Veterinary Parasitology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Research paper

Comparative study on secretome and transmembranome of immature and mature metacestodes of *Echinococcus multilocularis*

Fuqiang Huang[a,b,g,1], Zhisheng Dang[c,d,e,f,1,**], Haoji Zhang[g], Kinpei Yagi[h], Kyeongsoon Kim[a], Magona Joseph[i], Wenjun Pu[g], Yuzaburo Oku[a,*]

[a] Parasitology Laboratory, School of Veterinary Medicine, Faculty of Agriculture, Tottori University, 680-0945, Tottori, Japan
[b] The United Graduate School of Veterinary Science, Yamaguchi University, Yamaguchi 753-8515, Japan
[c] National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, People's Republic of China
[d] Key Laboratory of Parasite and Vector Biology, Ministry of Health, MOH, Shanghai 200025, People's Republic of China
[e] National Center for International Research on Tropical Diseases, Shanghai 200025, People's Republic of China
[f] WHO Collaborating Center for Tropical Diseases, Shanghai 200025, People's Republic of China
[g] School of Life Science and Engineering, Foshan University, Foshan 528231, People's Republic of China
[h] Department of Infectious Disease, Hokkaido Institute of Public Health, Sapporo, Hokkaido 060-0819, Japan
[i] Veterinary Epidemiologist AU-IBAR/ICPALD, African Union-Interafrican Bureau for Animal Resources, Nairobi 30786-00100, Kenya

ARTICLE INFO

Keywords:
Echinococcus multilocularis
Metacestode
In vivo
Transcriptome
ES protein
TM protein

ABSTRACT

Alveolar echinococcosis (AE) is a worldwide zoonosis caused by *E. multilocularis*. Humans become infected through oral ingestion of the eggs. Host of *E. multilocularis* produces immune responses that help to either reject and/or limit the growth of this parasite, and in response the parasite produces molecules against this immune attack. This study identifies candidate key molecules in the early infection phase and the chronic stage of the parasite infestation, through comparison of gene expression of 4- and 16-week metacestodes. First, RNA was isolated from 4- and 16-weeks metacestodes of *E. multilocularis* (Nemuro strain). Thereafter, clean reads with lengths of 50 bp or longer were compared against a reference genome using TopHat. Functional annotation of transcripts of *E. multilocularis* were investigated using multi-step bioinformatics tools. At the gene ontology (GO) level, 356 and 1774 transmembrane (TM) predicted proteins of the *E. multilocularis* were mapped to an enhanced 'hydrolase activity' and increased 'transmembrane transporter activity', respectively. In addition, comparison of gene expression level between 4- and 16-week metacestode revealed 168 different expression (DE) genes. This study has demonstrated that, the expression levels of predicted ES and TM proteins in *E. multilocularis* change in the transformation from one stage to another. Genes that are highly expressed in immature or mature metacestode could be explored as novel candidates for diagnostic antigens and vaccine targets.

1. Introduction

Hosts of *E. multilocularis* produces immune responses to reject and/or limit the growth of this parasite. The parasite can also produce molecules to avoid these immune attacks (Zhang et al., 2008, 2012). With immune responses to larval *Echoinococcus* spp. infections divided into "establishment" and "established metacestode" phases. The parasite is thought to be more susceptible to immune attack during the "establishment" phase (Siracusano et al., 2012).

DBA/2 mice are thought to be highly susceptibility to alveolar echinococcosis (AE) based on mature protoscolex formation and subsequent active growth of larval parasites in 4 inbred strains of mice (Matsumoto et al., 2010). Differential expression of stage-specific molecules in *in vivo* and *in vitro* 4-week metacestodes has been clearly demonstrated in this parasite, suggesting that differently expressed molecules may play an important role in the process of *E. multilocularis* infection and modulation of the immune response (Huang et al., 2016). Moreover, the specific IgG and IgM levels in DBA/2 mice against crude antigens became positive at 4 or 9 weeks post-infection and continued to increase until 16 weeks post-infection (Matsumoto et al., 2010) suggesting that metabolism of the parasite and host responses vary during different growth periods of metacestodes. However, gene expression profile data of metacestodes based on experimental infection through oral ingestion of parasite eggs (termed primary AE) remains

* Corresponding author.
** Corresponding author at: National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, People's Republic of China.
E-mail addresses: 447497081@qq.com (Z. Dang), oku@muses.tottori-u.ac.jp (Y. Oku).
[1] These authors contributed equally to this work and should be considered co-first authors.

http://dx.doi.org/10.1016/j.vetpar.2017.05.006
Received 20 January 2017; Received in revised form 1 May 2017; Accepted 11 May 2017

lacking.

To gain an understanding of metacestode development of the parasite and metabolic aspects, we analyzed the transcriptomes from 4-week metacestodes (4Wmet) and 16-week metacestodes (16Wmet) from experimentally infected DBA/2 mice.

2. Materials and methods

2.1. Ethics statement

This study was conducted in accordance with the recommendations for the Guidelines for Animal Experimentation of the Japanese Association for Laboratory Animal Science. In addition, the protocol for the animal experiments was approved by the ethics committee of the Hokkaido Institute of Public Health (permit number: K25-02).

2.2. Preparation of parasite samples

E. multilocularis isolated in Hokkaido (Nemuro strain) was routinely maintained through a dog–cotton rat life cycle at the Hokkaido Institute of Public Health (Sapporo, Japan). Dogs were orally administered with 5×10^5 *E. multilocularis* protoscoleces and the infection was terminated 35–77 days postinfection by administering two tablets of Droncit® (Kouguchi et al., 2016). Feces were collected from experimentally infected dogs at 35 days post-infection day of administration. Eggs were isolated from feces by filtering through mesh, with natural sedimentation and flotation with sugar solution. Then, mice were anesthetized and challenged orally with 0.5 ml *E. multilocularis* eggs (400 eggs/ml in physiological saline). Finally, DBA/2 mice were sacrificed after 4 and 16 weeks post-infections. Lesions with metacestodes were collected from the livers of three mice. Larvae collected after 4-weeks (4Wmet, Fig. 1A) and 16-weeks (16Wmet, Fig. 1B) were examined. The 4Wmet larvae contained the small vesicles. In 16Wmet larvae, many mature and immature protoscolexes, brood capsules, calcareous corpuscules and cells were contained in the cysts.

2.3. Construction of cDNA libraries

From metacestodes collected 4 and 16 weeks infection by oral administration of eggs, total RNA was extract using Trizol Reagent (Life Technologies). mRNA was prepared using Illumina mRNA Seq Sample Preparation Kit according to manufacturer instructions and RNA-seq libraries were prepared using a custom high throughput method (Kanematsu et al., 2014). Two RNA-seq libraries (4- and 16-weeks) were sequenced at Hokkaido System Science Co., Ltd. on the HiSeq 2000 platform (Illumina) for single-end sequences, trimmed using the Illumina Adaptor Sequence and the output reads with lengths of 50 bp or more were put in the cutadapt (version 1.1) program to simultaneously filter low quality reads and reads with “N”.

2.4. Read sequencing and quantification

Clean RNA-Seq reads larger than 50 were mapped to *E. multilocularis* genome retrieved from the WormBase ParaSite (http://parasite.wormbase.org/Echinococcus_multilocularis_prjeb122/Info/Index/, ParaSite version 5 (January 2016)) using TopHat with default parameters (Trapnell et al., 2009). The previously published reads of 4-weeks metacestodes *in vivo* by our group (Huang et al., 2016) were also mapped to the updated genome of the parasites using this method to obtain replication data for more reliable estimation of expression genes that are significant different between the two stages as described for edgeR (Chen et al., 2014). Then, the mapped read number for each gene was counted by htseq-count (Anders et al., 2015) which was integrated with Galaxy software (https://usegalaxy.org/) using the default parameters, and then transformed to counts per million (cpm) and reads per kilobase per million (rpkm).

2.5. In silico excretory-secretory (ES) and transmembrane (TM) proteins analysis

The translated transcripts of *E. multilocularis* was retrieved from WormBase ParaSite (http://parasite.wormbase.org/Echinococcus_multilocularis_prjeb122/Info/Index/, ParaSite version 5 (January 2016)). *In silico* prediction of ES proteins was carried out using the following programs: SignalP (version 4.1) (Petersen et al., 2011), Phobius (Käll et al., 2007), WegoLoc (weighted gene ontology term based subcellular locallization prediction) (Dataset, Kingdom: BaCelLo dataset, Animals and E-value cutoff: 1E-5) (Chi and Dougu, 2012) and

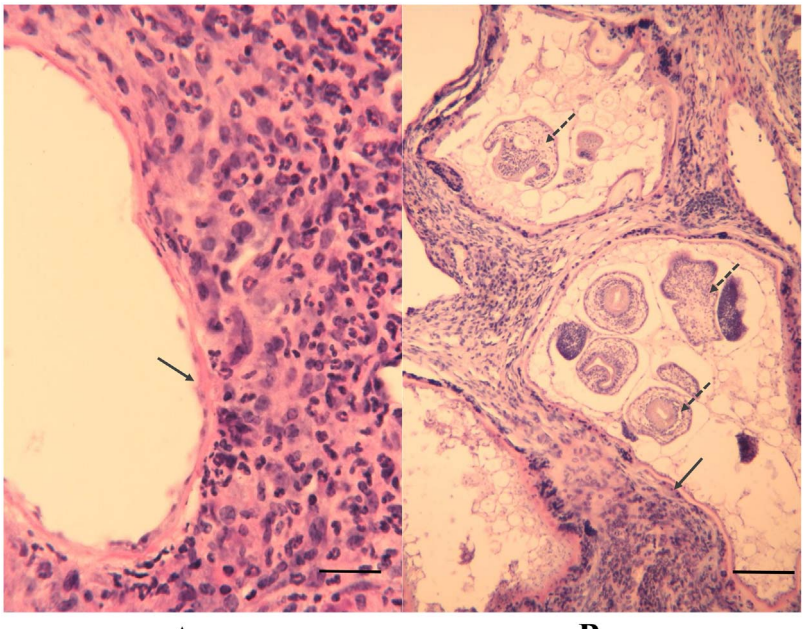

Fig. 1. Morphology at different life cycle stages of *E. multilocularis*. A: 4-week immature metacestodes; B: 16-week mature metacestodes; Bar:25 μm (A), 100 μm (B); Arrowhead: germinal (nucleated) inner layer; Dashed line arrowhead: protoscolex.

Table 1
Summary of reads identified in 4Wmet and 16Wmet libraries.

Sample Name	#Processed reads	#Mapped reads	#Unmapped reads	#Alignment quality < -10 reads
4Wmet(Single-end)	66,440,573	926,796	66,347,894	159,329
4Wmet(Pair-end)	132,558,666	595,927	131,962,739	476,287
16Wmet(Single-end)	69,707,919	6,276,152	63,431,767	5,221,960

Table 2
Different expressed ES proteins in 4Wmet and16Wmet libraries.

Transcript ID	Description	logFC (4Wmet/16Wmet)	logCPM	FRD
EmuJ_000140000.1	collagen alphaiv chain	−5.07	7.75	5.54E-07
EmuJ_001120900.1	diagnostic antigen gp50	5.79	7.16	5.54E-07
EmuJ_000681200.1	diagnostic antigen gp50	6.27	8.29	9.71E-07
EmuJ_000408200.1	expressed protein	9.06	7.66	9.71E-07
EmuJ_000049700.1	diagnostic antigen gp50	6.90	6.52	9.71E-07
EmuJ_000088000.1	expressed conserved protein	−6.18	6.47	9.71E-07
EmuJ_000604800.1	conserved hypothetical protein	4.80	6.28	5.47E-06
EmuJ_000381100.1	EmAgB2 (Taeniid Antigen)	−3.95	10.32	1.09E-05
EmuJ_000412500.1	expressed conserved protein	−5.56	8.19	1.09E-05
EmuJ_000849600.1	proteinase inhibitor I25 cystatin	−3.57	6.69	3.51E-05
EmuJ_000290500.1	glioma pathogenesis protein 1	5.11	10.00	9.53E-05
EmuJ_000342900.1	EG19 antigen	−5.47	8.86	1.08E-04
EmuJ_001137100.1	expressed protein	−5.33	5.79	3.48E-04
EmuJ_000524200.1	collagen type XI alpha 2	−5.27	5.74	5.74E-04
EmuJ_001120700.1	diagnostic antigen gp50	4.18	5.76	5.95E-04
EmuJ_000412600.1	expressed conserved protein	−4.37	5.82	7.03E-04
EmuJ_001051900.1	expressed conserved protein	−4.04	6.08	7.29E-04
EmuJ_000032300.1	diagnostic antigen gp50	8.63	7.64	1.04E-03
EmuJ_000638000.1	netrin 1	−4.05	5.51	1.84E-03
EmuJ_000596300.1	expressed conserved protein	−2.83	7.67	3.01E-03
EmuJ_000199900.1	RPE spondin	−3.71	5.83	5.34E-03
EmuJ_000295100.1	diagnostic antigen gp50	7.59	6.13	8.08E-03
EmuJ_000799300.1	insulin growth factor binding	−3.15	6.35	1.46E-02
EmuJ_000512300.1	diagnostic antigen gp50	7.07	4.78	1.50E-02
EmuJ_000701800.1	basement membrane specific heparan sulfate	−2.43	7.82	1.59E-02
EmuJ_000139900.1	abnormal EMB roylocus tagsis family member emb 9	−2.24	8.09	3.36E-02

Note: logFC: $\log_2$-fold change of expression between 4Wmet and 16Wmet; logCPM: the average $\log_2$-counts-per-millions of total samples; FDR: False Discovery Rate.

TargetP (organism: non-plant and cutoffs: specific > 95%) (Emanuelsson et al., 2007) for secretory signal peptide and subcellular location prediction. TMHMM (version 2.0) (Krogh et al., 2001) and Phobius (Käll et al., 2007) was used to identify proteins having transmembrane domains. The proteins that are predicted to be secreted or that contain a secretory signal peptide by these four tools are considered to be ES proteins excluding proteins that contain transmembrane domains by TMHMM and Phobius. Proteins that are predicted to have transmembrane domain by both TMHMM and Phobius are considered to be TM proteins, except for the predicted mitochondrial proteins predicted by WEGOLOC or TargetP.

2.6. Functional annotations

For functional annotation of the reference transcriptome, InterPro function domain annotations were directly extracted from the parasite GFF3 file retrieved from the WormBase ParaSite. Gene Ontology term annotation was performed by BlAST2GO (Götz et al., 2008). The BlastKOALA (Kanehisa et al., 2016) was used to map proteins to KEGG pathways for assigning orthologues, and a representative gene data set was selected for eukaryotes. The putative homologues proteases of *E. multilocularis* were identified using the complete set of core protease sequences from the MEROPS (release 11.0) database (Rawlings et al., 2016) using batch BLAST (E-value cutoff: 1E-5).

2.7. Real-time PCR validation

To validate the next-generation sequencing (NGS) data, six antigen candidate were selected for real-time PCR analysis. The primers used for amplification of six antigen candidate genes of *E. multilocularis* and the glyceraldehyde 3-phosphate dehydrogenase gene (EmuJ_000254600, internal control) were designed by OligoArchitect™ (http://www.oligoarchitect.com) as shown in Table S1. Real-time PCR was performed using the Applied Biosystems 7300 Real-time PCR System with SYBR-Green detection (SYBR Premix, TaKaRa) according to manufacturer' instructions. Each sample was assayed in triplicate, the average threshold cycle (Ct) was calculated per sample and relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

3. Result

3.1. Reads sequencing and quantification

Illumina sequencing of cDNA libraries yielded 66,440,573 (4Wmet, single-end), 132,558,666 (4Wmet, pair-end) and 69,707,919 (16Wmet, single-end) clean reads (length > 50 bp), characterized as shown in Table 1. For differentially expressed (DE) gene analysis, there were 168 DE genes between 4- and 16-week metacestodes (Table S2). And 26 and 59 genes are coding DE (FDR < 0.05) proteins ES (Table 2) and TM (Table 3), respectively.

3.2. Real-time PCR validation

To validate the expression profiles, antigen B subunits, EM95 and apomucins, which are candidate proteins for diagnostic antigens or vaccine against *E. multilocularis* were selected for real-time PCR analysis using the same RNA samples as for single-end sequencing (Fig. 2). The

Table 3
Different expressed TM proteins in 4Wmet and 16Wmet libraries.

Transcript ID	Description	logFC	logCPM	FDR
EmuJ_000681900.1	expressed protein	−5.62	6.88	5.54E-07
EmuJ_000140100.1	Collagen alpha 1V chain	−4.45	7.94	9.71E-07
EmuJ_000355000.1	tetraspanin	6.85	6.47	9.71E-07
EmuJ_000368620.1	EG95	9.35	6.53	9.71E-07
EmuJ_000823800.1	collagen alpha 2I chain	−4.83	6.70	1.13E-06
EmuJ_002194800.1	Leucine-rich repeat protein 1 (LRRP1)	6.13	8.42	1.23E-06
EmuJ_000354900.1	tetraspanin	8.34	6.29	2.27E-06
EmuJ_000276200.1	hypothetical transcript	4.50	7.29	3.37E-06
EmuJ_000312200.1	expressed protein	6.38	12.93	7.10E-06
EmuJ_000743100.1	hypothetical transcript	6.92	7.13	9.81E-06
EmuJ_000938200.1	hypothetical transcript	6.08	10.86	6.99E-05
EmuJ_000212300.1	tyrosine protein kinase otk	−3.45	8.18	1.22E-04
EmuJ_001008200.1	calcium transporting atpase	−2.96	6.82	3.85E-04
EmuJ_001190600.1	collagen alpha v chain	−3.41	6.38	3.85E-04
EmuJ_000642000.1	sn1 specific diacylglycerol lipase beta	5.20	5.45	6.21E-04
EmuJ_001105600.1	hypothetical transcript	5.42	11.20	6.70E-04
EmuJ_001115400.1	conserved hypothetical protein	3.58	7.60	8.03E-04
EmuJ_000800500.1	activin receptor type	−3.81	5.79	1.08E-03
EmuJ_000606100.1	oxalate:formate antiporter	−3.23	7.54	1.13E-03
EmuJ_000072600.1	conserved hypothetical protein	6.90	5.21	3.01E-03
EmuJ_000903100.1	calsyntenin 1	−2.77	7.70	3.11E-03
EmuJ_000192600.1	PQ loop repeat containing protein 2	3.77	11.66	3.15E-03
EmuJ_000618700.1	Glutamate receptor ionotropic kainate 2	−3.52	6.37	3.15E-03
EmuJ_000971100.1	sodium dependent phosphate transporter 1	−3.40	5.86	3.72E-03
EmuJ_000355900.1	tetraspanin	3.66	11.84	4.28E-03
EmuJ_000054300.1	cj cadherin	−3.53	5.57	5.02E-03
EmuJ_000991500.1	IQ calmodulin binding region	−4.01	5.54	7.01E-03
EmuJ_000804500.1	conserved hypothetical protein	−2.69	6.83	7.03E-03
EmuJ_000009500.1	solute carrier family 13	2.62	6.53	8.60E-03
EmuJ_000318600.1	Netrin receptor DCC	−3.94	5.49	9.48E-03
EmuJ_000355500.1	tetraspanin	3.08	8.38	1.05E-02
EmuJ_000583700.1	protocadherin 11	−3.30	5.84	1.06E-02
EmuJ_000731600.1	conserved hypothetical protein	2.34	6.83	1.18E-02
EmuJ_000641500.1	tyrosine protein kinase	−3.06	5.80	1.32E-02
EmuJ_000715500.1	protocadherin 9	−2.56	6.22	1.39E-02
EmuJ_001058200.1	expressed protein	3.99	5.15	1.39E-02
EmuJ_000656900.1	expressed conserved protein	−2.34	6.58	1.47E-02
EmuJ_000570400.1	protocadherin 1	−2.52	6.84	1.65E-02
EmuJ_000549500.1	discoidin domain containing receptor 2	−3.64	5.19	1.81E-02
EmuJ_000175300.1	GPI inositol deacylase	−4.59	5.22	1.98E-02
EmuJ_000744300.1	protein jagged 2	−4.60	5.23	1.98E-02
EmuJ_000634800.1	L lactate dehydrogenase B chain	−2.79	6.19	2.10E-02
EmuJ_001168100.1	neutral amino acid transporter A	5.46	4.91	2.24E-02
EmuJ_001080600.1	zinc transporter ZIP13	−3.58	5.14	2.66E-02
EmuJ_000826850.1	conserved hypothetical protein	−4.33	4.94	2.77E-02
EmuJ_000504200.1	expressed conserved protein	3.08	6.51	2.93E-02
EmuJ_001077400.1	tetraspanin (TSP3)	6.17	5.27	3.36E-02
EmuJ_000609000.1	equilibrative nucleoside transporter 3	−1.82	7.27	3.49E-02
EmuJ_000738800.1	expressed conserved protein	−2.25	6.89	3.68E-02
EmuJ_000071900.1	expressed protein	2.98	9.26	3.79E-02
EmuJ_000168700.1	synaptogyrin 2	−3.52	5.10	3.79E-02
EmuJ_001133800.1	receptor type tyrosine protein phosphatase	−3.52	5.77	3.82E-02
EmuJ_000544400.1	ryanodine receptor 44f	−2.35	6.72	4.12E-02
EmuJ_000758200.1	expressed protein	−4.15	4.80	4.17E-02
EmuJ_000661800.1	oxalate:formate antiporter	−2.24	9.13	4.36E-02
EmuJ_000986400.1	receptor type guanylyl cyclase	−2.94	5.45	4.38E-02
EmuJ_000443400.1	protein patched	−2.37	5.95	4.68E-02
EmuJ_000816800.1	fibroblast growth factor receptor 1	−2.87	5.46	4.72E-02
EmuJ_000758400.1	TGF beta receptor type 1	−2.26	6.17	4.90E-02

Note: logFC: $\log_2$-fold change of expression between 4Wmet and 16Wmet; logCPM: the average $\log_2$-counts-per-millions of total samples; FDR: False Discovery Rate.

real-time PCR results confirmed the result obtained from deep sequencing analysis and showed similar trends of up- or down-regulated genes.

3.3. Predicted *E. multilocularis* secretome and transmembranome size

Of the 10,669 putative *E. multilocularis* translated protein sequences retrieved from WormBase ParaSite, 775 sequences (7.26%) were predicted to include a signal peptide cleavage site by both SignalP4.1 and Phobius, but only 356 sequences (3.34%) are consider to be ES proteins based on subcellular location analysis by WEGOLOC and TargetP. A total of 1972 sequences (18.48%) were predicted to contain transmembrane helices by both TMHMM and Phobius, but to consider only proteins in the metacestode that have direct contact with host, 198 potential mitochondrial proteins predicted by WEGOLOC or TargetP were excluded, leaving 1774 *E. multilocularis* TM protein sequences in the datasets.

3.4. Functional annotation of *E. multilocularis* transcriptome

Functional annotation of the *E. multilocularis* transcriptome based on Gene Ontology (GO) terms, KEGG pathway, InterPro and MEROPS batch BLAST annotation. Of the 10,669 putative *E. multilocularis* translated protein sequences retrieved from the genome, 6515 (61.06%) gene coding proteins were retrieved as GO terms of the transcriptome. Of the 356 and 1774 *E. multilocularis* ES and TM protein, 109 (30.62%) and 1695, respectively, (95.54%) were assigned to GO terms (Table S2). The predominant GO terms of ES proteins at level 3 for 'molecular function' were 'hydrolase activity', 'enzyme regulator activity' and 'ion binding' (Fig. 3, Table S2). For TM proteins, the predominant GO terns at level 3 for 'molecular function' were 'transmembrane transporter activity', 'substrate-specific transporter activity', 'ion binding' and 'transferase activity'(Fig. 4, Table S2). Furthermore, enrichment analysis shows that a significant increase was observed in GO-terms associated with 'extracellular matrix structural constituent' (FDR = 1.1E-8) and 'calcium ion binding' (FDR = 6.1E-8) of different expression (DE) genes up-regulated in 16Wmet (Fig. 5). KEGG pathway analysis shows that all DE genes are up-regulated in the focal adhesion pathway as immature metacestodes transform to mature metacestes (Fig. S1; Table S2). InterPro annotation of predicted ES protein sequences generated 189 different assigned protein domains and families (Table S2), and the most represented domains showing high ratios of ES proteins between secretome and transcriptome are 'Pancreatic trypsin inhibitor Kunitz domain', 'Proteinase inhibitor I2', 'CAP domian' and 'Taeniidae antigen' (Table 2, Table S2). Interestingly, EmuJ_001137100 which assigned as ES proteins and contains pancreatic trypsin inhibitor kunitz domain and proteinase inhibitor I2 domain is significant highly expressed (FDR = 3.48E-4) at 16Wmet (Table S2). As for TM proteins, 'GPCR, rhodopsin-like', 'Ion transport domain', 'Tetraspanin/Peripherin' and 'Cadherin' are the most frequently occurring domains and show high ratios between transmembranome and transcriptome (Table S2). Furthermore, there are five tetraspanin domains containing proteins (EmuJ_000354900.1, EmuJ_000355000.1, EmuJ_000355500.1, EmuJ_000355900.1, EmuJ_001077400.1), which show significantly highly expression (FDR < 0.05) in 4Wmet when compared to 16Wmet (Table 3, Table S2). For proteinase analysis, there are 343 gene coding proteins that

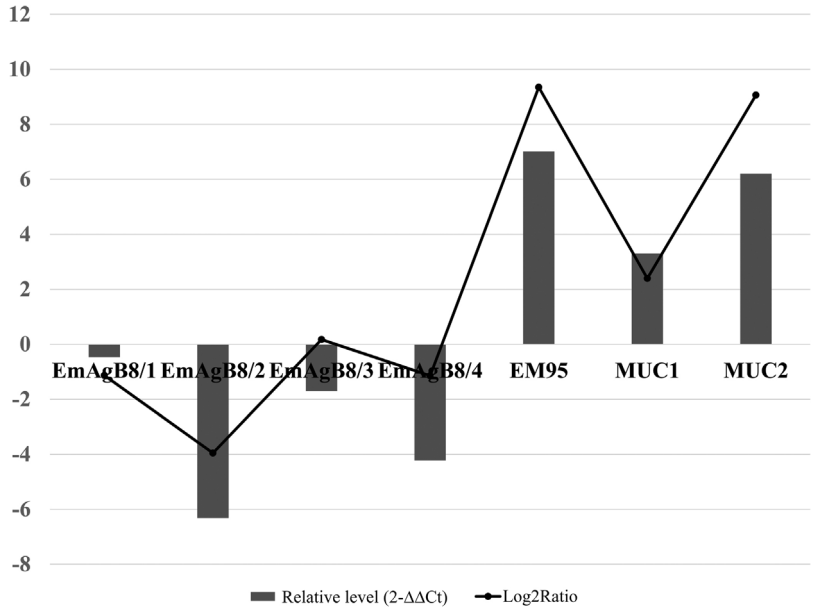

Fig. 2. Real-time PCR validation of genes at expression levels. The y-axis indicates the value of the relative expression level ($2^{-\Delta\Delta Ct}$) by real-time PCR and $\log_2$Ratio of 4Wmet/16Wmet by next-generation sequencing. GAPDH (EmuJ_000254600.1) as the internal control. Transcript IDs for each predicted protein are as follows: EmAgB8/1 (EmuJ_000381200.1), EmAgB8/2 (EmuJ_000381100.1), EmAgB8/3 (EmuJ_000381500.1), EmAgB8/4 (EmuJ_000381400.1), EM95 (EmuJ_000368620.1), MUC1 (EmuJ_000742900.1), and MUC2 (EmuJ_000408200.1).

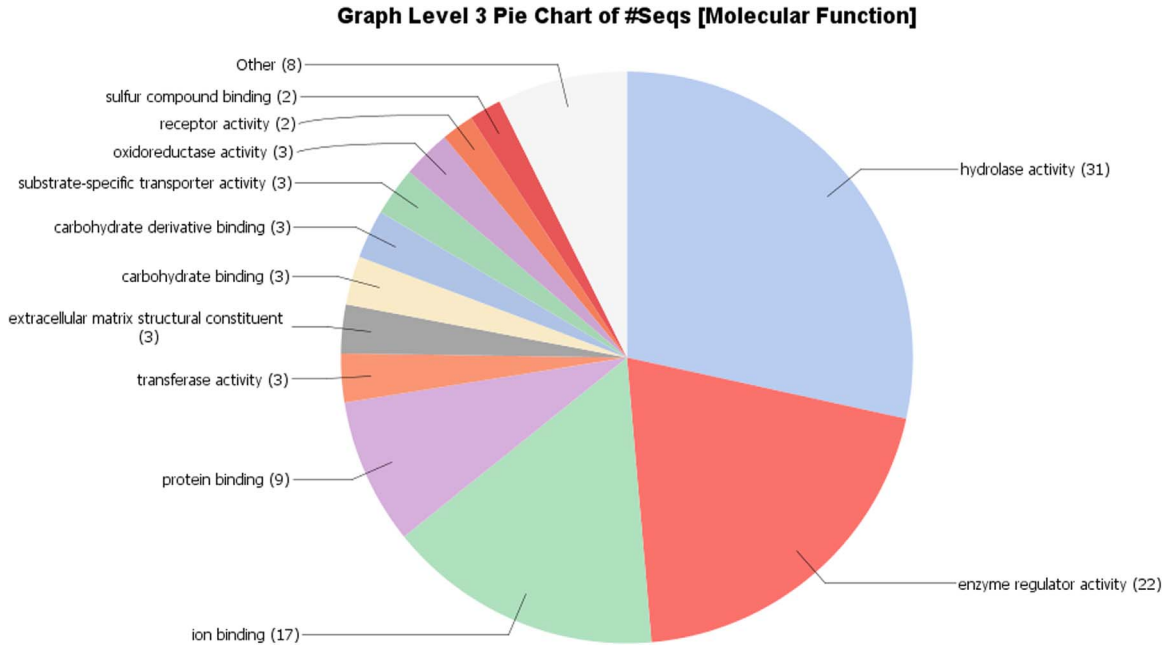

Fig. 3. Molecular function ontology distribution of third-level subcategory *E. multilocularis* predicted ES proteins in reference set.

contain the proteinase and/or peptidase domains at the e-value cutoff: 1E-5 and 5 out of 6 DE genes are up-regulated when immature metacestodes transform to mature metacestodes (Table S2).

4. Discussion

This study describes the genome of *E. multilocularis* and a bio-informational technical approach as a low cost way for predicting ES and TM proteins of *E. multilocularis*. In the present study, a low ratio of mapped reads was encountered due to contamination with host tissue that was difficult to separate from metacestodes in the host liver, especially for the 4-weeks metacestodes, but validation results of antigen B subunits, EM95 and apomucins, which are candidates for diagnostic antigen or vaccine targets, demonstrated similar trends with NGS approach.

Real-time PCR results of antigen B subunit 2 confirmed previous findings (Huang et al., 2016) that it is mostly sensitive to different cyst stages than to other antigen B subunits. For predicted ES proteins, gene EmuJ_000408200.1 which is described as an 'expressed protein' in WormBase annotation is the only MUC-2 sub-family member that was expressed in all sequenced stages (non-activated oncospheres, activated oncospheres, metacestodes and adult) of the parasite (Huang et al., 2016; Díaz et al., 2015; Tsai et al., 2013) and are significant highly expressed (FDR = 9.7E-07) in 4Wmet, in the present study. In addition, antigen EM95 which can produce significant levels of protection against challenge infection with *E. multilocularis* eggs in mice (Gauci et al., 2002) maintains expression in oncospheres (Huang et al., 2016) and immature metacestodes (4Wmet, Table S2) but almost no expression is

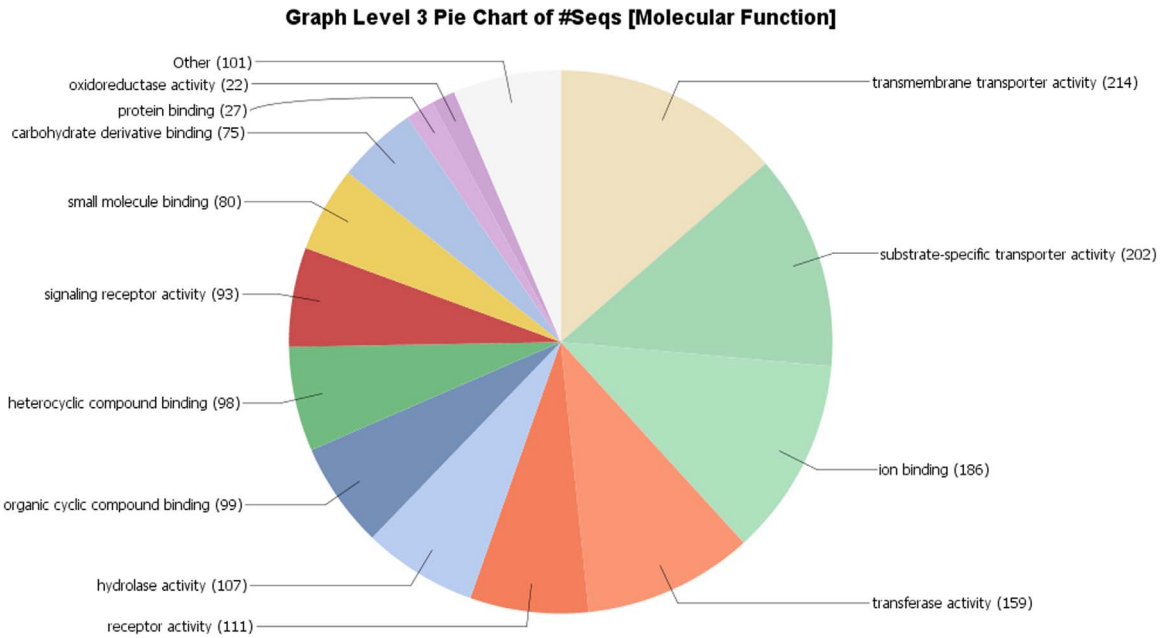

Fig. 4. Molecular function ontology distribution of third-level subcategory *E. multilocularis* predicted TM proteins in reference set.

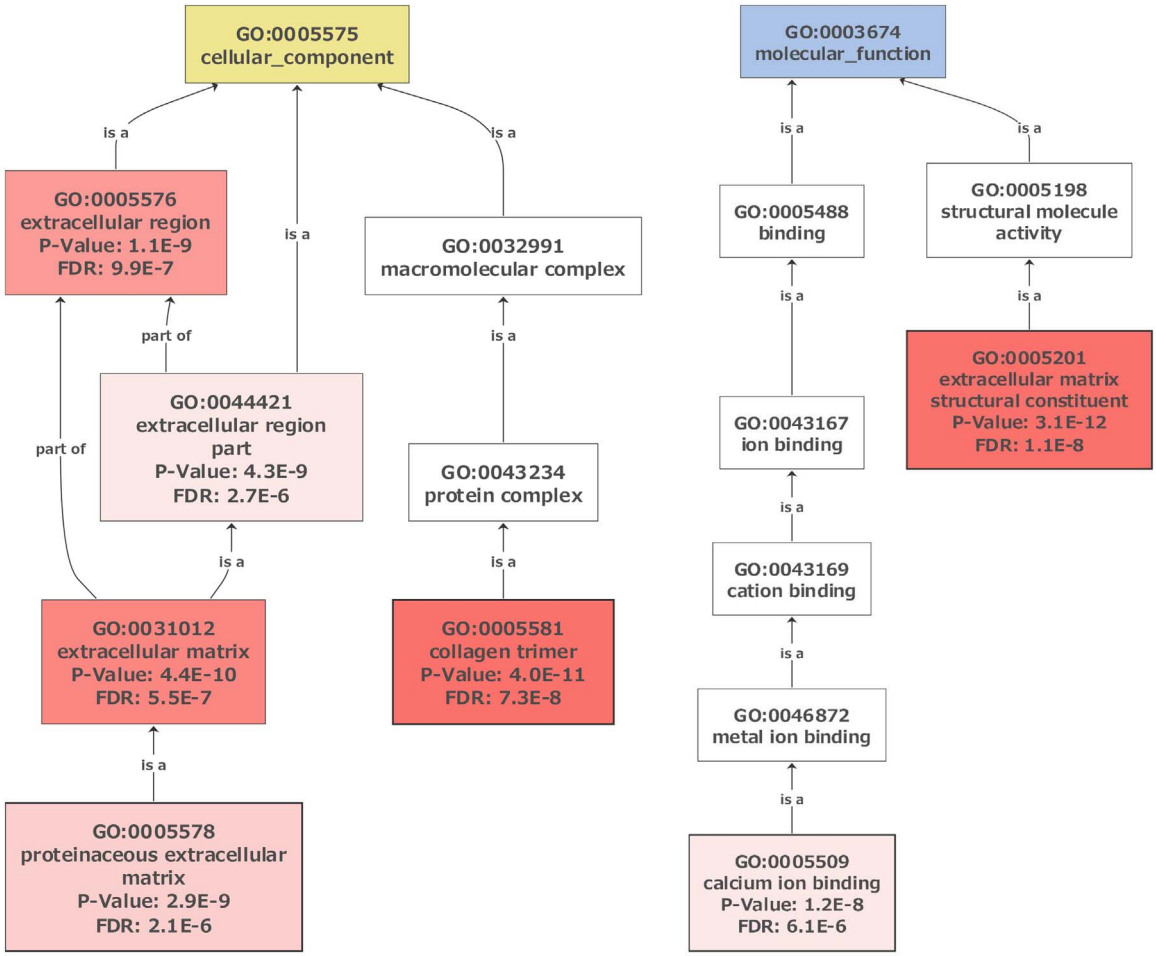

Fig. 5. GO enrichment of highly expressed genes in 16Wmet. Significant highly expression genes (FDR < 0.05) in 16Wmet as test set and whole transcriptome in reference set.

observed in mature metacestodes (16Wmet, Table S2) and adult parasites (Tsai et al., 2013). In contrast, the EG19 antigen, which was first detected in the *E. granulosus* hydatid fluid (Virginio et al., 2012) was highly expressed in both 4Wmet and 16Wmet (Table 2), but is a DE gene (FDR = 1.08E-4) between the two metacestodes stages. It was suggested that EG19 antigen may be a biomarker for cystic echinococcosis, since it is being more sensitive in the active disease development form (CE1–CE2) than during the inactive disease form (CE4–CE5)

(Virginio et al., 2012). Dynamic expression levels of this antigen in 4Wmet and 16Wmet also suggest that it can be a biomarker for AE. As for predicted TM proteins, it was shown that EmuJ_000312200.1, which had high Thr content (228/427) that offers multiple potential *O*-glycosylation sites had highly expression in both 4Wmet and 16Wmet but was significantly high (FDR = 7.10E-06) in 4Wmet. And it is proved EmA9 antigen, one of a coproantigens for *E. multilocularis*, is a surface glycoprotein with unique O-gycosylation (Hülsmeier et al., 2010). The high expression of the putative glycoprotein (EmuJ_000312200.1) in both 4Wmet and 16Wmet suggests that it can be a good candidate for the diagnosis of AE and may play a role in the localization and protection of the parasite during early infection and chronic stages.

The predicted secretomes of *E. multilocularis* that have high proportion of peptidases and peptidase regulating proteins could be potential antigens for vaccines and diagnostic assay (Wang et al., 2015). In the present study, genes that contain proteinase and/or peptidase domains of ES proteins have relatively higher expression levels in mature metacestodes. It has been suggested that mature metacestode are less susceptible to host effects due to the protective laminated layer (Siracusano et al., 2012) suggesting that proteinase and/or peptidase may play a more important role in regulating host immune response during the chronic stage of echinococcosis, as examined in the present study.

Transmembrane (TM) proteins are usually involved in many important biological processes such as cell signaling, transport of membrane-impermeable molecules and cell recognition (Nugent and Jones, 2009). Previous studies (Dang et al., 2009, 2012) have demonstrated that members of the tetraspanin family in *E. multilocularis* are potential vaccines for AE, and they may cross-protect (Huang et al., 2016). In the present study, 41 amino acid sequences containing the tetraspanin domain (Table S2) were found, and 5 gene coding tetraspanins had significantly different expression and down-regulated in 16Wmet. The highly expressed tetraspanins in 4Wmet are potential candidates for vaccine to provide protection in the early infection phase.

5. Conclusion

In this study, we sequenced the transcriptomes of *E. multilocularis* strain Nemuro from 4- and 16- week of metacestodes from DBA/2 mice using NGS technology for the first time, which provided some insights in to the key genes in the maturation of metacestodes. In addition, the expression levels of predicted ES and TM proteins in *E. multilocularis* may provide some novel candidates, especially highly expressed genes in immature or mature metacestodes, for development into diagnostics and vaccinations against the parasite.

Conflict of interest statement

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science (Grant No. 25450425) and The project of Echinococcosis Prevention and Contro-Ganzi Workstation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vetpar.2017.05.006.

References

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Chen, Y., McCarthy, D., Robinson, M., Smyth, G.K., 2014. edgeR: Differential Expression Analysis of Digital Gene Expression Data User's Guide. Available online at: https://www.genomatix.de/online_help/help_regionminer/edgeR.pdf.

Chi, S.M., Dougu, N., 2012. WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms. Bioinformatics 28, 1028–1030.

Dang, Z., Yagi, K., Oku, Y., Kouguchi, H., Kajino, K., Watanabe, J., Sugimoto, C., 2009. Evaluation of *Echinococcus multilocularis* tetraspanins as vaccine candidates against primary alveolar echinococcosis. Vaccine 27, 7339–7345.

Dang, Z., Yagi, K., Oku, Y., Kouguchi, H., Kajino, K., Watanabe, J., Sugimoto, C., 2012. A pilot study on developing mucosal vaccine against alveolar echinococcosis (AE) using recombinant tetraspanin 3: vaccine efficacy and immunology. PLoS Negl. Trop. Dis. 6, e1570.

Díaz, Á., Fernández, C., Pittini, Á., Seoane, P.I., Allen, J.E., Casaravilla, C., 2015. The laminated layer: recent advances and insights into *Echinococcus* biology and evolution. Exp. Parasitol. 158, 23–30.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2, 953–971.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435.

Gauci, C., Merli, M., Muller, V., Chow, C., Yagi, K., Mackenstedt, U., Lightowlers, M.W., 2002. Molecular cloning of a vaccine antigen against infection with the larval stage of *Echinococcus multilocularis*. Infect. Immun. 70, 3969–3972.

Huang, F., Dang, Z., Suzuki, Y., Horiuchi, T., Yagi, K., Kouguchi, H., Oku, Y., 2016. Analysis on gene expression profile in oncospheres and early stage metacestodes from *Echinococcus multilocularis*. PLoS Negl. Trop. Dis. 10, e0004634.

Hülsmeier, A.J., Deplazes, P., Naem, S., Nonaka, N., Hennet, T., Köhler, P., 2010. An *Echinococcus multilocularis* coproantigen is a surface glycoprotein with unique O-gycosylation. Glycobiology 20, 127–135.

Käll, L., Krogh, A., Sonnhammer, E.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res. 35, W429–W432.

Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.

Kanematsu, S., Tanimoto, K., Suzuki, Y., Sugano, S., 2014. Screening for possible miRNA–mRNA associations in a colon cancer cell line. Gene 533, 520–531.

Kouguchi, H., Irie, T., Matsumoto, J., Nakao, R., Sugano, Y., Oku, Y., Yagi, K., 2016. The timing of worm exclusion in dogs repeatedly infected with the cestode *Echinococcus multilocularis*. J. Helminthol. 90, 766–772.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle}$CT method. Methods 25, 402–408.

Matsumoto, J., Kouguchi, H., Oku, Y., Yagi, K., 2010. Primary alveolar echinococcosis: course of larval development and antibody responses in intermediate host rodents with different genetic backgrounds after oral infection with eggs of *Echinococcus multilocularis*. Parasitol. Int. 59, 435–444.

Nugent, T., Jones, D.T., 2009. Transmembrane protein topology prediction using support vector machines. BMC Bioinf. 10, 159.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Rawlings, N.D., Barrett, A.J., Finn, R.D., 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 44, D343–D350.

Siracusano, A., Delunardo, F., Teggi, A., Ortona, E., 2012. Host-parasite relationship in cystic echinococcosis: an evolving story. Clin. Dev. Immunol. 2012, 639362.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Aslett, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496, 57–63.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Virginio, V.G., Monteiro, K.M., Drumond, F., de Carvalho, M.O., Vargas, D.M., Zaha, A., Ferreira, H.B., 2012. Excretory/secretory products from in vitro-cultured *Echinococcus granulosus* protoscoleces. Mol. Biochem. Parasitol. 183, 15–22.

Wang, S., Wei, W., Cai, X., 2015. Genome-wide analysis of excretory/secretory proteins in *Echinococcus multilocularis*: insights into functional characteristics of the tapeworm secretome. Parasites Vectors 8, 666.

Zhang, W., Ross, A.G., McManus, D.P., 2008. Mechanisms of immunity in hydatid disease: implications for vaccine development. J. Immunol. 181, 6679–6685.

Zhang, W., Li, J., Lin, R., Wen, H., McManus, D.P., 2012. Recent Advances in the Immunology and Serological Diagnosis of Echinococcosis, Serological Diagnosis of Certain Human, Animal and Plant Diseases. In: Al-Moslih, M. (Ed.), InTech, . ISBN: 978-953-51-0370-7, Available from: http://www.intechopen.com/books/serological-diagnosis-of-certain-human-animal-and-plant-diseases/recentadvances-in-the-serological-diagnosis-of-echinococcosis.