ISSN: 03875547
Volume 45, Issue 01, February, 2022

Molecular diagnosis and genotyping of cystic echinococcosis from clinical samples of human in central and northern Iraq

Mohannad Hamid Jasim[1*], Abdulkhaliq Alwan Mohemeed[2], Muntaha M. Hasan Al-alouci[3]

Department of biology, college of education for pure sciences, Tikrit university, Iraq[1,2]
Department of microbiology, college of medicine, Anbar university, Iraq[3]

Corresponding Author: 1*

Keywords:
Hydatid cyst, Echinococcus granulosus, Genotype

ABSTRACT

The study included 45 samples of hydatid cysts isolated from humans for the period from April 2021 to October 2021 from hospitals in central and northern Iraq. This study was carried out at Almusaib bridge for scientific and lab equipments, The study aimed at the molecular diagnosis of *Echinococcus granulosus*, the identification of the causative strains of hydatid cyst disease using the polymerase chain reaction technique, the study of genetic sequencing of the strains, and the drawing of the phylogenetic tree to know the convergence of the strains causing the infection in the study area with the strains spread in the rest of the world , DNA was extracted from the protoscolices and the germ layer and amplified by the polymerase chain reaction technique. The COX1 gene with a molecular weight of 444bp was amplified. By analyzing the sequence of nucleotides for 17 samples of the cox1 gene and analyzing the phylogenetic tree and comparing it with the gene bank, it was found that the sheep strain (G1) is the strain that causes infection in humans in central and northern Iraq, The observed variations were respectively deposited in the NCBI-bank it database under the accession numbers (OL614464 to OL614480) to represent S1, to S17 samples respectively.

1. INTRODUCTION

Hydatidosis, which is caused by the larval stage of the worm *Echinococcus granulosus* and several other medically important species such as *E. multilocularis*, is one of the most important zoonotic diseases in the world and one of the most important animal diseases that infect humans and other animals [1].

This disease results in many medical, economic, and veterinary problems. It also represents a health problem that affects rural and poor communities, where livestock breeders are in close contact with dogs, which are the final hosts of the parasite, in addition to contact with domestic animals such as sheep, cows, and buffaloes, which are intermediate hosts, causing economic losses. As a result of damaging the vital organs, especially the liver, which causes great economic losses, In humans, the clinical symptoms are due to the pressure of the hydatid cyst on the organs within it, causing disruption of their functions and impeding their growth. Fatal

injuries resulting from this disease have also been recorded due to the explosion of the shock-causing sac, and some vital organs, including the brain, are difficult to surgically interfere with, so life may be threatening or fatal [2]. Hydatid disease spreads all over the world, especially in agricultural countries whose citizens raise sheep and cows, such as the countries of Southeast Asia and the Middle East, including the countries of the Arabian Peninsula, North Africa and South America, where there are large numbers of dogs that are carriers of the parasite mostly. Represents the middle hosts [3]. According to genetic markers in the nucleus and mitochondria, the parasite that causes hydatid cyst has a level of genetic diversity, as many types of strains of this parasite have been described in recent years [4]. These strains were identified using molecular affinity. According to research, there are ten different strains of the hydatid cyst parasite (G1 - G10) that differ from one another in characteristics of pathogenesis, epidemiology, control, and infection. The strains were identified using DNA sequences and genetic material analyses [5].

2. Material and method

Forty five hydatid cysts of human collected for the period from april 2021 to October 2021, samples were obtained after performing surgical operations in the liver and lung organs in hospitals in central and northern Iraq governorates, including Al-Ramadi Teaching Hospital, Fallujah Teaching Hospital, Al-Razi Specialized Hospital in Anbar Governorate, medical city and Al-Karkh National Hospital in Baghdad Governorate, among others. After collecting the hydatid cysts, the external surface was sterilized, and exam the hydatid fluid under microscope to determine fertility of cyst through present or absent of protoscolices, the protoscolices were separated and stored with ethyl alcohol at a 70% concentration until DNA extraction. A genomic Dna extraction kit (Geneaid, Taiwan) has been used to extract Dna from each sample, and nanodrop was used to analyze the concentration and purity of the extracted Dna sample. Until employed pcr to extract DNA that was kept at -20 °C. The genus of the hydatid cyst was determined using a primer, as well as the genotypes of human origin based on the DNA sequence of the Cytochrome oxidase 1 (COX1) gene following gene amplification using the PCR method. The initiator used in this study was designed using the following method [6]: The initiator used in this study was given by the Korean company Macrogen and was based on the NCBI-Gene bank data base site for the purpose of getting the nucleotide sequence of the gene used in the study pcr amplification was prepard by add 20µl including 5 µl of dna templete , 1 µl (10 pmol) of each forword: 5’ - TTT TTT GGG CAT CCT GAG GTT TAT - 3’ and reverse : 5’ -TAA AGA AAG AAC ATA ATG AAA ATG- 3’ primers , and 13µl of Free nuclease water to the pcr tube of accupower pcr premix kit (bioneer , Korea) which contain other pcr reaction requirements (dNTPs, KCl, MgCl2, Taq DNA polymerase, tris HCl pH: 9.0, tracking dye and stabilizer), The cox1 gene 444bp segment was amplified under the following conditions: initial denaturation at 95 °C for 3 min., followed by 35 cycles of denaturation at 95 °C for 30 sec., annealing at 56°C for 35 sec., elongation at 72°C for 55 sec. followed final extention at 72°C for 5 min. and holding at 4°C. Using agarose gel electrophoresis, the pcr product of the cox1 gene (444bp) was analyzed.

2.1 SEQUENCING METHODS

2.1.1 Nucleic acids Sequencing of PCR amplicons

The resolved PCR amplicons were commercially sequenced from forward directions, following to instruction manuals of the sequencing company (Macrogen Inc. Geumchen, Seoul, South Korea). Only clear chromatographs obtained from ABI (Applied Biosystem) sequence files were further analyzed, ensuring that the annotation and variations are not because of PCR or sequencing artifacts. By comparing the observed nucleic acid sequences of local samples with the retrieved nucleic acid sequences, the virtual positions, and other details of the retrieved PCR fragments were identified.

2.1.2 Interpretation of sequencing data

The sequencing results of the PCR products of the targeted samples were edited, aligned, and analyzed as long as with the respective sequences in the reference database using BioEdit Sequence Alignment Editor Software Version 7.1 (DNASTAR, Madison, WI, USA). The observed variations in each sequenced sample were numbered in PCR amplicons as well as in its corresponding position within the referring genome. The observed nucleic acids were numbered in PCR amplicons as well as in their corresponding positions within the referring genome. Each detected variant within the tapeworm sequences was annotated by SnapGene Viewer ver. 4.0.4 The observed variations were respectively deposited in the NCBI-bank it database under a unique accession number for each analyzed sample.

2.1.3 Comprehensive phylogenetic tree construction

A specific comprehensive tree was constructed in this study to unravel the possible effect of the detected single nucleic acid substitution on the phylogenetic positioning of the investigated samples. The observed variants were compared with their neighbour homologous reference sequences using the NCBI-BLASTn server [7]. Then, a full inclusive tree, including the observed variant, was built by the neighbour-joining method and visualized as a circular cladogram using the iTOL suit [8].

One specific PCR fragment partially covering the coding portion of the mitochondrial *COX1* sequences in human-infecting tapeworms was utilized in this study. The amplified fragments were directly exposed to direct sequencing experiments to assess the pattern of genetic polymorphism in the samples obtained from seventeen human hosts. Then, a specific comprehensive tree was built to assess the sort of phylogenetic distribution of the observed variants with their relative sequences.

3. Result and discussion

3.1 The results of the polymerase chain reaction technique for the cox1 gene

The results of this technique showed the success of the process of amplifying the DNA extracted from the protoscolices and the germ layer of the hydatid cyst of the cox1 gene, and this is in agreement with many studies, including [9]. Figure (1) shows the positive results of DNA samples that were amplified for the cox1 gene extracted from hydatid cysts obtained from surgical operations in humans after being transferred onto an agarose gel.

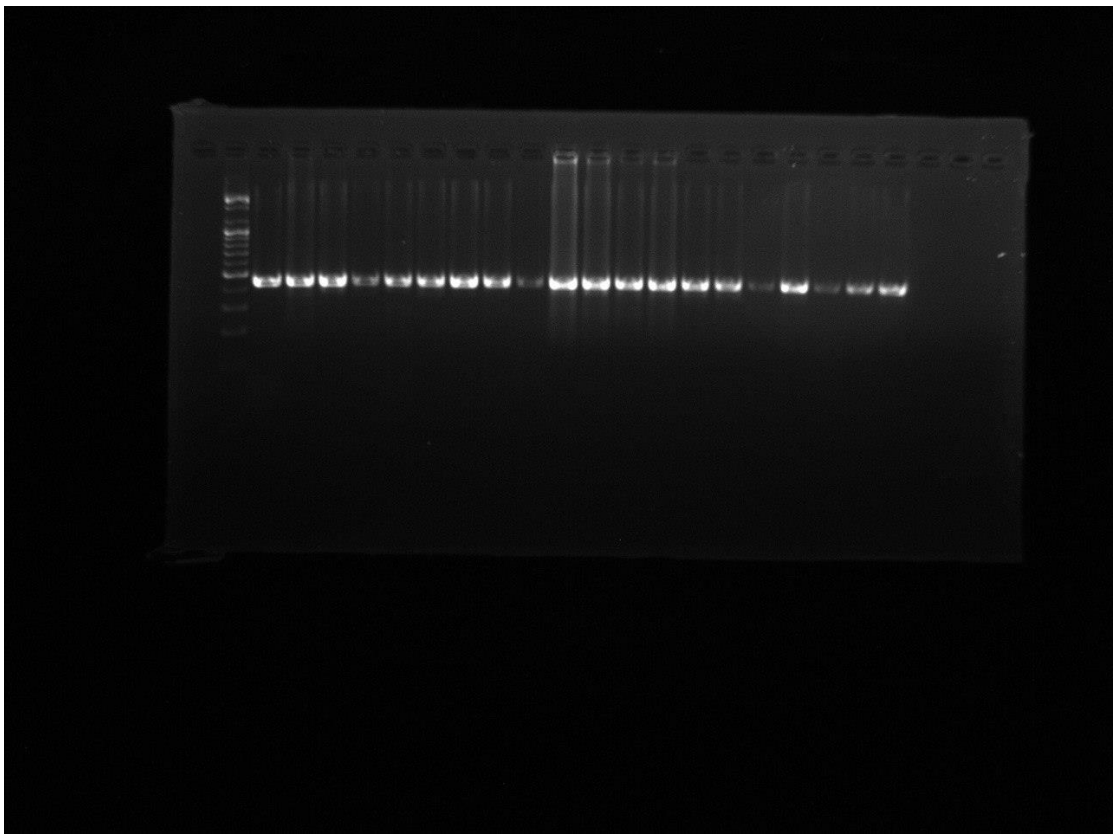

Figure (1) representing electrophoresis on agarose gel to amplify the COX1 gene extracted from the DNA of the *Echinococcus granulosus* parasite from hydatid cysts in humans.

The results of the analysis of the DNA sequence extracted from the hydatid cysts indicated that the sheep strain (G1) is responsible for human infections in the study area, and this is agree with the study in Alanbar governorate [10] and the results of the study also agree with [11] in northern Iraq, as it was found that the sheep strain (G1) is the most common cause of human infections, and the results of the study also agree with study [12], as it was found that the sheep strain (G1) is responsible for human infection in northern Iran. All of what was mentioned is due to the presence of the intermediate hosts represented by sheep in the study area as they constitute the host The only intermediate for sheep strain (G1).

While it differs with the study [13] whose study showed that the sheep strain (G1) and the buffalo strain (G3) are the strains responsible for human infection in the Muthanna Governorate, and this is due to the difference in the intermediate hosts in the study area, as there are no buffaloes in central and northern Iraq, and disagree Also with the study [14], whose study showed that the sheep strain (G1) and the buffalo strain (G3) are the strains responsible for human infection, but the percentage of sheep strain (G1) is the most widespread and the same result was reached [15] in Iran as well [16]. In Algeria, this diversity in breeds is due to the diversity of intermediate hosts.

3.2 SEQUENCING RESULTS

Our results indicated the presence of two nucleic acid variants in some of the investigated samples, namely C>T92 in S1, S3, and S14 and C>T147 in S15. Whereas the other tapeworm samples did not exhibit any nucleic acid alteration(s). Results inferred from the direct nucleic acid translation indicated that C>T92 showed a missense variant (p.Ala>Val267), while the other substitution of C>T147 showed a silent mutation (p.Ser286=) on the investigated COX1-encoded protein. Phylogenetic results showed the presence of several

major and minor clades of *Echinococcus granulosus*. Within these clades, our investigated samples showed three sorts of positioning within three distinct phylogenetic clades. It was inferred from the tree that both observed C>T92 and C>T147 variants induced noticeable alterations in the phylogenetic positioning of humans-infecting tapeworm samples respectively. However, C>T92 has induced a phylogenetic variation to a greater extent than that induced from C>T147. This was due to the distant positions that S1, S3, and S14 samples occupied with respect to the non-variable tapeworm samples and S15 sample respectively. However, such dramatic alteration of C>T92 was apparently not correlated with the geographical source from which these samples were collected because the majority of investigating humans-infecting samples have the same Asian source of *Echinococcus granulosus* that are deposited in the NCBI database. However, our investigated samples were suited in the immediate vicinity to variable geographical strains of *Echinococcus granulosus*.

Within this locus, seventeen samples were included in the present study. These samples were screened to partially amplify *COX1* sequences of the infecting tapeworms. Thus, the variation of the *COX1* sequences can be used for discrimination among various types of *Echinococcus granulosus* due to the possible roles of these sequences to adapt variable diversity in different host-infecting situations. The sequencing reactions indicated the exact identity of these infectious sequences after performing NCBI blastn for their PCR amplicons [7]. Concerning the 444 bp amplicons, the NCBI BLASTn engine showed up to 99% sequences similarities between the sequenced samples and the intended reference target sequences. By comparing the observed nucleic acid sequences of these investigated samples with the retrieved nucleic acid sequences (GenBank acc. CBLN010004653.1), the accurate positions and other details of the retrieved PCR fragments were identified. The total length of the targeted locus was localized in the NCBI server, and the positions of the start and end of the targeted locus were also confirmed within the most homologous tapeworm target.

After positioning the 444 bp amplicons' sequences within the genomic sequences of the tapeworm, the details of its sequences were highlighted, and the total length of the amplified fragments was also.

Interestingly, the alignment results of the 444 bp samples revealed the presence of two nucleic acid variations represented by two nucleic acid substitutions in comparison with the most similar referring reference nucleic acid sequences. These nucleic acid substitutions were found to be distributed specifically in the investigated tapeworm samples. Sequencing reactions revealed that one variant (C>T92) was only detected in the samples S1, S3, and S13. Another variant (C>T147) was only detected in the S15 sample. Whereas the other *Echinococcus granulosus* samples (S2, S4 – S13, S16, and S17) did not show any nucleic acid variation in comparison with the COX1 referring sequences (GenBank acc. no. CBLN010004653.1).

Both observed variants of C>T92 and C>T147 were not found in the corresponding reference sequences. To confirm these variations, the sequencing chromatograms of the investigated samples, as well as their detailed annotations, were verified and documented, and the chromatograms of their sequences were shown according to their positions in the PCR amplicons. The presence of each variant was confirmed in its original chromatogram and the absence of any possible technical error was also confirmed.

The observed nucleic acid variations were further analyzed to identify whether such substitutions induce possible alteration in their corresponding positions in the cytochrome c oxidase subunit I protein. All nucleic acid sequences of S1 to S17 were translated to their corresponding amino acid sequences using the Expasy translate suite. Amino acid alignment of these amino acid sequences with their references showed that these two variants exhibited two different effects on the cytochrome c oxidase subunit I protein. Results showed that the C>T92 induced a missense (nonsynonymous) effect on the translated nucleic acid sequences in the amplified fragment in position no. 31 (A>V31). Whereas the C>T147 variant exhibited a silent (synonymous)

effect on the translated nucleic acid sequences in the amplified fragment in position no. 31 (S50=). However, these missense and silent variants were exemplified and repositioned in the entire cytochrome c oxidase subunit I protein sequences to be p.Ala>Val267 and p.Ser286= respectively.

However, these alterations may be developed by the invading organism as an adaptation to drugs that are directed toward its targeted cytochrome c oxidase subunit I protein. The observed variations were respectively deposited in the NCBI-bank it database under the accession numbers (OL614464 to OL614480) to represent S1, to S17 samples respectively.

To give a phylogenetic understanding of the actual distances between our investigated samples and their relative sequences, a comprehensive phylogenetic tree was conducted in the present study according to nucleic acid variations observed in the amplified 444 bp of the *COX1* sequences amplicons. This phylogenetic tree was contained the currently analyzed samples (S1 to S17) alongside other relative nucleic acid sequences of *Echinococcus granulosus* sequences. Within this tree, our investigated samples were incorporated alongside relative sequences to constitute only one organism of incorporated sequences within the cladogram, which is *Echinococcus granulosus*. The total number of the aligned nucleic acid sequences in this comprehensive tree was 106. In the constructed cladogram, the investigated samples were clustered into several major and minor phylogenetic clades within the *Echinococcus granulosus* sequences. The most interesting observation in our investigated tapeworm isolates is correlated with the positioning of these samples and their neighbour sequences of the *Echinococcus granulosus* into several major clades. Within the biggest major clade of the tree, the majority of samples (S2, S4 – S13, S16, and S17) samples of *Echinococcus granulosus* sequences were incorporated (Fig. 2). Such a sort of positioning of these samples was due to the absence of any detectable nucleic acid variation compared with the reference sequences of the *COX1* gene. However, these samples were suited in the immediate vicinity to several sequences that were deposited from Kyrgyzstan (such as GenBank acc. no. MN787560.1), Algeria (such as GenBank acc. no. MG672282.1), Turkey (such as GenBank acc. no. KU925393.1), and China (such as GenBank acc. no. MH050613.1). Therefore, the multi-original sources of these biological samples were determined. Meanwhile, it was found that the S15 sample was suited beside this major clade but with a slight tilt. The reason for this tilt is attributed to the presence of the C>T92 variant that was induced such phylogenetic alteration. It was found that S15 was positioned in the vicinity of GenBank acc. no. of MG808284.1 that was deposited from Algeria. So, the presence of C>T92 may have a potential phylogenetic role. However, a larger phylogenetic role was found from the other observed variant, C>T147 that was detected in S1, S3, and S14. This variant induced much more alteration in the positioning of these samples, which enrolled a larger phylogenetic role than that found in samples having C>T92. However, these three samples were positioned in the vicinity to several sequences that belonged to several strains isolated from Kyrgyzstan (such as the GenBank acc. no. MN787561.1) and Algeria (such as the GenBank acc. no. MG808347). Whatever the source from which these samples were isolated, it was found that all the samples having C>T147 were occupied a further phylogenetic position than that found in the sample having C>T92. This necessitates greater importance of C>T147 in inducing noticeable phylogenetic alterations. This sort of positioning indicated a more remarkable role for the observed C>T147 variant in inducing a noticeable deviation with respect to the original positioning these tapeworm samples occupied within this cladogram. However, the multiple sources of all investigated samples were confirmed by other neighbouring sequences since it was these samples were deposited from different countries worldwide.

ISSN: 03875547
Volume 45, Issue 01, February, 2022

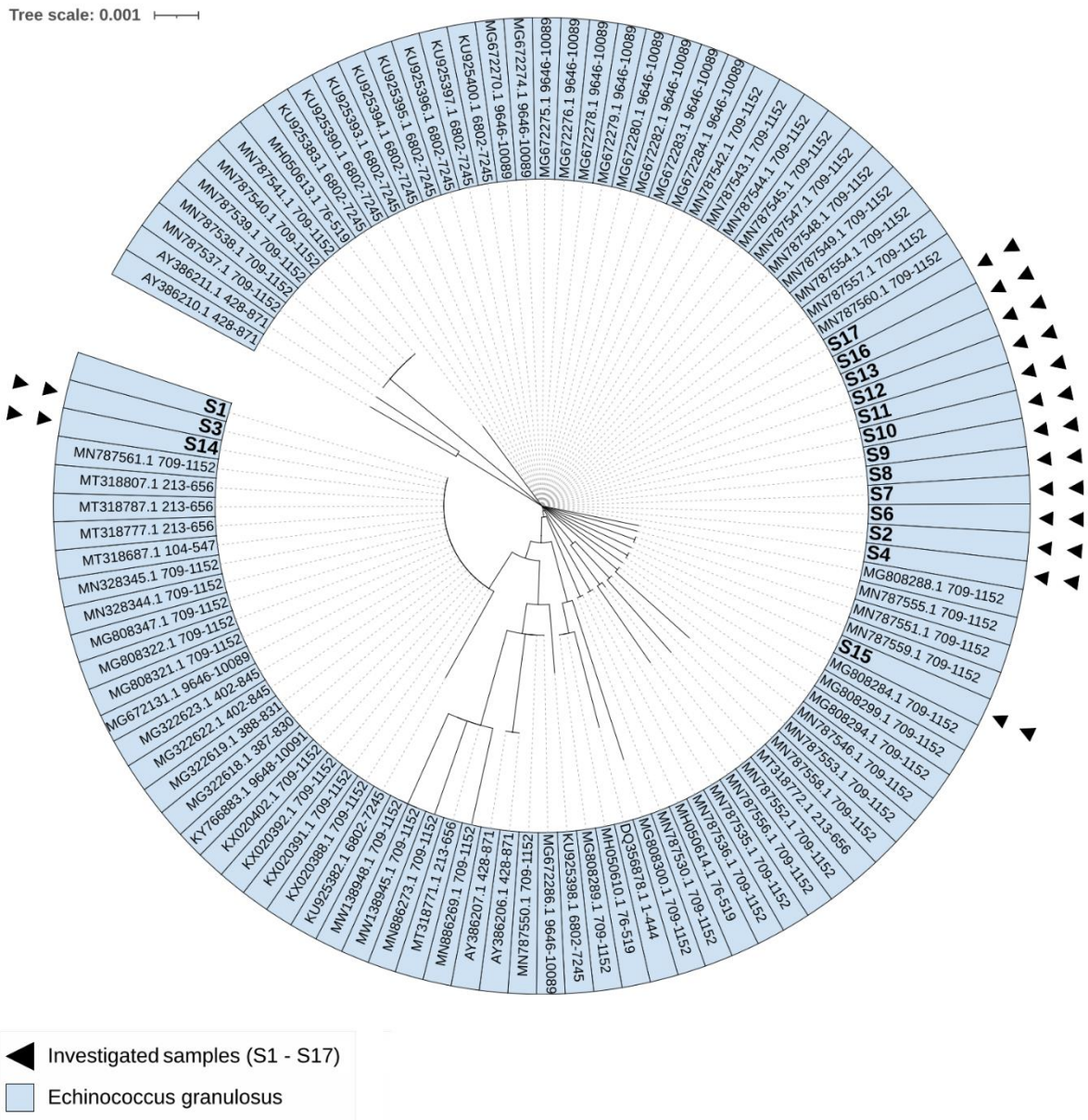

Fig. 2. The comprehensive cladogram phylogenetic tree of genetic variants of the *COX1* sequences fragment of seventeen *Echinococcus granulosus* samples. The black-colored triangle refers to the analyzed tapeworm variants. All the mentioned numbers referred to GenBank accession number of each referring species. The number "0.001" at the top portion of the tree refers to the degree of scale range among the comprehensive tree categorized organisms. The letter letters "S", followed by a number refers respectively to the investigated humans-infecting *Echinococcus granulosus* sequences

However, the aggregation of all investigated tapeworm samples in small phylogenetic distances may refer to the presence of close patterns of the phylogenetic distribution among the observed four clades. The current observation of this tree has confirmed sequencing reactions because it explained the actual neighbour-joining-based positioning in such observed variations. The importance of the medical implementation of *Echinococcus granulosus* has superseded other biological concerns, such as the effect of isolation source of the nucleic acid variations of these infectious organisms. This fact can be found in the tree because the majority of the incorporated organisms were found to have belonged to various strains isolated from human sources. Though a highly noticeable role of both observed variations was observed in this study in the

analyzed *COX1* sequences, no species deviation was observed among the incorporated tapeworm organisms within the tree.

Noteworthy, the utilization of the *COX1* sequences in this study has given further indication for the presence of the precise identification of the actual identity of these tapeworm organisms. Consequently, these pieces of evidence are in line with each other to support our observation of the divergence of these infectious pathogenic tapeworm sequences from variable sources. It was found that the investigated tapeworm samples were occupied variable positions within different major clades of the cladogram with few phylogenetic distances. This indicated the presence of high genetic homology among the tapeworm sequences of these strains. However, this *COX1* sequences-based comprehensive tree has provided an inclusive tool about the high ability of such genetic fragments to efficiently identify tapeworm genotypes using their *COX1* fragment. This, in turn, gives a further indication of the ability of the currently utilized *COX1* sequence-specific primers to describe the investigated *Echinococcus granulosus* tapeworm and their accurate phylogenetic positions. For this reason, the high ability of such genetic fragments to efficiently identify tapeworm sequences was proved by using these *COX1* sequences-based fragments. This, in turn, gives a further indication of the ability of the currently utilized mitochondrial sequence-specific primers to describe the currently investigated *Echinococcus granulosus* and their accurate phylogenetic positions.

4. Conclusions

Through the current study, it was found that the sheep strain (G1) is the strain that causes human infection in central and northern Iraq, and the cox1 gene is good from a diagnostic aspect as well as distinguishing polymorphism within the species.

5. References

[1] Lundström-Stadelmann, B., Rufener, R., Hemphill, A., 2020. Drug repurposing applied: activity of the anti-malarial mefloquine against Echinococcus multilocularis. Int. J. Parasitol. Drugs Drug Resist. 13, 121 – 129.

[2] Tamarozzi, F., Deplazes, P., Casulli, A., 2020. Reinventing the wheel of Echinococcus granulosus sensu lato transmission to humans. Trends Parasitol. 36 (5), 427-434.

[3] Casulli, A, 2020 Recognising the substantial burden of neglected pandemics cystic and alveolar echinococcosis. Lancet Glob. Health 8 (4), e470-e471.

[4] Terefe, Y.; Addy, R.N.F.; Alemu, S.; Mackenstedt, u.; Romig, T.; Wassermann, M. 2019, Genetic characterization of Echinococcus species in eastern Ethiopia. Vet. Parasitol. Reg. Stud. Rep., 17, 100302 .

[5] Vuitton, D.A., McManus, D.P., Rogan, M.T., Romig, T., Gottstein, B., Naidich, A., Avcioglu, A., 2020. International consensus on terminology to be used in the field of echinococcoses. Parasite 27, 41.

[6] Ai-Shimary, Zahraa Mohammed Salih Mahdi, 2019, Molecular study of Echinococcus granulosus for Hydatid Cysts its Organ some intermediate host and using the aqueous extracts for Annona aquamosa, Master thesis, College of scince for Women, University of Babylon, 118P.

[7] Zhang, L. H., Joshi, D. D., & McManus, D. P. (2000). Three genotypes of Echinococcus granulosus identified in Nepal using mitochondrial DNA markers. Transactions of the Royal Society of Tropical Medicine and Hygiene., 94(3): 258-260.

[8] Letunic I, Bork, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2;47(W1): W256-W259.

[9] Rahi, A. A., and Ali, M. A. (2016). A First DNA sequencing of hydatid agent isolated from human in Iraq. Journal of Pure and Applied Microbiology., 10(2): 1015-1020.

[10] Muhaidi, M. J., 2014, Molecular study on Cystic Echinococcosis in some Iraqi patients, Ph.d thesis, College of Science, University of Baghdad, 169 P.

[11] Hama, A.A ; Mero, W.M. and Jubrael, J.M. (2012). Molecular Characterization of Echinococcus granulosus, First Report of Sheep Strains in Kurdistan- Iraq. In: International Confreence on Ecological, Environmental and Biological Sciences (EEBSTM 2012)., 41-44pp.

[12] Gorgani-Firouzjaee, T; Kalantrai, N; Ghaffari, S; Alipour, J.,and Siadati, S. (2019). Genotype characterization of livestock and human cystic echinococcosis in Mazandaran province, Iran. Journal of helminthology., 93(2): 255-259.

[13] Al-Rishawi, K. M. Ali, 2019, Molecular detection of Echinococcus granulosus strains causative of human and some farm animals hydatidosis in Al-Muthana province, Master thesis , Collage of Education , Al-Qadisiyah Uni., 137 P.

[14] Pezeshki, A ; Akhlaghi, L; Sharbatkhori, M; Razmjou, E; Oormazdi, H;Mohebali, M., and Meamar, A. R. (2013). Genotyping of Echinococcus granulosus from domestic animals and humans from Ardabil Province, Northwest Iran. Journal of Helminthology, 87(4): .387-391.

[15] Khademvatan , S ; Majidiani , H ; Foroutan , M; Tappeh, K. H; Aryamand, S. and Khalkhali, H. R. (2019). Echinococcus granulosus genotypes in Iran: a systematic Review. Journal of helminthology, 93(2): 131-138.

[16] Zait, H; Kouidri, M; Grenouillet, F. E; Umhang, G; Millon, L., Hamrioui, B. and Grenouillet, F. (2016). Molecular characterization of Echinococcus granulosus sensu stricto and Echinococcus canadensis in humans and livestock from Algeria. Parasitology research, 115(6): 2423-2431.