RESEARCH ARTICLE

Comprehensive characterization of plasma cell-free *Echinococcus* spp. DNA in echinococcosis patients using ultra-high-throughput sequencing

Jingkai Ji[1,2,3☯], Bin Li[4,5☯], Jingzhong Li[5,6☯], Wangmu Danzeng[2,7], Jiandong Li[1,2,3], Yanping Zhao[2,3], Gezhen Qiangba[2,3], Qingda Zhang[4], Nibu Renzhen[4], Zhuoga Basang[4], Changlin Jia[4], Quzhen Gongsang[4,6], Jinmin Ma[2], Yicong Wang[8], Fang Chen[8,9], Hongcheng Zhou[2], Huasang[2,7], Jiefang Yin[2,3], Jiandan Xie[1,2,3], Na Pei[2,3], Huimin Cai[2,3,10], Huayan Jiang[2], Huanming Yang[2,11], Jian Wang[2], Asan[2,7], Xiumin Han[10], Junhua Li[2,3,12]*, Weijun Chen[1,2]*, Dong Yang[4,13]*

1 BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China, **2** BGI-Shenzhen, Shenzhen, China, **3** Shenzhen Key Laboratory of Unknown Pathogen Identification, Shenzhen, China, **4** Second People's Hospital of Tibet Autonomous Region, Lhasa, China, **5** Tibet Center for Disease Control and Prevention, Lhasa, China, **6** NHC Key Laboratory of Echinococcosis Prevention and Control (Xizang Center for Disease Control and Prevention), Lhasa, China, **7** BGI-Tibet, Lhasa, China, **8** MGI, BGI-Shenzhen, Shenzhen, China, **9** MGI-Wuhan, BGI-Shenzhen, Wuhan, China, **10** People's Hospital of Qinghai, Xining, China, **11** James D. Watson Institute of Genome Sciences, Hangzhou, China, **12** School of Biology and Biological Engineering, South China University of Technology, Guangzhou, China, **13** China-Japan Friendship Hospital, Beijing, China

☯ These authors contributed equally to this work.
* lijunhua@genomics.cn (JL); chenwj@bgi.com (WC); yangdong1975@sina.com (DY)

OPEN ACCESS

Citation: Ji J, Li B, Li J, Danzeng W, Li J, Zhao Y, et al. (2020) Comprehensive characterization of plasma cell-free *Echinococcus* spp. DNA in echinococcosis patients using ultra-high-throughput sequencing. PLoS Negl Trop Dis 14(4): e0008148. https://doi.org/10.1371/journal.pntd.0008148

Editor: Uriel Koziol, Universida de la República, URUGUAY

Received: October 10, 2019

Accepted: February 18, 2020

Published: April 13, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pntd.0008148

Data Availability Statement: The data reported in this study are available in the CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/

Abstract

Background

Echinococcosis is a life-threatening parasitic disease caused by *Echinococcus* spp. tapeworms with over one million people affected globally at any time. The *Echinococcus* spp. tapeworms in the human body release DNA to the circulatory system, which can be a biomarker for echinococcosis. Cell-free DNA (cfDNA) is widely used in medical research and has been applied in various clinical settings. As for echinococcosis, several PCR-based tests had been trialed to detect cell-free *Echinococcus* spp. DNA in plasma or serum, but the sensitivity was about 20% to 25%. Low sensitivity of PCR-based methods might be related to our limited understanding of the features of cell-free *Echinococcus* spp. DNA in plasma, including its concentration, fragment pattern and release source. In this study, we applied ultra-high-throughput sequencing to comprehensively investigate the characteristics of cell-free *Echinococcus* spp. DNA in plasma of echinococcosis patients.

Methodology/Principal findings

We collected plasma samples from 23 echinococcosis patients. Total plasma cfDNA was extracted and sequenced with a high-throughput sequencing platform. An average of 282 million read pairs were obtained for each plasma sample. Sequencing data were analyzed

cnsa; accession number CNP0000696). Control datasets of cell lines deep sequencing and simulated data were available with accession number CNP0000857. Control datasets of negative individuals were available with accession number CNP0000868.

Funding: This work was supported by the National Key Research and Development Program of China (No.2018YFC1200501), the Infectious Diseases Special Project, Ministry of Health of China (2018ZX10732-401) and the Science, Technology and Innovation Commission of Shenzhen Municipality under grant No. JCYJ20170817145915789. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

with bioinformatics workflow combined with *Echinococcus* spp. sequence database. After identification of cell-free *Echinococcus* spp. reads, we found that the cell-free *Echinococcus* spp. reads accounted for 1.8e-5 to 4.0e-9 of the total clean reads. Comparing fragment length distribution of cfDNA between *Echinococcus* spp. and humans showed that cell-free *Echinococcus* spp. DNA of cystic echinococcosis (CE) had a broad length range, while that of alveolar echinococcosis (AE) had an obvious peak at about 135 bp. We found that most of the cell-free *Echinococcus* spp. DNA reads were from the nuclear genome with an even distribution, which might indicate a random release pattern of cell-free *Echinococcus* spp. DNA.

Conclusions/Significance

With ultra-high-throughput sequencing technology, we analyzed the concentration, fragment length, release source, and other characteristics of cell-free *Echinococcus* spp. DNA in the plasma of echinococcosis patients. A better understanding of the characteristics of cell-free *Echinococcus* spp. DNA in plasma may facilitate their future application as a biomarker for diagnosis.

Author summary

Echinococcosis is one of the most neglected tropical diseases caused by the metacestodes of *Echinococcus* spp. tapeworms, which affect both humans and livestock. Plasma cell-free DNA (cfDNA) consists of nucleic acid fragments found extracellularly and may contain DNA released from the parasites. Research shows that a variety of parasites can be detected from plasma cfDNA. Cell-free *Echinococcus* spp. DNA in plasma or serum had been tested with PCR-based methods, but these PCR methods had low sensitivity ranged from 20% to 25%. Low sensitivity may be due to our limited understanding of cell-free *Echinococcus* spp. DNA in plasma. Here, we take advantage of high-throughput sequencing to get a comprehensive characterization of cell-free *Echinococcus* spp. DNA. Our results showed that with high-throughput sequencing we could detect cell-free *Echinococcus* spp. DNA in all samples, though at a very low level. Based on the sequencing data, we found that cell-free *Echinococcus* spp. DNA in plasma had a different fragment length distribution to cell-free human DNA, and fragment length distribution of cell-free *Echinococcus* spp. DNA is also different between cystic echinococcosis (CE) and alveolar echinococcosis (AE). The sequencing data can also help trace the release source of cell-free *Echinococcus* spp. DNA from the genome. According to the mapping results of cell-free *Echinococcus* spp. DNA reads, we found that most of them were from the nuclear genome rather than the mitochondrial genome, and their release position showed an even distribution on the genome. These characteristics of cell-free *Echinococcus* spp. DNA in echinococcosis patients' plasma could facilitate their future application in research or clinical settings.

Introduction

Echinococcosis is a life-threatening zoonosis caused by *Echinococcus* spp. tapeworms with a complex life cycle involving intermediate and definitive hosts. Definitive hosts of *Echinococcus* spp. tapeworms are mainly carnivores such as dogs, foxes, and wolves, and intermediate hosts are usually ungulates or rodents such as sheep, cattle, and pika [1]. Humans can be accidentally infected and develop echinococcosis [2]. As one of the most neglected diseases, at any given time, echinococcosis is affecting more than one million people globally [3–5]. Among the species in genus *Echinococcus*, there are two most important ones in terms of public health, *E. granulosus* and *E. multilocularis*, responsible for cystic echinococcosis (CE) and alveolar echinococcosis (AE) respectively [6,7]. CE is cosmopolitan, with high endemic areas include western China, Central Asia, eastern Africa, South America, and Mediterranean countries, and AE is mainly in the northern hemisphere [1,7].

The diagnosis of echinococcosis is based on clinical findings, imaging and serological test [7,8]. Imaging includes ultrasound, magnetic resonance imaging, and computed tomography, among which ultrasound is most widely used as the basis for screening and clinical diagnosis [8]. Based on ultrasound observations, the World Health Organization Informal Working Group on Echinococcosis classified CE cysts into six types (cystic lesion or CL, and CE1–5) and AE lesions into different PNM types (Parasite lesion, Neighbor organs, Metastases) [7,8]. Imaging techniques provide the clinician with important clinical information including the location, number, size, and stage of the cysts, which are crucial for the diagnosis of echinococcosis [7,9,10]. However, there are some unsolved issues with imaging techniques, especially the most commonly used ultrasound in diagnosing echinococcosis. The foremost problem is the late diagnosis. As the early phase of infection is generally asymptomatic, patients may remain asymptomatic for years, even permanently. Given the long incubation period (5–20 years), echinococcosis is not easy to be diagnosed in the early stage, and many asymptomatic patients are diagnosed by chance [11–13]. Besides, detecting small cystic lesions is also a challenge in imaging diagnosis of echinococcosis. It is not easy to distinguish echinococcosis cysts from cysts caused by other reasons, such as liver abscesses, Caroli disease, bilomas and cystadenomas [14–17]. The long incubation period and complex clinical manifestation of the disease also makes clinical findings difficult, and patients with symptoms are advised to undergo imaging and serological test immediately, thus clinical finding is of limited added value for diagnosis [7]. A serological test could serve as an auxiliary diagnostic tool, but its limitations include cross-reactivity and incompetence to differentiate present and past infections [18–20]. In consideration of the limitations of the existing diagnosis tools, detecting the cell-free DNA (cfDNA) released by *Echinococcus* spp. tapeworms may serve as a biomarker of the etiological agents [21,22].

CfDNA consists of nucleic acid fragments found extracellularly and mainly exists in the bloodstream, urine and other body fluids [22]. It has been widely used in clinical practice such as non-invasive prenatal testing (NIPT) [23], tumor monitoring [24] and pathogen detection [25]. As for parasite cfDNA, the metabolic activities of the parasites and attacks from the host's immune system may cause the parasites' DNA to be released into the host's circulatory system, and the possible mechanisms can be summarized as active secretion and passive release [22]. Several parasitic diseases have been successfully detected with cfDNA, including *Plasmodium* [26], *Trypanosoma* [27], *Leishmania* [28], *Schistosoma* [29] and *Wuchereria* spp. [30]. Cell-free *Echinococcus* spp. DNA had already been suggested as a biomarker for echinococcosis [21], and its existence in plasma or serum was proven with PCR-based methods [31–33], though with rather low sensitivities (20–25%) [31–33] Low sensitivity prevents further application of using plasma cfDNA in the diagnosis of echinococcosis. The unsatisfactory performance of the previous attempts could be due to three possible reasons. First, it was hypothesized that the cfDNA of the parasite did not enter the blood circulation unless the hydatid cyst(s) ruptured–thus non-existence of the parasite cfDNA in the host blood circulation made this detection method impossible [31]. Secondly, there is cfDNA from the parasite in the blood circulation, but its concentration is too low to be detected by the designed methods. Thirdly, the understanding of the characteristics of the cfDNA in circulation is limiting the application of cfDNA in detecting the parasitic infection. The better knowledge of cfDNA's characteristics in NIPT has facilitated its improvement from molecular-counting based first-generation testing strategy to global adopted size-based diagnostics [34]. There are studies and reviews on the characteristics of cfDNA in different conditions including cancer, pregnancy, and transplantation [35]. A detailed study on the existence, quantity, and characteristics of cell-free *Echinococcus* spp. DNA in echinococcosis patients' plasma is still missing.

The rapid development of high-throughput sequencing techniques made it feasible to sequence cfDNA in research and medical settings. Compared with target-based PCR methods, sequencing can provide more comprehensive information about cfDNA [25]. High-throughput sequencing of cfDNA has been widely used in tumor and prenatal diagnosis, which provides much more detailed information of cfDNA for clinical practice and research [23,24]. We initiated this study to explore the existence, quantity, and characteristics of cell-free *Echinococcus* spp. DNA in the plasma of echinococcosis patients with high-throughput sequencing. We collected plasma samples from clinically diagnosed echinococcosis patients, produced cfDNA sequencing data with high-throughput sequencing technology, and analyzed the massive data with bioinformatics workflow. The results revealed that high-throughput sequencing of plasma cfDNA could serve as a feasible tool for cell-free *Echinococcus* spp. DNA study and improve our understanding of *Echinococcus* spp. infection in the human body.

Materials and methods

Ethics statement

This research was reviewed and approved by the Ethics Committee of Second People's Hospital of Tibet Autonomous Region (SPHTAR-ERC-1), Center for Disease Control and Prevention of Tibet Autonomous Region Institutional Review Board (TCDCP-IRB001) as well as the Institutional Review Board of Beijing Genomics Institute in Shenzhen (BGI-IRB18157-T1). All samples were collected with written informed consent from adult participants, and minors' informed consent was given by their guardians.

Samples and processing

Blood samples from ultrasound-confirmed echinococcosis patients (N = 23) were collected at diagnosis and before any medical treatment. The patients' gender, age, and clinical classification are shown in Table 1. Type of echinococcosis was classified based on ultrasound observations and classification system of the World Health Organization Informal Working Group on Echinococcosis. Among these patients, 14 subsequently underwent surgical operations to remove the cystic lesions, and 9 received chemotherapy. The only AE case (S1) at the beginning was diagnosed as a cystic lesion with ultrasound examination, and the lesion sample of this case was confirmed with PCR to be *E. multilocularis* infection. All blood samples were collected with Ethylenediaminetetraacetic acid (EDTA) tubes. After collection, plasma samples were stored at 4˚C and centrifuged at 4˚C within four hours. The blood samples were centrifuged at 1600g for 10 min at 4˚C, and the plasma was recentrifuged at 16,000g for 10 min at 4˚C. After centrifugation, plasma samples were immediately stored at −80˚C for further experiments. Samples of the lesion from the 14 surgically treated patients were also collected and stored at −80˚C.

DNA extraction and high-throughput sequencing

Plasma samples stored at -80˚C were thawed, and cfDNA was immediately extracted from plasma using the cfDNA isolation kit. To yield high-quality cfDNA, two kits were used for

Table 1. Clinical data and sequencing results of each patient. A total of 23 echinococcosis patients were involved in the study. Plasma samples were performed with ELISA tests and cell-free DNA sequencing. Lesion samples from surgery patients were performed with PCR tests.

Patient characteristics			Diagnosis characteristics			Cell-free DNA sequencing				
ID	**Gender**	**Age (Year)**	**Clinical Type[a]**	**Lesion samples (PCR)**	**Plasma samples (ELISA)**	**Raw Data (PE)[b]**	**Clean Data (PE)[b]**	***Echinococcus* Reads (PE)[b]**	***Echinococcus* RPM[c]**	***Echinococcus* species**
S1	Male	34	CL	*E. multilocularis*	positive	742,927,842	615,603,566	11140	18.096	*E. multilocularis*
S2	Female	61	CE3	*E. granulosus*	positive	281,355,103	246,691,600	2	0.008	*E. granulosus*
S3	Female	30	CE1, CE4	*E. granulosus*	positive	250,618,825	227,517,883	3	0.013	*E. granulosus*
S4	Male	30	CE3	*E. granulosus*	positive	280,431,941	249,457,697	17	0.068	*E. granulosus*
S5	Male	40	CE3	*E. granulosus*	positive	279,312,347	248,762,189	1	0.004	*E. granulosus*
S6	Female	29	CE1	*E. granulosus*	positive	308,530,827	274,958,461	4	0.015	*E. granulosus*
S7	Female	44	CE2	*E. granulosus*	positive	263,251,023	236,679,083	2	0.008	*E. granulosus*
S8	Male	29	CE2	*E. granulosus*	positive	360,137,779	313,971,836	17	0.054	*E. granulosus*
S9	Male	15	CL	*E. granulosus*	negative	351,412,640	320,738,789	37	0.115	*E. granulosus*
S10	Female	30	CE2	*E. granulosus*	positive	306,363,351	269,830,299	13	0.048	*E. granulosus*
S11	Female	43	CL	*E. granulosus*	positive	262,203,537	231,673,428	173	0.747	*E. granulosus*
S12	Female	10	CL	*E. granulosus*	positive	231,127,477	205,492,100	1	0.005	*E. granulosus*
S13	Female	58	CE3	*E. granulosus*	positive	245,838,225	219,082,144	15	0.068	*E. granulosus*
S14	Female	36	CE1	*E. granulosus*	negative	256,759,640	224,126,469	13	0.058	*E. granulosus*
N1	Female	46	CE1, CE4	NA	positive	244,658,087	205,140,224	116	0.565	*E. granulosus*
N2	Female	59	CE2	NA	positive	364,203,245	281,310,703	129	0.459	*E. granulosus*
N3	Male	35	CE2, CE4	NA	positive	289,831,896	171,116,167	367	2.145	*E. granulosus*
N4	Male	58	CE5	NA	negative	367,706,652	213,450,366	540	2.530	*E. granulosus*
N5	Male	14	CE5	NA	positive	248,171,648	203,853,580	125	0.613	*E. granulosus*
N6	Male	47	CE5	NA	negative	211,410,500	153,887,667	234	1.521	*E. granulosus*
N7	Male	27	CE1	NA	negative	132,535,233	114,252,814	15	0.131	*E. granulosus*
N8	Female	49	CE2	NA	positive	117,991,568	104,563,714	10	0.096	*E. granulosus*
N9	Female	41	CE4	NA	negative	83,740,668	73,423,055	18	0.245	*E. granulosus*

[a] Clinical Type: CL, Cystic lesion. CE1-5, Cystic echinococcosis clinical stage.

[b] PE: Paired-end Reads.

[c] RPM: Read-Pairs Per Million.

https://doi.org/10.1371/journal.pntd.0008148.t001

cfDNA extraction according to the volume of plasma. Among the 23 plasma samples (S5 Table), 22 samples with volume 0.2 to 0.6 ml were extracted with MagPure Circulating DNA Mini KF Kit (Magen, Guangzhou, China), and one sample (N4) with volume 2.2 ml was extracted with QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany). The quantity and quality of cfDNA were assessed with Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). The concentration of cfDNA was quantified by Qubit Fluorometer (Invitrogen, Carlsbad, USA) and Qubit dsDNA HS Assay kit (Invitrogen, Carlsbad, USA) following the manufacturer's instructions. As the average fragment length of cfDNA was very short, the usual fragmentation step for library preparation was skipped. The qualified cfDNA was further used to construct sequencing libraries. The final quantified libraries were sequenced on the BGISEQ-500 platform (MGI, Shenzhen, China).

PCR test of lesion samples

Lesion samples stored at -80˚C were thawed, and DNA was extracted with phenol/chloroform methods. The presence of *Echinococcus* spp. tapeworms DNA in the lesion samples was confirmed with PCR assays which were based on the amplification of a fragment within the

NADH dehydrogenase subunit 1 (ND1) mitochondrial gene [36]. The specific primers and probes with fluorescence can also be used for qualitatively distinguishing *E. granulosus*, *E.* o*multilocularis*, and *E. shiquicus* [36]. PCR was assayed in a final volume of 30 ul, with 25 ul of master mix and 5 ul of DNA extract, in the ABI 7500 (Applied Biosystems, America) Real-Time PCR System. The thermal cycling condition was: 2 min at 50˚C, 5 min at 95˚C, followed by 40 cycles of 15 sec at 95˚C and 45 sec at 60˚C.

ELISA test of plasma

Plasma samples of the patients before any medical treatment were assayed with Echinococcosis ELISA IgG kit (Beijing BGI-GBI Biotech, Beijing, China) according to the manual. Briefly, phosphate buffered saline (PBS) diluted plasma samples (1:10) were added to the plates. The plates were incubated for 30 min at 37˚C and then washed five times with the PBS-Tween buffer. Peroxidase-conjugated goat anti-human IgG, diluted 1:2000 in a PBS buffer supplemented with 0.5% Tween-20 and 1.5% BSA, was added to each well and incubated at 37˚C for 30 min. Before the addition of the tetramethyl-benzidine (TMB) substrate, the plates were washed five times with the PBS-Tween buffer. The reaction was stopped by adding 2 mol/L H2SO4. The OD450/630 value was measured by a microtiter plate reader. A positive control sample, a negative control sample, and a blank control sample were included on each plate, with the cut-off value for IgG as 0.18.

Database construction

Sequences of *Echinococcus* spp. tapeworms were downloaded from the NCBI GenBank database. To reduce sequence contamination and get a high-quality sequence database, all sequences were quality controlled with the following steps. *Echinococcus* spp. sequences from GenBank were chopped into 100 bp short pseudo-reads (step size 50 bp), then mapped to the *Echinococcus* tapeworm common host genome sequences (sheep, cattle, pigs, humans, and mice) with BLASTn [37]. Pseudo-reads with high similarity (identity $\geq$ 97%, coverage $\geq$ 92%, and e-value $\leq$ 1e-5) to the host genome sequences were considered to be from host sequence contamination. These host-contaminated pseudo-reads were located to their original chopped sequence regions, and then the regions were masked with BEDTools [38]. After the above steps, we built a qualified *Echinococcus* tapeworm sequence database.

Workflow construction

Bioinformatics workflow was constructed to identify *Echinococcus* spp. reads with five main steps (Fig 1). 1) Raw data were first processed with SOAPnuke (v1.5.6) [39] and Fastp (v0.19.5) [40] to remove low-quality reads. 2) Clean data were mapped to *Echinococcus* spp. sequence database with Kraken (v0.10.5) [41], and the candidate *Echinococcus* spp. reads were extracted from mapping results. 3) Remove reads sourced from humans with Snap-aligner (1.0beta.23) [42]. 4) Low-complexity reads were difficult to be classified accurately, thus might introduce false-positive results, and were removed with PRINSEQ (v0.20.4) [43]. 5) Remove reads of other taxa. To remove reads of other microorganisms (such as bacteria, fungus, and viruses) either from plasma or introduced by the experimental process, left candidate reads were separately mapped to the *Echinococcus* database and comprehensive database (NCBI nt) by BLASTn [37]. Reads with poor mapping results (identity $<$ 97%, coverage $<$ 92%, and e-value $>$ 1e-5) to the *Echinococcus* spp. sequences would be removed and reads that had a better mapping result to other species would also be removed.

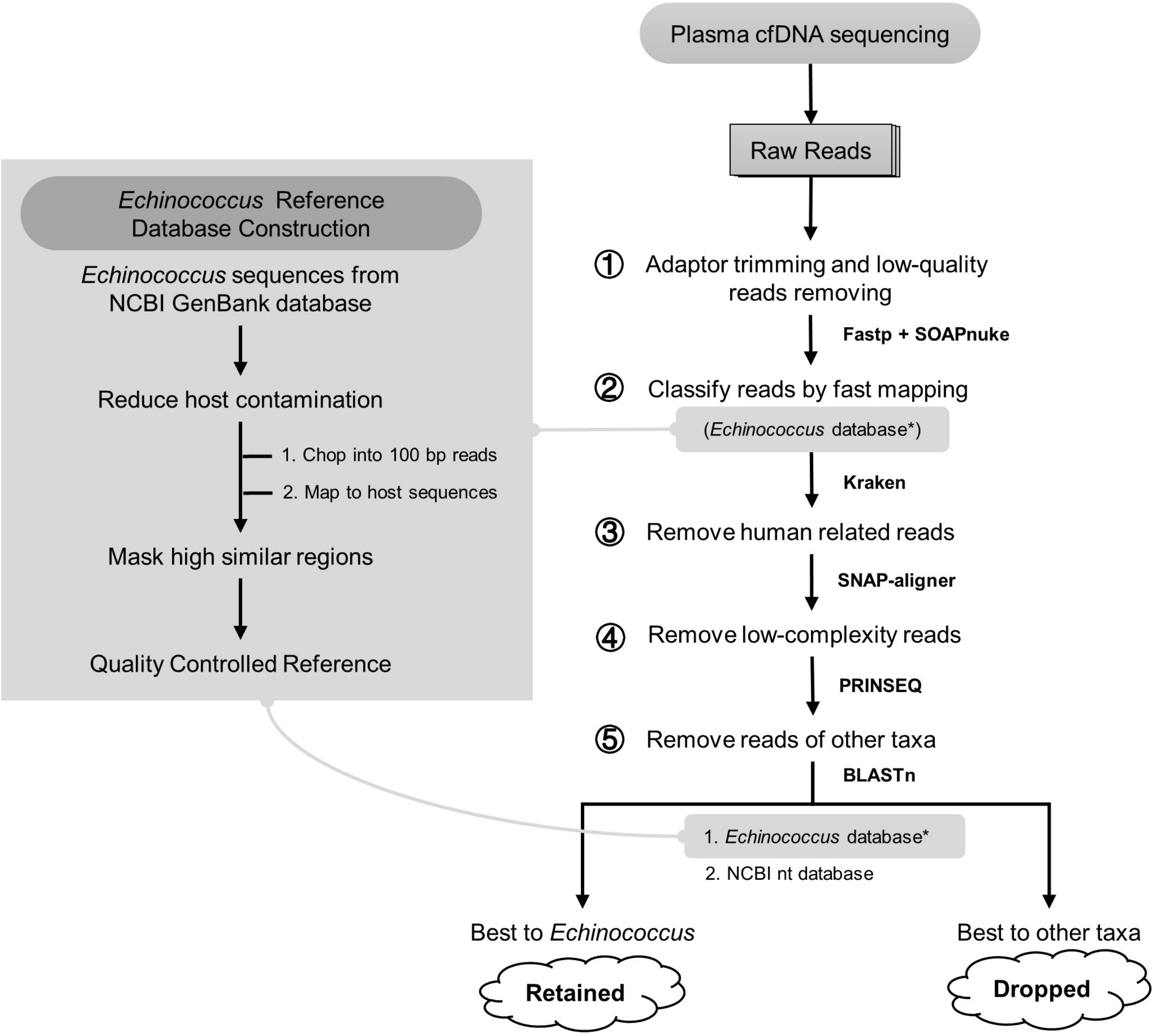
Echinococcus Reference Database Construction
Echinococcus sequences from NCBI GenBank database
Reduce host contamination
1. Chop into 100 bp reads
2. Map to host sequences
Mask high similar regions
Quality Controlled Reference
Plasma cfDNA sequencing
Raw Reads
① Adaptor trimming and low-quality reads removing
Fastp + SOAPnuke
② Classify reads by fast mapping
(Echinococcus database*)
Kraken
③ Remove human related reads
SNAP-aligner
④ Remove low-complexity reads
PRINSEQ
⑤ Remove reads of other taxa
BLASTn
1. Echinococcus database*
2. NCBI nt database
Best to Echinococcus
Retained
Best to other taxa
Dropped

Fig 1. Reference database construction and analysis workflow. Construction of *Echinococcus* spp. reference sequence database (left). Analysis workflow of cell-free *Echinococcus* spp. DNA reads identification (right).

https://doi.org/10.1371/journal.pntd.0008148.g001

Workflow evaluation

To evaluate the reliability of the workflow, we tested it with three datasets, including simulated data, cell lines deep sequencing data, and cfDNA sequencing data of individuals from non-endemic areas. Simulated data (paired-end 100bp) were produced by wgsim (https://github.com/lh3/wgsim) with human reference and *Echinococcus* spp. sequences. Data of cell lines produced with the same sequencing platform were used as a negative control. CfDNA sequencing data of 107 pregnant women from an ongoing study living in non-endemic areas were also used as a negative control. All three datasets were analyzed with the workflow to evaluate its performance.

Annotation and fragment length calculation

Identified cell-free *Echinococcus* spp. DNA reads were annotated with information from *Echinococcus* spp. sequence database. The annotation consisted of the release source and the species annotation. The read pairs were labeled according to their best mapping results to the mitochondrial or nuclear sequences which indicated their release source. The species annotation of the sample was determined similarly as the species with the most reads labeled. Based on the samples' species annotation results, *E*. *granulosus* and *E*. *multilocularis* [44] were chosen as reference for the cell-free *Echinococcus* spp. DNA features exploration. Since a more complete mitochondrial genome of *E*. *granulosus* has been published [45], we replaced the mitochondrial sequence of Tsai, *et al*. [44] with the most updated one. Read pairs from *E*. *granulosus* annotated samples and *E*. *multilocularis* annotated samples were pooled separately and remapped with BWA (v0.7.16) [46] to their corresponding references to get the mapping positions and fragment length. Based on the mapping results, the insert size was calculated with Picard (http://broadinstitute.github.io/picard). Fragment length distribution figures were produced with R version 3.3.2 (https://www.R-project.org/). Visualization of mapping positions of cell-free *Echinococcus* spp. DNA reads was achieved with Circos [47].

Analysis of sequencing data volume and positive detection

Based on *Echinococcus* spp. reads proportion, plasma cfDNA concentration, and statistical model, we analyzed the relationship between the amount of sequencing data and positive detection. Sequencing of cfDNA and cell-free *Echinococcus* spp. DNA detection can be regarded as a random sampling process. According to the hypergeometric distribution formula (1), where population size (*N*) = total number of cfDNA fragments, overall target number (*M*) = total number of cell-free *Echinococcus* spp. DNA. The number of draws (*n*) = sequencing reads amount, and the number of observed success (*x*) = detected cell-free *Echinococcus* spp. reads counts. Based on the concentration of cfDNA in the plasma of each sample, we converted the total quality of cfDNA contained in 1ml plasma to base pairs (bp) according to the formula 1pg = 978Mb [48]. According to the literature, the average length of cfDNA is 170 bp [49], and then we estimated total cfDNA fragment counts (*N*) of 1ml plasma. The total number of cell-free *Echinococcus* DNA (*M*) present in 1 ml plasma was estimated based on their proportion detected by sequencing. Then based on the formula (2), we can calculate the probability to get at least one cell-free *Echinococcus* spp. reads detection at a certain amount of sequencing data (S4 Table).

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{1}$$

$$P(X \geq 1) = 1 - P(X = 0) \tag{2}$$

Results

Samples collection and sequencing data production

Blood samples were collected from 23 echinococcosis patients. The average age of these patients (10 males and 13 females) was 38 years (Table 1). Plasma cfDNA was sequenced with the BGISEQ-500 platform and produced a total of 6,480,520,054 paired-end reads with the amount of data about 1.30 Tb. After quality control, an average of 235,025,384 paired-end clean reads per sample were left.

Performance evaluation of the analysis workflow

Simulated data, cell lines sequencing data and control human data were used to evaluate the workflow. The simulated data set included 300,000,000 paired-end reads from humans, 1,000 paired-end reads from *Echinococcus* spp. nuclear genome and 100 paired-end reads from *Echinococcus* spp. mitochondrial genome (S1 Table). After analysis with the workflow, 99.5% of the *Echinococcus* spp. nuclear genome reads were identified, 98.0% of the *Echinococcus* spp. mitochondrial genome reads were identified, and no human reads were wrongly identified (S1 Table).

As for the negative controls, DNA of cell lines was sequenced and 2,047,723,953 paired-end clean reads were produced. Evaluation of the cell lines data with the workflow showed that no *Echinococcus* spp. reads were detected. Besides, control data from 107 individuals with a total of 6,838,155,312 paired-end clean reads were used to evaluate the workflow and *Echinococcus* spp. reads were not detected from these data.

Detection of *Echinococcus* spp. infection

We used cfDNA sequencing and ELISA test to compare their performance in *Echinococcus* spp. infection detection with plasma samples from echinococcosis patients. Sequencing data of plasma cfDNA were analyzed with the analysis workflow. Cell-free *Echinococcus* spp. DNA reads were identified from all the sequencing data (23/23), with an average of 565 read pairs per sample (Table 1). To determine the *Echinococcus* species from cfDNA sequencing data, *Echinococcus* spp. reads were classified with taxonomic information. Species classification results of the identified *Echinococcus* spp. reads showed that 22 samples had most reads annotated to *E. granulosus*, and the remaining sample (S1) had most reads annotated to *E. multilocularis* (S3 Table). In comparison, the ELISA IgG kit identified 17 (73.9%) of the plasma samples of patients (N = 23) as positive. To be specific, out of the 14 surgically confirmed patients, 12 (85.7%) were positive. Out of the 9 non-surgery patients, 5 (55.6%) were positive (Table 1).

Lesion samples from surgery (n = 14) were tested with PCR methods [36] to validate the infection status and identify parasite species. All the 14 lesion samples were PCR positive (Table 1) which confirmed *Echinococcus* spp. tapeworm infection of these patients. According to PCR species differentiation results, 13 lesion samples were identified as *E. granulosus* infection and one lesion sample (S1) as *E. multilocularis* infection. The patient corresponding to S1 should be an AE patient, and other patients were confirmed as CE patients. Species identification results of PCR consisted of sequencing data analysis, which validated the plasma cfDNA sequencing methods.

Quantification of cell-free *Echinococcus* spp. DNA in plasma samples

To quantify cell-free *Echinococcus* spp. DNA in plasma, we calculated cell-free *Echinococcus* spp. DNA reads proportion in total clean reads of each sample, and the proportion ranged from 1.8e-5 to 4.0e-9 (Fig 2). Given the very low proportion of cell-free *Echinococcus* spp. DNA reads in the sequencing data, we normalized the identified *Echinococcus* spp. reads to total sequencing data with Read-Pairs Per Million (RPM) in order to facilitate comparison between samples. RPM was defined as *Echinococcus* spp. read counts per million sequencing data from one sample. Mean and median RPM of 22 CE patients were 0.433 and 0.082

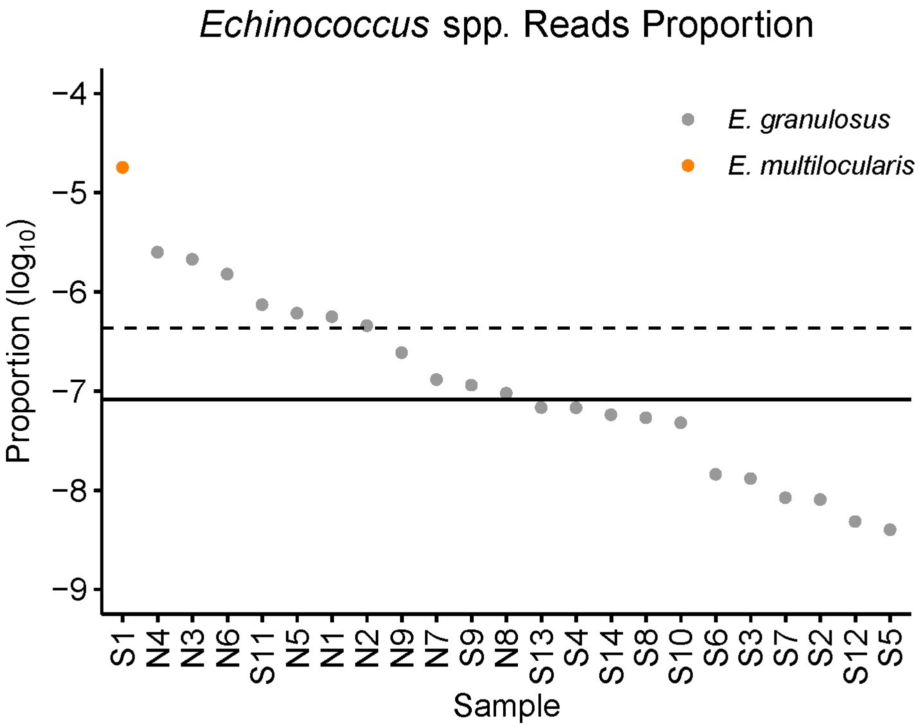

Fig 2. Cell-free *Echinococcus* spp. DNA reads proportion in total clean reads of the corresponding sample. A scatter plot shows the detected cell-free *Echinococcus* spp. read pairs proportion ($\log_{10}$) to all clean sequencing read pairs in each sample. The dashed line represents the mean value of 22 *E. granulosus* samples, and the solid line represents their median value. The results showed that the overall concentration of cell-free *Echinococcus* spp. DNA in plasma was at a low level.

https://doi.org/10.1371/journal.pntd.0008148.g002

respectively (ranging from 0.004 to 2.530) (Table 1), and the RPM of the only one AE sample was 18.096.

Based on the *Echinococcus* spp. DNA reads proportion, we calculated the probability to get at least one cell-free *Echinococcus* spp. read detection at different amounts of sequencing data (S4 Table). The results showed that sequencing with 50 million reads would make 72.73% (16/22) of CE samples with over 80% probability to get positive results, sequencing with 200 million reads would make 90.91% (20/22) of CE samples with over 80% probability to get positive results, and sequencing with 400 million reads would make all 22 CE samples with over 80% probability to get positive results.

Release source of cell-free *Echinococcus* spp. DNA

By reads mapping to the reference genomes, we traced cell-free *Echinococcus* spp. DNA to their genome release source. The analysis showed that most reads were from the nuclear genome, and only a small proportion was released from the mitochondrial genome. A small amount of mitochondrial sourced reads was identified in only 7 CE samples (7/22) and the average proportion was 2.08% (ranging from 0.74% to 7.69%) (S2 Table). To calculate reads per genome size, we normalized the reads counts by the genome size of nuclear and mitochondrial (S2 Table). For the seven CE samples detected with mitochondrial reads, reads per genome size of mitochondrial were from 5.66e-5 to 3.39e-4, and reads per genome size of nuclear were from 1.05e-7 to 4.68e-6. The reads per genome size value of mitochondria are all higher than that of nuclear, and the value of mitochondria was between 48.35 and 539.96

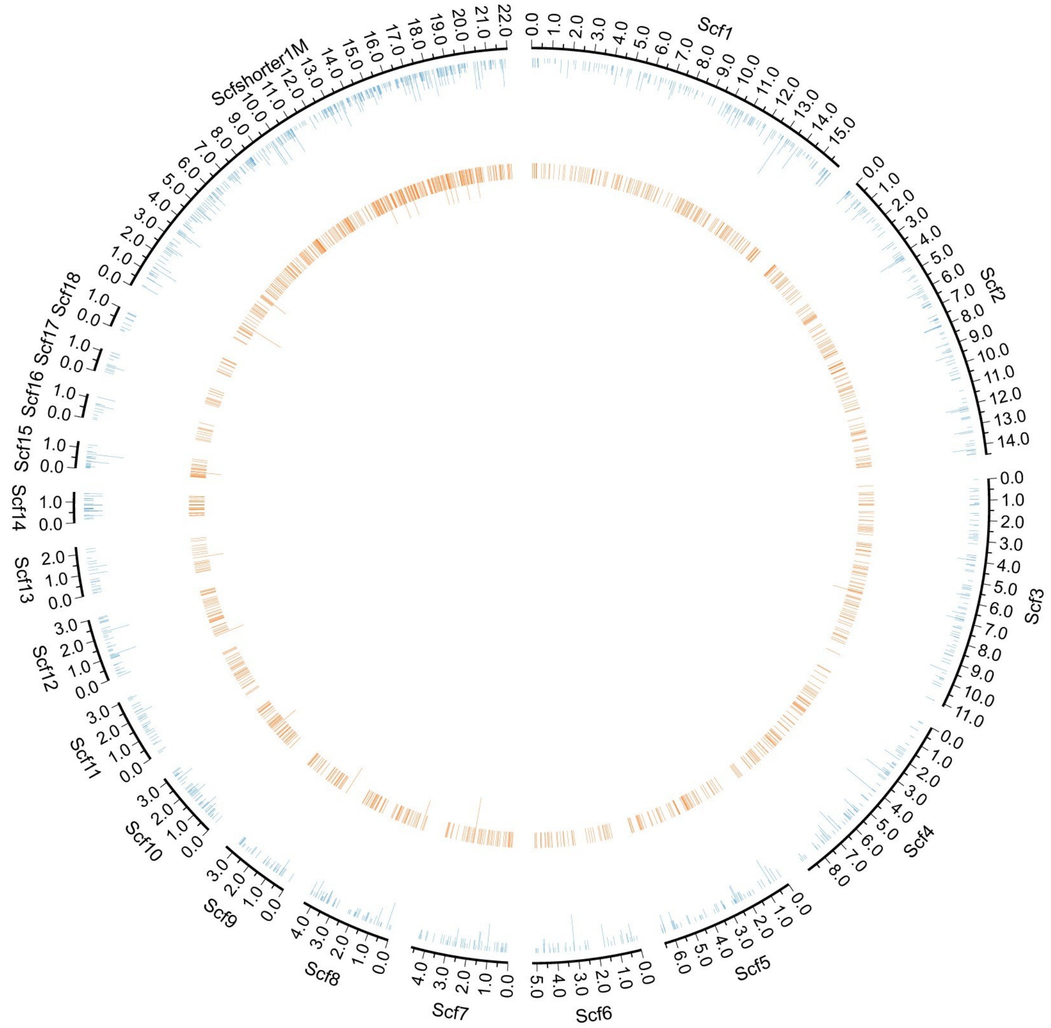

Fig 3. The distribution of cell-free *E. granulosus* DNA reads on the nuclear genome. The circulation genome visualization showed the *E. granulosus* reads mapping position on the nuclear genome (outermost blue circle). Eighteen scaffolds longer than 1Mb were displayed in the separate fragments (Scf1-Scf18). Scaffolds shorter than 1Mb were concatenated to display (Scfshort1M). The inner orange circle represents the count of patients with reads detected in the region. Circle figures of the *E. granulosus* mitochondrial genome were put in the supplementary materials (S1 Fig).

https://doi.org/10.1371/journal.pntd.0008148.g003

(median 75.78) times that of nuclear in the seven CE samples. For the AE sample, the mitochondrial sourced reads proportion was 0.19%, and reads per genome size of mitochondrial and nuclear were 1.54e-3 and 9.67e-5.

Based on the reads mapping, we further traced the release positions of cell-free *Echinococcus* spp. DNA from the genome. Given the low proportion of cell-free *Echinococcus* spp. DNA reads from the 22 CE samples, we pooled their reads and got a total of 1,852 read pairs. The number of cell-free *Echinococcus* spp. DNA read pairs of the AE sample was 11,140. These reads were mapped to the reference genomes of *E. granulosus* and *E. multilocularis* separately. The reads coverage of *E. granulosus* was 213,587 bp, which accounted for about 0.19% of the whole reference genome. The reads coverage of *E. multilocularis* was 1,232,072 bp, which accounted for about 1.07% of the whole reference genome. Mapping positions of cell-free *Echinococcus* spp. DNA reads showed that they appeared to be evenly distributed across the genomes (Fig 3, S1–S5 Figs)

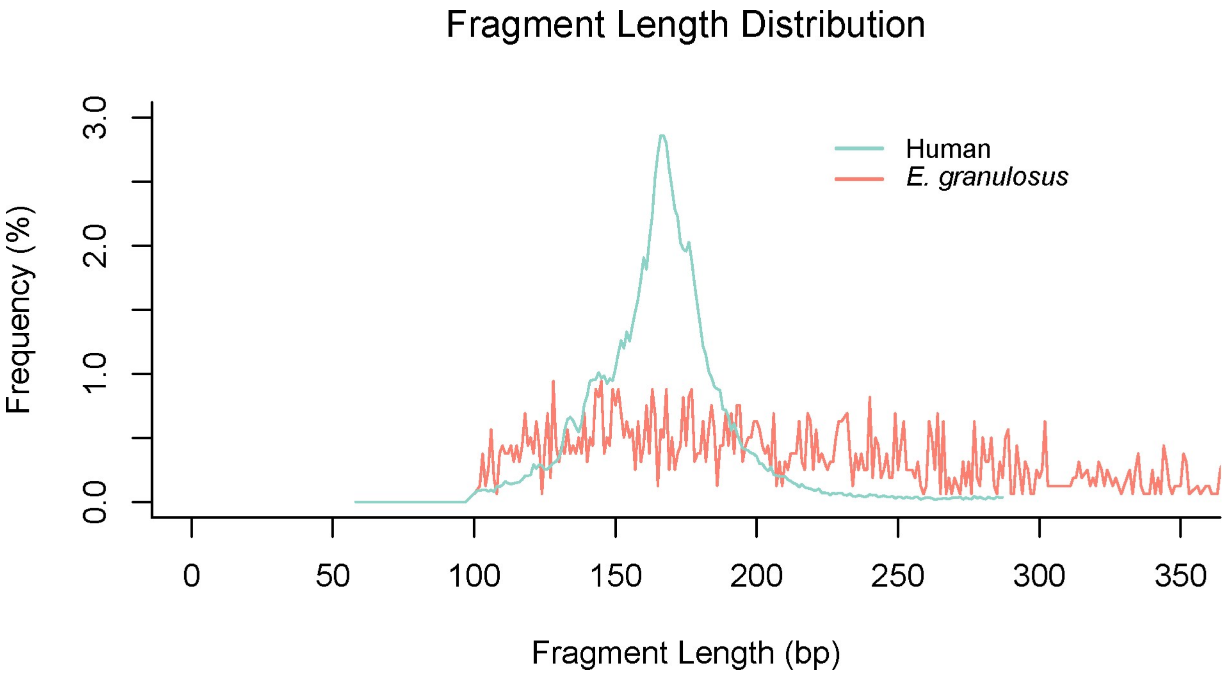

Fig 4. Fragment length distribution of cell-free *E. granulosus* DNA. The fragment length of cfDNA was calculated by the insert size of read pairs. The fragment length of cell-free *E. granulosus* DNA had a broad range than human cfDNA.

https://doi.org/10.1371/journal.pntd.0008148.g004

To analyze the distribution of identified reads in the genome, we calculated the coverage of these reads. Total coverage of 22 CE samples' reads were 213,587 bp. To analyze the overrepresented regions, we calculated the sample counts of the mapped regions. Most of the regions (211,342 bp, 98.95%) were covered with only one sample, less than 1% mapped length (1,983 bp, 0.93%) were covered with two samples, and very small region (262 bp, 0.12%) were covered with three samples. In order to intuitively compare the coverage between different samples, we plot the coverage of the samples detected with more than 100 read pairs (S2 Fig and S3 Fig).

Fragment length of cell-free *Echinococcus* spp. DNA

According to reads mapping to *Echinococcus* spp. genome references, fragment length of cell-free *Echinococcus* spp. DNA was inferred from the insert size of the read pairs. While human cfDNA showed an obvious peak at around 166 bp, cell-free *Echinococcus* spp. DNA fragment length distribution showed a different pattern. Cell-free *Echinococcus* spp. DNA fragment of CE showed a longer length range without an obvious peak (Fig 4). Cell-free *Echinococcus* spp. DNA fragment of the AE sample showed a more regular distribution pattern with an obvious peak at about 135 bp, which was shorter than human cfDNA (S6 Fig).

Discussion

With ultra-high-throughput sequencing technology, using plasma samples from clinically diagnosed echinococcosis patients, we identified the existence of cell-free *Echinococcus spp*. DNA in plasma, quantified the amount per sample, confirmed its low concentration and described its characteristics. The results revealed that high-throughput sequencing of plasma cfDNA could serve as a useful tool for cell-free *Echinococcus* spp. DNA studies and improve our understanding of *Echinococcus* spp. infection in the human body. Plasma cfDNA has shown its usefulness in NIPT [23], tumor monitoring [24], and pathogens detection [25].

Several attempts were made using cfDNA in *Echinococcus* spp. detection from plasma or serum with PCR-based methods, but their overall sensitivity was only 20% to 25% [31–33]. The low sensitivity could be due to non-existence, or low concentration of cfDNA of the parasite in the circulation, which showed our limited understanding of the cfDNA of *Echinococcus spp*. tapeworms. As one of the most neglected tropical diseases and zoonosis, echinococcosis poses serious public health threats to endemic areas. Given the increase of global trade, tourism, and immigration, people of non-endemic regions could also be diagnosed with echinococcosis [50–52]. Effective detection and diagnosis methods are the premises of controlling echinococcosis, and cfDNA could be a promising tool for clinical diagnosis. We are the first using high-throughput sequencing technology to evaluate the existence, quantity, and characteristics of cell-free *Echinococcus* spp. DNA in plasma of echinococcosis patients.

The existence of *E. granulosus* DNA in the blood circulation of the echinococcosis patients was questioned by Chaya *et al.* who believed that the cfDNA of the parasite would only enter the blood circulation when the hydatid cyst(s) ruptured [31]. Baraquin *et al.* confirmed the existence of cfDNA of *E. multilocularis* in AE patients and used the very low concentration of cfDNA to explain the low sensitivity of their PCR test [32]. Low concentration of target DNA in plasma is a common situation for cfDNA studies. Cell-free fetal DNA in maternal plasma cfDNA accounts for about 10% to 15% [53,54], and circulating-tumor DNA comprises about 0.01% to 10% or more in cancer patients plasma cfDNA [55,56]. Based on high-throughput sequencing data and bioinformatics workflow, we identified the cell-free *Echinococcus* spp. DNA reads from sequencing data of all the samples. Compared with cell-free fetal DNA and circulating-tumor DNA, cell-free *Echinococcus* spp. DNA in plasma is extremely low, whose proportion ranged 1.8e-5 to 4.0e-9 in these samples (Fig 2). Indeed, this low concentration may explain the low sensitivity of the PCR-based methods [31–33]. Besides the low concentration, we identified the difference between the cell-free *Echinococcus* spp. DNA from CE and AE samples. The AE sample had much more cfDNA identified than the CE samples, which could be due to the different developmental mechanisms of metacestode in the human body. Compared with *E. granulosus*, the metacestode of *E. multilocularis* is an infiltrating lesion composed of aggregated microvesicles, necrosis cells, and fibrosis cells, which have no clear edge to the host tissues [1] and relatively high concentration of cell-free *Echinococcus* spp. DNA in the AE sample could be due to the mixture of necrotic parasite tissue and actively proliferating tissues. This is in line with the previous finding that the sensitivity of PCR-based methods in AE samples was higher than in CE samples [31–33]. As we only collected one AE sample, it needed further verification with more samples. Plasma samples were also tested with ELISA assays to detect the antibody, and the positive results were found in 16 out of 22 CE patients (Table 1). Serological tests may be influenced by lots of factors, and difficult to standardize [18–20]. In contrast, DNA detection is a more direct and objective biomarker.

Low concentration is the major challenge to apply cell-free *Echinococcus* spp. DNA testing in routine clinical settings. To estimate the minimal number of reads needed to get cell-free *Echinococcus* spp. DNA, we treated sequencing as a random sampling process, and the number of sequencing reads regarded as sampling times. We estimated total cfDNA fragment counts and *Echinococcus* spp. fragment counts according to cfDNA concentration and existed detection results. By using hypergeometric distribution, we calculated the probability of each sample to get cell-free *Echinococcus* spp. DNA. detection with different sequencing amounts. We found that sequencing with 50 million reads would make 72.73% (16/22) of CE samples with over 80% probability to get positive results, while sequencing with 400 million reads would make all 22 CE samples identified with over 80% probability to get positive results (S4 Table). The lower the concentration, the harder it is to be detected, and increasing the amount of sequencing can increase the chance of positive detection. The concentration may vary greatly between individuals. Just like the cell-free DNA of fetus in maternal plasma, which are influenced by gestational age, maternal BMI, fetal aneuploidy status and other factors [57]. The concentration of cell-free *Echinococcus* spp. DNA might also be affected by many factors, such as disease status, parasite species, lesion size, and position, which need more comprehensive samples to explore its association with different patterns.

Cell-free DNA *Echinococcus* spp. DNA in plasma could not only detect the etiology of the patients' infection but also facilitate the species identification. Traditional species identification of *Echinococcus* spp. in echinococcosis patients is always invasive, which relies on the product of surgery or puncture. Surgery is only recommended for part of echinococcosis patients, puncture can assist to get specimens for confirming etiology. However, while puncture is of high diagnostic value and safe in most AE patients [58], it is not recommended for some CE patients, especially for CE4, CE5 and lung cysts, which may pose the risks of allergic reactions and anaphylaxis [1,8]. In this study, species annotation of cell-free *Echinococcus* spp. DNA was analyzed according to reads mapping results. Given the genome sequence similarity between *Echinococcus* species and limited reference sequences available, part of cell-free *Echinococcus* spp. DNA reads may be classified into closely related species of genus *Echinococcus*, but the majority of the reads should be classified correctly. Consistency of species classification between cell-free *Echinococcus* spp. DNA and lesion samples' PCR results proved their accuracy in species annotation. This cfDNA sequencing-based taxonomy annotation method may provide an innovative non-invasive alternative to obtain more detailed etiology information. Species identification of echinococcosis patients could provide more valuable information for guiding clinical management and research such as molecular epidemiology [59].

Tracing cell-free *Echinococcus* spp. DNA release sources could provide more background information. Based on cell-free *Echinococcus* spp. DNA reads mapping, we further analyzed their genome release source. Sequence origin analysis showed that much more cell-free *Echinococcus* spp. DNA was released from the nuclear genome than the mitochondrial genome. This phenomenon may be due to the fact that the genome size of nuclear is much larger than mitochondria. The overall low proportion of mitochondrial-derived cell-free *Echinococcus* spp. DNA in plasma may also partially explain the low positive rate of mitochondrial gene based PCR [31–33]. However, reads per genome size of mitochondria were about 75.78 times larger than that of nuclear, which could be due to the multi copies of mitochondria [44]. The position distribution of cell-free *Echinococcus* spp. DNA on the genome were analyzed with reads mapping to *E. granulosus* and *E. multilocularis* genome references. We found that the release positions of cell-free *Echinococcus* spp. DNA were nearly evenly distributed on the genome. It looks like there are some hotspots of cell-free *Echinococcus* spp. DNA release on the genome, but these spots are more gathered on the short and not well-assembled regions of the available genome references. With higher quality references in the future, the distribution of cell-free *Echinococcus* spp. DNA on the genome could be more evenly distributed.

Size characteristics of cfDNA is an important biological property [35]. To have a deep understanding of cell-free *Echinococcus* spp. DNA, we analyzed its fragment size with sequencing data. Literature shows that cfDNA could have different size pattern according to research settings [35]. Fetal cfDNA in maternal plasma has a shorter fragment size distribution compared with maternal cfDNA [60]. In certain types of cancer patients, tumor sourced cfDNA is concentrated in short fragments [61]. Fragment size analysis of cell-free *Echinococcus* spp. DNA in our study showed that they had a different length distribution to human-sourced cfDNA. We found that cell-free *Echinococcus* spp. DNA of CE had a broad length range (Fig 4), but that of AE had an obvious peak at about 135 bp (S6 Fig). The size profile of cfDNA is relevant to their release mechanism such as apoptosis, necrosis and actively release [62,63]. Quite different fragment size features of cell-free *Echinococcus* spp. DNA in CE and AE could be related to their developmental mechanism of metacestode in the human body. Tumor like AE lesions may give some explanation to its overall short fragment length, and similar phenomenon of tumor-derived DNA in plasma of hepatocellular carcinoma patients was also observed [64]. As there was only one accidental AE sample, this phenomenon needs more research to validate. Though the exact release mechanism of cfDNA is still unclear, it doesn't affect the application of size properties in diagnostics [35]. As for cell-free *Echinococcus* spp. DNA, their fragment size features may facilitate their detection in future studies.

The cfDNA sequencing-based method relies on high quality and comprehensive database, but existing genome references of *Echinococcus* spp. are limited, and only several genome references are available [44,65,66] whose quality is far from perfect. More importantly, sequence contamination is a serious problem for cell-free *Echinococcus* spp. DNA detection and contaminated sequence database might introduce false-positive results. Since the *Echinococcus* spp. tapeworm samples are always separated from host tissue [44,65,66], it is not easy to remove the contamination of host thoroughly by experimental processing. In the process of genome constructing, some host sequences may mix into the parasite sequence, which is a common problem for genomes construction [67]. It is essential to qualify the genome sequence with bioinformatics methods after downloading from the public database, instead of using it directly [67]. In our study, we filtered the *Echinococcus* spp. sequence database with their common host genomes such as sheep, humans, and mice, and evaluated the workflow with simulation data, cell line data, and negative control data, which all showed that qualified database introduced no false-positive results.

High-throughput sequencing facilitated identifying, quantifying and analyzing the characteristics of cell-free *Echinococcus* spp. DNA in human plasma. These comprehensive characteristics could help the application of cell-free *Echinococcus* spp. DNA in the future diagnosis of echinococcosis. However, for the very low concentration of cell-free *Echinococcus* spp. DNA, their even distribution on the genome, and the high sequencing depth and cost, the method requires further optimization. To increase the application of cell-free *Echinococcus* spp. DNA, we could think of some areas to be explored in the future study, for example, capturing cell-free *Echinococcus* spp. DNA with probes covered the whole genome and enriching the concentration of cell-free *Echinococcus* spp. DNA by host sequence removal. As for clinical application scenarios, massive sequencing of plasma cfDNA to detect cell-free *Echinococcus* spp. DNA may not be suitable for routine clinical examination yet, but it could be used for differential diagnosis, in which existing clinically methods cannot give clear conclusions. For example, the CL patients can be further diagnosed with plasma cfDNA sequencing and avoid the risk of invasive diagnosis.

Supporting information

S1 Table. Evaluation of analysis workflow with simulation data. Simulation data showed that no human reads appeared in the results, and most *Echinococcus* spp. reads were identified by the analysis workflow. Counts in the table were read pairs.
(XLSX)

S2 Table. Release source of identified *Echinococcus* spp. reads. Most of the identified *Echinococcus* spp. reads were released from the nuclear genome. Only eight samples were identified with mitochondrial reads.
(XLSX)

S3 Table. Species classification with cfDNA reads mapping. The table showed the species classification results from each sample with cfDNA sequencing read pairs. The sample was

classified to the species with most read pairs mapping.
(XLSX)

S4 Table. Amount of sequencing data and reads detection. Statistical analysis with hypergeometric distribution to estimate the probability to get positive results with different sequencing amount.
(XLSX)

S5 Table. Plasma volume and DNA extraction methods. The volume of plasma and kit used for each sample.
(XLSX)

S1 Fig. Circle figure of *E. granulosus* samples based on the mitochondrial genome. The circulation genome visualization showed the *E. granulosus* reads mapping position (outermost blue circle). The inner orange circle represents the count of patients with reads detected in the region.
(TIFF)

S2 Fig. Circle figure of multiple *E. granulosus* samples based on the nuclear genome. Seven *E. granulosus* samples detected with more than 100 *Echinococcus* spp. read pairs were displayed based on the nuclear genome. Green and red circles indicate different samples.
(TIF)

S3 Fig. Circle figure of multiple *E. granulosus* samples based on the mitochondrial genome. Seven *E. granulosus* samples detected with more than 100 *Echinococcus* spp. read pairs were displayed based on the mitochondrial genome. Green and red circles indicate different samples.
(TIF)

S4 Fig. Circle figure of the *E. multilocularis* sample based on the nuclear genome. The circulation genome visualization showed the *E. multilocularis* reads mapping position (outermost blue circle). Ten scaffolds longer than 1Mb were displayed in the separate fragment (Scf1-Scf10). Scaffolds shorter than 1Mb were concatenated to display (Scfshort1M).
(TIF)

S5 Fig. Circle figure of the *E. multilocularis* sample based on the mitochondrial genome. The circulation genome visualization showed the *E. multilocularis* reads mapping position (outermost blue circle).
(TIF)

S6 Fig. Fragment length distribution of the *E. multilocularis* sample. Overall fragment length distribution of *E. multilocularis* cfDNA was shorter than that of humans.
(TIFF)

Acknowledgments

The authors would like to thank China National GeneBank for sequencing the samples of this study. We would also like to thank Yue Zhao and Ying Yi for their previous analysis work on the project.

Author Contributions

Conceptualization: Bin Li, Jingzhong Li, Jinmin Ma, Asan, Xiumin Han, Weijun Chen, Dong Yang.

Data curation: Wangmu Danzeng, Yanping Zhao, Yicong Wang, Fang Chen.

Formal analysis: Jingkai Ji, Jiandan Xie, Huimin Cai, Junhua Li.

Funding acquisition: Bin Li, Jingzhong Li.

Investigation: Jingkai Ji, Qingda Zhang, Nibu Renzhen, Zhuoga Basang, Changlin Jia, Quzhen Gongsang, Jinmin Ma, Yicong Wang, Fang Chen, Huasang, Xiumin Han.

Methodology: Jingkai Ji, Jiandong Li, Gezhen Qiangba, Na Pei, Junhua Li.

Project administration: Jiefang Yin, Huayan Jiang, Weijun Chen, Dong Yang.

Resources: Jingzhong Li, Wangmu Danzeng, Nibu Renzhen, Zhuoga Basang, Changlin Jia, Quzhen Gongsang, Hongcheng Zhou, Huasang.

Software: Jingkai Ji, Jiandong Li, Jiandan Xie, Huimin Cai, Junhua Li.

Supervision: Qingda Zhang, Jiefang Yin, Huanming Yang, Jian Wang, Asan, Weijun Chen, Dong Yang.

Validation: Jiandong Li, Gezhen Qiangba, Na Pei.

Visualization: Jingkai Ji, Jiandong Li, Gezhen Qiangba, Huimin Cai.

Writing – original draft: Jingkai Ji.

Writing – review & editing: Bin Li, Jingzhong Li, Yanping Zhao, Junhua Li, Weijun Chen, Dong Yang.

References

1. Wen H, Vuitton L, Tuxun T, Li J, Vuitton DA, Zhang W, et al. Echinococcosis: Advances in the 21st century. Clin Microbiol Rev. 2019; 32: 1–39. https://doi.org/10.1128/CMR.00075-18 PMID: 30760475
2. Kern P, Menezes da Silva A, Akhan O, Müllhaupt B, Vizcaychipi KAA, Budke C, et al. The Echinococcoses:Diagnosis, Clinical Management and Burden of Disease. Advances in Parasitology. Academic Press; 2017. pp. 259–369. https://doi.org/10.1016/bs.apar.2016.09.006 PMID: 28212790
3. da Silva AM. Human Echinococcosis: A Neglected Disease. Gastroenterol Res Pract. 2010; 2010: 1–9. https://doi.org/10.1155/2010/583297 PMID: 20862339
4. Casulli A, Siles-Lucas M, Tamarozzi F. *Echinococcus* granulosus sensu lato. Trends Parasitol. 2019; 35: 5–6. https://doi.org/10.1016/j.pt.2019.05.006 PMID: 31182386
5. Casulli A, Barth TFE, Tamarozzi F. *Echinococcus* multilocularis. Trends Parasitol. 2019; 35: 738–739. https://doi.org/10.1016/j.pt.2019.05.005 PMID: 31182385
6. Agudelo Higuita NI, Brunetti E, McCloskey C. Cystic Echinococcosis. Kraft CS, editor. J Clin Microbiol. 2016; 54: 518–523. https://doi.org/10.1128/JCM.02420-15 PMID: 26677245
7. Brunetti E, Junghanss T. Update on cystic hydatid disease. Curr Opin Infect Dis. 2009; 22: 497–502. https://doi.org/10.1097/QCO.0b013e328330331c PMID: 19633552
8. Brunetti E, Kern P, Vuitton DA. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. Acta Trop. 2010; 114: 1–16. https://doi.org/10.1016/j.actatropica.2009.11.001 PMID: 19931502
9. Stojković M, Weber TF, Junghanss T. Clinical management of cystic echinococcosis: state of the art and perspectives. Curr Opin Infect Dis. 2018; 31: 383–392. https://doi.org/10.1097/QCO.0000000000000485 PMID: 30124496
10. Grüner B, Schmidberger J, Drews O, Kratzer W, Gräter T. Imaging in alveolar echinococcosis (AE): Comparison of *Echinococcus* multilocularis classification for computed-tomography (EMUC-CT) and ultrasonography (EMUC-US). Radiol Infect Dis. 2017; 4: 70–77. https://doi.org/10.1016/j.jrid.2017.05.001
11. Mihmanli M, Idiz UO, Kaya C, Demir U, Bostanci O, Omeroglu S, et al. Current status of diagnosis and treatment of hepatic echinococcosis. World J Hepatol. 2016; 8: 1169. https://doi.org/10.4254/wjh.v8.i28.1169 PMID: 27729953

12. Manzano-Román R, Sánchez-Ovejero C, Hernández-González A, Casulli A, Siles-Lucas M. Serological Diagnosis and Follow-Up of Human Cystic Echinococcosis: A New Hope for the Future? Biomed Res Int. 2015; 2015: 1–9. https://doi.org/10.1155/2015/428205 PMID: 26504805

13. Li B, Quzhen G, Xue C-Z, Han S, Chen W-Q, Yan X-L, et al. Epidemiological survey of echinococcosis in Tibet Autonomous Region of China. Infect Dis Poverty. 2019; 8: 29. https://doi.org/10.1186/s40249-019-0537-5 PMID: 31030673

14. Brunetti E, Tamarozzi F, Macpherson C, Filice C, Piontek M, Kabaalioglu A, et al. Ultrasound and Cystic Echinococcosis. Ultrasound Int Open. 2018; 04: E70–E78. https://doi.org/10.1055/a-0650-3807 PMID: 30364890

15. Polat P, Kantarci M, Alper F, Suma S, Koruyucu MB, Okur A. Hydatid Disease from Head to Toe. RadioGraphics. 2003; 23: 475–494. https://doi.org/10.1148/rg.232025704 PMID: 12640161

16. Cattaneo F, Graffeo M, Brunetti E. Extrahepatic Textiloma Long Misdiagnosed as Calcified Echinococcal Cyst. Case Rep Gastrointest Med. 2013; 2013: 1–5. https://doi.org/10.1155/2013/261685 PMID: 23533840

17. Engler A, Shi R, Beer M, Schmidberger J, Kratzer W, Barth TFE, et al. Simple liver cysts and cystoid lesions in hepatic alveolar echinococcosis: A retrospective cohort study with Hounsfield analysis. Parasite. 2019; 26. https://doi.org/10.1051/parasite/2019057 PMID: 31469072

18. Zhang W, McManus DP. Recent advances in the immunology and diagnosis of echinococcosis. FEMS Immunol Med Microbiol. 2006; 47: 24–41. https://doi.org/10.1111/j.1574-695X.2006.00060.x PMID: 16706785

19. Carmena D, Benito A, Eraso E. Antigens for the immunodiagnosis of *Echinococcus* granulosus infection: An update. Acta Trop. 2006; 98: 74–86. https://doi.org/10.1016/j.actatropica.2006.02.002 PMID: 16527225

20. Lissandrin R, Brunetti E, Tinelli C, Piccoli L, Tamarozzi F, Mariconti M, et al. Factors Influencing the Serological Response in Hepatic *Echinococcus* granulosus Infection. Am J Trop Med Hyg. 2016; 94: 166–171. https://doi.org/10.4269/ajtmh.15-0219 PMID: 26503271

21. Gottstein B, Wang J, Blagosklonov O, Grenouillet F, Millon L, Vuitton DA, et al. *Echinococcus* metacestode: in search of viability markers. Parasite. 2014; 21: 63. https://doi.org/10.1051/parasite/2014063 PMID: 25429386

22. Weerakoon KG, McManus DP. Cell-Free DNA as a Diagnostic Tool for Human Parasitic Infections. Trends Parasitol. 2016; 32: 378–391. https://doi.org/10.1016/j.pt.2016.01.006 PMID: 26847654

23. Norwitz ER, Levy B. Noninvasive prenatal testing: the future is now. Rev Obstet Gynecol. 2013; 6: 48–62. https://doi.org/10.3909/riog0201 PMID: 24466384

24. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem. 2015; 61: 112–123. https://doi.org/10.1373/clinchem.2014.222679 PMID: 25388429

25. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol. 2019; 4: 663–674. https://doi.org/10.1038/s41564-018-0349-6 PMID: 30742071

26. Ghayour Najafabadi Z, Oormazdi H, Akhlaghi L, Meamar AR, Nateghpour M, Farivar L, et al. Detection of *Plasmodium vivax* and *Plasmodium falciparum* DNA in human saliva and urine: Loop-mediated isothermal amplification for malaria diagnosis. Acta Trop. 2014; 136: 44–49. https://doi.org/10.1016/j.actatropica.2014.03.029 PMID: 24721227

27. Russomando G, Figueredo A, Almiron M, Sakamoto M, Morita K. Polymerase chain reaction-based detection of *Trypanosoma cruzi* DNA in serum. J Clin Microbiol. 1992.

28. Calderon F, Low DE, Ramos AP, Arevalo J, Veland N, Boggild AK, et al. Polymerase Chain Reaction Detection of *Leishmania* kDNA from the Urine of Peruvian Patients with Cutaneous and Mucocutaneous Leishmaniasis. Am J Trop Med Hyg. 2011; 84: 556–561. https://doi.org/10.4269/ajtmh.2011.10-0556 PMID: 21460009

29. Wichmann D, Panning M, Quack T, Kramme S, Burchard G-DD, Grevelding C, et al. Diagnosing schistosomiasis by detection of cell-free parasite DNA in human plasma. Correa-Oliveira R, editor. PLoS Negl Trop Dis. 2009; 3: e422. https://doi.org/10.1371/journal.pntd.0000422 PMID: 19381285

30. Ximenes C, Brandão E, Oliveira P, Rocha A, Rego T, Medeiros R, et al. Detection of Wuchereria bancrofti DNA in paired serum and urine samples using polymerase chain reaction-based systems. Mem Inst Oswaldo Cruz. 2014; 109: 978–983. https://doi.org/10.1590/0074-0276140155 PMID: 25424447

31. Parija S, Chaya D. Performance of polymerase chain reaction for the diagnosis of cystic echinococcosis using serum, urine, and cyst fluid samples. Trop Parasitol. 2014; 4: 43. https://doi.org/10.4103/2229-5070.129164 PMID: 24754027

32. Baraquin A, Hervouet E, Richou C, Flori P, Peixoto P, Azizi A, et al. Circulating cell-free DNA in patients with alveolar echinococcosis. Mol Biochem Parasitol. 2018; 222: 14–20. https://doi.org/10.1016/j.molbiopara.2018.04.004 PMID: 29679605

33. Moradi M, Meamar AR, Akhlaghi L, Roozbehani M, Razmjou E. Detection and genetic characterization of *Echinococcus* granulosus mitochondrial DNA in serum and formalin-fixed paraffin embedded cyst tissue samples of cystic echinococcosis patients. PLoS One. 2019; 14: e0224501. https://doi.org/10.1371/journal.pone.0224501 PMID: 31661532

34. Yu SCY, Chan KCA, Zheng YWL, Jiang P, Liao GJW, Sun H, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. Proc Natl Acad Sci. 2014; 111: 8583–8588. https://doi.org/10.1073/pnas.1406103111 PMID: 24843150

35. Jiang P, Lo YMD. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. Trends Genet. 2016; 32: 360–371. https://doi.org/10.1016/j.tig.2016.03.009 PMID: 27129983

36. Boufana B, Umhang G, Qiu J, Chen X, Lahmar S, Boué F, et al. Development of three PCR assays for the differentiation between *Echinococcus shiquicus*, *E. granulosus* (G1 genotype), and *E. multilocularis* DNA in the co-endemic region of Qinghai-Tibet plateau, China. Am J Trop Med Hyg. 2013; 88: 795–802. https://doi.org/10.4269/ajtmh.12-0331 PMID: 23438764

37. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 421. https://doi.org/10.1186/1471-2105-10-421 PMID: 20003500

38. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

39. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018; 7: 1–6. https://doi.org/10.1093/gigascience/gix120 PMID: 29220494

40. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018. https://doi.org/10.1093/bioinformatics/bty560 PMID: 30423086

41. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014; 15: R46. https://doi.org/10.1186/gb-2014-15-3-r46 PMID: 24580807

42. Zaharia Matei. Faster and More Accurate Sequence Alignment with SNAP. Data Struct Algorithms. 2011.

43. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27: 863–864. https://doi.org/10.1093/bioinformatics/btr026 PMID: 21278185

44. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013; 496: 57–63. https://doi.org/10.1038/nature12031 PMID: 23485966

45. Kinkar L, Korhonen PK, Cai H, Gauci CG, Lightowlers MW, Saarma U, et al. Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (sensu stricto) genotype G1. Parasit Vectors. 2019; 12: 238. https://doi.org/10.1186/s13071-019-3492-x PMID: 31097022

46. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

47. Connors J, Krzywinski M, Schein J, Gascoyne R, Horsman D, Jones SJ, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009; 19: 1639–1645. https://doi.org/10.1101/gr.092759.109 PMID: 19541911

48. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry A. 2003; 51: 127–8; author reply 129. https://doi.org/10.1002/cyto.a.10013 PMID: 12541287

49. Fernández-Carballo BL, Broger T, Wyss R, Banaei N, Denkinger CM. Toward the development of a circulating free DNA-Based in vitro diagnostic test for infectious diseases: A review of evidence for tuberculosis. J Clin Microbiol. 2018; 57: 1–9. https://doi.org/10.1128/JCM.01234-18_rfseq1

50. Zhang KJ, Schaldenbrand M, Turfah F. Multiorgan *Echinococcus* infection: Treatment of an immigrant in the United States. IDCases. 2017. https://doi.org/10.1016/j.idcr.2017.05.011 PMID: 28660127

51. Angheben A, Mariconti M, Degani M, Gobbo M, Palvarini L, Gobbi F, et al. Is there echinococcosis in West Africa? A refugee from Niger with a liver cyst. Parasit Vectors. 2017; 10: 232. https://doi.org/10.1186/s13071-017-2169-6 PMID: 28494818

52. Massolo A, Klein C, Kowalewska-Grochowska K, Belga S, MacDonald C, Vaughan S, et al. European *Echinococcus multilocularis* Identified in Patients in Canada. N Engl J Med. 2019; 381: 384–385. https://doi.org/10.1056/NEJMc1814975 PMID: 31340100

53. Wang E, Batey A, Struble C, Musci T, Song K, Oliphant A. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. Prenat Diagn. 2013; 33: 662–666. https://doi.org/10.1002/pd.4119 PMID: 23553731

54. Barrett AN, Xiong L, Tan TZ, Advani H V., Hua R, Laureano-Asibal C, et al. Measurement of fetal fraction in cell-free DNA from maternal plasma using a panel of insertion/deletion polymorphisms. PLoS One. 2017; 12: 1–16. https://doi.org/10.1371/journal.pone.0186771 PMID: 29084245

55. Elazezy M, Joosse SA. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. Comput Struct Biotechnol J. 2018; 16: 370–378. https://doi.org/10.1016/j.csbj.2018.10.002 PMID: 30364656

56. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012; 4. https://doi.org/10.1126/scitranslmed.3003726 PMID: 22649089

57. Kinnings SL, Geis JA, Almasri E, Wang H, Guan X, Mccullough RM, et al. Factors affecting levels of circulating cell-free fetal DNA in maternal plasma and their implications for noninvasive prenatal testing. Prenat Diagn. 2015; 35: 816–822. https://doi.org/10.1002/pd.4625 PMID: 26013964

58. Bulakci M, Ilhan M, Bademler S, Yilmaz E, Gulluoglu M, Bayraktar A, et al. Efficacy of ultrasound-guided core-needle biopsy in the diagnosis of hepatic alveolar echinococcosis: A retrospective analysis. Parasite. 2016; 23. https://doi.org/10.1051/parasite/2016019 PMID: 27101838

59. Cucher MA, Macchiaroli N, Baldi G, Camicia F, Prada L, Maldonado L, et al. Cystic echinococcosis in South America: systematic review of species and genotypes of *Echinococcus granulosus* sensu lato in humans and natural domestic hosts. Trop Med Int Heal. 2016; 21: 166–175. https://doi.org/10.1111/tmi.12647 PMID: 26610060

60. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. Sci Transl Med. 2010; 2: 61ra91–61ra91. https://doi.org/10.1126/scitranslmed.3001720 PMID: 21148127

61. Chan KCA, Zhang J, Chan ATC, Lei KIK, Leung S-F, Chan LYS, et al. Molecular characterization of circulating EBV DNA in the plasma of nasopharyngeal carcinoma and lymphoma patients. Cancer Res. 2003; 63: 2028–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/12727814 PMID: 12727814

62. Van Der Vaart M, Pretorius PJ. The origin of circulating free DNA [1]. Clin Chem. 2007; 53: 2215. https://doi.org/10.1373/clinchem.2007.092734 PMID: 18267930

63. Aucamp J, Bronkhorst AJ, Badenhorst CPS, Pretorius PJ. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. Biol Rev. 2018; 93: 1649–1683. https://doi.org/10.1111/brv.12413 PMID: 29654714

64. Jiang P, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong VWS, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci U S A. 2015; 112: E1317–E1325. https://doi.org/10.1073/pnas.1500076112 PMID: 25646427

65. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, et al. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nat Genet. 2013; 45: 1168–1175. https://doi.org/10.1038/ng.2757 PMID: 24013640

66. Maldonado LL, Assis J, Araújo FMG, Salim ACM, Macchiaroli N, Cucher M, et al. The *Echinococcus canadensis* (G7) genome: A key knowledge of parasitic platyhelminth human diseases. BMC Genomics. 2017; 18: 1–23. https://doi.org/10.1186/s12864-016-3406-7 PMID: 28049423

67. Lu J, Salzberg SL. Removing contaminants from databases of draft genomes. Sun F, editor. PLoS Comput Biol. 2018; 14: 1–13. https://doi.org/10.1371/journal.pcbi.1006277 PMID: 29939994