

**THE DEFINITION OF *ECHINOCOCCUS*  
*MULTILOCULARIS* DIFFERENTIALLY EXPRESSED  
MOLECULES USING DEEP SEQUENCING**

**YADONG ZHENG, MSc.**

**Thesis submitted to the University of Nottingham**

**for the degree of Doctor of Philosophy**

**May 2012**

## Abstract

*Echinococcus multilocularis* has recently been developed as a model for basic and applied studies on trematodes and cestodes. Deep understanding of molecular biology of this parasite is urgently required for efficient and extensive applications such as functional gene probing and anti-helminth drug screening. In our project, we aimed to unveil *E. multilocularis* miRNA repertoires and the identification of differentially expressed molecules from the mRNA and miRNA transcriptomes using next-generation sequencing technology. Furthermore, one family of *E. multilocularis* molecules identified, fatty acid-binding proteins that are involved in lipid metabolisms, was phylogenetically and functionally characterized.

Our data presented genome-widely developmental expression profiles of *E. multilocularis* miRNAs. Sixty-five potential miRNAs were predicted with nineteen of them being novel. The majority of differentially expressed miRNAs such as emu-miR-1, 16 and 71 were found to be significantly down-regulated in the primary neoblast-rich cells derived from the vesicles. Comparative analysis revealed that the miR-71/2 cluster was conserved in the platyhelminths. In addition, seven homologues to miRNAs likely linked to planarian neoblasts were present in *E. multilocularis* with the intact seed regions. Although their expression in *E. multilocularis* neoblasts remains to be fully certified, these miRNAs are plausible biomarkers for the germline stem cells.

For mRNA profiling, we firstly originated an enzymatic approach for the elimination of mitochondrial transcripts to increase the depth of mRNA

transcriptomes. In comparison with 2.7% in the control, the frequency of the sequences mapped to the mitochondria was dramatically reduced to 0.04~0.1% in the treated samples without significantly adverse effects on the targets. A number of differentially expressed genes were inferred from *E. multilocularis* mRNA transcriptomes and FABP-encoding genes were further phylogenetically and functionally characterised. *E. multilocularis* encoded at least four distinct FABPs and comparative analysis revealed that the current gene set of FABPs may have emerged before speciation of *E. multilocularis* and *E. granulosus*. The apparent divergence between Echinococcus and vertebrate FABPs in genomic organizations prompted us to further explore phylogenetically FABP genes across thirty-five invertebrates. The results demonstrated that FABP genes were organized in diverse ways, with a predominant pattern that is commonly present in vertebrates. Moreover, both gene duplication and alternative splicing might be most likely responsible for variety of invertebrate FABP functions.

Four *E. multilocularis* FABPs were developmentally regulated and FABP3 was highly expressed in the vesicles and secreted or released into the hydatid fluid. These proteins appeared to be able to weakly interact with *cis*-parinaric acid but not a fluorescent fatty acid DAUDA and retinol. *In vitro* analysis indicated that FABP3 was capable of the induction of cytokine secretion by bone marrow-derived macrophages and dendritic cells probably via Toll-like receptor 2. Additionally, FABP3 was also demonstrated to activate macrophage small reactive molecule production together with a ligand for TLR 2.

## **Acknowledgements**

I would first give my thanks to my supervisors Jan Bradley and Ed Louis for their supports in life and studies, and it is absolutely impossible for me to finish the studies without their guidance and encouragement. I am indebted to Matthew Berriman for giving me chance of one-year visiting work in his lab and other people in Wellcome Trust Sanger Institute, Cambridge, especially Anna Protasio and Minjie Ding, for technical help, and to Klaus Brehm from University of Würzburg, Germany, for his kind guidance and providing of material, and to Bruno Gottstein from University of Bern, Switzerland, for generous providing of material. I would like to give my gratitude to everyone in our research group in the university and to Angus Davison, Wenceslaus Luoga, Fadlul Mansur and Himanshu Kharkwal for their help in experiments and life, and to others not mentioned here for their help.

I extend my thanks to Xuepeng Cai, Hong Yin, Jianxun Luo, Xuenong Luo and other colleagues in Lanzhou Veterinary Research Institute, China, for their support.

My thanks are especially given to my wife Rui Xiong for abdication of her loved work to accompany me, her love and meticulous cares, and to my mother, father and sister for love and unwavering support. Their eager anticipation and smiles are main resources of impetus that drives me to surmount obstacles in life and study in the UK.

My study was financially supported by Overseas Research Students Awards Scheme (ORS), UK, The University of Nottingham, UK, and Lanzhou

Veterinary Research Institute, Chinese Academy of Agricultural Sciences  
(CAAS), China.

## **Authorship declaration**

The majority of work in this study was done by Yadong Zheng, with the exceptions of sequencing adaptor ligation and sequencing by Michael A. Quail and short sequence mapping by Jacqueline McQuillan in Chapter 2, and miRNA sequencing by Michael A. Quail and main perl script programming by Minjie Ding in Chapter 3.

# Contents

Abstract.....	I
Acknowledgements.....	III
Authorship declaration.....	V
Contents .....	VI
List of figures.....	X
List of tables .....	XII
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1. <i>Echinococcus multilocularis</i> .....	4
1.1 Life cycle and significance in public health of alveolar echinococcosis caused by <i>E. multilocularis</i> .....	4
1.2 Neoblasts and establishment of an <i>in vitro</i> cultivation system .....	9
1.3 Genomics of <i>E. multilocularis</i> .....	10
2. Transcriptome and RNA-seq .....	12
2.1 Transcriptome: an entire dynamic RNA profile .....	13
2.2 RNA-seq: a deep high-throughput technology for transcriptional characterization .....	14
3. microRNA .....	19
3.1 Biogenesis and genetic characteristics of miRNAs .....	20
3.2 Approaches to the identification of miRNAs .....	24
3.4 Distribution of miRNAs in parasites .....	25
3.5 Possible roles of parasite miRNAs .....	29
4. Fatty acid-binding proteins .....	33
4.1 FABPs in the animal kingdom.....	34
4.2 Gene structure and expression of FABPs .....	35
4.3 Structural and binding features of FABPs .....	37
4.4 Biological roles of FABPs.....	38
5. Summary.....	44
<b>Chapter 2 A robust strategy of depletion of polycistronic mitochondrial transcripts     in <i>Echinococcus multilocularis</i> for RNA-seq .....</b>	<b>46</b>
Abstract.....	46
1. Introduction .....	46
2. Materials and methods.....	48
2.1 Parasites and primary cells .....	48

2.2 Extraction of total RNA .....	48
2.3 Depletion of mitochondrial transcripts and in vitro amplification of mRNA.	49
2.4 Preparation of sequencing libraries.....	51
2.5 Sequence data quality control .....	52
2.6 Mapping and statistical analysis .....	52
3. Results and discussion .....	53
3.1 Rationale of the strategy for removal of <i>E. multilocularis</i> mitochondrial transcripts .....	53
3.2 Validation of efficiency of mitochondrial RNA elimination .....	54
3.3 Mapping reads to the <i>E. multilocularis</i> genome.....	56
3.4 Efficacy of the depletion of the mitochondrial RNA species .....	58
<b>Chapter 3 The genome-wide microRNA landscape of <i>Echinococcus multilocularis</i>.</b>	<b>64</b>
Abstract.....	64
1. Introduction .....	64
2. Material and methods .....	67
2.1 Parasites and cultured cells .....	67
2.2 Recovery of small RNA.....	67
2.3 Preparation of small RNA libraries for sequencing.....	68
2.4 Process of short reads .....	68
2.5 Annotation and comparative analysis of known and novel miRNAs.....	69
3. Results .....	70
3.1 Overview of three short sequence sets of small RNA libraries for vesicles, cell aggregates and activated protoscolecocytes.....	70
3.2 Known and novel miRNAs in <i>E. multilocularis</i> .....	71
3.3 Differential expression of miRNAs.....	74
3.4 miRNAs possibly related to regenerative capacity of neoblasts .....	75
3.5 miR-71b/2 cluster across the Phylum Platyhelminth .....	80
3.6 Comparative analysis with other platyhelminths miRNAs .....	83
4. Discussion.....	87
<b>Chapter 4 Diversity of gene structure and alternative splicing of fatty acid binding protein genes in invertebrates .....</b>	<b>91</b>
Abstract.....	91
1. Introduction .....	91
2. Methods .....	93
2.1 Identification and annotation of FABP genes .....	93

2.2 Sequence alignments and secondary structure prediction .....	94
3. Results .....	94
3.1 Identification and annotation of FABP genes across invertebrates .....	94
3.2 Diversity of FABP gene structures and exons across invertebrates .....	98
3.3 Intron loss in the Phylum Platyhelminth .....	98
3.4 Intronless FABP genes.....	102
3.5 Alternative splicing in FABP genes.....	102
4. Discussion.....	105
<b>Chapter 5 Expression, purification and fatty acid binding characteristics of fatty acid-binding proteins of <i>Echinococcus multilocularis</i>.....</b>	<b>109</b>
Abstract.....	109
1. Introduction .....	109
2. Materials and methods.....	111
2.1 Identification of FABP homologues in <i>E. multilocularis</i> and amplification of FABP genes .....	111
2.2 Construction of prokaryotic expression vector .....	112
2.3 Expression and purification of FABPs in <i>E. coli</i> .....	114
2.4 Digestion of rSUMO/FABPs to release rFABPs .....	115
2.5 Construction of yeast expression vectors.....	116
2.6 Transformation and identification of positive clones harbouring FABP3....	116
2.7 Expression and purification of rFABP3 produced by yeast.....	117
2.8 Delipidation of rFABPs .....	119
2.9 Removal of endotoxin in recombinant FABPs .....	120
2.10 Endotoxin test.....	121
2.11 Fatty acid binding assay.....	121
2.12 Preparation and purification of anti-FABP3 rabbit serum.....	121
2.13 Western blot .....	122
3. Results .....	123
3.1 Identification and gene structures of emFABPs .....	123
3.2 Comparison of emFABPs and <i>E. granulosus</i> FABPs (egFABPs) .....	125
3.3 FABP mRNA levels across life stages.....	128
3.4 Recombinant FABPs produced by <i>E. coli</i> and yeast.....	130
3.5 Fatty acid binding of rSUMO/FABPs and rFABP3 .....	135
3.6 Rabbit anti-rFABP3 antibody and Western blotting.....	138
4. Discussion.....	138

<b>Chapter 6 Modulation of macrophage and dendritic cell functions by <i>E. multilocularis</i> FABP3 in vitro</b> .....	<b>144</b>
Abstract.....	144
1. Introduction .....	144
2. Material and methods .....	146
2.1 TLR cell lines, ligands and mice.....	146
2.2 Stimulation of TLR transformants by rFABP3 and ligands .....	147
2.3 Preparation of bone marrow-derived macrophages and bone marrow-derived dendritic cells .....	147
2.4 Determination of cytokine levels in the supernatant of cell culture after stimulation .....	148
2.5 Assessment of nitric oxide by BMDMs .....	149
2.6 Preparation of BMDM total RNA .....	149
2.7 Qualitative PCR analysis of expression of arginase I by BMDMs.....	150
2.8 Statistical analysis.....	150
3. Results .....	150
3.1 Cytokine profiling of BMDMs and BMDDs treated by rFABP3 .....	150
3.2 Anti-rFABP3 antibody down-regulates the TNF $\alpha$ expression stimulated by rFABP3 .....	154
3.3 rFABP3 can activate Toll-like receptor 2.....	157
3.4 rFABP3 modulates BMDM/BMDD cytokine secretion potentially in a TLR 2-dependant manner.....	160
3.5 rFABP3 stimulates macrophages possibly by bypassing an alternative activation pathway.....	160
3.6 Effects of an adipocyte FABP inhibitor on cytokines induced by rFABP3....	163
3.7 FABP3 combines with HKLM to promote NO production by BMDMs in a dose-dependent manner .....	165
4. Discussion.....	168
<b>Chapter 7 General discussion and conclusions</b> .....	<b>173</b>
1. General discussion.....	173
2. Conclusions .....	179
<b>References</b> .....	<b>181</b>
<b>Appendix I</b> .....	<b>203</b>
<b>Appendix II</b> .....	<b>208</b>

## List of figures

Figure 1-1 Morphology of Echinococcus species.....	3
Figure 1-2 Life cycle of <i>E. multilocularis</i> .....	5
Figure 1-3 Global distribution of echinococcosis caused by <i>E. multilocularis</i> .....	8
Figure 1-4 Biogenesis of miRNAs in animals .....	22
Figure 1-5 The crystal structure of <i>E. granulosus</i> FABP1. ....	39
Figure 2-6 Depletion of <i>E. multilocularis</i> mitochondrial transcripts.....	55
Figure 2-7 Distribution of the paired-end reads of the PCR-free library with exonuclease treatment (A and B) or with RNase H and exonuclease double treatments (C and D). ....	59
Figure 2-8 Coverage of individual scaffold for two PCR-free sequencing libraries with exonuclease treatment (A) and with RNase H and exonuclease double treatments (B). ....	60
Figure 2-9 Correlation of the mean coverage of 438 supercontigs commonly covered by pf-mm and dpf-mm.....	62
Figure 3-10 Differential expression of miRNAs in <i>E. multilocularis</i> .....	77
Figure 3-11 Conservation of the miR-71b/2/752/13 cluster in platyhelminths .....	84
Figure 3-12 Alignment and putative secondary structure of the miR-71b/2/752/13 cluster in <i>Echinococcus</i> species.....	85
Figure 3-13 Phylogenetic distribution of miRNAs in platyhelminths .....	86
Figure 4-14 Fatty acid binding-related residues in intronless FABP genes of invertebrates.....	101
Figure 4-15 Alternative splicing in invertebrate FABP genes.....	104
Figure 5-16 Schematic presentation of genomic organization and gene structures of Echinococcus FABPs.....	124
Figure 5-17 Alignment of Echinococcus FABP sequences.....	126
Figure 5-18 Evolutionary relationship of Echinococcus FABPs inferred from amino acid or intron sequences.....	129
Figure 5-19 mRNA levels of emFABPs in vesicles, cell aggregates and dormant protoscoleces. ....	131
Figure 5-20 SDS-PAGE analysis of rSUMO/FABPs and efficacy of SUMO protease. ....	132
Figure 5-21 SDS-PAGE analysis of purified yeast-derive rFABP3 in a native form. ....	134
Figure 5-22 Binding assay using DAUDA and <i>cis</i> -parinaric acid.....	136
Figure 5-23 Binding assay of yeast-produced rFABP3 without any tags.....	137
Figure 5-24 Determination of rabbit anti-rFABP3 antibody level by ELISA. ....	139
Figure 5-25 Confirmation of native FABP3 present in the hydatid fluid by western blot. ....	139
Figure 6-26 rFABP3 effects on cytokine production by BMDMs and BMDDs. ....	153
Figure 6-27 Anti-rFABP3 neutralization of rFABP3 functions. ....	156
Figure 6-28 Activation of TLR transformant cells by rFABP3 via TLR 2.....	159
Figure 6-29 rFABP3-induced cytokine secretion by BMDMs via TLR 2.....	162
Figure 6-30 Effects of rFABP3 on macrophage arginase I expression. ....	162

Figure 6-31 Effects of adipocyte FABP inhibitor on cytokines induced by rFABP3.  
..... 164

Figure 6-32 Effects of rFABP3 on NO production by BMDMs and BMDDs. .... 167

## List of tables

Table 1-1 Strains and hosts of Echinococcus species.....	2
Table 1-2 Comparison of sequencing technologies and advantages of RNA-seq over other ways for investigation of transcriptomes.....	18
Table 1-3 miRNAs and/or RNA-induced silencing machinery in parasites.....	26
Table 1-4 Potential functions of parasite miRNAs.....	31
Table 2-5 Primers and probes used in this study. ....	50
Table 2-6 Mapping results of different <i>E. multilocularis</i> samples prepared by different methods.....	57
Table 2-7 First 120 supercontigs commonly covered by two PCR-free data sets .....	61
Table 3-8 Statistic results of short sequences generated for characterization of <i>E. multilocularis</i> miRNAs.....	72
Table 3-9 Potential novel miRNAs of <i>E. multilocularis</i> .....	73
Table 3-10 Some conserved miRNAs of <i>E. multilocularis</i> .....	76
Table 3-11 Differential expression of known and novel miRNAs of <i>E. multilocularis</i> .....	78
Table 3-12 Potential neoblast-specific miRNAs of <i>E. multilocularis</i> .....	81
Table 4-13 Distribution and features of FABP genes in invertebrates .....	96
Table 4-14 Diversity of FABP genomic structures in invertebrates.....	99
Table 5-15 Primer sets using for amplification and expression.....	113
Table 5-16 Comparison of emFABPs.....	124
Table 5-17 FABP homologues in <i>E. granulosus</i> .....	127

## Chapter 1 Introduction

The Phylum Platyhelminth is comprised of four classes with approximately 20,000 species, from completely free-living planarians to parasitic flatworms. Based on indirect evidence, the appearance of platyhelminths (Neodermata) may be predicted to have occurred around 560~540 million years ago (Maule and Marks, 2006). Generally, platyhelminths have no cuticle or protective covering and usually reside in damp or aquatic habitats. With a bilateral symmetry and a low level of cephalisation, flatworms are usually small but some of them can reach several metres. For example, *Taenia solium*, of the causative agent of cysticercosis in humans and pigs, can reach seven metres in length. During evolution, parasitic flatworms have evolved to extend their hosts from invertebrate to mammalian animals, commonly through one or more hosts to complete their life cycles. Some of parasitic platyhelminths, such as *Schistosoma mansoni*, *Echinococcus granulosus*, *E. multilocularis* and *T. solium*, are of medical or veterinary significance in developing or/and developed countries.

*Echinococcus* species are characterized by being very small, usually no more than seven millimetres, and have less than six segments. Moreover there is no gut in *Echinococcus spp.* and the teguments function for metabolic exchange. The adult worms have a scolex, which is comprised of four suckers and two rows of hooks, and segmented proglottids in which the eggs are produced. The taxonomy of *Echinococcus* species is still controversial. Initially, sixteen species and thirteen subspecies have been grouped into the genus *Echinococcus*, family Taeniidae, Cyclophyllidea, Cestoda, but only four species are well recognized: *E. vogeli*, *E. oligarthrus*, *E. granulosus* and *E. multilocularis* (Table 1-1) (Maule and Marks, 2006; Thompson, 2008). They are distinct in morphology (Fig. 1-1), useful for differential

**Table 1-1 Strains and hosts of Echinococcus species**

Species	Genotype/strain	Intermediate hosts	Definitive hosts
<i>E. granulosus</i>	G1/Sheep	sheep, cattle, human, goat, buffalo, cameral, pig, wild boar, kangaroo	dog, dingo, wolf, fox, jackal, hyena
	G2/Tasmanian sheep	sheep, human	dog, fox
	G3/Buffalo	buffalo	dog
	G4/Horse <sup>b</sup>	horse, donkey	dog
	G5/Cattle <sup>b</sup>	cattle, sheep, goat, buffalo, pig, human, zebra	dog
	G6/Camel	camel, cattle, human, sheep, goat	dog
	G7/Pig	pig, wild boar, goat, beaver, cattle, human	dog
	G8/Cervid	moose, human	wolf, dog
	G9/ND <sup>c</sup>	human	ND
	G10/Fennoscandian cervid	reindeer	wolf, dog
<i>E. multilocularis</i>	ND/Lion	zebra, wildebeest, warthog, bushpig, buffalo, antelope	lion
	Some isolate variations	Rodents <sup>d</sup> , pig, wild boar, dog, monkey (many types)	Fox, dog, cat, wolf, racoon-dog, coyote
<i>E. vogeli</i>	- <sup>e</sup>	rodents	bush dog
<i>E. oligarthrus</i>	-	rodents	wild felids
<i>E. shiquicus</i>	ND	pika	Tibetan fox

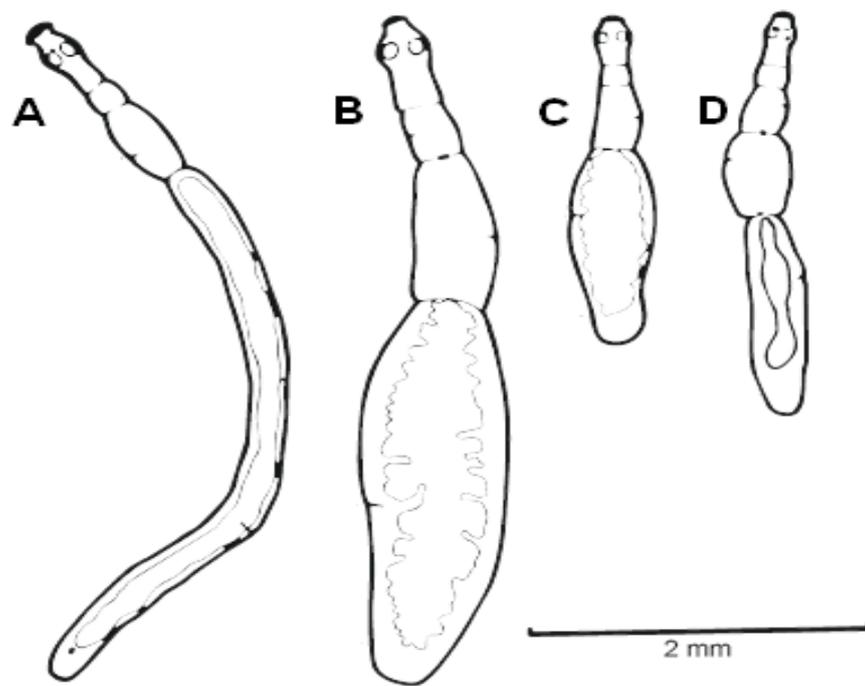
<sup>a</sup>The data from (Maule and Marks, 2006; Staebler et al., 2007; Thompson, 2008);

<sup>b</sup>These *E. granulosus* strains are also called as *E. equines* and *E. ortleppi* in some literatures, respectively;

<sup>c</sup>'ND': not determined;

<sup>d</sup>Including squirrel and chinchilla

<sup>e</sup>'-': not found.



**Figure 1-1 Morphology of Echinococcus species**

A: *E. vogeli*; B: *E. granulosus*; C: *E. orligarthrus*; D: *E. multilocularis*

Note: These figures were abridged from WHO/OIE manual on Echinococcosis.

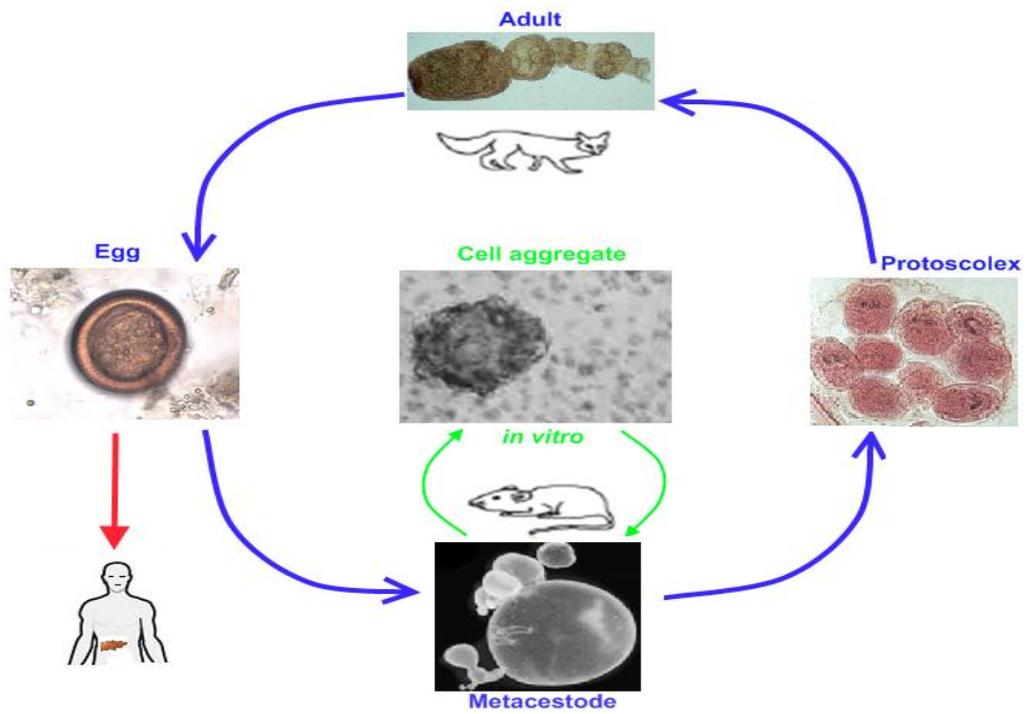
identification. Recently, a new species has been described: *E. shiquicus*, in pika, *Ochotona curzoniae*, in Sichuan and fox, *Vulpes ferrilata*, in Tibet, China, but its zoonotic potential remains to be determined (Xiao et al., 2005). Because of limited evidence, the taxonomical position of this parasite is still controversial.

Two Echinococcus parasites, *E. granulosus* and *E. multilocularis* that cause cystic echinococcosis (CE) and alveolar echinococcosis (AE) in animals and humans, respectively, are well studied. *E. granulosus* is characterized by the biological feature of intra-species divergence that seems to be associated with its physiology, pathology and development, so far up to eleven different strains or isolates G1~G10 identified (Table 1-1), which is rarely observed in other three Echinococcus species (Maule and Marks, 2006; McManus, 2009; McManus et al., 2003). Although the evolutionary relationship amongst *Echinococcus spp.* remains controversial (Maule and Marks, 2006; Saarma et al., 2009), *E. multilocularis* may have diverged from the other three parasites at an earlier time as shown by using a combination of nuclear or/and mitochondrial molecular markers.

## **1. *Echinococcus multilocularis***

### **1.1 Life cycle and and significance in public health of alveolar echinococcosis caused by *E. multilocularis***

*E. multilocularis*, regarded as an independent species in the 1950's (Tappe et al., 2010a), is a tiny tapeworm of the fox with four or five segments, around 1.2~4.5mm, and completes its life cycle via transmission between two different hosts (Fig 1-2). In the intestine of the definitive host such as foxes or dogs, the adult worm produces many of mature eggs that are released to environment with host faeces. When small rodents or humans eat or drink contaminated food or water, the eggs are digested,



**Figure 1-2 Life cycle of *E. multilocularis***

Note: Some miniatures were abridged from (Gottstein and Hemphill, 2008; Spiliotis et al., 2008).

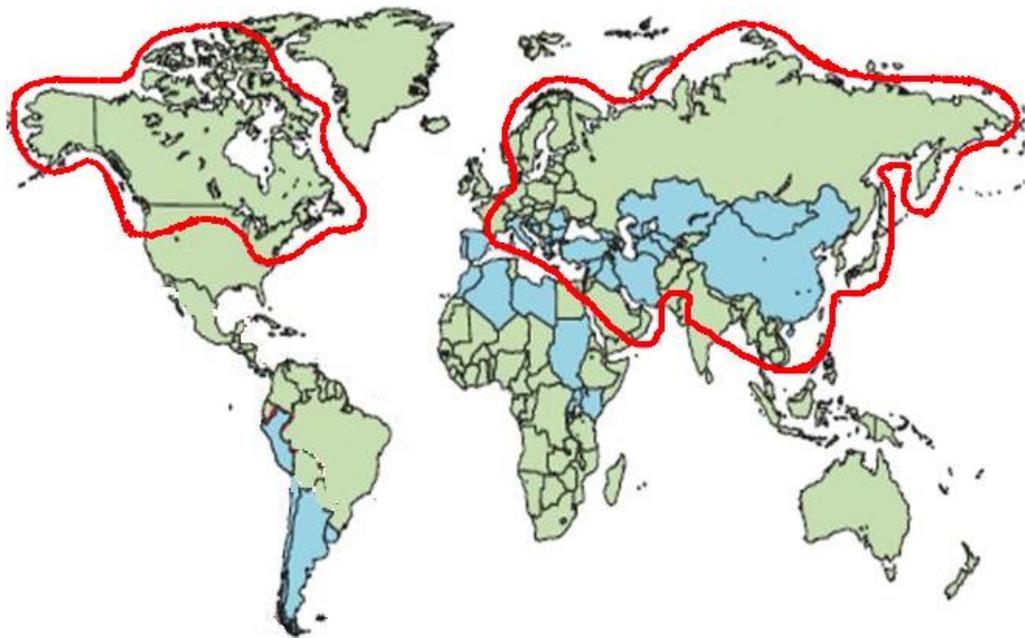
leading to the discharge of oncospheres in the presence of the stomach and intestinal fluids. Infective oncospheres develop into metacestodes in the liver or rarely other organs. The encysted metacestodes proliferate quickly in a tumor growth-like manner through new vesicles budding off from the germinal layer (metacestode, Fig. 1-2) to infiltrate the surrounding tissues. A large number of protoscoleces are generated by the metacestodes in the natural intermediate hosts but not in human beings for unknown reasons (Moro and Schantz, 2009). In the intestine of the definitive hosts, activated protoscoleces develop into the adults thus completing the life cycle (Fig. 1-2).

The larva of *E. multilocularis* is fully filled with transparent hydatid fluid and protoscoleces. The cyst is comprised of an inner germinal layer and an external carbohydrate-rich laminated layer that acts as a physical barrier against the host immune response (Eckert et al., 2002). Through digitally modelling a formation of vesicles, *E. multilocularis* larvae development has been revealed that the *in vivo* growth in the human liver exhibits a complex pattern (Tappe et al., 2010b). In contrast to *E. granulosus* larva that generates a single cyst, an individual *E. multilocularis* vesicle produces numerous offspring vesicles derived from the neoblasts, a class of undifferentiated cells, of the germinal layer. Whilst *E. granulosus* grows slowly and continually in its natural intermediate host, the growth and proliferation of *E. multilocularis* vesicles are rapid and they only grow for a short period. This is regarded to be an adaptation to the short life span of the host.

Alveolar echinococcosis (AE) has re-emerged and is considered as a neglected disease. It is estimated that there are nearly two billion humans infected with helminths in the world and around 2~3 million cases of Echinococcosis with 0.3~0.5 million being due to AE (Craig et al., 2007). AE is predominantly endemic in cold areas or regions in the northern hemisphere (Fig. 1-3), including Poland, Germany, Russia, China, Japan, Canada and United States of America. In Poland, epidemical investigations over three

years showed that the mean prevalence of *E. multilocularis* in fox populations reached 23.8% and it was as high as 62.9% in some areas (Malczewski et al., 2008). The fatality rate of untreated patients is very high, likely to exceed 90 percent. The prevalence and transmission of the disease depends on many factors, such as fox populations, rodent intensity, trade and breeding systems (Craig et al., 2007; Eckert and Deplazes, 2004; Guislain et al., 2008; McManus et al., 2003; Moro and Schantz, 2009; Pavlin et al., 2009). Correlated with the increase in fox population size in some parts of Europe, the cases of human AE increased by an average rate from  $1.0 \times 10^{-6}$  during 1993~2000 to  $2.6 \times 10^{-6}$  during 2001~2005 (Schweiger et al., 2007). Investigations have also confirmed the re-emergence of the disease in previously non-endemic regions due to the migration or re-distribution of fox populations.

Although the infectious rate of AE is much lower than CE, it is still a great threat to human health. Efficient control approaches including early diagnosis and treatment measures remain challenging (Brunetti et al., 2010; Eckert et al., 2002). The latent period of AE is up to 15 years, followed by a chronic infection, which makes earlier diagnosis and treatment difficult. The primary position of encystment of *E. multilocularis* larvae is the liver in humans and natural intermediate hosts and, clinically, hepatic vesicles are found in nearly 99% of AE cases (McManus et al., 2003). The metacestodes develop in a tumor-like growth manner and surgical dissection of the larval mass is preferable (Eckert and Deplazes, 2004), although there is a high risk of incomplete resection and recurrence. Although in many cases drugs such as mebendazole and albendazole are effective in prolonging the life of AE patients, the failure of chemotherapy has been reported (Liu et al., 2009b). Moreover, metacestode neoblasts can be released into host's blood and lymph systems (Hemphill et al., 2010), leading to re-infection in other tissues or organs. Whilst this disease is largely neglected (Budke et al., 2009), the rapid re-emergence and spread of AE across the world may prompt us to re-evaluate its public health significance.



**Figure 1-3 Global distribution of echinococcosis caused by *E. multilocularis***

AE is mainly distributed in the areas or regions circled in red. Note: this figure was modified from (Craig et al., 2007).

## 1.2 Neoblasts and establishment of an *in vitro* cultivation system

*Schmidtea mediterranea*, a free-living platyhelminth, possesses remarkable regeneration capacity even from little as 1/279 of a spliced body piece (Newmark and Sanchez Alvarado, 2002). This biological trait is attributed to neoblasts, somatic stem cells, which are distributed throughout the body. The neoblasts of *S. mediterranea* play roles in wound repair and in environmental adaptation via increasing or decreasing the cell population size (Sanchez Alvarado and Kang, 2005).

Neoblasts, which were initially and commonly called germinal cells in flatworms, are also found in other platyhelminths but the regenerative capacity seems limited in comparison with planarians. *E. multilocularis* neoblasts are distributed in the germinal layer towards the cavity of vesicles. Early studies have demonstrated that the neoblasts from *E. multilocularis* have the capacity to develop into larvae in susceptible hosts, and that these cells can survive *in vitro* for up to eighteen months and remain infective to cotton rats (Dieckmann and Frank, 1988; Furuya, 1991). With low oxygen and the addition of reducing reagents, *E. multilocularis* vesicles can survive *in vitro* several months and yield protoscoleces in the presence of unknown factor(s) secreted by mammalian cells (Spiliotis et al., 2004). Subsequently, similar methods have been applied for the cultivation of primary cells from vesicles. Under specific conditions, these cells, which include more than 30% of neoblasts, proliferate and aggregate, which is followed by the formation and enlargement of an inner cavity (Spiliotis et al., 2008). These vesicles mature and can reproduce and generate protoscoleces in mice once the outer laminated layer appears, illustrating that *E. multilocularis* neoblasts retain the potential to differentiate into all types of cells. Recently, this system has been successfully utilized to assess the efficacy of gene knockdown (Mizukami et al., 2010; Spiliotis et al., 2010), thus allowing reverse genetics.

In addition to the achievement of *in vitro* development from the primary cells to the infectious vesicles, system has been also modified to remove the necessity for feeder cells and thus decontaminating material from host (Spiliotis and Brehm, 2009). All the samples in our studies were therefore prepared using this host cell-free cultivation approaches.

### **1.3 Genomics of *E. multilocularis***

With a few exceptions from the Eucestoda and Turbellaria families, the majority of platyhelminths tend to have a haploid genome with a small number of chromosomes, generally less than ten with no sex chromosomes (Maule and Marks, 2006). The karyotypes of cestodes appear to be various, ranging from having between six and twenty chromosomes. For *E. granulosus* and *E. multilocularis*, the diploid chromosome number is 18 and there are no chromosome morphological differences between these two species (Petkeviciute and Ieshko, 1991; Rausch and Rausch, 1981). It is well known that changes in chromosome number and size occur during *in vitro* cultivation, largely due to endoreduplication. After 40 passages, the cells from *E. multilocularis* larvae were shown to have chromosome numbers between 14 to 104 that were classified into telocentric, subtelocentric and metacentric types but 91~100 chromosomes were dominant (Furuya, 1991). In addition to this, the chromosome morphology tends to be related to chromosome quantity and genome rearrangement may occur during cultivation, evidenced by morphological changes that are commonly discernible in cells with more than 59 chromosomes or more. Interestingly, these cells with abnormal karyotypes still exhibit infective ability and grew into cyst masses when inoculated into cotton rats (Furuya, 1991), suggesting the retention of differential capacity in these cells. However, the fate of these extra chromosomes post inoculation is unclear. The dramatic alteration in chromosome number has been also recently described in *in vitro* cultured cells of *E. granulosus* (Echeverria et al., 2010).

The completion of the draft sequence of human genome has fuelled studies on genomes and related fields of other eukaryotic organisms. In particular, the advent of next-generation sequencing technologies, such as 454, Illumina and SOLiD that are theoretically different from Sanger sequencing approach, renders the possibilities for sequencing entire genomes in an affordable way. In flatworms, the variation of nuclear genome sizes is striking, ranging from 50Mb for *Turbellaria Stenostomum spp.* to 18,390Mb for *Turbellaria Otomesostoma auditivum* (Maule and Marks, 2006). But the huge divergence in size requires careful interpretation due to the presence of polypoidy in *Turbellaria* species. In 2004, the *E. multilocularis* genome sequencing was initiated and, with combination of Sanger, 454 and Illumina sequencing platforms, more than 1,800 contigs and approximately 600 supercontigs are to date generated, a total size being up to 109Mb. At present, the genome sequencing is entering the finishing stage and the genome being annotated. Using flow cytometry, it is estimated that the genome size of *E. multilocularis* ranges from 270Mb to 390Mb (Spiliotis et al., 2008), similar to *E. granulosus* of which the haploid genome size is nearly 150Mb (Maule and Marks, 2006) and to *T. solium* whose genome is 250~270Mb (Aguilar-Diaz et al., 2006). The GC content of *E. multilocularis* genomic DNA is approximately 44% and a bias of codon usage is observed in the coding regions of highly expressed genes in *E. granulosus* (Fernandez et al., 2001; Smyth and McManus, 1989), which has also been reported in *Schistosoma spp.* (Maule and Marks, 2006).

The mitochondrial genome of *E. multilocularis* is 13,738bp long with a high AC content up to 69%, comprising 2 noncoding regions and 36 genes and lacking a coding ATPase subunit 8 that does not exist in nematode mitochondrial DNA (Nakao et al., 2002). With few or no introns or spacers, all the genes are tightly organised on the same strand, which is often conserved in the Phylum Platyhelminth (Le et al., 2000; Maule and Marks, 2006). Intriguingly, two noncoding regions contain inverted

repeats that are able to form hairpin structures and may function in replication and transcription (Nakao et al., 2002).

Numerous families of repetitive sequences, contributors to a genome size, are present in all parasitic flatworms investigated including *Echinococcus* species and they seem to be valuable in diagnosis and species identification because the repetitive sequences/elements are species- or genus-specific with regard to expression (Maule and Marks, 2006). *S. mansoni*, for example, possesses 40% repetitive sequences and *S. japonicum* 40.1% containing 72 families of long-terminus repeat (LTR) and non-LTR transposons that account for 15% and 5% of the genome (364Mb), respectively (2009b; Berriman et al., 2009). There are several types of repetitive sequences present in *E. granulosus*, of which the 186bp-long TREG element is genus-specific at strain-dependent copies from 120 to 23,000, while the EgG1 Hae III is species-specific and dispersed in the genome with approximately 7,000 copies (Marin et al., 1993; Rosenzvit et al., 2001; Rosenzvit et al., 1997). The functions of these repetitive elements, however, remain elusive.

Collectively, with the establishment of the *in vitro* cultivation system and the availability of the genome, *E. multilocularis* is a promising model to study evolutionarily close species. To facilitate an effective and wide use as a model, deep understanding of *E. multilocularis* biology especially at the molecular level is therefore necessary and imperative.

## **2. Transcriptome and RNA-seq**

It is believed that organisms maintain their stability against external and internal changes by regulating transcription or/and post-transcription and translation. The level of RNA expression can be considered as an indicator of the conditions to which an organism is exposed. Characterization of all the RNAs transcribed therefore helps us

understand profound bio-processes. Genome-wide spatial and temporal transcription is expected to be closely connected to the development, growth and physiology of *E. multilocularis*. In particular, the expression of gene-coding and non-coding regulatory RNA is essential and the baselines of their transcription at different cells and stages are needed for the understanding of gene expression patterns through the life cycles and the applications of *E. multilocularis* as a model, such as gene manipulation using loss-of-function methods.

## **2.1 Transcriptome: an entire dynamic RNA profile**

The term transcriptome was first used in yeast to describe the genes that were being expressed and their expression levels at distinct cell phases (Velculescu et al., 1997). The term is defined as all expressed transcripts in a tissue or organ or even a cell at a fixed time under specific physical conditions and is regarded as a dynamic junction between the genomes and physical features. The components of transcriptomes cover both non- and protein-coding RNAs, as is referred to as the mRNA transcriptomes. Transcriptomes can be complex due to tissue- or stage-specific expression and various flexible post-transcriptional splicing, especially in higher eukaryotic organisms. In human tissues, it is estimated that alternatively spliced transcripts from genes with multiple exons compose up to 95% and more than seven alternative splicing events occur in every multi-exon gene (Pan et al., 2008).

Transcriptome analysis provides a wealth of bioinformation as it records direct information of the transcription of an organism's genome. mRNA transcriptome analysis allows one to analyze such expression patterns as alternative splicing and find new exons, especially ones of a small size, or genes or SNPs, which facilitates the annotation of the genome (Sultan et al., 2008). In particular, it is quite useful for annotating genomes of organisms that are poorly understood (Yassour et al., 2009). For example, using mRNA transcriptome data, more than 98,000 splicing junctions

were found in *Caenorhabditis elegans*, much higher than that previously described (Hillier et al., 2009). Because organisms always up- or down-regulate the transcriptional levels of some mRNAs in response to changes in internal or/and external conditions, mRNA expression profiling will allow us to recognize an organism's strategies in response to the perturbation at the genetic level, possibly aiding to uncovering the signal or metabolic pathways involved (Jiao et al., 2009). Additionally, transcriptome analysis is a powerful robust tool for detection of such genetic aberrations that are implicated in tumour occurrence (Maher et al., 2009; Zhao et al., 2009). Moreover, transcriptome analysis is a powerful tool to be able to compare the up or down regulation of gene expression in cells or tissues exposed to different conditions. A study comparing normal and abnormal tissues in humans revealed that there are approximately 1,000 genes expressing in every cell type and only a small portion of genes that are expressed in a tissue-specific manner, with most of unique genes being expressed at a low level (Velculescu et al., 1999). This analysis gives us valuable data on the minimal genes necessary for fundamental functions.

## **2.2 RNA-seq: a deep high-throughput technology for transcriptional characterization**

The output of transcription of an organism's genome varies according to external environmental stimuli (Gibson, 2008) and many internal factors such as copy number variation (CNV) (Henrichsen et al., 2009), alternative splicing (Pan et al., 2008; Sultan et al., 2008) and other unknown post-transcriptional modifications (2009a). It is proposed that some mature mRNAs are either translated into proteins or, through unknown mechanisms, spliced into long RNAs and then further processed to produce small RNAs, under some circumstances, possibly with a capped structure (2009a).

Furthermore, albeit under the same stress, almost identical cells show different expression patterns (Cohen et al., 2008). Genes are randomly transcribed, giving rise

to variations of mRNA profiles from cell to cell. In the Burst hypothesis, it is believed that genes are in a randomly fluctuating state between activation and inactivation (Raj and van Oudenaarden, 2008). As for a given cell or tissue, the abundance among different transcripts is markedly diverse. It is well known that much more of the human genome is transcribed than previously expected (Preker et al., 2008). A global analysis of human transcripts indicates that the transcribed genes in a single cell type reach up to 43,500 and the expression levels are from 0.3 to 9,417 copies per cell (Velculescu et al., 1999). Consequently, the characterization of integral transcriptome of one organism poses a great challenge.

A pool of full-length cDNAs is the direct real reflection of transcription of one organism's genome, providing an invaluable resource for the most global survey of transcriptome (Okazaki et al., 2002). Such data, however, are quite limited. Initially, many approaches comprising hybridization- and tag-based ways have been developed for the characterization of transcriptomes. Serial analysis of gene expression (SAGE), one of the tag-based technologies, was successfully utilized for a comprehensive investigation of the yeast transcriptome, revealing the gene expression level from 0.3 to >200 copies per cell and a uniform distribution of transcription among the chromosomes (Velculescu et al., 1997). Afterwards, other Sanger sequencing-based approaches for quantitative gene expression analysis have effectively been used for a survey of transcriptomes. But these technologies have intrinsic limitations or/and disadvantages (Wang et al., 2009b), which greatly restrict the extensive application in transcriptome characterization. For instance, it is hard to distinguish isoforms from each other.

Different from Sanger sequencing (first generation), next-generation sequencing technologies, including 454, Illumina and SOLiD, directly sequence cDNA libraries without vector cloning and proliferation. These technologies are in principle different from the first generation sequencing technology: amplification by pyrosequencing

(454) or reversible dye-termination sequencing (Illumina) or sequencing by ligation (SOLiD) (Table 1-2). Next-generation sequencing technologies are characterized by massive throughput per run at an affordable cost and they generally generate several millions of reads ranging from 30 to 400 bp (Table 1-2). For example, an new Illumina platform, Hiseq2000, can generate up to 2 million paired-end 100bp sequences per run within 7~8 days (Zhao and Grant, 2011). Moreover, they are the most well suited to deep-coverage sequencing and investigation of a shotgun metagenome due to generation of the huge number of reads (Scholz et al., 2012). These platforms can be used to depict a digital single base-resolution transcriptional map of a genome and to uncover rare or extremely low frequency variations that are associated with disease occurrence (Kamb, 2011; Wang et al., 2009b). Because of the above advantages, these sequencing technologies have been successfully applied for characterization of mRNA profiles of humans (Sultan et al., 2008), mouse (Cloonan et al., 2008) and yeast (Nagalakshmi et al., 2008), and miRNA profiles of chicken (Glazov et al., 2008), human (Morin et al., 2008), murine leukemia progression model (Kuchenbauer et al., 2008), etc..

Great sequencing capacity of these approaches fulfils characterization of near-complete transcriptomes and a technology, called as RNA-seq, was developed. Ultrahigh-throughput sequencing has also been extended to probe genome-wide interactions between proteins and DNA (Johnson et al., 2007) and tissue-specific enhancers (Visel et al., 2009), which is named as chromatin immunoprecipitation assay, ChiP-seq. Cumulative studies show that RNA-seq is in a robust, reproductive, deep and quantitative manner (Marioni et al., 2008; t Hoen et al., 2008), revolutionizing our view towards transcriptome studies. RNA-seq is capable of producing single-base resolution transcription maps and shows strong strengths to circumvent problems or difficulties that previous approaches confronted (Wang et al.,

2009b). For example, compared with microarray, RNA-seq was able to identify 25% more genes (Sultan et al., 2008).

However, RNA-seq also has limitations. First, short reads generated by Illumina 1G can have base substitution, transversion, deletion and insertion errors, and the distribution of the reads is uneven, intensified in regions with increased GC content. Error rates tend to increase towards the end of the reads and they can be corrected by increase of coverage of the regions sequenced (Dohm et al., 2008). Coverage depth varies from the 5' end to the 3' end or from the sense strand to the antisense strand (Hillier et al., 2009). In addition, a bias is inevitably introduced in the construction of sequencing libraries. For example, RNA fragmentation causes a drop of coverage at both 5' and 3' ends, whereas cDNA fragmentation leads to sharp increase of coverage at the 3' end (Wang et al., 2009b). Fortunately, some biases seem to be minimized by technological improvements (Quail et al., 2008).

Table 1-2 Comparison of sequencing technologies and advantages of RNA-seq over other ways for investigation of transcriptomes

	Sanger sequencing/ABI	454/Roche	Illumina/Solexa	SOLiD/ABI
	Microassay	cDNA and EST sequencing		RNA-seq
Sequencing approach	Synthesis with the dye terminators	Pyrosequencing on solid support	Polymerase-based sequencing by synthesis with reversible terminators	Ligation-based sequencing
Read length	Up to 900 bp	200-300 bp	37-108 bp	35 bp
Mb/run	0.096	80-120	>1,300	3,000
Cost per Mb	High	Low	Extremely low	Extremely low
Method	Hybridization	Sanger sequencing		Next-generation sequencing
Resolution	Several to 100 bp	Single base		Single base
Throughput	High	Low		Extremely high
Quantification of gene expression level	Up to a few hundred fold	Not applicable		More than 8000 fold
Cost	High	High		Reasonably low

Note: the data mainly come from (Mardis, 2008; Morozova and Marra, 2008; Wang et al., 2009b) .

The Illumina 1G genome analyzer platform is based on sequencing by bridge synthesis with four different fluorescent reversible terminators (Mardis, 2008). Briefly, DNA fragments with specific adapters are immobilized on an eight-lane glass flow cell pre-bound with the same adapters followed by a few cycles of amplification to create the clusters with the nearly 1,000 identical copies. To trigger sequencing, all four labelled fluorescent nucleotides with a 3'-OH group, primers and polymerase are added and only one base is incorporated per cycle. Incorporated nucleotides in each cluster are captured by an imaging step using laser. After that, the fluorescent groups are removed, releasing the 3' end for addition of next labelled nucleotide. The raw sequences of reads are determined based on physical images and the valid reads are obtained through filtration. DNA fragments are sequenced, yielding single-end reads from one end or pair-end reads from both ends. Generally, it can generate at least one gigabase pairs sequence every run, albeit reads only 37~108 bp in length.

In summary, RNA-seq has been demonstrated to be a reliable, productive and robust approach for transcriptome investigation although challenges such as development of bioinformatic tools for sequence processes still exist. Considering many advantages mentioned above, this deep sequencing technology was employed to obtain transcriptional signals of mRNA and miRNAs in our project.

### **3. microRNA**

Noncoding RNAs (ncRNAs), which are not translated, are the main component of the repertoire of transcribed RNAs of eukaryotic organisms. ncRNAs are often grouped into long RNAs (lRNAs, longer than 200 bp) and short RNAs (sRNA, less than 200 bp). microRNAs (miRNAs) are one subtype of small endogenous single-stranded ncRNAs. The length of miRNAs varies from 19~30 bp but most are 21~24 bp. At present, there are no reports on experimental investigations of *E. multilocularis* miRNAs. Initially, we used *S. mansoni* key components for miRNA biogenesis to

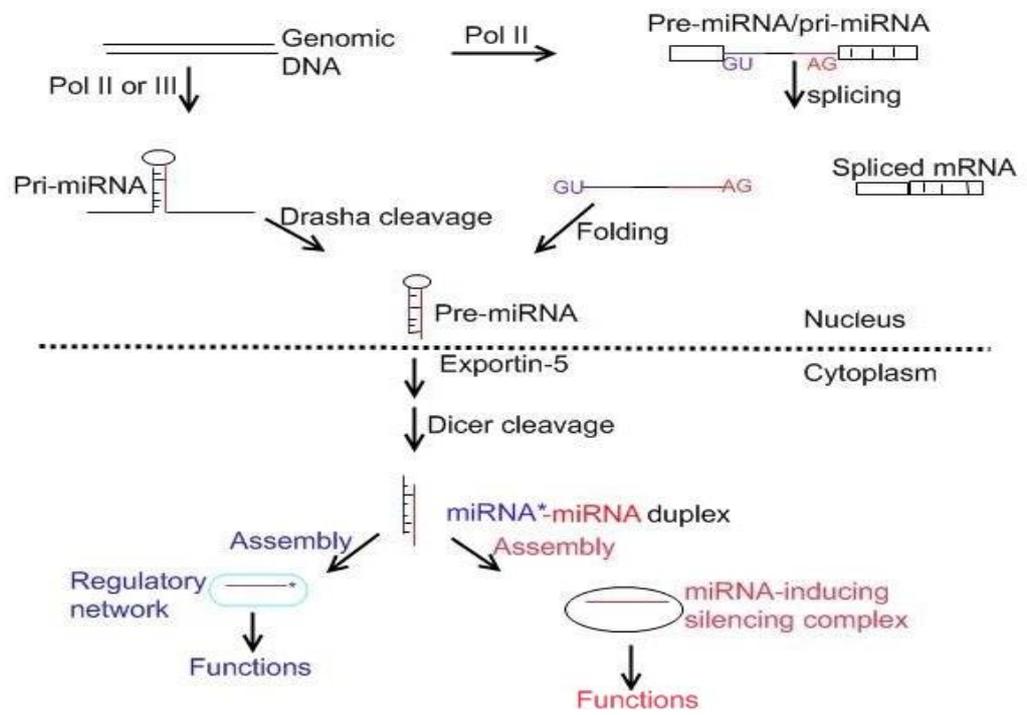
search *E. multilocularis* genome using Blast and the resulting hits with high scores suggested that this parasite encode miRNAs. Therefore our study attempted to profile genome-wide miRNAs, helping us to understand gene regulation and develop an alternative tool for manipulation of gene expression abundance.

### **3.1 Biogenesis and genetic characteristics of miRNAs**

In contrast to small interfering RNAs (siRNAs) that are primarily from mRNA, transposons or viral DNA, miRNAs are generally derived from exonic, intronic and intergenic regions or transposable elements, and a few miRNAs are derived from exons including untranslated regions (UTRs). In the canonical approach, miRNA genes are firstly transcribed into long pri-miRNA transcripts under the presence of RNA polymerase (RNA Pol), usually Pol II. These pri-miRNAs are specifically spliced by Drosha, a nuclear RNase III, giving rise to the precursors of miRNAs with a size of approximately 70 nucleotides (pre-miRNAs). With the help of a nuclear transport receptor (exportin-5) that acts with molecules with a stem and a short overhang at the 3' end, the pre-miRNAs in the nucleus are transported to the cytoplasm. Afterwards, the molecules exported are cleaved into smaller duplexes, which is mediated by RNase III Dicer. To be functional, double-stranded shortened miRNAs, which are swiftly changed into a single stand in the assembly, are loaded onto an Argonaut-containing RNA-induced silencing complex (Scaplehorn et al.) (Ruby et al., 2007). In most cases, the partner of the mature miRNA incorporated into the complex, named as miRNA\*, is degraded. Recently, it has been shown that miRNAs\* can also be present at abundant levels and have the capacity of repressing target sequences (Okamura et al., 2008). This finding adds to the complexity of regulatory networks where miRNAs or miRNAs\* dominate. It is noticed that the canonical synthesis of miRNAs is different in animals and plants, in that the duplexes of miRNA and miRNA\* are generated in the nucleus in plants but in the cytoplasm in animals (Bartel, 2004).

By contrast, the biogenesis of mirtron miRNAs, which are found in *C. elegans*, *Drosophila melanogaster* (Ruby et al., 2007), chicken (Glazov et al., 2008) and several mammals (Berezikov et al., 2007), is different only initially. The short intron-derived pri-miRNAs are spliced into spliced mRNAs and intron lariats. Without Drosha cleavage, the introns are folded directly to form pre-miRNAs with the hairpin structure (Ruby et al., 2007) and further processed as described above (Fig. 1-4). Murine herpes virus miRNAs are produced in yet another way (Bogerd et al., 2010). RNA Pol III instead of Pol II is used to produce pri-miRNAs that contain a tRNA-like structure at the 5' end and two stem loops 1 in the middle and 2 at the 3' end. The pri-miRNA is cleaved by tRNase Z but not by Drosha to yield the pre-miRNA that is then further edited in a traditional way to liberate mature viral miRNAs. Intriguingly, replacement of the viral tRNA and stem loop 1 seems no obvious effects on biogenesis of miRNAs. This finding may shed light on the mechanism whereby constructs comprising of tRNA and short hairpin RNA (shRNA) that have a similar structure of pre-miRNA are processed (Scherer et al., 2007). In *Giardia lamblia*, a parasitic protozoan, miRNAs are derived from small nucleolar RNAs (snoRNAs) that participate in modification of other types of RNA (Saraiya and Wang, 2008). Together with other facts that Drosha and Exportin 5 homologues are absent but two vital RISC factors Argonaute and Dicer do exist raise the possibility of another Drosha-independent pathway involved in miRNA biogenesis. The finding that snoRNAs can serve as miRNA precursors broadens our understanding of pre-miRNA origins.

The biogenesis of miRNAs from transcription to mature miRNAs is tightly controlled via modifications of miRNA-producing RNA and factors that form functional RISC or other complexes and these modifications can be offered by internal or even external molecules. Together with other proteins, such TRBP and PACT, Dicer forms the



**Figure 1-4 Biogenesis of miRNAs in animals**

machinery to process pre-miRNAs and its level can result from the stabilized TRBP through phosphorylation (Molnar et al., 2007). The connection between pre-mRNA splicing and miRNA biogenesis raises a question of what the temporal relationship is. Several studies have demonstrated that Drosha-dependent miRNA cleavage appears to be independent of pre-mRNA splicing and these two processes may occur simultaneously (Rhoads, 2010).

Based on the origins, miRNAs are grouped into exonic miRNAs in non-coding transcripts, intergenic miRNAs, intronic miRNAs in protein-coding transcripts or non-coding transcripts and mirtronic miRNAs. Previous analysis indicates that the majority of miRNAs are derived from introns in protein-coding genes (Kim and Nam, 2006). In contrast to siRNAs, many miRNAs are characterized by being well conserved (Bartel, 2004). miRNAs have many common remarkable features in the structure and genomic organization, although some characteristics of miRNAs are various from one species to another. Most miRNAs genes are clustered in the genomes and 70% of miRNA genes are distributed within transcription units of genes in the mammalian genomes (Kim and Nam, 2006). Except mirtronic miRNAs, beginning with U at the 5' end whose matching with the 3' A of mRNA promotes miRNA-inducing protein repression (Baek et al., 2008) is another striking trait of miRNAs with no conserved bases being commonly observed at the 3' end.

The secondary structure of miRNA precursor is a two-arm hairpin, always lacking large loops or bulges and bearing a 3' 2 bp overhang, which facilitates interactions with exportin-5 for transportation. In *C. elegans*, a large number of miRNAs are situated in the 3' arm of the hairpin structure (Lau et al., 2001). Mirtronic pre-miRNAs exhibit different characteristics and most are 60 nucleotides long, smaller than that of intronic, exonic and intergenic miRNAs. The bases at both 5' and 3' ends are pretty conserved in mirtronic precursors (Berezikov et al., 2007; Ruby et al., 2007). In addition to this, it is predicted that there are no perfect 2 bp overhangs at the 3'

terminus in most mammalian miRNAs. Furthermore, there is a strong bias in the GC content in mirtrons from different organisms.

Different miRNAs show a different expression pattern. The abundance of miRNAs greatly varies from 1 to hundreds of thousands of copies (Glazov et al., 2008; Kuchenbauer et al., 2008). Some miRNAs are expressed in a tissue- or stage-specific manner. For example, two *S. japonicum* miRNAs sja-miR-71 and sja-bantam are highly expressed in the cercarial stage and greatly repressed in the schistosomulum stage (Xue et al., 2008). Interestingly, although their functions are unclear, there are many miRNA isoforms from the same miRNA gene mostly by variations at the 3' end (Kuchenbauer et al., 2008; Morin et al., 2008), of which the non-templated variants are possibly introduced by modifications.

### **3.2 Approaches to the identification of miRNAs**

Direct sequencing is a main and convincing approach to investigate miRNAs and contributes to the discovery of most of miRNAs. This strategy involves the construction of small cDNA library (Hafner et al., 2008). Ultrahigh-throughput sequencing technologies satisfy the needs for characterization of entire miRNAs including ones that are expressed at a low level or in few cell types, impossible to be identified by conventional sequencing techniques.

Small RNAs are classified as miRNAs if they meet the following criteria (Berezikov et al., 2006; Kim and Nam, 2006): miRNAs and their precursor should be determined, usually by Northern blot hybridization; miRNA should be a part of one arm of the precursor hairpin structure without large internal loops; be processed by Dicer pathway; the structures of small RNAs and the hairpins are conserved in closely related species. Small homologous RNAs with known miRNAs in other species can be directly catalogued into miRNAs without further experimental validation.

### 3.4 Distribution of miRNAs in parasites

The miRNA *lin-4* was first reported in *C. elegans* in 1993 (Lee et al., 1993). However, the term ‘microRNA’ was coined to describe a subclass of small regulatory noncoding RNA with a size of 21 to 24 nucleotides in 2001 (Lau et al., 2001). miRNAs have been so far reported in viruses, animals and plants such as Marek's disease virus, fruit fly, human, zebra fish, chicken and Arabidopsis. But miRNA-inducing silencing mechanism by which gene expression is modulated is absent in yeast and some metazoan animals (Grimson et al., 2008). Gene expression mechanisms elicited by small regulatory RNAs may be widely distributed across the parasite kingdom. miRNAs were computationally or experimentally investigated in a number of parasite species, including parasitic protozoa, helminths and arthropods (Table 1-3), such as *Trypanosoma brucei* (Mallick et al., 2008), *Entamoeba histolytica* (De et al., 2006), *Clonorchis sinensis* (Xu et al., 2010), *Pediculus humanus humanus* (Kirkness et al., 2010), *Toxoplasma gondii* (Braun et al., 2010), *S. mansoni* (Gomes et al., 2009; Xue et al., 2008) and *S. japonicum* (Huang et al., 2009; Wang et al., 2010; Xue et al., 2008). It has, however, been confirmed that there are no miRNAs in the malaria parasite, *Plasmodium falciparum* (Rathjen et al., 2006). Moreover, the lack of coding genes of Dicer and Argonaute proteins, key components of RISC, in the genomes elucidates the absence of gene expression regulation triggered by miRNA in a few other protozoan parasites (Militello et al., 2008). Interestingly, apart from active Argonaute-coding genes, there are also Argonaute pseudogenes present in three Leishmania species (Table 1-3), suggesting a shrinking of RNA-inducing silencing functions during evolution. Conversely, the expansion of miRNA pathways has been demonstrated in the plant parasite, *Acyrtosiphon pisum*, although the biological significance of the miRNA machinery gene duplication events is elusive (Jaubert-Possamai et al., 2010). But, in some cases, the loss of this mechanism should be cautiously evaluated in that these coding genes may have undergone great variations

**Table 1-3 miRNAs and/or RNA-induced silencing machinery in parasites**

Species	Approach <sup>a</sup>	Num of miRNAs	Main RISC components <sup>b</sup>	Ref
<b>Protozoan<sup>c</sup></b>				
<i>Plasmodium falciparum</i>	Ex	No	No	(Rathjen et al., 2006)
<i>Trypanosoma brucei</i>	Com	1,162	Yes	(Mallick et al., 2008; Militello et al., 2008)
<i>Trypanosoma congolense</i>	-	ND	Yes	(Militello et al., 2008)
<i>Trypanosoma cruzi</i>	-	ND	Yes	(Garcia Silva et al., 2010; Militello et al., 2008)
<i>Trypanosoma gambiense</i>	-	ND	Yes (sanger)	
<i>Trypanosoma vivax</i>	-	ND	Yes (sanger)	
<i>Leishmania major</i>	-	ND	Yes/Pseudogene	
<i>Leishmania infantum</i>	-	ND	Yes/Pseudogene	
<i>Leishmania braziliensis</i>	-	ND	Yes (sanger)	
<i>Leishmania mexicana</i>	-	ND	Yes/Pseudogene	
<i>Leishmania tarentolae</i>	-	ND	Yes (TriTrypDB)	
<i>Leishmania braziliensis</i>	-	ND	Yes	(Militello et al., 2008)
<i>Cryptosporidium spp.</i>	-	ND	No	(Militello et al., 2008)
<i>Theileria spp.</i>	-	ND	No	(Militello et al., 2008)
<i>Babesia bovis</i>	-	ND	No	(Militello et al., 2008)
<i>Eimeria tenella</i>	-	ND	No	(Militello et al., 2008)
<i>Giardia lamblia</i>	Ex	4	Yes	(Saraiya and Wang, 2008)
<i>Neospora caninum</i>	-	ND	Yes	

Continued

<i>Giardia intestinalis</i>	Ex	10	Yes	(Chen et al., 2009; Militello et al., 2008)
<i>Trichomonas vaginalis</i>	Ex	11	Yes	(Chen et al., 2009; Militello et al., 2008)
<i>Entamoeba histolytica</i>	Com	17	Yes	(De et al., 2006; Militello et al., 2008)
<i>Toxoplasma gondii</i>	Ex	35	Yes	(Braun et al., 2010; Militello et al., 2008)
<b>Trematode</b>				
<i>Clonorchis sinensis</i>	Ex/Com	62,518	ND	(Xu et al., 2010)
<i>Schistosoma mansoni</i>	Ex	5	Yes	(Gomes et al., 2009; Xue et al., 2008)
<i>Schistosoma japonicum</i>	Ex	55	Yes	(Wang et al., 2010)
<b>Nematode</b>				
<i>Bursaphelenchus xylophilus</i>	Ex	810	ND	(Huang et al., 2010)
<i>Brugia malayi</i>	Ex/Com	32	Yes	(Poole et al., 2010)
<i>Haemonchus contortus</i>	-	ND	Yes (sanger)	
<i>Globodera pallida</i>	-	ND	Yes (sanger)	
<i>Trichuris trichiura</i>	-	ND	Yes (sanger)	
<i>Onchocerca volvulus</i>	-	ND	Yes (sanger)	
<i>Strongyloides ratti</i>	-	ND	Yes (sanger)	
<i>Trichuris muris</i>	-	ND	Yes (sanger)	
<i>Nippostrongylus brasiliensis</i>	-	ND	Yes (sanger)	
<b>Cestode</b>				
<i>Echinococcus multilocularis</i>	-	ND	Yes (sanger)	
<i>Echinococcus granulosus</i>	-	ND	Yes (sanger)	
<i>Hymenolepis microstoma</i>	-	ND	Yes (sanger)	
<b>Arthropod</b>				
<i>Pediculus humanus humanus</i>	Com	57	Yes (VectorBase)	(Kirkness et al., 2010)
<i>Acyrtosiphon pisum</i>	Com	163	Yes	(2010)
<i>Ixodes scapularis</i>	Ex	34	Yes (VectorBase)	(Wheeler et al., 2009)
<i>Aedes aegypti</i>	Ex	101	Yes (VectorBase)	(Li et al., 2009)
<i>Anopheles gambiae</i>	?	67 (miRBase)	Yes (VectorBase)	
<i>Anopheles stephensi</i>	Ex	27	ND	(Mead and Tu, 2008)

Continued

<i>Aedes albopictus</i>	Ex	65	ND	(Skalsky et al., 2010)
<i>Culex quinquefasciatus</i>	Ex	77	Yes (VectorBase)	(Skalsky et al., 2010)
<i>Rhodnius prolixus</i>	-	ND	Yes (WUGSC)	

<sup>a</sup>miRNAs are identified experimentally (Barth et al.) or computationally (Com) or both (Ex/Com);

<sup>b</sup>Argonaute or Dicer or both are present in species indicated. Argonaute homologue(s) in parasitic nematodes, cestodes and arthropods were searched using *C. elegans* Argonaute (ABA18180) in the databases shown in brackets;

<sup>c</sup>Drosha homologue(s) in protozoans were searched using *C. elegans* Drosha (NP\_001122460) and there were no hits with an e-value > 10<sup>-10</sup>. Due to the fact that protozoan Drosha homologues may be highly heterogeneous in an amino acid level, it can't be ruled out the possibility of factor(s) that are functionally similar to Drosha;

'ND': not determined; '-': not applicable; '?': unclear.

in domain(s) during evolution, leading to low similarity and difficulties in identification of putative proteins. For instance, *T. cruzi*, the causative pathogen of Chagas disease, was commonly thought to be unable to utilize siRNAs to alter gene expression (Militello et al., 2008). The presence of an Argonaute/Piwi protein-coding gene has recently been found to be expressed and translated in all life cycle stages (Garcia Silva et al., 2010). Although no potential miRNAs were found in a small RNA subpopulation (Garcia-Silva et al., 2010), it is still hard to rule out the possibility of that this protozoan parasite has ability to produce miRNAs because of the limitations of the methods used.

### **3.5 Possible roles of parasite miRNAs**

An exact spatial and temporal control of gene expression is crucial for animals where the life cycles have several different developmental stages. miRNAs such as the let-7 family and lin-4 were first revealed to be a master switch of development in *C. elegans* (Brehm and Spiliotis, 2008), targeting hb1-7 and lin-14 mRNAs respectively to ensure proper larval developmental transition. The transcription of the let-7 family is controlled by the interaction between the nuclear hormone receptor DAF-12 and its ligands, dafachronic acids that are naturally produced in favourable environments (Bethke et al., 2009). The striking discrepancy of expression of Dicer and Argonaute at different stages in *S. mansoni* is an indication that small RNA-induced silencing represents a mechanism of modulation of developmental transitions (Gomes et al., 2009). The idea that *S. japonicum* let-7 may associate with transition from miracidium to sporocyst is indirectly supported by higher expression of let-7 in the miracidium. Along with let-7, the miRNA bantam that is extraordinarily highly expressed in cercaria is supposed to be active in developmental processes through regulation of a cellular population size (Xue et al., 2008). However, a recent study supports the notion that let-7 has other functions rather than controlling developmental timing (Knapp et al., 2009). It is clear that these presumptions need to be further tested.

Some miRNAs are anticipated to take part in other processes of parasites, such as reproduction and tissue development (Table 1-4). Of interest is the very high expression of miR-71(a/b) in stages of trematodes (Huang et al., 2009; Knapp et al., 2009; Xue et al., 2008) and a cestode (our unpublished data) but not in a nematodes (Huang et al., 2010). Apart from functional connections with neoblast biology (Knapp et al., 2008; Lin et al., 2009; Liu et al., 2009b), miR-71 is expected to have an additional role probably related to development. Further experiments will be worthwhile specifying the connections between its expression and functions.

miRNAs have also been shown to be involved in immune evasion, affecting differentiation, development, homeostasis and functions of immune cells (Hemphill et al., 2002; Spiliotis et al., 2004). It is well studied that miRNAs play important roles in viral infection (Cazalla et al., 2010; Deplazes et al., 2005; Ghosh et al., 2009; Triboulet et al., 2007; Vuitton and Gottstein, 2010). For example, viral miRNAs are expressed during the infection of herpes viruses to target stress-induced natural killer cell ligand, MICB, giving rise to repression of MICB and thus attenuated recognition of infected cells by natural killer cells (Nachmani et al., 2009). The temporal regulation of variant surface proteins (VSGs), one of which covers the entire surface of parasite at any time, is essential to survive host immune attacks in free living protozoans such as Trypanosomes and *G. lamblia* (Eckert and Deplazes, 2004; Mallick et al., 2008). Various VSGs expressed on the surface were observed in individual *G. lamblia* with silenced Dicer and RNA-dependent RNA polymerase, indicating the key role of silencing machinery in VSG expression regulation (Gottstein et al., 2010). miRNAs may be partly responsible for this antigenic variation by targeting 3' VSG-coding gene UTRs that contain highly conserved fragments (Eckert and Deplazes, 2004; Saraiya and Wang, 2008). The mechanism of how miRNA(s) modulate several hundreds of VSG genes to ensure that only one is retained is unknown.

**Table 1-4 Potential functions of parasite miRNAs**

miRNA	Species	Possible functions <sup>a</sup>	Expression <sup>b</sup>	Ref.
let-7 Bantam <sup>c</sup>	<i>S. japonicum</i>	Transition from miracidium to sporocyst? Regulation of cell proliferation and apoptosis/sex development or reproduction	? Cercaria, female	(Knapp et al., 2009; Xue et al., 2008) (Knapp et al., 2009; Xue et al., 2008)
miR-7 miR-36 miR-71 miR-nov-70 <sup>d</sup>	<i>B. xylophilus</i>	Tissue development	Cercaria	(Knapp et al., 2009)
		Developmental transition	Cercaria	
		Sexual development	Male	
		Sexual development/reproduction	Female	
let-7		Regulation of worm activity	Cold-stressed worm	(Huang et al., 2010)
miR-1 miR-nov-10 <sup>d</sup> miR-29 miR-40 miR-72	<i>T. gondii</i>	Response to cold stress	Down-regulated in cold- stressed worm	
miR-4		Related to virulence?	Lethal strain	(Braun et al., 2010)
miR-49 miR-60 miR-40 miR-56			Hypo-virulent strain	
		Antigenic variation	ND	(Mallick et al., 2008)
miR-2		<i>G. lamblia</i>		

<sup>a</sup>None of miRNA functions was experimental validated;

<sup>b</sup>miRNA(s) are highly or exclusively expressed in tissues or strains listed, if not clearly stated;

<sup>c</sup>miR-nov-110 is highly homogenous to bantam and therefore the latter is listed here;

<sup>d</sup>Novel miRNAs;

<sup>e</sup>Many miRNAs are predicted to target VSG mRNA and therefore not listed here;

‘?’: controversial or not clear; ‘ND’: not determined.

In summary, although the function of the vast majority of miRNAs remains an enigma, it is clear that the miRNA-induced silencing machinery directly or indirectly affects many processes of organisms and their responses to the environment. Parasite and host miRNA profiles can be served as a probe to investigate underlying mechanisms and thus deeply understand pathogen-host interplay. Likewise, the miRNA silencing network will be an alternative to help us understand drug resistance in parasitic nematodes (Devaney et al., 2010). The intervention of miRNA pathways to control diseases is in infancy but successful attempts to control viral infection in chimpanzees or tumorigenesis in animal models (Vuitton, 2003; Zhang et al., 2008) have shed light on the potential of miRNA pathways as therapeutic targets.

#### **4. Fatty acid-binding proteins**

From the mRNA transcriptome data, we identified a number of known and hypothetical protein-coding genes that exhibit differential expression patterns. In our project, attention was paid to a subfamily of lipid-binding proteins fatty acid-binding proteins (FABPs) that are essential for lipid metabolism in parasitic platyhelminths which are incapable of *de novo* synthesis of long-chain fatty acids.

Fatty acids are key molecules that serve as energy and as mediators that can affect lipid metabolism and inflammatory reactions via the regulation of gene expression and cellular signalling pathways. Omega 3 fatty acids, for example, are anti-inflammatory, partially acting through the nuclear factor-kappa B transcriptional network to inactivate pro-inflammatory cytokine production (Boutros et al., 2010; Martin and Stapleton, 2010). In addition, fatty acids are thought to contribute to the protection of organs or tissues from infection because a high concentration of fatty acids is detrimental to pathogens (Tchkonia et al., 2010). Dysregulation of fatty acid uptake, trafficking and metabolism appears to be closely related to occurrence of many inflammatory and metabolic disorders including aging, atherosclerosis, obesity and

diabetes. The deficiency of fatty acid metabolism-associated proteins in peroxisomes has been linked to several known human diseases including cancers (Van Veldhoven, 2010).

FABPs are small molecules that are responsible for the transportation of fatty acids. These lipid chaperons are abundantly expressed in majority of tissues and characterized by binding to lipid ligands including saturated or unsaturated long-chain fatty acids with high avidity. Together with retinol- or retinoic acid-binding proteins (RBPs or RABPs) that are likely to have evolved from liver/intestinal FABPs in vertebrates (Schleicher et al., 1995), FABPs constitute a multigene subfamily of intracellular lipid-binding proteins (Haunerland and Spener, 2004). This lipid chaperon subfamily has evolved to give rise to tissue-specific FABPs via successive gene duplication events, which exert functions specifically or complementarily (Schaap et al., 2002; Storch and Thumser, 2010).

#### **4.1 FABPs in the animal kingdom**

FABPs are extensively expressed in animals including invertebrates and vertebrates but not in fungi and plants. It has been shown that an ancestral FABP gene may have diverged from lipocalins and arose from separation of animals from fungi and plants approximately 1,000 to 1,200 million years ago (Ganfornina et al., 2000; Schaap et al., 2002). In yeasts, there are several types of lipid-binding proteins, such as sterol carrier protein 2 (SCP2), which can interact with *cis*-parinaric acid in a nM range (Ferreya et al., 2006). This protein is also found in plants, bacteria, archaea and multicellular organisms but, although they possess similar binding spectra, SCP2s do not belong to the FABP family and are a completely different structure from these of FABPs (Edqvist and Blomqvist, 2006; Ferreya et al., 2006).

In the 1970's a FABP was isolated from rats and found to be extensively expressed in many organs or tissues including adipose tissue (Ockner et al., 1972). Since then, nine

types of FABPs, known as FABP1 to 9, have been characterized in mammals (Furuhashi and Hotamisligil, 2008). Recently, a new FABP gene, FABP12, was described in humans, mouse and rat but not chicken or zebra fish (Liu et al., 2008). FABP12 is clustered with other four FABP genes on the same chromosome and distributed in retina and testis, the latter in which FABP9 is known to be highly expressed.

So far more than thirty different FABPs have been reported in a number of invertebrates over several phyla (Esteves and Ehrlich, 2006). In the Phylum Platyhelminth, FABP genes have been described in the following species: *S. mansoni* (Moser et al., 1991), *S. japonicum* (Becker et al., 1994), *F. hepatica* (Rodriguez-Perez et al., 1992), *F. gigantica* (Chunchob et al., 2010), *E. granulosus* (Esteves et al., 1993; Esteves et al., 2003) and *Mesocostoides vogae* (Alvite et al., 2008).

## **4.2 Gene structure and expression of FABPs**

Vertebrate FABPs have relatively similar gene structures and usually have four exons and three introns with comparable positions, but the gene organization of FABP genes varies considerably in invertebrates (Esteves and Ehrlich, 2006; Schaap et al., 2002). There are, however, a few intronless FABPs that have been reported in mammals such as rats (Bonne et al., 2003), mice (Treuner et al., 1994) and humans (Prinsen et al., 1997). However, there are up to date no reports for intronless FABPs in other species. These mammalian single-exon FABPs appear not to function. In humans, an intronless FABP pseudogene, FABP3-ps, is located on chromosome 13, sharing 85% similarity to heart FABP (also known as FABP3) that resides on chromosome 1. Moreover, it lacks a TATA box in the 5' upstream and its expression is not detected in skeletal muscle and fetal brain (Prinsen et al., 1997). Compared with FABP3, the key residues for interaction with fatty acids are found to be altered, potentially leading to functional loss. To better understand the evolution of FABP genes, it is important to

systematically and extensively analyze the gene structure over a broad range of animals.

There are three conserved motifs present in FABP family members (Furuhashi and Hotamisligil, 2008). Furthermore, there exist a large number of regulatory elements in the upstream and downstream regions of FABP genes, such as a TATA box found in the upstreams in all FABP genes, a hepatic nuclear factor 1 binding site in FABP1, conserved consensus motifs for activator protein 1/2, and mammary active factor in FABP3, for glucocorticoid half site receptor in *E. granulosus* FABP2 (Chmurzynska, 2006; Esteves and Ehrlich, 2006; Haunerland and Spener, 2004). The presence of these elements suggests that the expression of FABPs is tightly regulated, exhibiting different temporal and spatial expression patterns. In *M. vogae*, FABPs are predominantly distributed in the tegument and areas around the calcareous corpuscles (Alvite et al., 2008). Using quantitative PCR, expression of crab FABP is present in most of tissues studied with exception of the stomach and eyestalk. Interestingly, it appears to be down- or up-regulated at different developmental points in ovary and hepatopancreas (Gong et al., 2010).

*Ascaris suum* FABP, known as As-p18, is spliced leader *trans*-spliced and exhibits two different forms with or without a hydrophobic signal. As-p18 expression shows a developmentally regulated pattern: it is absent in unembryonated eggs, commences to express at day 3 of development and reaches at a peak when L1 are formed (Mei et al., 1997). As-p18 protein is also confirmed not to be expressed in L3 and adult tissues but Northern blot results reveal that an appreciable amount of As-p18 mRNA is detected in the ovaries and L3, especially the latter, suggesting post-transcriptional regulation of protein expression (Mei et al., 1997).

### 4.3 Structural and binding features of FABPs

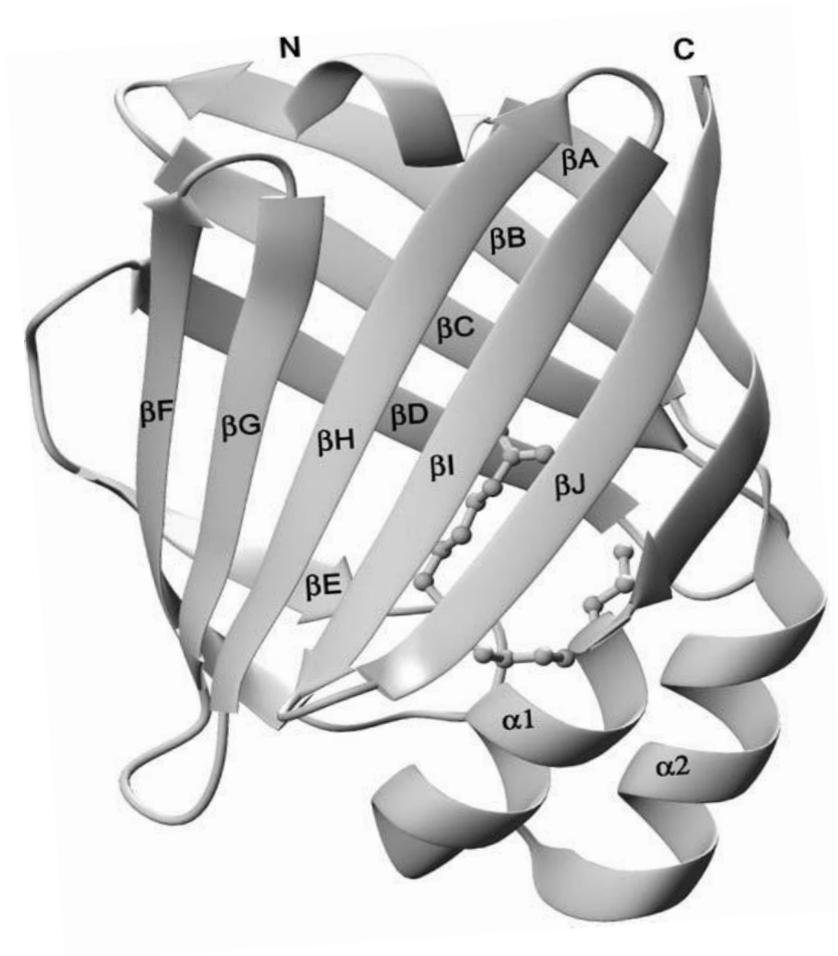
The crystal structures of several FABPs have been resolved. Although they have modest homologies, all members of the multigene family of FABPs share conserved secondary structures that are characterized by an elliptical  $\beta$ -barrel, which is comprised of ten  $\beta$ -strands and two  $\alpha$ -helixes (Storch and McDermott, 2009). The two  $\alpha$ -helixes have no interactions with  $\beta$ -barrel, together with the turns between  $\beta$ C/ $\beta$ D and  $\beta$ E/ $\beta$ F, shapes a portal that allows fatty acid(s) entry and egress (Fig. 1-5) (Jakobsson et al., 2003; Storch and McDermott, 2009). The  $\beta$ -barrel contains a binding pocket, the size of which is remarkably larger than bound lipid(s). FABPs interact with the carboxylate terminus of bound fatty acid(s) and the binding affinity relies on non-covalent molecular forces between fatty acid and several essential residues such as Arg<sup>127</sup>, Tyr<sup>129</sup> and Ala<sup>76</sup> (Jakobsson et al., 2003). Under fatty acid-free and -binding circumstances, FABPs have substantially similar structures but the conformation of the helix-turn-helix portal moieties is slightly changed. Compared with that of FABP with a bound fatty acid, the portal structure of the fatty acid-free FABP is more open relative to the  $\beta$ -barrel, favouring the model that hypothesizes diffusion-controlled trafficking of hydrophobic molecules (He et al., 2007).

FABPs are featured by binding to unsaturated or saturated long-chain fatty acids with high affinity in the nM range. Although it is controversial, FABPs do not exhibit binding to fatty acids in a selective way and the binding affinity is closely related to ligand hydrophobicity (Angelucci et al., 2004; Richieri et al., 2000). It is of interest to understand how multiple FABPs with similar spectra function in a single organism. It has been proposed that the functions of distinct FABPs may largely depend on the structural properties of the FABP's surfaces but not selectivity of binding to fatty acids (Storch and McDermott, 2009). It has been shown that nearly almost all of FABPs reported bind to one fatty acid molecule except FABP1. FABP1 offers two

binding sites for fatty acids and this is consistent with the fact that the FABP1 binding cavity is noticeably larger in comparison with that of others (Furuhashi and Hotamisligil, 2008).

#### **4.4 Biological roles of FABPs**

FABPs are proposed to be involved in numerous processes, including energy storage, signalling, membrane synthesis, oxidation in mitochondria or peroxisomes and the regulation of enzyme activities. Although the exact roles and mechanisms and pathways with regard to lipid metabolism are still enigmatic, recent studies on FABP-deficient mice have demonstrated that not all individual FABPs are vital and there is some redundancy (Furuhashi and Hotamisligil, 2008; Storch and Corsico, 2008). For example, FABP3 (heart FABP), abundantly distributed in myocardium, has been experimentally demonstrated to be responsible for fatty acid uptake and oxidation in heart and skeletal muscle (Binas et al., 1999; Schaap et al., 1999). In the FABP3-knockout mice, the heart used glucose as energy source instead of long-chain fatty acids, and the fatty acid oxidation in the liver was enhanced (Binas et al., 1999). Under some conditions such as low temperature, FABP3 was up-regulated in rat and mouse brown adipose tissue that functions in thermogenesis in rodents. FABP3<sup>-/-</sup> mice had significantly lower body temperatures and brown adipose tissue, compared to wild-type, suggesting FABP3 roles in resistance to cold. Moreover, a significantly reduced mass of brown adipose tissue and enhanced glycolysis was observed in FABP3<sup>-/-</sup> mice. These observations suggest that FABP3 plays a key role in fatty acid oxidation (Vergnes et al., 2011). FABP3 is also distributed in the neurons of the brain and involved in arachidonic acid trafficking. It can interact *in vitro* with the dopamine D2 receptor long isoform and co-locates with dopamine D2 receptor in the dorsal striatum in the brain. In FABP3-deficient mice, the dysfunction of the dopamine D2 receptor in the brain occurred, leading to abnormal release of acetylcholine



**Figure 1-5 The crystal structure of *E. granulosus* FABP1.**

Numbered ten  $\beta$ -sheets (from A to J) and two  $\alpha$  helices ( $\alpha 1$  and  $\alpha 2$ ) are noted.

Similarly, one bound fatty acid is shown in the binding pocket of EgFABP1. This figure was modified from (Jakobsson et al., 2003).

and depolarization. These observations demonstrate that FABP3 plays a regulatory role in the normal functions of the dopamine D2 receptor (Shioda et al., 2010).

#### **4.4.1 Immunological roles of adipocyte FABP**

Our understanding of the relationship between immune responses and the functions of FABPs are largely derived from studies on FABP4, known as adipocyte FABP or P2. FABP4 is tightly modulated during the process of adipocyte differentiation and its expression can be influenced by factors such as fatty acids and insulin (Furuhashi and Hotamisligil, 2008). Interestingly, early studies have shown that FABP4 mRNA levels are increased in human monocytes upon activation with peroxisome proliferator-activated receptor- $\gamma$  (PPAR- $\gamma$ ) activators *in vitro*, parallel to the activation of PPAR- $\gamma$  (Pelton et al., 1999). Similarly, an increase in FABP4 was also observed in human macrophage cells treated by oxidized low density lipoprotein, which was suppressed by inhibitors for nuclear factor- $\kappa$  B and protein kinase C, suggesting that these signalling pathways are associated with transcriptional regulation of lipoprotein-induced FABP4 (Fu et al., 2000). In FABP4-deficient macrophages, expression of several cytokines including TNF- $\alpha$ , IL-1 $\beta$  and IL-6 was abolished in comparison with wild type cells. In addition, intracellular cholesterol ester accumulation was also severely impaired (Makowski et al., 2001). These results suggest that there are potential connections between FABP4 and cytokine production by macrophages. Likewise, FABP4 also appears to be closely associated with cytokine secretion by dendritic cells and the impairment of IL-12 and TNF secretion occurred in FABP4-deficient dendritic cells stimulated with LPS (Rolph et al., 2006).

FABP4 is detected in human airway epithelial cells and its expression is up-regulated upon addition of IL-4 or IL-13 but downregulated by IFN- $\gamma$ , which depends on the activity of the transcription factor STAT6 that participates in transcriptional regulation of genes in allergic responses. In the allergic airway inflammation model, FABP4 was

significantly expressed in bronchial epithelium and cell populations, especially eosinophils, were pronouncedly increased in the bronchoalveolar lavage fluid in the wild type mice. However, increased cell populations were not observed in FABP4<sup>-/-</sup> mice. Consistent with an attenuated infiltration of airway eosinophils, the levels of cytokines IL-5 and IL-13 in the bronchoalveolar lavage fluid were considerably lower than that in FABP4<sup>+/+</sup> mice. These findings illustrate that FABP4 may be functional in airway inflammation and bridge a connection between lipid metabolism and asthma (Shum et al., 2006).

The FABP4-cytokine network has been further investigated. In LPS-induced macrophages, the deficiency of FABP4 causes reduction of I $\kappa$ B kinase and NF- $\kappa$ B activities without alteration of their protein levels (Makowski et al., 2005). This finding implies that FABP4 is likely to influence macrophage cytokine secretion via regulation of the I $\kappa$ B-NF $\kappa$ B pathway, which is well known to be crucial to modulation of pro-inflammatory activity of macrophages. Recently, it has been shown that LPS activation can also lead to an elevated level of FABP4 in macrophages, which could be initiated by the binding of activator protein-1 to the promoter region of the FABP4 gene. Intriguingly, downregulation of macrophage FABP4 induced a substantial decrease of phosphorylated c-Jun NH2-terminus kinase, activator protein-1 activity and TNF- $\alpha$  and IL-6 synthesis (Hui et al., 2010). This evidence suggests a positive feedback network, whereby FABP4 is involved in the augmentation of macrophage cytokine responses to LPS stimulation.

Several lines of evidence have shown that lipid-binding proteins are allergens. In the 1990's, an allergic hydrophobic molecule-binding protein, ABA-1, was firstly described in swine roundworm *A. suum* (Spence et al., 1993). It is abundantly present in the pseudocoelomic fluid in *Ascaris* species and possesses ability to bind with a number of fatty acids and retinol (Xia et al., 2000). Such allergens with capacity of fatty acid binding are present in many nematodes (Kennedy, 2000a, b). Nematode

polyprotein allergens/antigens (NPAs) belong to tandemly repetitive polyproteins and are comprised of many similar or identical repeated segments. The polyproteins are post-translationally cleaved to generate numerous small proteins with a molecular mass weight of approximately 15 kDa, which have been found to bind fatty acids or retinol with high affinity (Kennedy, 2000a) and may take part in lipid trafficking and distribution (Meenan et al., 2011). Structurally, like *O. volvulus* Ov-FAR-1 (Kennedy et al., 1997), NPAs are enriched with  $\alpha$ -helix and show four-helix bundles that accommodate bound ligands (Kennedy et al., 1995; Meenan et al., 2011; Moore et al., 1999), which is completely distinct from the canonically structural elements of FABPs.

Up to now, FABPs have been reported to be allergens in the following species: *Blomia tropicalis* (Caraballo et al., 1997; Puerta et al., 1999), *Acarus siro* (Eriksson et al., 1999), *Lepidoglyphus destructor* (Eriksson et al., 2001), *Dermatophagoides farinae* (Chan et al., 2006) and *Tyrophagus putrescentiae* (Jeong et al., 2005). *B. tropicalis* FABP shares 42.3% identity with *S. mansoni* FABP and 36% identity with mammalian FABPs and it can selectively bind to *cis*-parinaric acid and oleic acid but not retinol and others tested (Caraballo et al., 1997; Puerta et al., 1999). Structural analysis indicates that this lipid-binding protein contains high  $\beta$ -sheet content with 13%  $\alpha$ -helix, supporting the anti-parallel 10-strand  $\beta$ -barrel typical of FABP structures (Puerta et al., 1999). It has also been shown that IgE of approximately 11% of allergic sera binds to this allergen (Caraballo et al., 1997) and this binding capacity largely depends on several surface charged residues of FABPs (Chan et al., 2006).

#### **4.4.2 Other roles of FABPs**

Phylogenetically, invertebrate FABPs are much closer to the mammalian FABP3 cluster than the FABP2 cluster (known as intestinal FABP) (Esteves and Ehrlich, 2006; Esteves et al., 1997). Up to now, few invertebrate FABPs have been functionally investigated *in vivo* thus their exact functions are largely unknown. Many

biological actions have been proposed for FABPs in invertebrates. In *A. suum*, FABP expression is restricted to early development in eggs, consistent with the increasing use of triglycerides as energy during developmental periods. Therefore FABP is postulated to protect parasites from toxicity derived from fatty acids and their peroxidation products. It is also plausible that this protein may be associated with permeability changes of the eggshell lipid layer upon hatching (Mei et al., 1997). As they are unable to synthesize *de novo* fatty acids, parasitic helminths are thought to completely rely on lipid resources from hosts where they reside and FABPs, at least partially, function in fatty acid transport. Using fluorescence resonance energy transfer assay, the trafficking mechanisms of different lipid-binding proteins including trematode FABP and the nematode polyprotein ABA-1A1 and Ov-FARs, one of which the crystal structure was recently resolved (Jordanova et al., 2009), were investigated. Although all of them directly interacted with membranes during transport, these proteins exhibited different properties with regard to transfer rates and mechanisms (McDermott et al., 2002). An analysis of the crystal structure reveals that *S. mansoni* FABP has evolved to selectively bind unsaturated fatty acids, especially arachidonic acid. This protein showed higher affinity with fatty acids at acidic pH than neutral or basic pH and therefore it is considered to transport lipids from the contact interface between parasite and host to parasites because the microenvironment around the membrane is more acidic than the cytoplasm (Angelucci et al., 2004). In the locust, *S. gregaria*, FABP was also found in the nuclei (Haunerland et al., 1993). Although the underlying mechanism remains unclear, its distribution is in agreement with the concept that FABPs are likely to be involved in fatty acid-induced transcriptional regulation in the nucleus (Furuhashi and Hotamisligil, 2008).

In conclusion it seems apparent that the gene structures of vertebrate FABP genes are more conserved than that of invertebrate FABP genes. Moreover, gene duplication followed by divergence is most likely to contribute to variety of FABP functions in

mammals. But whether or not the FABP copy numbers are also responsible for multiple FABP functions in invertebrates remains largely unknown and the diversity of genomic organizations of invertebrate FABP genes are not extensively explored. To address these questions, it is clear that more sampling of invertebrates is required to phylogenetically analyze FABP genes.

Albeit low sequence homology, FABP genes show a conserved tertiary structure with similar lipid-binding spectra. It has emerged that FABP functions are various and mammalian and mite FABPs are closely linked with inflammation. Interestingly, trematode and nematode FABPs are present in excretory/secretory products but the roles of these secretory proteins are not understood. Whether or not FABPs are involved in infection or immunology during the infection remains to be experimentally investigated.

## **5. Summary**

The identification and understanding of neoblasts have recently led to establishment of the *in vitro* cultivation system (Spiliotis et al., 2008), whereby the primary cells from vesicles can grow into mature larval parasites with infectivity. Equipped with the completion of the genome, *E. multilocularis* has been recently developed as an animal model to study other platyhelminths. However the limited knowledge on *E. multilocularis* biology including baseline gene and non-coding regulatory RNA expression impedes its efficient and extensive applications. Using next-generation sequencing technology, our project was initiated to explore miRNA and mRNA transcriptomes for the investigation of miRNA expression patterns and the identification of developmentally regulated molecules, which are able to be used as stage-specific biomarkers. As *E. multilocularis* lacks the biogenesis components for *de novo* synthesis of long-chain fatty acids and lipid metabolism is essential in the parasitic platyhelminths in terms of physiology (Smyth and McManus, 1989), one

family of differentially expressed molecules, fatty acid binding proteins (FABPs) that bind to fatty acid(s) with high affinity, was further experimentally analysed.

## **Chapter 2 A robust strategy of depletion of polycistronic mitochondrial transcripts in *Echinococcus multilocularis* for RNA-seq**

### **Abstract**

In *Echinococcus multilocularis*, polyadenylated mitochondrial RNA is very abundant, and it is therefore important to remove such transcripts to increase the depth of coverage when sequencing the mRNA transcriptome. Here, we conceived a practical strategy using ribonuclease H and 5'-phosphate dependent exonuclease in combination to efficiently purify mRNA for RNA-seq using the Illumina sequencing platform. In total, we obtained 28,829,936 reads for the control sample which had only been treated with 5'-phosphate dependent exonuclease and 43,166,442 for the sample after treatment of 5'-phosphate dependent exonuclease and ribonuclease H. Mapping results showed a significant decrease of mitochondrial RNA contamination from 2.68% in the control to 0.09% in the treated sample. Moreover, the proportion of mitochondrial reads from transcriptome data obtained for *E. multilocularis* protoscoleces, cultured primary cells and mature metacestode vesicles ranged from 0.04% to 0.09%, respectively, indicating its reproducibility. This strategy is expected to be of general use to deplete polyadenylated message for mRNA transcriptome analysis.

### **1. Introduction**

In the Phylum Platyhelminth the ~360 Mb genomes of *S. japonicum* and *S. mansoni* have been recently published and contain 13,000 and 12,000 predicted protein-coding genes, respectively (Berriman et al., 2009; Zhou et al., 2009). The *E. multilocularis*

genome is smaller at 106 Mb and its draft sequence is now entering completion (Brehm, 2010a). Considering the complexity of stage-specific gene expression in the parasite, there is an urgent need for baseline transcriptome data to provide a detailed view of genes expressed through the complex life cycle.

Polyadenylated mitochondrial transcripts are abundant in both *E. granulosus* (Fernandez et al., 2002) and *E. multilocularis* (Brehm, 2010a). A previous study on *E. granulosus* showed that nearly 70% of randomly sequenced clones in a cDNA library constructed using the poly T priming were mitochondrial transcripts. Of them, 63% (156/248) were either large or small ribosomal subunit-coding fragments, likely due to the existence of a transcriptional termination signal downstream of the large rRNA which was also found in *E. multilocularis* (Nakao et al., 2002). Moreover, it was found that the 3' terminal of the mitochondrial transcripts investigated was also polyadenylated (Fernandez et al., 2002). It is therefore unlikely to be able to separate nuclear mRNA without contamination of mitochondrial RNA using poly (A)-based systems.

RNA-seq is an approach for global investigation of transcriptional patterns using second-generation sequencing technologies to produce single-base resolution transcription maps in a robust, reproductive, deep and quantitative manner (Marioni et al., 2008; t Hoen et al., 2008). In order to increase the depth of mRNA coverage, elimination of the super-abundant *E. multilocularis* mitochondrial RNA transcripts is required. Biotinylated probes combined with protein A or G-coated magnetic beads have previously been used to remove ribosomal or other RNA transcripts in mRNA transcriptional profiling studies (Lister et al., 2009; Raghavachari et al., 2009). In this study, we developed a practical and economical way to prepare mRNA for expression studies, through specific digestion of mitochondrial RNA with enzymes RNase H and 5'-dependent exonuclease. The protocol can be used for any transcriptomic

investigation, in which mRNA signals are masked by the presence of other super-abundant polyadenylated RNA species.

## **2. Materials and methods**

### **2.1 Parasites and primary cells**

The *E. multilocularis* isolates H95 and JAVA were used in this study (Jura et al., 1996; Tappe et al., 2007). In both cases, larval material was propagated and continuously kept in mongolian jirds (*Meriones unguiculatus*) as previously described (Spiliotis and Brehm, 2009). Mature metacystode vesicles (mm) of isolate H95 were cultivated under axenic conditions (Spiliotis et al., 2004) and were used to set up primary parasite cell cultures (agg) essentially as described previously (Spiliotis et al., 2008). Dormant protoscoleces (psano) were isolated from *in vivo* cultivated parasite material of isolate JAVA (Brehm et al., 2003).

### **2.2 Extraction of total RNA**

Total RNA was extracted with modified protocols using Trizol (Invitrogen) and RNeasy kit (Qiagen). Briefly, parasites were homogenized in 1 ml Trizol on ice and after addition of chloroform and centrifugation, 225 µl of absolute ethanol were added drop by drop per 445 µl of the supernatant containing RNA. Then the mixture was transferred onto the columns (Qiagen) and the flow-through was reloaded onto the columns, centrifuged, and washed. Total RNA was eluted into nuclease-free water. To get rid of contaminating genomic DNA, the recovered RNA was incubated with DNase I (Ambion), precipitated, and suspended into RNase-free water. The purity and concentration of total RNA were determined by a Bioanalyzer (Agilent 2100).

## **2.3 Depletion of mitochondrial transcripts and *in vitro* amplification of mRNA**

Four 26bp specific probes absolutely complementary to the mitochondrial genes *cox3*, *cytb*, *lrRNA* and *srRNA*, respectively, were designed by OligoRankPick software (Hu et al., 2007) (Table 2-5). 25 µg of the suspended RNA were gently mixed with the probes to final concentration of 4 µM, incubated for 10 min at 65°C and then cooled slowly to room temperature for 10 min. 2.5 µl ribonuclease H (RNase H) was then added, heated for 30 min at 37°C, precipitated and resolved into 44 µl nuclease-free water followed by incubation with DNase I to decontaminate the DNA oligonucleotides and precipitation. The resolved RNA was incubated with 2 µl of 5'-phosphate dependent exonuclease (Epicenter) for 60 min at 30°C. After precipitation, the treated product was re-suspended into nuclease-free water and its quality and concentration were determined by Nanodrop (Thermo Scientific).

In order to get sufficient mRNA for RNA-seq, amplification of mRNA was conducted *in vitro* with T7 RNA polymerase (Illumina) according to the instructions with a few modifications. 500 ng of the above RNA were utilized as starting material. Reverse transcription was performed with T7 Oligo (dT) primer for 2 h at 42°C in hybridization oven and ds cDNA was synthesized for 2 h at 16°C in a PCR cycler. After purification, 16 µl of ds cDNA were mixed with 3 µl T7 10x Reaction Buffer, 3 µl T7 Enzyme Mix and 8 µl rNTP (25 mM each) and then incubated for 6 h at 37°C in hybridization oven. In the end, amplified complementary RNA (cRNA) was purified and precipitated.

**Table 2-5 Primers and probes used in this study.**

Probe/Primer	Targeting gene	Sequence (5'-3')	Length	Function
lrnaF	lrRNA	GAATATTTGGCATTTGATTGAAATTG	26 bp	PCR validation; lrnaR used as a probe
lrnaR <sup>a, b</sup>	lrRNA	CCAAAAATCTCAACTACTACACCAAG	26 bp	
lrnaP <sup>c</sup>	tRNA	ATCACAACAACCCTAATCAGCTTTTG	26 bp	PCR identification of the polycistronic RNA
coxR	cox1	CAAACCAGTAATCAACGGTCACCATC	26 bp	PCR validation
coxFb	cox1	GGGAGTAGTGTGGGGTCATCATAT	26 bp	
actinF	actin II	GAGAAGATGACACAAATCATGTTTG	25 bp	PCR validation
actinR	actin II	GCTATTTTCGCGCTCTGCTGTAGTAG	25 bp	
actinFg	actin II	TCGTCCAAGACATCAGGATAGTT	23 bp	Determination of genomic DNA
actinRg	actin II	AAAGATTATTTTCGGAATAGGTACGTTG	27 bp	
P1 <sup>a</sup>	cox3	TAAGCCAACAAAAGAAGCACCAAACA	26 bp	Removal of the abundant mitochondrial RNA
P2 <sup>a</sup>	Cytb	ATACACCGAAGAATAGCATAAAAAGGC	26 bp	
P3 <sup>a</sup>	srRNA	GTATCTAATCCCTGTCACCACATATA	26 bp	

<sup>a</sup>these probes used for depletion of the mitochondrial transcripts have the same GC content (38.4%);

<sup>b</sup>the primers were used to identify the polycistronic RNA;

<sup>c</sup>this tRNA is located directly downstream the lrRNA-coding region. Coupled with the primer lrnaF or coxF, it was also used to amplify the polycistronic RNA.

## 2.4 Preparation of sequencing libraries

In this study we constructed sequencing libraries by two different methods: PCR-free and PCR-enriched (standard Illumina protocol). If biological material is sufficient, the former will be a priority for sequencing library construction because the PCR-enriched method may introduce artefacts after several rounds of PCR amplifications (Kozarewa et al., 2009; Schadt et al., 2010). For PCR-free libraries (Kozarewa et al., 2009), 5µg cRNA was gently mixed with 1 µl random primer (3 µg/µl, Invitrogen) and denatured for 5min at 65°C. First strand cDNA was generated by addition of 5µl Super II reverse transcriptase (Invitrogen) under the following condition: 25°C for 10min, 42°C for 50min and 70°C for 15min, followed by ds cDNA synthesis for 2h at 16°C in the presence of 1µl DNA ligase, 4 µl DNA polymerase and 1 µl RNase H in a total volume of 150 µl. 1µl T4 DNA polymerase (Promega) was added and then incubated for 10 min at 16°C. ds cDNA was purified using QIAquick PCR Kit (Qiagen) and physically fragmented. Prior to analysis, the quality and quantity of each sample were assessed by Nanodrop and PCR. Amplifications of Actin1 (control), LrRNA and Cox1 were performed using Fast Cycling PCR Kit (Qiagen) under the recommended conditions: 95°C for 5min, 20 cycle of 96°C for 5s, 55°C for 5s and 68°C for 15s, and a final extension at 72°C for 1min. End-repair, adenine addition and adaptor ligation were conducted according to the standard Illumina protocol for RNA-seq. The final product was separated on an agarose gel and fragments with a size of 300bp to 500bp were selected, eluted and analyzed by Bioanalyzer (Agilent) and 37bp pair-end sequences were produced using Genome Analyzer (Illumina).

In contrast, standard Illumina method involves a step of PCR amplification to enrich the fragments selected. 1.5µl cRNA was fragmented chemically not physically as in preparation of the PCR-free libraries. For first strand synthesis, 1.5µl of Super II reverse transcriptase was used. Other procedures were conducted according to the

standard II recommended RNA-seq protocol (<http://www.illumina.com/applications/sequencing/rna.ilmn>). Herein, the fragments ranged from 200bp to 300bp were recovered and enriched by PCR amplification and sequenced.

## **2.5 Sequence data quality control**

37 or 54bp paired-end reads were produced for each sample. Data from each Illumina flowcell lane was assessed for quality based on GC content, average base quality, Illumina adapter contamination and mitochondrial contamination. To assess the quality of the lane, the mean base quality at each base position in the read was computed over all reads from the lane. To assess GC content of the reads the GC percent of each read was calculated and a frequency distribution of the values was plotted. If there is only a single sample in the lane, then we expect the GC plot to be a normal distribution around the expected proportion of GC for the organism.

To screen for adapter contamination, the reads were aligned to the set of Illumina adapter sequences using BLAT v.34 with default parameters (Kent, 2002). Similarly, the reads were screened for mitochondrial contamination by aligning to the mitochondrial sequence of *E. multilocularis* (Nakao et al., 2002). Any reads matching these sequences were reported as being contaminated with either adapter or mitochondrial sequence.

## **2.6 Mapping and statistical analysis**

Reads from each life-cycle stage were mapped to the draft assembly using MAQ v0.7.1 (Li et al., 2008). The default mapping parameters for reads were used, but with the *e* parameter raised to 80 and 120 for the 37 bp and 54bp reads respectively. This was to allow for a larger number of mismatches in read alignments of longer length. All alignments produced by MAQ were converted to BAM format (Berriman et al., 2009).

Pearson correlation was used to analyze the mean coverage of two PCR-free samples. The mean coverage is calculated by the following formula: mean coverage = total length of reads mapped/ length of supercontig, where the total length of reads mapped is generated by length of read multiplied by the number of reads mapped onto the supercontig. To compensate for differences in sequence outputs between different sequencing runs, the dataset for exonuclease-treated metacestode library (pf-mm) was normalized against that for exonuclease and RNase H-treated metacestode library (dpf-mm) as follows: normalized mean coverage = mean coverage of the pf-mm multiplied by coverage ratio, where the coverage ratio (mapped reads in dpf-mm/mapped reads in pf-mm)=  $37,015,693/24,165,997 = 1.53$ . After calibration, difference of the mean coverage of the first 120 supercontigs, which are covered by two data sets, was determined by using paired, two-tailed *Student's t*-test.

### **3. Results and discussion**

#### **3.1 Rationale of the strategy for removal of *E. multilocularis* mitochondrial transcripts**

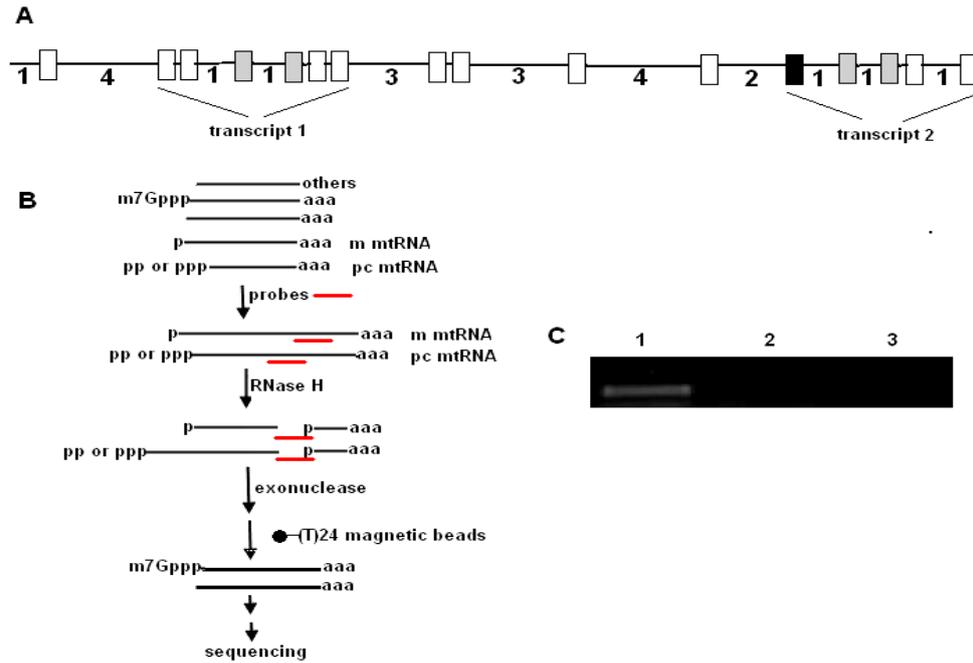
The *E. multilocularis* mitochondrial genome is a double-stranded closed circle DNA with 13,738 base pairs, encoding 12 proteins, 2 rRNAs and 22 tRNAs that are organized compactly with no introns and few, small intergenic spacers (Nakao et al., 2002). The mitochondrial genome is transcribed from few promoters in a polycistronic manner in some organisms, yielding several long primary transcripts with a poly A stretch at the 3' end and a 5' bi- or tri-phosphate terminus which are further processed to form short multiple gene-containing RNA, tRNA, and mature mono- or bi-cistronic species (Barth et al., 2001; Gissi and Pesole, 2003; Montoya et al., 2006; Yoza and Bogenhagen, 1984). Based on the tRNA punctuation model (Montoya et al., 2006), it is supposed that transcription of the *E. multilocularis* mitochondrial genome gives rise to two main potential polycistronic transcripts

(transcripts or RNA 1 and 2, Fig. 2-6 A). We used one probe against the large rRNA (LrRNA) to deplete RNA 2, which includes the LrRNA-coding gene, and PCR results showed that the majority of the LrRNA-containing transcripts were eliminated compared to the control (data not shown). To enforce the efficacy of depletion, an additional probe complementary to small rRNA, which is also abundant in *E. granulosus* (Fernandez et al., 2002), was used. This approach was employed to digest RNA 1 using the probes against Cox3 and Cytb, which are relatively highly expressed in other organisms (Gissi and Pesole, 2003).

In this approach, two restriction treatments with RNase H and 5'-phosphate dependent exonuclease were used. The endonuclease RNase H specifically digests RNA within RNA-DNA hybrids, and generate a monophosphate group at the 5' end of a cleavage site (Donis-Keller, 1979). It has previously been applied for second strand cDNA synthesis, removal of polyadenylated tails of mRNA and investigation of molecules with complicate interior structures (Wilusz et al., 2008). After RNase H treatment, truncated mitochondrial RNA fragments with a 3' poly A tail are produced and then specifically hydrolyzed by a 5'-monophosphate dependent exonuclease. Finally, mRNA can be purified using oligo d(T)-based technologies, such as oligo d(T) capture beads (Fig. 2-6 B).

### **3.2 Validation of efficiency of mitochondrial RNA elimination**

Using treated mRNA purified with oligo d(T)-labelled magnetic beads as template, complete depletion of LrRNA and Cox1, which compose the portion of RNA 2, was ascertained by reverse-transcription PCR (Fig. 2-6 C). In this study, no probe against Cox1 was utilized and the amplification for Cox1 was negative, indicating efficacy of the approach to deplete the mitochondrial polycistronic and mature RNA and indirectly demonstrating a 5' monophosphate terminus of the mitochondrial mature transcripts of *E. multilocularis*. Electrophoresis results of each sample for sequencing showed little or no contamination of the mitochondrial transcripts.



**Figure 2-6 Depletion of *E. multilocularis* mitochondrial transcripts.**

(A) Schematic representation of the mitochondrial genome. Protein-coding genes, rRNA and two non-coding regions (boxes) are spaced along the mitochondrial DNA based on the number of tRNA among them, which is shown underneath the line. Two potential long polycistronic transcripts are indicated. Four probes (Table 2-5) were used to target Cox3, Cytb, LrRNA and SrRNA (grey filled boxes, from left to right), respectively, of which every two are situated in the different transcripts 1 and 2. PCR amplification for Cox1 (black filled box) was conducted to assess efficiency of decontamination of the mitochondrial RNA. (B) Workflow of the strategy. During hybridization, the mature (m mtRNA) and polycistronic mitochondrial RNA (pc mtRNA) are specifically bound by the probes (short red bars). After treatments with RNase H and exonuclease, respectively, mRNA is purified using oligo dT-labelled magnetic beads. It is worth noting that RNA with a poly A stretch is not shown from hybridization to exonuclease digestion to simplify the figure. (C) PCR validation of efficacy of the approach. Using purified mRNA, amplification for Actin (lane 1), Cox1 (lane 2) and LrRNA (lane 3) was performed by reverse transcription-PCR using specific primer sets (Table 2-5).

### 3.3 Mapping reads to the *E. multilocularis* genome

In this study, two PCR-free sequencing libraries were constructed. One was for the sample of mature metacystode vesicles with double treatments of RNase H and exonuclease (dpf-mm) and the other only with exonuclease treatment (pf-mm) was used as the control. The number of 37bp sequences for pf-mm and dpf-mm was 28,829,936 and 43,166,442, respectively, and the majority of the reads from both libraries (>83%) were successfully mapped to the genome of *E. multilocularis*. Moreover, 461 and 472 of the 644 supercontigs of the current assembly version, respectively, were covered by the data set, accounting for 19.26% and 18.40% of the genomic sequence (Table 2-6).

All three standard Illumina libraries were prepared based on double restriction digestion of RNase H and exonuclease (Table 2-6). Except in the case of the non-activated protoscoleces library (ds-psno), more than 34 million 54bp reads were obtained for both mature metacystode vesicles (ds-mm) and cultured primary cells (ds-agg). In contrast to the two PCR-free samples, there was a lower proportion of the sequences for these Illumina libraries to align to the genome, from 62.61% in the ds-agg to 72.28% in the ds-psno. As for the ds-agg, however, much more genomic DNA was covered, reaching up to 25.32%. Moreover, 87.42% of the supercontigs, obviously higher than the rest including the PCR-free data sets, were covered in the ds-agg. These results indicate that more genomic regions are transcribed in *E. multilocularis* primary cell aggregates than in the mature metacystode vesicles or in protoscoleces.

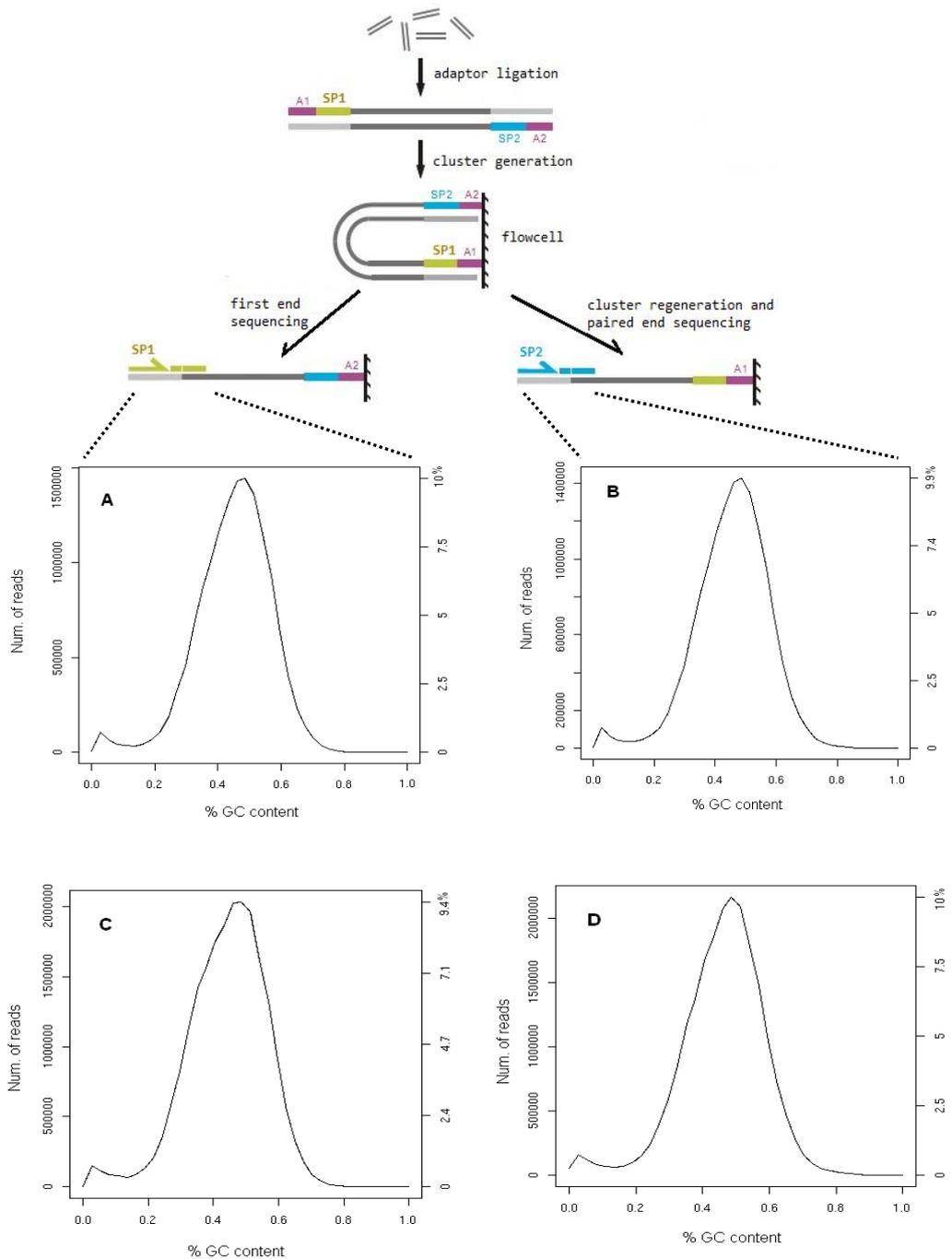
**Table 2-6 Mapping results of different *E. multilocularis* samples prepared by different methods**

Method/samples	Total reads	Read length	Reads mapped	Mt RNA contamination	Num. of bases covered	Supercontigs covered
<b>PCR-free</b>						
pf-mm	28,829,936	37	24,165,997 (83.8%)	771,902 (2.7%)	20,520,731 (19.3%)	71.6% (461/644)
dpf- mm	43,166,442	37	37,015,693 (85.8%)	37,191 (0.1%)	19,604,455 (18.4%)	73.3% (472/644)
<b>Standard</b>						
ds-mm	34,808,454	54	22,460,249 (64.5%)	30,957 (0.1%)	16,282,042 (15.3%)	69.9% (450/644)
ds-agg	34,039,470	54	21,311,351 (62.6%)	28,215 (0.1%)	26,983,637 (25.3%)	87.4% (563/644)
ds-psno	18,693,774	54	13,510,969 (72.3%)	7,227 (0.0%)	21,362,238 (20.0%)	70.3% (453/644)

### 3.4 Efficacy of the depletion of the mitochondrial RNA species

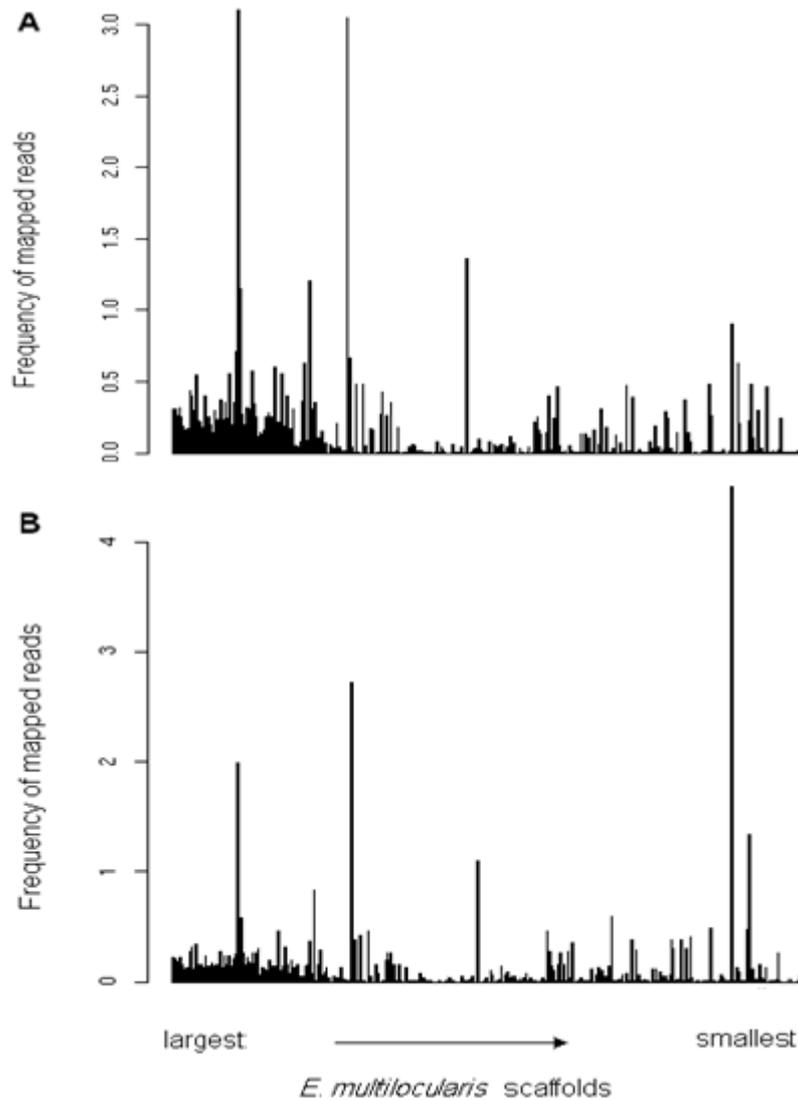
To determine the level of contamination, all reads from the five samples were aligned to the mitochondrial genome of *E. multilocularis*. In the PCR-free libraries, the proportion of mitochondrial sequences decreased significantly from 2.68% in the control pf-mm to 0.09% in the dpf-mm (Table 2-6), suggesting efficient elimination of the mitochondrial RNA using the approach developed in this study. As shown in the Table 2-6, the contamination rate was still low in the three Illumina data sets, ranging from 0.04% to 0.09%. Together with the above results, the strategy of combined use of RNase H and exonuclease is utilized to decontaminate the mitochondrial RNA in an easy, efficient and reproducible manner. In this strategy, which involves several rounds of precipitation and incubation, the loss of a certain amount of mRNA of interest is inevitable. Ideally, this issue can be circumvented by performance of the restriction treatments at the same time in one tube under optimal conditions.

As expected, the overall GC content of the paired-end reads for the dpf-mm was very similar to that for the pf-mm (Fig. 2-7). The same GC pattern was also discerned in the Illumina sequences except for the ds-agg sample that might have adaptor contamination. With a few exceptions, the frequency of mapped reads per supercontig was comparable between pf-mm and dpf-mm (Fig. 2-8). For the Illumina data sets, however, there were noticeable differences of the read frequency per supercontig, especially for the ds-agg. There were 461, 472 and 438 supercontigs covered by respective reads of the pf-mm and dpf-mm and by both.



**Figure 2-7 Distribution of the paired-end reads of the PCR-free library with exonuclease treatment (A and B) or with RNase H and exonuclease double treatments (C and D).**

The bottom scale is for percentage of the GC content. The vertical scale on the left stands for the number of reads with a certain GC content, whereas the one on the right for the proportion of the reads. A and C: first end sequences; B and D: paired-end sequences.



**Figure 2-8 Coverage of individual scaffold for two PCR-free sequencing libraries with exonuclease treatment (A) and with RNase H and exonuclease double treatments (B).**

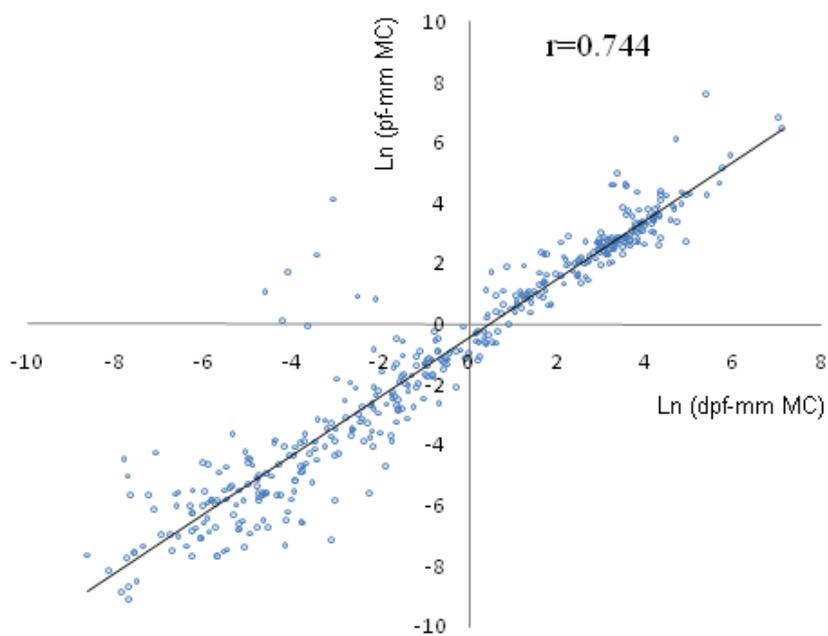
The vertical scale represents the frequency of mapped reads along *E. multilocularis* scaffolds from largest to smallest (the horizontal axis) and it is calculated by the number of reads mapped to a scaffold divided by the length of the scaffold.

**Table 2-7 First 120 supercontigs commonly covered by two PCR-free data sets**

Sample	Length of the supercontigs	Read mapped <sup>a</sup>	Num. of nucleotides covered <sup>b</sup>
pf-mm	94.99%	97.81%	97.46% (19,998,685/20,520,731)
dpf-mm	(100,974,788/106,300,000)	(23,637,396/24,165,997) 98.37%	97.33% (19,081,525/19,604,455)
		(36,412,116/37,015,693)	

<sup>a</sup>Number of the reads that were fallen onto the first 120 supercontigs;

<sup>b</sup>Number of the covered genomic DNA of the first 120 supercontigs.



**Figure 2-9 Correlation of the mean coverage of 438 supercontigs commonly covered by pf-mm and dpf-mm.**

The plot shows log (base 2)-transformed data of the mean coverage (MC). Correlation coefficient and a theoretical trend line are also shown.

Statistical results showed that the mean coverage per supercontig was positively correlated between pf-mm and dpf-mm ( $r(436)=0.744$ ,  $P<0.01$  (two-tailed)) (Fig. 2-9). After calibration, the first 120 supercontigs shared by the two data sets, of which the total length account for 94.99% of the genomic DNA sequenced (Table 2-7), were selected for further statistical analysis. A good positive correlation of the mean coverage of these supercontigs was observed ( $r(118)=0.975$ ,  $p<0.01$  (two-tailed)) and the difference of the mean coverage of the samples pf-mm and dpf-mm was not significant ( $P=0.864 > 0.05$ ). Taken together, the results suggest the reliability of this approach with no or slight adverse effects on mRNA for transcriptome analyses.

In summary, second-generation sequencing is revolutionizing our ability to profile tissue-, cell- or environment-specific gene expression. The mRNA transcriptome of *E. multilocularis* aims to provide a deep understanding of biological characteristics that may lead to identification of targets for vaccines or chemotherapy to control echinococcosis. By using two enzymes RNase H and exonuclease, we efficiently depleted the abundant mitochondrial RNA species and thereby increased the depth of mRNA coverage in transcriptome analysis. Mapping results illustrate the robustness of the strategy for purification of *E. multilocularis* mRNA, with little contamination of the mitochondrial transcripts and thereby enhance the signal that is obtainable from high throughput platforms for surveying the transcriptome.

## Chapter 3 The genome-wide microRNA landscape of *Echinococcus multilocularis*

### Abstract

microRNAs (miRNAs) are a family of small regulatory RNAs that play multiple roles in controlling cellular mRNA levels. Here we report the stage- or cell-specific miRNA profiling of the parasitic platyhelminth, *E. multilocularis*. In total 65 potential unique miRNAs and 27 unique miRNA\* sequences were identified and grouped into 50 families of which 47 represent conserved families. These are spread across 46 unique loci. 38.5 % (25/65) of miRNAs appear to be commonly expressed in activated protoscoleces, metacystodes and cell aggregates, whereas 15.4 % (10/65) show a stage-specific pattern. Interestingly, we found that the cluster miR-71b/2/752/1 is conserved across platyhelminths but miR-752 is completely lost in *E. multilocularis*. Taken together, our data confirm that there exists a miRNA-induced regulatory mechanism in *E. multilocularis* and provide informative data for miRNA function investigation in development and neoblast biology.

### 1. Introduction

microRNAs are a subclass of small RNA with a size of ~22 bp, which regulate gene expression through the degradation or translational inhibition of targeted mRNAs. miRNAs are derived from intergenic regions, introns and even the exons of pseudogenes (Kim and Nam, 2006). Canonically, primary transcripts containing miRNA(s) (pri-miRNA) are transcribed by RNA polymerase II or III and release ~70 bp miRNA precursors (pre-miRNA) with a 5' phosphate and 3' overhang after cleavage by Drosha, a RNA III endonuclease (Bartel, 2004). Recently, another group

of miRNAs, mirtronic miRNAs (mirtrons), has been found to mature independently of the Drosha pathway with the pre-miRNAs formed from activity of the splicing machinery (Okamura et al., 2007; Ruby et al., 2007). In addition, a herpes virus bypasses Drosha cleavage to generate pre-miRNAs by tRNase Z from RNA polymerase III-transcribed pri-miRNAs with a 5' tRNA sequence (Bogerd et al., 2010). Transportation of pre-miRNAs from the nucleus to the cytoplasm is fulfilled by Exportin-5A. The cytoplasmic precursors are then processed to yield miRNA duplexes by another RNA III endonuclease Dicer (Lee et al., 2004), which also functions in the biogenesis of small interfering RNAs (siRNAs). Together with other molecules such as Argonaute, one selective fragment of the miRNA duplex is retained to form the RNA-induced silencing complex (Scaplehorn et al.), which mediates post transcriptional regulatory effects on mRNA targets (Bartel, 2004). It is possible for the opposite strand of a pre-miRNA hairpin to also form an active miRNA\*, and these are either degraded or actively function in gene regulation (Okamura et al., 2008). miRNA expression can exhibit stage-, tissue- or cell-specific patterns and are likely controlled by their own *cis*-regulatory elements for intergenic miRNAs or promoters of protein-coding genes harbouring miRNAs in intronic regions (Kim and Nam, 2006; Saini et al., 2007).

In animals, most miRNAs induce fine changes of expression of many protein-coding genes by means of perfect binding of the seed sequence (2~8 at the 5' end) with 3' UTRs or, occasionally, 5' UTRs or coding regions (Bartel, 2009). There is redundancy in the function of individual miRNA, with individual miRNAs repressing hundreds or even thousands of genes and individual mRNAs subject to regulation by potentially several miRNAs (Baek et al., 2008; Selbach et al., 2008). One possibility is that many target sites exist as buffers for miRNA functions on authentic targets of significant interaction (Seitz, 2009). Although numerous individual miRNAs are not essential for development or viability (Miska et al., 2007), increasing evidence

supports critical roles of some miRNAs in development, immunity, and tissue differentiation and proliferation (Bartel, 2004; Xiao and Rajewsky, 2009). Consistent with the finding that the distribution of mouse miRNAs are often near loci that are associated with development of tumours (Sevignani et al., 2007), the dysregulation of these miRNAs is attributed to disease occurrence and the change of miRNA expression pattern has therefore been proposed to be an alternative for diagnosis of diseases such as cancer (Lu et al., 2005). An synthetic antagonist of miR-122 has been recently used to successfully control the infection of hepatitis C virus in chimpanzee (Lanford et al., 2010); moreover, large-scale production of anti-miRNA oligonucleotides with high quality has been technically solved (Wang et al., 2010), altogether paving a plausible way towards therapeutic treatment of human diseases via intervention involving miRNA pathways.

In the Phylum Platyhelminth, miRNAs have only been described in the planarian, *Schmidtea mediterranea* (Lu et al., 2009) and two trematodes, *S. japonicum* and *S. mansoni* (Huang et al., 2009; Xue et al., 2008). Recently, the miRNA-induced silencing mechanism has been shown to be present in *E. granulosus* (Cucher et al., 2011). Considering the potential functions of miRNAs in embryogenesis (Medeiros et al., 2011), pathogenesis (Jopling et al., 2005) and neoblast biology (Friedlander et al., 2009; Gonzalez-Estevez et al., 2009; Lu et al., 2009), it is important to investigate miRNA profiling in *E. multilocularis*. In this work, we unveil the miRNA transcriptomes of metacestodes, primary cell aggregates and activated protoscolexes using next-generation sequencing technology. Comparative analysis reveals stage- or cell-specific miRNAs and the miRNAs that are probably involved in *E. multilocularis* neoblast functions. Our study gives us an overview of miRNAs in *E. multilocularis* and clues to a deeper understanding of their roles in *E. multilocularis* biology.

## **2. Material and methods**

### **2.1 Parasites and cultured cells**

Metacestode vesicles (Staebler et al.) of *E. multilocularis* isolates H95 and JAVA were kept in mongolian jirds (*Meriones unguiculatus*) (Jura et al., 1996; Tappe et al., 2007). Primary cells were prepared using mature metacestodes of isolate H95 and the cell aggregates (Agg) were formed and harvested after 1 week as described previously (Spiliotis and Brehm, 2009; Spiliotis et al., 2008). Activation of the protoscoleces (Psa), which were separated from *in vivo* metacestodes of isolate JAVA, was performed at 37 °C for 3 hours in a Hank's solution containing 0.5 mg/ml pepsin (pH 2.0).

### **2.2 Recovery of small RNA**

Small RNA species were enriched using mirVana miRNA Isolation Kit (Ambion, USA) with the method provided in the manual with a few modifications. Briefly, after washing with sterile normal saline water, parasites or cell aggregates were homogenized in the lysis buffer, followed by vigorous vortexing. The RNA-containing supernatant with addition of miRNA Homogenate Additive was incubated on ice for 15 min to facilitate the recovery of small RNA, followed by Acid-Phenol:Chloroform extraction. Small RNA species were then enriched using two-step ethanol precipitation. First one third volume of absolute ethanol was added into the aqueous phase, mixed gently and loaded into columns. The two third volume of 100% ethanol were added into the filtrate, mixed and transferred into new columns. The RNA-bound columns were washed by Washing Solution 1 one time and then Washing Solution 2/3 twice. The enriched small RNA was eluted into 50 µl pre-heated nuclease-free water. The quality of the small RNA extracted was assessed by Bioanalyzer (Agilent 2100).

## 2.3 Preparation of small RNA libraries for sequencing

100 ng of small RNA were employed to construct the libraries using Small RNA Sample Prep Kit (Illumina, USA) according to an alternative V1.5 protocol with some modifications. In brief, specific adapters were concatenated to small RNA species at the 3' and 5' ends, respectively. Following reverse transcription, enrichment of the products was performed using PCR as described previously (Morin et al., 2008). The PCR products were separated on 6% PAGE gel and the bands ranging from 90 bp to 100 bp were cut, followed by elution of small DNA from the gel in ultra pure water. The eluted RNA was qualified using Agilent Bioanalyzer and then loaded onto Genome Analyzer (Illumina) for single-end sequencing.

## 2.4 Process of short reads

Three lanes of 54 bp short reads for metacestodes, cell aggregates and activated protoscolecids were generated. Reads with ambiguous base(s) were discarded. The 3' adapters (5'-ATCTCGTATGCCGTCTTCTGCTT-3') were removed and the identical sequences were counted using a custom script. We found the majority of short sequences in the three data sets were perfectly aligned to *E. multilocularis* genome (<ftp://ftp.sanger.ac.uk/pub/pathogens/Echinococcus/multilocularis/transcriptome/>).

The sequences in a size range from 19 bp to 25 bp were recounted and the non-redundant sequence datasets were generated. Using Blastall searching against eukaryotic tRNA database (<http://gtrnadb.ucsc.edu/download.html>) and the mitochondrial genome (GenBank Accession Number: NC\_000928) and ribosomal RNA genes (5S, 18S, 5.8S and 28S) of *E. multilocularis*, short tRNA, rRNA and mitochondrial sequences were eliminated to yield 'pure' short sequences for further analyses.

## 2.5 Annotation and comparative analysis of known and novel

### miRNAs

Prediction of known and novel miRNAs was performed using miRNA biogenesis-based miRDeep package ([http://www.mdc-berlin.de/en/research/research\\_teams/systems\\_biology\\_of\\_gene\\_regulatory\\_elements/projects/miRDeep](http://www.mdc-berlin.de/en/research/research_teams/systems_biology_of_gene_regulatory_elements/projects/miRDeep)) combined with Vienna RNA Secondary Structure Prediction Program (<http://www.tbi.univie.ac.at/RNA>), which has been already employed successfully to uncover new miRNAs in dogs, *C. elegans* and planarian flatworms (Friedlander et al., 2009; Friedlander et al., 2008). In short, the purified sequences were firstly mapped to the genome supercontigs using Megablast (Zhang et al., 2000). The secondary structures of possible pre-miRNAs were calculated by RNAfold, and conserved and novel miRNAs of *E. multilocularis* were predicted and scored using the miRDeep core script with default settings.

To further identify potential known miRNAs, searches against miRNA database (<http://www.mirbase.org>, Release 14) were performed with an e-value cut-off of  $1.0 \times 10^{-2}$  using Blastall algorithm and then similarity of the conserved miRNAs across interspecies was manually checked. All potential miRNAs of *E. multilocularis* were annotated with the expression and biogenesis criteria described previously (Ambros et al., 2003): “identification of the ~22bp sequence in a cDNA library of small RNA that exactly matches a genomic sequence (expression criterion); a potential fold-back precursor structure that contains the ~22-nt miRNA sequence within one arm of the hairpin (biogenesis criterion); in cases where a cDNA sequence does not match the available genomic sequences, if that precise cDNA was nevertheless cloned multiple times and is close in sequence to a known miRNA, it can be annotated as a variant form of the known miRNA”.

We analyzed differential expression patterns of miRNAs from two different stages or cells using the Bayesian approach reported previously (Audic and Claverie, 1997). For easy presentation, miRNA abundance in vesicles (Met) or activated protoscolecetes (Psa) was compared to that in cell aggregates (Agg). To minimize the artefacts introduced by sequencing, expression fold change was calculated by the following formula: (a read count of a miRNA in Met or Psa  $\times$  a) / a read count of the corresponding miRNA in Agg, where  $a = 14,203,285/2,511,879 = 5.65$  for Met or  $14,203,285/10,763,266 = 1.32$  for Psa. In the statistical analysis, the total number of the counts of all mature miRNA variants and the total number of short sequences after removal of rRNA, tRNA and mitochondrial RNA were used. Only the same miRNAs with a count ratio of no less than 1.5 from two different sequence sets were retained for investigation and only the  $p$  values of less than 0.01 inferred by this method were interpreted as significance (Morin et al., 2008).

### **3. Results**

#### **3.1 Overview of three short sequence sets of small RNA libraries for vesicles, cell aggregates and activated protoscolecetes**

Using the Illumina Genome Analyzer, we generated 2,511,879, 14,203,285 and 10,763,266 54 bp single-ended sequences for vesicles (Staebler et al.), cell aggregates (Agg) and activated protoscolecetes (Psa), respectively. Sequences containing an adaptor were first chosen and then adaptor sequences removed. After removal of mitochondrial RNA, tRNA, ribosomal RNA and potential degraded mRNA, the sequences in a size from 19 bp to 25 bp were selected and recounted, resulting in respective 217,216, 598,568 and 521,373 of the short sequence sets for Met, Agg and Psa, which represented 10,559, 32,740 and 16,055 unique sequences with counts of 2

to 87,747 (Table 3-8). These sequences were then used for prediction of known and novel miRNAs using Megablast or/and miRDeep (Friedlander et al., 2008).

### **3.2 Known and novel miRNAs in *E. multilocularis***

We firstly predicted miRNAs using miRDeep, which scores the probability based on the seed region and quantity and position of the reads mapped on potential pre-miRNAs (Friedlander et al., 2008). In total, 28, 40 and 34 pre-miRNA loci were identified for Met, Agg and Psa, respectively, representing 46 unique loci in *E. multilocularis*. Of them, 8, 15 and 10 encoded novel miRNAs and represented 20 unique loci (Table 3-9). Among the novel miRNAs, 16 were grouped into 14 existing miRNA families, one with the same seed site as miR-877\*, and the rest unknown (Table 3-9).

**Table 3-8** Statistic results of short sequences generated for characterization of *E. multilocularis* miRNAs.

Sample	Total reads	Subtotal seq <sup>a</sup>	Unique seq./Count range <sup>a</sup>	Subtotal seq. <sup>b</sup>	Unique seq./Count range <sup>b</sup>
Agg	14,203,285	1,097,798	42,197/2~289,011	598,568	32,740/2~48,253
Met	2,511,879	318,168	14,255/2~37,242	217,261	10,559/2~21,250
Psa	10,763,266	590,173	20,809/2~87,747	521,373	16,055/2~87,747

<sup>a</sup>The number of the entire or unique sequences ranging from 19 bp to 25 bp.

<sup>b</sup>The number of the sequences left after removal of ribosomal RNA, tRNA and mitochondrial RNA homologues.

**Table 3-9 Potential novel miRNAs of *E. multilocularis*.**

miRNA	Sequences (5'-3')	Expression	miRNA*(5'-3')	Expression	Location/strand	Family
emu-nov-1	TATTGCACGTTCTTTCGCCATC	Agg, Psa, Met	CGGTGAAAGTTTATGCATTTACA	Agg, Met	contig_3457: 319713-34/+	miR-25
emu-nov-3	TCCTGGGACTTATACCGGGGCTGT	Agg	/	/	contig_5358: 24125-48/-	miR-1015
emu-nov-4	AGGTGACTCTAAAACCTTTCC	Agg, Met	/	/	contig_2743: 161973-93/-	miR-1032
emu-nov-5	TTGTGCGTCGTTTCAGTGACCGA	Agg, Psa, Met	/	/	contig_5447: 14332-54/-	miR-210
emu-nov-6	CCTTCTCCCTCGTCGCTCCAAGCCG	Agg	/	/	contig_2780: 12078-102/+	miR-748
emu-nov-7	GGGACGGAAGTCTGAAAGTTT	Agg, Psa, Met	ACCTATCACACTTCAGTCCAGT	Agg, Psa, Met	contig_3426: 469414-35/+	miR-184
emu-nov-8	GATTGCACTACCCATCGCCACA	Agg, Psa	TGGCGGTGCGCGGTGCAATTTCTG	Psa	contig_5467: 25310-32/+	miR-25
emu-nov-10	CCCTCTTCTTCGTCCACTAAGA	Agg	/	/	contig_5574: 41922-43/+	miR-877*
emu-nov-11	TAAATGCAAAATATCTGGTTATG	Agg, Psa, Met	/	/	contig_5582: 85003-25/-	miR-277
emu-nov-14	CCCACCTCAGCATGGTCACTCTTCC	Agg	/	/	contig_4550: 104334-58/+	miR-1224
emu-nov-16	TGGTGGTGGTGGTGGGGGT	Agg, Met	/	/	contig_5557: 252258-76/-	miR-876
emu-nov-18	GTAGTCTTCCGAGCAGTATATGG	Agg, Psa, Met	TAATACTGTTCGGTTAGGACGCCA	Agg, Psa, Met	contig_1840: 64696-718/+	miR-7
emu-nov-21	TGGCGCTTGATTTCAACACTGT	Agg	/	/	contig_5380: 23936-57/+	miR-2160
emu-nov-22	TGGCGCTTTCTAACTTTACTGA	Agg, Psa	/	/	contig_5380: 25635-56/-	miR-2160
emu-nov-23	AGTGTTGATGTCAGGTTGCTTCT	Agg	/	/	contig_5380: 24161-83/+	miR-505
emu-nov-24	TCGATGCCTGTCGACGCATC	Psa	TGCTTCGACAGCTAAGATC	Psa	contig_4454: 26282-301/-	/
emu-nov-25	TCTCGATCCCGGCACTACGATGC	Psa	AACGTGGTCTTGGGTCGGTGGT	Psa	contig_2898: 31190-212/+	/
emu-nov-26	TGGCGCTTAATGTCATCACGG	Psa	/	/	contig_5380: 25376-96/-	miR-2160
emu-nov-27	TGCCATCTATCTATCTGTCCG	Met	GGATGGGTGGGTGGGTGCGTGCGTG	Met	contig_6083: 3006-28/- contig_1790: 7863-85/-	/

To further investigate known miRNAs, we employed the three data sets to search the homologues of known miRNAs retrieved from miRNA database (Griffiths-Jones et al., 2008). As a result, we obtained respectively 10, 11 and 7 potentially conserved miRNAs, bringing the total number of miRNAs to 38, 51 and 41 for Met, Agg and Psa, representing a total of 65 unique miRNA sequences (Appendix I). In our data, 14, 23 and 18 sequences from the respective Met, Agg and Psa were predicted to be miRNAs\*. In summary, our data unveiled 65 potential unique miRNAs and 27 unique miRNA\* sequences of *E. multilocularis* (Appendix I).

We also noticed isomiRs, miRNA variants, in the data sets, which have also been described previously (Morin et al., 2008). Intriguingly, some conserved miRNAs of *E. multilocularis* were homologues of known miRNAs only found in plant and mammalian animals (Table 3-10). For instance, the homologues of emu-miR-2284 and 1134 have been so far described only in cattle and wheat, respectively.

### **3.3 Differential expression of miRNAs**

Expression of miRNAs is tightly regulated in a stage-, cell- or tissue-specific manner and, to some extent, the expressional status of some miRNA or small population of miRNAs exactly reflects real-time physiology of organisms (Lu et al., 2005). In our study, 38.5 % (25/65) of *E. multilocularis* miRNAs were commonly expressed in all the stages and cells investigated, whereas 15.4 % (10/65) were expressed in a stage-specific pattern (Fig. 3-10). With comparison of miRNAs in the Agg sample, which is composed of a great number of neoblasts (Spiliotis et al., 2008), we analyzed the differential expression patterns of miRNAs in Met and Psa. Consistent with the finding that miRNAs in early embryogenesis and stem cells are largely expressed at a low level (Chen et al., 2006), the results showed the down-regulation of most of the miRNAs in Agg, compared to these in Met and Psa (Table 3-11).

miRNAs are frequently clustered at the genomic level. It is estimated that nearly a half of human miRNAs are clustered, and they are generally transcribed into a long primary transcript, followed by liberation of individual pre-miRNA duplexes (Kim and Nam, 2006). We thus checked the genomic arrangements of miRNAs in *E. multilocularis* and found some miRNAs were indeed clustered. For example, three miRNAs emu-miR-71, 2b and 2c resided within approximately 260 bp DNA in the same orientation. Surprisingly, all these miRNAs were expressed in Met and Agg but with considerably different counts from 79 to 9,777. Furthermore, emu-miR-71 was highly expressed with a count of 23,088 in Psa but the expression of emu-miR-2b and 2c was completely abolished. This expression pattern that individual miRNA in miRNA clusters is tightly regulated at the same or different developmental stages was common and also observed in other clustered miRNAs in *E. multilocularis*. The similar expression profiling of clustered miRNAs has previously been described in other organisms (Lu et al., 2009). This finding strongly suggests the existence of post-transcriptional mechanism(s) for modulation of miRNA abundance.

### **3.4 miRNAs possibly related to regenerative capacity of neoblasts**

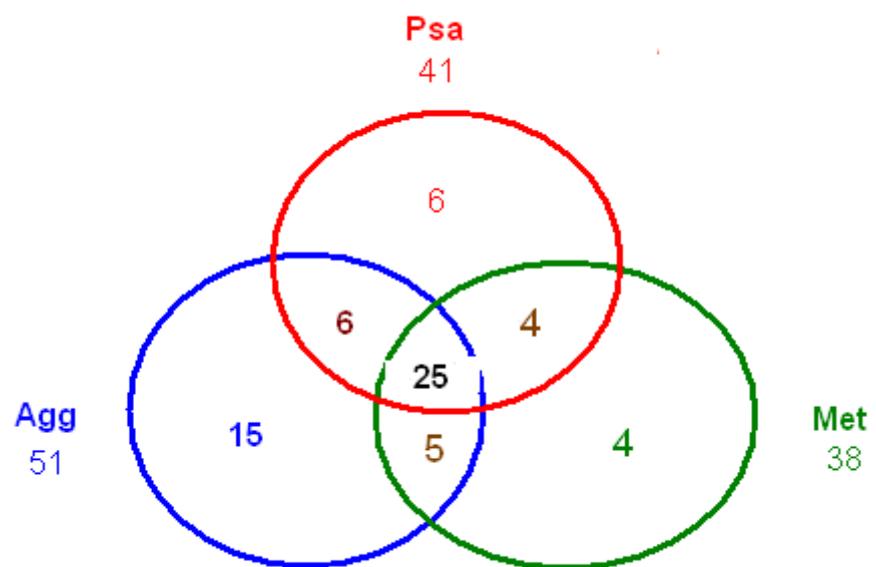
In the Phylum Platyhelminth, planarians are well characterized by remarkable regeneration, being a useful model for studies on stem cell biology and regeneration (Aboobaker, 2011; Palakodeti et al., 2006). Although some studies are controversial, likely due to different methodologies used, 15 miRNAs including let-7a and 7b are expressed in planarian neoblasts (Friedlander et al., 2009; Gonzalez-Estevez et al., 2009; Lu et al., 2009; Palakodeti et al., 2006). The neoblasts of *E. multilocularis* also

**Table 3-10** Some conserved miRNAs of *E. multilocularis*

miRNA <sup>a</sup>	Length	Homologue (5'-3')	Species	Expression
emu-miR-2162	21	guauuugcaaaauucacaau	<i>S. mediterranea, S. japonicum</i>	Agg, Met, Psa
emu-miR-2478	19	guauccacuucugacacca	cattle	Agg, Psa
emu-miR-1134	24	caacaacaacaagaagaagau	wheat	Agg
emu-miR-2284	20	bta-miR-2284q, r	cattle	Agg
emu-miR-341	21	/	mouse, rat	Agg
emu-miR-1260	19	aucccaccucgccacca	human	Agg

<sup>a</sup>*E. multilocularis* miRNAs are conserved in no more than three species using Blastn with default parameters (Release 14, miRBase).

‘/’: not applicable.



**Figure 3-10** Differential expression of miRNAs in *E. multilocularis*

**Table 3-11 Differential expression of known and novel miRNAs of *E. multilocularis***

miRNA	Count in Agg <sup>a</sup>	Met				Psa			
		Count <sup>a</sup>	Fold change <sup>b</sup>	<i>p</i> -value	Expression <sup>b</sup>	Count <sup>a</sup>	Fold change	<i>p</i> -value	Expression <sup>c</sup>
emu-miR-9	76,387	30,252	2.24	1.46E-37	UP	/	/	/	/
emu-bantam	12,831	6,479	2.85	2.03E-100	UP	3,686	2.64	0	DW
emu-miR-10b	27,629	14,384	2.94	6.64E-259	UP	11,747	1.78	0	DW
emu-miR-71	9,777	4,404	2.55	4.28E-32	UP	23,088	3.128	0	UP
emu-miR-281	10	24	13.56	4.14E-8	UP	102	13.46	8.11E-24	UP
emu-nov-5	32	61	10.77	1.89E-15	UP	611	25.20	1.68E-158	UP
emu-miR-190	1,114	2,627	13.32	0	UP	5,898	6.99	0	UP
emu-nov-18	10	19	10.74	5.25E-6	UP	45	5.94	4.40E-8	UP
emu-miR-16	1,974	10,498	30.05	0	UP	22,848	15.28	0	UP
emu-miR-87	300	/	/	/	/	3,819	16.80	0	UP
emu-miR-31	408	18	4.01	1.23E-34	DW	103	3.00	4.65E-36	DW
emu-miR-125	215	/	/	/	/	3,921	24.07	0	UP
emu-let-7	715	/	/	/	/	1,780	3.29	6.04E-139	UP
emu-miR-2a	771	88	1.55	5.48E-33	DW	/	/	/	/
emu-miR-2c	816	79	1.83	4.03E-41	DW	/	/	/	/
emu-miR-2b	1,162	612	2.98	1.35E-13	UP	/	/	/	/
emu-miR-277	2,676	1,192	2.52	3.13E-9	UP	/	/	/	/
emu-miR-36	15	92	34.65	3.25E-38	UP	/	/	/	/
emu-miR-2162	192	/	/	/	/	1,212	8.33	8.15E-214	UP
emu-miR-1	6,950	/	/	/	/	114,589	21.76	0	UP
emu-miR-133	9	/	/	/	/	346	50.75	5.01E-101	UP
emu-miR-307	662	/	/	/	/	2,561	/	2.57E-322	UP
emu-miR-219	35	/	/	/	/	564	2.64	4.75E-141	UP

emu-nov-1	302	/	/	/	/	1,087	1.78	1.19E-129	UP
emu-miR-4989	3,138	/	/	/	/	11,692	3.12	0	UP
emu-miR-124	33	/	/	/	/	328	13.46	4.60E-72	UP
emu-nov-7	494	/	/	/	/	1,036	25.20	9.19E-63	UP
emu-miR-7	51	/	/	/	/	344	6.99	2.87E-64	UP
emu-nov-22	169	/	/	/	/	470	5.94	6.15E-44	UP

<sup>a</sup>The total number of isomiRs;

<sup>b</sup>The fold changes were calculated by the following formula: (a read count of a miRNA in Met or Psa  $\times$  a) / a read count of the corresponding miRNA in Agg, where a = 14,203,285/2,511,879 = 5.65 for Met or 14,203,285/10,763,266 = 1.32 for Psa;

<sup>c</sup>Compared to Agg, some miRNAs from Met or Psa are up-regulated (UP) and others down-regulated (DW);

“/”: not applicable.

possess ability to *in vitro* generate mature metacestodes under specific conditions (Spiliotis et al., 2008). To classify the miRNAs possibly associated with this regenerative capacity, we performed a comparison between *E. multilocularis* miRNAs and miRNAs expressed in *S. mediterranea* neoblasts (Friedlander et al., 2009; Gonzalez-Estevez et al., 2009; Lu et al., 2009). In total, we found 7 homologues in *E. multilocularis* and, although there were no complete matches, all the seed sites were conserved (Table 3-12).

In these potentially neoblast-related miRNAs, it was evident that emu-miR-71 was extraordinarily highly expressed in all samples studied, especially in the activated protoscolecetes, and the low expression of the others with exception of emu-miR-2 family in Agg. Although no direct evidence was provided in our study, these miRNAs were potential biomarkers for *E. multilocularis* neoblasts.

### **3.5 miR-71b/2 cluster across the Phylum Platyhelminth**

It is remarkable that the clustering organization of miR-71b/2 is conserved in all platyhelminths investigated so far (Huang et al., 2009; Lu et al., 2009) (Fig. 3-11). The difference was the replacement of miR-752 by miR-2b in *S. japonicum* and its removal in *E. multilocularis*. The conservation of the cluster miR-71b/2 was also observed in *E. granulosus*, in which the organization of the cluster was identical to that of *E. multilocularis* (Fig. 3-12), suggesting that the current structure of the cluster miR-71b/2 may have formed before speciation of both Echinococcus species but after divergence between trematodes and cestodes. Although the seed sites were intact in each miRNA of this cluster (Fig. 3-11 B), mutations outside the seed regions may offer distinct effects on miRNA/mRNA target interactions. It is noticeable that the miR71/2 cluster has been expanded into two copies in *S. japonicum*, residing on two different contigs (Huang et al., 2009), but not in *E. multilocularis*, suggesting additional roles of this cluster during the development in immature flukes.

**Table 3-12 Potential neoblast-specific miRNAs of *E. multilocularis***

miRNA <sup>a</sup>	Homologue	Length <sup>b</sup>	Expression/Count	Alignment (5'-3') <sup>c</sup>	Ref.
sme-miR-36b	emu-miR-36	23/22	Agg (15), Met (92), Psa (5)	TCACCGGGTAGACATTAATCATG -----CC-TGC?	(Friedlander et al., 2009; Gonzalez-Estevez et al., 2009)
sme-miR-2a	emu-miR-2a	24/24	Agg (771), Met (88)	TATCACAGCCCCGCTTGGAAACGCT A-----T-----C--	(Friedlander et al., 2009)
sme-miR-2d	emu-miR-2b <sup>d</sup>	21/22	Agg (1162), Met (612)	TCACAGCCAAATTTGATGTCC? -----TA-----AA-G	(Friedlander et al., 2009; Gonzalez-Estevez et al., 2009)
sme-miR-13	emu-miR-2c <sup>d</sup>	22/24	Agg (816), Met (79)	TATCACAGTCATGCTAAAGAGC?? -----C-C-----TGG--CACA	(Friedlander et al., 2009; Lu et al., 2009)
sme-miR-71b	emu-miR-71 <sup>d</sup>	22/21	Agg (9777), Met (4404), Psa (23088)	TGAAAGACACAGGTAGTGGGAC -----GAT-----A--?	(Friedlander et al., 2009; Gonzalez-Estevez et al., 2009; Lu et al., 2009)
sme-miR-7b	emu-miR-7	23/24	Met (2)	TGGAAGACTGTCGATTTTCGTTGT? -----AGT-----T-----T	(Gonzalez-Estevez et al., 2009)
sme-miR-7c		22/24		TGGAAGACTGATGATTT?GCTGA? -----AG-----T-T--TT	(Gonzalez-Estevez et al., 2009)
sme-miR-124a	emu-miR-124	21/21	Agg (72), Met (2), Psa (33)	TAAGGCACGCGGTGAATGCTT -----A-CA	(Gonzalez-Estevez et al., 2009)
sme-miR-124b		22/21		TAAGGCACGCGGTGAATGCTGA -----A-CA?	(Gonzalez-Estevez et al., 2009)

<sup>a</sup>*S. mediterranea* miRNAs closely related to neoblast biology. Besides from the listed in this table, other six sem-miR-752, 92, let-7a, let-7b, 2160 and 756 are proposed to participate in neoblast functions.

<sup>b</sup>Length of *S. mediterranea* miRNAs (before slash) and homologues in *E. multilocularis* (after slash).

<sup>c</sup>Sequences of miRNAs of both species are aligned. Consensus nucleotides are masked by dash and gaps are filled by question markers. Variants in homologues of *E. multilocularis* are shown and highlighted in yellow.

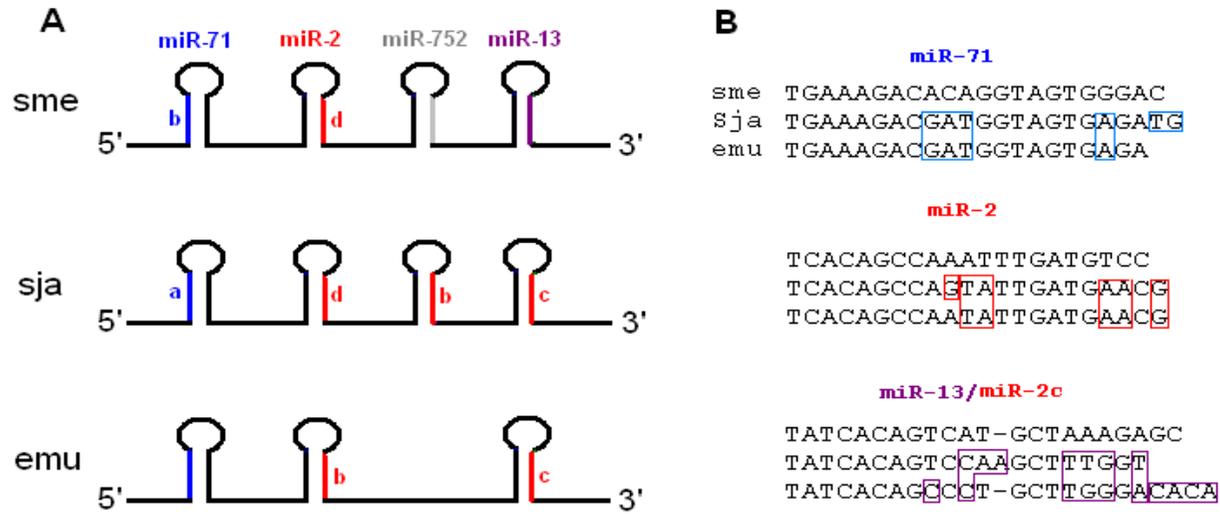
<sup>d</sup>These three miRNAs are clustered together at the genomic level.

This notion is consistent with the abruptly up-regulated expression of emu-miR-71 in the activated protoscolexes (our data).

### **3.6 Comparative analysis with other platyhelminths miRNAs**

Based on 5' seed sites (nucleotides from 2 to 8 at the 5' terminus), 65 possible miRNAs of *E. multilocularis* were grouped into 50 families with 47 of them being known (Bartel, 2004). By comparison with *S. mediterranea*, *Schistosoma* and *Echinococcus* species, we found that these parasites shared 18 common families, while *Echinococcus* species shared 26 families with *S. mediterranea* (Fig. 3-13). It is worthwhile noting that one *E. multilocularis* miRNA, emu-miR-2162, has been so far described only in a planarian and a trematode (Table 3-10), indicating an ancient origination.

To explore miRNAs in *E. granulosus*, we employed *E. multilocularis* miRNAs to interrogate the genome and evaluate their corresponding structures as mentioned previously. Totally, 40 potential miRNAs were identified (Appendix II) and 20 of them have been recently determined experimentally in this parasite using a canonical sequencing approach (Cucher et al., 2011). There were 32 miRNA families shared by both *E. granulosus* and *E. multilocularis*. The former expressed at least two species-specific miRNAs miR-4991 and miR-4990, while the latter had 18 miRNA families absent in *E. granulosus* (Fig. 3-13). It should be aware that this difference may be partially due to the incomplete genome of *E. granulosus*.



**Figure 3-11 Conservation of the miR-71b/2/752/13 cluster in platyhelminths**

The genomic organization of this clustered miRNAs in *E. multilocularis* (emu) is compared with *S. japonicum* (sja) and *S. mediterranea* (sme) (**A**). Sequences of miR-71 and miR-2 and miR-13 from three species above are aligned (**B**). The nucleotides different from ones of sme are boxed and gaps are filled by dash (-).

**A**

```

emul  TGTGTCCCAAGCAGGGCTGTGATACTCCAACTGAGTAGATCACCACCCTTGGGACACCGA 60
egra  TGTGTCCCAAGCAGGGCTGTGATACTCCAACTGAGTAGATCACCACCCTTGGGACACCGA 60
      miR-2c

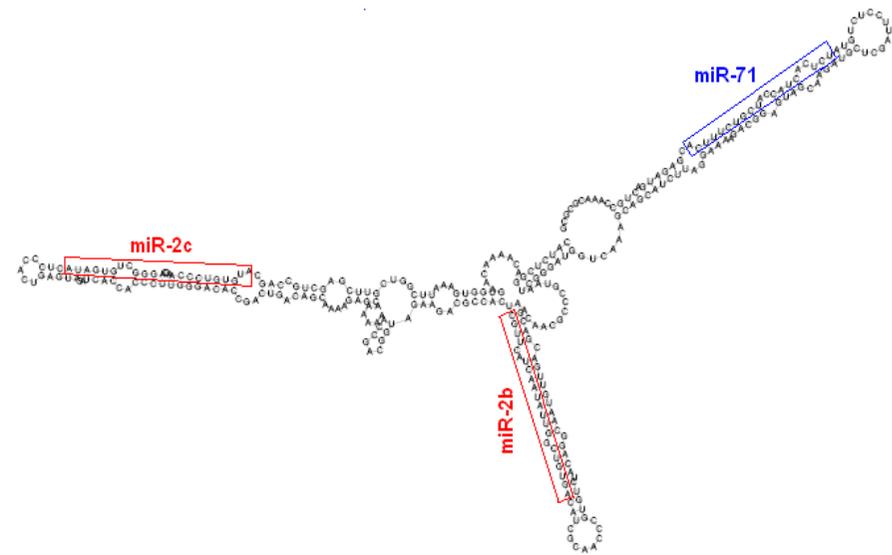
emul  CTGACAGCAAAAAGAGCAGAAA GAACCGACGGTAGAAGACGCCACTCGTTCATCAATATTG 120
egra  CTGACAGCAAAAAGAGCAAAA -AACCGACGGTAGAAGACGCCACTCGTTCATCAATATTG 119
      miR-2b

emul  GCTGTGACATCGCAACCCGTGTCTACAGGCAATGTTGACGACGAACAACGCCCGTAATCG 180
egra  GCTGTGACATCGCAACCCGTGTCTACAGGCAATGTTGACGACGAACAACGCCCGTAATCG 179

emul  GGATGGTCAAAGCAGCATCTTAGGAAAAGACGGAGTAGCAAGATGCTCGATTCCCTCTGTA 240
egra  GGATGGTCAAAGCAGCATCTTAGGAAAAGACGGAGTAGCAAGATGCTCGATTCCCTCTGTA 239

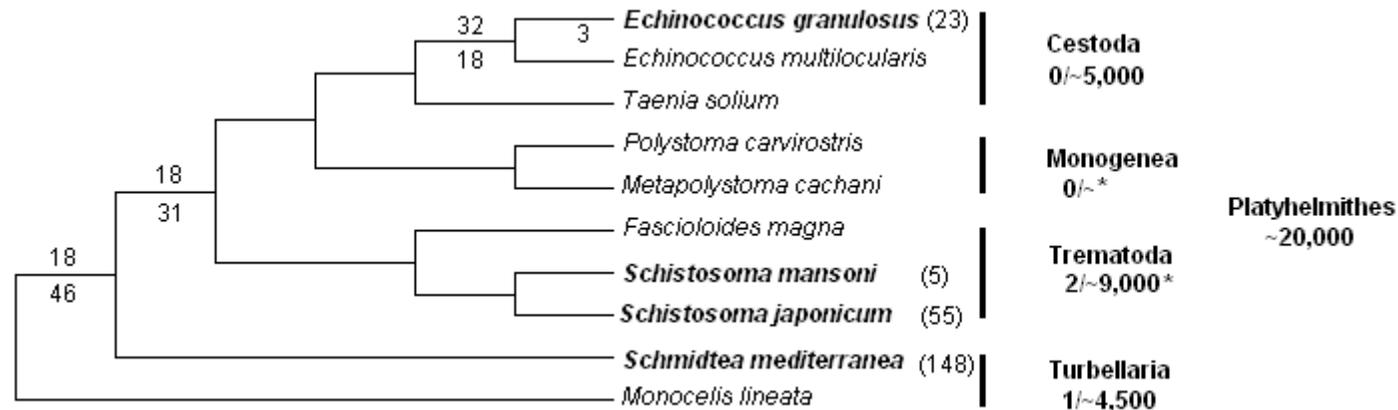
emul  TCTCACTACCATCGTCTTTCA 261
egra  TCTCACTACCATCGTCTTTCA 260
      miR-71

```

**B**

**Figure 3-12 Alignment and putative secondary structure of the miR-71b/2/752/13 cluster in *Echinococcus* species.**

The genomic sequences for the the miR-71b/2/752/13 cluster in *E. multilocularis* (emul) and *E. granulosus* (egra) were aligned using Clustral W (A). The individual miRNA is boxed in red or blue and the length is also shown on the left. The different nucleotides are highlighted in gray and the gap is filled with ‘-’. The structure of the partial primary transcript of *E. granulosus* miR-71b/2/752/13 cluster was predicted using RNAfold (B).



**Figure 3-13 Phylogenetic distribution of miRNAs in platyhelminths**

The Phylum Platyhelminth comprises four classes with approximate 20,000 species. At present, miRNAs have been reported in Trematoda with ~9,000 species plus Monogenea (\*), and in Cestoda and Turbellaria. The numbers of miRNAs (Release 17) are indicated in the brackets after the species in bold. The numbers above the branches stand for miRNA families commonly shared by along the phylogeny of platyhelminths, whereas the ones under the branches for miRNA families specific to the below branching species.

## 4. Discussion

With a few exceptions, miRNAs are ubiquitously found in organisms from the unicellular alga *Chlamydomonas reinhardtii* to humans (Molnar et al., 2007). Recent studies have shown that miRNAs are also present in free-living and parasitic platyhelminths (Cucher et al., 2011; Huang et al., 2009; Lu et al., 2009; Xue et al., 2008). In this study, we showed genome-wide miRNA expression patterns in a cestode, *E. multilocularis*, using a deep-sequencing approach. Moreover, we also bioinformatically uncovered additional 20 miRNAs for *E. granulosus*. To corroborate this result, we searched orthologues related to miRNA pathway. Indeed, all potential key components, such as Drosha, Dicer and Argonaute family, were present in *E. multilocularis* (data not shown). Taken together these results confirm the miRNA-induced silence mechanism in *E. multilocularis* that acts as a part of a gene regulation network.

Our study identified 27 miRNA\* and 65 potential miRNAs. Of them, two miRNAs emu-miR-2a and emu-nov-27 had two loci and sixteen miRNAs were identified by homologous searching lack genomic locations. Among these miRNAs, emu-miR-1134, 2478 and 1260 shared high similarity with corresponding miRNAs in wheat, cattle and human, respectively. This study does not provide compelling evidence to rule out the possibility of that these miRNAs may be contaminants. *E. multilocularis* miRNAs identified belonged to 47 known families and 19 families shared the same seed regions with miRNAs from dogs, mice and humans. Three miRNAs emu-nov-24, 25 and 27, one of which has a homologue in *E. granulosus*, appear to be families specific to Echinococcus species. A recent study revealed that four novel miRNAs miR-4988, 4989, 4990 and 4991 were found in *E. granulosus* (Cucher et al., 2011). With the exception of miR-4989, expression of the rest was not seen in our datasets. In *E. multilocularis* genome, the DNA sequenc completely identical to *E. granulosus*

pre-miR-4988 was present (pathogen\_EMU\_contig\_62302: 3441643-3441724/-), whereas the sequence sharing high similarity with pre-miR-4990 was not found (data not shown). Moreover, *E. multilocularis* genome had two genomic DNAs that were identical to *E. granulosus* miR-4991 but the free energy of the putative miRNA precursors (-14.2 kcal/mol) indicated that they were not stable. These results suggest that miR-4990 is likely to be species-specific. Intriguingly, *E. multilocularis* seems to express a few miRNAs that are reported in no more than three species. In particular, one miRNA emu-miR-1134 sharing high similarity with a miRNA exclusively reported in plant was also found in this parasite. This miRNA lacked genomic information and was expressed at very low level in all the samples and there was not convincing evidence to rule out the possibility of contamination.

It has been recently established that, under specific conditions, the primary cells derived from *E. multilocularis* vesicles are able to form cellular aggregates that contain plenty of neoblasts and then grow into mature vesicles with a laminated layer (Spiliotis et al., 2008). This suggests the retention of regenerative capacity in *E. multilocularis* neoblasts. A finding in this study was that seven miRNAs potentially associated with neoblast functions were present in *E. multilocularis*. It has been shown that miRNA pathways are required for a stem cell cycle by means of regulation of the normal G1-S phase transition, in which miRNAs are proposed to directly or indirectly repress the G1/S inhibitor Dap (Hatfield et al., 2005; Wang et al., 2008). In addition, the loss of germline stem cells maintenance or division occurs in the ovaries in *Drosophila* with a mutated double-stranded RNA-binding domain protein Loquacious, which is indispensable for biogenesis and normal functions of miRNAs (Forstemann et al., 2005). Similarly, a miR-290 cluster that is a prominently expressed in mouse embryonic stem cell miRNAs has been shown to increase the efficacy of pluripotent stem cell reprogramming (Judson et al., 2009). The miRNA-induced silencing mechanism is supposed to be involved in the normal functions of neoblasts

in the free-living platyhelminths, planarians (Thatcher and Patton, 2010). In line with this idea, the expression of a number of miRNAs has been found to be enriched in planarian neoblasts (Friedlander et al., 2009; Gonzalez-Estevez et al., 2009; Lu et al., 2009). In our datasets, with exception of emu-miR-7, all the miRNAs potentially related to the neoblasts were expressed in the cell aggregates, which compose a large number of *E. multilocularis* neoblasts (Spiliotis et al., 2008). However, the expression status of these miRNAs in the neoblast remains to be tested experimentally.

miR-71 is organized in a cluster with other miRNAs across the platyhelminths studied and appears to be developmentally regulated in *E. multilocularis*. In comparison with the cell aggregates, emu-miR-71 was up-regulated in the vesicles and particularly in the activated protoscolexes. Of interest is the similar expression pattern of miR-71(a/b) in the trematodes (Huang et al., 2009; Knapp et al., 2009; Xue et al., 2008), free-living nematode (Kato et al., 2009) and cestode (our data) but not in a plant nematode (Huang et al., 2010). In addition to functional connections with neoblast biology (Knapp et al., 2008; Lin et al., 2009; Liu et al., 2009b), miR-71 is expected to play a role probably related to developmental transition from the larval to adult stages (our data). Further experiments will be worthwhile specifying the connections between its expression and functions.

Several miRNAs, such as emu-miR-9, 10, bantam, 71 and 1, were highly expressed in all the samples investigated. In particular, the expression of emu-miR-1 was significantly up-regulated from metacestodes with a count of 7678 to activated protoscolexes with a count of 114,589. A line of evidence has shown that, together with miR-133, miR-1 cooperatively functions in muscle proliferation and differentiation (Chen et al., 2006; Luo et al., 2008). In human and mouse, the expression of miR-1 and 133 are restricted to cardiac and skeletal muscle. They are clustered together on two chromosomes in mice, flanked by respective 9.3 and 2.5 kb DNA, and transcribed into one polycistronic transcript (Chen et al., 2006). Likewise,

*E. multilocularis* miR-1 and 133 were located in the same supercontig (contig\_6122), which were separated by approximate 12 kb, but the expression of emu-miR-133 was very low and not detected in the metacystode. Conversely, such genomic arrangement of miR-133 and 1 is not observed in other two platyhelminths *S. mediterranea* and *S. japonicum* (Friedlander et al., 2009; Huang et al., 2009; Lu et al., 2009). Considering imbalance of expression level between emu-miR-133 and miR-1, we can't rule out the possibility of that two miRNAs have different functions to act independently not cooperatively. Increasing expression of miR-1 indicates its important roles in the development from activated protoscoleces to adults, for instance in muscle biogenesis.

## **Chapter 4 Diversity of gene structure and alternative splicing of fatty acid binding protein genes in invertebrates**

### **Abstract**

FABPs are a family of fatty acid-binding small proteins essential for lipid trafficking, energy storage and gene regulation. Although they have low primary sequence identity, these proteins share a conserved tertiary structure comprised of ten beta sheets and two alpha helices. The complete genomes of 35 invertebrates, together with transcriptomes and ESTs, allow us to systematically investigate the gene structure and alternative splicing of FABP genes over a wide range of phyla. The genomic loci for FABP genes were diverse and their genomic structure varied throughout invertebrates. In particular, the intronless FABP genes, most of which key residues involved in fatty acid binding were altered, were extensively distributed over five phyla. Furthermore, an apparent intron loss of FABP genes seemed to have occurred in some arthropod species and particularly in platyhelminths. Interestingly, several invertebrates including one trematode, one nematode and four arthropods generated FABP mRNA variants via alternative splicing. These results demonstrate that gene copy numbers and post-transcriptional modifications act as important mechanisms contributing to diverse functions of FABPs in these invertebrates.

### **1. Introduction**

Lipids are a very important subclass of constituents in the maintenance of normal physiology in organisms and a delicate balance of these hydrophobic molecules is partially regulated by lipid chaperons, a family of high affinity fatty acid-binding proteins. In the 1990's, FABPs were firstly described in the muscle of the locust *Schistocerca gregaria* (Haunerland and Chisholm, 1990) and so far more than 30

FABP genes have been found in a wide range of invertebrates (Esteves and Ehrlich, 2006). Fatty acid-binding proteins (FABPs), small proteins of approximately 15 kDa, execute fatty acid transport and, together with intracellular retinol- and retinoic acid-binding proteins, comprise the lipid binding protein (iLBP) subfamily of calycins that also contain two subfamilies lipocalin and avidin. iLBPs are ubiquitously expressed in animals and ancestral iLBP genes are supposed to have arisen after separation of animals from fungi and plants (Schaap et al., 2002). Multiple gene duplications have occurred in this family, giving rise to 16 iLBP types in vertebrates (Schaap et al., 2002). It has been shown that FABPs are expressed in invertebrates and vertebrates but absent in archaeobacteria and yeast (Esteves and Ehrlich, 2006; Storch and Corsico, 2008).

Mammalian FABP genes generally consist of four exons with a comparable intron structure and most of them are dispersed on distinct chromosomes in human beings, rats and mice (Furuhashi and Hotamisligil, 2008; Schaap et al., 2002). Analysis of a few invertebrates shows that, unlike mammals, invertebrates have different genomic organizations for FABP genes, in aspects of size, exon and intron numbers (Esteves and Ehrlich, 2006; Schaap et al., 2002). For example, *Caenorhabditis elegans* expresses nine FABPs, also known as lipid binding proteins (LBP), and these FABP genes reside on different chromosomes. Of them, LBP-5 and LBP-6 genes are comprised two exons and one intron and positioned on chromosome I, suggesting that they might have arisen from direct gene duplication. However, invertebrate and vertebrate FABP genes commonly share some similar response elements in the promoter regions, such as a glucocorticoid half-site and peroxisome proliferator response elements (Cook et al., 1988; Esteves et al., 2003).

Although FABPs share moderate homology at the nucleotide and amino acid levels across invertebrate interspecies and even within species, their tertiary structures are highly conserved, characterized by the formation of a cavity by ten anti-parallel sheets

and two helices that accommodates fatty acid (s). With a few exceptions, the residues related to ligand binding appear to be conserved in both invertebrate and vertebrate FABPs. In the  $\beta$ -barrel cavity, the bound fatty acids interacts with some residues R...R-x-Y, called the P2 motif. Moreover, residues Phe on the first helix and Ala/Pro-Asp in the turn between  $\beta$ E and  $\beta$ F are also critical for binding affinity in FABPs (Jakobsson et al., 2003).

A number of previous studies have been conducted to resolve the phylogenetic relationship of hydrophobic molecule-binding proteins (Esteves and Ehrlich, 2006; Esteves et al., 1997; Ganfornina et al., 2000; Schaap et al., 2002). A systematic analysis of lipocalins demonstrates that FABPs are not a sister group of the lipocalin subfamily and they might have evolved from already-existing lipocalins (Ganfornina et al., 2000). However, the systematic phylogeny of invertebrate FABP genes remains incomplete.

With the availability of complete genomes and transcriptome data, it is feasible to systematically and extensively perform phylogenetic analyses, such as gene organization and post-transcriptional modification, across 35 invertebrates from 8 phyla. This study will provide insights into a new understanding the diversity of the genomic organization of invertebrate FABPs. Additionally, it is also shown that both increase of gene copy numbers followed by divergence and alternative splicing are the mechanisms likely responsible for functional expansion of FABPs in invertebrate species.

## **2. Methods**

### **2.1 Identification and annotation of FABP genes**

Invertebrate FABP genes were acquired using TblastN to search each of genome databases (Table 4-13). Since these genes generally share moderate similarity at both

nucleotide and amino acid levels, TblastN searching was performed using already reported FABP sequence(s) from evolutionarily close species where applicable. Two sequential approaches were utilized for determination of the exons and exon boundaries. Firstly the exons and their boundaries were determined from TblastN outcomes as highly-scored segment pairs or gaps within the segment pairs as described previously (Lek et al., 2010). FABP gene models were then verified and finely modified using transcriptome data or expression sequence tags (ESTs). Segment pairs that dispersed over two or more supercontigs were not considered to build gene models in this study. The intron-exon boundaries were manually checked based on consensus splicing acceptor and donor sites.

## **2.2 Sequence alignments and secondary structure prediction**

The FABP protein sequences were aligned using Clustal W algorithm (MEGA 4.0) with default parameters (Tamura et al., 2007) and then manually checked. The secondary structures of FABPs were predicted using Psipred (Bryson et al., 2005).

# **3. Results**

## **3.1 Identification and annotation of FABP genes across invertebrates**

In our study two criteria were applied for identification of FABP genes. First, considering that most of known FABPs are ~130 amino acids (aa) in length, here we artificially set the size limit of FABPs from 80 to 180aa ( $130 \pm 50$ aa). Meanwhile the sequences within the size limit were used for secondary structure prediction. The sequences with the putatively typical structural elements were considered as FABP genes. The FABP genes were searched using TblastN with experimentally or putatively identified FABP gene(s) in a closely related species. Otherwise, *S. mansoni* FABPs (Smp\_095360 and Smp\_046800) or *C. elegans* FABPs (NP\_505016,

NP\_508558, NP\_508557, NP\_491928, NP\_506440, NP\_491926, NP\_001041249, NP\_506444 and NP\_001033511) were used as queries. During sequence searching we indeed obtained high-scoring hits that encoded more than 180aa or less than 80aa, which all were excluded in further analyses in this study. For instance, a *Branchiostoma floridae* hypothetical protein (987aa, XP\_002589099) contained a region at the C terminal that shared 96% identities with *Branchiostoma belcheri* FABP (136aa, ADD10136). In total, 118 sequences were collected from 33 invertebrate species including one placozoan, two annelids, one mollusc, six platyhelminths, seven nematodes, twelve arthropods, one echinoderm and three chordates (Table 4-13). No homologues of FABPs were identified in two Cnidaria species, *Hydra magnipapillata* and *Nematostella vectensis*. Notably, four *Haemonchus contortus* FABP genes were identified by TBlastN searches against transcriptome (NCBI) but none of them was found in the genome, largely due to incomplete genomic data (Sanger). Five FABP transcripts that were not found in the genome sequences were derived from respective transcriptome or ESTs for the following species: *Helobdella robusta*, *Lottia gigantea*, *Schistosoma japonicum*, *Heterorhabditis bacteriophora* and *Saccoglossus kowalevskii*. It is noticed that, with the exception of the body louse, *Pediculus humanus corporis*, partial or all FABP genes were validated by EST or transcriptomic data.

Invertebrate genomes contained various numbers of the genomic loci for FABP, ranging from one to fifteen in Chordata, *B. floridae*. Only one FABP locus for the following invertebrates was found: *S. japonicum*, *Drosophila melanogaster*, *Aedes aegypti*, *Tribolium castaneum*, *Culex pipiens quinquefasciatus* and *Rhodnius prolixus* (Table 4-13). It was also found that each of *Echinococcus multilocularis*, *Anopheles gambiae* and *B. floridae* had two distinct loci that encoded the identical FABPs.

**Table 4-13 Distribution and features of FABP genes in invertebrates**

Species	Num. loci	Length <sup>a</sup>	Evidence/EST <sup>b</sup>	Species Abbreviation	Data origin <sup>c</sup>
<b>Cnidaria</b>					
<i>Nematostella vectensis</i>	/	/	/	/	JGI
<i>Hydra magnipapillata</i>	/	/	/	/	Metazome
<b>Placozoa</b>					
<i>Trichoplax adhaerens</i>	5	120~178	1/5	Tadh	JGI NCBI
<b>Annelida</b>					
<i>Capitella teleta</i>	7	135~167	7/7	Ctel	JGI
<i>Helobdella robusta</i>	3	119~143	3/4	Hrob	JGI NCBI
<b>Mollusca</b>					
<i>Lottia gigantea</i>	7	132~163	7/8	Lgig	JGI NCBI
<b>Platyhelminth</b>					
<i>Schmidtea mediterranea</i>	3	123~168	2/3	Smed	SmedGD NCBI
<i>Schistosoma mansoni</i>	2	132, 133	2/2	Sman	Sanger NCBI
<i>Schistosoma japonicum</i>	1	130	2/2	Sjap	Sanger NCBI
<i>Echinococcus granulosus</i>	5	124~143	2/5	Egra	NCBI Sanger
<i>Echinococcus multilocularis</i>	5	124~143	4/4	Emul	Sanger
<i>Mesocestoides vogae</i>	2	131, 121	2/2	Mvog	NCBI
<b>Nematoda</b>					
<i>Caenorhabditis elegans</i>	9	135~165	9/9	Cele	NCBI
<i>Pristionchus pacificus</i>	4	118~163	4/4	Ppac	NCBI WUGSC
<i>Heterorhabditis bacteriophora</i>	3	133~164	4/4	Hbac	NCBI WUGSC
<i>Trichinella spiralis</i>	3	133~143	3/3	Tspi	NCBI WUGSC
<i>Haemonchus contortus</i>	0 <sup>d</sup>	133~164	4/4	Hcon	Sanger NCBI
<i>Strongyloides ratti</i>	4	132~165	4/4	Srat	Sanger
<i>Brugia malayi</i>	3	130~180	3/3	Bmal	NCBI

Continued

<b>Arthropod</b>					
<i>Daphnia pulex</i>	2	130, 131	2/2	Dpul	wFleaBase
<i>Pediculus humanus corporis</i>	3	132~135	0/3	Phum	NCBI VectorBaseFlyBase
<i>Bombyx mori</i>	5	95~142	4/5	Bmor	SilkDB
<i>Tribolium castaneum</i>	1	136	1/1	Tcas	NCBI
<i>Nasonia vitripennis</i>	2	132	2/2	Nvit	NCBI
<i>Acyrtosiphon pisum</i>	3	135, 136	3/3	Apis	NCBI
<i>Apis mellifera</i>	2	132~134	2/2	Amel	NCBI
<i>Drosophila melanogaster</i>	1	117~157	1/1	Dmel	NCBI FlyBase
<i>Anopheles gambiae</i>	2	131	1/1	Agam	NCBI
<i>Aedes aegypti</i>	1	132	1/1	Aaeg	NCBI
<i>Culex pipiens quinquefasciatus</i>	1	132	1/1	Cpip	NCBI
<i>Rhodnius prolixus</i>	1	134	1/1	Rpro	NCBI
<b>Echinodermata</b>					
<i>Strongylocentrotus purpuratus</i>	2	130	2/2	Spur	NCBI JGI
<b>Chordata</b>					
<i>Branchiostoma floridae</i>	15	135~151	7/13	Bflo	JGI NCBI
<i>Ciona savignyi</i>	3	127~133	3/3	Csav	Broad NCBI
<i>Saccoglossus kowalevskii</i>	3	132~138	4/4	Skow	Baylor NCBI

<sup>a</sup>Amino acid length; <sup>b</sup>In invertebrates, the number of different FABP genes (the number after '/') is various and expression of partial or all them (the number before '/') is validated by transcriptomic data; <sup>c</sup>JGI: Joint Genome Institute; NCBI: National Centre for Biotechnology Information; SmedGD: *Schmidtea mediterranea* Genome Database; Sanger: Wellcome Trust Sanger Institute; WUGSC: Washington University Genome Sequencing Centre; wFleaBase: Daphnia Water Flea Genome Database; FlyBase: Drosophila database; SilkDB: silkworm database; Broad: Broad Institute; Baylor: Baylor College of Medicine; <sup>d</sup>no genomic loci for FBAPs were found using Blast with EST sequences.

### **3.2 Diversity of FABP gene structures and exons across invertebrates**

Although intronless FABP pseudogenes have been described in several species including humans (Bonne et al., 2003; Prinsen et al., 1997; Treuner et al., 1994), mammalian FABP genes exhibit similar gene structures, containing four exons and three introns (Schaap et al., 2002). An analysis of FABP gene organization revealed diversity in invertebrates, especially in the Phyla Platyhelminth and Nematoda, although the canonical organization predominated (Table 4-14).

A six-exon five-intron structure for FABP was only found in the early branching invertebrate *Trichoplax adhaerens*. The genes comprised of five exons and four introns were absent in Arthropoda, Chordata, Annelida and Echinodermata species. In contrast to other metazoans, platyhelminths and arthropods encoded FABP genes predominantly having two-exon one-intron and three-exon two-intron structures, respectively (Table 4-14).

### **3.3 Intron loss in the Phylum Platyhelminth**

In the Phylum Platyhelminth, the free-living flatworm, *S. mediterranea*, was intron-rich in FABP genes, while the trematodes with a free-living stage, *S. mansoni* and *S. japonicum*, did not contain the five-exon four-intron FABP genes present in *S. mediterranea*. In parasitic cestodes including *E. granulosus*, *E. multilocularis* and *Mesocestoides vogae*, FABP genes had two exons and one intron and/or only one exon (Table 4-14). This pronounced variety of FABP gene structures is along with the evolutionary history of platyhelminths, suggesting that introns tend to have lost in flatworm FABPs during evolution.

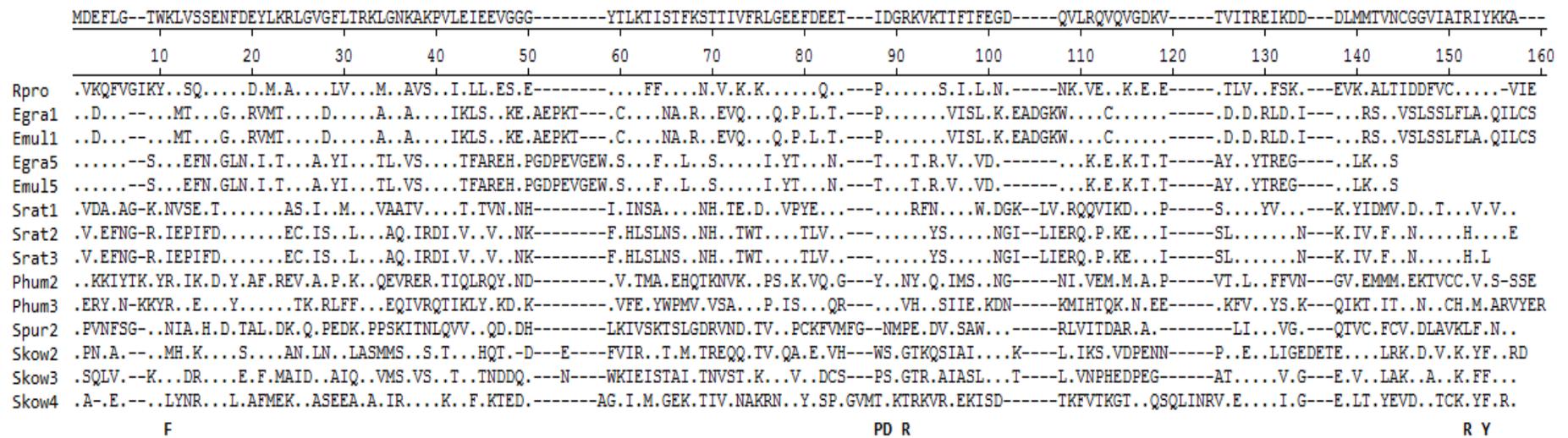
**Table 4-14 Diversity of FABP genomic structures in invertebrates**

Species <sup>1</sup>	Number of exons					
	6	5	4	3	2	1
<b>Placozoa</b>						
<i>Trichoplax adhaerens</i>	1	1	3			
<b>Annelida</b>						
<i>Capitella teleta</i>			7			
<i>Helobdella robusta</i>			2			
<b>Mollusca</b>						
<i>Lottia gigantea</i>		1	7			
<b>Platyhelminth</b>						
<i>Schmidtea mediterranea</i>		1	2			
<i>Schistosoma mansoni</i>			2			
<i>Schistosoma japonicum</i>				1		
<i>Echinococcus granulosus</i>					3	2
<i>Echinococcus multilocularis</i>					3	2
<i>Mesocestoides vogae</i>					2	
<b>Nematoda</b>						
<i>Caenorhabditis elegans</i>			1	4	4	
<i>Pristionchus pacificus</i>		1	2			
<i>Heterorhabditis bacteriophora</i>		2	1			
<i>Trichinella spiralis</i>			3			
<i>Strongyloides ratti</i>					1	3

Continued

<i>Brugia malayi</i>		2	1			
<b>Arthropod</b>						
<i>Daphnia pulex</i>		1	1			
<i>Pediculus humanus corporis</i>			1			2
<i>Bombyx mori</i>		4	1			
<i>Tribolium castaneum</i>			1			
<i>Nasonia vitripennis</i>			2			
<i>Acyrtosiphon pisum</i>			3			
<i>Apis mellifera</i>		1	1			
<i>Drosophila melanogaster</i>			1			
<i>Anopheles gambiae</i>					2	
<i>Aedes aegypti</i>					1	
<i>Culex pipiens quinquefasciatus</i>					1	
<i>Rhodnius prolixus</i>						1
<b>Echinodermata</b>						
<i>Strongylocentrotus purpuratus</i>					1	1
<b>Chordata</b>						
<i>Branchiostoma floridae</i>		11	4			
<i>Ciona savignyi</i>		2	1			
<i>Saccoglossus kowalevskii</i>		1				3
<b>Total</b>	1	6	52	22	18	14

<sup>1</sup>In total thirty-five species from eight phyla were included in this study.



**Figure 4-14 Fatty acid binding-related residues in intronless FABP genes of invertebrates.**

Invertebrate intronless FABP amino acid sequences were aligned using Clustal W. The amino acids identical to the consensus are shown as dots and the gaps are filled with dashes (-). Numbers above the alignment represent positions of amino acids. The key amino acids responsible for interaction with lipid ligands are directly indicated beneath the alignment.

### 3.4 Intronless FABP genes

Single-exon FABP genes were found in the following species: *E. granulosus*, *E. multilocularis*, *Strongyloides ratti*, *P. humanus corporis*, *R. prolixus*, *Strongylocentrotus purpuratus* and *Saccoglossus kowalevskii* (Table 4-14). Expression of most of these intronless genes was confirmed either by transcriptome analysis or analysis of ESTs. With exception of *R. prolixus*, the rest encoding intronless FABPs also encoded other FABP genes with two or more exons. It is worth mentioning that the intronless FABP gene architecture dominated in other three metazoans *S. ratti*, *P. humanus corporis* and *S. kowalevskii*.

Alignment of the intronless FABP has revealed that *R. prolixus*'s FABP gene contained intact key residues which are important in defining fatty acid binding (Jakobsson et al., 2003), while absence or alterations in these sites occurred in the other one-exon FABP genes (Fig. 4-14). This suggests that the *R. prolixus* FABP is still functional, but the others may no longer be so. Alternatively, they have different binding spectra for fatty acids in comparison with those that have been characterised.

### 3.5 Alternative splicing in FABP genes

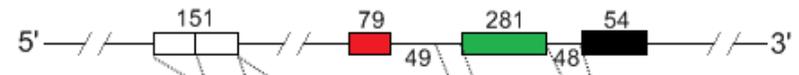
Alternative splicing is an important post-transcriptional modification in eukaryotic pre-mRNAs, accounting for the complexity and variety of proteomes. FABP genes underwent alternative splicing to generate isoforms in several invertebrates including *S. mansoni*, *C. elegans*, *Tribolium castaneum*, *Acyrtosiphon pisum*, *Apis mellifera* and *Drosophila melanogaster*. Furthermore, all the transcripts derived from alternative splicing were confirmed by transcriptomic data.

Comparative analysis of FABP gene structure and transcripts revealed that FABP pre-mRNAs were alternatively spliced in different patterns (Fig. 4-15). In *S. mansoni*, FABP variants were produced by exon skipping, which has been described in cestodes (Agorio et al., 2003; Gauci and Lightowlers, 2001; Spiliotis et al., 2006).

**A**

9a SL1 --- 152 aa (2)

9b SL1 --- 246 --- poly (A) 137 aa (5)

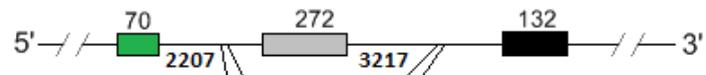
**B**

a --- 31 --- 48 --- poly (A) 137 aa (1)

b --- 76 --- poly (A) 161 aa (0)

c --- 76 --- poly (A) 136 aa (7)

d --- poly (A) 137 aa (3)

**C**

a --- 31 --- 51 --- poly (A) 117 aa (2)

b --- poly (A) 130 aa (101)

c --- poly (A) 157 aa (6)

**Figure 4-15 Alternative splicing in invertebrate FABP genes.**

Typical alternative splicing patterns in *C. elegans* (A), red flour beetle (B) and fruit fly (C) are represented. *C. elegans* LBP-9 pre-mRNA is spliced to generate two variants 9a and 9b by addition of a short spliced leader sequence (SL1: 5'-GGTTTAATTACCCAAGTTTGAG-3') at the 5' end. Blank or filled boxes and straight lines represent exons and introns, respectively, and a poly (A) stretch present in each FABP cDNA clone or EST sequence is directly shown. In each group, an annotated FABP gene is placed above the variants that are indicated by a, b, c or/and d. Numbers above the boxes and under the lines show the sizes of corresponding exons and introns, respectively. The sizes of exons with a different size are indicated above the corresponding exons in the spliced transcripts. The length of variants is also shown after each transcript and the number of EST is shown in the brackets.

Both arthropods *A. pisum* and *A. mellifera* utilized the same approach to yield different FABP transcripts. In contrast to the exon exclusion mechanism in *S. mansoni*, spliced leader *trans*-splicing (SL *trans*-splicing), a type of alternative splicing whereby a spliced leader serves as a mini-exon to be added onto 5' pre-mRNA ends (Hastings, 2005), was used in *C. elegans* FABP (also known as lipid binding proteins) mRNA precursors (Fig. 4-15).

Both *T. castaneum* and *D. melanogaster* genomes contained only one FABP gene locus but variants were found in transcriptome or EST datasets. Mapping of these transcripts revealed that another alternative splicing mechanism, intron retention, was involved in the post-transcriptional splicing of FABP genes together with exon skipping (Fig. 4-15). These results indicate that both invertebrates have adopted alternative splicing rather than gene duplication to produce a variety of FABP genes. Noticeably, the last exons of the FABP genes were retained during alternative splicing in all the invertebrates studied except *D. melanogaster* who also used the partial sequence of the second intron as the last exon to generate FABP isoforms.

## 4. Discussion

In this study, it was shown that the number of the FABP genomic loci identified was remarkably variable, from 1 in several invertebrates to 15 in amphioxus. Interestingly, no FABP loci were identified in both *Hydra magnipapillata* and *N. vectensis* genomes but present in the simplest known free-living organism, *T. adhaerens*, which is considered as a basal metazoan (Dellaporta et al., 2006; Halanych, 2004). Consistent with the gene structure, *T. adhaerens* FABP genes seems to be a prototype of this family. *T. adhaerens* is only the recognized species in the Phylum Placozoa and, although it is still controversial, this invertebrate seems to phylogenetically form the 'eumetazoan' clade with cnidarians and bilaterians (Halanych, 2004; Srivastava et al., 2008). Moreover, the Placozoa lineage has diverged before segregation of Cnidaria

and Bilateria but after the divergence of demosponges (Srivastava et al., 2008). The deficiency of FABP expression in cnidarians may be explained by gene loss but the possibility remains that these species may express extremely heterogeneous FABPs and investigations of more Cnidaria species are required. Furthermore, each of three invertebrate genomes contained two loci to encode the same proteins, suggesting that they might have arisen from gene duplication. In comparison with *E. granulosus* FABPs, the current gene set in *E. multilocularis* may have been generated before the speciation of *Echinococcus* species (data not shown), supporting the idea that FABP gene duplication may have occurred in their common ancestor. This finding does not fully support the previous assumption that *E. granulosus* FABPs 2 and 4 arose from a recent duplication event (Esteves et al., 2003).

The ancestral gene of FABPs might evolve from lipocalins and undergo the first duplication approximate 930 million years ago with subsequent duplications and divergence (Ganfornina et al., 2000; Schaap et al., 2002). The FABP sequences annotated in this study were heterogeneous with regard to length, composition and identity, possibly driven by needs for the transport of numerous different fatty acids (Schaap et al., 2002). In contrast to fifteen FABP copies in *B. floridae*, only one FABP locus was predominantly distributed in the arthropods including mosquitoes. Relative to one locus in *A. aegypti* and *C. pipiens quinquefasciatus*, a malaria mosquito, *A. aegypti*, had two copies for FABP genes that resided on the same scaffold. It is noticeable that these mosquito sequences have been identified as allergens. Secondary structural prediction with high confidence showed that all of these mosquito allergen genes had two alpha helices and ten beta sheets typical of FABP structures (data not shown). In addition to the conserved secondary structures, they contained fatty acid binding-related key residues except Val-Asp instead of Pro-Asp (Fig. 4-14). We therefore propose that these allergens are functionally specified FABPs. This idea is in agreement with the facts that lipid-binding proteins from *mites*

(Jeong et al., 2005; Puerta et al., 1999) and nematodes (Kennedy, 2000b; Spence et al., 1993) are recognized by IgE. Then it raises the question of why these mosquito FABPs have evolved to be allergic. It is still not clear but the fact that a high level of FABP is present in normal human blood suggest that mosquitoes can acquire bound fatty acids directly via blood feeding (Bao et al., 2011; Cabre et al., 2008).

In contrast to relatively analogous structures for the mammalian FABP genes, invertebrate FABP genes were genetically organized in a wide range of patterns with a dominance of the four-exon three-intron structure. A key finding in this study was that invertebrate FABP genes tended to be intron loss with few intron gains during evolution. This idea is enhanced by the fact that in most of the invertebrate FABPs investigated have matched intron positions (Esteves and Ehrlich, 2006). Compared to the early branching invertebrate, *T. adhaerens*, FABP gene from cestodes and mosquitoes were intron-poor and the tendency of intron loss became more apparent in the Phylum Platyhelminth. Our findings strongly argue against the speculation that the first and second introns of FABP genes might have occurred later (Esteves et al., 2003). Surprisingly, a number of invertebrates encoded intronless FABPs with most of the key residues that participate in lipid binding being altered. However, here no evidence was provided to suggest that these FABPs remain functional. Such intronless FABPs have also been reported in several mammals and they have lost their capacity to bind ligands (Bonne et al., 2003; Prinsen et al., 1997; Treuner et al., 1994).

A wealth of data has revealed that numerous introns are present in early multicellular organisms and alterations of intron positions occur at a very low frequency during evolution (Irimia and Roy, 2008). It was shown that intron gain and intron loss occurred at various points in the past 100 million years and these genetic events were considered to be a very low process in some lineages (Irimia and Roy, 2008). Several mechanisms have been proposed for intron gain or loss (Belshaw and Bensasson, 2006; Roy and Gilbert, 2006). In comparison with a canonical three-intron structure,

the first intron (17/31) was more likely to be preferentially retained in two- or one-intron FABP genes in invertebrates. This suggests that reverse transcription followed by gene conversion may have been involved in the FABP intron loss as this mechanism tends to remove 3' introns from the genes (Belshaw and Bensasson, 2006). An analysis of 684 gene introns from eight organisms has showed that loss of most ancestral introns has occurred in worms and arthropods but not in humans (Rogozin et al., 2003). This result may give us some clues, but the selective forces that have driven intron loss in platyhelminths remain unclear.

Alternative splicing, a substantial mechanism for the modification of pre-mRNA, ubiquitously exists in nearly all eukaryotic organisms and accounts for the complexity and diversity of protein functions. In contrast to mammals where alternative splicing of FABP genes has rarely been observed, here it was shown that FABP genes in some invertebrates were alternatively spliced, leading to generation of FABP variants. In particular, these various transcripts were produced in different splicing patterns. In *C. elegans*, only FABP genes 5, 6 and 9 were confirmed to mature by means of SL *trans*-splicing using spliced leader 1 (SL1). There are two distinct spliced leader sequences, SL1 and SL2, and the former is used to mainly generate monocistronic pre-mRNA (Pettitt et al., 2010). It is estimated that approximately 70% of all genes in this free-living nematode are post-transcriptionally modified by this mechanism (Ross et al., 1995). It is still not clear why *C. elegans* FABP 1, 2, 3 and 8 pre-mRNAs are not matured via SL *trans*-splicing. Although the SL *trans*-splicing mechanism is also extensively present in the Phyla Cnidaria, Platyhelminth and Chordata (Hastings, 2005), the occurrence of this specific alternatively splicing in FABP genes was not observed in other invertebrates collected in this study. These results suggest that the SL *trans*-splicing modification in *C. elegans* FABPs may have been acquired during evolution.

# Chapter 5 Expression, purification and fatty acid binding characteristics of fatty acid-binding proteins of *Echinococcus multilocularis*

## Abstract

*E. multilocularis* is incapable of the *de novo* synthesis of long-chain fatty acids and needs to acquire these lipids entirely from their hosts. FABPs are small lipid chaperons that are involved in lipid metabolism. Here, we showed *E. multilocularis* had five loci encoding four distinct emFABPs, which were located on two different scaffolds. Among them emFABP3 and 4 were constitutively expressed in vesicles, protoscoleces and cell aggregates. In particular, emFABP3 appeared to be highly expressed in the vesicles. emFABPs had atypical gene structures and emFABP1 and 5 were intronless. Comparison of emFABPs and *E. granulosus* FABPs (egFABPs) revealed that the current genomic organization of FABPs might have formed before the speciation of *E. multilocularis* and *E. granulosus*. Although it lacked a signal peptide, native FABP3 was confirmed to be a secretory/excretory protein present in the hydatid fluid. *E. coli*- or yeast-derived recombinant FABPs appeared not to bind to fatty acids including DAUDA and retinol but exhibited weak binding to *cis*-parinaric acid.

## 1. Introduction

Fatty acid-binding proteins (FABPs) are small proteins with a mass weight of ~15kDa and, together with intracellular retinol- and retinoic acid-binding proteins, comprise an intracellular lipid binding protein (iLBP) subfamily of calycins. iLBPs are present in animals but not plants and fungi and are supposed to have arisen via gene duplication

and divergence after the separation of animals from fungi and plants, generating 16 iLBP types in vertebrates (Haunerland and Spener, 2004). FABPs were initially isolated from rat liver and other tissues in the 1970's (Ockner et al., 1972) and twenty years later invertebrate FABPs were identified in the muscle of the locust, *Schistocerca gregaria* (Haunerland and Chisholm, 1990). In platyhelminths, FABPs have been already reported in *Schistosoma mansoni*, *S. japonicum*, *S. bovis*, *Fasciola hepatica*, *F. gigantica*, *Clonorchis sinensis*, *Mesocestoides vogae* and *E. granulosus* (Alvite et al., 2008; Esteves and Ehrlich, 2006; Esteves et al., 2003; Lee et al., 2003).

FABPs seem to be expressed in a tissue- or stage-specific manner (Haunerland and Spener, 2004; Storch and Thumser, 2010). In *F. gigantica*, FABP1 is abundantly distributed in testes and spermatozoa, whereas FABP3 is mainly expressed in vitellines, eggs and caecal epithelium (Chunchob et al., 2010). *S. mansoni* FABP, also called Sm14, is distributed in the tegument tubercles and regions between the muscles and the body area (Moser et al., 1991). Similarly, *S. japonicum* FABPc is expressed in lipid droplets in males and in the vitelline eggs in females (Gobert et al., 1997). In *E. granulosus*, expression of FABP1 is found in protoscoleces (Esteves et al., 1993). Although they lack signal peptides, FABPs in some species have been shown to be secreted or released in some way. Recently, proteomic analyses have confirmed the presence of FABP in the *in vitro* ES products of *S. japonicum* (Liu et al., 2009a), *F. hepatica* (Morphew et al., 2007) and *F. gigantica* (Chunchob et al., 2010).

FABPs are characterized by binding to saturated and unsaturated long-chain fatty acids ( $\geq^{14}\text{C}$ ) with high affinity. Although FABPs have low homology at the nucleotide and amino acid levels, the tertiary structures are conserved, containing a central cavity by 10-stranded  $\beta$ -barrel and 2  $\alpha$  helices, which harbours bound ligand(s), and guaranteeing the similar binding capacity of heterogeneous FABPs (Jakobsson et al., 2003; Storch and McDermott, 2009). With a few exceptions, the majority of FABPs can bind a single fatty acid with its carboxylate group and conformational changes

seem to occur during fatty acid binding and release. Moreover, FABPs do not exhibit selective binding for a specific ligands although it is obvious that the binding affinities are positively associated with fatty acid hydrophobicity (Storch and Corsico, 2008). FABPs are involved in the trafficking and storage of lipids and closely linked with metabolic disorders and other lipid-related biological processes such as gene transcription regulation via the peroxisome proliferator-activated receptor gamma network. Since cestodes are unable to synthesize long-chain fatty acids (Smyth and McManus, 1989), FABPs are vital, playing an indispensable role in the absorption process of lipids from hosts.

In this study we showed that there were five genomic loci for four distinct FABP genes, designated as emFABP1, emFABP3, emFABP4 and emFABP5, residing on two different scaffolds of *E. multilocularis*. *Echinococcus* species shared the same gene structures and genomic organization of FABPs. Compared to emFABP1 and emFABP5, emFABP3 and emFABP4 were abundantly expressed in the vesicles, dormant protoscoleces and cell aggregates of vesicle primary cells. Yeast-derived recombinant FABP3 possessed weak binding ability with *cis*-parinaric acid but not with fluorescent fatty acid 11-[[[5-(dimethylamino)-1-naphthalenyl]sulfonyl]amino]-undecanoic acid (DAUDA) and retinol. The result of Western-blot analysis showed that the native FABP3 was present in the hydatid fluid, suggesting that it is a secretory/excretory protein.

## **2. Materials and methods**

### **2.1 Identification of FABP homologues in *E. multilocularis* and amplification of FABP genes**

*E. granulosus* FABPs (AAK12095.1 and AAK12096.1) and *S. mansoni* FABPs (Smp\_046800 and Smp\_095360) were used to interrogate the genome database of *E.*

*multilocularis* (<http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus>). The FABP gene structures were then inferred from our transcriptome data. The open reading frames of each FABP genes were determined using ORF finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) and sequence alignment and phylogenetic analysis were conducted using MEGA (Version 4). N-glycosylation sites and signal peptides were predicted using NetNGlyc 1.0 and SignalP 3.0, respectively. The putative secondary structures were modelled using Psipred.

Primers for amplification of the FABP genes were designed using Oligo (version 4.1). Amplicons of four FABP genes were generated using one step reverse transcription-PCR kit (Qiagen) according to the instructions. Briefly, 1.5 µg of total RNA (see 2.2, Chapter 2) were mixed with corresponding primer sets (Table 5-15), 2 µl of enzyme mix and 10 µl of 5× Buffer. The mixture was subject to incubation at 50°C for 30 min, 95°C for 15 min and 30 cycles of 94°C for 30 sec, 56°C for 30 sec and 72°C for 30 sec and a final extension at 72°C for 10 min. PCR products were separated on 1.2% agarose gel containing 0.005% ethidium bromide (EB), followed by purification using Gel Purification Kit (Qiagen) according to the manual. The bands of interest were cut out under ultraviolet light, followed by addition of Solution Q and incubation at 50°C to release the amplicon. The solutions were then loaded onto columns to allow DNA binding and then the target fragments were eluted in nuclease-free water. The purity and concentration of purified amplicons were determined using Nanodrop (Thermo Scientific).

## **2.2 Construction of prokaryotic expression vector**

Ligation of the purified PCR products with pET-SUMO vector (Invitrogen) was conducted in 10 µl reaction system comprising of 1.5 µl PCR product, 1.5 µl pET-SUMO, 1 µl T4 DNA ligase and 1 µl 10× Reaction Buffer. After gentle agitation, the mixture was briefly centrifuged and incubated at room temperature for 30min,

**Table 5-15 Primer sets using for amplification and expression**

Primer	Sequence (5'-3')	Target	Note
1F	ATG GAT GAC TTT CTG GGC ACC T	emFABP1	These primers were used for cloning and expression in <i>E. coli</i> .
1R	CTA TGA ACA TAA GAT TTG GTA TGC AAG A		
3F	ATG GAG GCG TTC CTC GGT AC	emFABP3	
3R	TTA CGA CAC CTT TGA GTA GGT TCG		
4F	ATG GAG CCA TTC ATC GGT AC	emFABP4	
4R	TTA CAT CCC TCT TGA GTA GGT TC		
5F	ATG GAT GAA TTC CTG GGA TCC	emFABP5	
5R	TTA ACT CAC CGT TTT TAG CAA ATC		
Y-F	<u>GGAATTC</u> ATGGAGGCGTTCCTC	emFABP3	These primers were used for expression in yeast. The restriction sites were underlined. The primer Y-R2 contains a stop codon but Y-R1 does not.
Y-R1 <sup>a</sup>	GCTCTAGAAACGACACCTTTGAGTAGGTTTC		
Y-R2 <sup>b</sup>	GCTCTAGATTACGACACCTTTGAGTAGG		

<sup>a</sup>The primers Y-F and Y-R1 were used for expression of His-tagged rFABP3 in yeast;

<sup>b</sup>The primers Y-F and Y-R2 were used for expression of rFABP3 in a native form in yeast.

followed by transformation. The ligated product was briefly mixed with competent *E. coli*, incubated on ice for 30 min and then at 42°C for 30 sec, and agitated at 220rpm at 37°C for 1h after addition of 800 µl of pre-heated MSO media. 50 µl aliquots of the media were put onto LB agar plates with a final concentration of 50 µg/ml kanamycin and placed inversely in a 37°C incubator overnight.

The transformants were inoculated into 5 ml LB containing 50 µg/ml kanamycin and incubated at 250rpm at 37°C overnight. The cells were collected by centrifugation at 12,000×g and plasmids were extracted. Briefly, the pellets were completely resuspended in Solution I (50mM Tris pH 8.0 with HCl, 10mM EDTA) containing RNase A, followed by addition of Solution II (200mM NaOH, 1%SDS). Solution III (3.0M Potassium Acetate, pH5.5) was then quickly added, mixed, placed on ice for at least 5min, and spun at 12,000×g for 10 min. The supernatant was transferred into new tubes, mixed with a 2.5× volume of ethanol, and then incubated at -20°C for 30 min. After centrifugation the supernatant was discarded and the pellets were washed with 75% ethanol, followed by incubation at room temperature to air-dry for at least 10 min. The pellets were resolved in 50 µl nuclease-free water and the purity and concentration determined by Nanodrop. The positive plasmids comprising of FABP fragments, designated as pET-SUMO/FABP, were identified using PCR as described above. Insertions of FABP fragments were then further confirmed through sequencing.

### **2.3 Expression and purification of FABPs in *E. coli***

For expression plasmids containing the target with a correct orientation were transformed into competent BL21 cells as described above. The transformants were inoculated into kanamycin-containing LB media and induced by addition of a final concentration of 1mM IPTG when the OD value reached at 0.4~0.6. After 4h agitation at 250rpm at 37°C cells were collected by centrifugation, washed three times with cold PBS (pH 7.2), and resuspended in Native Purification Buffer (Invitrogen) containing

10 mM imidazole and 1 mg/ml lyzosome. After incubation on ice for 30 min, the resuspended cells were sonicated, centrifuged at 5,000×g at 4°C for 1h, and the supernatant was transferred to new tubes for late use. In this study, plasmid pET-SUMO/CAT (Invitrogen) was used as a control.

Before loading onto a nickel-charged agarose column, the lysate was filtered to get rid of remaining particles. Purification was performed according to the manufacturer's instructions with a few modifications. Firstly, the resin was suspended thoroughly and gently and aliquoted into the column, followed by three washes with 5 multiple volumes of distilled water. The beads were equilibrated by two washes with Native Purification buffer containing 10 mM imidazole. The filtered lysate was loaded onto the column, gently rotated at 4°C for 30 min, and eluted at 1 ml/min. The beads immobilized with the target proteins were thorough rinsed 5 times in the washing buffer containing 20 mM imidazole and the recombinant SUMO/FABPs (rSUMO/FABPs) were released using elution buffer containing 150 mM imidazole.

A buffer change was conducted using centrifugal concentrator (3K, Millipore). The proteins were finally resolved in 20 mM Tris buffer or PBS (pH7.2). The purity of these proteins was evaluated by SDS-PAGE. 15% gel was utilized for separation of rSUMO/FABPs and visualized using Protein Staining Solution (Ferment) following the manufacturers recommendation.

#### **2.4 Digestion of rSUMO/FABPs to release rFABPs**

SUMO protease was used to release rFABPs and rSUMO/CAT was utilized as control. The reaction was optimized as follows: 20 µg rFABP and 10 units of SUMO protease were combined with the different concentration of NaCl: 150 mM, 200 mM, 250 mM and 300 mM; the mixture was then incubated at 4°C, 15°C, 23°C and 37°C for 6 h and aliquoted at 2-h intervals for SDS-PAGE analysis as previously described.

## **2.5 Construction of yeast expression vectors**

Using sequence-validated pET-SUMO/FABP3 DNA as a template, the fragments containing two different restriction sites were amplified by PCR under the conditions as described above. The amplicons were purified by Gel Purification Kit (Qiagene) and the concentration was determined as mentioned previously, followed by double-enzyme digestion at 37°C for at least 4h. Simultaneously, pPICZ alpha-A DNA (kept in our lab) was also digested using the same enzymes and purified. Insertion of FABP3 into digested plasmid vector was performed at 4°C overnight, followed by transformation into competent DH5 alpha cells (NEB) as described previously. The cells were spread onto low salt LB agar plates containing 25 µg/ml Zeocin and then incubated at 37°C for overnight. Plasmid DNA was extracted as previously mentioned and the positive construct, designated as pPICZ-FABP3, was screened out by PCR and then sequenced.

## **2.6 Transformation and identification of positive clones harbouring FABP3**

For expression, *Pichia pastoris* strains X-33 and GS-115 were used as hosts and the competent cells for transformation were prepared according to the instructions. When OD value at 600 nm reached at 1.3, the cells were collected by centrifugation at 4°C, followed by washes with sterile ice-cold water and then with sterile ice-cold 1 M sorbitol. The cells were finally resuspended in 1 ml of ice-cold 1 M sorbitol. pPICZ-FABP3 DNA was linearized using enzyme *Sac I* and subject to phenol-chloroform extraction after assessment of linearization of the digest on 1% agarose gel. Plasmid DNA was then precipitated by addition of 1/10 volume of 3 M sodium acetate and 2.5 volumes of absolute ethanol. Consequently, the pellet was washed with 80% ethanol, air-dried, and resuspended in nuclease-free water.

10 µg of the linearized DNA was added into 50 µl of freshly prepared yeast competent cells and mixed by pipetting. The mixture was then transferred to a pre-chilled 0.2 cm cuvette and placed on ice for 5 min. Electroporation transformation was performed using MicroPulser (Bio-Rad) with default parameters. After one pulse, 1 ml of ice-cold sorbitol was immediately added into the cuvette. The diluted cells were incubated at 30°C for 2h without shaking and then employed onto YPDS plates containing 100 µg/ml Zeocin, followed by incubation at 30°C until colonies were visible. In the transformation, blank plasmid pPICZ alpha-A was also used as control.

The integration of FABP3 into the yeast genome was validated by PCR. Yeast DNA was recovered as previously reported (Zhang et al., 2010). Yeast cells were firstly suspended in the lysis buffer (50 mM sodium phosphate, pH 7.4, 1 mM PMSF, 1 mM EDTA and 5% glycerol) and incubated at 85°C for 30 min. After centrifugation, the supernatant containing the genomic DNA was pipetted to new tubes. The reaction system was constituted as follows: 2 µl 10x buffer, 0.5 µl 10mM dNTP, 0.5 µl of 5' AOX1 primer and 3'AOX1 primer each (Invitrogen), 0.2 µl Taq DNA polymerase (NEB) and 1 µl yeast lysis supernatant, and PCR was performed under the following conditions: 95°C for 5 min, followed by 30 cycles of 95°C for 1 min, 54°C for 1 min and 72°C for 1 min, and finally 72°C for 7 min. PCR products were resolved on 1.2% agarose gel. Nuclease-free water and genomic DNA of integrants with blank plasmid pPICZ alpha-A were utilized as controls.

## **2.7 Expression and purification of rFABP3 produced by yeast**

To optimize expression conditions, the phenotypes of PCR-validated yeast transformants were determined. Positive integrants were patched first on minimal methanol histidine plates (MMH) and then on minimal dextrose histidine plates (MDH) and incubated at 30°C for 2~3 days. To obtain individual pure clone isolates,

single colonies were streaked on yeast extract peptone dextrose plates (YPD) and incubated at 30°C until colonies formed.

An individual clone was inoculated into buffered glycerol-complex medium (BMGY) and agitated at 29°C at 270rpm until the OD value at 600nm was in a range from 2 to 6, followed by collection of cells. The cell pellet was then resuspended in buffered methanol-complex medium (BMMY) for expression induced by addition of absolute methanol to a final concentration of 0.5% at 24-h intervals. Simultaneously, 1 ml of media was aliquoted for SDS-PAGE analysis before addition of methanol. These samples at different time points were concentrated using resin (Stratagene) before loading onto the gel. Briefly, 100 µl expression media was gently mixed with 10 µl resin and incubated at room temperature for 1 min. After one wash with PBS, the resin coupled with proteins was resuspended in loading buffer. As previously 15% gel was used for SDS-PAGE and transformants harbouring blank plasmid pPICZ alpha-A were utilized as control.

To optimize precipitation by ammonium sulphate, expression media supernatant was brought to 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90% saturation, respectively, at 4°C with a gentle agitation for 2h. After centrifugation, the pellets were retained for electrophoresis analysis as mentioned above. As for rFABP3 tagged with 6×His at the C end, it was purified using two approaches. The media supernatant containing rFABP3 was brought to 80% saturation of ammonium sulphate and incubated as described above and the pellets were collected and resolved in Native Purification Buffer (Invitrogen). Before loading onto nickel-coupled columns, washing on a centrifugal concentrator was performed to get rid of trace ammonium sulphate. Further purification with the columns was conducted according to the methods mentioned previously.

rFABP3 without a His tag was purified by combination of ammonium sulphate precipitation and chromatography techniques. After precipitation by ammonium sulphate, the pellets were resolved in starting buffer (20 mM Tris, 50 mM NaCl, pH 8.3) and concentrator (Millipore) was used to decontaminate ammonium sulphate. Before loading to chromatography columns, samples were centrifuged at 4°C at 10,000×g for 10 min to get rid of any particles. An anion exchange column (HiTrap Q FF, GE Healthcare) was equipped with AKTAprimer plus (GE Healthcare) to load, wash and elute proteins with pre-programmed parameters. The bound proteins were eluted using a salt gradient with the starting buffer containing 0 to 1 M NaCl at 1 ml/min and different fractions were analyzed on 15% SDS-PAGE. In order to screen out better approaches for purification, a hydrophobic interaction chromatography column (HiTrap Phenyl FF (high sub), GE Healthcare) was also connected to AKTAprimer plus to purify the target protein. As previously, the pellets were suspended in binding buffer (50 mM sodium phosphate, 1.0 M ammonium sulphate, pH 7.0) and subject to loading onto the column, followed by wash with 30 volumes of the binding buffer. The proteins were then eluted in a gradient manner in 0 to 100% of 50 mM sodium phosphate, pH 7.0, at 5 ml/min and different fractions were separated on 15% SDS-PAGE.

The diluted fractions containing rFABP3 were collected together and then concentrated using concentrator (Millipore). The mixture was further polished using size exclusion chromatography (Superdex-75, GE Healthcare) at 1 ml/min. Filtered and degassed PBS (pH 7.4) was used as running buffer.

## **2.8 Delipidation of rFABPs**

Pure rFABPs were delipidated using Lipidex-1000 (PerkinElmer). First Lipidex-1000 was completely suspended and aliquoted to a new tube, followed by two washes with distilled water and three washes with 20 mM Tris buffer (pH 7.2). Protein samples

were then combined with an equal volume of Tris buffer-equilibrated Lipidex-1000, end-over-end mixed at 37°C for 45 min and centrifuged at 4°C at 1,000×g for 5min. The supernatant containing recombinant FABP was transferred into new tubes. The buffer of delipidated FABPs was changed into PBS (pH 7.4) using concentrator as mentioned above.

## **2.9 Removal of endotoxin in recombinant FABPs**

For *E. coli*-derived FABPs, endotoxin was removed using two different approaches. First FABP samples were combined with 1% of Triton-114 on ice and agitated by end-over-end mix at 4°C for 30 min. Afterwards the mixture was quickly placed in a 37°C incubator for 10 min, followed by centrifugation at 12,000rpm at 4°C for 5min. The supernatant was then transferred into new tubes and repeated additional four rounds of an incubation-and-centrifugation cycle. Polymixin B-immobilized gel (Berriman et al.) was employed to further remove endotoxin from the preparations. The gel was washed with 5 volumes of distilled water and then with 10 volumes of 1% sodium dextyocholate, followed by equilibration by two washes with 5 volumes of sterile PBS (pH 7.4). The protein samples above were loaded onto the equilibrated columns and the flow-through containing the target protein was collected. The gel was regenerated by washes with 10 volume of 1% sodium dextyocholate and then with 5 volumes of sterile PBS (pH 7.4). The protein solutions were applied again for a total of five rounds of purification.

With regard to rFABP3 produced in yeast, the removal of endotoxin was conducted only using polymixin B-immobilized gel (Berriman et al.) as described above.

The endotoxin-removed FABPs were concentrated using a concentrator as mentioned above and the concentration was determined by Bradford method according to the manual (Bio-Rad).

## **2.10 Endotoxin test**

The level of endotoxin in protein solutions was determined using Gel Clot Endotoxin Assay Kit (GenScript) according to the instructions. Briefly, 10  $\mu$ l of the pre-diluted protein samples were completely and gently mixed with 100  $\mu$ l of reconstituted Limulus amoebocyte lysate (LAL), followed by incubation for 1 h at 37°C. The positive reaction was determined by formation of a solid gel clot. In this assay endotoxin-free water was used as a negative control and lipopolysaccharide (LPS) as a positive control.

## **2.11 Fatty acid binding assay**

To check the binding capacity of recombinant FABPs, a binding assay was conducted using fatty acids including DAUDA, retinol and *cis*-parinaric acid (Cayman Chemical). For DAUDA and *cis*-parinaric acid, 1  $\mu$ M of protein was mixed with 1  $\mu$ M of fatty acid by pipetting and then transferred in a cuvette. The fluorescent intensity was recorded from 350 nm or 400 nm to 600 nm after excitation at 345 nm and 320 nm, respectively, using fluorescence spectrometer (Varian). Because it easily degrades in aqueous solutions, retinol was directly added into a cuvette containing 1  $\mu$ M of protein in PBS, briefly pipetted, and the emission was read from 400 nm to 600 nm after excitation at 350 nm. In this study PBS and bovine serum albumin (BSA) were used as negative and positive controls, respectively.

## **2.12 Preparation and purification of anti-FABP3 rabbit serum**

The immunization of rabbits with yeast-expressed rFABP3 was conducted by Harlan Laboratories, Leicester, UK. Before immunization, animals were pre-bled. Boosts were given 14d and 28d post-immunization and 100  $\mu$ g of pure rFABP3 antigen emulsified with adjuvant Titermax were administered each time. The immunized rabbit was kindly slaughtered 7d post-boost 2.

The serum was brought to 40% saturation of ammonium sulphate and gently agitated at 4°C overnight, followed by centrifugation at 4°C at 5,100rpm for 30 min. The pellet was resolved in Tris Buffer Saline (20 mM Tris, 150 mM NaCl, pH 7.4, TBS). The resuspended serum was dialyzed to remove ammonium sulphate at 4°C overnight. The concentration was determined by Bradford as previously described.

The anti-FABP3 titre was determined by enzyme-linked-immunosorbent serologic assay (ELISA). rFABP3 (5 µg/ml) was coated and incubated at 4°C overnight. After three washes with PBS containing 0.05% Tween-20, the coated plate was blocked in 1% BSA-containing PBS (pH 7.4, B-PBS) and incubated with diluted sera in B-PBS at room temperature for 1h. For the blank wells, B-PBS was added. The plate was rinsed as previously, followed by addition of peroxidase-conjugated anti-rabbit IgG (Sigma) 10,000 diluted in B-PBS and then incubation at room temperature for 45 min. After wash, the substrate solution (Sigma) was added into each well for colour development in the dark at room temperature for 15 min, followed by addition of H<sub>2</sub>SO<sub>4</sub>. The OD values were read using microplate reader (Bio-Rad).

### **2.13 Western blot**

The protein samples were resolved on 15% SDS-PAGE gel and transferred onto Hybond nitrocellulose membrane (Amersham, GE). Transfer was conducted at 90 voltages for 70 min using cool transferring buffer (25 mM Tris, 190 mM glycine and 20% methanol). The protein-bound membrane was subject to blocking in TBS containing Tween-20 (20 mM Tris, 150 mM NaCl and 0.1% Tween-20, TBST) and 1% BSA. After three washes with TBST, the membrane was incubated with rabbit purified anti-FABP3 antibody 1:1000 diluted in BSA-containing TBST at 4°C overnight with gentle agitation, followed by incubation with goat horseradish peroxidase-conjugated anti-rabbit IgG (1:2000, Sigma) at room temperature for 60 min.

The colour was developed using 3, 3', 5, 5'-Tetramethylbenzidine (TMB) Liquid Substrate System (Sigma).

## 3. Results

### 3.1 Identification and gene structures of emFABPs

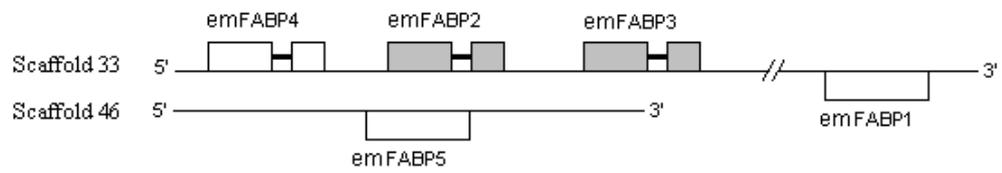
Using *E. granulosus* FABPs (AAK12095.1 and AAK12096.1) and *S. mansoni* FABPs (Smp\_046800 and Smp\_095360) as queries, we obtained five potential homologues with an e-value of less than 1e-10. With comparison of *E. granulosus* FABPs, their open reading frames were determined and validated by our transcriptome data (<ftp://ftp.sanger.ac.uk/pub/pathogens/Echinococcus/multilocularis/transcriptome/>).

These FABP genes were designated as emFABP1, 2, 3, 4 and 5, which were located on two different scaffolds (Fig. 5-16 A) It was notable that emFABP2, 3 and 4 were on the forward strand, while emFABP1 and 5 on the reverse. Alignment analyses showed that emFABP2 and emFABP3 encoded an identical protein and the others being distinct. In addition, only emFABP3 and emFABP4 shared the highest identity, up to 73% (Table 5-16).

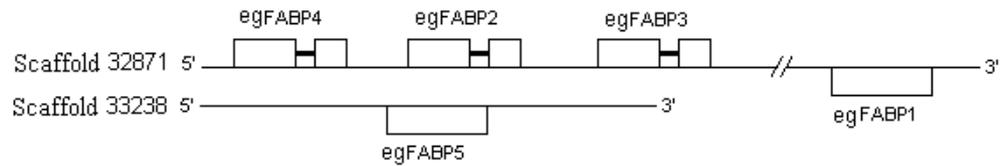
Unlike mammalian FABP genes that commonly comprise four exons and three introns (Schaap et al., 2002), emFABPs had different gene structures and genomic organizations. Three emFABPs 2, 3 and 4 were organized in two exons and one intron, whereas emFABP1 and 5 were intronless. Further comparison of introns and untranslated regions (UTRs) of both emFABP2 and 3 showed that the two introns and the upstream region (from -615 to ATG) were completely identical. It was remarkable that the length of the last exon of emFABPs and egFABPs was invariable, which is also observed in most of other invertebrate and vertebrate FABPs (see Chapter 4).

Structural prediction showed that emFABP1, 3 and 4 had typical secondary structure elements except emFABP5, two alpha helixes and ten beta sheets (Fig. 5-17).

**A**



**B**



**Figure 5-16 Schematic presentation of genomic organization and gene structures of *Echinococcus* FABPs.**

The exons are depicted as boxes and the introns as thick short lines between boxes. The grey-filled boxes in **A** represent the genes that are identical at the nucleotide level. The loci for *E. multilocularis* (**A**) and *E. granulosus* (**B**) FABPs are localized on the forward strand (above the lines) or on the reverse strand (beneath the lines)

**Table 5-16 Comparison of emFABPs.**

	Length (aa/ mw <sup>a</sup> )	Identity (%)		
		FABP1	FABP3	FABP4
emFABP1	143/16.21			
emFABP3	133/15.10	42		
emFABP4	133/15.36	45	<b>73</b>	
emFABP5	124/14.00	37	40	40

<sup>a</sup>aa: amino acid; mw: molecular mass weight expressed as kDa.

None of the emFABP genes contained known signal peptides and O-linked glycosylation sites. It was predicted that there were one (Asn<sup>128</sup>) and two (Asn<sup>34</sup> and Asn<sup>81</sup>) potential N-linked glycosylation sites in emFABP1 and 5, respectively, but not in emFABP3 and 4.

### **3.2 Comparison of emFABPs and *E. granulosus* FABPs (egFABPs)**

We took advantage of the genomic sequence of *E. granulosus* to analyze the evolutionary relationship of FABPs through comparison of genomic organisations, exons and introns. FABP homologues of *E. granulosus* were firstly identified through interrogating the genome database using four emFABPs as queries (<http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus>). As a result, five orthologues with low e-values (< 1e-10) were confirmed and the nucleotide identities ranged from 80% to 99% (Table 5-17). Interestingly, the genomic organization of egFABPs was similar to that of emFABPs and egFABP homologues had the same gene structure as corresponding emFABPs (Fig. 5-16 B). However, egFABP2 and egFABP3 were heterogeneous at the nucleotide and amino acid levels. Moreover, unlike emFABP2 and 3, the upstream regions (from -615 to ATG) of egFABP2 and egFABP3 had 84% identity. It was worth mentioning that FABP1, 4 and 5 in *E. multilocularis* and *E. granulosus* were conserved, whereas emFABP3 and egFABP3 only shared 63% identity at the amino acid level (Table 5-17).

Since introns tend to easily mutate and intron gain or loss appears to be a very slow process (Irimia and Roy, 2008), intron compositions and structures are informative markers for the evolutionary history of orthologues and paralogues across species or in individual species. Using the introns of emFABPs and egFABPs, the evolutionary relationship of these FABP-encoding genes was further investigated. Although the clade of egFABP2-intron and egFABP3-intron was poorly supported, the position



**Table 5-17 FABP homologues in *E. granulosus***

Homologue	Gene	Length (aa)	Identity (% , nu/aa) <sup>a</sup>	E-value	Position/strand
egFABP1	emFABP1	143	98/100	3.2e-89	Egu.pathogen_EMU_contig_32871/-
egFABP2 <sup>b</sup>	emFABP3	133	91/95	2.6e-73	Egu.pathogen_EMU_contig_32871/+
egFABP3		133	80/63	1.2e-44	Egu.pathogen_EMU_contig_32871/+
egFABP4 <sup>c</sup>	emFABP4	133	99/100	4.1e-78	Egu.pathogen_EMU_contig_32871/+
egFABP5	emFABP5	124	99/100	2.2e-77	Egu.pathogen_EMU_contig_33238/-

Note: <sup>a</sup>Identity to corresponding emFABPs at the nucleotide (nu) or amino acid (aa) level;

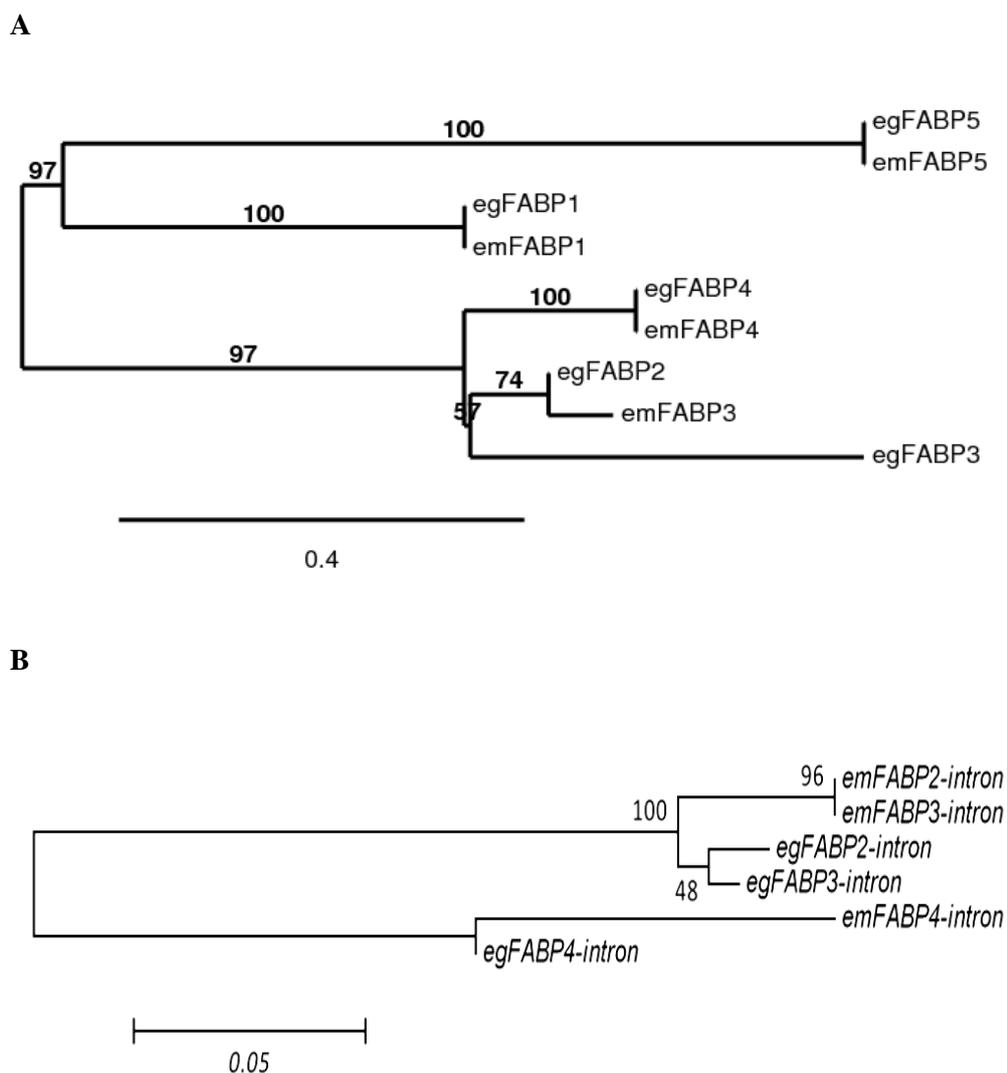
<sup>b</sup>With 98% identity to reported *E. granulosus* FABP1 (AAK12096.1);

<sup>c</sup>Identical to reported *E. granulosus* FABP2 (AAK12095.1).

between emFABP3-intron and the egFABP2-intron lineage was clearly resolved; furthermore, emFABP4-intron and egFABP4-intron formed one clade different from the cluster comprising of emFABP3-intron, egFABP2-intron and egFABP3-intron (Fig.5-18 **B**). These results are consistent with Echinococcus FABP positions in the evolutionary tree inferred from amino acid sequences (Fig.5-18 **A**). These results demonstrate that egFABP2 and egFABP3 may have derived from the same origin as emFABP3, and that egFABP4 and emFABP4 may have descended from a common ancestor. It is possible that the significant divergence of egFABP3 may have resulted from mutation during evolution after speciation.

### **3.3 FABP mRNA levels across life stages**

*E. multilocularis* has three different developmental stages. Recently, it has been demonstrated that the primary cells derived from vesicles that are comprised of neoblasts, a type of stem cells, enable to form cellular aggregates in the early cultivation phase and later develop mature vesicle with the laminated layer under specific conditions (Spiliotis et al., 2008). It is worth investigating the transcriptional status of four emFABPs across the life cycle. Here we used our transcriptome data to evaluate their mRNA levels in the cellular aggregates, vesicles and dormant protoscoleces. Technically, it is impossible to distinguish the reads mapped to emFABP2 and emFABP3 because these two genes including the 5' UTR are identical. Therefore the reads mapped to both emFABP2 and emFABP3 were summed up for calculation of emFABP3 abundance. The expression levels were transformed into reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008). For easy presentation the levels of each emFABPs in the different samples were normalized to the level of emFABP4 in each sample.



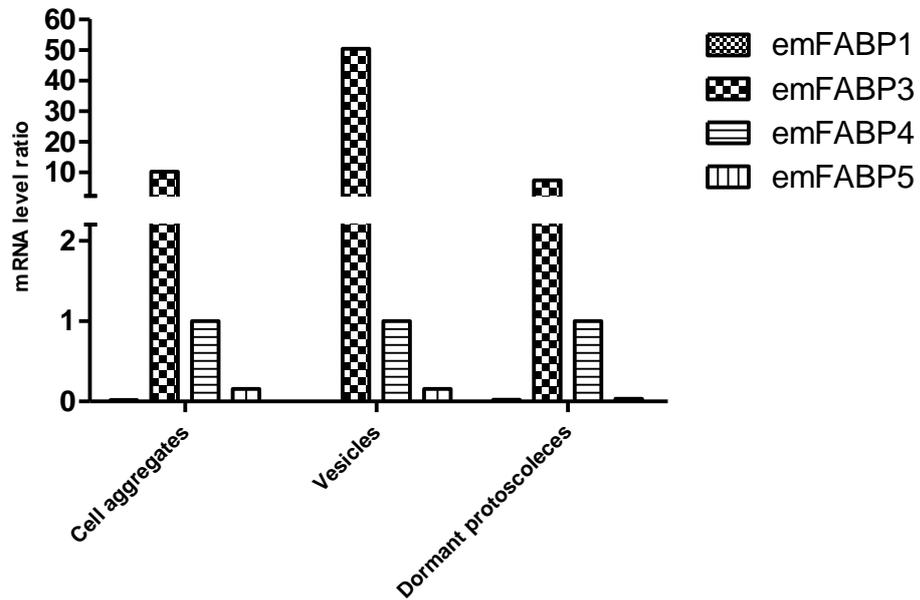
**Figure 5-18 Evolutionary relationship of Echinococcus FABPs inferred from amino acid or intron sequences.**

The amino acid and intron sequences were aligned using Clustal W with default parameters and checked by eyes. Maximum-likelihood and Neighbour-joining were used to construct phylogenetic trees inferred from amino acid (**A**) and intron (**B**) sequences, respectively. The trees were tested by bootstrap with 1,000 replications. The bootstrap values are shown on or below the branches and the substitution frequency of the branch is also indicated below the trees.

In all the samples emFABP3 was highly expressed while emFABP1 and 5 had low expression levels. In particular, emFABP3 appeared to be remarkably up-regulated in the vesicles (Fig. 5-19). In contrast to the emFABP3 mRNA level, emFABP4 mRNA was approximately 10 times less in the cell aggregates and dormant protoscoleces and 50 in the vesicles. In summary, the elevated expression of emFABP3 may provide a clue that it plays a role in the infection of the vesicles.

### **3.4 Recombinant FABPs produced by *E. coli* and yeast**

All of emFABP genes were subcloned into pET-SUMO vector and induced to express in *E. coli* hosts. It was shown that purified recombinant FABPs fused with a 6×His tag and a SUMO moiety (~13 kDa) at the N terminal (rSUMO/FABPs) had the same molecular mass weights as theoretical ones (Fig. 5-20 A). rSUMO/FABP3 and 4 were present mainly in a soluble form (lanes 2 and 3, Fig. 5-20 A), whereas rSUMO/FABP1 and 5 (lanes 1 and 4, Fig. 5-20 A), especially rSUMO/FABP5, in inclusive bodies. It was noted that some ambiguous bands with molecular mass weights of nearly 20 kDa and less than 14.3 kDa were observed in rSUMO/FABP1 (arrowed in the lane 1, Fig. 5-20 A), possibly due to the instability and easily to hydrolyze. Moreover, there existed a single band with a molecular mass weight of more than 55.6 kDa in each rSUMO/FABP3, 4 and 5 (marked with arrowheads, Fig. 5-20 A). Based on their molecular mass, these proteins may be the dimers of individual rSUMO/FABP.

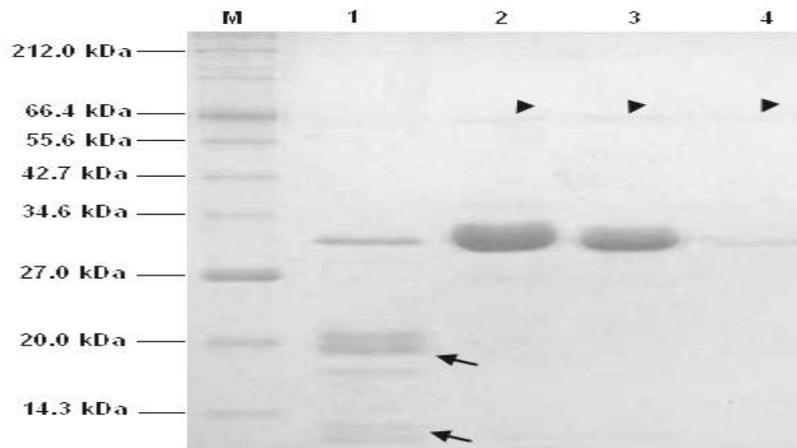


Developmental stage	total mapped reads	Num. of reads mapped to each gene			
		FABP1	FABP3	FABP4	FABP5
Cell aggregates	21311351	43(0.016)	25161(10.25)	2454(1)	362(0.158)
Vesicles	22460249	15(0.002)	264740(50.42)	5250(1)	760(0.155)
Dorm. protoscolecuses	13510969	33(0.020)	11562(7.47)	1547(1)	48(0.033)

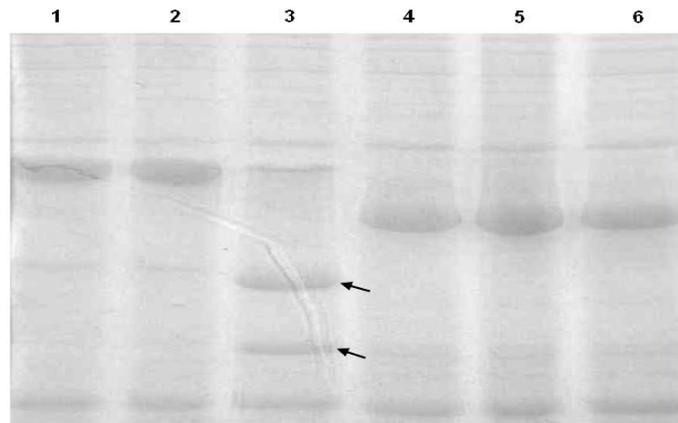
**Figure 5-19 mRNA levels of emFABPs in vesicles, cell aggregates and dormant protoscolecuses.**

In the table below the figure, the number of reads mapped to each emFABP genes is shown and the relative abundance of individual emFABP (numbers in the brackets) is normalized to the emFABP4 level in the corresponding sample.

**A**



**B**

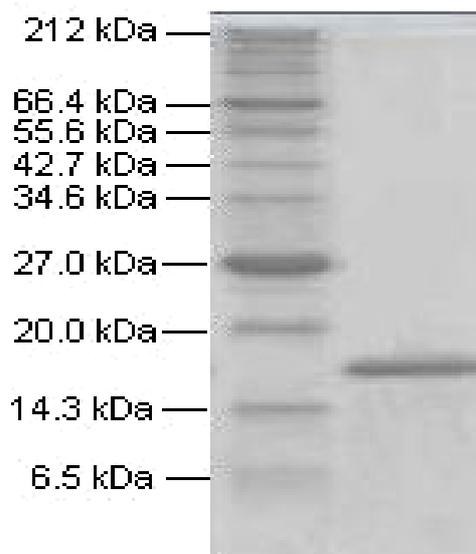


**Figure 5-20 SDS-PAGE analysis of rSUMO/FABPs and efficacy of SUMO protease.**

Purified *E.coli*-produced FABP1, 3, 4 and 5 (from lane 1 to 5 in **A**) were resolved on 15% SDS-PAGE gel. The molecular mass weight markers are shown on the far left (lane M). The potentially degraded rSUMO/FABP1 is indicated by arrows and the dimers of rSUMO/FABP3, 4 and 5 are pointed with arrowheads. rSUMO/FABP3 (lane 4 in **B**) and rSUMO/CAT (lane3 in **B**) were treated with SUMO protease at room temperature. The parallel samples that were incubated without the protease (lanes 2 and 5) and the original lysates (lanes 1 and 6 in **B**) are also loaded for comparison. The digested products are indicated by arrows.

We then tried to remove the SUMO segment from rSUMO/FABPs to release rFABPs using SUMO protease. The SDS-PAGE results showed that SUMO protease had very limited capacity to cleave the target from the fused proteins under different salt concentrations, incubation time and temperature (data not shown). It is well known that the cleavage efficiency of SUMO protease depends on the correct conformation of the SUMO moiety. It might be that the conformation of rSUMO/FABPs changes during the processes of purification, delipidation or/and endotoxin removal. Therefore the original lysate was subject to SUMO protease digestion. In the presence of SUMO protease, the control rSUMO/CAT was efficiently digested into two parts but not for rSUMO/FABP3 (Fig. 5-20 B). This evidence rules out the possibility that the limited digestion efficacy may be derived from the conformational changes that occur during the periods of purification, delipidation or/and endotoxin removal. We also tried to denature and renature rSUMO/FABP3 using a gradient urea buffer but this approach did not enhance digestion (data not shown). Considering the binding ability of rSUMO/FABPs (see the section fatty acid binding of rSUMO/FABPs and rFABP3), we thought *E. coli* might not be a suitable host for expression and thus used yeast as a host.

For expression in yeast, only emFABP3 was focused on to produce a recombinant protein. For ease of purification, a 6×His tag was fused at the C terminal. Purified yeast-derived His-tagged FABP3 showed no obvious binding affinity with a fatty acid DAUDA and retinol. The potential explanation is that this protein may be unfolded due to the fused His tag. Thereby we managed to express FABP3 in a native form. The purity of yeast-produced recombinant FABP3 (rFABP3) was checked using SDS-PAGE (Fig. 5-21) and its concentration was up to 678 µg/ml. In this endotoxin-removed rFABP3 sample, the endotoxin level was confirmed to be less than 5 EU/ml (equal to 0.5 ng/ml).

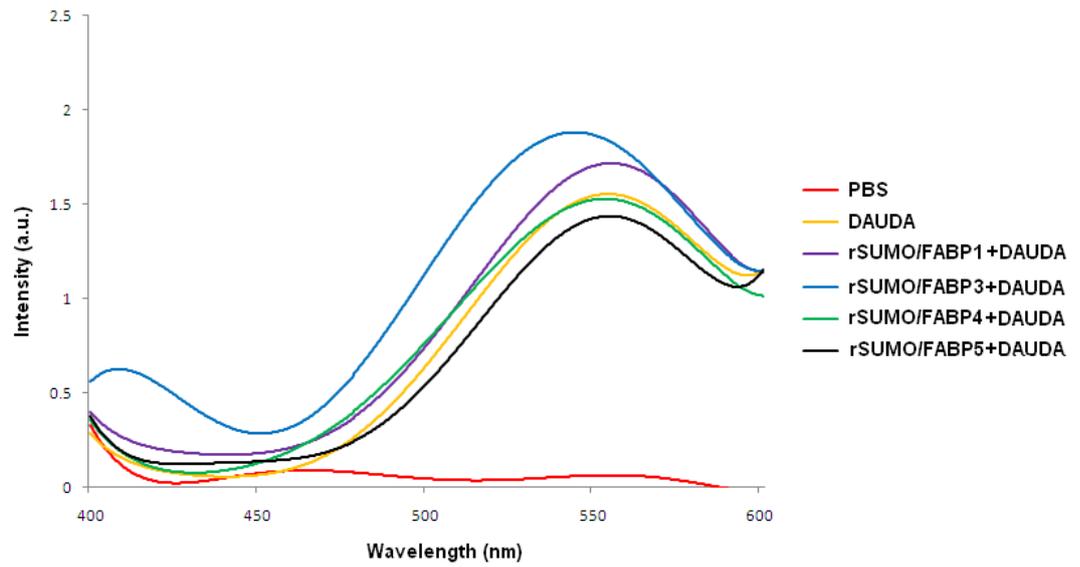
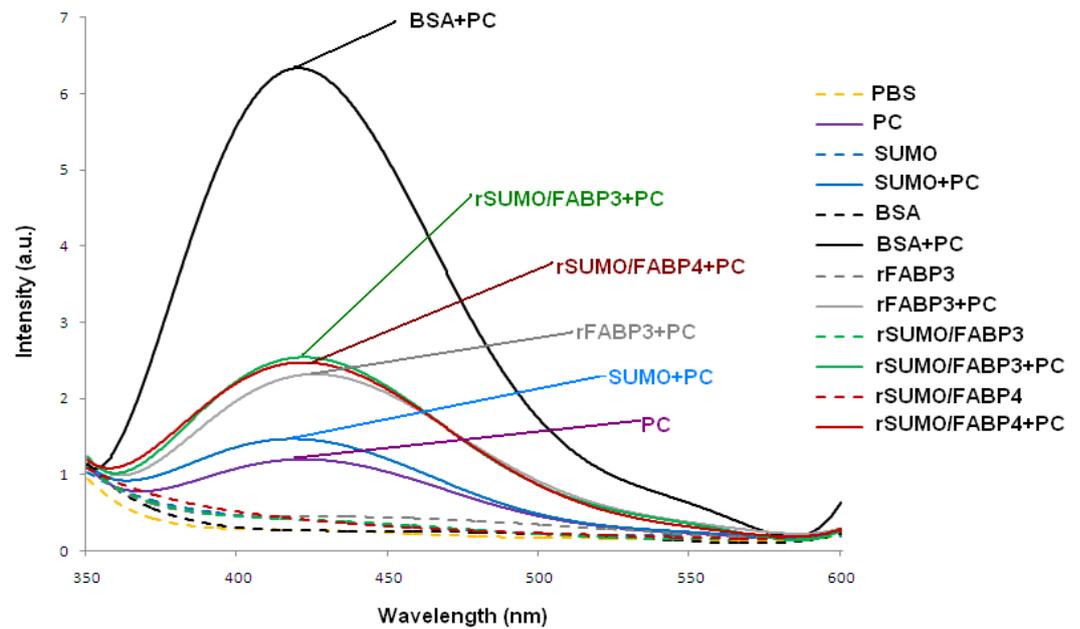


**Figure 5-21 SDS-PAGE analysis of purified yeast-derived rFABP3 in a native form.**

### 3.5 Fatty acid binding of rSUMO/FABPs and rFABP3

The fluorescent fatty acid DAUDA gives maximum emission at 550 nm ( $E_m$  max) after excitation at 350 nm (Barth et al.). The binding of this acid to hydrophobic ligand binding proteins results in increase of fluorescent intensity and a shift of  $E_m$  max wavelength from 550 nm to 500 nm (Saghir et al., 2001; Timanova-Atanasova et al., 2004; Timanova et al., 1999). In our study, the addition of rSUMO/FABPs did not give rise to a pronounced fluorescence increase nor did the the  $E_x$  max shift compared with DAUDA alone (Fig. 5-22 A), indicating no appreciable binding. Likewise, there was not distinguishable difference in the retinol with or without rSUMO/FABPs, which has a natural fluorescent property ( $E_x$  at 350 nm and  $E_m$  max at 480 nm). As for *cis*-parinaric acid that gives a natural fluorescence at 430 nm after excitation at 320 nm, a modest fluorescent increase was observed in both rSUMO/FABP3 and 4 samples compared to the fluorescent probe alone. In contrast, the addition of SUMO used as a negative control only gave slight enhancement of fluorescence. These suggest that the increased fluorescence is attributed to the binding of *cis*-parinaric acid to rSUMO/FABP2 or 4. Apparently, the fluorescent enhancement of rSUMO/FABP2 or 4 was still significantly lower than BSA which was used as a positive control (Fig. 5-22 B). Collectively these results demonstrated that these *E. coli*-derived FABPs had no evident binding capacity with DAUDA and retinol but exhibited weak binding to *cis*-parinaric acid.

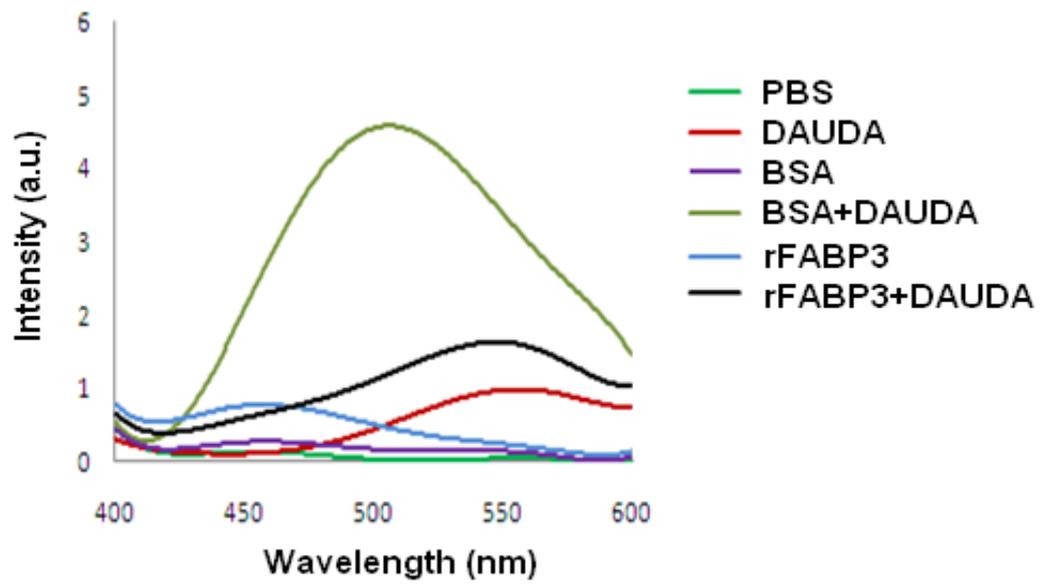
The binding assay with DAUDA showed that the addition of purified His-tagged rFABP3 did not affect the fluorescent intensity and emission wavelength compared to the fatty acid alone. With the exception of slight increase of intensity in the samples combined with DAUDA and retinol, yeast-derived rFABP3 without a His tag presented the same binding patterns as did rSUMO/FABP3 and 4 (Fig. 5-23). We cannot rule out the possibility that the increase of DAUDA and retinol fluorescence results from other factors rather than bound fatty acids.

**A****B**

**Figure 5-22 Binding assay using DAUDA and *cis*-parinaric acid.**

The binding ability of rSUMO/FABPs and yeast-produced rFABP3 without any tags were determined using DAUDA (in **A**) or/and *cis*-parinaric acid (PC, in **B**). The fluorescent intensity was recorded after excitation at 350 nm and 320 nm, respectively.

A



B

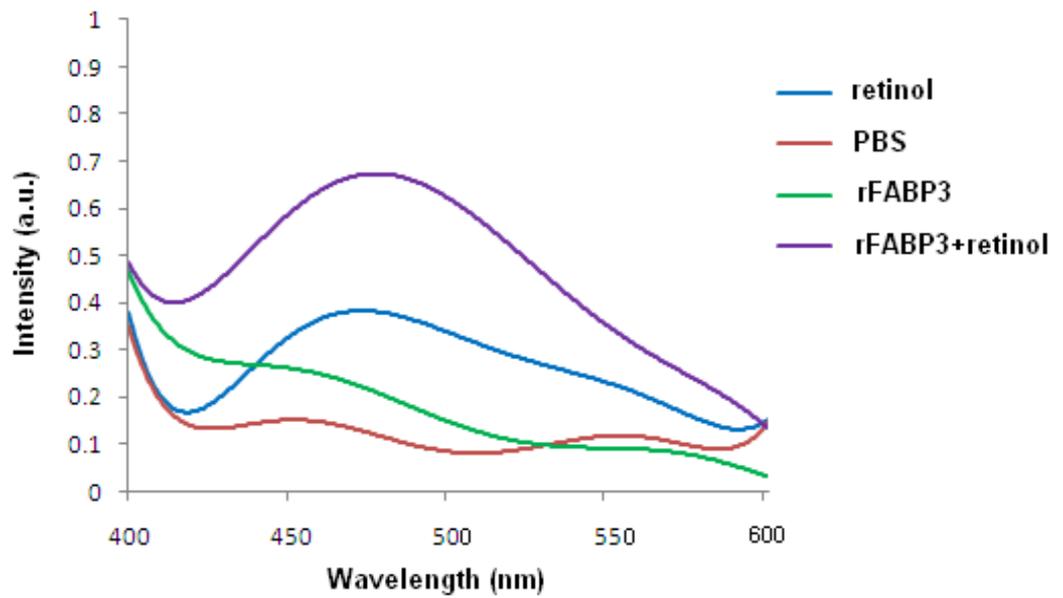


Figure 5-23 Binding assay of yeast-produced rFABP3 without any tags.

### 3.6 Rabbit anti-rFABP3 antibody and Western blotting

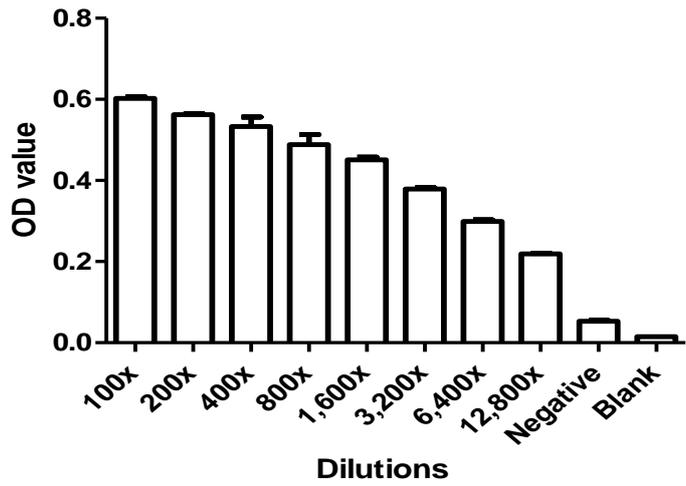
After ammonium sulphate purification, rabbit serum against rFABP3 was subject to determination of a specific anti-rFABP3 IgG level using direct ELISA. As shown in Fig. 5-24, OD values were significantly higher in 12,800× diluted anti-rFABP3 serum than serum collected before immunization (indicated as negative).

Using proteomic approaches, it has been reported that *E. granulosus* FABP(s) are expressed in the germinal layer (Monteiro et al., 2010). But it is not clear whether or not *Echinococcus* FABPs are secretory/excretory. Here, Western blot result showed that anti-FABP3 antibody strongly reacted with the lysate of vesicles or the hydatid fluid (Fig. 5-25), the bands of which had the similar molecular mass weight as rFABP3. This suggests that native FABP3 of *E. multilocularis* is secreted or released into the hydatid fluid.

## 4. Discussion

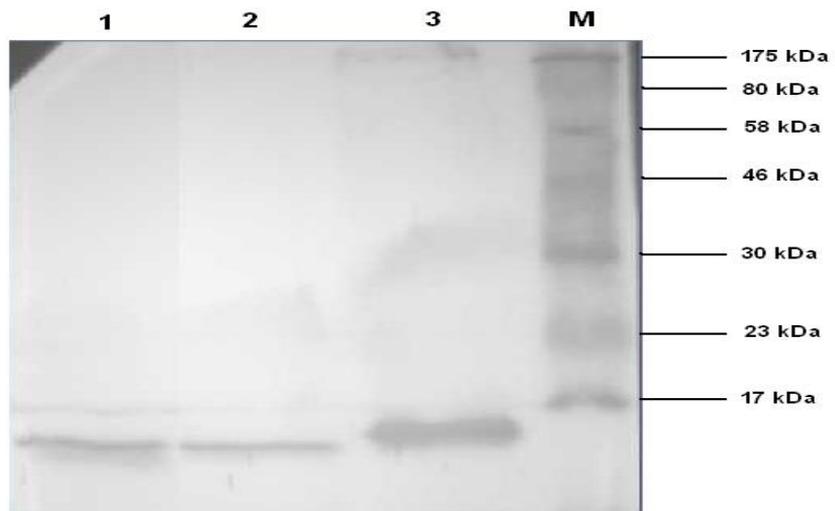
Fatty acids play numerous roles in energy storage, metabolism, gene expression regulation and cellular signalling in organisms. During infection, the normal physiological processes of *E. multilocularis* largely rely on the hosts in which it resides. Due to lack of ability to *de novo* synthesize long-chain fatty acids, this parasite has to acquire lipids entirely from the hosts, in which FABPs are essential players.

Here, it was shown that *E. multilocularis* expressed at least four distinct FABPs without typical gene structures commonly found in vertebrate and other invertebrate FABPs. In contrast to *S. mansoni* and *S. japonicum* which have two and one FABP, respectively, the FABP-encoding genes were expanded in *Echinococcus* species, suggesting the important functions of this protein family. It is worth highlighting



**Figure 5-24 Determination of rabbit anti-rFABP3 antibody level by ELISA.**

The data are expressed as mean  $\pm$  S.D. Each dilution was tested in triplicate. The pre-immunization serum (50 $\times$  diluted) was used as negative.



**Figure 5-25 Confirmation of native FABP3 present in the hydatid fluid by western blot.**

Reaction of anti-rabbit FABP to vesicle lysate (lane1), hydatid fluid (lane 2) and to rFABP3 (lane3). Positive reactions are arrowed at an appropriate molecular weight to predictions from the amino acid sequence. The protein markers are shown on the far right (M).

that FABP-coding genes of both trematodes can be spliced to yield isoforms that may possess variant properties and execute different functions. There was no evidence to support such a post-transcriptional modification in maturation of *Echinococcus* FABPs. Intriguingly, the *E. multilocularis* genome contained two loci for intronless FABP genes, emFABP1 and emFABP5. Intronless FABP genes have already been described in mice (Treuner et al., 1994), rat (Bonne et al., 2003) and humans (Prinsen et al., 1997), all of which have lost activity during evolution, being pseudogenes. The human intronless FABP gene is localized on chromosome 13 and shares 85% identity to human heart-FABP, which resides on chromosome 1. Alterations of some amino acid residues, which are indispensable for the interaction with fatty acids, lead to the loss of function in lipid binding (Prinsen et al., 1997). It is still not clear what the binding spectra of *E. multilocularis* FABP1 and 5 are and whether or not emFABP1 and emFABP5 are pseudogenes. But the fact that both genes are absolutely conserved at the amino acid level between *E. multilocularis* and *E. granulosus* favours the idea that emFABP1 and emFABP5 may function normally although they seem to be expressed at low levels in the samples investigated in this study.

The genomic organization and gene structures of FABP-coding genes were similar in both *Echinococcus* species. Integral analyses of coding region, 5' UTR or/and intron sequences illustrated that these features may have derived from their common ancestor. Simultaneously, it is also supported that duplication of FABP2 or FABP3, which are identical at the nucleotide and amino acid levels in *E. multilocularis*, may have occurred before speciation of *E. multilocularis* and *E. granulosus*. Among five orthologues, FABP1, 4 and 5 were highly conserved in both parasites, while FABP2 and FABP3 were various. The divergence of both egFABP2 and egFABP3 is proposed to have stemmed from mutations in the evolution of *E. granulosus*. We therefore argue against the concept that egFABP2 and egFABP3 may have proceeded from a recent duplication event (Esteves et al., 2003). Given egFABP3 plays roles in

the fatty acid metabolism in *E. granulosus* infection, the heterogeneity of egFABP3 may reflect variety of its hosts because different animals including human beings are various in an aspect of lipid metabolism (Nafikov and Beitz, 2007). For instance, lipids are digested in a unique way in ruminants, giving rise to fermentation of triacylglycerols- and phospholipids-derived glycerol to volatile fatty acids and to hydrogenation of unsaturated fatty acids to saturated fatty acids (Nafikov and Beitz, 2007).

In *E. multilocularis*, emFABP3 was highly expressed in cell aggregates, protoscoleces and vesicles, particularly the latter, suggesting that emFABP3 may play a role in vesicle infection or/and development. Conversely, emFABP1 and 5 appeared to be expressed at low levels. As discussed before, these two genes may be functional and it will be interesting to investigate their expression status in the infectious stages. It is well established that *E. multilocularis* can release numerous molecules to increase the chance of survival by modulating the parasitic microenvironment in hosts. So it is worthwhile investigating whether or not emFABPs are secretory/excretory. Not surprisingly, our study confirmed that native emFABP3 was secreted or released into the hydatid fluid. Although their precise functions are unclear, a wealth of studies has shown that some FABPs are present in the excretory/secretory (ES) products in trematodes (Chunchob et al., 2010; Liu et al., 2009a; Morpew et al., 2007) and the blood in healthy humans (Tso et al., 2010). In *E. granulosus*, the expression of FABP(s) was found in the germinal layer but not in the hydatid fluid using proteomic approaches (Monteiro et al., 2010). Vertebrate FABPs have been shown to be closely related with inflammation (Makowski et al., 2001; Shum et al., 2006). Notwithstanding the low identity between emFABPs and vertebrate FABPs, they do have similar tertiary structures, whereby these FABPs function. Therefore it can be envisaged that emFABPs released into surrounding tissues may participate in the host-parasite interaction during infection.

Although the function remains unclear, some studies indicate that different FABPs may function cooperatively as a protein complex and individual FABPs can form dimers. For instance, *F. hepatica* FABP from adults comprises at least eight isoforms with the same molecular weight but not pI (Espino and Hillyer, 2001). Another example is a bovine mammary-derived growth inhibitor, which depresses the growth of breast cancer cells. This was revealed to be a mixture of heart and adipocyte FABPs, with heart FABP being the active factor responsible for suppression of cell growth and proliferation (Borchers et al., 1997). For *E. coli*-derived emFABP3, 4 and 5 but not yeast-produced FABP3, we noticed that there was a single band with a molecular mass weight of less than 66.4 kDa, suggesting that these proteins are dimers. Such self-aggregation has already been described in other FABPs and it is assumed that this self-aggregation occur upon ligand binding (Lucke et al., 2002; Timanova-Atanasova et al., 2004). It is possible that delipidated emFABP3, 4 and 5 still harbour bound fatty acids due to inefficient lipid depletion, which was also reported previously (Jakobsson et al., 2003).

Endotoxin, known as lipopolysaccharide (LPS), is one of cell wall constituents of Gram-negative bacteria. It can activate immune cells to secrete a large number of pro- and anti-inflammation cytokines via Toll-like receptors. To eliminate artifacts, recombinant proteins that are used to explore immunological functions are required to be endotoxin-free. There are several approaches that have been extensively used for endotoxin elimination (Magalhaes et al., 2007). In our experiments Triton-114 was first employed to remove LPS but the efficiency was limited especially when LPS concentration in the samples was low. Then we tried to utilize Triton-114 during purification steps as described previously (Reichelt et al., 2006). However the samples still contained a high level of endotoxin (100 EU/ml~1000 EU/ml). Finally, a combination of Triton-114 and polymyxin B-labelled beads was used, which was

found to efficiently deplete LPS. The main disadvantage of the combined method is relatively lower recovery in comparison with the approaches described above.

The crystal structure of egFABP2 has been deciphered at a resolution of 1.6 Å, revealing a canonical barrel comprising of 10 β sheets that is structurally similar to myelin FABP (Jakobsson et al., 2003). Three residues Arg<sup>107</sup>, Arg<sup>127</sup> and Tyr<sup>129</sup> are supposed to be involved in interactions with fatty acids. Unexpectedly, a fatty acid molecule, possibly palmitic acid, was found to bind to purified and delipidated egFABP2. This protein is capable of binding to *cis*-parinaric acid with a dissociation constant at a nano-molar range (Alvite et al., 2001). It may offer two binding sites for *cis*-parinaric acid, which were not supported by structural data (Jakobsson et al., 2003). In our study the binding results revealed that emFABP3 and 4, which shared 93% and 76% identities with egFABP2, respectively, were incapable of DAUDA and retinol but both exhibited weak binding to *cis*-parinaric acid. Likewise, yeast-derived FABP3 had the similar binding spectrum. In both emFABP3 and 4, the amino acids that are directly associated with ligand binding were intact. Moreover, another three residues, such as Phe<sup>16</sup>, Pro<sup>76</sup> and Asp<sup>77</sup> that are critical for fatty acid binding (Jakobsson et al., 2003), were also invarious. It is still not clear why these proteins possess limited capacity for binding to DAUDA and retinol. One possible explanation is that these proteins lack the proper conformation due to unknown factors. Another possibility is that the benzenoid moiety of DAUDA prevents it from entry into the binding pockets. Alternatively, emFABP3 and 4 may have lost the binding capacity during evolution. Elucidation of the crystal structures will help us to address these issues.

## **Chapter 6 Modulation of macrophage and dendritic cell functions by *E. multilocularis* FABP3 *in vitro***

### **Abstract**

The lack of adipocyte FABP gives rise to the dramatic abolition of cytokine expression in macrophages and dendritic cells. It suggests that this protein could be involved in the pathway associated with cytokine synthesis. Here rFABP3 was shown to induce the secretion of pro- and anti-inflammatory cytokines by murine bone marrow-derived macrophages (BMDMs) and/or dendritic cells (BMDCs). This was significantly attenuated by addition of specific antibody against rFABP3. rFABP3 was also shown to activate Toll-like receptor (TLR) 2 but not TLR 4, 1/2, 6/2, 5 and 9 in HEK transfected cells. Moreover, the expression of IL-6 and TNF  $\alpha$  by BMDMs stimulated by rFABP3 was significantly down-regulated by the addition of anti-TLR2 antibody but no effects were observed in LPS-activated and anti-TLR 2 antibody-blocked BMDMs. We also showed that rFABP3 did not possess the capacity to activate BMDM nitric oxide (NO) or suppress of NO production by LPS-activated BMDMs. Interestingly, the combination of rFABP3 and heat-killed *Listeria monocytogenes* elicited high levels of NO in a dose-dependent manner. These results demonstrate that FABP3 induces BMDM and BMDC cytokine secretion feasibly through a TLR 2 pathway.

### **1. Introduction**

During infection, *E. multilocularis* dynamically interacts with hosts to favour its invasion and growth via the release of a wide range of molecules, some of which interfere with the normal functions of the host immune system. It has been shown that

*E. multilocularis* FABP3 is highly expressed in protoscoleces and is present in the hydatid cyst fluid (Chapter 5).

FABPs are proposed to participate in a wide range of processes, including lipid trafficking, signalling, oxidation and gene transcription regulation, and deficiencies in this protein family are strongly associated with metabolic syndrome or disorders (Atshaves et al., 2010; Furuhashi and Hotamisligil, 2008; Furuhashi et al., 2007; Liu et al., 2010). For instance, three mammalian heart, epidermal and brain FABPs that are expressed in different spatial and temporal patterns are involved in neuron-glia interactions during brain development (Liu et al., 2010). Genetic analysis has revealed that brain FABP is likely linked to prepulse inhibition, which is a biological hallmark for schizophrenia, and abnormal expression or/and a single nucleotide polymorphism of the brain FABP-coding gene are observed in schizophrenia patients.

A link between FABP and inflammation emerges from the findings that adipocyte FABP (aFABP) is expressed in macrophages upon stimulation and cytokine expression is strikingly reduced in aFABP-deficient macrophages (Makowski et al., 2001). Knockout of aFABP reduces the activity of I $\kappa$ B kinase and NF- $\kappa$ B, the activity of which relies on the availability of fatty acids, giving rise to an impairment of cytokine production (Makowski et al., 2005). It has been recently demonstrated that aFABP influences macrophage cytokine production via a positive feedback network, which is comprised of c-Jun NH<sub>2</sub>-terminal kinase and activator protein-1 (Hui et al., 2010). After LPS stimulation, aFABP transcription is activated, which is initiated by c-Jun NH<sub>2</sub>-terminal kinase that is recruited to the binding region of the activator protein 1 in the aFABP promoter.

Dendritic cells show similar expression profiles of FABPs to macrophages. During differentiation, aFABP is also up-regulated but is not crucial to dendritic cell differentiation (Rolph et al., 2006). Deficiency of aFABP adversely affects the

degradation of I $\kappa$ B kinase in the LPS-stimulated dendritic cells, which mediates the migration of NF- $\kappa$ B to the nucleus, resulting in a reduced transcription of inflammatory genes. Furthermore, aFABP is required for dendritic cells to activate T cells (Rolph et al., 2006).

Considering the diversity of FABP functions, it is worthwhile investigating the immunological properties of *E. multilocularis* FABP3. Here we showed that this antigen modulated the expression of cytokines TNF  $\alpha$ , IL-6 and IL-10 by bone marrow-derived macrophages (BMDMs) and dendritic cells (BMDDs) potentially through TLR 2. Moreover, rFABP3 did not induce macrophages or dendritic cells directly to generate nitric oxide (NO) or suppress NO by LPS-stimulated BMDMs and BMDDs. However, the combination of rFABP3 and heat-killed *Listeria monocytogenes* induced NO synthesis in a dose-dependent pattern. These results raise the possibility that FABP3 has immunomodulatory functions during *Echinococcus* species infection.

## **2. Material and methods**

### **2.1 TLR cell lines, ligands and mice**

Human embryonic kidney epithelial cell line 293 (HEK293) TLR transformants including TLR 1/2, 2/6, 2, 4 and 5 and CD14 transformant cells (kept in our lab) were used in this study and corresponding ligands were purchased from Invivogen. For TLR 2, 4 and CD 14, the cells were maintained in DMEM (Sigma) supplemented with 10% fetal bovine serum (FBS, endotoxin-free, sigma), 5  $\mu$ g/ml puromycin (Invivogen) and 100  $\mu$ g/ml normocin (Invivogen). The other TLR cells were cultured in the same media except puromycin that was replaced by 10  $\mu$ g/ml blasticidin.

6~8 weeks old healthy C57 BL/6 mice were purchased from Charles River Laboratories, UK. Mice were kept in animal facilities (Queen's Medical Centre, the

University of Nottingham) until they were humanely sacrificed by carbon dioxide and cervical dislocation before use.

## **2.2 Stimulation of TLR transformants by rFABP3 and ligands**

One day before activation, cells at approximately 80% confluence were harvested using trypsin detachment solution (Sigma) and cell viability was checked by trypan blue exclusion. 100  $\mu$ l of  $\sim 3.5 \times 10^5$  cells were seeded in 96-well plates (Costar) and then placed for 12h in a 37°C incubator supplemented with 5% CO<sub>2</sub>. Next day, the medium in each well was gently removed and 100  $\mu$ l of the fresh medium containing rFABP3 or/and ligands added and incubated for additional 24h. The medium was then stored at -80°C for future use.

## **2.3 Preparation of bone marrow-derived macrophages and bone marrow-derived dendritic cells**

Bone marrow-derived macrophages (BMDMs) and bone marrow-derived dendritic cells (BMDDs) were prepared according to previously described methods with some modifications (Flores et al., 2007; Weischenfeldt and Porse, 2008). Briefly, mouse tissue-free femurs and tibias were flushed with ice-cold endotoxin-free PBS (Sigma) to collect bone marrow. Red blood cells were disrupted by the addition of lysis buffer (Sigma), followed by centrifugation at 400 $\times$ g for 10 min. The cells were then resuspended in RPMI-1640 media (Sigma) supplemented with 10% FBS and the cell viability and concentration determined as before.

For BMDMs,  $1.5 \times 10^6$  cells per well were seeded in 24-well plates (Costar) in 10% FBS RPMI-1640 containing 20 ng/ml macrophage-colony stimulation factor (M-CSF, R&D Systems) and incubated overnight at 37°C with 5% CO<sub>2</sub>. On day 2, the plates were gently and briefly shaken to get rid of dead and unattached cells and the same fresh medium was added, followed by incubation as described previously. The

medium was also changed on day 3 and 5. BMDMs normally appeared to be confluent and adherent on day 6 to 9.

To prepare BMDDs, 2 ml of bone marrow cells ( $6.0 \times 10^5$  cells per ml) were added into each well in 6-well plates in 10% FBS RPMI-1640 containing 50  $\mu$ M 2-mercaptoethanol and 20 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D Systems) and incubated as before. The medium was changed on day 3 and 6. On day 7, the cells were thoroughly washed with endotoxin-free PBS (Sigma) to detach cells. The viable cells were counted and 500  $\mu$ l of the cells ( $3.0 \times 10^5$ ) were added into each well in 24-well plates and incubated for 6h for adherence. Afterwards, BMDDs were stimulated by rFABP3 and other ligands.

## **2.4 Determination of cytokine levels in the supernatant of cell culture after stimulation**

The concentration of cytokines including interleukin-8 (IL-8) by TLR transformant cells, TNF  $\alpha$ , IL-6 and IL-10 by BMDMs or/and BMDDs was determined by DuoSet ELISA Kit according to the instructions (R&D Systems).

Firstly, 50  $\mu$ l of capture antibody diluted in filtered PBS were added per well in 96-well plates and incubated overnight at room temperature. After three washes with PBS containing 0.05% Tween-20, the coated plates were blocked in 1% BSA-containing PBS (pH 7.4, B-PBS) at room temperature for at least 1h, followed by addition of 50  $\mu$ l of cell culture supernatant or standards and then incubation at room temperature for additional 2h. For the blank wells, B-PBS or reagent diluent (0.1% BSA, 0.05% Tween-20, 20 mM Tris, 150 mM NaCl, pH 7.2, only for IL-8 and IFN gamma) was used. The plates were rinsed as mentioned before and incubated with 50  $\mu$ l detection antibody for 2h at room temperature. 50  $\mu$ l of 200 $\times$  diluted streptavidin conjugated to peroxidase were added to each well and placed in the dark for 20 min. After three washes, 50  $\mu$ l of substrate solution (3,3',5,5'-tetramethylbenzidine, Sigma) were

added into each well for colour development in the dark at room temperature for 20 min, followed by addition of 25  $\mu$ l of 2M H<sub>2</sub>SO<sub>4</sub>. The OD values were read using microplate reader (Bio-Rad) and the concentration calculated using the 4PL program.

## **2.5 Assessment of nitric oxide by BMDMs**

Generally, nitric oxide (NO) is unstable and tends to be easily oxidized into nitrite (NO<sub>2</sub><sup>-</sup>) (Tarpey et al., 2004). The concentration of nitrite in the supernatant of cell culture can indirectly reflect the output of NO. Nitrite was detected using modified Griess reagent (Sigma). Simply, 50  $\mu$ l of culture supernatant were mixed with an equal volume of Griess reagent and incubated at room temperature for 15 min. The absorbance was recorded at 570 nm using microplate reader (Bio-Rad). A serial of seven double dilutions (from 100  $\mu$ m to 1.56  $\mu$ m) was used as standards.

## **2.6 Preparation of BMDM total RNA**

Total RNA of BMDM was extracted using mirVana miRNA Isolation Kit (Ambion) according the method recommended for total RNA extraction protocol. Briefly, BMDMs were washed with ice-cold PBS and then disrupted by addition of Lysis/Binding Solution, followed by vortexing vigorously to entirely lyse cells. The lysate was subject to extraction using Acid-Phenol: Chloroform. The aqueous phase was transferred to new tubes and 1.25 volume of absolute ethanol were added and vortexed. This mixture was loaded onto Filter Cartridge, followed by washes with Wash Solution I and then Wash Solution 2/3. Total RNA species were eluted using pre-heated nuclease-free water.

To decontaminate genomic DNA, the extracted RNA was subject to DNase I (Ambion) treatment for at 37°C 20 min, followed by addition of DNase Inactivation Reagent (Ambion). The quality of total RNA was assessed by Nanodrop (Thermo Scientific).

## **2.7 Qualitative PCR analysis of expression of arginase I by**

### **BMDMs**

In this study, two-step quantitative PCR (qPCR) was applied. Complementary DNA (cDNA) was synthesized using High Capacity RNA-to-cDNA Kit (Applied Biosystems). 200 ng of genomic DNA-free total RNA were gently mixed with 10  $\mu$ l 2 $\times$  RT Buffer and 1  $\mu$ l of 20 $\times$  Enzyme Mix. The reaction was then incubated at 37°C for 1h, followed by inactivation of enzyme at 95°C for 5 min.

Using 1  $\mu$ l of cDNA as template, qPCR was performed in a 10  $\mu$ l reaction system containing 5  $\mu$ l of Fast SYBR Green Master Mix (Applied Biosystems) and 0.25  $\mu$ l of arginase I primers each (2  $\mu$ M, forward: 5'-CCTTGGCTTGCTTCGGAAC-3'; reverse: 5'-ATGTGGCGCATTACAGTCA-3'). The reaction was run using 7500 Fast (Applied Biosystems) under the following conditions: 95°C for 20 sec, followed by 40 cycles of 95°C for 3 sec and 60°C for 30 sec. Relative expression levels of arginase I were normalized to that of actin beta (2  $\mu$ M, forward: 5'-GCATTGCTGACAGGATGCAG-3'; reverse: 5'-GCCACCGATCCACACAGAGT-3').

## **2.8 Statistical analysis**

Graphpad Prism 5 software was employed for data analysis. Statistical significance was analyzed using two-tailed unpaired *t*-test if not clearly indicated. In all comparisons, a *p* value less than 0.05 was considered statistically significant.

## **3. Results**

### **3.1 Cytokine profiling of BMDMs and BMDDs treated by rFABP3**

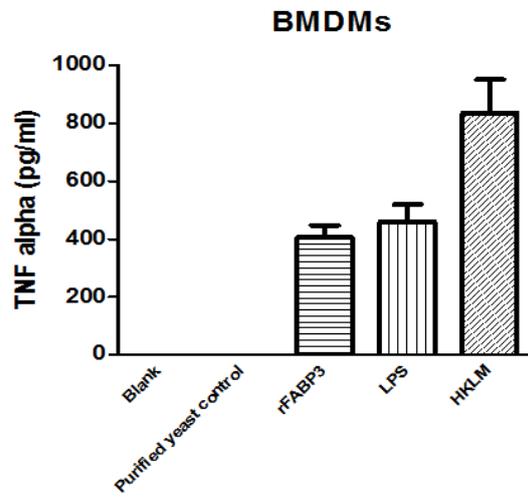
We first investigated the effects of FABP3 on the cytokine production of macrophages and dendritic cells. rFABP3 induced the production of cytokines including TNF  $\alpha$ , IL-

IL-10 and IL-6, with no appreciable TNF  $\alpha$  secretion in the BMDMs treated with purified yeast control or in the blank (Fig. 6-26 A, B and C). In BMDMs, production of TNF  $\alpha$  and IL-6 in the rFABP3-treated group was comparable to that produced by the cells stimulated by LPS (1  $\mu\text{g/ml}$ ,  $p > 0.05$ ), an agonist for TLR 4. By contrast, rFABP3 elicited much less IL-10 than did LPS ( $p < 0.01$ ). It was noted that, with exception of IL-6, heat-killed *L. monocytogenes* (HKLM), a ligand for TLR 2, significantly induced higher levels of TNF  $\alpha$  and IL-10 in comparison with rFABP3 ( $p < 0.01$ ).

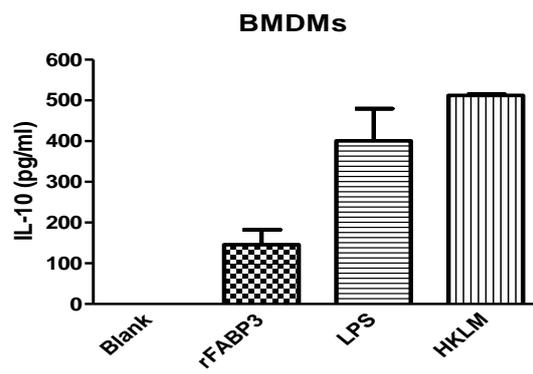
In BMDDs, elevated TNF  $\alpha$  was also observed after stimulation and was significantly higher in the rFABP3-stimulated cells than that in the blank ( $p < 0.01$ ). Likewise, LPS also induced a high level of TNF  $\alpha$  (its concentration was out of detection limits, data not shown). It was noticed that the appreciable production of TNF  $\alpha$  was observed in the normal BMDDs (Fig. 6-26 D).

In our experiments, the different order of addition of rFABP3 (5  $\mu\text{g/ml}$ ) or LPS (100  $\mu\text{g/ml}$ ) exhibited different stimulation efficacy in BMDMs. The TNF  $\alpha$  levels were significantly higher in the group that were treated by rFABP3 first and LPS added 12h later (rFABP3-LPS) than the LPS- and LPS-rFABP3-stimulated cells (Fig. 6-26 E).

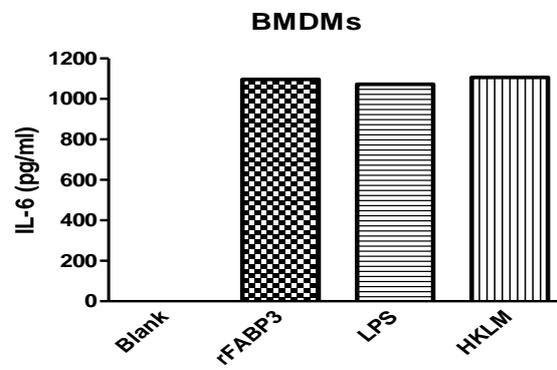
A



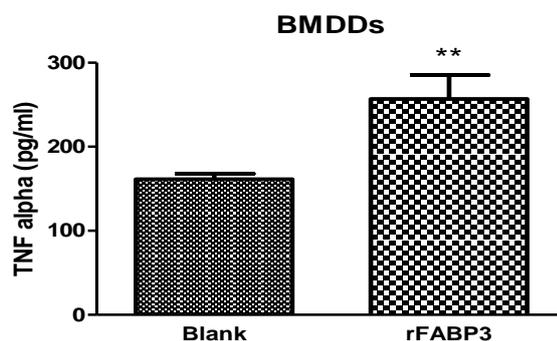
B



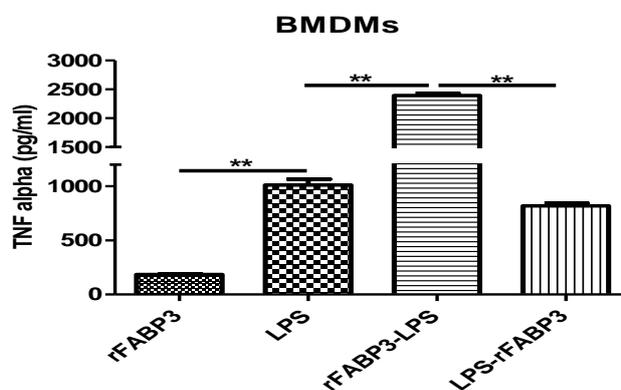
C



**D**



**E**



**Figure 6-26 rFABP3 effects on cytokine production by BMDMs and BMDDs.**

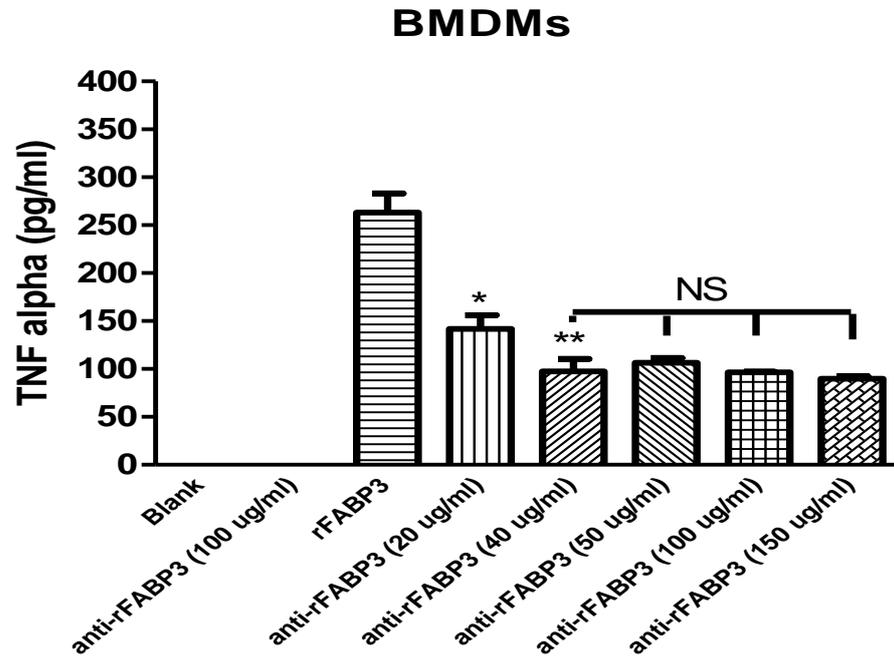
Except **E**, BMDMs and BMDDs were activated by the following stimuli: rFABP3 (10  $\mu\text{g/ml}$ ), LPS (1  $\mu\text{g/ml}$ ), rFABP3 (10  $\mu\text{g/ml}$ ) + LPS (1  $\mu\text{g/ml}$ ), HKLM ( $1 \times 10^8$ ) and purified yeast control (2 or 5 volumes of rFABP3). After 24h incubation, culture supernatant was subject to ELISA for determination of TNF  $\alpha$  (**A** and **D**), IL-10 (**B**) and IL-6 (**C**). Purified yeast control was prepared as the methods used for preparation of rFABP3. In **E**, BMDMs were treated with rFABP3 (5  $\mu\text{g/ml}$ ), LPS (100 ng/ml), first rFABP3 (5  $\mu\text{g/ml}$ ) and 12h later LPS (100 ng/ml) (rFABP3-LPS), or first LPS (100 ng/ml) and 12h later rFABP3 (5  $\mu\text{g/ml}$ ) (LPS-rFABP3). The data are expressed as the mean  $\pm$  S.E. \*\* $p < 0.01$ .

### **3.2 Anti-rFABP3 antibody down-regulates the TNF $\alpha$ expression stimulated by rFABP3**

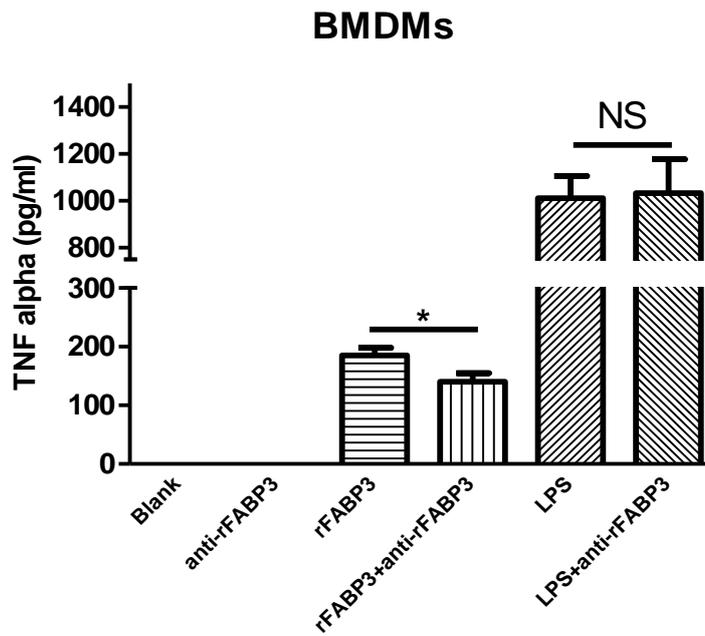
To confirm the effects of rFABPs on cytokine production, anti-rFABP3 antibody was used to block the actions of the recombinant protein. First of all, the concentration of anti-rFABP3 for neutralization was optimized. The secretion of TNF  $\alpha$  or IL-10 production significantly decreased in all the treated cells upon the addition of anti-rFABP3 antibody (from 20  $\mu\text{g/ml}$  to 150  $\mu\text{g/ml}$ ) compared to the rFABP3 group without addition of anti-rFABP3 antibody ( $p < 0.05$ , Fig. 6-27 A) and pre-bleed serum group (data not shown). Moreover, the significant decrease in TNF  $\alpha$  concentration ( $p < 0.01$ ) occurred in the groups with the addition of 40  $\mu\text{g/ml}$  or higher but there were no significant differences among these groups. TNF  $\alpha$  secretion was not detected in the cells in the presence of anti-rFABP3 antibody. These results suggest that 40  $\mu\text{l/ml}$  is an optimal concentration for blocking rFABP3 function.

In BMDMs, the production of TNF  $\alpha$  induced by rFABP3 was significantly attenuated by anti-rFABP3 blocking ( $p < 0.05$ ), and the difference between LPS- and LPS+anti-rFABP3-costimulated cells was not statistically significant ( $p > 0.05$ , Fig. 6-27 B). Similarly, the significant decrease of TNF  $\alpha$  levels was observed in the rFABP3-treated dendritic cells in the presence of anti-rFABP3 antibody in comparison with that in the blank (Fig. 6-27 C). Statistically, there were no differences between the blank, the anti-rFABP3 group and the rFABP3+anti-rFABP3 group. It is assumed that rabbit anti-rFABP3 antibody is capable of neutralizing the cytokine production of rFABP3 by macrophages and dendritic cells.

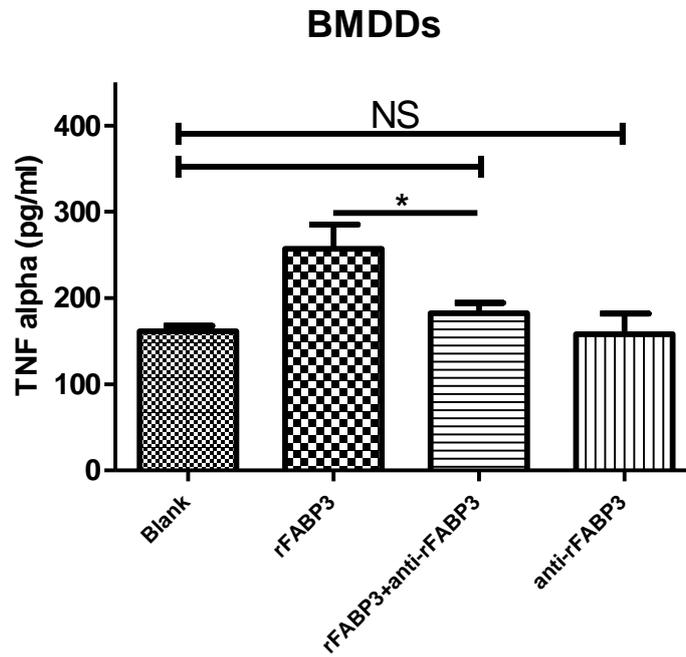
A



B



C



**Figure 6-27 Anti-rFABP3 neutralization of rFABP3 functions.**

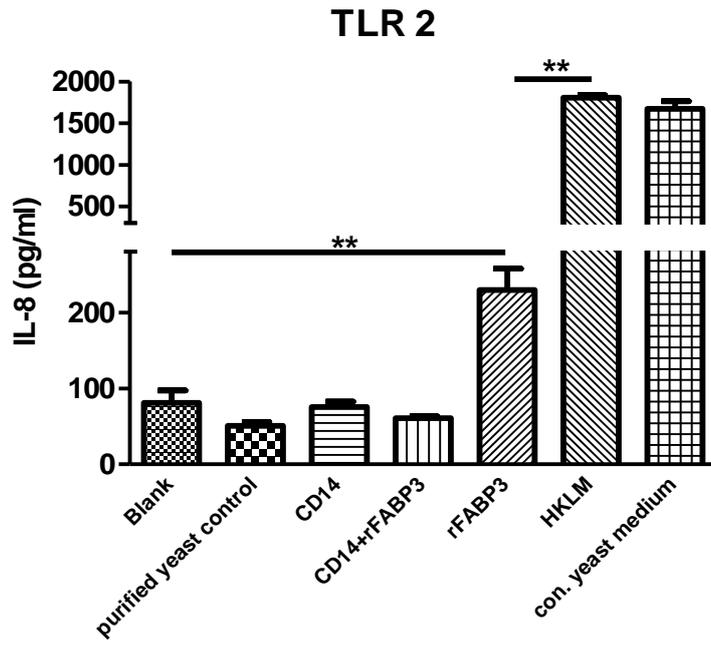
Before blockade of rFABP3 using specific antibody, the conditions were optimized using ELISA (A). In the experiments, rFABP3 (5  $\mu\text{g/ml}$ ) and LPS (100  $\text{ng/ml}$ ) were used to stimulate BMDMs (B) and BMDDs (C). For rFABP3 + or LPS + anti-rFABP3, rFABP3 or LPS was mixed with 40  $\mu\text{g/ml}$  anti-rFABP3 and then incubated at 37°C for 1h prior to activation. The data are expressed as the mean  $\pm$  S.E. \*\* $p < 0.01$ , \* $p < 0.05$ , NS: not significant.

### 3.3 rFABP3 can activate Toll-like receptor 2

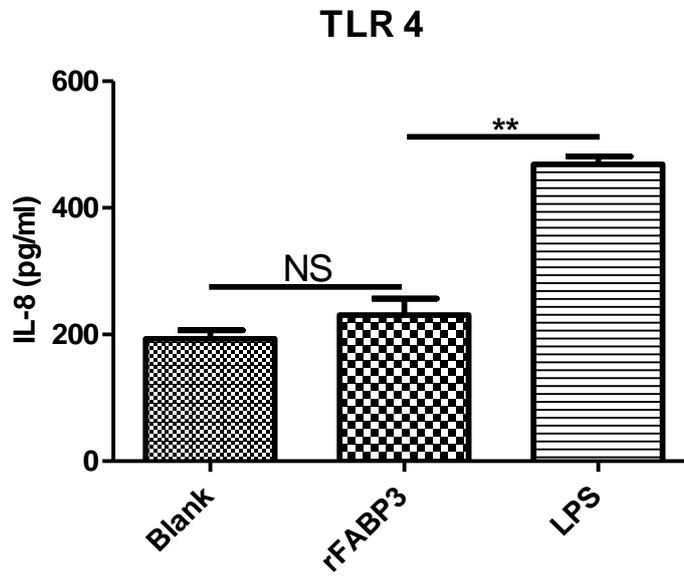
To explore whether rFABP3 can stimulate BMDMs and BMDDs to generate cytokines via TLR(s), TLR transformants including TLR 1/2, 6/2, 2, 4 and 5 and CD14 transformant cells as a control were exposed to rFABP3. The data presented here showed that rFABP3 activated transformant cells via TLR 2, giving rise to significant increase in IL-8 production in comparison with the blank group ( $p < 0.01$ ) and the purified yeast control (Fig. 6-28 A). In contrast to TLR 2 ligand, HKLM, rFABP3 seemed to have relatively limited ability to activate TLR2 ( $p < 0.01$ ). Concentrated growth medium (5× concentrated) induced a high level of IL-8, comparable to that induced by HKLM-activated cells.

Compared to the blank, there was no significant increase of IL-8 in the rFABP3-treated TLR 4 transformant cells, while expression of IL-8 was significantly up-regulated after stimulation with LPS, a ligand for TLR4 (Fig. 6-28 B). This suggests that the endotoxin level in the rFABP3 solution is low. None of the rFABP3-treated TLR 1/2, 2/6 and 5 cell lines secreted appreciable amounts of IL-8, while corresponding ligands for these transformants stimulated the secretion of IL-8 (the levels of IL-8 by the blank controls of these three cell lines were under the detectable limit and then not shown in the figure) (Fig. 6-28 C). Together with these results suggest that rFABP3 activates TLR 2 but not TLR 4, 1/2, 2/6 and 5.

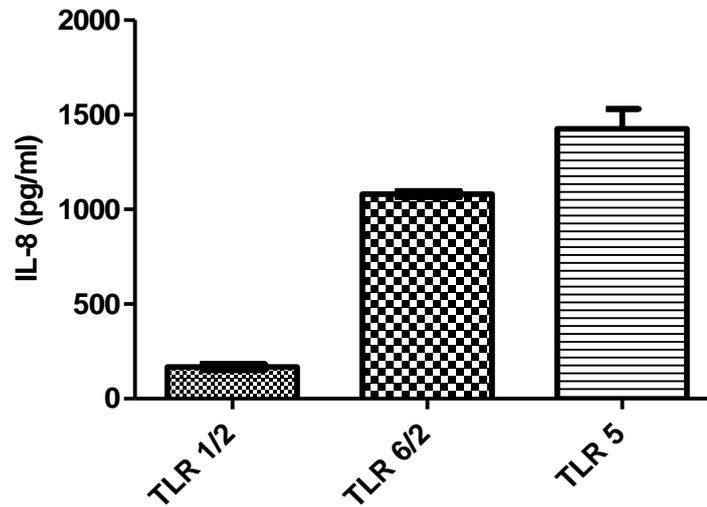
A



B



C



**Figure 6-28 Activation of TLR transformant cells by rFABP3 via TLR 2.**

TLR 2 and 4 transformant cells were treated by rFABP3 (5  $\mu\text{g/ml}$ ), LPS (100 ng/ml), HKLM ( $1 \times 10^8$ ) and concentrated yeast medium (5 $\times$ , an equal volume of rFABP3) (A and B). Pam3CSK4, FSL-1 and Flagellin were used to activate TLR 1/2, 6/2 and 5 transformants, respectively (C). The data are expressed as the mean  $\pm$  S.E. **\*\*** $p < 0.01$ , NS: not significant.

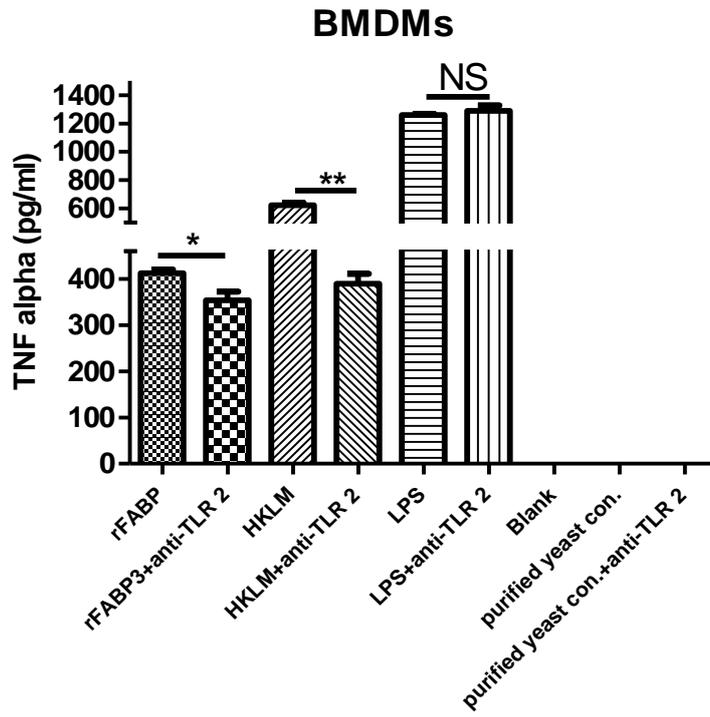
### **3.4 rFABP3 modulates BMDM/BMDD cytokine secretion potentially in a TLR 2-dependant manner**

To test if rFABP3 mediates BMDMs and BMDDs to increase the secretion of TNF  $\alpha$  via TLR 2, we showed that anti-TLR 2 pre-treatment gave rise to a pronounced decrease in TNF  $\alpha$  induced by rFABP3 ( $p < 0.05$ ) as well as HKLM ( $p < 0.01$ ). However there was no significant difference amongst the LPS-treated cells with or without anti-TLR 2 (Fig. 6-29 A) and the HKLM-treated cells with or without isotype control, IgG2a (data not shown). No appreciable secretion of TNF  $\alpha$  was found in the blank and purified yeast control groups with or without anti-TLR 2. A significant reduction of IL-10 and IL-6 was also observed in the anti-TLR2-blocked and rFABP3-mediated BMDMs in comparison with that in the treated cells without blocking of TLR 2 functions (the levels of IL-10 and IL-6 by the blank controls of these three cell lines were under the detectable limit and then not shown in the figures) (Fig. 6-29 B and C).

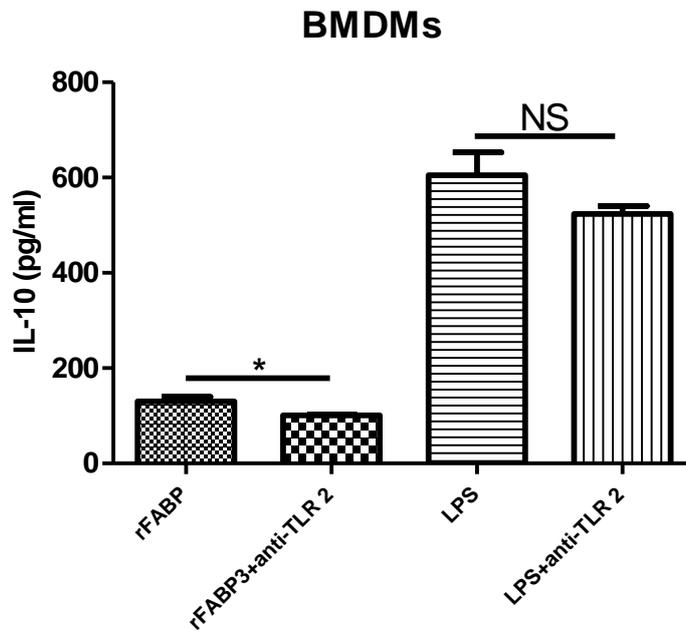
### **3.5 rFABP3 stimulates macrophages possibly by bypassing an alternative activation pathway**

Macrophages from different sources are heterogeneous in aspects of morphology and phenotype and there are two distinct pathways for macrophage activation: classical and alternative. Macrophage alternative activation can be initiated directly or indirectly by several cytokines such as IL-4, IL-13, IL-33 and IL-25 (Gordon and Martinez, 2010). The distinction between these two activation patterns can be determined by phenotypical markers, including arginase I, Fizz1, Ym1 and Ym2. Although its functions are not fully understood, arginase I is up-regulated in the IL-4- or IL-13-activated mouse macrophages.

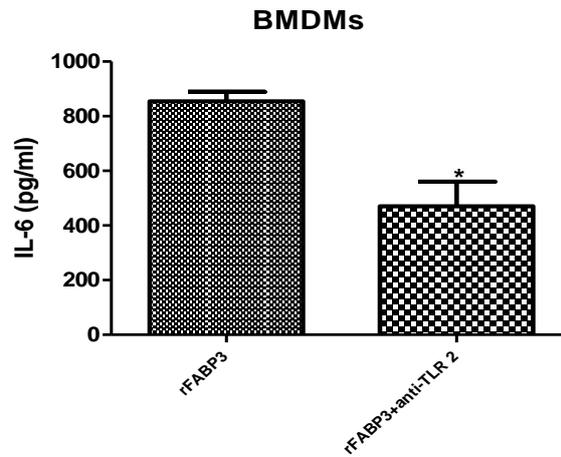
A



B

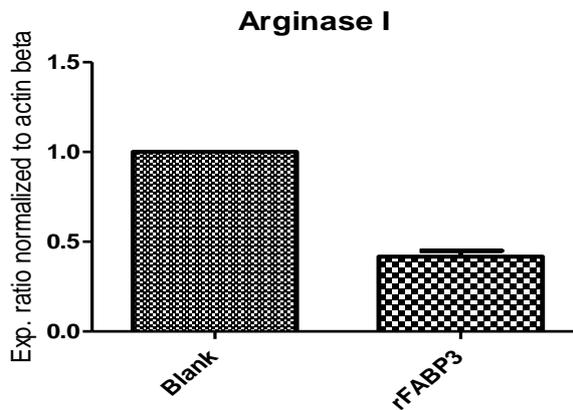


C



**Figure 6-29 rFABP3-induced cytokine secretion by BMDMs via TLR 2.**

BMDMs were activated using rFABP3 (5  $\mu$ g/ml), LPS (100 ng/ml) and HKLM ( $1 \times 10^8$ ). For the groups blocked using anti-TLR2 antibody, cells were pre-treated with 500 ng/ml anti-TLR 2 and then incubated at 37°C for 1h prior to activation. The culture supernatant was used to determine the concentration of TNF  $\alpha$  (A), IL-10 (B) and IL-6 (C) using ELISA. The data are expressed as the mean  $\pm$  S.E. \*\* $p < 0.01$ , \* $p < 0.05$ , NS: not significant.



**Figure 6-30 Effects of rFABP3 on macrophage arginase I expression.**

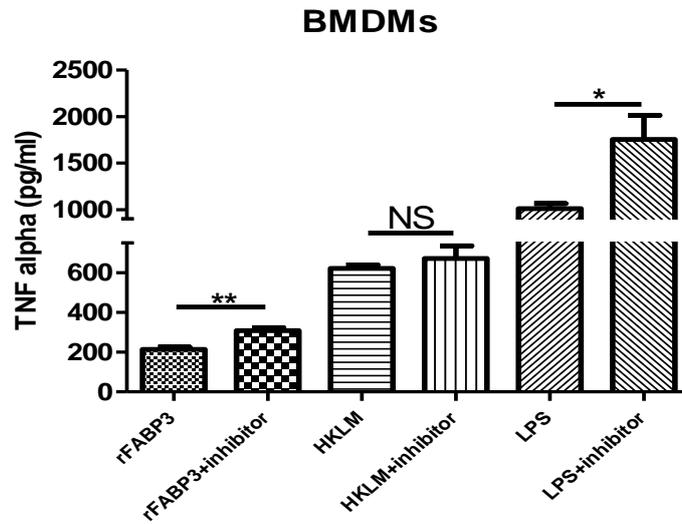
Here we employed real-time PCR to investigate the expression status of arginase I after rFABP3 treatment. As indicated in Fig. 6-30, the arginase I mRNA abundance in the rFABP3-treated BMDMs decreased down to approximate half of that in the blank, indicating that these macrophages are not alternatively activated. In summary, these results suggest that rFABP3 activates BMDMs to promote cytokine production via bypassing the alternative activation pathway.

### **3.6 Effects of an adipocyte FABP inhibitor on cytokines induced by rFABP3**

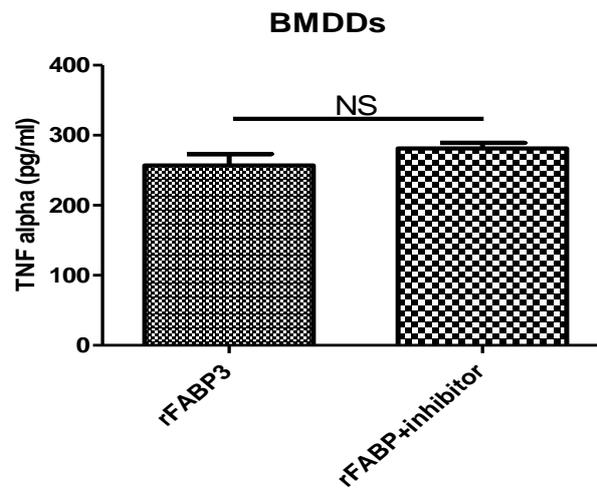
Adipocyte FABP (aFABP) has been shown to be closely linked to macrophage cytokine output (Makowski et al., 2001). A selective inhibitor of aFABP, BMS309403, can bind to its fatty acid binding cavity in a nM range, indirectly inducing reduction of aFABP and cytokine mRNA expression (Hui et al., 2010). In my study, this inhibitor was used to try to understand the relationship between aFABP and rFABP3-induced cytokines by BMDMs.

Unexpectedly, TNF  $\alpha$  secretion did not reduce in the activated cells in the presence of aFABP inhibitor. On the contrary, the inhibitor significantly increased the level of TNF  $\alpha$  in the rFABP3-treated ( $p < 0.01$ ) and LPS-activated cells ( $p < 0.05$ ) but did not affect HKLM-activated cells (Fig. 6-31 A). In the rFABP3-treated BMDDs, there was no significant difference with or without addition of the inhibitor with regard to TNF  $\alpha$  production (Fig. 6-31 B).

A



B



**Figure 6-31 Effects of adipocyte FABP inhibitor on cytokines induced by rFABP3.**

BMDMs (A) and BMDDs (B) were activated using rFABP3 (5  $\mu$ g/ml), LPS (100 ng/ml) and HKLM ( $1 \times 10^8$ ). For the groups treated with the inhibitor, cells were pre-incubated with 50  $\mu$ M inhibitor at 37°C for 30 min prior to activation. The data are expressed as the mean  $\pm$  S.E. \*\* $p < 0.01$ , \* $p < 0.05$ , NS: not significant.

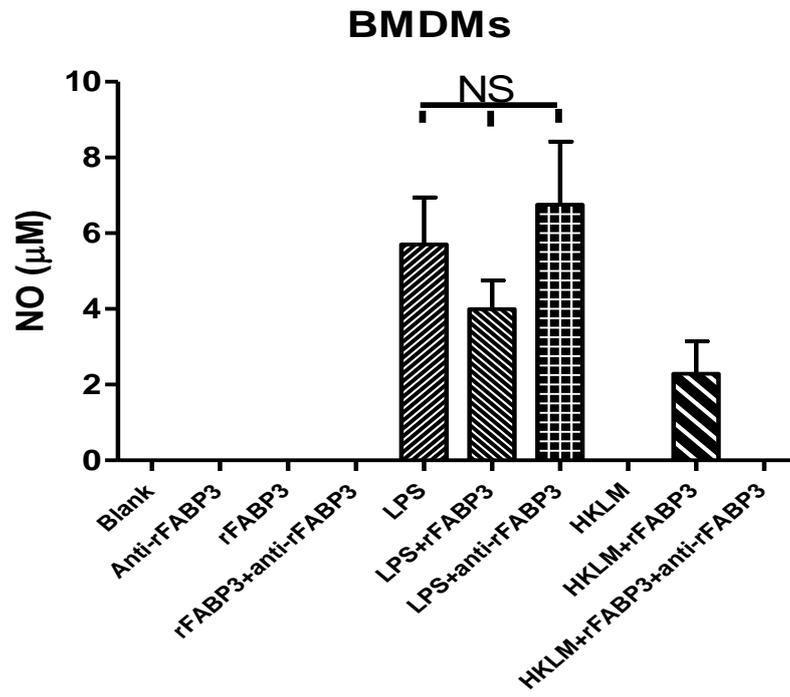
### **3.7 FABP3 combines with HKLM to promote NO production by BMDMs in a dose-dependent manner**

NO is a radical molecule that plays a role in the infection of *Echinococcus* species. Previous studies have showed that it appears to execute double functions by suppressing of parasite growth and modulating host immune response against parasites (Dai and Gottstein, 1999; Kanazawa et al., 1993). In this experiment, rFABP3 did not induce NO production by BMDMs nor did it promote NO production in the LPS-stimulated cells (Fig. 6-32 A and B). After stimulation with rFABP3, the treated dendritic cells did not synthesize appreciable NO, whereas there was a considerable amount of NO produced by the LPS-mediated cells (Fig. 6-32 C).

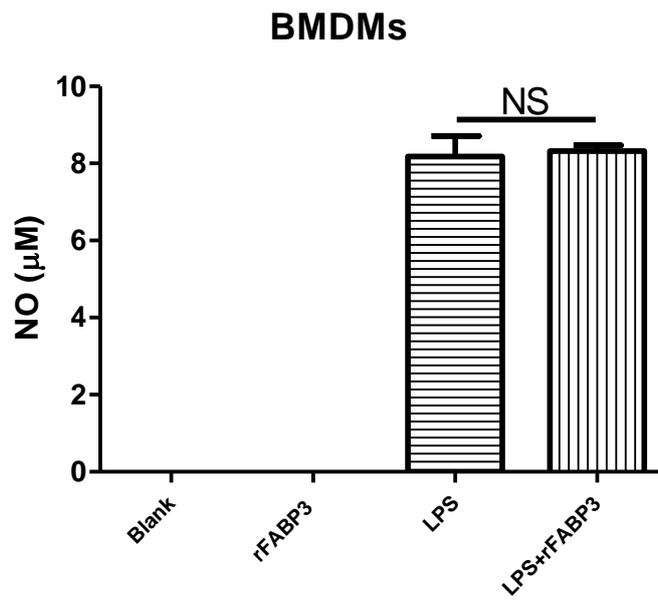
Interestingly, HKLM was shown not to induce NO synthesis but the combination of HKLM and rFABP3 did induce detectable NO. This NO was completely eliminated by addition of specific antibody against rFABP3 (Fig. 6-32 A). Moreover, in the HKLM-treated cells a higher dose of rFABP3 (15 µg/ml) appeared to induce a higher level of NO than a lower dose (10 µg/ml) (Fig. 6-32 D).

Taken together, it is concluded that rFABP3 combines with HKLM to induce macrophage NO production in a dose-dependent manner although both lack the capacity to stimulate NO synthesis.

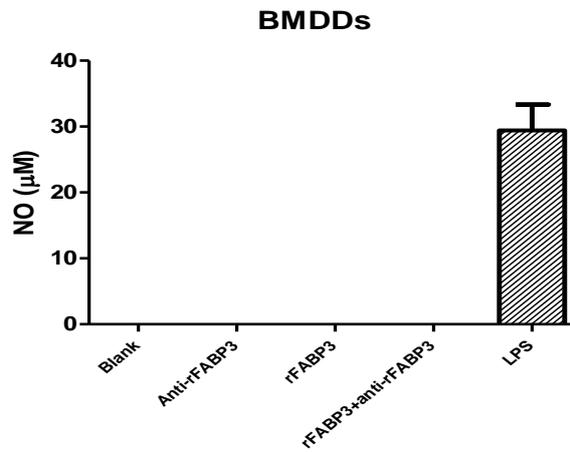
A



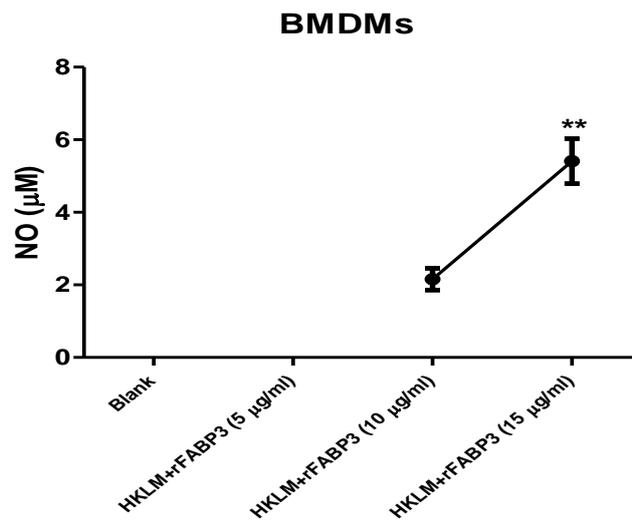
B



C



D



**Figure 6-32 Effects of rFABP3 on NO production by BMDMs and BMDDs.**

Except **D**, BMDMs and BMDDs were activated using rFABP3 (10  $\mu\text{g/ml}$ ), LPS (100 ng/ml) and HKLM ( $1 \times 10^8$ ). For the groups treated using anti-rFABP3, rFABP3, LPS or HKLM was mixed with 40  $\mu\text{g/ml}$  anti-rFABP3 and then incubated at 37°C for 1h prior to activation. The data are expressed as the mean  $\pm$  S.E. \*\* $p < 0.01$ , NS: not significant.

## 4. Discussion

In *Echinococcus* infection, there are a variety of parasite-released molecules that have been shown to have immunomodulatory functions, such as antigen B (AgB) (Shepherd et al., 1991; Siracusano et al., 2008) and a cysteine protease (Mejri and Gottstein, 2009; Sako et al., 2007), both of which are involved in skewing the Th1-type response in the early infectious phase towards Th2-type response in the chronic phase. This response alteration is partially driven by progressive subversion of production from Th1-associated cytokines to Th2-associated cytokines, such as IL-13 and IL-4. Here, we showed that *E. multilocularis* rFABP3 was capable of the stimulation of macrophages and dendritic cells to yield high levels of IL-6 and appreciable levels of IL-10 and TNF  $\alpha$ . Moreover, anti-rFABP3 antibody reversely affected significantly the cytokine secretion induced by rFABP3 but not LPS. In future experiments, it will be interesting to check if anti-rFABP3 antibody reverses the secretion of TNF  $\alpha$  induced by zymosan or HKLM due to the concern that the effects observed might be attributed to yeast contaminants in the protein preparation. A similar cytokine pattern has also been reported in dendritic cells stimulated by AgB (Rigano et al., 2007), a dominant small antigen in the hydatid fluid that also exhibits high binding affinity to fatty acids (Chemale et al., 2005). Unlike rFABP3, however, AgB can also suppress the secretion of cytokines by LPS-stimulated dendritic cells, including IL-6, IL-12p70, IL-10 and TNF  $\alpha$ .

In microbial infection, it is well known that antigen-presenting cells (APCs) including dendritic cells and macrophages are involved in the initiation of immune responses. This process is initiated by pattern-recognition receptors such as Toll-like receptors (TLRs), which can recognize a wide range of molecular structures commonly expressed by pathogens (Kawai and Akira, 2011). Although it is still controversial, a number of studies have shown that helminth- or protozoa-derived products including

excretory/secretory products seem to be able to modulate the hosts' immune systems possibly via TLRs (Carvalho et al., 2008; Perrigoue et al., 2008; Venugopal et al., 2009). An example is a phosphorylcholine-containing glycoprotein, ES-62, secreted by nematode *Acanthocheilonema viteae*. In response to ES-62, secretion of IL-12 and TNF  $\alpha$  was elevated in macrophages and dendritic cells from TLR 2-knockout mice but not from TLR4<sup>-/-</sup> mice, suggesting the induction of cytokines by ES-62 in a TLR4-dependant manner (Goodridge et al., 2005).

In this study, it was shown that rFABP3 activated TLR 2 transformant cells but not those transfected with other TLR receptors. Moreover, the addition of anti-TLR 2 antibody significantly down-regulated cytokine production in BMDMs. These findings support the idea that rFABP3-induced cytokine secretion is in a TLR 2-dependant manner. However, we do not have sufficient evidence to rule out that possible other mechanisms exist, whereby rFABP3 affects macrophage or/and dendritic cell cytokine expression. A few helminth antigens have been shown to trigger Th2-type responses by immune cells potentially via interacting with TLRs, but very little about the underlying pathways are understood (MacDonald and Maizels, 2008). Recently, it has been shown that phosphatidylserine fractionated from *S. mansoni* eggs and adult worms has been shown to polarize dendritic cell maturation, leading to reduced secretion of IL-12p70. The phosphatidylserine-containing fraction activated TLR 2 transfected human cells and the macrophages from TLR2<sup>-/-</sup> mice did not respond to stimulation of phosphatidylserine compared to these from wild-type mice (van der Kleij et al., 2002). Helminth-induced signalling cascades appear to differ from those of conventional ligands such as TLR 4 agonist, LPS (Goodridge et al., 2005; MacDonald and Maizels, 2008). It is still feasible that extracellular rFABP3 directly stimulates endogenous factors or other receptors of macrophages and dendritic cells to influence cytokine secretion, such as eosinophil-derived neurotoxin, an endogenous agonist for TLR 2 (Yang et al., 2008), or C-type lectin receptors

through which schistosomal soluble antigens have been shown to modulate the normal functions of dendritic cells (van Liempt et al., 2007). Future experiments need to be done to define the connections of these factors in rFABP3-induced cytokine secretion. As mentioned previously a concern that yeast contaminants may be involved remains. Whilst the purified yeast control that was prepared in parallel did not induce secretion of TNF  $\alpha$  by BMDMs, it is still possible that the elevated levels of cytokines in rFABP3-treated groups are attributed to contamination of the medium components because of strikingly increased secretion of cytokines induced by the concentrated medium.

In our experiments, we noticed that the combination of rFABP3 and LPS stimulation induced comparable levels of TNF  $\alpha$  when they were given individually, and the sequential addition of these two stimuli had significant outcomes with regard to cytokine production. It may be due to macrophage exhaustion, a phenomenon that cells progressively lose response upon stimulation (Hernandez-Ruiz et al., 2010).

In rFABP3-treated macrophages, expression of arginase I, a prototypic marker for murine alternatively-activated macrophages that are involved in wound healing, was lower than that in the untreated cells, suggesting classical activation. Mounting evidence shows that arginase I is commonly induced during helminth infection and that alternatively-activated macrophages can both enhance parasite survival or even clearance (Stempin et al., 2010). Noticeably, arginase I expression tends to be influenced by other factors, such as animal age (Sugawara et al., 2011).

An intriguing finding in this study was that the inhibitor against macrophage aFABP significantly promoted TNF  $\alpha$  secretion in both rFABP3- and LPS-stimulated BMDMs. These results seem to not be consistent with the previous data that suppression of aFABP activity by the inhibitor leads to decreased transcription or secretion of TNF  $\alpha$  (Furuhashi et al., 2007; Hui et al., 2010). A positive feedback

pathway was proposed in the aFABP-involved inflammatory cytokine regulation, as this inhibitor suppresses not only c-Jun NH<sub>2</sub>-terminal kinases phosphorylation, activator protein-1 transactivation, but also aFABP promoter activity (Hui et al., 2010). Here no evidence was provided to clarify how the aFABP inhibitor was able to increase cytokine production. A feasible explanation is that the translation efficacy is actually promoted although the transcription of TNF  $\alpha$  is downregulated. Future experiments will be required to evaluate the transcriptional and translational efficacies of this pro-inflammatory cytokine.

NO is one of radical molecules produced in response to the infection of *Echinococcus* species. Previous studies have shown that NO is detrimental to *E. multilocularis* protoscoleces (Dai and Gottstein, 1999) and *E. granulosus* vesicles and protoscoleces (Steers et al., 2001; Zeghir-Bouteldja et al., 2009). However, NO also serves as a modulator to suppress host immune responses against *Echinococcus* infection (Dai et al., 2003), allowing parasite survival. These data suggest that NO plays double roles in echinococcosis. A key question is how parasites keep the delicate balance of NO in their milieu. It is so far documented that extracts of cyst wall of *Echinococcus* species can suppress macrophage NO production and 14-3-3 antigen appears to contribute to NO suppression (Andrade et al., 2004; Steers et al., 2001). Here, we showed that *E. multilocularis* recombinant FABP3 did not affect NO synthesis in both resting and LPS-activated macrophages but induced macrophages to release NO in the presence of HKLM, an agonist for TLR 2, and this elevated NO was attenuated upon the addition of anti-rFABP3 antibody *in vitro*. It is unclear whether or not the cooperative induction of rFABP3 is specific for HKLM. It has been shown that NO can be generated through activations of TLRs, including TLR 1/2, 2, 3, 4, 7 and 9 (Tukel et al., 2009; Tumurkhuu et al., 2010). Moreover, NO production may be induced in different signalling pathways in distinct type cells (Ganster et al., 2001; Kleinert et al., 2004). In mosquitoes, resistance to the malaria parasite *Plasmodium falciparum*

infection was observed and attributed to bacterium-derived reactive oxygen species, which results from the interactions between the parasite and *Enterobacter* bacterium that resides in the gut of wild mosquitoes (Cirimotich et al., 2011). It is possible that FABP3 may be involved in immune-modulation in the intestinal tracts of definite hosts via NO stemming from the interactions of FABP3 and intestinal residents. It will be interesting to test the combinations of this antigen and other microbes in an aspect of NO production, especially commensal microbiota residing in the guts of hosts.

## Chapter 7 General discussion and conclusions

### 1. General discussion

Although the biology, biochemistry and genomics of *E. multilocularis* have been extensively investigated (Brehm, 2010b; Brehm and Spiliotis, 2008; Gottstein and Hemphill, 2008; McManus, 2009; Vuitton, 2003), studies on the dynamic transcription of the chromosomes are required to bridge the gaps between genotype and phenotype. To address this question, this project aimed to track the developmental expression profiling of mRNA and miRNA in order to identify the stage-specific molecules via comparison of miRNA and mRNA transcriptomes across different developmental stages.

As part of this work a new approach to increase mRNA transcriptome coverage by removal of polyadenylated mitochondrial transcripts was designed (Chapter 2). In addition genome-wide miRNA transcriptomes were derived (Chapter 3) and a phylogenetic and functional analysis of FABPs was undertaken (Chapters 4-6), which were found to be differentially expressed across the developmental stages of *E. multilocularis*. This general discussion chapter gives a broader overview of the integrated results of Chapters 3-6 and their significance, and insights to direct future experiments by comparative review of current knowledge in the field. For more detailed and specific consideration of the results, the readers therefore need to refer to individual chapters.

In combined use with the latest '-omics' technologies, the developmental regulation and site specific location of *Echinococcus* species proteins have been recently investigated in the excretory/secretory products (Virginio et al., 2012) and different metacestode components, including the germinal layer (Monteiro et al., 2010), hydatid cyst fluid (Aziz et al., 2011; Monteiro et al., 2010) and protoscoleces (Wang et al.,

2009a). In addition real-time changes of hepatic gene expression in response to the infection of *E. multilocularis* in mice have also been studied (Gottstein et al., 2010; Lin et al., 2011). Genome-wide gene profiles of Echinococcus species have not been explored so far. Here the studies on the transcriptome of the different developmental stages using next-generation sequencing technology add new knowledge, which greatly impacts major areas of Echinococcus biology together with a wealth of proteomic and metabolomic data.

Our results uncover that stage-specific molecules that are expressed exclusively at one of the developmental stages just constitute only a minor proportion of the total (Chapter 3). Consistent with this finding, a survey using RNA-seq revealed that approximately 90% of 12,088 transcripts were stably expressed (expression fold-change less than 1.25) in sexual and asexual freshwater planarian *S. mediterranea*, whilst only 6% specifically expressed in either sexual or asexual animals (Resch et al., 2012). This raises the question whether these stage- or strain-specific molecules are only contributors responsible for their phenotypic divergence or not. These specifically expressed molecules may be essential, but their biological functions seem hard to fully explain individual phenotype. Using two different budding yeast *Saccharomyces cerevisiae* strains  $\Sigma$ 1278b and S288c whose genomes are almost identical, a recent study demonstrated that the genes essential only in either  $\Sigma$ 1278b or S288c were enriched in distinct annotated cellular functions, suggesting that genetic interactions are also attributed to divergence of phenotypes between individuals (Dowell et al., 2010). Therefore it is possible that deep understanding of *E. multilocularis* biology will benefit from genetic interaction studies based on transcriptome profiling.

In this thesis the transcriptomes of the early developmental stage (cell aggregate) and two larval stages (vesicles and dormant or activated protoscoleces) of *E. multilocularis* have been studied but the other stages including adult are of particular

interest to be explored in future experiments, thus establishing the entire gene profiling landscape at the life cycle. Here it is worth mentioning that some transcriptome profiling, such as in activated protoscolexes, does not present the real chromosome activities in response to host's microenvironment during the infection. This can be indirectly shown by the findings that the components of the hydatid cyst fluid from humans, sheep and cattle were remarkably distinguishable (Aziz et al., 2011), and mice with different genetic background displayed different susceptibility to *E. multilocularis* infection (Nakao et al., 2011).

As introduced in Chapter 1, *E. multilocularis* neoblasts are regarded to be a cell population that is able to undergo mitosis, which may be essential in the development, differentiation and growth (Spiliotis et al., 2008). Moreover, the *in vitro* development of mature vesicles from the primary cells is attributed to the presence of these mitotically-active cells. So it is envisaged that neoblast studies will allow us to have a better understanding of *E. multilocularis* biological features using functional genomics approaches such as loss- or gain-of-function. Like planarian neoblasts, *E. multilocularis* neoblasts display similar morphological characters, containing a large nucleus with a spherical nucleolus and very little cytoplasm and being small with an average diameter of 5  $\mu\text{m}$  (Reddien and Sanchez Alvarado, 2004; Spiliotis et al., 2008). Based on these morphological features, it is feasible to efficiently isolate *E. multilocularis* neoblasts (Hayashi et al., 2006). The transcriptome profiling of the neoblasts has not been studied in this thesis but of importance and interest is to probe the gene expression patterns of this sub-population in future experiments.

RNA variation is an interesting area of study as it will help us understand the complexity of transcriptome and proteome. As shown in Chapter 3, RNA variants were extensively present in miRNA data, called as isomiRs that refer to miRNAs that are different in a size and component, especially at the 3' or 5' terminal (Kuchenbauer et al., 2008; Morin et al., 2008), in contrast to miRNA reference sequences. It has

been showed that isomiRs expression is closely linked with physiological status and these miRNA variants are likely to cooperatively target common biological pathways with canonical miRNAs (Fernandez-Valverde et al., 2010; Wyman et al., 2011). Of particular interest are 3' variations not matching the genome, predominantly adenosine and uridine additions, which were commonly reported in several animals (Kuchenbauer et al., 2008; Wyman et al., 2011). Although their biological roles are not fully understood, these non-templated additions seem to enhance miRNA stability and miRNA:target interactions (Fernandez-Valverde et al., 2010). Recently such non-templated RNA variations were also reported in human protein-coding genes (Li et al., 2011). Rather than sequencing artefacts, these differences between RNA and DNA sequences (RDD) have been non-randomly found extensively in two or more individuals. Moreover, it was proved the presence of translated peptide sequences that correspond to mRNA variant sequences. Systematic analysis also revealed that RDD sites were not evenly located in chromosomes and mRNA transcripts, dominantly in coding exons and 3' untranslated regions (Li et al., 2011). From the characteristics of RDD sites and isomiRs, non-templated variants in miRNAs and mRNA transcripts are not likely to be generated by the same mechanism (Wyman et al., 2011).

Although they are functionally enriched in helicase activity and protein and nucleotide binding (Wyman et al., 2011), biological implications of these genes containing RDD sites remain unclear. However it is clear that some of non-templated mRNA variations give rise to production of various proteins, adding a more layer of proteomic complexity. This raises the question whether parasites alter antigen properties via the same mechanism or not. To date, coding-gene RDD sites have not been described in cestode parasites that comprise many gene families such as Eg/Em95 in *Echinococcus* spp. (Haag et al., 2009), 45W in *Taenia* spp. (Gauci and Lightowlers, 2001; Waterkeyn et al., 1995) and 8kDa in both (Jia et al., 2011). It was shown that *T. solium* 45W transcripts exhibited diversity at the nucleotide and amino acid levels,

and that with exception of 45W-B and -C transcripts the genomic loci that correspond to the other transcripts were not found (Gauci and Lightowlers, 2001; Zheng et al., 2008). The number of 45W transcripts was far more than three of the genomic loci that were determined experimentally (Gauci and Lightowlers, 2001) and validated via Blast-searching *T. solium* genome (our unpublished data). This great discrepancy between 45W mRNA and genomic locus numbers may be due to introduction of nucleotide variants in transcripts, which differ from corresponding DNA sequences. It is necessary to systematically assess RNA variations using transcriptome data by deep sequencing technologies, which has implications in illustration of antigen diversity and possibly leads to discovery of invasive or evasive mechanisms during the infection of cestode parasites.

Another interesting research area not included in our study is to integrate miRNA and mRNA transcriptome profiling to establish a miRNA target database, which is designed to display expression levels of both miRNAs and their target genes across the life cycle, and then evaluate the expression connections between miRNAs and corresponding mRNA targets. Such an integrated database will allow one to quickly define which putative targets are required to validate and facilitate identification of miRNAs essential for parasite development and growth.

The other core part of this thesis is with regard to phylogenetic analysis of invertebrate FABPs (Chapter 4) and molecular and functional characterization of *E. multilocularis* FABPs (Chapters 5 and 6), which were identified to be differentially expressed at the developmental stages.  $\beta$ -sheet-containing FABPs are featured by binding to lipids with high affinity, structurally distinguishable from other lipid-binding proteins such as fatty acid- and retinol-binding proteins (FAPs) enriched with  $\alpha$ -helixes in the secondary structure (Jordanova et al., 2009; Kennedy et al., 1997) and  $\alpha$ -helix-enriched polyprotein allergens/antigens (Gomes et al.) (Kennedy, 2000a;

McDermott et al., 1999). In contrast to FABPs and FAPs, in the PA biogenesis mRNAs are organized in tandemly repeated similar or identical units and translated into large polyproteins, followed by cleavage at the consensus sites to generate individual PAs with a mass molecular weight of approximately 15kDa (Kennedy, 2000a). In this study we showed that recombinant FABPs were able to weakly bind to *cis*-parinaric acid but not to DAUDA and retinol. Moreover, it was also demonstrated that native FABP3 was present in the hydatid cyst fluid, consistent with the recent findings that *E. granulosus* FABP was one of the excretory/secretory components of protoscoleces (Virginio et al., 2012) and present in the hydatid cyst fluid (Aziz et al., 2011). These results raise the question of what the role of Echinococcus FABPs is. Our study revealed that *E. multilocularis* FABP3 had effects on the cytokine and nitric oxide production of bone marrow-derived macrophages. This would imply that this lipid chaperon protein plays a role in controlling microenvironment or countering immune defence responses as supposed for NPAs and other lipid-binding proteins (McDermott et al., 1999), offering parasites more chances to survive.

In *E. multilocularis*, *in vivo* functional analyses have been conducted using loss-of-function approaches such as RNA interference (RNAi) (Mizukami et al., 2010; Spiliotis et al., 2010). It will be very interesting to explore the functions of FABPs in aspects of lipid metabolism, infection and immunity in *E. multilocularis*. Although they are found to have a relatively low expression in the samples studied, *E. multilocularis* FABP1 and FABP5 are the most likely to be functional. Future experiments will be needed to explore their expression during the infection. Previous studies have shown macrophage cytokine production is closely linked with adipocyte FABP (Hui et al., 2010; Makowski et al., 2001; Makowski et al., 2005; Shum et al., 2006). To further understand the mechanism whereby *E. multilocularis* FABP promote cytokine secretion, it is worthwhile investigating adipocyte FABP and its pathway in macrophages mediated by *E. multilocularis* FABP.

Relative to previous studies on invertebrate FABP genes (Esteves and Ehrlich, 2006; Esteves et al., 1997), the main strength in our study is to extensively explore FABP genes, especially genomic organizations and post-transcriptional splicing. Obviously, more species needs to be sampled in order to systematically analyze the phylogeny of FABP genes. To help us understand tissue-specific expression and functions of FABPs, future experiments are also required to systematically study the response elements or motifs in the promoter regions of FABP genes, which have been found in several FABP genes (Chmurzynska, 2006; Esteves and Ehrlich, 2006).

## 2. Conclusions

The results presented in this thesis suggest that the miRNA-induced silencing gene regulation systems are present in *E. multilocularis* and miRNA expression exhibits a complex developmental pattern. Of the differentially expressed molecules inferred from transcriptome data, some have been shown to be potential neoblast biomarkers. The FABPs that were identified may have lost fatty acid-binding properties that are commonly reported in trematodes and nematodes. The results of *in vitro* studies demonstrate that FABP3 may play a part in modulation of immune cells during infection. A finding to emerge from this study is that invertebrate FABP variants and functional diversity have shown to be due to gene duplication and alternative splicing rather than only gene duplication that is largely responsible for FABP variants in mammals.

The empirical findings in this study make several contributions to the current literature. First our study will provide informative data to facilitate annotation of *E. multilocularis* genome and to understand dynamic genome transcription during development, which in turn make it possibility to be efficiently used as a model for gene function investigations. The findings on miRNA expression enhance our understanding of miRNA evolution in platyhelminths and serve as a base for

antihelminthic development through targeting the miRNA pathway. Finally, we have described potential roles of extracellular FABP antigens during infection and a role of alternative splicing in invertebrate FABP variant generation.

## References

- 2009a. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* 457, 1028-1032.
- 2009b. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 460, 345-351.
2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8, e1000313.
- Aboobaker, A.A., 2011. Planarian stem cells: a simple paradigm for regeneration. *Trends Cell Biol* 21, 304-311.
- Agorio, A., Chalar, C., Cardozo, S., Salinas, G., 2003. Alternative mRNAs arising from trans-splicing code for mitochondrial and cytosolic variants of *Echinococcus granulosus* thioredoxin Glutathione reductase. *J Biol Chem* 278, 12920-12928.
- Aguilar-Diaz, H., Bobes, R.J., Carrero, J.C., Camacho-Carranza, R., Cervantes, C., Cevallos, M.A., Davila, G., Rodriguez-Dorantes, M., Escobedo, G., Fernandez, J.L., Fragoso, G., Gaytan, P., Garciarubio, A., Gonzalez, V.M., Gonzalez, L., Jose, M.V., Jimenez, L., Laclette, J.P., Landa, A., Larralde, C., Morales-Montor, J., Morett, E., Ostoa-Saloma, P., Sciutto, E., Santamaria, R.I., Soberon, X., de la Torre, P., Valdes, V., Yanez, J., 2006. The genome project of *Taenia solium*. *Parasitol Int* 55 Suppl, S127-130.
- Alvite, G., Canclini, L., Corvo, I., Esteves, A., 2008. Two novel *Mesocestoides vogae* fatty acid binding proteins--functional and evolutionary implications. *FEBS J* 275, 107-116.
- Alvite, G., Di Pietro, S.M., Santome, J.A., Ehrlich, R., Esteves, A., 2001. Binding properties of *Echinococcus granulosus* fatty acid binding protein. *Biochim Biophys Acta* 1533, 293-302.
- Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., Tuschl, T., 2003. A uniform system for microRNA annotation. *RNA* 9, 277-279.
- Andrade, M.A., Siles-Lucas, M., Espinoza, E., Perez Arellano, J.L., Gottstein, B., Muro, A., 2004. *Echinococcus multilocularis* laminated-layer components and the E14t 14-3-3 recombinant protein decrease NO production by activated rat macrophages in vitro. *Nitric Oxide* 10, 150-155.
- Angelucci, F., Johnson, K.A., Baiocco, P., Miele, A.E., Brunori, M., Valle, C., Vigorosi, F., Troiani, A.R., Liberti, P., Cioli, D., Klinkert, M.Q., Bellelli, A., 2004. *Schistosoma mansoni* fatty acid binding protein: specificity and functional control as revealed by crystallographic structure. *Biochemistry* 43, 13000-13011.
- Atshaves, B.P., Martin, G.G., Hostetler, H.A., McIntosh, A.L., Kier, A.B., Schroeder, F., 2010. Liver fatty acid-binding protein and obesity. *J Nutr Biochem* 21, 1015-1032.
- Audic, S., Claverie, J.M., 1997. The significance of digital gene expression profiles. *Genome Res* 7, 986-995.
- Aziz, A., Zhang, W., Li, J., Loukas, A., McManus, D.P., Mulvenna, J., 2011. Proteomic characterisation of *Echinococcus granulosus* hydatid cyst fluid from sheep, cattle and humans. *J Proteomics* 74, 1560-1572.
- Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P., 2008. The impact of microRNAs on protein output. *Nature* 455, 64-71.

- Bao, Y., Lu, Z., Zhou, M., Li, H., Wang, Y., Gao, M., Wei, M., Jia, W., 2011. Serum levels of adipocyte fatty acid-binding protein are associated with the severity of coronary artery disease in Chinese women. *PLoS One* 6, e19115.
- Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.
- Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215-233.
- Barth, C., Greferath, U., Kotsifas, M., Tanaka, Y., Alexander, S., Alexander, H., Fisher, P.R., 2001. Transcript mapping and processing of mitochondrial RNA in *Dictyostelium discoideum*. *Curr Genet* 39, 355-364.
- Becker, M.M., Kalinna, B.H., Waine, G.J., McManus, D.P., 1994. Gene cloning, overproduction and purification of a functionally active cytoplasmic fatty acid-binding protein (Sj-FABPC) from the human blood fluke *Schistosoma japonicum*. *Gene* 148, 321-325.
- Belshaw, R., Bensasson, D., 2006. The rise and falls of introns. *Heredity* 96, 208-213.
- Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., Lai, E.C., 2007. Mammalian mirtron genes. *Mol Cell* 28, 328-336.
- Berezikov, E., Cuppen, E., Plasterk, R.H., 2006. Approaches to microRNA discovery. *Nat Genet* 38 Suppl, S2-7.
- Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C., Mashiyama, S.T., Al-Lazikani, B., Andrade, L.F., Ashton, P.D., Aslett, M.A., Bartholomeu, D.C., Blandin, G., Caffrey, C.R., Coghlan, A., Coulson, R., Day, T.A., Delcher, A., DeMarco, R., Djikeng, A., Eyre, T., Gamble, J.A., Ghedin, E., Gu, Y., Hertz-Fowler, C., Hirai, H., Hirai, Y., Houston, R., Ivens, A., Johnston, D.A., Lacerda, D., Macedo, C.D., McVeigh, P., Ning, Z., Oliveira, G., Overington, J.P., Parkhill, J., Pertea, M., Pierce, R.J., Protasio, A.V., Quail, M.A., Rajandream, M.A., Rogers, J., Sajid, M., Salzberg, S.L., Stanke, M., Tivey, A.R., White, O., Williams, D.L., Wortman, J., Wu, W., Zamanian, M., Zerlotini, A., Fraser-Liggett, C.M., Barrell, B.G., El-Sayed, N.M., 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460, 352-358.
- Bethke, A., Fielenbach, N., Wang, Z., Mangelsdorf, D.J., Antebi, A., 2009. Nuclear hormone receptor regulation of microRNAs controls developmental progression. *Science* 324, 95-98.
- Binas, B., Danneberg, H., McWhir, J., Mullins, L., Clark, A.J., 1999. Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization. *FASEB J* 13, 805-812.
- Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlars, M., Cullen, B.R., 2010. A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Mol Cell* 37, 135-142.
- Bonne, A., Gosele, C., den Bieman, M., Gillissen, G., Kreitler, T., Pravenec, M., Kren, V., van Lith, H., van Zutphen, B., 2003. Sequencing and chromosomal localization of Fabp6 and an intronless Fabp6 segment in the rat. *Mol Biol Rep* 30, 173-176.
- Borchers, T., Hohoff, C., Buhlmann, C., Spener, F., 1997. Heart-type fatty acid binding protein - involvement in growth inhibition and differentiation. *Prostaglandins Leukot Essent Fatty Acids* 57, 77-84.
- Boutros, C., Somasundar, P., Razzak, A., Helton, S., Espat, N.J., 2010. Omega-3 fatty acids: investigations from cytokine regulation to pancreatic cancer gene suppression. *Arch Surg* 145, 515-520.
- Braun, L., Cannella, D., Ortet, P., Barakat, M., Sautel, C.F., Kieffer, S., Garin, J., Bastien, O., Voinnet, O., Hakimi, M.A., 2010. A complex small RNA repertoire is

- generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. *PLoS Pathog* 6, e1000920.
- Brehm, K., 2010a. *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. *Parasitology* 137, 537-555.
- Brehm, K., 2010b. The role of evolutionarily conserved signalling systems in *Echinococcus multilocularis* development and host-parasite interaction. *Med Microbiol Immunol* 199, 247-259.
- Brehm, K., Spiliotis, M., 2008. The influence of host hormones and cytokines on *Echinococcus multilocularis* signalling and development. *Parasite* 15, 286-290.
- Brehm, K., Wolf, M., Beland, H., Kroner, A., Frosch, M., 2003. Analysis of differential gene expression in *Echinococcus multilocularis* larval stages by means of spliced leader differential display. *Int J Parasitol* 33, 1145-1159.
- Brunetti, E., Kern, P., Vuitton, D.A., 2010. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Trop* 114, 1-16.
- Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T., 2005. Protein structure prediction servers at University College London. *Nucleic Acids Res* 33, W36-38.
- Budke, C.M., White, A.C., Jr., Garcia, H.H., 2009. Zoonotic larval cestode infections: neglected, neglected tropical diseases? *PLoS Negl Trop Dis* 3, e319.
- Cabre, A., Lazaro, I., Girona, J., Manzanares, J.M., Marimon, F., Plana, N., Heras, M., Masana, L., 2008. Plasma fatty acid binding protein 4 is associated with atherogenic dyslipidemia in diabetes. *J Lipid Res* 49, 1746-1751.
- Caraballo, L., Puerta, L., Jimenez, S., Martinez, B., Mercado, D., Avjiouglu, A., Marsh, D., 1997. Cloning and IgE binding of a recombinant allergen from the mite *Blomia tropicalis*, homologous with fatty acid-binding proteins. *Int Arch Allergy Immunol* 112, 341-347.
- Carvalho, L.P., Pearce, E.J., Scott, P., 2008. Functional dichotomy of dendritic cells following interaction with *Leishmania braziliensis*: infected cells produce high levels of TNF-alpha, whereas bystander dendritic cells are activated to promote T cell responses. *J Immunol* 181, 6473-6480.
- Cazalla, D., Yario, T., Steitz, J.A., 2010. Down-regulation of a host microRNA by a Herpesvirus *saimiri* noncoding RNA. *Science* 328, 1563-1566.
- Chan, S.L., Ong, S.T., Ong, S.Y., Chew, F.T., Mok, Y.K., 2006. Nuclear magnetic resonance structure-based epitope mapping and modulation of dust mite group 13 allergen as a hypoallergen. *J Immunol* 176, 4852-4860.
- Chemale, G., Ferreira, H.B., Barrett, J., Brophy, P.M., Zaha, A., 2005. *Echinococcus granulosus* antigen B hydrophobic ligand binding properties. *Biochim Biophys Acta* 1747, 189-194.
- Chen, J.F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., Wang, D.Z., 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38, 228-233.
- Chen, X.S., Collins, L.J., Biggs, P.J., Penny, D., 2009. High throughput genome-wide survey of small RNAs from the parasitic protists *Giardia intestinalis* and *Trichomonas vaginalis*. *Genome Biol Evol* 1, 165-175.
- Chmurzynska, A., 2006. The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism. *J Appl Genet* 47, 39-48.
- Chunchob, S., Grams, R., Viyanant, V., Smooker, P.M., Vichasri-Grams, S., 2010. Comparative analysis of two fatty acid binding proteins from *Fasciola gigantica*. *Parasitology* 137, 1805-1817.

- Cirimotich, C.M., Dong, Y., Clayton, A.M., Sandiford, S.L., Souza-Neto, J.A., Mulenga, M., Dimopoulos, G., 2011. Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. *Science* 332, 855-858.
- Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., Bruce, S.J., Lee, C.C., Ranade, S.S., Peckham, H.E., Manning, J.M., McKernan, K.J., Grimmond, S.M., 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5, 613-619.
- Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., Cohen, L., Danon, T., Perzov, N., Alon, U., 2008. Dynamic proteomics of individual cancer cells in response to a drug. *Science* 322, 1511-1516.
- Cook, J.S., Lucas, J.J., Sibley, E., Bolanowski, M.A., Christy, R.J., Kelly, T.J., Lane, M.D., 1988. Expression of the differentiation-induced gene for fatty acid-binding protein is activated by glucocorticoid and cAMP. *Proc Natl Acad Sci U S A* 85, 2949-2953.
- Craig, P.S., McManus, D.P., Lightowers, M.W., Chabalgoity, J.A., Garcia, H.H., Gavidia, C.M., Gilman, R.H., Gonzalez, A.E., Lorca, M., Naquira, C., Nieto, A., Schantz, P.M., 2007. Prevention and control of cystic echinococcosis. *Lancet Infect Dis* 7, 385-394.
- Cucher, M., Prada, L., Mourglia-Ettlin, G., Dematteis, S., Camicia, F., Asurmendi, S., Rosenzvit, M., 2011. Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. *Int J Parasitol* 41, 439-448.
- Dai, W.J., Gottstein, B., 1999. Nitric oxide-mediated immunosuppression following murine *Echinococcus multilocularis* infection. *Immunology* 97, 107-116.
- Dai, W.J., Waldvogel, A., Jungi, T., Stettler, M., Gottstein, B., 2003. Inducible nitric oxide synthase deficiency in mice increases resistance to chronic infection with *Echinococcus multilocularis*. *Immunology* 108, 238-244.
- De, S., Pal, D., Ghosh, S.K., 2006. *Entamoeba histolytica*: computational identification of putative microRNA candidates. *Exp Parasitol* 113, 239-243.
- Dellaporta, S.L., Xu, A., Sagasser, S., Jakob, W., Moreno, M.A., Buss, L.W., Schierwater, B., 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A* 103, 8751-8756.
- Deplazes, P., Grimm, F., Sydler, T., Tanner, I., Kapel, C.M., 2005. Experimental alveolar echinococcosis in pigs, lesion development and serological follow up. *Vet Parasitol* 130, 213-222.
- Devaney, E., Winter, A.D., Britton, C., 2010. microRNAs: a role in drug resistance in parasitic nematodes? *Trends Parasitol* 26, 428-433.
- Dieckmann, A., Frank, W., 1988. Isolation of viable germinal cells from *Echinococcus multilocularis*. *Parasitol Res* 74, 297-298.
- Dohm, J.C., Lottaz, C., Borodina, T., Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 36, e105.
- Donis-Keller, H., 1979. Site specific enzymatic cleavage of RNA. *Nucleic Acids Res* 7, 179-192.
- Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., Nislow, C., Giaever, G., Phillips, P.C., Fink, G.R., Gifford, D.K., Boone, C., 2010. Genotype to phenotype: a complex problem. *Science* 328, 469.

- Echeverria, C.I., Isolabella, D.M., Prieto Gonzalez, E.A., Leonardelli, A., Prada, L., Perrone, A., Fuchs, A.G., 2010. Morphological and biological characterization of cell line developed from bovine *Echinococcus granulosus*. *In Vitro Cell Dev Biol Anim* 46, 781-792.
- Eckert, J., Deplazes, P., 2004. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. *Clin Microbiol Rev* 17, 107-135.
- Eckert, J., Gemmell, M.A., Meslin, F.-X., Pawlowski, Z.S., 2002. WHO/OIE manual on echinococcosis in humans and animals: a public health problem of global concern.
- Edqvist, J., Blomqvist, K., 2006. Fusion and fission, the evolution of sterol carrier protein-2. *J Mol Evol* 62, 292-306.
- Eriksson, T.L., Rasool, O., Huecas, S., Whitley, P., Cramer, R., Appenzeller, U., Gafvelin, G., van Hage-Hamsten, M., 2001. Cloning of three new allergens from the dust mite *Lepidoglyphus destructor* using phage surface display technology. *Eur J Biochem* 268, 287-294.
- Eriksson, T.L., Whitley, P., Johansson, E., van Hage-Hamsten, M., Gafvelin, G., 1999. Identification and characterisation of two allergens from the dust mite *Acarus siro*, homologous with fatty acid-binding proteins. *Int Arch Allergy Immunol* 119, 275-281.
- Espino, A.M., Hillyer, G.V., 2001. Identification of fatty acid molecules in a *Fasciola hepatica* immunoprophylactic fatty acid-binding protein. *J Parasitol* 87, 426-428.
- Esteves, A., Dallagiovanna, B., Ehrlich, R., 1993. A developmentally regulated gene of *Echinococcus granulosus* codes for a 15.5-kilodalton polypeptide related to fatty acid binding proteins. *Mol Biochem Parasitol* 58, 215-222.
- Esteves, A., Ehrlich, R., 2006. Invertebrate intracellular fatty acid binding proteins. *Comp Biochem Physiol C Toxicol Pharmacol* 142, 262-274.
- Esteves, A., Joseph, L., Paulino, M., Ehrlich, R., 1997. Remarks on the phylogeny and structure of fatty acid binding proteins from parasitic platyhelminths. *Int J Parasitol* 27, 1013-1023.
- Esteves, A., Portillo, V., Ehrlich, R., 2003. Genomic structure and expression of a gene coding for a new fatty acid binding protein from *Echinococcus granulosus*. *Biochim Biophys Acta* 1631, 26-34.
- Fernandez-Valverde, S.L., Taft, R.J., Mattick, J.S., 2010. Dynamic isomiR regulation in *Drosophila* development. *RNA* 16, 1881-1888.
- Fernandez, C., Gregory, W.F., Loke, P., Maizels, R.M., 2002. Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Mol Biochem Parasitol* 122, 171-180.
- Fernandez, V., Zavala, A., Musto, H., 2001. Evidence for translational selection in codon usage in *Echinococcus* spp. *Parasitology* 123, 203-209.
- Ferreyra, R.G., Burgardt, N.I., Milikowski, D., Melen, G., Kornblihtt, A.R., Dell'Angelica, E.C., Santome, J.A., Ermacora, M.R., 2006. A yeast sterol carrier protein with fatty-acid and fatty-acyl-CoA binding activity. *Arch Biochem Biophys* 453, 197-206.
- Flores, R.R., Diggs, K.A., Tait, L.M., Morel, P.A., 2007. IFN-gamma negatively regulates CpG-induced IL-10 in bone marrow-derived dendritic cells. *J Immunol* 178, 211-218.
- Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., Zamore, P.D., 2005. Normal microRNA maturation and

- germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 3, e236.
- Friedlander, M.R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T.A., Hirst, M., Marra, M., Nusbaum, C., Lee, W.L., Jenkin, J.C., Sanchez Alvarado, A., Kim, J.K., Rajewsky, N., 2009. High-resolution profiling and discovery of planarian small RNAs. *Proc Natl Acad Sci U S A* 106, 11546-11551.
- Friedlander, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N., 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* 26, 407-415.
- Fu, Y., Luo, N., Lopes-Virella, M.F., 2000. Oxidized LDL induces the expression of ALBP/aP2 mRNA and protein in human THP-1 macrophages. *J Lipid Res* 41, 2017-2023.
- Furuhashi, M., Hotamisligil, G.S., 2008. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nat Rev Drug Discov* 7, 489-503.
- Furuhashi, M., Tuncman, G., Gorgun, C.Z., Makowski, L., Atsumi, G., Vaillancourt, E., Kono, K., Babaev, V.R., Fazio, S., Linton, M.F., Sulsky, R., Robl, J.A., Parker, R.A., Hotamisligil, G.S., 2007. Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2. *Nature* 447, 959-965.
- Furuya, K., 1991. An established cell line of larval *Echinococcus multilocularis*. *Int J Parasitol* 21, 233-240.
- Ganformina, M.D., Gutierrez, G., Bastiani, M., Sanchez, D., 2000. A phylogenetic analysis of the lipocalin protein family. *Mol Biol Evol* 17, 114-126.
- Ganster, R.W., Taylor, B.S., Shao, L., Geller, D.A., 2001. Complex regulation of human inducible nitric oxide synthase gene transcription by Stat 1 and NF-kappa B. *Proc Natl Acad Sci U S A* 98, 8638-8643.
- Garcia-Silva, M.R., Frugier, M., Tosar, J.P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S., Cayota, A., 2010. A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. *Mol Biochem Parasitol* 171, 64-73.
- Garcia Silva, M.R., Tosar, J.P., Frugier, M., Pantano, S., Bonilla, B., Esteban, L., Serra, E., Rovira, C., Robello, C., Cayota, A., 2010. Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene* 466, 26-35.
- Gauci, C.G., Lightowers, M.W., 2001. Alternative splicing and sequence diversity of transcripts from the oncosphere stage of *Taenia solium* with homology to the 45W antigen of *Taenia ovis*. *Mol Biochem Parasitol* 112, 173-181.
- Ghosh, Z., Mallick, B., Chakrabarti, J., 2009. Cellular versus viral microRNAs in host-virus interaction. *Nucleic Acids Res* 37, 1035-1048.
- Gibson, G., 2008. The environmental contribution to gene expression profiles. *Nat Rev Genet* 9, 575-581.
- Gissi, C., Pesole, G., 2003. Transcript mapping and genome annotation of ascidian mtDNA using EST data. *Genome Res* 13, 2203-2212.
- Glazov, E.A., Cottee, P.A., Barris, W.C., Moore, R.J., Dalrymple, B.P., Tizard, M.L., 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res* 18, 957-964.
- Gobert, G.N., Stenzel, D.J., Jones, M.K., McManus, D.P., 1997. Immunolocalization of the fatty acid-binding protein Sj-FABPc within adult *Schistosoma japonicum*. *Parasitology* 115 ( Pt 1), 33-39.

- Gomes, M.S., Cabral, F.J., Jannotti-Passos, L.K., Carvalho, O., Rodrigues, V., Baba, E.H., Sa, R.G., 2009. Preliminary analysis of miRNA pathway in *Schistosoma mansoni*. *Parasitol Int* 58, 61-68.
- Gong, Y.N., Li, W.W., Sun, J.L., Ren, F., He, L., Jiang, H., Wang, Q., 2010. Molecular cloning and tissue expression of the fatty acid-binding protein (Es-FABP) gene in female Chinese mitten crab (*Eriocheir sinensis*). *BMC Mol Biol* 11, 71.
- Gonzalez-Estevez, C., Arseni, V., Thambyrajah, R.S., Felix, D.A., Aboobaker, A.A., 2009. Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int J Dev Biol* 53, 493-505.
- Goodridge, H.S., Marshall, F.A., Else, K.J., Houston, K.M., Egan, C., Al-Riyami, L., Liew, F.Y., Harnett, W., Harnett, M.M., 2005. Immunomodulation via novel use of TLR4 by the filarial nematode phosphorylcholine-containing secreted product, ES-62. *J Immunol* 174, 284-293.
- Gordon, S., Martinez, F.O., 2010. Alternative activation of macrophages: mechanism and functions. *Immunity* 32, 593-604.
- Gottstein, B., Hemphill, A., 2008. *Echinococcus multilocularis*: the parasite-host interplay. *Exp Parasitol* 119, 447-452.
- Gottstein, B., Wittwer, M., Schild, M., Merli, M., Leib, S.L., Muller, N., Muller, J., Jaggi, R., 2010. Hepatic gene expression profile in mice perorally infected with *Echinococcus multilocularis* eggs. *PLoS One* 5, e9779.
- Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36, D154-158.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degan, B.M., Rokhsar, D.S., Bartel, D.P., 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455, 1193-1197.
- Guislain, M.H., Raoul, F., Giraudoux, P., Terrier, M.E., Froment, G., Ferte, H., Poulle, M.L., 2008. Ecological and biological factors involved in the transmission of *Echinococcus multilocularis* in the French Ardennes. *J Helminthol* 82, 143-151.
- Haag, K.L., Gottstein, B., Ayala, F.J., 2009. The EG95 antigen of *Echinococcus* spp. contains positively selected amino acids, which may influence host specificity and vaccine efficacy. *PLoS One* 4, e5362.
- Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., Tuschl, T., 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* 44, 3-12.
- Halanych, K.M., 2004. The new view of animal phylogeny. *Annu Rev Ecol Evol Syst* 35, 229-256.
- Hastings, K.E., 2005. SL trans-splicing: easy come or easy go? *Trends Genet* 21, 240-247.
- Hatfield, S.D., Shcherbata, H.R., Fischer, K.A., Nakahara, K., Carthew, R.W., Ruohola-Baker, H., 2005. Stem cell division is regulated by the microRNA pathway. *Nature* 435, 974-978.
- Hauerland, N.H., Chen, X., Andolfatto, P., Chisholm, J.M., Wang, Z., 1993. Developmental changes of FABP concentration, expression, and intracellular distribution in locust flight muscle. *Mol Cell Biochem* 123, 153-158.
- Hauerland, N.H., Chisholm, J.M., 1990. Fatty acid binding protein in flight muscle of the locust, *Schistocerca gregaria*. *Biochim Biophys Acta* 1047, 233-238.
- Hauerland, N.H., Spener, F., 2004. Fatty acid-binding proteins--insights from genetic manipulations. *Prog Lipid Res* 43, 328-349.

- Hayashi, T., Asami, M., Higuchi, S., Shibata, N., Agata, K., 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev Growth Differ* 48, 371-380.
- He, Y., Yang, X., Wang, H., Estephan, R., Francis, F., Kodukula, S., Storch, J., Stark, R.E., 2007. Solution-state molecular structure of apo and oleate-liganded liver fatty acid-binding protein. *Biochemistry* 46, 12543-12556.
- Hemphill, A., Stadelmann, B., Scholl, S., Muller, J., Spiliotis, M., Muller, N., Gottstein, B., Siles-Lucas, M., 2010. Echinococcus metacestodes as laboratory models for the screening of drugs against cestodes and trematodes. *Parasitology* 137, 569-587.
- Hemphill, A., Stettler, M., Walker, M., Siles-Lucas, M., Fink, R., Gottstein, B., 2002. Culture of Echinococcus multilocularis metacestodes: an alternative to animal use. *Trends Parasitol* 18, 445-451.
- Henrichsen, C.N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., Ruedi, M., Kaessmann, H., Reymond, A., 2009. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41, 424-429.
- Hernandez-Ruiz, J., Salaiza-Suazo, N., Carrada, G., Escoto, S., Ruiz-Remigio, A., Rosenstein, Y., Zentella, A., Becker, I., 2010. CD8 cells of patients with diffuse cutaneous leishmaniasis display functional exhaustion: the latter is reversed, in vitro, by TLR2 agonists. *PLoS Negl Trop Dis* 4, e871.
- Hillier, L.W., Reinke, V., Green, P., Hirst, M., Marra, M.A., Waterston, R.H., 2009. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res* 19, 657-666.
- Hu, G., Llinas, M., Li, J., Preiser, P.R., Bozdech, Z., 2007. Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy. *BMC Bioinformatics* 8, 350.
- Huang, J., Hao, P., Chen, H., Hu, W., Yan, Q., Liu, F., Han, Z.G., 2009. Genome-wide identification of Schistosoma japonicum microRNAs using a deep-sequencing approach. *PLoS One* 4, e8206.
- Huang, Q.X., Cheng, X.Y., Mao, Z.C., Wang, Y.S., Zhao, L.L., Yan, X., Ferris, V.R., Xu, R.M., Xie, B.Y., 2010. MicroRNA discovery and analysis of pinewood nematode Bursaphelenchus xylophilus by deep sequencing. *PLoS One* 5, e13271.
- Hui, X., Li, H., Zhou, Z., Lam, K.S., Xiao, Y., Wu, D., Ding, K., Wang, Y., Vanhoutte, P.M., Xu, A., 2010. Adipocyte fatty acid-binding protein modulates inflammatory responses in macrophages through a positive feedback loop involving c-Jun NH2-terminal kinases and activator protein-1. *J Biol Chem* 285, 10273-10280.
- Irimia, M., Roy, S.W., 2008. Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucleic Acids Res* 36, 1703-1712.
- Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A., Kleywegt, G.J., 2003. The crystal structure of Echinococcus granulosus fatty-acid-binding protein 1. *Biochim Biophys Acta* 1649, 40-50.
- Jaubert-Possamai, S., Rispe, C., Tanguy, S., Gordon, K., Walsh, T., Edwards, O., Tagu, D., 2010. Expansion of the miRNA pathway in the hemipteran insect Acyrthosiphon pisum. *Mol Biol Evol* 27, 979-987.
- Jeong, K.Y., Kim, W.K., Lee, J.S., Lee, J., Lee, I.Y., Kim, K.E., Park, J.W., Hong, C.S., Ree, H.I., Yong, T.S., 2005. Immunoglobulin E reactivity of recombinant allergen Tyr p 13 from Tyrophagus putrescentiae homologous to fatty acid binding protein. *Clin Diagn Lab Immunol* 12, 581-585.
- Jia, W.Z., Yan, H.B., Lou, Z.Z., Ni, X.W., Liu, H.X., Li, H.M., Guo, A.J., Fu, B.Q., 2011. Genetic variation of the 8-kDa glycoprotein family from Echinococcus

- granulosus, *Taenia multiceps* and *Taenia hydatigena*. *Chin Med J (Engl)* 124, 2849-2856.
- Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.Y., Zhao, H., Deng, X.W., Nelson, T., 2009. A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat Genet* 41, 258-263.
- Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B., 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497-1502.
- Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., Sarnow, P., 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* 309, 1577-1581.
- Jordanova, R., Groves, M.R., Kostova, E., Woltersdorf, C., Liebau, E., Tucker, P.A., 2009. Fatty acid- and retinoid-binding proteins have distinct binding pockets for the two types of cargo. *J Biol Chem* 284, 35818-35826.
- Judson, R.L., Babiarz, J.E., Venere, M., Billelloch, R., 2009. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol* 27, 459-461.
- Jura, H., Bader, A., Hartmann, M., Maschek, H., Frosch, M., 1996. Hepatic tissue culture model for study of host-parasite interactions in alveolar echinococcosis. *Infect Immun* 64, 3484-3490.
- Kamb, A., 2011. Next-generation sequencing and its potential impact. *Chemical research in toxicology* 24, 1163-1168.
- Kanazawa, T., Asahi, H., Hata, H., Mochida, K., Kagei, N., Stadecker, M.J., 1993. Arginine-dependent generation of reactive nitrogen intermediates is instrumental in the in vitro killing of protoscoleces of *Echinococcus multilocularis* by activated macrophages. *Parasite Immunol* 15, 619-623.
- Kato, M., de Lencastre, A., Pincus, Z., Slack, F.J., 2009. Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during *Caenorhabditis elegans* development. *Genome Biol* 10, R54.
- Kawai, T., Akira, S., 2011. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637-650.
- Kennedy, M.W., 2000a. The nematode polyprotein allergens/antigens. *Parasitol Today* 16, 373-380.
- Kennedy, M.W., 2000b. The polyprotein lipid binding proteins of nematodes. *Biochim Biophys Acta* 1476, 149-164.
- Kennedy, M.W., Brass, A., McCrudden, A.B., Price, N.C., Kelly, S.M., Cooper, A., 1995. The ABA-1 allergen of the parasitic nematode *Ascaris suum*: fatty acid and retinoid binding function and structural characterization. *Biochemistry* 34, 6700-6710.
- Kennedy, M.W., Garside, L.H., Goodrick, L.E., McDermott, L., Brass, A., Price, N.C., Kelly, S.M., Cooper, A., Bradley, J.E., 1997. The Ov20 protein of the parasitic nematode *Onchocerca volvulus*. A structurally novel class of small helix-rich retinol-binding proteins. *J Biol Chem* 272, 29442-29448.
- Kent, W.J., 2002. BLAT--the BLAST-like alignment tool. *Genome Res* 12, 656-664.
- Kim, V.N., Nam, J.W., 2006. Genomics of microRNA. *Trends Genet* 22, 165-173.
- Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elisk, C.G., Gaur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubio, J.M., Ribeiro, J.M., Rozas, J., Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raoult, D., Reed, D.L., Tomoyasu, Y., Krause, E., Mittapalli, O., Margam, V.M., Li, H.M., Meyer, J.M., Johnson, R.M., Romero-Severson, J., Vanzee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguade, M., Guirao-Rico, S., Anzola, J.M., Yoon, K.S., Strycharz,

- J.P., Unger, M.F., Christley, S., Lobo, N.F., Seufferheld, M.J., Wang, N., Dasch, G.A., Struchiner, C.J., Madey, G., Hannick, L.I., Bidwell, S., Joardar, V., Caler, E., Shao, R., Barker, S.C., Cameron, S., Bruggner, R.V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D., Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov, E.M., Pittendrigh, B.R., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* 107, 12168-12173.
- Kleinert, H., Pautz, A., Linker, K., Schwarz, P.M., 2004. Regulation of the expression of inducible nitric oxide synthase. *Eur J Pharmacol* 500, 255-266.
- Knapp, J., Bart, J.M., Giraudoux, P., Glowatzki, M.L., Breyer, I., Raoul, F., Deplazes, P., Duscher, G., Martinek, K., Dubinsky, P., Guislain, M.H., Cliquet, F., Romig, T., Malczewski, A., Gottstein, B., Piarroux, R., 2009. Genetic diversity of the cestode *Echinococcus multilocularis* in red foxes at a continental scale in Europe. *PLoS Negl Trop Dis* 3, e452.
- Knapp, J., Guislain, M.H., Bart, J.M., Raoul, F., Gottstein, B., Giraudoux, P., Piarroux, R., 2008. Genetic diversity of *Echinococcus multilocularis* on a local scale. *Infect Genet Evol* 8, 367-373.
- Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., Turner, D.J., 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* 6, 291-295.
- Kuchenbauer, F., Morin, R.D., Argiropoulos, B., Petriv, O.I., Griffith, M., Heuser, M., Yung, E., Piper, J., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Hansen, C.L., Marra, M.A., Humphries, R.K., 2008. In-depth characterization of the microRNA transcriptome in a leukemia progression model. *Genome Res* 18, 1787-1797.
- Lanford, R.E., Hildebrandt-Eriksen, E.S., Petri, A., Persson, R., Lindow, M., Munk, M.E., Kauppinen, S., Orum, H., Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* 327, 198-201.
- Lanford, R.E., Hildebrandt-Eriksen, E.S., Petri, A., Persson, R., Lindow, M., Munk, M.E., Kauppinen, S., Orum, H., 2010. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* 327, 198-201.
- Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858-862.
- Le, T.H., Blair, D., Agatsuma, T., Humair, P.F., Campbell, N.J., Iwagami, M., Littlewood, D.T., Peacock, B., Johnston, D.A., Bartley, J., Rollinson, D., Herniou, E.A., Zarlenga, D.S., McManus, D.P., 2000. Phylogenies inferred from mitochondrial gene orders—a cautionary tale from the parasitic flatworms. *Mol Biol Evol* 17, 1123-1125.
- Lee, J.S., Lee, J., Park, S.J., Yong, T.S., 2003. Analysis of the genes expressed in *Clonorchis sinensis* adults using the expressed sequence tag approach. *Parasitol Res* 91, 283-289.
- Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843-854.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W., 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117, 69-81.

- Lek, M., MacArthur, D.G., Yang, N., North, K.N., 2010. Phylogenetic analysis of gene structure and alternative splicing in alpha-actinins. *Mol Biol Evol* 27, 773-780.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18, 1851-1858.
- Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., Cheung, V.G., 2011. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333, 53-58.
- Li, S., Mead, E.A., Liang, S., Tu, Z., 2009. Direct sequencing and expression analysis of a large number of miRNAs in *Aedes aegypti* and a multi-species survey of novel mosquito miRNAs. *BMC Genomics* 10, 581.
- Lin, R., Lu, G., Wang, J., Zhang, C., Xie, W., Lu, X., Manton, G., Martin, H., Richert, L., Vuitton, D.A., Wen, H., 2011. Time course of gene expression profiling in the liver of experimental mice infected with *Echinococcus multilocularis*. *PLoS One* 6, e14557.
- Lin, R.Y., Wang, J.H., Lu, X.M., Zhou, X.T., Manton, G., Wen, H., Vuitton, D.A., Richert, L., 2009. Components of the mitogen-activated protein kinase cascade are activated in hepatic cells by *Echinococcus multilocularis* metacystode. *World J Gastroenterol* 15, 2116-2124.
- Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*.
- Liu, F., Cui, S.J., Hu, W., Feng, Z., Wang, Z.Q., Han, Z.G., 2009a. Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. *Mol Cell Proteomics* 8, 1236-1251.
- Liu, R.Z., Li, X., Godbout, R., 2008. A novel fatty acid-binding protein (FABP) gene resulting from tandem gene duplication in mammals: transcription in rat retina and testis. *Genomics* 92, 436-445.
- Liu, R.Z., Mita, R., Beaulieu, M., Gao, Z., Godbout, R., 2010. Fatty acid binding proteins in brain development and disease. *Int J Dev Biol* 54, 1229-1239.
- Liu, Y.H., Wang, X.G., Gao, J.S., Qingyao, Y., Horton, J., 2009b. Continuous albendazole therapy in alveolar echinococcosis: long-term follow-up observation of 20 cases. *Trans R Soc Trop Med Hyg* 103, 768-778.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. *Nature* 435, 834-838.
- Lu, Y.C., Smielewska, M., Palakodeti, D., Lovci, M.T., Aigner, S., Yeo, G.W., Graveley, B.R., 2009. Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*. *RNA* 15, 1483-1491.
- Lucke, C., Huang, S., Rademacher, M., Ruterjans, H., 2002. New insights into intracellular lipid binding proteins: The role of buried water. *Protein Sci* 11, 2382-2392.
- Luo, X., Lin, H., Pan, Z., Xiao, J., Zhang, Y., Lu, Y., Yang, B., Wang, Z., 2008. Down-regulation of miR-1/miR-133 contributes to re-expression of pacemaker channel genes HCN2 and HCN4 in hypertrophic heart. *J Biol Chem* 283, 20045-20052.
- MacDonald, A.S., Maizels, R.M., 2008. Alarming dendritic cells for Th2 induction. *J Exp Med* 205, 13-17.

- Magalhaes, P.O., Lopes, A.M., Mazzola, P.G., Rangel-Yagui, C., Penna, T.C., Pessoa, A., Jr., 2007. Methods of endotoxin removal from biological preparations: a review. *J Pharm Pharm Sci* 10, 388-404.
- Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., Chinnaiyan, A.M., 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458, 97-101.
- Makowski, L., Boord, J.B., Maeda, K., Babaev, V.R., Uysal, K.T., Morgan, M.A., Parker, R.A., Suttles, J., Fazio, S., Hotamisligil, G.S., Linton, M.F., 2001. Lack of macrophage fatty-acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis. *Nat Med* 7, 699-705.
- Makowski, L., Brittingham, K.C., Reynolds, J.M., Suttles, J., Hotamisligil, G.S., 2005. The fatty acid-binding protein, aP2, coordinates macrophage cholesterol trafficking and inflammatory activity. Macrophage expression of aP2 impacts peroxisome proliferator-activated receptor gamma and I $\kappa$ B kinase activities. *J Biol Chem* 280, 12888-12895.
- Malczewski, A., Gawor, J., Malczewska, M., 2008. Infection of red foxes (*Vulpes vulpes*) with *Echinococcus multilocularis* during the years 2001-2004 in Poland. *Parasitol Res* 103, 501-505.
- Mallick, B., Ghosh, Z., Chakrabarti, J., 2008. MicroRNA switches in *Trypanosoma brucei*. *Biochem Biophys Res Commun* 372, 459-463.
- Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet* 24, 133-141.
- Marin, M., Garat, B., Pettersson, U., Ehrlich, R., 1993. Isolation and characterization of a middle repetitive DNA element from *Echinococcus granulosus*. *Mol Biochem Parasitol* 59, 335-338.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., Gilad, Y., 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18, 1509-1517.
- Martin, J.M., Stapleton, R.D., 2010. Omega-3 fatty acids in critical illness. *Nutr Rev* 68, 531-541.
- Maule, A.G., Marks, N.J. 2006. Parasitic flatworms : molecular biology, biochemistry, immunology and physiology, In: CABI, Wallingford.
- McDermott, L., Cooper, A., Kennedy, M.W., 1999. Novel classes of fatty acid and retinol binding protein from nematodes. *Mol Cell Biochem* 192, 69-75.
- McDermott, L., Kennedy, M.W., McManus, D.P., Bradley, J.E., Cooper, A., Storch, J., 2002. How helminth lipid-binding proteins offload their ligands to membranes: differential mechanisms of fatty acid transfer by the ABA-1 polyprotein allergen and Ov-FAR-1 proteins of nematodes and Sj-FABPc of schistosomes. *Biochemistry* 41, 6706-6713.
- McManus, D.P., 2009. Reflections on the biochemistry of *Echinococcus*: past, present and future. *Parasitology* 136, 1643-1652.
- McManus, D.P., Zhang, W., Li, J., Bartley, P.B., 2003. Echinococcosis. *Lancet* 362, 1295-1304.
- Mead, E.A., Tu, Z., 2008. Cloning, characterization, and expression of microRNAs from the Asian malaria mosquito, *Anopheles stephensi*. *BMC Genomics* 9, 244.
- Medeiros, L.A., Dennis, L.M., Gill, M.E., Houbaviv, H., Markoulaki, S., Fu, D., White, A.C., Kirak, O., Sharp, P.A., Page, D.C., Jaenisch, R., 2011. Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proc Natl Acad Sci U S A* 108, 14163-14168.

- Meenan, N.A., Ball, G., Bromek, K., Uhrin, D., Cooper, A., Kennedy, M.W., Smith, B.O., 2011. Solution structure of a repeated unit of the ABA-1 nematode polyprotein allergen of *Ascaris* reveals a novel fold and two discrete lipid-binding sites. *PLoS Negl Trop Dis* 5, e1040.
- Mei, B., Kennedy, M.W., Beauchamp, J., Komuniecki, P.R., Komuniecki, R., 1997. Secretion of a novel, developmentally regulated fatty acid-binding protein into the perivitelline fluid of the parasitic nematode, *Ascaris suum*. *J Biol Chem* 272, 9933-9941.
- Mejri, N., Gottstein, B., 2009. *Echinococcus multilocularis* metacestode metabolites contain a cysteine protease that digests eotaxin, a CC pro-inflammatory chemokine. *Parasitol Res* 105, 1253-1260.
- Militello, K.T., Refour, P., Comeaux, C.A., Duraisingh, M.T., 2008. Antisense RNA and RNAi in protozoan parasites: working hard or hardly working? *Mol Biochem Parasitol* 157, 117-126.
- Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., Horvitz, H.R., 2007. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet* 3, e215.
- Mizukami, C., Spiliotis, M., Gottstein, B., Yagi, K., Katakura, K., Oku, Y., 2010. Gene silencing in *Echinococcus multilocularis* protoscoleces using RNA interference. *Parasitol Int* 59, 647-652.
- Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., Baulcombe, D.C., 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* 447, 1126-1129.
- Monteiro, K.M., de Carvalho, M.O., Zaha, A., Ferreira, H.B., 2010. Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* 10, 1985-1999.
- Montoya, J., Lopez-Perez, M.J., Ruiz-Pesini, E., 2006. Mitochondrial DNA transcription and diseases: past, present and future. *Biochim Biophys Acta* 1757, 1179-1189.
- Moore, J., McDermott, L., Price, N.C., Kelly, S.M., Cooper, A., Kennedy, M.W., 1999. Sequence-divergent units of the ABA-1 polyprotein array of the nematode *Ascaris suum* have similar fatty-acid- and retinol-binding properties but different binding-site environments. *Biochem J* 340 ( Pt 1), 337-343.
- Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., Marra, M.A., 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* 18, 610-621.
- Moro, P., Schantz, P.M., 2009. Echinococcosis: a review. *Int J Infect Dis* 13, 125-133.
- Morozova, O., Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255-264.
- Morphew, R.M., Wright, H.A., LaCourse, E.J., Woods, D.J., Brophy, P.M., 2007. Comparative proteomics of excretory-secretory proteins released by the liver fluke *Fasciola hepatica* in sheep host bile and during in vitro culture ex host. *Mol Cell Proteomics* 6, 963-972.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621-628.
- Moser, D., Tandler, M., Griffiths, G., Klinkert, M.Q., 1991. A 14-kDa *Schistosoma mansoni* polypeptide is homologous to a gene family of fatty acid binding proteins. *J Biol Chem* 266, 8447-8454.

- Nachmani, D., Stern-Ginossar, N., Sarid, R., Mandelboim, O., 2009. Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell Host Microbe* 5, 376-385.
- Nafikov, R.A., Beitz, D.C., 2007. Carbohydrate and lipid metabolism in farm animals. *J Nutr* 137, 702-705.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344-1349.
- Nakao, M., Yokoyama, N., Sako, Y., Fukunaga, M., Ito, A., 2002. The complete mitochondrial DNA sequence of the cestode *Echinococcus multilocularis* (Cyclophyllidae: Taeniidae). *Mitochondrion* 1, 497-509.
- Nakao, R., Kameda, Y., Kouguchi, H., Matsumoto, J., Dang, Z., Simon, A.Y., Torigoe, D., Sasaki, N., Oku, Y., Sugimoto, C., Agui, T., Yagi, K., 2011. Identification of genetic loci affecting the establishment and development of *Echinococcus multilocularis* larvae in mice. *Int J Parasitol* 41, 1121-1128.
- Newmark, P.A., Sanchez Alvarado, A., 2002. Not your father's planarian: a classic model enters the era of functional genomics. *Nat Rev Genet* 3, 210-219.
- Ockner, R.K., Manning, J.A., Poppenhausen, R.B., Ho, W.K., 1972. A binding protein for fatty acids in cytosol of intestinal mucosa, liver, myocardium, and other tissues. *Science* 177, 56-58.
- Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., Lai, E.C., 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130, 89-100.
- Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., Lai, E.C., 2008. The regulatory activity of microRNA\* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* 15, 354-363.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusica, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziński, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y., 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563-573.

- Palakodeti, D., Smielewska, M., Graveley, B.R., 2006. MicroRNAs from the Planarian *Schmidtea mediterranea*: a model system for stem cell biology. *RNA* 12, 1640-1649.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413-1415.
- Pavlin, B.I., Schloegel, L.M., Daszak, P., 2009. Risk of importing zoonotic diseases through wildlife trade, United States. *Emerg Infect Dis* 15, 1721-1726.
- Pelton, P.D., Zhou, L., Demarest, K.T., Burris, T.P., 1999. PPARgamma activation induces the expression of the adipocyte fatty acid binding protein gene in human monocytes. *Biochem Biophys Res Commun* 261, 456-458.
- Perrigoue, J.G., Marshall, F.A., Artis, D., 2008. On the hunt for helminths: innate immune cells in the recognition and response to helminth parasites. *Cell Microbiol* 10, 1757-1764.
- Petkeviciute, R., Ieshko, E.P., 1991. The karyotypes of *Triaenophorus nodulosus* and *T. crassus* (Cestoda: Pseudophyllidea). *Int J Parasitol* 21, 11-15.
- Pettitt, J., Harrison, N., Stansfield, I., Connolly, B., Muller, B., 2010. The evolution of spliced leader trans-splicing in nematodes. *Biochem Soc Trans* 38, 1125-1130.
- Poole, C.B., Davis, P.J., Jin, J., McReynolds, L.A., 2010. Cloning and bioinformatic identification of small RNAs in the filarial nematode, *Brugia malayi*. *Mol Biochem Parasitol* 169, 87-94.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., Jensen, T.H., 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851-1854.
- Prinsen, C.F., Weghuis, D.O., Kessel, A.G., Veerkamp, J.H., 1997. Identification of a human heart FABP pseudogene located on chromosome 13. *Gene* 193, 245-251.
- Puerta, L., Kennedy, M.W., Jimenez, S., Caraballo, L., 1999. Structural and ligand binding analysis of recombinant blot 13 allergen from *Blomia tropicalis* mite, a fatty acid binding protein. *Int Arch Allergy Immunol* 119, 181-184.
- Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J., 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* 5, 1005-1010.
- Raghavachari, N., Xu, X., Munson, P.J., Gladwin, M.T., 2009. Characterization of whole blood gene expression profiles as a sequel to globin mRNA reduction in patients with sickle cell disease. *PLoS One* 4, e6484.
- Raj, A., van Oudenaarden, A., 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-226.
- Rathjen, T., Nicol, C., McConkey, G., Dalmay, T., 2006. Analysis of short RNAs in the malaria parasite and its red blood cell host. *FEBS Lett* 580, 5185-5188.
- Rausch, V.R., Rausch, R.L., 1981. The karyotype of *Echinococcus multilocularis* (Cestoda: Taeniidae). *Can J Genet Cytol* 23, 151-154.
- Reddien, P.W., Sanchez Alvarado, A., 2004. Fundamentals of planarian regeneration. *Annu Rev Cell Dev Biol* 20, 725-757.
- Reichelt, P., Schwarz, C., Donzeau, M., 2006. Single step protocol to purify recombinant proteins with low endotoxin contents. *Protein Expr Purif* 46, 483-488.
- Resch, A.M., Palakodeti, D., Lu, Y.C., Horowitz, M., Graveley, B.R., 2012. Transcriptome Analysis Reveals Strain-Specific and Conserved Stemness Genes in *Schmidtea mediterranea*. *PLoS One* 7, e34447.

- Rhoads, R.E., 2010. miRNA regulation of the translational machinery. Preface. *Prog Mol Subcell Biol* 50, v-vi.
- Richieri, G.V., Ogata, R.T., Zimmerman, A.W., Veerkamp, J.H., Kleinfeld, A.M., 2000. Fatty acid binding proteins from different tissues show distinct patterns of fatty acid interactions. *Biochemistry* 39, 7197-7204.
- Rigano, R., Buttari, B., Profumo, E., Ortona, E., Delunardo, F., Margutti, P., Mattei, V., Teggi, A., Sorice, M., Siracusano, A., 2007. *Echinococcus granulosus* antigen B impairs human dendritic cell differentiation and polarizes immature dendritic cell maturation towards a Th2 cell response. *Infect Immun* 75, 1667-1678.
- Rodriguez-Perez, J., Rodriguez-Medina, J.R., Garcia-Blanco, M.A., Hillyer, G.V., 1992. *Fasciola hepatica*: molecular cloning, nucleotide sequence, and expression of a gene encoding a polypeptide homologous to a *Schistosoma mansoni* fatty acid-binding protein. *Exp Parasitol* 74, 400-407.
- Rogozin, I.B., Wolf, Y.I., Sorokin, A.V., Mirkin, B.G., Koonin, E.V., 2003. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr Biol* 13, 1512-1517.
- Rolph, M.S., Young, T.R., Shum, B.O., Gorgun, C.Z., Schmitz-Peiffer, C., Ramshaw, I.A., Hotamisligil, G.S., Mackay, C.R., 2006. Regulation of dendritic cell function and T cell priming by the fatty acid-binding protein AP2. *J Immunol* 177, 7794-7801.
- Rosenzvit, M.C., Canova, S.G., Kamenetzky, L., Guarnera, E.A., 2001. *Echinococcus granulosus*: intraspecific genetic variation assessed by a DNA repetitive element. *Parasitology* 123, 381-388.
- Rosenzvit, M.C., Canova, S.G., Kamenetzky, L., Ledesma, B.A., Guarnera, E.A., 1997. *Echinococcus granulosus*: cloning and characterization of a tandemly repeated DNA element. *Exp Parasitol* 87, 65-68.
- Ross, L.H., Freedman, J.H., Rubin, C.S., 1995. Structure and expression of novel spliced leader RNA genes in *Caenorhabditis elegans*. *J Biol Chem* 270, 22066-22075.
- Roy, S.W., Gilbert, W., 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet* 7, 211-221.
- Ruby, J.G., Jan, C.H., Bartel, D.P., 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83-86.
- Saarma, U., Jogisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., Simsek, S., Andresiuk, V., Denegri, G., Gonzalez, L.M., Ferrer, E., Garate, T., Rinaldi, L., Maravilla, P., 2009. A novel phylogeny for the genus *Echinococcus*, based on nuclear data, challenges relationships based on mitochondrial evidence. *Parasitology* 136, 317-328.
- Saghir, N., Conde, P.J., Brophy, P.M., Barrett, J., 2001. Biochemical characterisation of a hydrophobic ligand binding protein from the tapeworm *Hymenolepis diminuta*. *Int J Parasitol* 31, 653-660.
- Saini, H.K., Griffiths-Jones, S., Enright, A.J., 2007. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A* 104, 17719-17724.
- Sako, Y., Yamasaki, H., Nakaya, K., Nakao, M., Ito, A., 2007. Cloning and characterization of cathepsin L-like peptidases of *Echinococcus multilocularis* metacestodes. *Mol Biochem Parasitol* 154, 181-189.
- Sanchez Alvarado, A., Kang, H., 2005. Multicellularity, stem cells, and the neoblasts of the planarian *Schmidtea mediterranea*. *Exp Cell Res* 306, 299-308.
- Saraiya, A.A., Wang, C.C., 2008. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog* 4, e1000224.

- Scaplehorn, N., Holmstrom, A., Moreau, V., Frischknecht, F., Reckmann, I., Way, M., 2002. Grb2 and Nck act cooperatively to promote actin-based motility of vaccinia virus. *Curr Biol* 12, 740-745.
- Schaap, F.G., Binas, B., Danneberg, H., van der Vusse, G.J., Glatz, J.F., 1999. Impaired long-chain fatty acid utilization by cardiac myocytes isolated from mice lacking the heart-type fatty acid binding protein gene. *Circ Res* 85, 329-337.
- Schaap, F.G., van der Vusse, G.J., Glatz, J.F., 2002. Evolution of the family of intracellular lipid binding proteins in vertebrates. *Mol Cell Biochem* 239, 69-77.
- Schadt, E.E., Turner, S., Kasarskis, A., 2010. A window into third-generation sequencing. *Human molecular genetics* 19, R227-240.
- Scherer, L.J., Frank, R., Rossi, J.J., 2007. Optimization and characterization of tRNA-shRNA expression constructs. *Nucleic Acids Res* 35, 2620-2628.
- Schleicher, C.H., Cordoba, O.L., Santome, J.A., Dell'Angelica, E.C., 1995. Molecular evolution of the multigene family of intracellular lipid-binding proteins. *Biochem Mol Biol Int* 36, 1117-1125.
- Scholz, M.B., Lo, C.C., Chain, P.S., 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current opinion in biotechnology* 23, 9-15.
- Schweiger, A., Ammann, R.W., Candinas, D., Clavien, P.A., Eckert, J., Gottstein, B., Halkic, N., Muellhaupt, B., Prinz, B.M., Reichen, J., Tarr, P.E., Torgerson, P.R., Deplazes, P., 2007. Human alveolar echinococcosis after fox population increase, Switzerland. *Emerg Infect Dis* 13, 878-882.
- Seitz, H., 2009. Redefining microRNA targets. *Curr Biol* 19, 870-873.
- Selbach, M., Schwanhauser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58-63.
- Sevignani, C., Calin, G.A., Nnadi, S.C., Shimizu, M., Davuluri, R.V., Hyslop, T., Demant, P., Croce, C.M., Siracusa, L.D., 2007. MicroRNA genes are frequently located near mouse cancer susceptibility loci. *Proc Natl Acad Sci U S A* 104, 8017-8022.
- Shepherd, J.C., Aitken, A., McManus, D.P., 1991. A protein secreted in vivo by *Echinococcus granulosus* inhibits elastase activity and neutrophil chemotaxis. *Mol Biochem Parasitol* 44, 81-90.
- Shioda, N., Yamamoto, Y., Watanabe, M., Binas, B., Owada, Y., Fukunaga, K., 2010. Heart-type fatty acid binding protein regulates dopamine D2 receptor function in mouse brain. *J Neurosci* 30, 3146-3155.
- Shum, B.O., Mackay, C.R., Gorgun, C.Z., Frost, M.J., Kumar, R.K., Hotamisligil, G.S., Rolph, M.S., 2006. The adipocyte fatty acid-binding protein aP2 is required in allergic airway inflammation. *J Clin Invest* 116, 2183-2192.
- Siracusano, A., Margutti, P., Delunardo, F., Profumo, E., Rigano, R., Buttari, B., Teggi, A., Ortona, E., 2008. Molecular cross-talk in host-parasite relationships: the intriguing immunomodulatory role of *Echinococcus* antigen B in cystic echinococcosis. *Int J Parasitol* 38, 1371-1376.
- Skalsky, R.L., Vanlandingham, D.L., Scholle, F., Higgs, S., Cullen, B.R., 2010. Identification of microRNAs expressed in two mosquito vectors, *Aedes albopictus* and *Culex quinquefasciatus*. *BMC Genomics* 11, 119.
- Smyth, J.D., McManus, D.P., 1989. *The physiology and biochemistry of cestodes*. Cambridge University Press.
- Spence, H.J., Moore, J., Brass, A., Kennedy, M.W., 1993. A cDNA encoding repeating units of the ABA-1 allergen of *Ascaris*. *Mol Biochem Parasitol* 57, 339-343.

- Spiliotis, M., Brehm, K., 2009. Axenic in vitro cultivation of *Echinococcus multilocularis* metacestode vesicles and the generation of primary cell cultures. *Methods Mol Biol* 470, 245-262.
- Spiliotis, M., Konrad, C., Gelmedin, V., Tappe, D., Bruckner, S., Mosch, H.U., Brehm, K., 2006. Characterisation of EmMPK1, an ERK-like MAP kinase from *Echinococcus multilocularis* which is activated in response to human epidermal growth factor. *Int J Parasitol* 36, 1097-1112.
- Spiliotis, M., Lechner, S., Tappe, D., Scheller, C., Krohne, G., Brehm, K., 2008. Transient transfection of *Echinococcus multilocularis* primary cells and complete in vitro regeneration of metacestode vesicles. *Int J Parasitol* 38, 1025-1039.
- Spiliotis, M., Mizukami, C., Oku, Y., Kiss, F., Brehm, K., Gottstein, B., 2010. *Echinococcus multilocularis* primary cells: improved isolation, small-scale cultivation and RNA interference. *Mol Biochem Parasitol* 174, 83-87.
- Spiliotis, M., Tappe, D., Sesterhenn, L., Brehm, K., 2004. Long-term in vitro cultivation of *Echinococcus multilocularis* metacestodes under axenic conditions. *Parasitol Res* 92, 430-432.
- Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L., Rokhsar, D.S., 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454, 955-960.
- Staebler, S., Steinmetz, H., Keller, S., Deplazes, P., 2007. First description of natural *Echinococcus multilocularis* infections in chinchilla (*Chinchilla laniger*) and Prevost's squirrel (*Callosciurus prevostii borneoensis*). *Parasitol Res* 101, 1725-1727.
- Steers, N.J., Rogan, M.T., Heath, S., 2001. In-vitro susceptibility of hydatid cysts of *Echinococcus granulosus* to nitric oxide and the effect of the laminated layer on nitric oxide production. *Parasite Immunol* 23, 411-417.
- Stempin, C.C., Dulgerian, L.R., Garrido, V.V., Cerban, F.M., 2010. Arginase in parasitic infections: macrophage activation, immunosuppression, and intracellular signals. *J Biomed Biotechnol* 2010, 683485.
- Storch, J., Corsico, B., 2008. The emerging functions and mechanisms of mammalian fatty acid-binding proteins. *Annu Rev Nutr* 28, 73-95.
- Storch, J., McDermott, L., 2009. Structural and functional analysis of fatty acid-binding proteins. *J Lipid Res* 50 Suppl, S126-131.
- Storch, J., Thumser, A.E., 2010. Tissue-specific functions in the fatty acid-binding protein family. *J Biol Chem* 285, 32679-32683.
- Sugawara, Y., Azuma, N., Onodera, S., Tsunoka, Y., Morimoto, M., 2011. Th2 immune responses and alternatively activated macrophages (AAMacs) in helminth infection in aged mice. *J Vet Med Sci* 73, 511-516.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., Yaspo, M.L., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956-960.
- t Hoen, P.A., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H., de Menezes, R.X., Boer, J.M., van Ommen, G.J., den Dunnen, J.T., 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36, e141.

- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24, 1596-1599.
- Tappe, D., Brehm, K., Frosch, M., Blankenburg, A., Schrod, A., Kaup, F.J., Matz-Rensing, K., 2007. *Echinococcus multilocularis* infection of several Old World monkey species in a breeding enclosure. *Am J Trop Med Hyg* 77, 504-506.
- Tappe, D., Kern, P., Frosch, M., 2010a. A hundred years of controversy about the taxonomic status of *Echinococcus* species. *Acta Trop* 115, 167-174.
- Tappe, D., Zidowitz, S., Demmer, P., Kern, P., Barth, T.F., Frosch, M., 2010b. Three-dimensional reconstruction of *Echinococcus multilocularis* larval growth in human hepatic tissue reveals complex growth patterns. *Am J Trop Med Hyg* 82, 126-127.
- Tarpey, M.M., Wink, D.A., Grisham, M.B., 2004. Methods for detection of reactive metabolites of oxygen and nitrogen: in vitro and in vivo considerations. *Am J Physiol Regul Integr Comp Physiol* 286, R431-444.
- Tchkonia, T., Morbeck, D.E., Von Zglinicki, T., Van Deursen, J., Lustgarten, J., Scrable, H., Khosla, S., Jensen, M.D., Kirkland, J.L., 2010. Fat tissue, aging, and cellular senescence. *Aging Cell* 9, 667-684.
- Thatcher, E.J., Patton, J.G., 2010. Small RNAs have a big impact on regeneration. *RNA Biol* 7, 333-338.
- Thompson, R.C., 2008. The taxonomy, phylogeny and transmission of *Echinococcus*. *Exp Parasitol* 119, 439-446.
- Timanova-Atanasova, A., Jordanova, R., Radoslavov, G., Deevska, G., Bankov, I., Barrett, J., 2004. A native 13-kDa fatty acid binding protein from the liver fluke *Fasciola hepatica*. *Biochim Biophys Acta* 1674, 200-204.
- Timanova, A., Muller, S., Marti, T., Bankov, I., Walter, R.D., 1999. *Ascaridia galli* fatty acid-binding protein, a member of the nematode polyprotein allergens family. *Eur J Biochem* 261, 569-576.
- Treuner, M., Kozak, C.A., Gallahan, D., Grosse, R., Muller, T., 1994. Cloning and characterization of the mouse gene encoding mammary-derived growth inhibitor/heart-fatty acid-binding protein. *Gene* 147, 237-242.
- Triboulet, R., Mari, B., Lin, Y.L., Chable-Bessia, C., Bennasser, Y., Lebrigand, K., Cardinaud, B., Maurin, T., Barbry, P., Baillat, V., Reynes, J., Corbeau, P., Jeang, K.T., Benkirane, M., 2007. Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science* 315, 1579-1582.
- Tso, A.W., Lam, T.K., Xu, A., Yiu, K.H., Tse, H.F., Li, L.S., Law, L.S., Cheung, B.M., Cheung, R.T., Lam, K.S., 2010. Serum adipocyte fatty acid-binding protein associated with ischemic stroke and early death. *Neurology* 76, 1968-1975.
- Tukel, C., Wilson, R.P., Nishimori, J.H., Pezeshki, M., Chromy, B.A., Baumler, A.J., 2009. Responses to amyloids of microbial and host origin are mediated through toll-like receptor 2. *Cell Host Microbe* 6, 45-53.
- Tumurkhuu, G., Koide, N., Dagvadorj, J., Noman, A.S., Khuda, I., Naiki, Y., Komatsu, T., Yoshida, T., Yokochi, T., 2010. B1 cells produce nitric oxide in response to a series of toll-like receptor ligands. *Cell Immunol* 261, 122-127.
- van der Kleij, D., Latz, E., Brouwers, J.F., Kruize, Y.C., Schmitz, M., Kurt-Jones, E.A., Espevik, T., de Jong, E.C., Kapsenberg, M.L., Golenbock, D.T., Tielens, A.G., Yazdanbakhsh, M., 2002. A novel host-parasite lipid cross-talk. Schistosomal lyso-phosphatidylserine activates toll-like receptor 2 and affects immune polarization. *J Biol Chem* 277, 48122-48129.
- van Liempt, E., van Vliet, S.J., Engering, A., Garcia Vallejo, J.J., Bank, C.M., Sanchez-Hernandez, M., van Kooyk, Y., van Die, I., 2007. *Schistosoma mansoni* soluble egg antigens are internalized by human dendritic cells through multiple C-

- type lectins and suppress TLR-induced dendritic cell activation. *Mol Immunol* 44, 2605-2615.
- Van Veldhoven, P.P., 2010. Biochemistry and genetics of inherited disorders of peroxisomal fatty acid metabolism. *J Lipid Res* 51, 2863-2895.
- Velculescu, V.E., Madden, S.L., Zhang, L., Lash, A.E., Yu, J., Rago, C., Lal, A., Wang, C.J., Beaudry, G.A., Ciriello, K.M., Cook, B.P., Dufault, M.R., Ferguson, A.T., Gao, Y., He, T.C., Hermeking, H., Hiraldo, S.K., Hwang, P.M., Lopez, M.A., Luderer, H.F., Mathews, B., Petroziello, J.M., Polyak, K., Zawel, L., Kinzler, K.W., et al., 1999. Analysis of human transcriptomes. *Nat Genet* 23, 387-388.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B., Kinzler, K.W., 1997. Characterization of the yeast transcriptome. *Cell* 88, 243-251.
- Venugopal, P.G., Nutman, T.B., Semnani, R.T., 2009. Activation and regulation of toll-like receptors (TLRs) by helminth parasites. *Immunol Res* 43, 252-263.
- Vergnes, L., Chin, R., Young, S.G., Reue, K., 2011. Heart-type fatty acid-binding protein is essential for efficient brown adipose tissue fatty acid oxidation and cold tolerance. *J Biol Chem* 286, 380-390.
- 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.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854-858.
- Vuitton, D.A., 2003. The ambiguous role of immunity in echinococcosis: protection of the host or of the parasite? *Acta Trop* 85, 119-132.
- Vuitton, D.A., Gottstein, B., 2010. *Echinococcus multilocularis* and its intermediate host: a model of parasite-host interplay. *J Biomed Biotechnol* 2010, 923193.
- Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L., Brelloch, R., 2008. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* 40, 1478-1483.
- Wang, Y., Cheng, Z., Lu, X., Tang, C., 2009a. *Echinococcus multilocularis*: Proteomic analysis of the protoscoleces by two-dimensional electrophoresis and mass spectrometry. *Exp Parasitol* 123, 162-167.
- Wang, Z., Gerstein, M., Snyder, M., 2009b. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57-63.
- Wang, Z., Xue, X., Sun, J., Luo, R., Xu, X., Jiang, Y., Zhang, Q., Pan, W., 2010. An "in-depth" description of the small non-coding RNA population of *Schistosoma japonicum* schistosomulum. *PLoS Negl Trop Dis* 4, e596.
- Waterkeyn, J.G., Lightowers, M.W., Coppel, R., Cowman, A.F., 1995. Characterization of the gene family encoding a host-protective antigen of the tapeworm *Taenia ovis*. *Mol Biochem Parasitol* 73, 123-131.
- Weischenfeldt, J., Porse, B., 2008. Bone Marrow-Derived Macrophages (BMM): Isolation and Applications. *CSH Protoc* 2008, pdb prot5080.
- Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., Peterson, K.J., 2009. The deep evolution of metazoan microRNAs. *Evol Dev* 11, 50-68.
- Wilusz, J.E., Freier, S.M., Spector, D.L., 2008. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135, 919-932.

- Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J., Tewari, M., 2011. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res* 21, 1450-1461.
- Xia, Y., Spence, H.J., Moore, J., Heaney, N., McDermott, L., Cooper, A., Watson, D.G., Mei, B., Komuniecki, R., Kennedy, M.W., 2000. The ABA-1 allergen of *Ascaris lumbricoides*: sequence polymorphism, stage and tissue-specific expression, lipid binding function, and protein biophysical properties. *Parasitology* 120 ( Pt 2), 211-224.
- Xiao, C., Rajewsky, K., 2009. MicroRNA control in the immune system: basic principles. *Cell* 136, 26-36.
- Xiao, N., Qiu, J., Nakao, M., Li, T., Yang, W., Chen, X., Schantz, P.M., Craig, P.S., Ito, A., 2005. *Echinococcus shiquicus* n. sp., a taeniid cestode from Tibetan fox and plateau pika in China. *Int J Parasitol* 35, 693-701.
- Xu, M.J., Liu, Q., Nisbet, A.J., Cai, X.Q., Yan, C., Lin, R.Q., Yuan, Z.G., Song, H.Q., He, X.H., Zhu, X.Q., 2010. Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance. *BMC Genomics* 11, 521.
- Xue, X., Sun, J., Zhang, Q., Wang, Z., Huang, Y., Pan, W., 2008. Identification and characterization of novel microRNAs from *Schistosoma japonicum*. *PLoS One* 3, e4034.
- Yang, D., Chen, Q., Su, S.B., Zhang, P., Kurosaka, K., Caspi, R.R., Michalek, S.M., Rosenberg, H.F., Zhang, N., Oppenheim, J.J., 2008. Eosinophil-derived neurotoxin acts as an alarmin to activate the TLR2-MyD88 signal pathway in dendritic cells and enhances Th2 immune responses. *J Exp Med* 205, 79-90.
- Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., Nusbaum, C., Thompson, D.A., Friedman, N., Regev, A., 2009. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc Natl Acad Sci U S A* 106, 3264-3269.
- Yoza, B.K., Bogenhagen, D.F., 1984. Identification and in vitro capping of a primary transcript of human mitochondrial DNA. *J Biol Chem* 259, 3909-3915.
- Zeghir-Bouteldja, R., Amri, M., Aitaissa, S., Bouaziz, S., Mezioug, D., Touil-Boukoffa, C., 2009. In Vitro Study of Nitric Oxide Metabolites Effects on Human Hydatid of *Echinococcus granulosus*. *J Parasitol Res* 2009.
- 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, Y.J., Zhang, S., Liu, X.Z., Wen, H.A., Wang, M., 2010. A simple method of genomic DNA extraction suitable for analysis of bulk fungal strains. *Lett Appl Microbiol* 51, 114-118.
- Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. *Journal of computational biology : a journal of computational molecular cell biology* 7, 203-214.
- Zhao, J., Grant, S.F., 2011. Advances in whole genome sequencing technology. *Current pharmaceutical biotechnology* 12, 293-305.
- Zhao, Q., Caballero, O.L., Levy, S., Stevenson, B.J., Iseli, C., de Souza, S.J., Galante, P.A., Busam, D., Leversha, M.A., Chadalavada, K., Rogers, Y.H., Venter, J.C., Simpson, A.J., Strausberg, R.L., 2009. Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. *Proc Natl Acad Sci U S A* 106, 1886-1891.

- Zheng, Y., Cai, X., Luo, X., Zhang, D., Jing, Z., 2008. Genetic variability of the 45W gene family between Chinese and Mexican *Taenia solium*. *Am J Trop Med Hyg* 78, 946-948.
- Zhou, Y., Zheng, H, Yangyi, Chen, L.Z., Kai Wang, Jing Guo, Zhen Huang, Bo Zhang, Wei Huang, Ke Jin,, Tonghai Dou, M.H., Li Wang, Yuan Zhang, Jie Zhou, Lin Tao,, Zhiwei Cao, Y.L., Tomas Vinar, Brona Brejova, Dan Brown, Ming Li, David J., Miller, D.B., Yang Zhong , Zhu Chen; Feng Liu, Wei Hu, Zhi-Qin Wang,, Qin-Hua Zhang, H.-D.S., Saijuan Chen, Xuenian Xu, Bin Xu, Chuan Ju,, Yucheng Huang, P.J.B., Donald P. McManus, Zheng Feng , Ze-Guang Han; Gang Lu, Shuangxi Ren, Yuezhu Wang, Wenyi Gu, Hui Kang, Jie Chen, Xiaoyun, Chen, S.C., Lijun Wang, Jie Yan , Biyun Wang, Xinyan Lv, Lei Jin, Bofei, Wang, S.P., Xianglin Zhang, Wei Zhang, Qiuping Hu, Genfeng Zhu, Jun, Wang, J.Y., Jian Wang, Huanming Yang, Zemin Ning, Matthew Beriman,, Chia-Lin Wei, Y.R., Guoping Zhao, Shengyue, Wang ; Feng Liu, Y.Z., Zhi-Qin Wang,, Gang Lu, H.Z., Paul J. Brindley, Donald P. McManus, David Blair,, Qin-hua Zhang, Y.Z., Shengyue Wang, Ze-Guang Han, Zhu Chen,, 2009. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 460, 345-351.

## Appendix I

### *E. multilocularis* miRNAs

#### Agg(Cell aggregates, total: 51)

miRNA <sup>a</sup>	Sequence (5'-3')	Length	Position	Location <sup>b</sup>	Expression
emu-nov-1	TATTGCACGTTCTTTTCGCCATC	22	3'	contig_3457: 319713-34/+	Agg, Psa, Met
emu-miR-281	TGTCATGGAGTTGCTCTCTA	20	3'	contig_4166: 257035-54/+	Agg, Psa, Met
emu-miR-61	TGACTAGAAAAGAGCACTCACATC	23	3'	contig_4626: 394679-701/+	Agg, Psa, Met
emu-miR-9	TCTTTGGTTATCTAGCTGTGTGA	23	5'	contig_3425: 19838-60/-	Agg, Psa, Met
emu-nov-5	TTGTGCGTCGTTTTAGTGACCGA	23	3'	contig_5447: 14332-54/-	Agg, Psa, Met
emu-miR-307	TCACAACCTACTTGATTGAGG	21	3'	contig_2724: 4638-58/+	Agg, Psa, Met
emu-nov-7	GGGACGGAAGTCTGAAAGGTTT	22	3'	contig_3426: 469414-35/+	Agg, Psa, Met
emu-miR-745	TGCTGCCTGGTAAGAGCTGTGA	22	3'	contig_3450: 430482-503/-	Agg, Psa, Met
emu-miR-219	TGATTGTCCATTTCGCATTTCTTG	23	5'	contig_2850: 21638-60/-	Agg, Psa, Met
emu-miR-71	TGAAAGACGATGGTAGTGAGA	21	5'	contig_2878: 28973-93/-	Agg, Psa, Met
emu-miR-4989	AAAATGCACCAACTATCTGAGA	22	3'	contig_4588: 50012-33/+	Agg, Psa, Met
emu-miR-277	TAAATGCATTTTCTGGCCCCGTA	22	3'	contig_4588: 50145-66/+ contig_5573: 81662-83/- ; contig_5490: 82925-46/-	Agg, Psa, Met
emu-miR-2a	AATCACAGCCCTGCTTGGAACC	22	3'		Agg, Psa, Met
emu-miR-1	TGGAATGTTGTGAAGTATGT	20	3'	contig_6122: 187447-66/-	Agg, Psa, Met
emu-miR-2162	TATTATGCAACTTTTCACTCC	21	3'	contig_3087: 30249-69/+	Agg, Psa, Met
emu-nov-11	TAAATGCAAAATATCTGGTTATG	23	3'	contig_5582: 85003-25/-	Agg, Psa, Met
emu-miR-36	TCACCGGGTAGACATTCCTTGC	22	3'	contig_1750: 37182-203/-	Agg, Psa, Met
emu-let-7	TGAGGTAGTGTTCGAATGTCT	22	5'	contig_5162: 8005-26/-	Agg, Psa, Met

emu-nov-18	GTAGTCTTCCGAGCAGTATATGG	23	5'	contig_1840: 64696-718/+	Agg, Psa, Met
emu-miR-125	TCCCTGAGACCCTAGAGTTGTC	22	5'	contig_5445: 196794-815/+	Agg, Psa, Met
emu-miR-10b	CACCCTGTAGACCCGAGTTTGA	22	5'	contig_5356: 7529-50/+	Agg, Psa, Met
emu-bantam	TGAGATCGCGATTACAGCTGAT	22	3'	contig_1932: 7345-66/+	Agg, Psa, Met
emu-miR-31	TGGCAAGATACTGGCGAAGCTGA	23	5'	contig_1973: 28921-43/-	Agg, Psa, Met
emu-miR-190	AGATATGTTTGGGTTACTTGG	21	5'	contig_5464: 166568-88/+	Agg, Psa, Met
emu-miR-133	TTGGTCCCCATTAACCAGCCGC	22	3'	contig_6122: 175720-41/-	Agg, Psa
emu-miR-7	TGGAAGACTGGTGATATGTTGT	22	5'	contig_5582: 283014-35/+	Agg, Psa
emu-miR-124	GTATTCTACGCGATGTCTTGG	21	5'	contig_1718: 116067-87/-	Agg, Psa
emu-nov-22	TGGCGCTTTCTAACTTTACTGA	22	5'	contig_5380: 25635-56/-	Agg, Psa
emu-nov-8	GATTGCACTACCCATCGCCCACA	23	3'	contig_5467: 25310-32/+	Agg, Psa
emu-miR-2b	TCACAGCCAATATTGATGAACG	22	3'	contig_2878: 28838-59/-	Agg, Met
emu-miR-2c	TATCACAGCCCTGCTTGGGACACA	24	3'	contig_2878: 28733-56/-	Agg, Met
emu-nov-4	AGGTGACTCTAAAACCTTTTCC	21	3'	contig_2743: 161973-93/-	Agg, Met
emu-nov-16	TGGTGGTGGTGGTGGGGGT	19	5'	contig_5557: 252258-76/-	Agg, Met
<b>emu-nov-10</b>	CCCTCTTCTTCGTCCACTAAGA	22	5'	contig_5574: 41922-43/+	Agg
<b>emu-nov-6</b>	CCTTCTCCCTCGTCGCTCCAAGCCG	25	3'	contig_2780: 12078-102/+	Agg
<b>emu-nov-3</b>	TCCTGGGACTTATACCGGGGCTGT	24	3'	contig_5358: 24125-48/-	Agg
<b>emu-miR-10c</b>	TAATTCGAGTCAACAGGGTCGTT	23	3'	contig_1659: 99447-69/+	Agg
<b>emu-nov-14</b>	CCCACCTCAGCATGGTCACTCTTCC	25	3'	contig_4550: 104334-58/+	Agg
<b>emu-nov-21</b>	TGGCGCTTGATTTCAACACTGT	22	5'	contig_5380: 23936-57/+	Agg
<b>emu-miR-2284</b>	TGGCCAAAAGTTCATTCGG	20	3'	contig_4623: 214357-76/+	Agg
<b>emu-miR-341</b>	TGTCGGTCGGTCGGTCGGTCG	21	5'	contig_2208: 65-85/+	Agg
<b>emu-nov-23</b>	AGTGTGATGTCAGGTTGCTTCT	23	5'	contig_5380: 24161-83/+	Agg
emu-miR-87	GTGAGCAAAGTTTCAGGTGTGC	22		contig_3363: 135483-504/-	Agg, Psa, Met
emu-miR-2478	TATCCCACCTTCTGACACCA	19			Agg, Psa
emu-miR-8	TAATACTGTTAGGTAAAGATGC	22			Agg, Met
<b>emu-miR-184</b>	TGGACGGAGAACTGATAAGGGC	22			Agg

<b>emu-miR-1134</b>	TTCTTCTTCTTGTTCCTTGTGTTTC	24			Agg
<b>emu-miR-315</b>	TTTTGATTGTTGCTCAGAGAGT	22			Agg
<b>emu-miR-1260</b>	AATCCCACCGCTGCCACCA	19			Agg
<b>emu-miR-981</b>	TTCGTTGTCGTCAAAACCTGTAT	23			Agg
<b>emu-miR-279</b>	TGACTAGATCTCACACTCATCC	22			Agg

### Psa (Activated protoscolec, total 10)

<b>emu-nov-24</b>	TCGATGCCTGTGACGCATC	20	3'	contig_4454: 26282-301/-	Psa
<b>emu-nov-25</b>	TCTCGATCCCGGCACTACGATGC	23	3'	contig_2898: 31190-212/+	Psa
<b>emu-nov-26</b>	TGGCGCTTAATGTCATCACGG	21	5'	contig_5380: 25376-96/-	Psa
<b>emu-miR-153</b>	TTGCATAGTCTCATAAGTGCCA	22	3'	contig_4610: 270031-52/-	Psa
<b>emu-miR-1992</b>	TCAGCAGTTGTACCATTGA	19	3'	contig_5992: 43175-93/-	Psa
emu-let-7b	GGAGGTAGTTTCGTTGTGTGGT	21			Psa, Met
emu-miR-125b	TCCCTGAGACTGATAATTGCTC	22			Psa, Met
emu-miR-125a	TCCCTGAGACCCTTTGATTGCC	22			Psa, Met
emu-miR-159	TTTGGATTGAAGGGAGCTCTA	21			Psa, Met
<b>emu-miR-455</b>	CAGTCCACGGGCATATACACTT	22			Psa

### Met (Mature metacestode vesicles, total 4)

<b>emu-nov-27</b>	TGCCCATCTATCTATCTGTCCGC	23	5'	contig_6083: 3006-28/-; contig_1790: 7863-85/-	Met
<b>emu-miR-7</b>	TGGAAGACTAGTGATTTTGTGTT	24			Met
<b>emu-miR-15</b>	TAGCAGCACATAATGGTTTGTGGA	24			Met
<b>emu-miR-192</b>	TGACCTATGAATTGACAGCCAG	22			Met

### Star miRNAs predicted by miRDeep

---

**Agg (total 14)**

emu-nov-1*	CGGTGAAAGTTTATGCATTTACA	23	Agg, Met
emu-miR-9*	ACAAGGCTAGATTTCCAAACAAA	23	Agg, Psa, Met
emu-nov-7*	ACCTATCACACTTCAGTCCAGT	22	Agg, Psa, Met
emu-miR-4989*	TGGGTAGTCGTTGCATTTCCA	21	Agg, Psa, Met
emu-miR-2162*	AGTGGATTTGTTGCATATTATA	22	Agg, Psa, Met
emu-nov-13*	CAGTGACCAAACATATTCAC	20	Agg, Psa, Met
emu-let-7*	ACATCCGTTTCACTATCTGCATA	23	Agg, Psa, Met
emu-nov-18*	TAATACTGTTCCGGTTAGGACGCCA	24	Agg, Psa, Met
emu-miR-124*	AAGGCACGCGGTGAATACCA	20	Agg, Psa, Met
emu-miR-2b*	TCGTCAACATTGCCTGTAGACA	22	Agg, Met
emu-miR-2c*	TGTCCCAAGGGTGGTGATCTACT	23	Agg, Met
emu-miR-10b*	AAGCTCGTGCTTCAAGGAATCA	23	Agg, Met
emu-miR-31*	AGCTTCGTCTGGTCTTGCTGCA	22	Agg, Met
<b>emu-miR-10c*</b>	GGACTCGTTTACTCGAATTGGT	22	Agg

**Psa (total 10)**

emu-nov-24*	TGCTTCGACAGCTAAGATC	19	Psa
emu-nov-25*	AACGTGGTCTTGGGTCCGGTGGT	22	Psa
emu-miR-61*	CGTGAGGCCCTTTCTTGTGCATG	23	Psa
emu-miR-307*	CCTCATCCCGTGGGTTGAGCGATG	24	Psa
emu-nov-8*	TGGCGGTGCGCGGTGCAATTTCTG	24	Psa
emu-miR-277*	GTGGGTAGAAAGTGCCTTTTACA	23	Psa
emu-miR-1*	GTGCTTCCCAATAGGCCATA	20	Psa
emu-miR-133*	GGCTGATTTGTGGGGCTCAGAA	22	Psa

---

<b>emu-miR-7*</b>	AGCGTATCGCAGTTTTTCCCTA	22	Psa
<b>emu-miR-125*</b>	CAACTCTAATGTCCCGGGTTA	21	Psa
<b>Met (total 3)</b>			
<b>emu-nov-27*</b>	GGATGGGTGGGTGGGTGCGTGCGTG	25	Met
<b>emu-miR-745*</b>	CCAGTTCTTATCGGGTATATCATG	24	Met
<b>emu-miR-2a*</b>	GTCCAAGGGTGGTGATCTACT	22	Met

<sup>a</sup>Stage-specific miRNAs are highlighted in bold and different colors;

<sup>b</sup>miRNAs residing more than one loci are colored;

<sup>c</sup>the majority of pre-miRNA sequences were predicted using miRDeep but some highlighted in yellow not; ND' for not determined; 'OM' for omitted.

## Appendix II

### Potential *E. granulosus* miRNAs<sup>a</sup>

miRNA	Sequence (5'-3')	Length	Position	Location
egr-miR-133	TTGGTCCCCATTAACCAGCCGC	22	5'	Egu.pathogen_EMU_contig_32174: 1119795-1119816/-
egr-miR-2162	TATTATGCAACTTTTCACTCC	21	5'	NODE_321480_length_348629_cov_11.370583:13872-13892/-
egr-nov-8	GATTGCACTACCCATCGCCCACA	23	5'	Egu.pathogen_EMU_contig_31533: 2054200-2054222/-
egr-miR-36	TCACCGGGTAGACATTCCTTGC	22	5'	Egu.pathogen_EMU_contig_27961: 34839-34860/-
egr-miR-10b	CACCCTGTAGACCCGAGTTTGA	22	3'	Egu.pathogen_EMU_contig_32606:4289790-4289811/-
egr-nov-20	TGGCAAGATACTGGCGAAGCTGA	23	3'	Egu.pathogen_EMU_contig_32798: 170194-170216/-
egr-bantam	TGAGATCGCGATTACAGCTGAT	22	3'	Egu.pathogen_EMU_contig_30682: 9021-9042/+
egr-miR-307	TCACAACCTACTTGATTGAGG	21	5'	Egu.pathogen_EMU_contig_30969: 267950-267970/-
egr-miR-153	TTGCATAGTCTCATAAGTGCCA	22	3'	Egu.pathogen_EMU_contig_26745: 1261050-1261071/+
egr-miR-281	TGTCATGGAGTTGCTCTCTA	20	5'	Egu.pathogen_EMU_contig_32606: 8147735-8147754/-
egr-nov-26	TGGCGCTTAATGTCATCACGG	21	5'	NODE_200032_length_99975_cov_11.312288: 50981-51001/+
egr-miR-745	TGCTGCCTGGTAAGAGCTGTGA	22	3'	Egu.pathogen_EMU_contig_32606: 3523056-3523077/+
egr-nov-5	TTGTGCGTCGTTTCAGTGACCGA	23	3'	Egu.pathogen_EMU_contig_30969: 1337103-1337125/+
egr-nov-16	TGGTGGTGGTGGTGGGGAT	19	5'	Egu.pathogen_EMU_contig_32606: 2316451-2316469/+
egr-nov-10	CCCTCTTCTTCGTCCATTAAGA	22	5'	Egu.pathogen_EMU_contig_30717: 771090-771111/+
egr-nov-6	CCTTCTCCCTCGTCGCTCCAAGCCG	25	5'	Egu.pathogen_EMU_contig_32489: 1496771-1496795/-
egr-nov-25	TCTCGATCCCGGCACTACGATGC	23	5'	NODE_77180_length_78635_cov_11.537534: 59015-59037/-
egr-miR-2284	TGGCCAAAAGTTCATCCGG	20	3'	NODE_115985_length_127129_cov_11.231324: 17759-17778/+
egr-nov-15	GTATTCTACGCGATGTCTTGG	21	5'	Egu.pathogen_EMU_contig_32174: 1841428-1841448/+
egr-nov-18	GTAGTCTTCCGAGCAGTATGTGG	23	3'	Egu.pathogen_EMU_contig_32840: 229190-229212/-

<sup>a</sup> *E. granulosus* miRNAs were computationally identified using Blastn against its genome (Sanger Institute) with *E. multilocularis* miRNAs and the miRNAs

that were not reported previously were listed here.

<sup>b</sup>The secondary structures of putative pre-miRNAs were predicted using RNAfold.