPs Genotyping Assays. Pathogens 10:125.

Bowles J, Blair D and McManus DP, 1994. Molecular genetic characterization of the cervid strain ('northern form') of Echinococcus granulosus. Parasitology 109 (Pt 2):215–221.

Busi M, Snábel V, Varcasia A, *et al.*, 2007. Genetic variation within and between G1 and G3 genotypes of Echinococcus granulosus in Italy revealed by multilocus DNA sequencing. Vet Parasitol 150:75–83.

Cantara A, Luo Y, Dobrovolná M, *et al.*, 2022. G-quadruplexes in helminth parasites. Nucleic Acids Res 50:2719–2735.

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

Casulli A, Siles-Lucas M and Tamarozzi F, 2019. Echinococcus granulosus sensu lato. Trends in Parasitol 35:663–664.

Casulli A, Massolo A, Saarma U, *et al.*, 2022. Species and genotypes belonging to Echinococcus granulosus sensu lato complex causing human cystic echinococcosis in Europe (2000–2021): a systematic review. Parasites & Vectors 15:109.

Clare EL, Kerr KCR, von Königslöw TE, *et al.*, 2008. Diagnosing mitochondrial DNA diversity: applications of a sentinel gene approach. J Mol Evol 66:362–367.

Craig PS, McManus DP, Lightowlers MW, *et al.*, 2007. Prevention and control of cystic echinococcosis. Lancet Infect Dis 7:385–394.

Cucher MA, Macchiaroli N, Baldi G, *et al.*, 2016. Cystic echinococcosis in South America: systematic review of species and genotypes of Echinococcus granulosus sensu lato in humans and natural domestic hosts. Tropical Med & Int Health 21:166–175.

Donadeu M, Gauci CG, Poudel I, *et al.*, 2020. A hyperendemic focus of porcine cystic echinococcosis in the Banke District of Nepal. Acta Tropica 201:105203.

Eckert J and Deplazes P, 2004. Biological, Epidemiological, and Clinical Aspects of Echinococcosis, a Zoonosis of Increasing Concern. Clin Microbiol Rev 17:107–135.

Eom KS, Jeon H-K, Kong Y, *et al.*, 2002. Identification of Taenia asiatica in China: Molecular, Morphological, and Epidemiological Analysis of a Luzhai Isolate. J Parasitol 88:758–764.

Faruk MO, Siddiki AM a. MZ, Masuduzzaman M, *et al.*, 2017. Identification and molecular characterization of Echinococcus granulosus from domestic goat in Chittagong, Bangladesh. Trop Biomed 34:831–840.

Galeh TM, Spotin A, Mahami-Oskouei M, *et al.*, 2018. The seroprevalence rate and population genetic structure of human cystic echinococcosis in the Middle East: A systematic review and meta-analysis. Int J of Surgery 51:39–48.

Galtier N, Enard D, Radondy Y, *et al.*, 2006. Mutation hot spots in mammalian mitochondrial DNA. Genome Res 16:215–222.

Gao Y, Wang W, Lyu C, *et al.*, 2021. Meta-Analysis of the Prevalence of Echinococcus in Sheep in China from 1983 to 2020. Frontiers in Cellular and Infection Microbiology 11.

Guo B, Zhang Z, Zheng X, *et al.*, 2019. Prevalence and Molecular Characterization of Echinococcus granulosus Sensu Stricto in Northern Xinjiang, China. Korean J Parasitol 57:153–159.

Han X, Jian Y, Zhang X, *et al.*, 2019. Genetic characterization of Echinococcus isolates from various intermediate hosts in the Qinghai-Tibetan Plateau Area, China. Parasitology 146:1305–1312.

Hijjawi NS, Al-Radaideh AM, Rababah EM, *et al.*, 2018. Cystic echinococcosis in Jordan: A review of causative species, previous studies, serological and radiological diagnosis. Acta Trop 179:10–16.

Himsawi N, Hijjawi N, Al-Radaideh A, *et al.*, 2019. Seroprevalence of cystic echinococcosis in a high-risk area (Al-Mafraq Governorate) in Jordan, using indirect hemagglutination test. Parasite Epidemiol Control 5:e00104.

Issa HS, Abdel-Hafez SK, Hijjawi NS, *et al.*, 2018. Molecular Characterization of Echinococcus granulosus sensu stricto Cysts of Domestic Ruminants in Jordan. 11.

Jeon H-K, Yong T-S, Sohn W-M, *et al.*, 2011. Molecular Identification of Taenia Tapeworms by Cox1 Gene in Koh Kong, Cambodia. Korean J Parasitol 49:195–197.

Joanny G, Mehmood N, Dessì G, *et al.*, 2021. Cystic echinococcosis in sheep and goats of Lebanon. Parasitology 148:871–878.

Kamhawi S, 1995. A retrospective study of human cystic echinococcosis in Jordan. Ann Trop Med Parasitol 89:409–414.

Kinkar L, Laurimäe T, Simsek S, *et al.*, 2016. High-resolution phylogeography of zoonotic tapeworm Echinococcus granulosus sensu stricto genotype G1 with an emphasis on its distribution in Turkey, Italy and Spain. Parasitology 143:1790–1801.

Kinkar L, Laurimäe T, Sharbatkhori M, *et al.*, 2017. New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm Echinococcus granulosus sensu stricto. Infection, Genet and Evolution 52:52–58.

Kinkar L, Laurimäe T, Acosta-Jamett G, *et al.*, 2018a. Global phylogeography and genetic diversity of the zoonotic tapeworm Echinococcus granulosus sensu stricto genotype G1. Int J for Parasitol 48:729–742.

Kinkar L, Laurimäe T, Acosta-Jamett G, *et al.*, 2018b. Distinguishing Echinococcus granulosus sensu stricto genotypes G1 and G3 with confidence: A practical guide. Infect Genet Evol 64:178–184.

Larrieu E, Gavidia CM and Lightowlers MW, 2019. Control of cystic echinococcosis: Background and prospects. Zoonoses and Public Health 66:889–899.

Laurimäe T, Kinkar L, Romig T, *et al.*, 2019. Analysis of nad2 and nad5 enables reliable identification of genotypes G6 and G7 within the species complex Echinococcus granulosus sensu lato. Infect Genet Evol 74:103941.

Meadows JRS, 2014. Sheep: Domestication. In: Encyclopedia of Global Archaeology (Smith C, ed). Springer: New York, NY, pp:6597–6600.

Mehmood N, Muqaddas H, Arshad M, *et al.*, 2020. Comprehensive study based on mtDNA signature (nad1) providing insights on Echinococcus granulosus s.s. genotypes from Pakistan and potential role of buffalo-dog cycle. Infect Genet Evol 81:104271.

Mehmood N, Dessì G, Ahmed F, *et al.*, 2021. Genetic diversity and transmission patterns of Echinococcus granulosus sensu stricto among domestic ungulates of Sardinia, Italy. Parasitol Res 120:2533–2542.

Mehmood N, Muqaddas H, Ullah MI, *et al.*, 2022. Genetic structure and phylogeography of Echinococcus granulosus sensu stricto genotypes G1 and G3 in Pakistan and other regions of the world based on nad5 gene. Infection, Genet and Evolution 98:105223.

Metwally DM, Al-Enezy HA, Al-Turaiki IM, *et al.*, 2019. Gene-based molecular characterization of cox1 and pnad5 in Hymenolepis nana isolated from naturally infected mice and rats in Saudi Arabia. Biosci Rep 39:BSR20181224.

Miman Ö, Aycan ÖM, Aydın C, *et al.*, 2010. What Should Be The Concentration of Eosin to Qualification of İdeal Staining for Viability Determination on Hydatid Cyst ? Turkish Bulletin of Hygiene and Experimental Biol 67:21–26.

Monge-Maillo B, Olmedo Samperio M, Pérez-Molina JA, *et al.*, 2019. Osseous cystic echinococcosis: A case series study at a referral unit in Spain. PLoS Negl Trop Dis 13:e0007006.

Morgan-Richards M, Bulgarella M, Sivyer L, *et al.*, 2017. Explaining large mitochondrial sequence differences within a population sample. Royal Society Open Science 4.

Moses AM, Chiang DY, Kellis M, *et al.*, 2003. Position specific variation in the rate of evolution in transcription factor binding sites. BMC Evol Biol 3:19.

Nakao M, Li T, Han X, *et al.*, 2010. Genetic polymorphisms of Echinococcus tapeworms in China as determined by mitochondrial and nuclear DNA sequences. Int J Parasitol 40:379–385.

Ohiolei JA, Xia C-Y, Li L, *et al.*, 2019. Genetic variation of Echinococcus spp. in yaks and sheep in the Tibet Autonomous Region of China based on mitochondrial DNA. Parasites & Vectors 12:608.

Romig T, Ebi D and Wassermann M, 2015. Taxonomy and molecular epidemiology of Echinococcus granulosus sensu lato. Vet Parasitol 213:76–84.

Romig T, Deplazes P, Jenkins D, *et al.*, 2017. Ecology and Life Cycle Patterns of Echinococcus Species. Adv Parasitol 95:213–314.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, *et al.*, 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol and Evolution 34:3299–3302.

Saitou N and Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Samari H, Laurimäe T, Reghaissia N, *et al.*, 2022. Molecular characterization of Echinococcus granulosus sensu lato genotypes in dromedary camels from extreme Sahara of Algeria based on analysis of nad2 and nad5 genetic markers. Acta Tropica 234:106616.

Selcuk MA, Celik F, Kesik HK, *et al.*, 2022a. In Silico Evaluation of the Haplotype Diversity, Phylogenetic Variation and Population Structure of Human *E. granulosus* sensu stricto (G1 Genotype) Sequences. Pathogens 11:1346.

Selcuk MA, Celik F, Kesik HK, *et al.*, 2022b. In Silico Evaluation of the Haplotype Diversity, Phylogenetic Variation and Population Structure of Human E. granulosus sensu stricto (G1 Genotype) Sequences. Pathogens 11:1346.

Siyadatpanah A, Anvari D, Emami Zeydi A, *et al.*, 2019. A systematic review and meta-analysis of the genetic characterization of human echinococcosis in Iran, an endemic country. Epidemiol Health 41:e2019024.

Sun C-H, Liu H-Y, Xu N, *et al.*, 2021. Mitochondrial genome structures and phylogenetic analyses of two tropical characidae fishes. Front Genet 12.

Tamarozzi F, Legnardi M, Fittipaldo A, *et al.*, 2020. Epidemiological distribution of Echinococcus granulosus s.l. infection in human and domestic animal hosts in European Mediterranean and Balkan countries: A systematic review. PLOS Neglected Trop Dis 14:e0008519.

Tamura K, Stecher G and Kumar S, 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022–3027.

Uchiumi L, Mujica G, Araya D, *et al.*, 2021. Prevalence of human cystic echinococcosis in the towns of Ñorquinco and Ramos Mexia in Rio Negro Province, Argentina, and direct risk factors for infection. Parasites & Vectors 14:262.

Vuitton DA, McManus DP, Rogan MT, *et al.*, 2020. International consensus on terminology to be used in the field of echinococcoses. Parasite 27:41.

Wang J, Wang N, Hu D, *et al.*, 2014. Genetic diversity of Echinococcus granulosus in southwest China determined by the mitochondrial NADH dehydrogenase subunit 2 gene. ScientificWorldJournal 2014:867839.

Yanagida T, Mohammadzadeh T, Kamhawi S, *et al.*, 2012. Genetic polymorphisms of Echinococcus granulosus sensu stricto in the Middle East. Parasitol Int 61:599–603.

Yang Z, 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314.