Deanship of Graduate Studies

Al-Quds University

Development of PCR systems for detection of Taeniidae family species.

George Samir Farah Kokaly

Reg. No. 21710041

M.Sc. Thesis

Jerusalem-Palestine

1441/ 2020

Deanship of Graduate Studies

Al-Quds University

Development of PCR systems for detection of Taeniidae family species.

Prepared By: George Samir Farah Kokaly

Reg. No. 21710041

BSc. Medical Laboratory Science /2016

Arab American University in Jenin.

Supervisor: Dr. Ibrahim Abbasi

Co-supervisor: Dr. Rasmi Abu Helu

A thesis submitted in partial fulfillment of requirement for the Master degree in Medical Laboratory science / Diagnostic microbiology and Immunology Track/ Deanship of Graduate Studies / Al-Quds University.

Al – Quds University

Deanship of Graduate Studies

Medical Laboratory Science

Thesis Approval

Development of PCR systems for detection of Taeniidae family species.

Prepared By: George Samir Farah Kokaly

Reg. No. 21710041

Supervisor: Ibrahim Abbasi, PhD

Co-supervisor: Rasmi Abu Helu, PhD

Master thesis submitted and accepted Date: **3/6/2020**

Names and Signatures of Examining Committee Members are as follow:

1. **Head of committee**: Dr. Ibrahim Abbasi **Signature:**
2. **Co-supervisor:** Dr. Rasmi Abu Helu **Signature:**
3. **Internal Examiner:** Prof. Sameer Al Barghouthi **Signature:**
4. **External Examiner:** Dr. Majdi Dwikat **Signature:**

Jerusalem – Palestine

1441/ 2020

Dedication

I dedicate my thesis work to my family, fiancée and many friends. A special feeling of gratitude to my loving parents and sisters whose words of encouragement did not stop until I have finished this work. I will appreciate and be thankful to my many colleagues who have supported me throughout the process.

Declaration

I certify that this thesis submitted for the degree of master in Medical Laboratory Science/ Diagnostic microbiology and immunology track is the result of my own research accept were otherwise cited, and this thesis is not to be submitted for a higher degree to any other University or institution.

George Samir Farah Kokaly

15-4-2020

ACKNOWLEDGMENT

I would like first to thank the God on how he is giving me the strength and knowledge to reach my goals and achieve my ambition.

I would like to express my deep and sincere gratitude to my research supervisor, Dr. Ibrahim Abassi and co-supervisor Dr. Rasmi Abu Helu for giving me the opportunity to do research and providing invaluable guidance throughout this research. They have taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under their guidance. I am extremely grateful for what they have offered me. Also I would like to express my thank to Prof. Sameer Al-Barghouthi our program coordinator who always give me advices and his experience in this field.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my sisters and fiancée for their love, understanding and continuing support to complete this research work. Also I express my thanks to my friends and colleagues in work for their support and valuable prayers.

Abstract:

Echinococcosis, caused by the larval stage of tapeworms of the genus Echinococcus, is one of the most important zoonotic diseases worldwide. *Echinococcus granulosus* and *Echinococcus multilocularis* are the most prevalent species infecting humans, resulting in cystic echinococcosis (CE) and alveolar echinococcosis (AE), respectively. In this study we investigated the presence of E. *granulosus*-specific DNA in dog's feces by detecting the 18s r-DNA and cytochrome C oxidase I (cox1) mitochondrial gene, beside a repetitive *E. granulosus* DNA segment.

From the total of seven PCR systems that were designed to amplify *E. granulosus* DNA that could be found in infected dog fecal samples only three systems were proved to be effective for this purpose. The three PCR systems were adapted first to amplify DNA from other cestode species of *Taenidae* family to be used in next generation DNA sequence analysis.

The details of the PCR systems were as follow: PCR system 1 and 2 were based on universal primers from 18s rDNA gene, and amplified a DNA segment of 224bp and 216 bp respectively. PCR 3 based on *E. granulosus* repetitive region and amplified a region of 152 bp. PCR sensitivity test of the three candidate PCR systems was done, the sensitivity limit was 0.0025 ng/ml of *E.granulosus* genomic DNA.

The three PCR systems were used to detect the presence of *E. granulosus* in 50 dog fecal samples collected from Yatta town, the produced amplicons were subjected to NGS MiSeq DNA analysis. A bioinformatics workflow was developed through Galaxy/europe online software for identifying *E. granulosus* specific sequences from thousands of obtained sequences for each sample. PCR system 1 proved to be the most effective and it detected target DNA in 12 from a total of 50 fecal samples (24%). This study will support future research that should reveal a better understanding of the Echinococcus-host interplay, and suggests new avenues for the identification of additional targets for diagnosis and chemotherapy.

الملخص:

يعتبر داء المشوكات ، الذي تسببه مرحلة اليرقات من الديدان الشريطية من جنس المكورات، أحد أهم الأمراض الحيوانية في العالم. وتعتبر الطفيليات *Echinococcus granulosus* و *Echinococcus multilocularis* هما من أكثر الأنواع انتشارًا من التي تصيب البشر، مما يؤدي إلى داء المشوكات الكيسي (CE) والمكورات المتعددت الأكياس (AE)، على التوالي. في هذه الدراسة ، تحققنا من وجود الحمض النووي الخاص بالطفيلي في براز الكلب من خلال الكشف عن جينات 18s rDNA, COX1 في الميتوكوندريا، إلى جانب شريحة الحمض النووي المتكرر من طفيل *E. granulosus*.

من مجموع ثمانية أنظمة PCR التي تم تصميمها لتضخيم الحمض النووي للطفيل *E. granulosus* التي يمكن العثور عليها في عينات البراز الكلاب المصابة ثبت أن ثلاثة أنظمة فقط فعالة لهذا الغرض. تم تكييف أنظمة PCR الثلاثة أولاً لتضخيم الحمض النووي أيضاً من أنواع أخرى من cestode تنتمي إلى عائلة Taenidae وثانيًا ليتم استخدامها في الجيل التالي من تحليل تسلسل DNA.

كانت تفاصيل أنظمة PCR على النحو التالي: اعتمد نظام PCR 1 و 2 على بادئات عامة من جين 18 rDNA وتضخيم جزء DNA من 224 bp و 216 bp على التوالي، و PCR الثالث يعتمد على المنطقة المتكررة للطفيل *E. granulosus* وتضخيم المنطقة من 152bp. تم إجراء اختبار حساسية PCR لأنظمة PCR الثلاثة المرشحة، وكان حد الحساسية يصل إلى 0.0025 نانوغرام DNA/ مل E.granulosus.

تم استخدام أنظمة PCR الثلاثة للكشف عن وجود *E. granulosus* في 50 عينة من براز الكلاب التي تم جمعها من بلدة يطا، وتم إخضاع القطع المضاعفة لتحليل NGS MiSeq DNA. تم تطوير سير عمل المعلوماتية الحيوية من خلال برنامج Galaxy/europe لتحديد تسلسل *E. granulosus* المحدد من خلال كشفه من ضمن ألف التسلسلات التي تم الحصول عليها لكل عينة. نظام PCR 1 أثبت بأنه الأكثر فاعلية وكشف عن 12 عينة مصابة من إجمالي 50 عينة براز (24%) خضته لهذا الفحص. نتوقع أن تقوم هذه الدراسة بدعم الأبحاث المستقبلية التي ستكشف عن فهم أفضل للتفاعل بين مضيفات *Echinococcus* ، وتستخدم كأدوات جديدة لتحديد أهداف إضافية للتشخيص والعلاج الكيميائي.

Table of Contents

List of tables

List of figures

List of Abbreviation

AE	alveolar echinococcosis
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CDC	Center for disease control
CE	Cystic echinococcosis
cox1	Cytochrome C oxidase subunit 1
CT	computerized tomography
DC	Dendritic cell
DNA	Deoxyribonucleic acid Deoxyribonucleic acid
DNTPs	Deoxynucleoside triphosphates
EDTA	Ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
IFN- γ	Interferon gamma
IgE	Immunoglobulin E
IGF	Insulin-like growth factor
IL-12	Interleukin 12
MOH	Ministry of health
mt-DNA	Mitochondrial
NCBI	National Center for Biotechnology Information
NGS	Next generation sequence
PAIR	Percutaneous aspiration, injection of chemicals and re-aspiration
PCR	Polymerase chain reaction
TAE	Tris-acetate EDTA
Th1	T helper cells

1. Introduction:

Cystic echinococcosis (CE) is one of the world's major zoonotic diseases, which is characterized by the development of large cysts in liver, lungs and other organs of humans and livestock. The infection is caused by the dog tapeworm, *E. granulosus*, in humans and animals (Heidari, Z. et al., 2019). Echinococcus, is one of the most important zoonotic parasitic diseases worldwide and in the Mediterranean region much due to extensive home slaughtering and presence of numerous stray dogs with access to offal (Heidari, Z. et al., 2019), (Gottstein et al., 1992). *Echinococcus granulosus* and *Echinococcus multilocularis* are the most prevalent species infecting humans, resulting in cystic echinococcosis (CE) and alveolar echinococcosis (AE), respectively (Heidari, Z. et al., 2019). *E. granulosus* is known to be endemic in all continents, while *E. multilocularis* limited to the northern hemisphere. Both AE and CE are considered neglected zoonosis, with a global distribution and higher prevalence for CE, but a higher pathogenicity and mortality for AE, especially in Asia (Heidari, Z. et al., 2019).

Herbivores are the intermediate hosts for *E. granulosus* , and canids, including dogs, wolves, foxes and jackals, act as definitive hosts and they hosting the adult worms in their small intestine. Humans is considered an accidental dead-end intermediate host for Echinococcus species and he get infected via close contact with the definitive host or by indirect ingestion of eggs through contaminated water or vegetables and herbs (Heidari, Z. et al., 2019), (Jacey Roche Cerda et al., 2018). Alveolar echinococcosis is of great public health importance because it is associated with high morbidity and mortality rate, and is therefore considered to be the most pathogenic zoonosis in temperate and arctic regions (Eckert, Johannes, 2004). The parasite is endemic in central and western Europe, parts of the near East, Russia, and the central Asian Republics, China northern Japan and Alaska (Eckert, Johannes, 2004).

During the past decades, molecular studies, mainly based on mitochondrial genes, have described several genotypes or species within *E. granulosus*, as follows: *E. granulosus* (Genotypes G1–G3), *E. equinus* (G4), *E. ortleppi* (G5), *E. canadensis* (G6–G10) and E. felidis ('lion strain'); the existence of the humanspecifc genotype G9 is controversial (Sharma, M., et al., 2013). Recently, (Kinkar, L., et al, 2018) showed that G1 and G3 are two distinct mitochondrial genotypes and can be considered as a single species of *E. granulosus* and G2 is more related to G3 genotype (Sharma, M., et al., 2013). Also it was confirmed that based on six nuclear loci, G6/G7 and G8/G10 genotypes can be considered as two distinct species (Laurimäe, 2018)Additionally, (Thompson et al., 2002) proposed to consider camel and pig strains of *E. granulosus* as a single species (*E. intermedius*). Two new species have recently been identified, *E. shiquicus* in small mammals from the Tibetan plateau and *E. felidis* in African lions, but their zoonotic transmission potential is unknown (Moro, P. L., et al., 2008). Unlike *E. granulosus* , there is limited variation in *E. multilocularis* isolates, and only minor variations have been detected in the cytochrome c oxidase subunit 1 (cox1) and other site in mitochondrial DNA sequences (Sharma, M., et al., 2013). This little variation produces two groups, namely M1 and M2. M1 originates in China and North America and M2 in Europe (Jabbar, Abdul et al., 2011).

1.1 Prevalence of cystic echinococcosis in Palestine

Hydatidosis, of which 95% is cystic echinococcosis (CE), is widely spread in all Mediterranean basin countries with incidence determined in humans at surgery ranging from 0.28 in France to 15 cases per 100,000 inhabitants in Tunisia. The disease is highly endemic in the whole of the Mediterranean Basin and adjacent countries including Greece, Turkey, Syria, Palestine, Lebanon, Jordan, Egypt, Libya, Morocco, Tunisia and Algeria (Craig, et al., 2007). Cystic echinococcosis (CE) is still one of the most important zoonotic diseases in the Middle East (Craig, et al., 2007). In some of the mentioned countries, both rural and urban communities exhibit high transmission rates to humans. This probably results in part from uncontrolled stray-dogs populations with access to offal, lack of meat inspection and the persistence of extensive home slaughtering such in some of the Palestinian areas.

In Palestine, studies on *E. granulosus* are lacking. However, a total of 390 surgically confirmed CE cases were found in the records of surgical hospitals of the West-Bank for an 8-year period 1990-1997. The overall mean annual surgical incidence was 3.1 per 100,000. The highest annual surgical incidence (16.8 per 100,000) was found in Yata town near Hebron (Abu-Hasan, N., et al., 2002). These data, however, are limited to certain villages and most likely represent minimal values.

In Palestine, hydatidosis is a reportable disease appearing in the Ministry of Health annual report. A study in 2002 revealed the sero-prevalence of CE among school children in Palestine to be 2.4% (Adwan et al., 2013). A study in the Nagab area in and around the city of Rahat showed that sero-prevalence among human subjects was 1.5 per 100,000 persons (Youngster et al., 2002). The genotyping of isolated *Echinococcus granulosus* strains is of paramount importance due to difference in life cycles, parasite transmission patterns, host susceptibility to different genotypes, different clinical picture in terms of the size of cyst and severity as well as geographical distribution of each genotype. (Youngster et al., 2002) (Table 1).

Table 1: Human Echinococcus according to MOH in Palestine:

Year	**Frequency (Number of cases)**
1999	6
2000	4
2001	4
2002	4
2003	13
2004	10
2005	15
2006	9
2007	27
2007	17
2009	9

1.2 Pathology and disease symptoms

Infected human being by hydatidosis may go on for months or years without showing any signs or symptoms of infection. The disease is considered a chronic and cysts in human organs may persistent for years without noticing, others cysts may rupture spontaneously or due to trauma and disappear entirely (Almulhim et al., 2020). If a cyst continues growing, patients may exhibit symptoms gradually due to the cyst making pressure on the surrounding tissue. Sudden signs and symptoms are more likely due to rupture rather than the growth of the cyst. Rupture of the cyst can induce an IgE antibody mediated hypersensitivity reaction which considered as life-threatening (Almulhim et al., 2020).

The nature of signs and symptoms varies depending on the site of the cyst, which is most common in the liver and lung, but can affect other sites including bones, spleen, central nervous system, and the heart (kammerer and Schantz, 1993). Most primary infections consist of one cyst, but 20 to 40% of patients display multiple organ involvement (kammerer and Schantz, 1993). While infections can be acquired in childhood, most causes of hydatidosis present later in life due to the slow growing nature of cysts, which is approximately about 1cm/year (Almulhim et al., 2020). However, cysts in the brain or eye can cause symptoms early on and thus mostly present in childhood. Common findings in liver hydatidosis include abdominal pain, decreased appetite, hepatomegaly, a palpable mass, abdominal distention. Common findings in lung hydatidosis include chronic cough, chest pain, and shortness of breath (Almulhim et al., 2020).

1.3 **Morphology of *E. granulosus* adult stage**

E. granulosus adult stage is about 1-2 cm in length that composed from 3-4 segments including scolex. As many other cestode species and specifically those belong to Taenidae family, adult worms have cyclophylidean type of scolex with four cup like suckers and two rows of circular arranged hooks (armed scolex). *E. granulosus* worm contain the following strobila segments: neck region, immature, and mature segments. In some worms, it is possible still to see the gravid segment (proglottid) that harbor the viable eggs (figure 1).

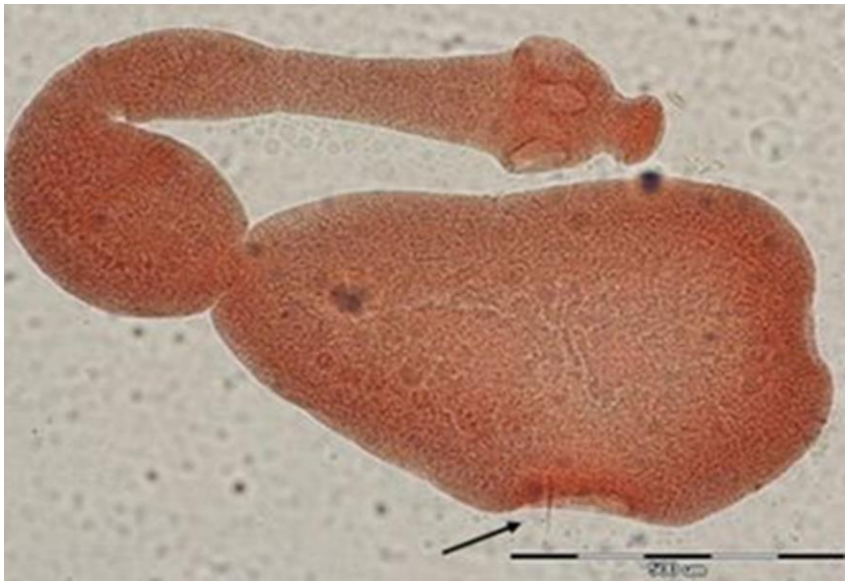


Figure 1: Adult of *E. granulosus* showing the gravid proglottid, mature and immature proglottid, and the anterior armed scolex. (Scale-bars: 500 μm, (Heidari, Z. et al., 2019).

1.4 *E. granulosus* Life Cycle

E. granulosus life cycle involves dogs as final hosts and herbivores as definitive hosts, while the life cycle of *E. multilocularis* involves several carnivores such as foxes, coyotes, dogs and cats as definitive hosts, and rodents as intermediate hosts (CDC, 2012). The exact stages of *E. granulosus* are depicted in the figure 2 that can be summarized in the following points:

1- *Echinococcus granulosus* adult worm resides in the small intestine of the definitive host, and release infectious eggs that are passed in the dogs' feces.

2- Eggs are ingested by a suitable intermediate host, they hatch in the small intestine and release six-hooked oncospheres that penetrate the intestinal wall and migrate into various organs, such as liver and lungs.

3- In the infected organs, the oncosphere develops into a thick-walled hydatid cyst that enlarges gradually, producing protoscolices and daughter cysts that fill the cyst interior.

4- If the infected cyst-containing organs from the intermediate host were eaten by the definitive host it becomes infected and protoscolices from the cyst stage will develop into adult stages within 32 to 80 days.

5- Human being considered as accidental intermediate hosts, and become infected by ingesting eggs that liberate oncospheres which develops into hydatid cysts.

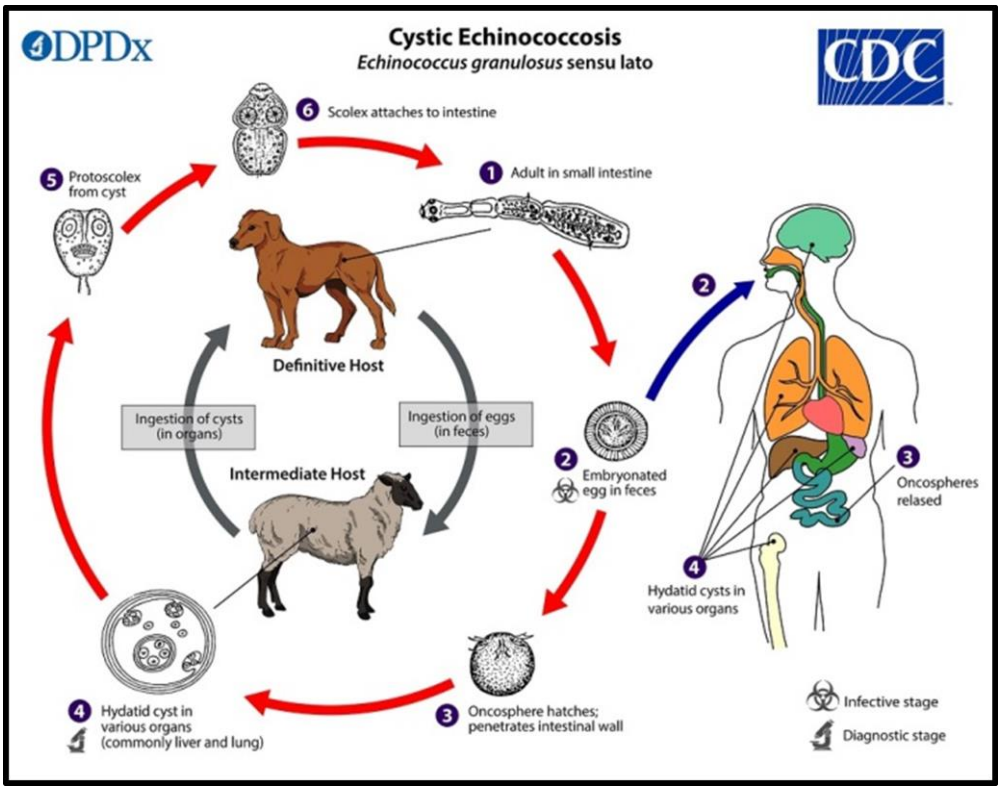


Figure 2: Schematic representation of *E. granulosus* life cycle (CDC, 2012). *Adapted from CDC/USA.* https://www.cdc.gov/dpdx/echinococcosis/index.html

1.5 Diagnosis of hydatidosis in human

Evaluation and diagnosis of echinococcosis starts from a thorough history, including a history of exposure or immigration from endemic areas, followed by a combination of serology and imaging (Symeonidis, N., et al, 2013). Liver function tests are unreliable in detecting the severity of disease and are abnormal in only about 40% of patients. Alkaline phosphatase is the one typically elevated, while AST, ALT, and bilirubin levels typically remain within normal range. The complete blood count may show eosinophilia (Symeonidis, N., et al, 2013).

Serological testing for *Echinoccocus* antibodies is a common method for diagnosis in which ELISA tests are usually used for this. A confirmatory immunoblot assay for echinococcal antigens is carried out following a positive result. However, a significant number of echinococcosis patients do not elicit an immune response. Strength and detectability of the immune response depend on various factors including cyst viability and cyst wall intactness (Zhang, W., et al, 2012). Imaging is extremely useful in detecting and monitoring hydatidosis cases, especially those that are seronegative. Ultrasound CT imaging method is the preferred modality for liver and intraabdominal hydatidosis, but its accuracy remains operator dependent. It remains the modality of choice for screening due to its accessibility and portability. It is also helpful in post-treatment monitoring of the disease (Group, W. H. O. I. W., 2003).

1.6 Detection of eggs in definitive hosts

E. granulosus are morphologically indistinguishable from *E. multilocularis* and most Taenia species eggs, so it is not possible to detect the presence of eggs based on microscopic examination from fecal samples. Canids cannot be diagnosed by microscope, because these eggs are morphologically indistinguishable from those of and the Taenia species. Furthermore, egg excretion is often irregular. Proglottids of *E. granulosus* spontaneously discharged by dogs and detected mostly on the surface of fecal samples may allow a correct morphological diagnosis (Gottstein et al., 1992). A standard method currently used for surveys of *E. granulosus* infection in dog populations is Arecoline purging which affecting smooth muscle of the small intestine. It includes the application of Arecoline to dogs and the examination of fecal material discharged after purging (Gottstein et al., 1992).

Determining the infection rate in dogs is important for epidemiologic studies and surveillance of control programs, and is also useful for assessing the dynamics of transmission and the danger of infection. Traditionally, infection in dogs has been determined by identifying worms in intestinal washes post mortem, or following Arecoline purgation. More recently, an enzyme immunoassay-based coproantigen test has been developed for this purpose (Abbasi et al, 2003). Coproantigen tests are genus specific with a specificity of approximately 97% (when worm burdens are more than 50−100 worms) (Abbasi et al, 2003). However, sensitivity is relatively limited, resulting in an overall average test sensitivity of only approximately 60% for natural canine *E.granulosus* infection (Abbasi et al, 2003).

E. granulosus is a parasite, which exhibit diversity in their life cycle patterns and host range. To date, 10 genotypes of Echinococcus have been identified by molecular genetic analysis using mainly mtDNA sequences. *E. granulosus* genotypes grouped into 4 species: *E. granulosus* (genotypes G1–G3), *E. equinus* (G4), *E. ortleppi* (G5) and *E. canadensis* (G6–G10). *E. felidis* (lion strain) isolated from South Africa has been identified as independent taxon (Sharma, M., et al., 2013).

1.7 Immune response to *Echinococcus*

Parasites generally have invasion methods to infect the host system, it is easier to the immune system to response to the early stages of infection. Future survival of parasites indicates that they have developed some mechanisms of evasion from host protective immune mechanisms to preserve their expansion (Schmid-Hempel, 2009). At the beginning of the infection it was found that the immune response is characterized by T helper 1 (Th1) immune responses which involves the release of interferon-γ (IFN-γ) and priming by dendritic cells (DCs) with IL-12 (Rigano et al., 2001). In other studies, it has shown that the secretory antigens of the parasite actively influences the host immune response, leading it to Th2 response that is characterized by Th2 cytokine profile of IL-4, IL-5, and transforming growth factor beta (TGF-β), with main production of IgE antibodies (Rigano et al., 2001). Innate immune cells of polymorphonuclear leukocyte, basophil-mast cell and monocyte also participate in local inflammatory reaction to parasites.

1.8 Treatment

In the past, surgery was the only treatment for cystic echinococcal cysts. Chemotherapy, cyst puncture, and PAIR (percutaneous aspiration, injection of chemicals and re-aspiration) have been used to replace surgery as effective treatments for cystic echinococcosis (Brunetti et al, 2010). However, surgery remains the most effective treatment to remove the cyst and can lead to a complete cure. Benzimidazole carbamates (mebendazole and albendazole) are anti-helmintic

drugs that inhibit the assembly of tubulin into microtubules, thus impairing uptake of glucose and interfering with the homeostasis of the parasite. Since their introduction in the 1970s, Benzimidazoles have proved effective against the larval stages of *E. granulosus* (Franchi et al, 1999). The treatment of alveolar echinococcosis is more difficult than cystic echinococcosis and usually requires radical surgery and long-term chemotherapy.

1.9 Characterization of *E. granulosus* genotypes

Strain differentiation of *E.granulosus* is of importance in order to determine the source of infection in dogs and the risk of human infection . While the possibility has been raised that the sheep strain (G1) is responsible for most human infections by *E.granulosus* in the Middle East where *it* was identified in sheep, camel, horse and cattle; no molecular strain identification studies have been done so far in the Middle-East strain variation on disease in man, transmission or control.

1.10 Next generation DNA sequencing

Next Generation Sequencing (NGS) is relatively new technology that allows mass sequencing of genetic material, and enables the production of a vast array of genomic information from many organisms. NGS also called high throughout sequencing method is relatively new technology that allows mass sequencing of different DNA fragments from an amplification reaction of fragmentation of whole genome (Besser, 2018); (Bonk et al., 2018);

(Mardis, 2013).This technology enables the production of a vast array of genomic information from many organisms in parallel and it is provides a separate quantitative counting measurement for each sequenced DNA segment type (Bonk et al., 2018); (Salipante et al., 2013).

Adapting NGS technology reduces the cost of DNA sequencing compared to Sanger traditional method and this by avoiding time-consuming and tedious traditional cloning steps. In NGS sequencing method it is possible to perform millions of sequencing reactions for part of whole genes or reactions that involve different amplified PCR segments.

Sanger sequencing was developed by Fred Sanger in 1977 (Sanger et al., 1977),it can determine nucleotide sequence in one amplicon only. The method can sequence a regions of DNA up to about 900 base pairs in length. In Sanger sequencing the target DNA is copied many times, making fragments of different lengths. Fluorescent "chain terminator" nucleotides mark the ends of the fragments and allow the sequence to be determined.

Next-generation sequencing (NGS), is a modern sequencing technology also known as high-throughput sequencing. These technologies allow for sequencing of DNA and RNA much more quickly and cheaply than the previously used Sanger sequencing. Development of next generation sequencing (NGS) technologies have added to our knowledge greater information that exceeded many data obtained by Sanger sequencing, continuous use of this technology is expected to secure the establishment of DNA databases for both human genes and plant pathogens (Roossinck, 2017) (Maxwell et al., 2018). NGS emerged in 2005 using commercial

Solexa sequencing technology and expanded rapidly to different systems (Liu et al., 2012) (Massart et al., 2014). NGS technologies include the following main methods:

1- Illumina sequencing (sequencing by synthesis) works by simultaneously identifying DNA bases, as each base emits a unique fluorescent signal, and adding them to a nucleic acid chain.

2- Roche 454 sequencing: This method is based on pyrosequencing, a technique which detects pyrophosphate release using fluorescence, after nucleotides are incorporated by polymerase to a new strand of DNA.

3- Ion Torrent sequencing that measures the direct release of H+ (protons) from the incorporation of individual bases by DNA polymerase and therefore differs from the previous two methods, as it does not measure light.

Although NGS technology initially was used to study whole genomes, a variety of approaches that address defined regions of the genome have emerged. There are essentially two technical preparatory approaches to explore selected regions of the genome with NGS. The first is by PCR, typically involving multiple primer pairs in a mixture that are combined with genomic DNA of interest in a multiplex approach to preserve precious DNA. The use of multiplex primer pairs couples the high throughput of NGS platforms and the fact that each sequence read represents a single DNA product in the mixture due to the nature of the sequencing platforms (Mardis, 2013).The second approach involves hybrid capture, which has been developed by several groups and commercialized (Albert, 2007)__(Gnirke et al., 2009); (Hodges et al, 2007).

Illumina NGS sequencing system is considered the most widely used system; solid-phase bridge amplification is used where each end of a DNA template is ligated with adapters. While one end of adapter-conjugated DNA fragment is attached to a substrate, the other end makes a bridge with immobilized primers and generates clusters of identical template in order to enhance the chemiluminescent signal (Ambardar et al., 2016). This process continues in a cycle in the presence of a mixture of four nucleotides, followed by image capture while each nucleotide is labeled with a different fluoro-phore. This cycle is repeated until the DNA fragment is synthesized to its target length (Maxwell et al., 2018).The basic principle of Ion Torrent system, by sequentially adding nucleotides, the incorporated nucleotide is detected by measuring pH change due to the release of H+ ions (Chen et al., 2018), (Abed et al, 2019). Next generation sequencing systems are able to simultaneously read the sequence of millions of short DNA fragments (typically 25-400 bps in length) (Maxwell et al., 2018), (Ravi et al., 2018).

2.Objectives:

The main goal of the proposed study is to develop a PCR system that can specifically identify cestode parasites in copro-DNA samples. The study based on designing new primers that can amplify a DNA segment suitable for sequencing by NGS technology of Illumina platform. The specific objectives are:

1. To design new PCR system that is based on shared mitochondrial sequences among different cestode species and to be suitable for DNA amplification from most *Teanidae* species.
2. To test the suitability of the designed primers in terms of its sensitivity and specificity to amplify *E. granulosus* mitochondrial DNA and other cestode DNA.
3. To use the designed and primers to test the presence of *E. granulosus* target DNA in dog fecal samples using PCR technology followed by NGS DNA analysis.

3. Materials and methods:

3.1 Samples

3.1.1. Dog fecal samples:

Dog fecal samples were collected from Yatta town located in the southern part of Hebron district. A total of 50 samples were collected from home dogs and sheep owner farmers. For each collected sample, 5 grams of fecal materials were collected in 50 ml sterile screw-cap tube containing 20 ml 70% alcohol.

3.1.2. DNA controls:

DNA samples preserved worms from Al-Quds university collection were used, included different *Echinococcus* species and *Teania* species.

3.2 DNA extraction

DNA was extracted from Dog feces in triplicate preparations in order to maximize the concentration of the prepared DNA, each time about 0.2 grams of fecal sample were mixed with lysis buffer as indicated below. The exact steps of DNA extraction were as the following steps:

1. 0.2 gram of fecal material was dispensed in a sterile microfuge tube and treated with 400 µl of DNA lysis buffer (50 mM NaCl, 10 mM EDTA, 50 mM Tris-HCL pH 7.4, 1% Triton X-100) to each microfuge tube followed by vortexing for 1 minute at room temperature. Then 20 µl of 10 mg/ml of dissolved Proteinase K were added (Sigma/Aldrich, Sant-Louis, USA).
2. Samples were incubated at 60^{o}C for 2 hours to insure complete digestion.

3. Addition of 0.4 ml Phenol (pH 8.0) (Sigma/Aldrich, Sant-Louis, USA), followed by Vortexing for 1 minute and microcentrifugation (14,000 rpm)| for 3 minutes.
4. Transfer of the top upper aqueous layer (200 µl) into new 1.5 ml microfuge tube.
5. Addition of 8 µl of 5M NaCl (to bring the final concentration of NaCl in the tube to 0.2 M) .Followed by direct addition of three volumes of cold absolute ethanol.
6. Tubes were incubated overnight at -70 ^{o}C.
7. Tubes were centrifuged at (14,000 rpm) for 10 minutes to precipitate the DNA.
8. Tubes were washed in 0.3 ml 70% ethanol.
9. Tubes were decanted and the precipitated DNA was air dried for about 10 minutes to remove all the residual alcohol.
10. To each tube 100 µl of sterile double distilled water was added. DNA was suspended in the water by 1 minute vortexing.
11. At the end each of the three preparations belong to each sample were pooled in one tube, and samples were kept at -20 ^{o}C.

3.3 DNA amplification by polymerase chain reaction (PCR)

3.3.1. Primers: Parasite's DNA was amplified from the extracted DNA using newly designed primers as indicated in the results section. The primers targeted cytochrome C oxidase gene, 18s r-DNA gene, and EgG1 *Hae* III repeat. The designed reverse and direct primers were indicated and separate PCR systems, the exact sequences of these primers and their putative amplified DNA amplicon size are indicated in Table 2.

3.3.2. Primers designing: Primers were designed first by preparing DNA alignment on the website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to all the species that we looked for. (figure 3) shows that we selected the areas that have the same sequence between the parasites that the primers can be made. Between these primers, there is almost unique sequence to each parasite that can be identified by next generation sequencing.

```
E.granulosus      -ctatggtttattggatcgtgcccgttagatggataactgtaataactctagagctaata       59
E.canadensis      -ctatggtttattggatcgtgcccgttagatggataactgtaataactctagagctaata       59
E.equinus         -ctatggtttattggatcgtgcccgctagatggataactgtaataactctagagctaata       59
T.Solium          -ctatggtttattggatcgtacccgttaaatggataactgtaataactctagagctaata       59
T.saginata        -ctatggtttattggatcgtacccgttaaatgggtaactgtaataactctagagctaata       59
T.asiatica        -ctatggtttattggatcgtacccgttaaatggataactgtaataactctagagctaata       59
                   ***************** * * ** ** **** **************************


H.d                catgccacgaagccctgaccccgctcg---------------------------------      155
H.nana            catgccacgaagccctgaccccgggctccc------------------------------      210
D.caninum         catgcctcgatgccctgaccgcgctgaccgcttggcgggtctctcgacct----------      109
Em                catgcctcgatgccctgaccctgcttgc---------------ttgcgt-----------      93
E.granulosus      catgcctcgatgccctgaccctgctcgtgtttgccgtgc----ttgcat-----------      104
E.canadensis      catgcctcgatgccctggccctgctcgtgtttgccttgcttgcttgcat-----------      108
E.equinus         catgcctcgatgccctgaccctgctcgtgtttgccttgcttgcttgcat-----------      108
T.hydatigena      catgctttgatgccctgactctgctgtttgcttgcttgcccgctgctgtctagtagtggt      119
T.Solium          catgcctcgatgccctgactctgtcggcttgctgctt-------gcctgt---------g      103
T.saginata        catgccccgatgccctgactctgttagcctgctgctgcttgcttgtgtgt---------g      110
T.asiatica        catgcctcgatgccctgactctgttagcctgctgctgcttgcctgtgtgt---------g      110
                  *****   ** ****** *   *


H.d                ---------------tcggggaatgggtgcacttattagaacagaagccaaccagtacat      200
H.nana            ---------------tcggggaatgggtgcacttattagaacagaagccaaccagtctcc      255
D.caninum         ----gccggctggacggcggggatgggtgcacttattagatcagaagccaaccgactgcg      165
Em                ------------gcggtgggggatgggtgcacttattagaccagaagccaaccaaccaac      141
E.granulosus      ------------gtggtgggggatgggtgcacttattagatcagaagc----caaccaac      148
E.canadensis      ------------gtggtgggggatgggtgcacttattagatcagaagc----cagccaac      152
E.equinus         ------------gtgatgggggatgggtgcacttattagatcagaagc----caaccaac      152
T.hydatigena      ggcagtaggtaggcagtaggggatgggtgcacttattagatcagaagccaaccaattgtt      179
T.Solium          ggtaggtggtggcaggcagggtgtgggtgcacttattagatcagaagccaaccaactgct      163
T.saginata        ggtggtggcgggtaggcagggtgtgggtgcacttattagatcagaagccaaccaactgct      170
T.asiatica        ggtggtggcaggcaggcagggtgtgggtgcacttattagatcagaagccaaccagctgct      170
                                  ***  ***************** *******    *
```

Figure 3: DNA aligned 18s rDNA sequences for different cestode species. The yellow highlighted DNA sequence region represents the primers chosen sites that shows homology among all other species.

Table 2: PCR systems and the different primers that were used in DNA amplification.

PCR system	Primers name and sequence	Amplicon size (bp)	Accession No.	Melting temp.
PCR system 1	**Eg18S3D:** TGGGTGCACTTATTAGATC **Eg18S3R:**CTGTAACAATTATCCAGAGTC	108 bp	AB731639	55 ^{o}C
PCR system 2	**EgC1D:** TTGATCCDTTRGGWGGK **EgC1R:** AACATATGATTDSCYCMCA	180 bp	KM014610	55 ^{o}C
PCR system 3	**Eg18S5D:** GGTTTATTGGATCGTACCC **Eg18S5R:**CTGTAACAATTATCCAGAGTC	219 bp	AB731639	55 ^{o}C
PCR system 4	**EgC2D:** AACATATGATTDSCYCMCA **EgC2R:** GTATCATGDARAAYWTTATCC	240 bp	KM014610	55 ^{o}C
PCR system 5	**Eg18S4D:** GGTTTATTGGATCGTACCC **Eg18S4R:** GHTCTAATAAGTGCACCCA	111 bp	AB731639	55 ^{o}C
PCR system 6	**Eg18S2D**: AGATACCGYCCTAGTTCTG **Eg18S2R:** CAGCTTTGCYACCATACT	135 bp	AB731639	55 ^{o}C
PCR system 7	**Eg18S1D:** TTGAAHAAATTWGAGTGCTC **Eg18S1R:** TGAARACATVCTTGRCRAA	216 bp	AB731639	55 ^{o}C

The IUPAC code is a 16-character code which allows the ambiguous specification of nucleic acids . The code can represent states that include single specifications for nucleic acids (A, G, C, T/U) or allows for ambiguity among 2, 3 or 4 possible nucleic acid states. The IUPAC code is, in principle, case insensitive, but its established uses generally default to the capital case (Johnson, 2010).

Table 3: IUPAC code for incomplete nucleic acid specification.

Symbol	Translation
A	A
C	C
G	G
T	T
U	U
R	A
Y	C or T/U
M	A or C
K	G or T/U
S	C or G
W	A or T/U
H	A, C or T/U
B	C, G or T/U
V	A, C or G
D	A, G or T/U
N	A,C,G or T/U

3.3.3. Polymerase chain reaction:

DNA amplification of *Echinococcus* DNA was achieved using ready to use dry Taq DNA polymerase (Syntezza, Jerusalem). This type of Taq DNA polymerase is ready to use and all needed buffer, dNTPs, and a standard concentration of $MgCl_2$ salts. The exact quantities needed for one PCR reaction using this type of Taq polymerase tubes, forward and reverse primers: 1µl (20 pmoles/µl) each, pure water and 5 µl of DNA sample and double distilled water to a final volume of 25 µl. In most of the cases more than one sample is tested by each PCR system, so enough master mix without the DNA was prepared, aliquot as 20 µl into each dry Taq DNA polymerase PCR tube, and then adding 5µl DNA from each sample is added.

The used thermal cycler program was as follows:

- Complete denaturation: 5 min at 95^{o}C.
- Amplification: 35 cycles: each composed of
    - Quick denaturation for 30 seconds at 95^{o}C.
    - Primers annealing to target DNA for 30 seconds at the Tm of the used primer.
    - Extension for 1 min at 72^{o}C.
- A final elongation step at 72 ^{o}C for 10 min.
- The amplified DNA was lastly incubated at 4^{o}C until PCR tubes were removed from the PCR machine (Bio-Rad, USA).

3.3.4. PCR amplification for next generation (NGS) DNA sequence analysis (first PCR):

For this purpose, the candidate primers that were most effective as judged by band intensity in DNA amplification were adapted to be used for NGS analysis. This was done by the addition to the 5'-prime end of specific primers of the following forward and reverse overhang adaptors. (Referred to the guideline of 16S Metagenomic Sequencing Library Preparation).

- **Forward adaptor:** TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
- **Reverse adaptor:** GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC

So the average length of each primer used in NGS analysis was about 45 bp. The used annealing temperatures in NGS first PCR were those shown in Table 2.

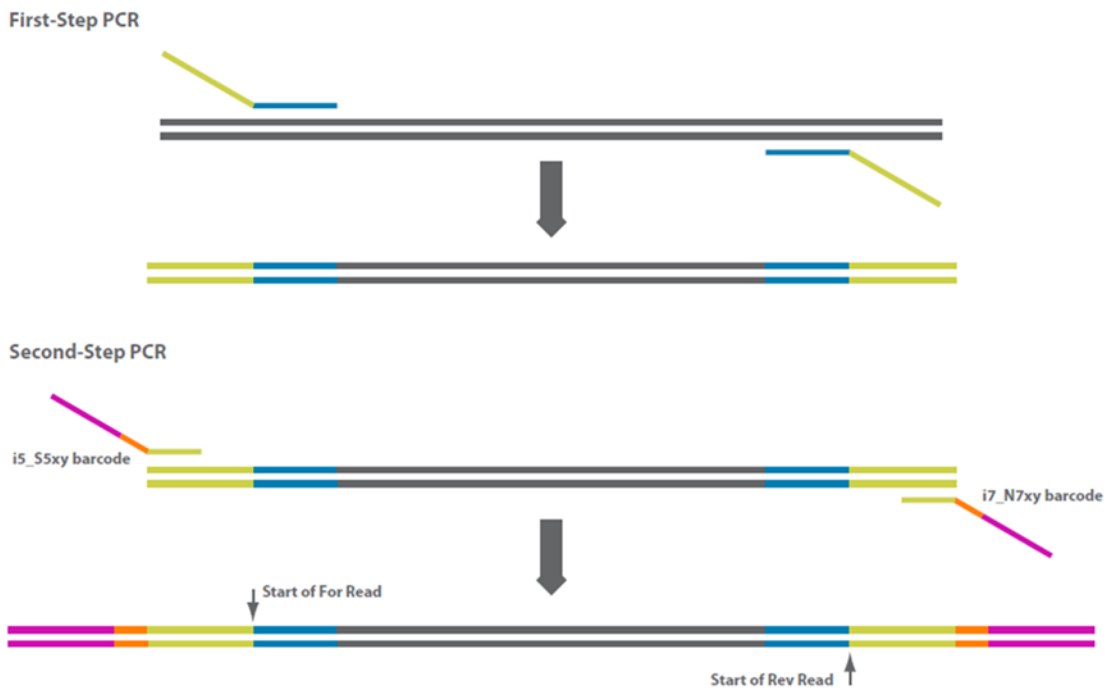
3.4 Agarose Gel Electrophoreses

The produced PCR products were analyzed on a 1.5% agarose gel prepared in 1X TAE stock 50X running buffer is : (242g Tris base, 57.1ml glacial acetic acid and 100ml 0.5M EDTA (pH 8.0)). The Gene Ruler 50bp DNA ladder (Thermo Scientific, USA) used as DNA size marker. The DNA was visualized on imaging documentation system (Dyner, UK) after the addition of 10µl of Ethidium bromide to the running buffer.``````````````

3.5 Next generation DNA sequence analysis

For this type of DNA sequence analysis Illumina Miseq (Illumina, USA) NGS strategy as used for Next era microbiome MiSeq DNA analysis was used (Ravi et al., 2018). The protocol is based on multiple PCR amplification for each sample followed by DNA sequence analysis of all amplified DNA amplicons. In short, different PCR reactions were performed for each individual extracted DNA sample using the PCR systems that were adapted to be used in the Illumina MiSeq next generation sequencing (NGS) system. Each NGS primer is composed of two parts: The first part: Direct and reverse primers that are specifically designed to target the specific DNA to be amplified (*Echinococcus* or *Taenia* parasites). The second part: Universal 5'-tailed oligonucleotides (complementary region for R1 connected to the direct primer or R2 connected to reverse primer) known as Read 1 and Read 2. These sites will be used later to add sample barcodes (indices) and for DNA sequencing from both sides, Figure 4.

Figure 4: Illumina-MiSeq primers startegy.



The first primer (used in the first PCR) composed from two parts:

1. blue= for the target DNA,
2. Yellow= for R1 and R2 primers that starts DNA sequence later.

The second primer is of three parts:

1. Yellow: complementary to R1 and R2 site (used as sticky sites).
2. Orange: is the index (i5 Sxx or i7 Nxx).
3. Pink: region that is complementary to the flow cell oligos.

3.5.1. First PCR: Amplification of Parasites DNA from fecal samples and preparation of MiSeq DNA library:

Each extracted DNA sample of the total 50 fecal extracted DNA samples was subjected to three PCR systems using specific primers with Illumina overhang region as indicated in figure 4. The used PCR systems were: PCR system 1, 2 and 3.After agarose gel-electrophoresis analysis

of each amplified DNA fragments; the three PCR systems for each tested sample were pooled in one PCR tube. The PCR reactions (60 µl) were later purified using magnetic beads as indicated below.

3.5.2 Purification of first PCR pools using magnetic beads:

PCR pools for each individual sample were purified uisng AMPure XP magnetic beads (AMPure XP beads kit / Beckman coulter, USA). Purification was performed to remove the incorporate dNTPs and the un-used PCR primers used in each PCR reaction. The purification was performed according to the following steps:

1- To each sample (60 µl) add 40 µl of AMPure XP magnetic beads were added and mixed well.

2- The mixed PCR and the beads were then kept at room temperature for 5 minutes.

3- Then the tubes were transferred to 96 wells magnetic plate stand, and left for another 5 minutes until the beads attach to tube side. (DNA is supposed to bind to the magnetic beads).

4- The PCR reaction solution was remove by gentile pipetting making without disturbing the attached beads.

5- The remained bound beads then were washed twice with 200 µl of freshly prepared 70% ethanol.

6- While the tubes standing in the magnetic plate, all the ethanol used in the washing steps was removed, then the magnetic beads left in the magnetic plate and were air dry for about 5 minutes.

7- After ethanol was dried from the tubes, DNA was eluted by the addition of 30 µl of double distilled water. The tubes were left for about 2-3 minutes and then the eluted DNA (20 µl) was transferred into fresh tubes to be used in the second PCR for NGS library preparation. It

is important to note that the transferred 20µl supposed to have a representation of the three PCR amplified materials for each analysed single sample.

3.5.3. NGS library preparation: the second PCR used for index addition:

The second stage PCR was used to attach the dual indices (i5 and i7) linked to Illumina sequencing adapters. For this PCR two types of indices were used that represent direct and reverse side of each PCR amplicon, the indices are commonly known as: N7XX and S5XX. Index additions were performed in ready-mix PCR kit (Syntezza, Jerusalem). The following is the composition of each PCR reaction: (reaction total volume= 25 µl).

NGS library preparation: the second PCR used for index addition:

- For each index primer S5 XX and N7 XX (20 pmoles/µl), 5µl of each primer was added.
- 15 µl of DNA (quantity to be transferred directly to this tube at the end of magnetic bead purification).
- Total volume 25 µl.

The PCR was performed on Thermo-cycler using the following program:

- 5 min at 95^{o}C.

- 12 cycles: each composed of:

- 30 seconds at 95^{o}C.
- 30 seconds at 55^{o}C.
- 30 seconds at 72^{o}C.

- A final elongation step at 72^{o}C for 5 min.

Representative samples were analysed by Agarose gel electrophoresis in order to prove the success of dual index addition.

3.5.4. Final PCR pooling and preparation of Miseq NGS library:

After indices addition by the second PCR, all the PCR reactions were purified for the second time using the AMPure XP magnetic beads as indicated above. Then and after the final purification all the individual eluted PCRs were pooled into one tube (pooling all the 3 PCR systems applied for each samples; a total of 50 samples). This pooling was done by mixing 10 µl from each eluted PCR product. At this stage the library is ready for next generation DNA sequencing (NGS), the NGS DNA sequence was done as outsourcing service Company (sequencing was done on Miseq machine using 500 cycle kit from Illumina, USA).

3.6 Bioinformatics analysis

Raw Illumina sequencing data were generated from all analyzed PCR amplicons as FASTQ files of read1 (forward) and read2 (reverse) for each individual sample. These sequence reads were uploaded to Galaxy platform at (usegalaxy.org) for further sequence processing and analysis (Afgan, 2016). Initially raw sequences were filtered for quality control at a phred score of 20 equivalent to 99% confidence of each nucleotide, followed by merging forward and reverse reads, the amplified specific genes were selected according to their specific sequence length and sequence identity. The selected sequence reads from each soil and plant leaves were analyzed for sequence homology above 97% using BLAST analysis tools in order to determine number of reads related to specific microbiome or fungi operational taxonomic unit.

4. Results

4.1 Problem identification of *E. granulosus* diagnosis in dogs definitive host

The main objective of this study was based on obtained results related to epidemiological survey handling the prevalence of *echinococcus* parasite in dogs using previously known molecular method normally used for this purpose (Abbasi et al, 2003). Surprisingly the two used PCR systems reveled large number of infected dogs after examining DNA that was extracted from the collected dog fecal samples. Figure 5 shows agarose gel electrophoresis an alysis of the amplified *E. granulosus* mitochondrial *cox1* gene, the results clearly indicated the presence of 39 positive samples from the total analyzed 40 samples (97.5%) . The same exact results were obtained upon retesting the same samples (an amplification band of 446bp size) with another already published and widely used PCR system that targets a repetitive gene (EgG1 *Hae* III repeat) in *E. granulosus* genome (Abbasi et al, 2003). Figure 6, shows the results of the obtained bands using this PCR that indicates the presence of 38 positive samples out of 40 tested samples (95%). A positive PCR was indicated by the presence of the main repetitive band that measured about 269 bp and the second repetitive fragment band measured about 538 bp. Efforts were done to confirm the positivity of these results by applying DNA sequence analysis. Having sequencing information from these bands was not possible, it was obvious for the second PCR targeting the *E. granulosus* repetitive gene; since many bands were amplified and conventional Sanger DNA sequencing method was not effective. However, in case of the COX1 amplified bands that were discrete and representing one type of amplicon, the sequence information was containing many unknown bases (N), and this indicated the presence of several amplicon types with the same band size. For this reason, more specific primers that could give accurate and reliable results, without the need for further DNA sequence analysis were sought.

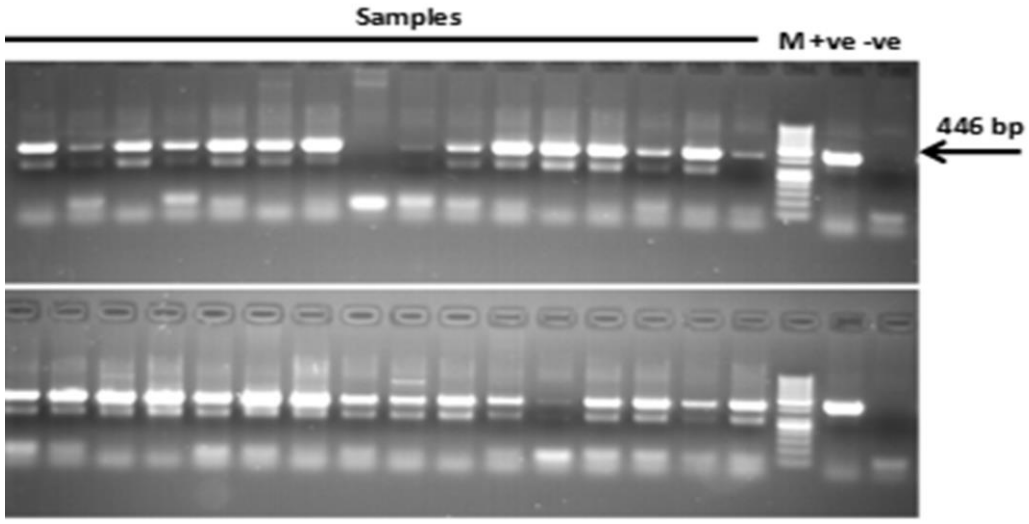


Figure 5: PCR amplification of *E. granulosus* COX1 gene fragment targeting DNA extracted from dog fecal samples. Arrow indicates the expected band size (446 bp).

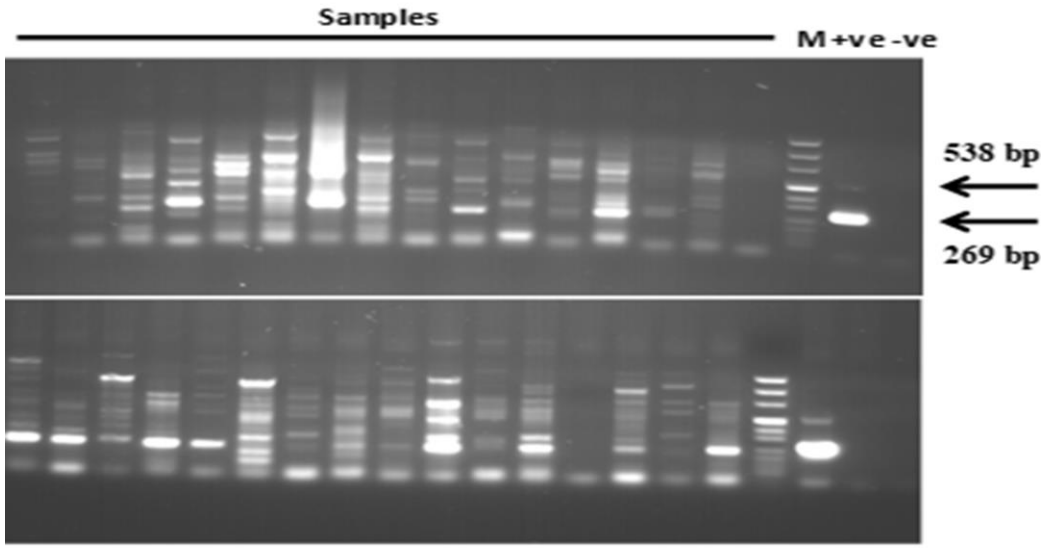


Figure 6: PCR amplification of *E. granulosus* repetitive gene fragment (EgG1 *Hae* III repeat) targeting DNA extracted from dogs fecal samples. Arrow indicates the expected band size (446 bp).

4.2 Designing new specific primers for *E. granulosus* DNA amplification suitable for NGS analysis

Attempts were done to design new primers to be used in NGS DNA sequence analysis, these primers were selected from the mitochondrial genes. For this purpose 18s rDNA genes and COX1 mitochondrial genes for different cestode species were selected from GenBank and aligned for primer selection. Selected genes were for the following cestode parasites: *E.granulosus, Echinococcus equinus, hymenolepis nana, Dipylidium caninum, Taenia hydatigena, T.Solium, T.saginata, T.asiatica*. The main criterion and criteria for the selected primers were to have a conserved sequence among all aligned sequences. This will enable amplification of most cestode species and then further discrimination upon DNA sequence analysis. Figure 7 shows an example of aligned 18s r-DNA genes for different cestode species.

Based on different aligned sequences it was possible to identify 7 sets of direct and reverse primers that represent 7 different PCR systems. The sequence details of these PCR systems and the length (in bp) of the amplified DNA fragment are shown in Table 2 in material and methods. For NGS DNA based amplicon sequencing analysis a PCR system that amplified a DNA segment between 200-300 bp is the best size for this type of analysis, taking into account that the maximum length of any DNA segment to be analyzed by such method is a maximum of 500 bp. (Referred to the guideline 16S Metagenomic Sequencing Library Preparation).

Beside the indicated 7 PCR systems targeting 18s r-DNA or the COX1 genes, another extra PCR systems was designed that was based on the known *E. granulosus* repetitive DNA segment The primers were designed to amplify a DNA fragment of about 200 bp (Figure 8).

Part of E.granulosus repetitive DNA segment (EgG1 Hae III repeat).

```
1     ACTGGGGAGG GCTTTTGGTG TTAGGAATTG CTGATATCAG GCGTTAGGGG
51    AATGGGAGAA GAGATCGCCA GACTATCCTT GAACAACCGA CAAAGACATC
101   ATAACCACGA AAAATGATAG GGATAAGGCA AATCCAACCA ATTGAAAATA
151   AGACACCACG CATGAGGATT ACTGACTTGA ACGATCACTC CTATTCTAGC
201   ATGTGTGTGA ATGCAAGCAG CAGATGCCTA CCCATCCGGA AAAGCATTTA
251   ACTTACCAGT GGGCCGTGCT GTGGAGGTAG TTTCGGCATC TAGTGAAGGA
```

Alignment between E.granulosus repetitive DNA segment and E. multilocularis genome:

```
Query:        1 ACTGGGGAGGGCTTTTGGTGTTAGGAATTGCTGATATCAGGCGTTAGGGGAATGGGAGAA 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 15487011 ACTGGGGAGGGCTTTTGGTGTTAGGAATTGCTGATATCAGGCGTTAGGGGAATGGGAGAA 15487070

Query:       61 GAGATCGCCAGACTATCCTTGAACAACCGACAAAGACATCATAACCACGAAAAATGATAG 120
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 15487071 GAGATTGCCAGACTATCCTTGAACAACCGACAAAGACATCATAACCACGAAAAATGATAG 15487130

Query:      121 GGATAAGGCAAATCCAACCAATTGAAAATAAG 152
                ||||||||| ||||||||||||||||||||
Sbjct: 15487131 GGATAAGGCGAATCCAACCAATTGAAAATAAG 15487162
```

Figure 7: *E. granulosus* repetitive DNA segment that was used for new primer design. The yellow highlighted region represents the sites of direct and reverse primers.

4.3 Testing the utility of the candidate primers in amplification of target *E. granulosus* DNA

The selected primers were examined if they were able to produce one nanogram of pure DNA of *E. granulosus*, *T. saginata*, and *T. hydatigenia*. From the newly designed seven primers only two primers were to have successful amplification . PCR system 1 and PCR system 3 (Table 2). These PCR systems were able to purify DNA from the indicated three cetsode species, representing different groups of Taenia parasites. Similarly, the designed primers that

were based on the known repetitive *E. granulosus* DNA segment, amplified 1ng of pure genomic DNA from the three tested species (Figure 8). This extra PCR system was named PCR system 8.

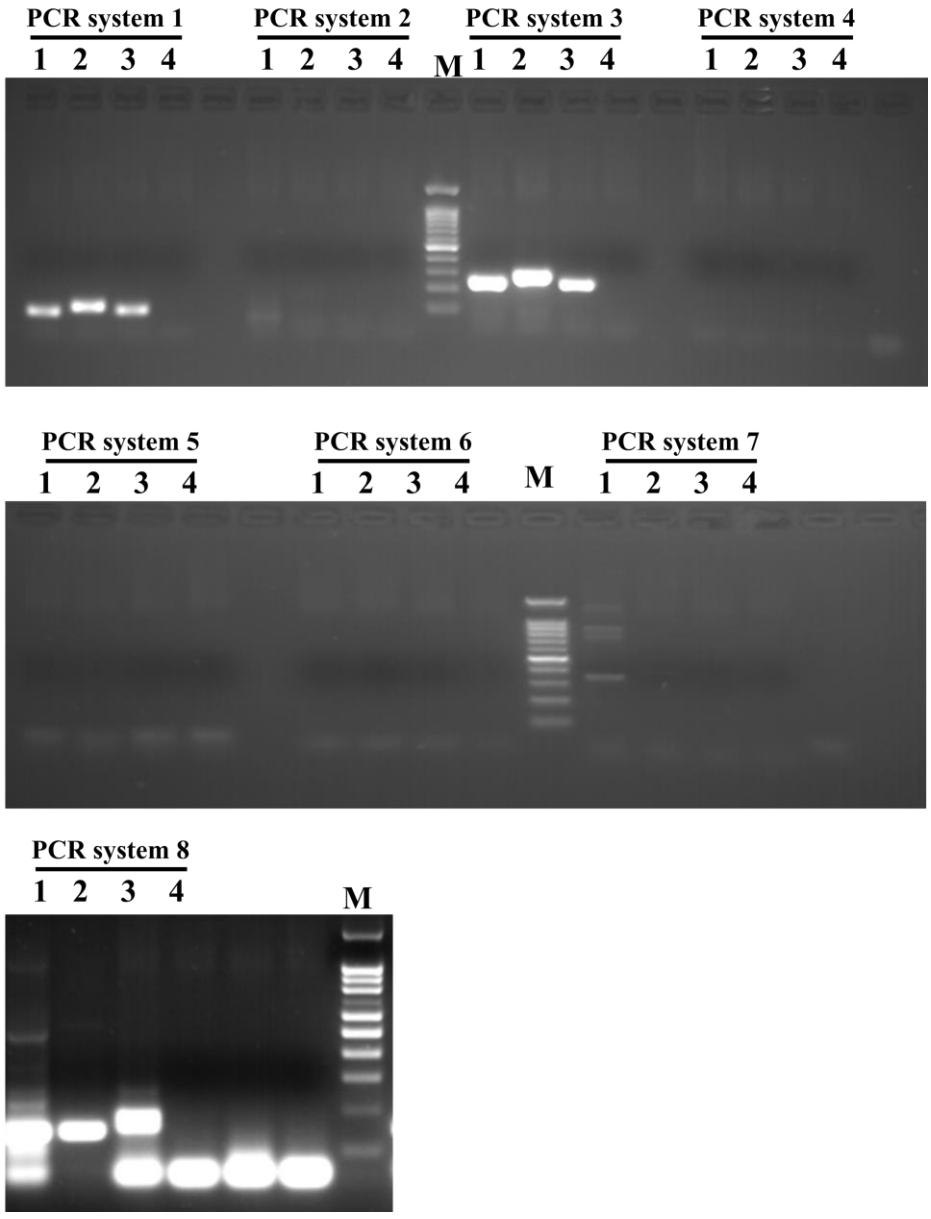


Figure 8: PCR amplification of pure genomic DNA using the newly designed primers. Pure genomic DNA (1ng) was used in this amplification from different cestode species: 1: *E. granulosus*, 2: *T. saginata*, 3: *T. hydatigenia*, 4: negative control, M: DNA size marker.

Table 4 shows the names of the newly designed primers of the three PCR systems that were used in this study. In order to simplify and have easier follow up of the PCR systems, they were renamed as the following: PCR sys 1, sys 2, and sys 3. In (table 4) new primers are indicated.

Table 4: PCR systems and their primers' DNA sequence information.

PCR system	Primers name and sequence	Amplicon size (bp)	Target gene
PCR SYS 1	**Sys1DNGS**:TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGAAHAAATTWGAGTGCTC **Sys1RNGS:**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAARACATVCTTGRCRAA	**108 bp**	**Mitochondrial 18s r-DNA**
PCR SYS 2	**Sys2DNGS:**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTTTATTGGATCGTACCC **Sys2RNGS:**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGTAACAATTATCCAGAGTC	**219 bp**	**Mitochondrial 18s r-DNA**
PCR SYS 3	**SYS5DNGS:**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTGGGGAGGGCTTTTGGTGTT **SYS5RNGS:**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTATTTTCAATTGGTTGGATTTGCC	**152**	**EgG1 *Hae* III repeat**

The Tm of these primers was 60 °C.

Note: The exact primers sequences is highlighted in yellow, while the other sequence region represents the reverse and direct adapters that are used in adapting these primers to be used in NGS analysis.

4.4 Sensitivity test of the new designed primers

The three designed PCR systems were tested for their sensitivity limit in amplifcation of pure *E. granulosus* genomic DNA. Several two fold dilutions of *E granulosus* were made starting from 0.1 ng/µl, the used final DNA concentrations were (0.1, 0.05. 0.025. 0.01. 0.005, 0.0025 ng/µl). PCR system 1 proved to be the most sensitive test used among the three systems, since it was possible to amplify all the diluted DNA concentration and the end-point of the lowest dilution was not reached (Figure 9). PCR system 2 and system 3 showed less amplification sensitivity than PCR system 1 and both reached to a sensitivity limit of 0.01 (ng/µl).

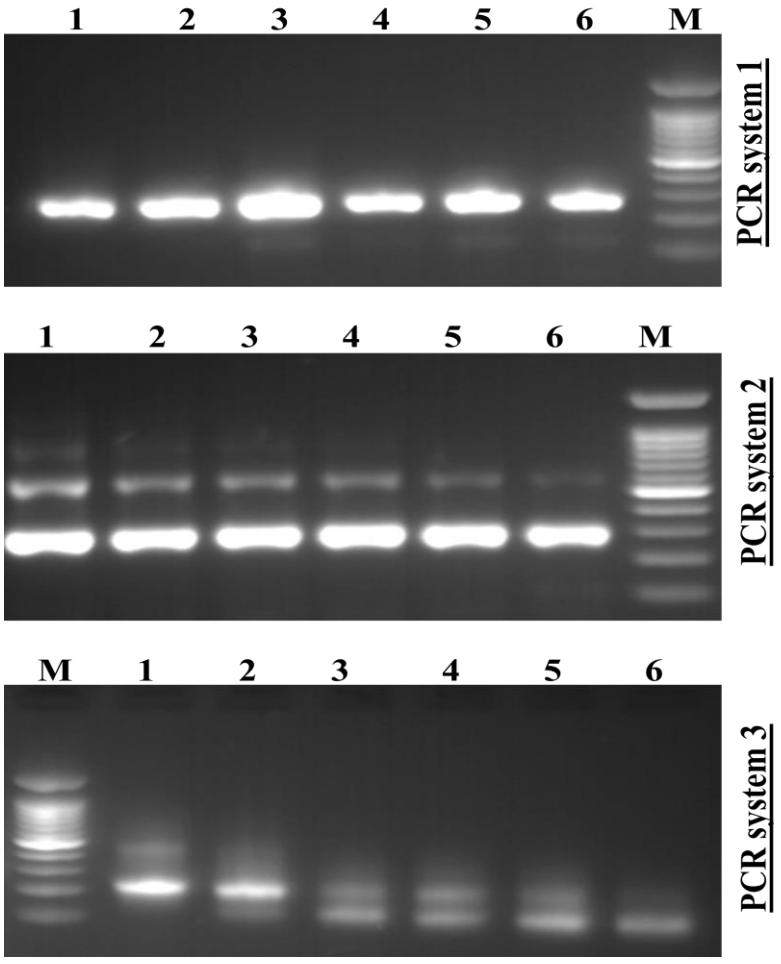


Figure 9: PCR sensitivity test of the three candidate PCR systems. Several dilutions of *E. granulosus* genomic DNA control was used (1:0.1 ng/µl, 2: 0.05 ng/µl. 3: 0.025 ng/µl. 4: 0.01 ng/µl. 5: 0.005 ng/µl, 6: 0.0025 ng/µl). M= DNA size marker.

4.5 Specificity test of the candidate PCR systems

The three PCR systems were examined for their potential to amplify pure genomic DNA from different cestodes species, namely the following species were used in this specificity study (*E. granulosus. E. multilocularis, Taenia serialis, T. pisiformis, T. ovis, T. hydatigenia, Multiceps multiceps, D. caninum*). PCR system 1 that targeting 18s rDNA fragment was able to amplify all the tested species with relatively similar strength of amplification (Figure 10). Similarly, and as seen in (figure 10) PCR system 2 that also targeted 18s rDNA fragment could amplify DNA from all the tested species. PCR system 3 was less effective than the other two systems, although amplification of most examined species was positive but amplified bands were less intense than the other systems suggesting sensitivity as observed in (figure 10). It is worth to mention that the amplified bands from most species have relatively the same size, and this is not affecting the purpose of this study, as NGS of DNA should will discriminate between the different species. In addition, it is known from the primer design that there is enough sequence differences between all cestode species and only the primers were constructed from the conserved sequences.

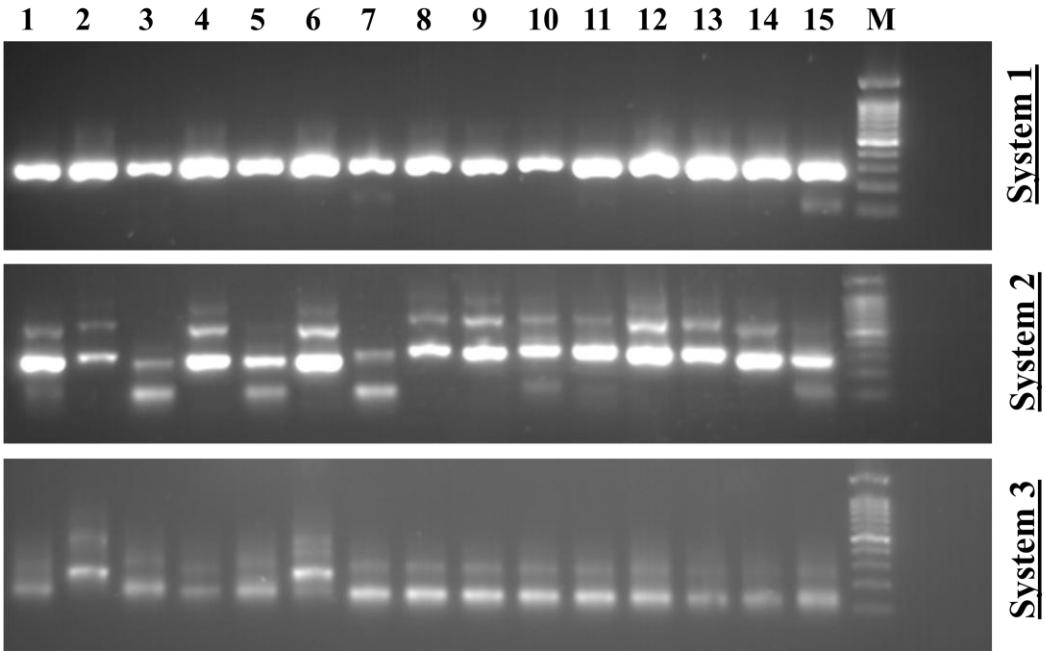


Figure 10: Specificity test of the three candidate PCR systems. The three PCR systems were used to amplify 1ng/µl of different genomic DNA extracted from the following cestode worms: (1-5: *E. granulosus.* 6: *E. multilocularis,* 7*: Taenia serialis,* 8: *T. pisiformis,* 9: *T. ovis,* 10-11: *T. hydatigenia,* 12-13: *Multiceps multiceps,* 14-15: *D. caninum*). M: DNA size marker.

4.6 Appling PCR systems to DNA extracted from dog-fecal samples

The three developed PCR systems were applied in PCR amplification targeting DNA extracted from dog fecal samples. A total of 50 samples that were collected from Yatta town were used in this part of the study. Figures 11-13 shows agarose gel electrophoresis analysis of the produced amplicons from the three different PCR systems. It is clearly seen that PCR system 1 showed the highest number of amplified bands from the total examined samples, and almost no sample showed amplification products using PCR system 1 and only few amplified bands were seen upon the use of PCR system 3. At that stage, it was not possible to calculate the number of positive samples, for accuracy and confirmation positivity had to be based on DNA sequence analysis.

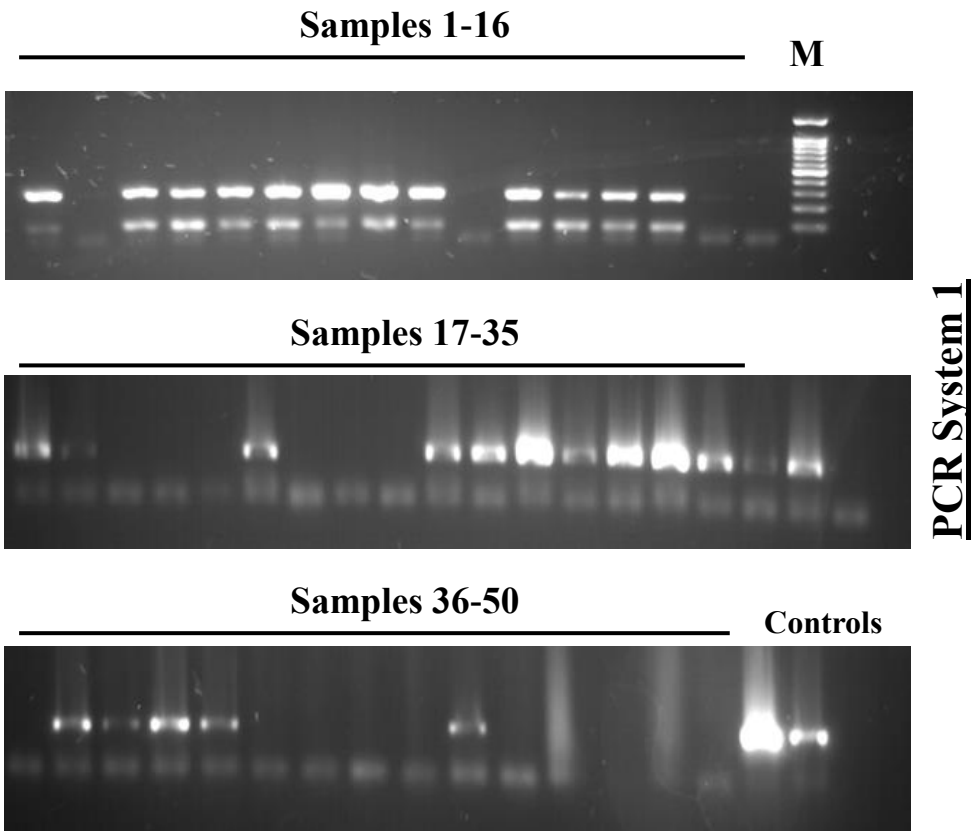


Figure 11: PCR amplification targeting 50 DNA samples extracted from dog fecal samples using PCR system 1.

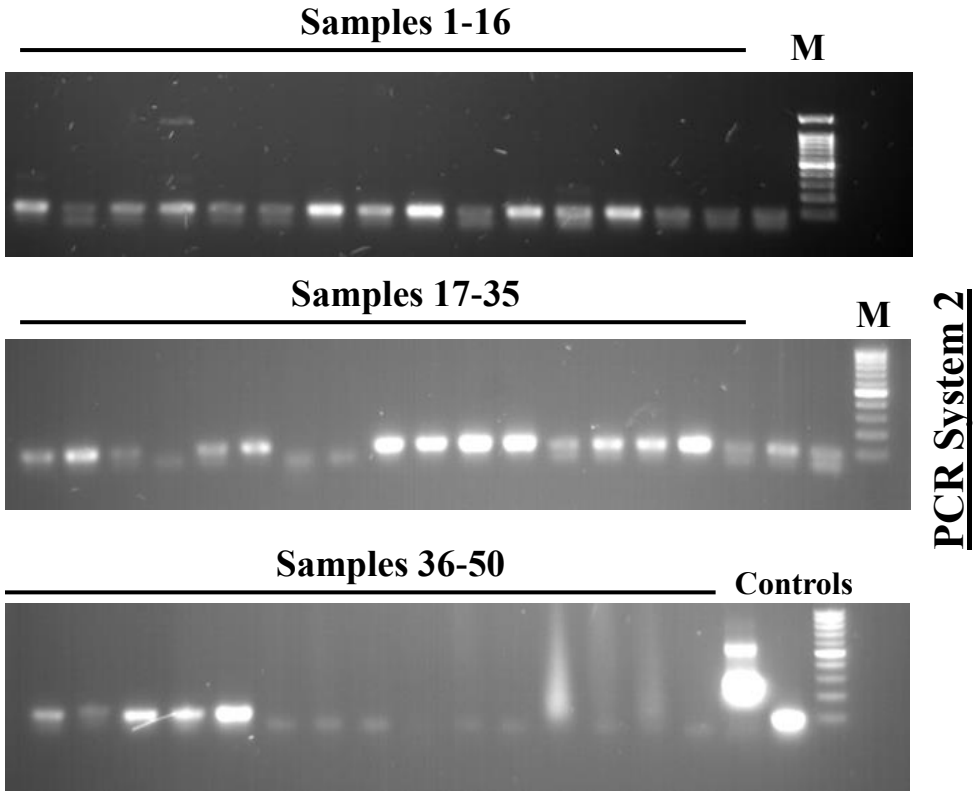


Figure 12: PCR amplification targeting 50 DNA samples extracted from dog fecal samples using PCR system 2.

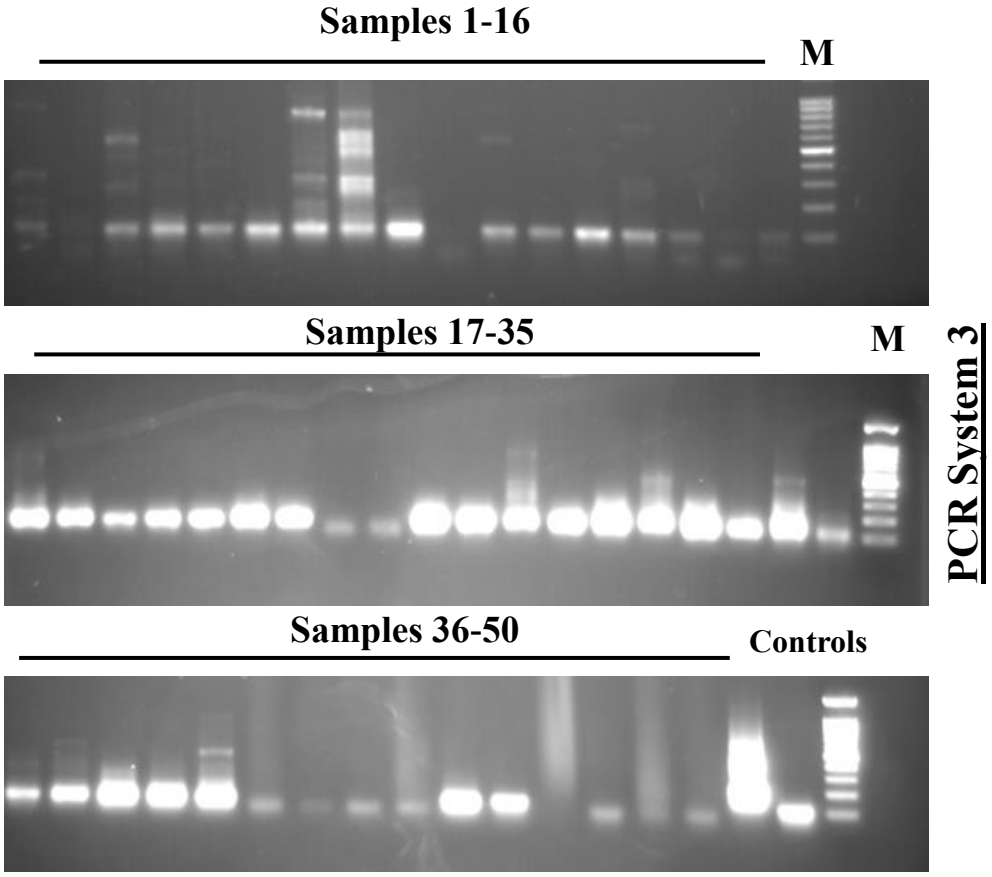


Figure 13: PCR amplification targeting 50 DNA samples extracted from dog fecal samples using PCR system 3.

4.7 NGS library preparation and bioinformatics

4.7.1. Library pooling and purification:

As it was indicated in material and methods, NGS library was prepared at two stages: the first PCR, which was represented by the three used PCR systems, and then the second PCR that included individual sample labeling by specific barcodes. So, for library preparation all the products of the three PCR systems for each individual sample from of 50 samples were pooled together, then each sample was labeled by two indices (Nxx and Sxx). All the labeled samples were then pooled in one tube after primer cleaning from all samples using magnetic beads and analyzed for their suitability for NGS analysis. (Figure 14) shows the (Tap-station) fluorescent DNA gel electrophoresis analysis of the pooled and purified library.

The analysis shows an abundance of most amplified bands by the three PCR systems, it is worth mentioning that all the expected bands had gained about 80 bp after the addition of the dual indices on both sides.

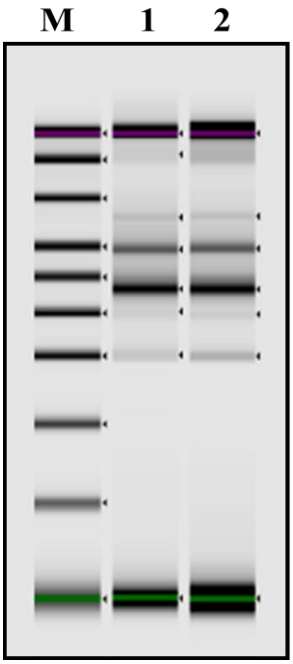


Figure 14: Tap-station fluorescent DNA gel electrophoresis analysis. 1 and 2 represents two different trials of analysis for the same NGS prepared and purified library. Bands are concentrated in the a range between 200-500 bp.

4.7.2. DNA sequence data analysis and bioinformatics:

MiSeq DNA sequence analysis was performed for each pooled sample, the obtained sequencing data received as FASTQ files for each individual sample. For each of the 50 analyzed samples, direct and reverse reads files were received (known as Read 1 and Read 2 respectively), so in a total of 2 FASTQ files were processed for each individual sample. The files were uploaded on Galaxy.org bioinformatics online software, and sequences of each sample

were processed separately. A workflow was used in this analysis that included the following steps:

1- Compiling read 1 and read 2 amplicons for each sample (at least one thousand sequences for each sample were joined).

2- Transforming FASTQ files into FASTA files.

3- Size selection for amplicons from PCR system 1 (223-224bp) and PCR system 2 (216).

4- Collapsing all sequences in one file.

5- Fetching similarity in GenBank using BLAST tool from NCBI/NIH.

After analysis it was possible to detect two amplicon types that were related to PCR system 1 and PCR system 2. The precise identification was based on selection based on sequence length and then selection based on DNA sequence of each selected amplicon. Table 5, represents a summary for the obtained results from PCR system 1 and PCR system 2. The total number of samples that had *E. granulosus* DNA amplicons was found to be 12 in PCR systems 1 and it was only three in PCR system 2. No *Echinococcus* DNA sequences were recovered using PCR system 3, and this has two reasons as it will be discussed later. The obtained sequences showed (97-99)% similarity with 18s rDNA gene of Echinococcus species as seen in figures 16 and 17.

Table 5: Summary results of NGS of Echinoccocus DNA sequence analysis for both PCR system 1 and PCR system 2.

PCR system	Amplicon size (bp)	Total positives	Positive samples
1	223-225	12	1, 4, 8, 18, 24, 27, 30, 31, 35, 48, 49,50.
2	216	3	48, 49, 50

>2-129
GGTTTATTGGATCGTACCCGTTAGATGGATAACTGTAATAACTCTAGAGCTAATACATGCCTCGATGCCCTGACC
CTGCTCATGTTTGCCTTGCTTGCATGTGGGGGGGATGGGTGCACTTATTAGATCAGAAGCCAACCAACTGTGCGT
GTGCAGCTCACTGCGAGGTGGACGTCGCGTGCGTTGAGACACTTACTTCTGGTGACTCTGGATAATTGTTACAG

Echinococcus granulosus 18S ribosomal RNA gene, complete sequence
Sequence ID: U27015.1

Score	Expect	Identities	Gaps
398 bits(215)	7e-107	221/224(99%)	0/224(0%)

```
Query  1    GGTTTATTGGATCGTACCCGTTAGATGGATAACTGTAATAACTCTAGAGCTAATACATGC  60
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  113  GGTTTATTGGATCGTGCCCGTTAGATGGATAACTGTAATAACTCTAGAGCTAATACATGC  172

Query  61   CTCGATGCCCTGACCCTGCTCATGTTTGCCTTGCTTGCATGTgggggggATGGGTGCACT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  173  CTCGATGCCCTGACCCTGCTCATGTTTGCCTTGCTTGCATGTGGGGGGGATGGGTGCACT  232

Query  121  TATTAGATCAGAAGCCAACCAACTGTGCGTGTGCAGCTCACTGCGAGGTGGACGTCGCGT  180
            |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  233  TATTAGATCAGAAGCCAACCAACTGTGCGGGTGCAGCTCACTGCGAGGTGGACGTCGCGT  292

Query  181  GCGTTGAGACACTTACTTCTGGTGACTCTGGATAATTGTTACAG  224
            |||||||||||||||||||||| ||||||||||||||||||||
Sbjct  293  GCGTTGAGACACTTACTTCTGGAGACTCTGGATAATTGTTACAG  336
```

Sequences producing significant alignments — Manage Columns — Show 100

select all *100 sequences selected* — GenBank — Graphics — Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Echinococcus granulosus 18S ribosomal RNA gene, complete sequence	398	398	100%	7e-107	98.66%	U27015.1
Echinococcus felidis gene for 18S rRNA, partial sequence, sample code: EfelUg	375	375	100%	3e-100	96.89%	AB731638.1
Echinococcus granulosus 18S ribosomal RNA gene, complete sequence	370	370	100%	2e-98	96.44%	GQ260092.1
Echinococcus granulosus isolate XJ-h 18S ribosomal RNA gene, partial sequenc	370	370	100%	2e-98	96.44%	EU704119.1
Echinococcus granulosus gene for 18S rRNA, partial sequence	359	359	100%	4e-95	95.56%	AB731639.1

Figure 15: PCR system 1 amplicon DNA sequence analysis. The figure include 224 bp sequence of one of the samples, its alignment with *E. granulosus* , and its similarity to the highest five BLAST scores from the GenBank.

>7-2
TTGAATAAATTTGAGTGCTCAAGTCAGGCCGATGTTGCCTGTAAAGTTTTGCATGGAATAATGGAAGAGGACTTCGGT
TCTATTTCGTTGGTTTTCGGATCCGAAGTAATGATCAAAAGGAGCAGGCGGGGGCGTTTGTATGGCTGCGCTAGAGGT
GAAATTCATGGACCGTAGCCAGACAAACTAAAGCGAAGGCATTCGTCAAGCATGTCTTCA

Echinococcus felidis gene for 18S rRNA, partial sequence, sample code: EfelUganda
Sequence ID: AB731638.1

Score	Expect	Identities	Gaps
377 bits(204)	9e-101	212/216(98%)	0/216(0%)

```
Query  1     TTGAATAAATTTGAGTGCTCAAGTCAGGCCGATGTTGCCTGTAAAGTTTTGCATGGAATA  60
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1052  TTGAACAAATTTGAGTGCTCAAGTCAGGCCGATGTTGCCTGTAAAGTTTTGCATGGAATA  1111

Query  61    ATGGAAGAGGACTTCGGTTCTATTTCGTTGGTTTTCGGATCCGAAGTAATGATCAAAAGG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1112  ATGGAAGAGGACTTCGGTTCTATTTCGTTGGTTTTCGGATCCGAAGTAATGATCAAAAGA  1171

Query  121   AGCAGGCGGGGGCGTTTGTATGGCTGCGCTAGAGGTGAAATTCATGGACCGTAGCCAGAC  180
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1172  AACAGGCGGGGGCGTTTGTATGGCTGCGCTAGAGGTGAAATTCATGGACCGTAGCCAGAC  1231

Query  181   AAACTAAAGCGAAGGCATTCGTCAAGCATGTCTTCA  216
             |||||||||||||||||||||||||||||| ||||
Sbjct  1232  AAACTAAAGCGAAGGCATTCGTCAAGCATGTTTTCA  1267
```

Echinococcus canadensis gene for 18S rRNA, partial sequence	388	388	100%	4e-104	99.07%	AB731642.1
Echinococcus ortleppi gene for 18S rRNA, partial sequence	388	388	100%	4e-104	99.07%	AB731641.1
Echinococcus granulosus gene for 18S rRNA, partial sequence	388	388	100%	4e-104	99.07%	AB731639.1
Echinococcus felidis gene for 18S rRNA, partial sequence, sample code: EfelUganda	388	388	100%	4e-104	99.07%	AB731638.1
Echinococcus equinus gene for 18S rRNA, partial sequence	383	383	100%	2e-102	98.61%	AB731640.1
Echinococcus shiquicus gene for 18S rRNA, partial sequence	383	383	100%	2e-102	98.61%	AB731635.1
Echinococcus multilocularis gene for 18S rRNA, partial sequence	383	383	100%	2e-102	98.61%	AB731634.1
Echinococcus granulosus 18S ribosomal RNA gene, complete sequence	383	383	100%	2e-102	98.61%	GQ260092.1

Figure 16: PCR system 2 amplicon DNA sequence analysis. The figure include 216 bp sequence of one of the samples, its alignment with *E. granulosus* , and its similarity to the highest five BLAST scores from the GenBank.

5. Discussion:

Cystic echinococcosis or hydatidosis is a serious and chronic zoonotic disease in humans with cosmopolitan distribution and is especially prevalent in sheep-raising countries. The causative organism, the dog tapeworm *Echinococcus granulosus*, is transmitted cyclically between canines and numerous herbivorous livestock animals, which can serve as intermediate hosts. The disease is accidently diagnosed in children and normally it is diagnosed in older individuals and this due to the time needed for the larval stage of the parasite (hydatid cysts) to develop in the infected organ (liver, lungs, spleen, and in some cases the brain). So, the early detection of cystic echinococcosis in infected individual may prevent further development of the hydatid cysts in the human body and avoiding future hospitalization and surgical removal of the cyst. Identification of intermediate infected animal hosts and subsequent parasite control should prevent human infections.

In the southern West Bank, especially in the town of Yatta, it is the largest flat, open area in the West Bank where there are large population of stray dogs and other wild carnivores living near human habitations, as well as several sheep, and cow farms. Farmers slaughter livestock at home, and inappropriate discarding of remains to the dog contributes to spreading the disease. The dangerous situation is clearly indicated in the study in previous studies, sero-prevalence of at least 2.6/ per 100,00 among school-aged children (Abu-Hasan, N., et al., 2002). Also, this was emphasized by the statistics information of the Ministry of Health for 2011 in the West –Bank, since the prevalence of hydatidosis was shown to be (3.1 per 100,000) in human.

Determination of *E.granulosus* infection in dogs is very important for evaluating the risk of infection and for identifying new foci of active infections and for evaluating the efficacy of control programs. Identification of *E. granulosus* in intestinal washes or following arecoline purgation of canines has been traditionally used for this purpose. The coproantigen tests facilitated large-scale screening of definitive hosts, the need for improved detection sensitivity and for species-specific detection prompted the development of suitable molecular tools.

This study started by epidemiological survey of echinococcosis in Yatta town using a well-known two PCR systems that were reported to specifically amplify *E. granulosus* COX1 DNA from dog fecal samples (Abbasi et al, 2003), (Craig, et al., 2007), (Adwan et al., 2013). The results of this study showed a high positivity for the detection of the parasite from the examined samples (Figure 5 and 6). To confirm these results further DNA sequence analysis was done, and it was not easy to have DNA sequence data from the amplified products by Sanger DNA sequence analysis. Sanger DNA sequence analysis is not possible if there is more than one type of amplicons of the same size or of different sizes, and this is the situation in the used two PCR system (repetitive PCR and COXI PCR system), many un-identified nucleotides sequences appeared as (N). This problem can be resolved only of the expected DNA fragment of the parasites (if it is there) was cloned and then sequenced, and this is very difficult to be done for many samples at once.

For accurate evaluation of dog infection by *E. granulosus* new primers were designed and adapted in NGS DNA analysis. The selected target genes for this purpose were mainly the 18s rDNA and COX1 mitochondrial genes, beside the known *E. granulosus* repetitive DNA segment (EgG1 *Hae* III repeat). Seven candidate PCR systems were developed based on the

mitochondrial genes and one based on *E. granulosus* repetitive segment. Upon experimental examination of the designed primers for these PCR systems, only two systems were based on 18s r-DNA showed a successful DNA amplification targeting pure *E. granulosus* DNA and a third system that was based on *E. granulosus* repetitive region. The designed primers in this study were constructed from shared regions by from cestodes that belong to Taenidae family (*E.granulosus, D.caninum, E.equinus, H.nana, Dipylidium caninum, Taenia hydatigena, T. Solium, T. saginata, T. asiatica*) (Figure 3). This strategy was followed in constructing primer pairs that can amplify DNA for most possible cestode species, since this PCR system will followed by NGS analysis that is able to identify the specific parasitic infection even if more than one species co-existed. Based on personal communications with veterinary doctors working in the same field, they reported the presence of *Dipylidium caninum, Taenia hydatigena*, beside dog infections by *E. granulosus*. Therefore, in the absence of specific PCR system that amplifies *E. granuolosus* DNA found in dog fecal materials it may result in over-estimation of dog infections by *E. granulosus*. To avoid this situation, use of general primers that could amplify many Taenidae species (Figure 10) and to do further DNA sequencing for each amplicon by NGS analysis could resolve this problem.

Regarding the sensitivity test, tow fold of serial dilution did not show us the final concentration that the primer can amplify in PCR sys 1, so 10 fold of serial dilution is preferred to be done to the three PCR systems to reach the sensitivity limit.

Using NGS sequencing approach it is possible to use universal primers that amplify a group of pathogens such as the 16S rRNA genes for bacterial species (Barghouthi, 2011),the 18S rRNA for fungal pathogens (Colabella et al., 2018); (Imabayashi et al., 2016), primers that amplify all related DNA segments, and later all produced amplicons could be sequenced independently from each other. The obtained results will include thousands of sequences from each type of amplicon that reflects its abundance and nature.

The received sequencing data for the 50 analysed samples exceeded 1 Gigabit of DNA sequencing data, and this because all the represented DNA amplicons in each sample was sequenced. These files are of FASTQ type that shows the DNA sequence information with its quality score determined by the MiSeq sequencing machine. All these files were transformed into FASTA files that can be managed by the online bioinformatics software (Galaxy.org). As indicated in material and methods; a workflow was developed to select *E. granulosus* specific sequences based on sequence size selection and further confirmation by BLAST analysis. It was noticed first that there are many sequences that don't have the exact DNA amplicon size amplified by the it's specific PCR, these sequence were belong to fungal organisms that its DNA was extracted as well in from the dog fecal samples.

The exact length of *E. granulosus* specific amplified sequences were (223-225) bp for PCR system 1 and 216 bp for PCR system 2. Based on this finding the total number of infected dogs by *E. granulosus* was found to be 12 from a total of 50 revelled by PCR system 1 and 3 from a total of 50 revelled by PCR system 2. All the positive samples that were revelled by PCR system 2 were also detected positive by PCR system 1. So the total number of *E. granulosus* positive

samples is 12 out 50, which equals to 24%, and this is consistent with many studies in the region that estimated the number of dogs infectivity by *E. granulosus* using traditional methods for worm detection including Arecoline purging tests.

The NGS DNA sequence data did not show any specific sequences of PCR system 3 that has an amplicon size of about 150 bp. This could be due to the our strategy that was used in the NGS library preparation and the use of magnetic beads for amplicon purifications. Upon magnetic beads purification we used a ratio of 0.6% of magnetic beads to PCR products, and it is known that this ratio will eliminate all bands that are less than 200 bp and this included the 150 bp of the PCR system 1 amplicons. We were urged to do this harsh purification and this in order to clean all the produced primer dimers in this range, because these primers can affect adversely the quality of the produced DNA sequence data.

6. References:

1. Abbasi, I., Branzburg, A., Campos-Ponce, M., Abdel Hafez, S. K., Raoul, F., Craig, P. S., & Hamburger, J. (2003). Copro-diagnosis of Echinococcus granulosus infection in dogs by amplification of a newly identified repeated DNA sequence. *The American journal of tropical medicine and hygiene*, *69*(3), 324–330.

2. Abed, A., Légaré, G., Pomerleau, S., St-Cyr, J., Boyle, B., & Belzile, F. J. (2019). Genotyping-by-Sequencing on the Ion Torrent Platform in Barley. *Methods in molecular biology (Clifton, N.J.), 1900*, 233–252. https://doi.org/10.1007/978-1-4939-8944-7_15.

3. Abu-Hasan, N., Daragmeh, M., Adwan, K., Al-Qaoud, K., & Abdel-Hafez, S. K. (2002). Human cystic echinococcosis in the West Bank of Palestine: surgical incidence and seroepidemiological study. *Parasitology research*, *88*(2), 107–112.

4. Adwan, G., Adwan, K., Bdir, S., & Abuseir, S. (2013). Molecular characterization of Echinococcus granulosus isolated from sheep in Palestine. *Experimental parasitology*, *134*(2), 195–199. https://doi.org/10.1016/j.exppara.2013.03.024

5. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.

6. Almulhim, A. M., & John, S. (2020). Echinococcus Granulosus (Hydatid Cysts, Echinococcosis). In *StatPearls*. StatPearls Publishing.Ambardar et al. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry.

7. Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics, 107(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.

8. Barghouthi, S.A. A Universal Method for the Identification of Bacteria Based on General PCR Primers. *Indian J Microbiol* **51,** 430–444 (2011). https://doi.org/10.1007/s12088-011-0122-5.

9. Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, *24*(4), 335–341. https://doi.org/10.1016/j.cmi.2017.10.013.

10. Bonk, F., Popp, D., Harms, H., & Centler, F. (2018). PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. Journal of microbiological methods, 153, 139–147. https://doi.org/10.1016/j.mimet.2018.09.015.

11. Brunetti, E., Kern, P., Vuitton, D. A., & Writing Panel for the WHO-IWGE (2010). Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta tropica*, *114*(1), 1–16. https://doi.org/10.1016/j.actatropica.2009.11.001.

12. Cerda, J. R., Buttke, D. E., & Ballweber, L. R. (2018). Echinococcus spp. Tapeworms in North America. *Emerging infectious diseases*, *24*(2), 230–235. https://doi.org/10.3201/eid2402.161126.

13. CDC. (2012). Cystic Echinococcosis (Echinococcus granulosus sensu lato).

14. Chen, G., Qiu, Y., Zhuang, Q., Wang, S., Wang, T., Chen, J., & Wang, K. (2018). Next-generation sequencing library preparation method for identification of RNA viruses on the Ion Torrent Sequencing Platform. *Virus genes*, *54*(4), 536–542. https://doi.org/10.1007/s11262-018-1568-x.

15. Colabella, C., Corte, L., Roscini, L., Bassetti, M., Tascini, C., Mellor, J. C., Meyer, W., Robert, V., Vu, D., & Cardinali, G. (2018). NGS barcode sequencing in taxonomy and diagnostics, an application in "*Candida*" pathogenic yeasts with a metagenomic perspective. *IMA fungus*, *9*(1), 91–105. https://doi.org/10.5598/imafungus.2018.09.01.07.

16. Craig, P. S., McManus, D. P., Lightowlers, M. W., Chabalgoity, J. A., Garcia, H. H., Gavidia, C. M., Gilman, R. H., Gonzalez, A. E., Lorca, M., Naquira, C., Nieto, A., & Schantz, P. M. (2007). Prevention and control of cystic echinococcosis. *The Lancet. Infectious diseases*, *7*(6), 385–394. https://doi.org/10.1016/S1473-3099(07)70134-2.

17. Eckert, J., & Deplazes, P. (2004). Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. *Clinical microbiology reviews*, *17*(1), 107–135. https://doi.org/10.1128/cmr.17.1.107-135.2004.

18. Franchi, C., Di Vico, B., & Teggi, A. (1999). Long-term evaluation of patients with hydatidosis treated with benzimidazole carbamates. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, *29*(2), 304–309. https://doi.org/10.1086/520205.

19. Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology*, *27*(2), 182–189. https://doi.org/10.1038/nbt.1523.

20. Gottstein B. (1992). Molecular and immunological diagnosis of echinococcosis. *Clinical microbiology reviews*, *5*(3), 248–261. https://doi.org/10.1128/cmr.5.3.248.

21. Grubor, N. M., Jovanova-Nesic, K. D., & Shoenfeld, Y. (2017). Liver cystic echinococcosis and human host immune and autoimmune follow-up: A review. *World journal of hepatology*, *9*(30), 1176–1189. https://doi.org/10.4254/wjh.v9.i30.1176.

22. Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J., & McCombie, W. R. (2007). Genome-wide in situ

exon capture for selective resequencing. *Nature genetics*, *39*(12), 1522–1527. https://doi.org/10.1038/ng.2007.42.

23. Heidari, Z., Sharbatkhori, M., Mobedi, I., Mirhendi, S. H., Nikmanesh, B., Sharifdini, M., Mohebali, M., Zarei, Z., Arzamani, K., & Kia, E. B. (2019). Echinococcus multilocularis and Echinococcus granulosus in canines in North-Khorasan Province, northeastern Iran, identified using morphology and genetic characterization of mitochondrial DNA. *Parasites & vectors*, *12*(1), 606. https://doi.org/10.1186/s13071-019-3859-z.

24. Hui, W., Jiang, S., Tang, J., Hou, H., Chen, S., Jia, B., & Ban, Q. (2015). An Immediate Innate Immune Response Occurred In the Early Stage of E. granulosus Eggs Infection in Sheep: Evidence from Microarray Analysis. *PloS one*, *10*(8), e0135096. https://doi.org/10.1371/journal.pone.0135096.

25. Imabayashi, Y., Moriyama, M., Takeshita, T., Ieda, S., Hayashida, J. N., Tanaka, A., Maehara, T., Furukawa, S., Ohta, M., Kubota, K., Yamauchi, M., Ishiguro, N., Yamashita, Y., & Nakamura, S. (2016). Molecular analysis of fungal populations in patients with oral candidiasis using next-generation sequencing. *Scientific reports*, *6*, 28110. https://doi.org/10.1038/srep28110.

26. Jabbar, A., Narankhajid, M., Nolan, M. J., Jex, A. R., Campbell, B. E., & Gasser, R. B. (2011). A first insight into the genotypes of Echinococcus granulosus from humans in Mongolia. *Molecular and cellular probes*, *25*(1), 49–54. https://doi.org/10.1016/j.mcp.2010.11.001.

27. Johnson A. D. (2010). An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics (Oxford, England)*, *26*(10), 1386–1389. https://doi.org/10.1093/bioinformatics/btq098.

28. Kammerer, W. S., & Schantz, P. M. (1993). Echinococcal disease. *Infectious disease clinics of North America*, *7*(3), 605–618.

29. Kilic, D., Tercan, F., Sahin, E., Bilen, A., & Hatipoglu, A. (2006). Unusual radiologic manifestations of the echinococcus infection in the thorax. *Journal of thoracic imaging*, *21*(1), 32–36. https://doi.org/10.1097/01.rti.0000186994.92705.2d.

30. Kinkar, L., Laurimäe, T., Acosta-Jamett, G., Andresiuk, V., Balkaya, I., Casulli, A., Gasser, R. B., González, L. M., Haag, K. L., Zait, H., Irshadullah, M., Jabbar, A., Jenkins, D. J., Manfredi, M. T., Mirhendi, H., M'rad, S., Rostami-Nejad, M., Oudni-M'rad, M., Pierangeli, N. B., Ponce-Gordo, F.,Saarma, U. (2018). Distinguishing Echinococcus granulosus sensu stricto genotypes G1 and G3 with confidence: A practical guide. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, *64*, 178–184. https://doi.org/10.1016/j.meegid.2018.06.026.

31. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38. https://doi.org/10.1016/j.cell.2013.09.006.

32. Laurimäe, T., Kinkar, L., Moks, E., Romig, T., Omer, R. A., Casulli, A., Umhang, G., Bagrade, G., Irshadullah, M., Sharbatkhori, M., Mirhendi, H., Ponce-Gordo, F., Soriano, S. V., Varcasia, A., Rostami-Nejad, M., Andresiuk, V., & Saarma, U. (2018). Molecular phylogeny based on six nuclear genes suggests that Echinococcus granulosus sensu lato genotypes G6/G7 and G8/G10 can be regarded as two distinct species. *Parasitology*, *145*(14), 1929–1937. https://doi.org/10.1017/S0031182018000719.

33. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology*, *2012*, 251364. https://doi.org/10.1155/2012/251364Mardis, E. R. (2013). Next-generation sequencing platforms.

34. Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus research*, *188*, 90–96. https://doi.org/10.1016/j.virusres.2014.03.029.

35. Maxwell, P., Hynes, S. O., Fuchs, M., Craig, S., McGready, C., McLean, F., McQuaid, S., James, J., & Salto-Tellez, M. (2019). Practical guide for the comparison of two next-generation sequencing systems for solid tumour analysis in a universal healthcare system. *Journal of clinical pathology*, *72*(3), 225–231. https://doi.org/10.1136/jclinpath-2017-204917.

36. Moro, P. L., Cavero, C. A., Tambini, M., Briceño, Y., Jiménez, R., & Cabrera, L. (2008). Identification of risk factors for cystic echinococcosis in a peri-urban population of Peru. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(1), 75–78. https://doi.org/10.1016/j.trstmh.2007.09.010.

37. Ravi, R. K., Walton, K., & Khosroheidari, M. (2018). MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. *Methods in molecular biology (Clifton, N.J.)*, *1706*, 223–232. https://doi.org/10.1007/978-1-4939-7471-9_12.

38. Riganò, R., Profumo, E., Bruschi, F., Carulli, G., Azzarà, A., Ioppolo, S., Buttari, B., Ortona, E., Margutti, P., Teggi, A., & Siracusano, A. (2001). Modulation of human immune response by Echinococcus granulosus antigen B and its possible role in evading host defenses. *Infection and immunity*, *69*(1), 288–296. https://doi.org/10.1128/IAI.69.1.288-296.2001.

39. Roossinck M. J. (2017). Deep sequencing for discovery and evolutionary analysis of plant viruses. *Virus research*, *239*, 82–86. https://doi.org/10.1016/j.virusres.2016.11.019.

40. Saarma, U., Jõgisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., Simsek, S., Andresiuk, V., Denegri, G., González, L. M., Ferrer, E., Gárate, T., Rinaldi, L., & Maravilla, P. (2009). A novel phylogeny for the genus Echinococcus, based on nuclear data, challenges relationships based on mitochondrial evidence. *Parasitology*, *136*(3), 317–328. https://doi.org/10.1017/S0031182008005453.

41. Salipante, S. J., Sengupta, D. J., Rosenthal, C., Costa, G., Spangler, J., Sims, E. H., Jacobs, M. A., Miller, S. I., Hoogestraat, D. R., Cookson, B. T., McCoy, C., Matsen, F. A., Shendure, J., Lee, C. C., Harkins, T. T., & Hoffman, N. G. (2013). Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PloS one*, *8*(5), e65226.

42. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463.

43. Schmid-Hempel P. (2009). Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1513), 85–98. https://doi.org/10.1098/rstb.2008.0157.

44. Sharma, M., Sehgal, R., Fomda, B. A., Malhotra, A., & Malla, N. (2013). Molecular characterization of Echinococcus granulosus cysts in north Indian patients: identification of G1, G3, G5 and G6 genotypes. *PLoS neglected tropical diseases*, *7*(6), e2262. https://doi.org/10.1371/journal.pntd.0002262.

45. Symeonidis, N., Pavlidis, T., Baltatzis, M., Ballas, K., Psarras, K., Marakis, G., & Sakantamis, A. (2013). Complicated liver echinococcosis: 30 years of experience from an endemic area. *Scandinavian journal of surgery : SJS : official organ for the Finnish Surgical Society and the Scandinavian Surgical Society*, *102*(3), 171–177. https://doi.org/10.1177/1457496913491877.

46. Thompson, R. C., & McManus, D. P. (2002). Towards a taxonomic revision of the genus Echinococcus. *Trends in parasitology*, *18*(10), 452–457. https://doi.org/10.1016/s1471-4922(02)02358-9.

47. WHO Informal Working Group (2003). International classification of ultrasound images in cystic echinococcosis for application in clinical and field epidemiological settings. *Acta tropica*, *85*(2), 253–261. https://doi.org/10.1016/s0001-706x(02)00223-1.

48. Youngster, I., Hoida, G., Craig, P. S., Sneir, R., & El-On, J. (2002). Prevalence of cystic echinococcosis among Muslim and Jewish populations in southern Israel. *Acta tropica*, *82*(3), 369–375. https://doi.org/10.1016/s0001-706x(02)00026-8.

49. Zhang, W., Wen, H., Li, J., Lin, R., & McManus, D. P. (2012). Immunology and immunodiagnosis of cystic echinococcosis: an update. *Clinical & developmental immunology*, *2012*, 101895. https://doi.org/10.1155/2012/101895.