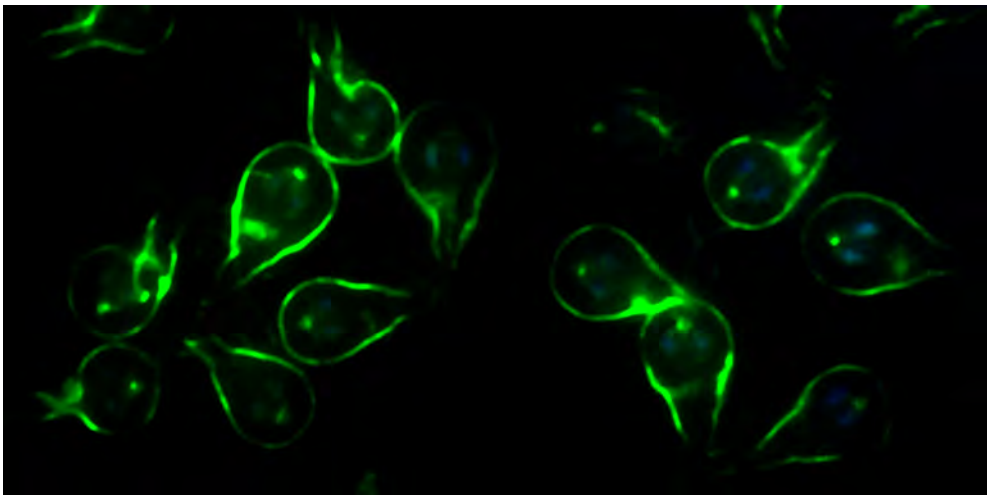
UPPSALA
UNIVERSITET

Characterization of putative diagnostic proteins from *Giardia intestinalis* and *Spironucleus salmonicida*



Adeel ur Rehman

Degree project in biology, Master of science (2 years), 2012
Examensarbete i biologi 30 hp till masterexamen, 2012
Biology Education Centre and Department of Cell and Molecular Biology, Uppsala University
Supervisor: Staffan G Svärd
External opponent: Mattias Andersson

Contents

1. INTRODUCTION

1.1. *Spironucleus salmonicida*

The fish infectious parasite *Spironucleus salmonicida*, is a diplokaryotic, mitochondrion-lacking flagellate belonging to the order diplomonodida. *S. salmonicida* was formerly known as *Spironucleus barkhanus.*[1-3] *S. salmonicida* may cause systemic spironucleosis in farmed and ornamental salmonids. The parasite causes bulk mortality, enormous economical loses and the fish become unsuitable for human consumption.[4-5] Recently, a large *Spironucleus salmonicida* outbreak that caused systemic infection was observed in northern Norway leading to destruction of 640 tons of salmon,[3] all of these outbreaks came from the same farm and consequently spironucleosis is a recurring problem in farmed Atlantic salmon.[6]

Diplomonodida are described as amitochondriate, flagellated, protists found in microaerophilic environments.[7] Diplomonads (e.g. *Giardia intestinalis)* has a simple intracellular composition, lacks mitochondria and Golgi apparatus.[8-9] In addition diplomonads can be found as commensals and in parasitic forms.[2,7] Parasites among Diplomonads can infect large variety of hosts among them different fishes such as salmonis, cichilds, gadids and cyprininds. The parasites are usually found in sea water or fresh water in Europe, Asia and North America.[10] Many species of genus *Spironucleus* cause systemic infection in both ornamental and farmed fish, and infectious disease in birds and mice.[1,6] The virulence mechanism of *Spironucleus* is still unknown. Though, *Giardia intestinalis* and *S. salmonicida* belong same order diplomonodida, there are few publications about *S. salmonicida.*

The life cycle of *S. salmonicida* is still unknown. *S. salmonicida* grow as trophozoites *in vitro*. Eleven cyst wall proteins have been identified in the *S. salmonicida* genome (unpublished data) and this suggests that the parasite is spread via cysts, but cysts remain to be identified. Thus this parasite would have a similar life cycle to *Giardia*.

1.2. *Giardia intestinalis*

The renowned intestinal parasite *Giardia intestinalis* (also called *Giardia lamblia* or *Giardia duodenalis*) is a flagellated, binucleated protozoan found in human and other mammals.[9] Though it was discovered over 300 years ago by Antony van Leeuwenhoek, the disease causing mechanism is still poorly understood.[9,11] It is the common cause of giardiasis worldwide and it is estimated around 280 million people are infected by symptomatic giardiasis per annum[11-12] with 0.5 million new cases per year.[13] Recently, Giardiasis has been included in the neglected disease initiative by WHO. The symptoms are characterized by acute watery diarrhea, dehydration, weight loss and abdominal discomfort.[11] *Giardia* is the major human intestinal pathogen globally, and the common cause of diarrhea in developed countries.[13] The infection spreads via the fecal oral route. It is generally caused by contaminated drinking water and only 10 cysts is enough to cause infection.[11,14] Enteric infection with *Giardia* species causes microvillus shortening, ion hypersecretion, malabsorption and intestinal hypermotility.[15-16] Metronidazole frequently used for the treatment of giardiasis, is an old and effective drug though it has

common side effects which include nausea, headache, vertigo and predisposed to carcinogenic in mice upon high dose.[17]

Giardia intestinalis can be divided into eight different assemblages (A-H) on the basis of host specificity and genetic diversity.[18] Assemblage A and B infect human and other animals (e.g., livestock, dogs, cat and rats) and assemblage B has high allelic sequence heterogeneity and more commonly found in human.[18] Assemblage C and D cause disease in dogs, cats, wolves and coyotes. Assemblage E has been identified in cattle, goat, sheep, pigs and water buffaloes. Aeemblage F and G have been identified in cats and rats respectively.[9] Assemblages A (WB), B (GS) and E (P15) are the three *Giardia* isolates from which the genome has been sequenced.[19] Whole genome comparison generates large amount of information, e.g. genome conservation, genome evolution, gene content, gene regulatory elements and assemblage specific gene.

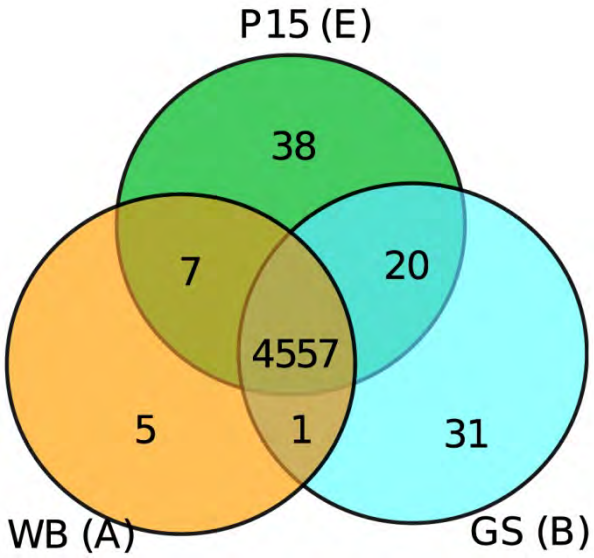


Figure 1: Comparative analysis of shared and non-shared genes among three *Giardia* assemblages. *Giardia's* genome consist of 4557 genes out of which 31 genes found to be specific for assemblage B (GS), 38 genes for assemblage E (P15) and 5 genes for assemblage A (WB). Picture is taken from Jerlström-Hultqvist et al. 2010.

Comparative genomic of three *Giardia* species revealed 74 assemblage specific genes.[19] These assemblage specific putative genes were selected for characterization in this study. Here, new strategies are suggested for the development of diagnosis and epidemiologic tools for giardiasis. Therefore the protein products of these assemblage specific genes were used for antibody generation and subsequently these polyclonal antibodies were used for characterization studies, e.g. *Giardia* genes ID GL50803_10192 is the only isolate specific gene shared between assemblage A (WB) and assemblage B (GS), while assemblage A and B are the only assemblages that cause disease in humans, whereas GLP15_874 were selected on the basis of organism best hits in gene banks.

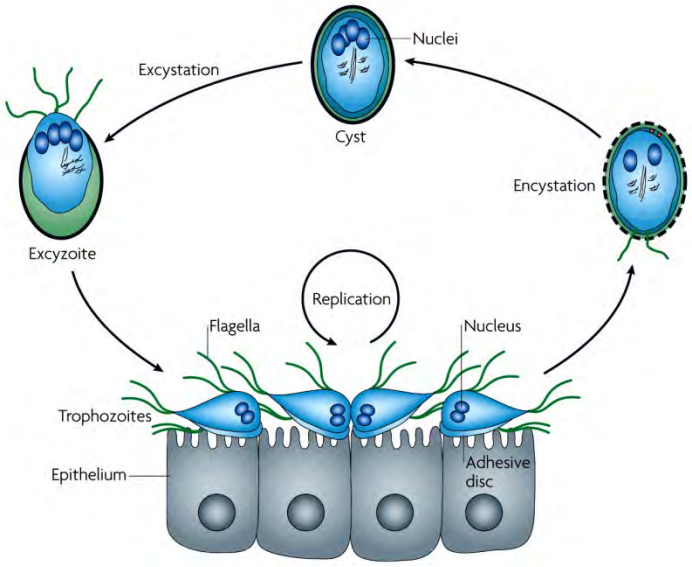


Figure 2: The life cycle of *Giardia intestinalis*. When the *Giardial* cysts are exposed to the host gastric acid, that trigger the *Giardial* cysts to differentiate into trophozoites via the excyzoites process. The trophozoites proliferate in small intestine and when the trophozoites travel further down in the intestine, the process of cyst wall formation begins (encystation). Encystation leads to the formation of water resistant cysts. This picture is taken from Ankarklev et al. 2010.

The life cycle of *G. intestinalis* can be divided into vegetative growing trophozoites and infective cysts (Figure 2). The cyst is highly resistant to the surrounding environment, dormant life form and important for disease transmission whereas, trophozoite represents the motile, vegetative form, colonizing the small intestine and cause diarrhea and malabsorption. In order to complete the life cycle, the trophozoites form cysts in the small intestine and are passed through feces. Cysts can survive outside the host for several months.[9,12].

G. intestinalis has adapted two main strategies for immune evasion transmission and survival termed as encystation and excystation.[20-21] Encystation is the gradual transmission of trophozoite to cyst, during this process, trophozoites loses ability to attach, the flagella becomes internalized, there is decrease in metabolism (dormancy) and encasement in extracellular cyst wall (CW).[21-22] CW is made up of cyst wall proteins (CWPs) and glycopolymers. The CW protects *Giardia cysts* from disinfectants and host stomach acids.[21] CWPs are synthesized during early encystation, were large encystation specific vesicles (ESVs) are formed, which export CWPs to construct the cysts wall.[21] The main function of Golgi apparatus is protein processing and/or post translational modification. It suggested that the ESVs works like a Golgi apparatus. It supports the idea that ESVs is a primordial form of Golgi.[23]

1.3. Cyst wall proteins

Giardia cyst wall is important for the survival (outside host environment) and infection.[24] It is around 0.3-0.5 µm thick, and covered by a double inner membrane.[9] The *Giardial* cyst is composed of around 57% protein and 43% carbohydrates of which *N*-acetyl Galactosamine is the main constituent.[25] Three CWPs have been identified yet, which include CWP1, CWP2, and

CWP3.[9] Leucine-rich repeats and positionally conserved cysteine residues are the desirable characteristics of CWPs[9] *Giardial* cyst wall is mainly made up of CWP1 and CWP2.[26] Encysting trophozoites express CWP1 and CWP2 with the same kinetics.[26] CWP2 is important for the formation and biogenesis of ESVs and induce ligand sorting with the help of CWP1 and CWP3. Therefore CWP2 is the most important protein target in *G. intestinalis.*[27]

Currently, there is no vaccine available against *Giardia intestinalis*, despite the fact it has great clinical importance globally.[28] A couple of studies have been done to measure the efficacy of live and DNA vaccines.[28] Comparative analyses demonstrated that encystation specific antibodies against CWP2 reduce cyst shedding and transmission.[29,30] Transmission-blocking vaccine against *Giardia* only protect against cysts but not trophozoites.[28] Recently α1-giardin (conserved human and murine antigen) is indicating a suitable vaccine candidate against giardiasis.[28]

1.4. The *Giardia* mitosomes

Giardia intestinalis possess the simplest form of mitochondrion-related organelles termed as mitosome.[31] If mitosomes participate in ATP (adenosine triphosphate) generation is still unknown, however, it is established that the *Giardia* mitosomes have role in iron sulfur cluster assembly machinery.[31-32] The mitosomes distribution varies from 25 to 100 per cell. Mitosome populations can be divided in two types on the basis of their location central mitosomes and peripheral mitosomes.[9] Several proteins have been localized in the *Giardia* mitosome.[9,33] *G. intestinalis* has a minimal proteome with reduced mitochondrial metabolism.[9,33] The *Giardia* mitosomes are confined to biogenesis of iron sulfur cluster pathway and reflect simple protein import pathway for organ biogenesis.[9,33] Mitosomal protein targeting into the organelle is governed by N-terminal presequences or internal targeting signals. N-terminal presequences are composed of 8-10 amino acid and arginine residues (positively charged at the cleavage site motif recognized by giardial processing peptidase). The positive charge of mitosomal presequences is comparatively lower (mitochondrial presequences contain several positive charge residues) therefore, the mitosomal membrane reflects low membrane potential. Several proteins are imported into the *Giardial* mitosome which lacks N-terminal presequences. Cpn 60 and Cpn 10 (homolog) are chaperonins present in the mitosomes of *Giardia*. Cpn 60 seems to have lost the N-terminal targeting presequences. Conversely Cpn 60 and Cpn 10 mediate proper protein folding into naive conformation in the mitosomal matrix.[34] ATP is required for mitosomal function but it is unknown how this compound is exported to cytosolic machinery or mechanism behind the FeS cluster machinery.

1.5. Aims of study

The aims of this study were to identify encystation associated genes in *S. salmonicida* by comparative bioinformatics analysis and to characterize putative diagnostic proteins in *S. salmonicida* and *G. intestinalis.* Cyst wall Protein-D (CWP-D), part of the putative cyst wall and Cpn 60 has been localized to the mitosome of *Spironucleus salmonicida.* These proteins were over expressed in *E.coli* and specific antibodies will be raised for biological characterization of these protein. In addition, the efficiency for these different proteins over expression protocols were tested. Another part of this project was to identify conserved genes between *S.*

salmonicida and *Giardia* spp. by bioinformatics analysis. Finally, Putative protein GLP15_874 and GL50803_10192 were characterized in A, B and E assemblage of *Giardia intestinalis* trophozoites.

2. MATERIALS AND METHODS

2.1. Bioinformatics

A list of *Giardia* isolate specific genes was constructed. This list contains predictive putative genes belonging to three *Giardia* isolates, -'P15, GS and WB'-. Predicted protein coding sequences was generated using *Giardia* DB (http://giardiadb.org/giardiadb) for each open reading frame (ORF). If ORF start with an invalid start codon the ORFs were trimmed. Using these predicted protein coding sequences, Blast-P searches was performed using *S. salmonicida* (unpublished genome) for each ORF, the E-values and the score were arranged manually in Table 4.

2.2. Over-Expression, Purification, Detection and Localization of Recombinant Protein CWP-D and Cpn 60

The Glutathione S-transferase (GST) Gene Fusion System was used for the over expression, purification, and detection of fusion proteins produced in *Escherichia coli*. BL21 is an *E. coli* strain, which is protease-deficient and engineered to maximize expression of full-length fusion protein. The expression of the fused gene is under control of the *lac* promoter which is induced by isopropyl β-D thiogalactoside (IPTG). The vector also carries an internal *lacI*^{q} gene which codes for a repressor protein that binds the operator region of the *lac* promoter to prevent induction of expression until IPTG is present. Fusion protein was purified from bacterial lysate by affinity chromatography using Glutathione Sepharose 4Bmedium (GST gene fusion system; GE Healthcare; 18-1157-58 AB 41). Cleavage of the desired protein from GST was achieved by site-specific PreScission protease. Detection of fusion protein was done by immunoblotting.

To get maximum solublization of recombinant protein for the production of antibodies, the efficiency of different protocols in small scale expression experiments were tested. Therefore different media, IPTG concentrations, temperatures, and induction times were tested (Table 1 and 2). On the basis of the small scale expression, large scale protein purification protocols were developed, in order to get maximum amount of the proteins of interest.

2.2.1. Small scale protein expression

The over expression plasmid, pGEX-6P-3 (GE Healthcare), which contain desired gene encoding protein Cpn 60 or CWP-D were subcloned using *BamHI* and *XhoI* restriction sites of pGEX-6P-3 were transformed into *E.coli* BL21 cells. A single colony of *E.coli* containing pGEX-6P-3 vector carrying the gene of interest was used to inoculate 5 ml of LB (containing 50 µg/ml of ampicillin), and the culture was allowed to grow overnight (O/N) on a shaker at 37^{o}C. The O/N culture was diluted (1:100) with 50 ml of fresh pre warmed LB (containing 50 µg/ml of ampicillin). The culture was grown at 37^{o}C until an optical density (OD_{600nm}) reached 0.5 (Table 1 and Table 2). Thereafter 100 µl of un-induced sample was mixed with 100 µl of 2X SDS loading dye (see

Appendix for details) and was boiled (100°C; 5min) for SDS-PAGE analysis (see Appendix). Production of the protein was induced by adding 1 mM IPTG and culture was placed on a shaker at room temperature (RT) for additional 2 hours (Cpn 60) and 3 hours (CWP-D). Then 100 µl of induced sample was prepared for SDS-PAGE, as previously described. Cells were harvested by centrifugation (4500 rpm for 20 min) of 10 ml of the culture. The cell pellets were resuspended in 1 ml ice cold sonication buffer (see Appendix). The resuspended cells were disrupted using sonication for 30 seconds on ice. Thereafter 1% of Triton-X was added to aid solubilization of soluble proteins and incubated in end-over-end rotations (30 min; 4°C). The cell debris was removed by centrifugation at 14000 rpm for 20 min at 4°C. Samples for SDS-PAGE were prepared by taking 200 µl from supernatant and pellet and equal amount of 2X SDS loading dye. The samples were then loaded on 10% SDS-PAGE gels and ran at 150 V for approximately 45 minutes. Thereafter the gels were pre-fixed for at least 30 minutes followed by staining with Coomassie Brilliant Blue for four hours and then the gels were destained overnight. See appendix for staining solution compositions.

Table 1: Overexpression of Cpn 60 in LB media at 20°C with 0.1mM IPTG and different induction times were tested for the fused protein.

Sample no	**Induce at OD**	**OD**	**Time***	**OD**	**Time***	**OD**	**Time***
1	0.5	1.3	3	2.4	15	2.4	45
2	1.2	1.8	3	2.3	15	2.4	45
3	1.8	1.9	1	2.3	15	2.4	45

* Culture harvesting time (hours) after induction.

Table 2: Overexpression of CWP-D grown in LB media at 20°C and different IPTG concentrations and induction time were tested.

Sample no	**Induce at OD**	**IPTG mM**	**Time***	**OD after induction**	**Time***
1	0.5	0.1	3	0.75	6
2	0.5	0.5	3	0.69	6
3	0.5	1	3	0.75	6
4	1	0.1	2.5	1.08	5.5
5	1	0.5	2.5	1.21	5.5
6	1	1	2.5	1.32	5.5
7	2	0.1	2.5	1.66	4.5
8	2	0.5	2.5	1.53	4.5
9	2	1	2.5	1.65	4.5

*Culture harvesting time (hours) after induction.

2.2.2. Large scale protein purification

On the basis of small scale purification, protocol for large scale purification was designed. An overnight pre-culture of *E.coli* BL21 containing pGEX-6P-3 plasmid, which contain desired gene encoding protein Cpn 60 or CWP-D was used to inoculate 4 liters of LB (containing 50 µg/ml of ampicillin) and grown at 37°C until an OD_{600nm} reached 0.5. Protein production was induced by

adding 1 mM in IPTG and the culture was grown at room temperature for and additional 3 hours. The cells were harvested by centrifugation (20min at 4500 rpm) and resuspended in 50 ml sonication buffer (see appendix) followed by sonication for a total of 3 minutes per 25 ml of cell suspension. After adding 1% Triton-X to the tubes they were incubated at 4 degrees at end-over-end rotations for 30 minutes to aid solubilization of fused protein. The cell suspension was centrifuged 4500 rpm (20 min; 4^{o}C) and supernatant containing fusion protein was transferred into a new tube. Glutathione Sepharose 4Bmedium (GST gene fusion system; GE Healthcare; 18-1157-58 AB 41) was used for batch purification i.e. the resin was added to the supernatant containing the fusion protein and incubated at 4 degrees over night for binding. The resin was then spun down (500 xg, 10 minutes) followed by washing with PBS to remove as much of unbound protein and protease inhibitors as possible. After the washing of the resin containing the bound fusion protein, the resin was equilibrated in elution buffer and then the site specific PreScission Protease (GE Healthcare) was added. Elution of the fusion protein took place at 4 degrees over night. Thereafter the protein concentration was measured by Nanodrop and also by Bradford assay. At least 2.5 mg of purified protein was sent to Capra Science (Ängelholm, Sweden) for polyclonal antibody production.

2.4. Western blot

All incubation steps were performed at room temperature. The PVDF membrane was activated in 100% methanol and then put in transfer buffer. The blot was prepared by placing the membrane on the gel and using filter paper and pads to keep membrane moist. The stack was placed in electrophoresis unit containing with transfer buffer. The transfer was done overnight at 4^{o}C at 35 mV with continuous stirring.

After the transfer the membrane was incubated in blocking solution (see appendix) and the membrane was blocked for one hour. The membrane was washed three times for five minutes using PBS-Tween 20 (0.1%). The membrane was incubated with the primary antibody either WB_10192 or P15_874 (diluted 1: 2000 in PBS containing 1% BSA) for two hours. The membrane was washed three times with PBS-Tween, followed by incubation with secondary antibody anti-rabbit coupled with HRP (DAKO) (diluted1:7500 in PBS containing 3% non-fat dry milk) for one hour. The membrane was washed three times with PBS-Tween for five minutes. The membrane was treated with peroxide phosphate substrate detection solution following the manufacturer's protocol (Amersham ECL Plus detection kit, GE Health care). Excess liquid was drained off and the membrane was placed on saran foil. Wrapped blot was place in an x-ray film cassette. The images of the immunoblot were captured using Bio RAD Gel Doc XR and images were analyzed and optimized by Bio-Rad Image LAB^{TM} software.

2.5. Paraformaldehyde fixation

Giardia trophozoites was cultivated at 37^{o}C in 10 ml culture tube, in order to detach the cells from the walls of tubes, they were put on ice for 20 mintes. Cells were harvested by centrifugation (5min, 500 x g, 4^{o}C) and the cells were washed twice using 1 ml ice cold HBS+ Glucose buffer. 20 µl cells were transferred on the poly-L-lysine coated slides (Thermo Scientific) and cells were allowed to attach on the slide for 5 min at 37^{o}C. Cells were then fixed with 20 µl

ice cold 4% Paraformaldehyde (PFA) in PBS and incubated (25 min; 37^{o}C). The cells were washed twice with PBS, quenched with 0.1 M glycine dissolve in PBS and washed with PBS twice. The cells were permeabilized with 0.2% TritionX-100 in PBS (RT; 30min), washed thrice with PBS (5min), and blocked with 2% BSA in PBS (O/N; 4^{o}C).

2.6. Immunofluorescence

The blocking solution was removed by vacuum suction. 15 µl of primary antibody anti-WB10192 polyclonal antibody or anti-P15_874 polyclonal antibody was diluted 1:500 in PBS containing 0.1% TritonX-100 with 3% serum. Cells were incubated for 1h, washed six times with 15 µl of PBS and incubated (in dark) for 2h with 15 µl FITC-conjugated anti-rabbit antibody (Sigma) diluted (1:50) in PBS containing 0.1% TritonX-100 with 3% serum. Cells were washed six times with PBS to remove unbound antibody and then 5 µl Vectashield mounting medium with DAPI (Vector laboratories Inc., Burlingame CA) was added. A coverslip was placed on the slide and sealed with nail polish and stored at 4^{o}C in dark awaiting microscopy.

Florescence microscopy was performed on Zeiss Axioplan2 microscope and images were processed with the software Axiovision Rel. 4.8 (http://www.micro-shop.zeiss.com) and Adobe Photoshop CS5.

3. RESULTS

3.1. Comparative analysis of upregulated encystation genes in *G. intestinalis* trophozoites and *S. salmonicida*

Comparative microarray analysis previously identified 54 genes whose expression is upregulated during the first eleven hours of encystation in *Giardia intestinalis* (Morf et al, 2010). To further identify common ortholog genes between *G. intestinalis* and *S. salmonicida*, a bioinformatics Blastp search was performed in *S. salmonicida* genome database (unpublished). Comparative analysis of the upregulated genes in the *G. intestinalis* and *S. salmonicida* genome data base revealed a large number of genes with strong hits in *S. salmonicida* genome data-base. The following results suggest that *S. salmonicida* might have same life cycle pattern as *Giardia* possess.

Table 3. Comparative analysis of upregulated encystation genes in *G. intestinalis* and *S. salmonicida.* *Giardia* gene ID (http://www.giardiadb.org/giardiadb/) and annotation were listed in column 1 and column 2 respectively. *Giardia* genes ID were used as queries in *S. salmonicida* genome against the 2010 release of *Giardia intestinalis* protein dataset. *S. salmonicida* homologs of these proteins are listed with *S. salmonicida* gene ID (column 3) and annotation (column 4). The score of *S. salmonicida* is shown in column 5 and E-value in column 6. The score values in column 5 is the measure of similarity between two sequences whereas; The E-values in column 6 represent the probability of the alignment occurring by chance. Same or similar annotations are displayed bold letters.

Giardia Gene ID	Giardia DB Annotation	*Salmonicida* Gene ID	Annotation	Score	E-value
GL50803_9115	**Glucose-6-phosphate isomerase**	**SS50380_12284**	**Glucose-6-phosphate isomerase**	**453**	**3.00E-128**
GL50803_92729	Fatty acid elongase 1	SS50380_18628	3-ketoacyl-CoA synthase	362	6.00E-101
GL50803_7982	**UDP-glucose 4-epimerase**	**SS50380_17325**	**UDP-glucose 4-epimerase**	**315**	**6.00E-87**
GL50803_17043	**Glyceraldehyde 3-phosphate dehydrogenase**	**SS50380_13399**	**Glyceraldehyde 3-phosphate dehydrogenase**	**309**	**6.00E-85**
GL50803_16069	**Phosphoacetylglucosamine mutase**	**SS50380_14999**	**Phosphoacetylglucosamine mutase**	**288**	**1.00E-78**
GL50803_7260	**Aldose reductase**	**SS50380_13549**	**Aldo/keto reductase**	**285**	**9.00E-78**
GL50803_8245	**Glucosamine-6-phosphate deaminase**	**SS50380_10983**	**Glucosamine-6-phosphate deaminase**	**261**	**6.00E-71**
GL50803_10552	Hypothetical protein	SS50380_13890	GCC2 and GCC3 domain containing protein	260	4.00E-70
GL50803_5638	**Cyst wall protein 1**	**SS50380_17413**	**Cyst wall protein**	**221**	**9.00E-59**
GL50803_14247	TM efflux prot	SS50380_12119	Major facilitator superfamily protein	217	4.00E-57
GL50803_5435	**Cyst wall protein 2**	**SS50380_17413**	**Cyst wall protein**	**214**	**2.00E-56**
GL50803_14259	**Glucose 6-phosphate N-acetyltransferase**	**SS50380_14553**	**Glucose 6-phosphate N-acetyltransferase**	**156**	**2.00E-39**
GL50803_88581	Synaptic glycoprotein SC2	SS50380_16582	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	149	5.00E-37
GL50803_137680	**Cathepsin L-like protease**	**SS50380_10919**	**Cathepsin L**	**150**	**6.00E-37**
GL50803_3063	Hypothetical protein	SS50380_14467	Conserved hypothetical protein	142	1.00E-34
GL50803_9355	Hypothetical protein	SS50380_17283	Redoxin domain protein	132	2.00E-32

GL50803_8722	**Myb 1-like protein**	**SS50380_12979**	**Protein containing Myb-like DNA-binding domain**	**130**	**6.00E-31**
GL50803_14626	Oxidoreductase, short chain dehydrogenase/reductase family	SS50380_18340	3-ketoacyl-CoA reductase	117	2.00E-27
GL50803_9620	High cysteine membrane protein Group 2	SS50380_10984	Conserved hypothetical protein	119	2.00E-27
GL50803_112432	High cysteine membrane protein Group 5	SS50380_16894	Conserved hypothetical protein	117	7.00E-27
GL50803_4846	protein 21.1	SS50380_18810	Protein containing ankyrin repeats	110	1.00E-24
GL50803_2421	**Cyst wall protein 3**	**SS50380_17413**	**Cyst wall protein**	**105**	**7.00E-24**
GL50803_21924	Kinase, NEK	SS50380_10401	Protein containing ankyrin repeats	99,8	1.00E-21
GL50803_15250	High cysteine membrane protein Group 6	SS50380_14690	Cysteine-rich protein	94,4	4.00E-20
GL50803_24412	protein 21.1	SS50380_18810	Protein containing ankyrin repeats	94	5.00E-20
GL50803_137701	Kinase, NEK	SS50380_13477	Protein containing protein tyrosine kinase domain	88,6	8.00E-19
GL50803_102813	protein 21.1	SS50380_12798	Conserved hypothetical protein	88,6	1.00E-18
GL50803_8987	Hypothetical protein	SS50380_18766	Conserved hypothetical protein	85,1	5.00E-18
GL50803_11149	Hypothetical protein	SS50380_11910	Conserved hypothetical protein	81,3	2.00E-16
GL50803_101699	protein 21.1	SS_FX_049	Protein containing ankyrin repeat	83,6	2.00E-16
GL50803_5800	Hypothetical protein	SS50380_19064	Conserved hypothetical protein	70,9	2.00E-13
GL50803_88814	Protein kinase	SS50380_12700	Hypothetical protein	65,5	6.00E-12
GL50803_106496	Hypothetical protein	SS50380_12700	Hypothetical protein	60,8	1.00E-10
GL50803_6492	Hypothetical protein	SS50380_16722	Glycerophosphoryl diester phosphodiesterase family Protein	61,2	5.00E-10
GL50803_12082	Hypothetical protein	SS50380_11530	Biotin-(acetyl-CoA-carboxylase) ligase	50,1	3.00E-07
GL50803_23015	Serine palmitoyltransferase 1	SS50380_12195	Conserved hypothetical protein	50,4	7.00E-07
GL50803_7139	Hypothetical protein	SS50380_10683	Phosphopantetheine adenylyltransferase	40,8	2.00E-04
GL50803_10425	Hypothetical protein	SS50380_11360	Protein containing transmembrane domains	37	0.002
GL50803_5810	Hypothetical protein	SS50380_15691	Protein containing pyridoxamine 5'-phosphate oxidase Domain	34,3	0.006
GL50803_7134	Hypothetical protein	SS50380_15459	Hypothetical protein	33,9	0.068
GL50803_32419	Hypothetical protein	SS50380_18133	Conserved hypothetical protein	33,9	0.068
GL50803_14759	6-phosphogluconate dehydrogenase, decarboxylating	SS50380_10436	Membrane occupation and recognition nexus (Morn) repeat protein	32,7	0.12
GL50803_8652	Hypothetical protein	SS50380_14996	Hypothetical protein	31,6	0.25
GL50803_89849	Hypothetical protein	SS50380_13920	Conserved hypothetical protein	27,3	0.44
GL50803_103785	Hypothetical protein	SS50380_13745	Hypothetical protein	28,9	0.87
GL50803_9046	Sugar transport family protein	SS50380_14281	Major facilitator superfamily protein	29,6	0.99
GL50803_32657	Hypothetical protein	SS50380_18951	Protein containing transmembrane domains	28,1	1.5
GL50803_7388	Hypothetical protein	SS50380_12340	Conserved hypothetical protein	25,4	1.7

GL50803_93488	Hypothetical protein	SS50380_14401	Kinase, CDC7	24,6	2.6
GL50803_28112	Hypothetical protein	SS50380_16145	Hypothetical protein	23,5	6.5
GL50803_27028	Hypothetical protein	SS50380_16845	Protein containing papain family cysteine protease Domains	23,5	7.1
GL50803_35999	Hypothetical protein	No hits found			
GL50803_37010	Hypothetical protein	No hits found			

3.2. Identification of conserved genes between *Giardia intestinalis* and *Spironucleus salmonicida* by bioinformatics analysis

To determine the most promising genes for future assemblage specific diagnosis in *Giardia,* a bioinformatics analysis was performed. In total, 74 *Giardia* isolates specific ORFs were selected and a BlastP search was performed against *Spironucleus salmonicida* genome database (unpublished). 34% (25 in total) of the selected *Giardia* ORFs were displayed as strong hits in BlastP searches performed against the *S. salmonicida* genome database (unpublished genome), 5 ORFs were identified with significant E-value (Table 4). However 24% (18) of the ORFs found hypothetical while 28% (21) of the ORFs were found unique for *Giardia* (no hits found). 14% (10) genes with annotated function conserved hypothetical protein.

Table 4. Conserved genes between *Giardia* and *Spironucleus salmonicida*; *Giardia* assemblage specific genes that were found to be expressed in *S. salmonicida*. E-value and score are indicated in outer right.

Giardia GENE ID	ANNOTATION	Organism with best hit in genbank and E-value	Domains or signatures	*Spironucleus salmonicida* GENE ID	ANNOTATION	E-Value	Score
GL50581_4508	Hypothetical protein	Entamoeba dispar SAW760 4.00E-29	Methyltransferase TRM13 superfamily domain	SS50380_13668	Methyltransferase TRM13 family protein	3E-26	114
GL50581_3637	Hypothetical protein			SS50380_13188	Conserved hypothetical protein	6E-11	65.9
GL50581_3038	Hypothetical protein		Uncharacterized conserved protein, contains RING Zn-finger (COG5219)	SS50380_12274	Protein containing Zinc finger, C3HC4 type (RING finger)	2.00E-04	40.4
GLP15_4996	Hypothetical protein		Reverse transcriptase (RT) catalytic domaim	SS50380_13706	Reverse transcriptase, putative	4.00E-04	38.5
GL50581_2039	Hypothetical protein			SS50380_15986	Protein containing %09CHY zinc finger domain	4.00E-04	38.5

3.3. Immunoblotting against different Giardia assemblages by using cell extractions to evaluate affinity purified antibodies

Western blot was used to determine the specificity of polyclonal antibody produced. Affinity purified anti P15_874 antibody revealed a specific bands for P15 trophozoites and the antibody did not express against *Giardia* isolate WB (cyst and trophozoites) and GS trophozoites (Figure 3A). Even higher specificity was detected when P15 trophozoites were incubated at higher dilution (1: 1000). This experiment confirms that P15_874 is an assemblage specific protein and it secreted by Giardia trophozoites assemblage E (in vitro).

Anti WB_10192 antibody showed strong but unspecific binding against WB trophozoites, WB recombinant and P15 trophozoites. WB cysts and GS trophozoites did not express against anti P15_874 antibody (Figure 3B). Therefore WB_10192 antibody is not specific against WB assemblage, probably due to low expression of protein in trophozoites.

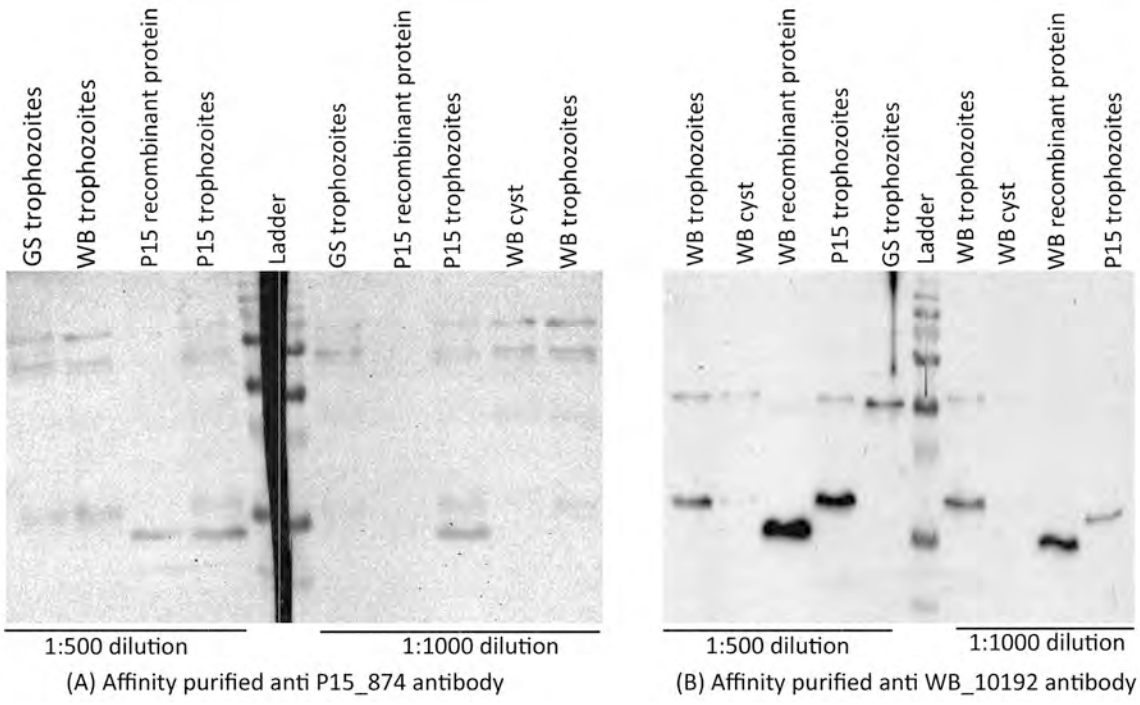


Figure 3: Antibody validation by Western blot analyses, Figure 3(A) P15_874 immunoblotting, Figure 3(B) WB_10192 immunoblotting. Protein were separated by SDS-PAGE (10%) and transferred onto PVDF membranes.

3.4. Localization of Putative Diagnostic Proteins in *Giardia*

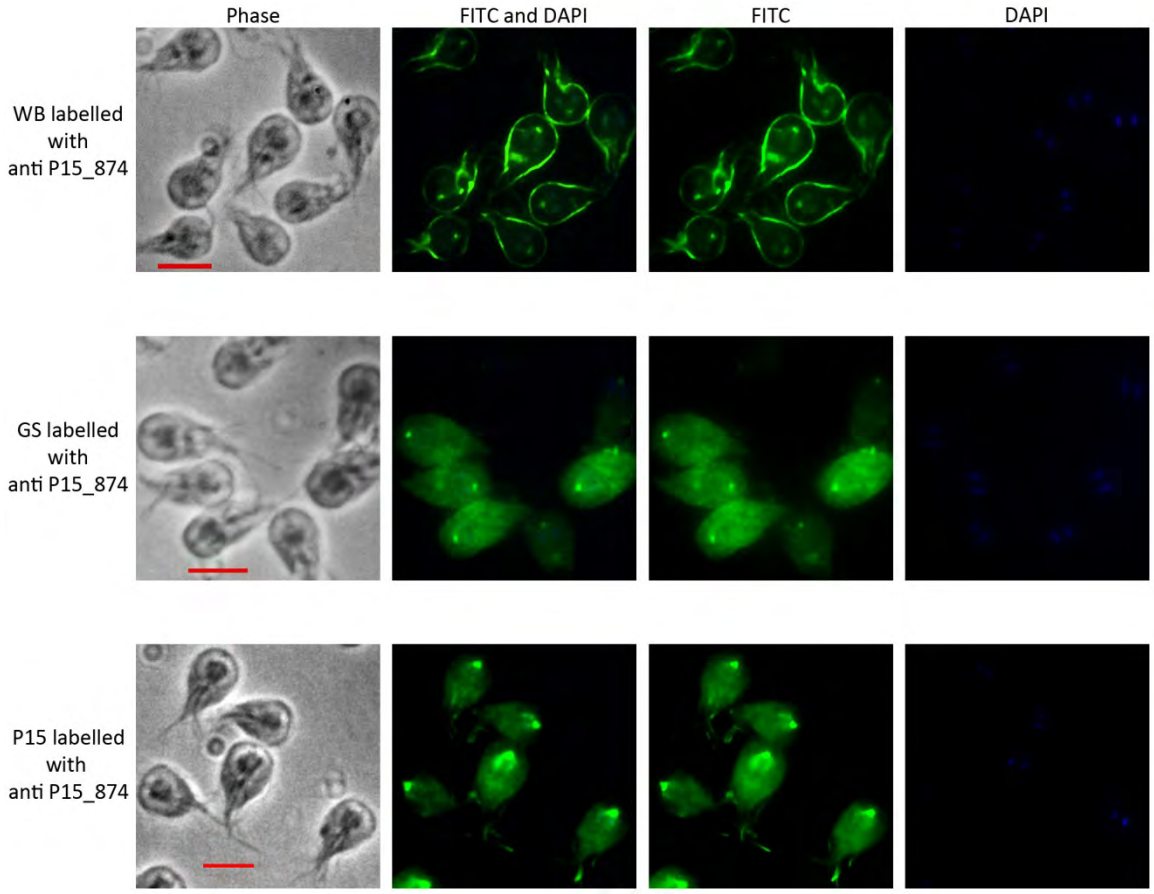


Figure 4: Immunolocalization of P15_874 in WB, GS and P15 *Giardia* trophozoites. *Giardia* trophozoites were probed with the anti-P15_874 antibody and FITC-conjugated anti-rabbit antibody. Each assemblage had a distinct localization pattern. In addition, anti P15 localized to plasma membrane in WB assemblage. Morphology is shown in phase contrast image of all the *Giardia* trophozoites on the left. FITC and DAPI are shown in the merged image. Anti-P15_874 antibody is shown in green and Nuclei are labelled with DAPI on the far right. Scale bar = 10µm

P15_874 and WB_10192 protein were characterized by Western blot (Figure 3) and immunofluorescence in *G. intestinalis* WB (assemblage A), GS (assemblage B) and P15 (assemblage E) trophozoites using antibodies (Figure 4, 5). Differential localization pattern of anti-P15_874 was observed among assemblages A, B and E, but localization of anti-WB_10192 revealed same localization in assemblage A and B, while weak localization in assemblage P15 was observed (Figure 4, 5).

Giardia assemblage E specific gene GLP15_874 is annotated as acetyltransferase. WB trophozoites labelled with anti P15_875 antibody showed strong stain in plasma membrane with weak nuclear stain while, GS and P15 trophozoites showed a dot like or circular structure at the anterior part of the cells (Figure 4).

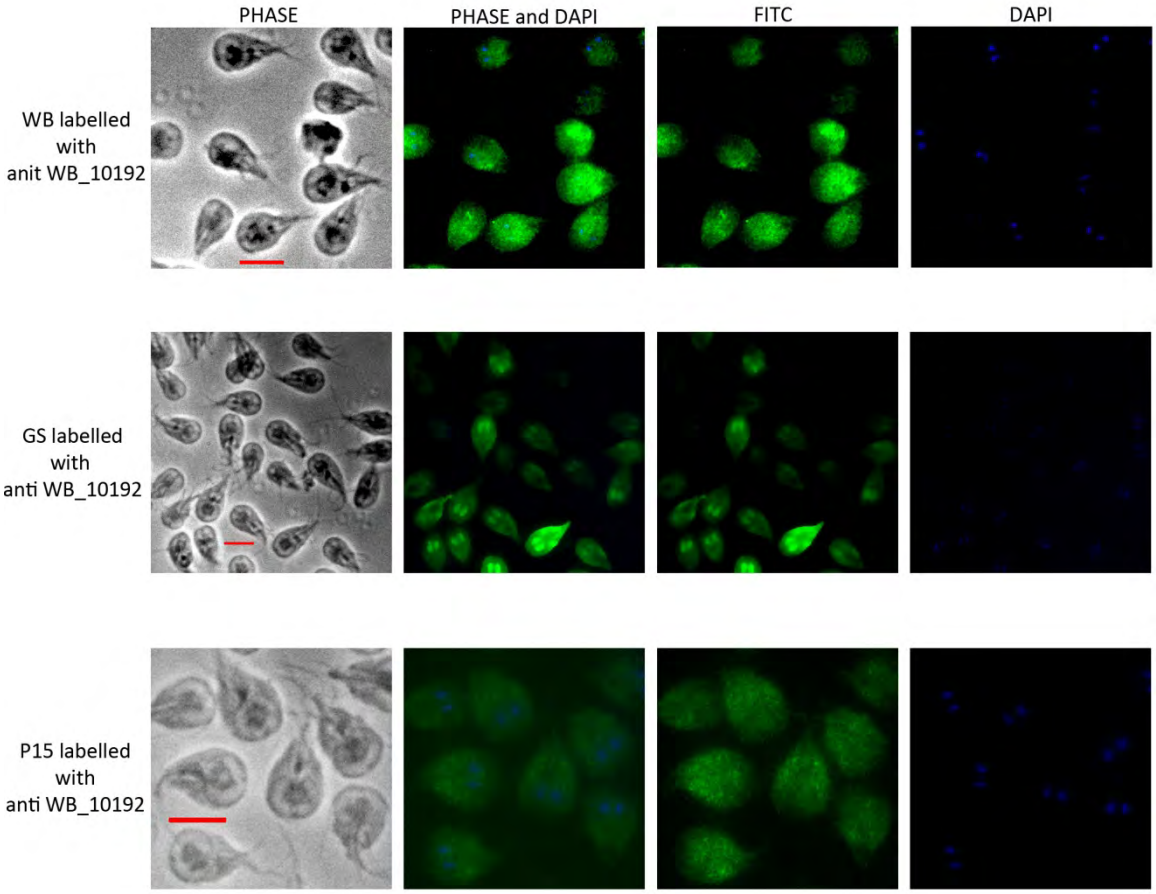


Figure 5: Immunolocalization of WB_10192 in WB, GS and P15 *Giardia* trophozoites. *Giardia* trophozoites were probed with anti-WB_10192 antibody and FITC-conjugated anti-rabbit antibody. Each assemblage had a distinct localization pattern. Morphology is shown in phase image of all the *Giardia* trophozoites on the left. FITC and DAPI are shown in the merged image. Anti-WB_10192 antibody is shown in green and the nuclei are labelled with DAPI (blue) on the far right. Scale bar=10µm

Assemblage A specific gene GL50803_10192 is annotated as a hypothetical protein and showed localization in the entire cytoplasm in WB, GS and P15 isolates. In addition strong nuclei staining observed in WB trophozoites (figure 5).

3.5. Optimization of *Spironucleus salmonicida* CWP-D and Cpn 60 recombinant Protein expression

Over-expression levels were optimized by testing different conditions such as length of induction time, induction temperatures, IPTG concentrations and different growth media (Table 1 and Table 2). Using LB media with an IPTG concentration of 1mM and induction at 0.5 OD (OD_{600nm}) for 2h and 3h was the optimal condition on the basis of Coomassie blue gel band for CWP-D and Cpn 60, respectively (Figure 6). Distinct band separation at 82.4 kDa and 71.7 kDa for supernatants containing the soluble tagged protein suggest that 2 hours and 3 hours induction time was optimum for Cpn 60 and CWP-D, respectively.

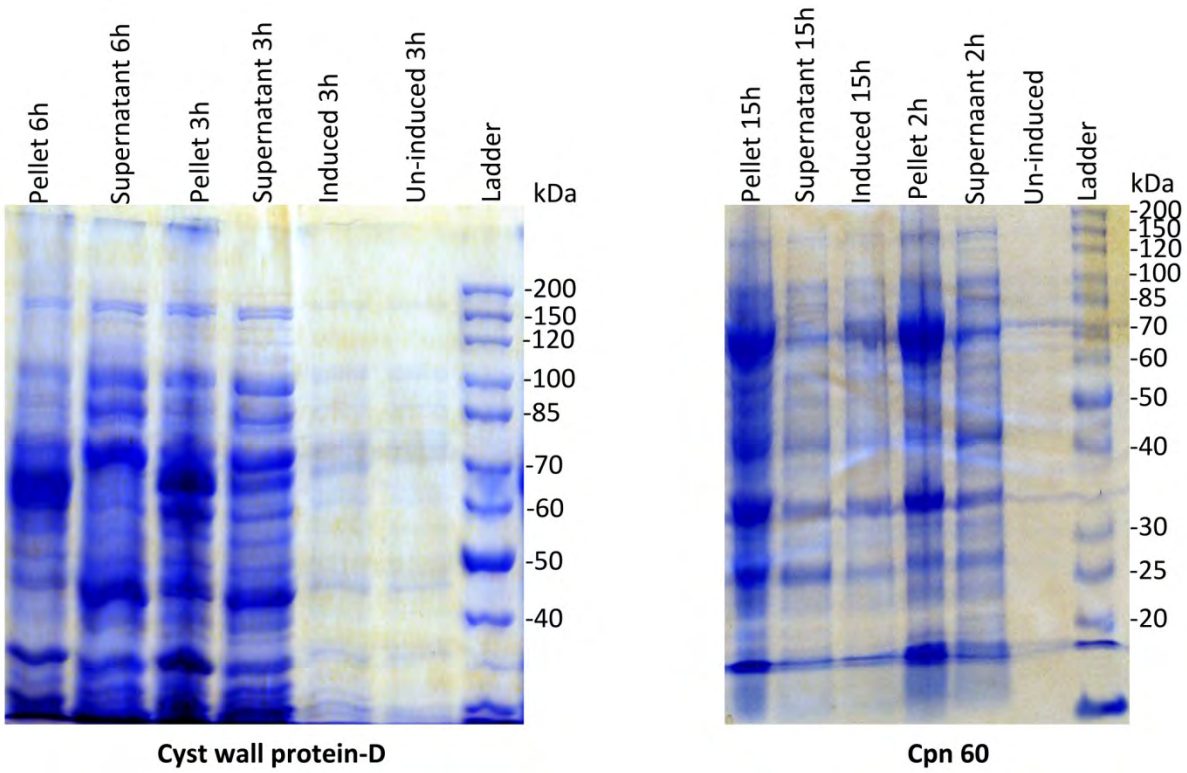


Figure 6: SDS-PAGE representing the purification of recombinant proteins, Left purification of recombinant CWP-D, right purification of recombinant Cpn 60. 20 µl/lane samples from un-induced, induced, supernatant and pallet were separated on 10% SDS-PAGE gel and visualized with Coomassie blue staining.

4. DISCUSSION

Giardia's genome comprise many lateral gene transfer (LGT) candidates, indicating that that LGT play an important role in influencing metabolic pathways. Protein from *Giardia* with similarity to *S. salmonicida* proteins at a Blast significance level of e^{-24} or better were identified within top 21 matches. These include cyst wall proteins and others are glucosamine 6-phosphate deaminase (GNP; glucosamine 6-P isomerase) these were shown to be relics of LGT.[36]

Previous comparative analysis showed that a core set of genes was upregulated during encystation in *Giardia intestinails*.[35] Blastp searches in the *S. salmonicida* genome identified few proteins known to be crucial for cyst wall formation and encystation.[19] There were 11 proteins with similar annotation identified in *S. salmonicida* genome (Table 3), these might be interesting to characterize more in the future. Similar to *Giardia*, the rodent parasite *Spironucleus muris* also transmits via cysts in the fecal material and also possess four nuclei.[1] Based on previous suggestions that *S. salmonicida* have the same kind of life cycle pattern as *Giardia* possess. This comparative analysis of encystation genes will not only be valuable for investigating evolution of encystation in diplomonads, but will also be applied for the search of life cycle and new therapies for these parasites. The antibody against *Spironucleus salmonicida*'s cyst wall protein might also be very useful in diagnostics of infected fish.

The structure and composition of the cysts wall has been studied in depth in *Giardia*.[9] Blastp searches in the *S. salmonicida* genome detected all three homologs of *Giardia* cysts wall proteins (Table 3). *S. salmonicida* also seem to have several more gene copies of cyst wall proteins than Giardia, in total 11 CWPs are found, and it would be useful to investigate if all these are expressed during the encystation process.

The cyst wall of *Giardia intestinalis* made up of proteins and N-acetyl acetylgalatosamine (GaINAc) polysaccharides, which is the major constituent. GaINAc synthetic enzymes seem to occur at the level of transcription during encystation via the pathways of inducible enzymes in *Giardia*. These five enzymes glucosamine 6-phosphate deaminase, glucosamine 6-phosphate N-acetyltransferase, phosphoacetylglucosamine mutase, UDP-N-acetylglucosamine pyrophosphorylase and UDP-N-acetylglucosamine 4-epimerase are controlled by the activation of transcription of each of their genes in *Giardia*.[37] However four enzymes were present in *S. salmonicida*, UDP-N-acetylglucosamine pyrophosphorylase was not listed in this Table 3. It is possible that all these enzymes are important for encystation in *S. salmonicida*. E-value indicates that the existance of above proteins is probable because a clear ortholog exist in *Giardia*.

'Glucose-6-phosphate isomerase' (GPI) is a LGT candidate in *S. salmonicida* and it is classified as being involved in carbohydrate metabolism; glycolysis / gluconeogenesis.[1] In addition, GPI has been described as glycolytic enzyme in diplomonads *Giardia intestinalis* and *Spironucleus barkhanuas* and the parabasalid *Tricomonas vaginalis*.[38] Therefore, it could be present in the mitosomes of *Giardia* and *Spironucleus*.

Comparative genomic analysis of the two human isolates WB (assemblages A) and GS (assemblages B) and porcine strain P15 (assemblage E) resulted in the identification of 74 putative assemblage specific genes (Figure 1).[19] Conserved genes between three the *Giardia* assemblages and *Spironucleus salmonicida* were cataloged using BLASTP searches of predicted protein coding sequences from *Giardia* assemblages to *S. salmonicida* genome database (unpublished genome).

WB, GS, and P15 are sharing the same genus name despite that their genomes are quite divergent. The average protein identity between WB and P15, GS and P15 were 90% and 81% respectively. The average protein identity between WB and GS was 81%.[19] The average protein identity among A, B and E assemblage indicating that assemblage E (non-human isolate) is more closely related to assemblage A that to assemblage B.

The transmission of genetic information across normal mating barrier, between more or less distantly related organisms termed as Lateral gene transfer (LGT). LGT has had significant role in eukaryotic genome evolution. Diplomonads or their closely related ancestors acquired numerous genes via LGT.[1,19] Recently, *S. salmonicida* genomic survey identified the largest category of genes annotated as 'conserved hypothetical proteins', using combined sequencing approaches for highly expressed and diverse gene set[1]. These results suggest that, genes with annotated function probably not have the appropriate coding potential.[1] However, the above results show that the genes annotated as conserved genes was the smallest category. Probably, *Giardia* spp. and *S. salmonicida* acquired these genes by LGT. On the other hand *Giardia*'s isolate specific genes were usually located in nonsyntenic regions of the genome. Such region primarily encodes variable surface proteins (VSPs) and high cysteine membrane protein (HCMP). These proteins are likely to be linked to antigenic variation or immune evasion.

Five conserved genes between *Giardia* spp. and *Spironucleus salmonicida* were cataloged (table 4) with significant E value. These ORFs might be linked to host specificity and antigenic variation or taken by same ancestor. One previously predicted ORF (GL50803_4508) in *Giardia* was annotated as Methyltransferase TRM13 and this annotation is shared in *S. salmonicida*. These bioinformatic results can be used for future localization studies and potentially provide new diagnostic tools for these parasites.

Giardia is considered to be a zoonotic parasite and therefore it is important to type form which assemblage the parasite belongs to. Assemblage specific antibodies against these proteins could be used for instant and easy typing of *Giardial* infection in clinical laboratories.

The localization of assemblage specific gene WB_10192 and P15_874 were tested in *Giardia* trophozoites. These putative genes were previously expressed and selected on the basis of organism best hit in the gene bank.

GL50803_10192 is the only ortholog shared between *Giardia*'s assemblage A and B (human isolate). It does not exist in the non-human isolate P15,[19] therefore it is speculated that WB_10192 may associate with human specific *Giardial* infection. Blast searches against non-

redundant genome database with ORF WB_10192 show strong sequence homology with actinobacterium *Collinsella aerofaciens* (*e*-value: e^{-27}). *C. aerofaciens* reside in human intestinal tract, therefore this gene may have been transferred via lateral gene transfer in the gut. The affinity purified antibody WB_10192 shows unspecific and weak signal in both localization and Western blot images (Figure 3B and 5). Probably the target protein was present in too low concentration to be detected in trophozoites. For diagnostic purposes the antibody would be used for IMF on cysts but that remains to be tested. In addition, the introduction of epitope tag did not affect the localization of WB_10192.[39]

Assemblage specific gene GLP15_874 encode for acetyltransferase, Blast searches display high sequence similarity to bacterial homologs of the gene.[19] The localization of P15_874 was not identical to that of previously epitope tag results.[39] It could be due to the transient transfected trophozoites. WB, GS and P15 *Giardia* trophozoites were labelled with anti P15_874 antibody. Surprisingly, each *Giardia* assemblage show different localization pattern (Figure 4), anti P15_874 antibody localize to cell wall and flagella in WB vegetative trophozoites. The localization of GS and P15 trophozoites with anti P15_874 exhibited a dot like structure at the anterior part of the cell as well as cytoplasm, however pronounced staining were observed in P15 trophozoites. In addition affinity purified anti P15_874 antibody was analyzed by Western blot which also validate antibody (Figure 3a). Therefore these polyclonal antibodies have potential for future diagnostic purposes.

Future studies could also include localization of the unique proteins in cyst and co-localization of CWPs to ESVs. This could open the door for in depth and specific localization of unique proteins in Giardia and other protozoan parasites.

To study assemblage specific proteins an efficient protocol was needed for inclusion protein's solublization. An important issue for this protein purification was the question of how much is needed for antibody production. In the present study we needed optimally 2.5mg of each protein for successful polyclonal antibody production. The protein purification protocol varies for each recombinant protein and therefore different overexpression protocols have to be tested repeatedly to evaluate their efficiency. The data of different overexpression protocols were shown in Table 1 and 2. Bacterial expression of protein can be carried out at low temperature (4-20^{o}C) to increase the yield of soluble proteins by slowing down the rate of protein expression and allowing time for their proper folding.[40] However for *S. salmonicida* Cpn60 recombinant protein, overexpression condition at 4^{o}C was tested but this condition didn't work (data not shown). Probably, molecular folding modulators might be limiting factor[40] because translation of the expressed protein is slowed down or folding accuracy might increase at lower temperature (4^{o}C).

5. ABBREVIATIONS

WHO: World Health Organization; CW: cyst wall; CWP: cyst wall protein; CWP-D: cyst wall protein-D; ESV: encystation specific vesicles; ORF: open reading frame; IPTG: isopropyl β-D thiogalactoside; GST: Glutathione S-transferase; O/N: overnight; RT: room temperature; Paraformaldehyde: PFA; *Spironucleus salmonicida:* SSK; lateral gene transfer: LGT; glucosamine 6-phosphate deaminase: GNP; glyceraldehyde 3-phosphate dehydrogenase: GAPDH; N-acetyl acetylgalatosamine: GaiNAc; Glucose-6-phosphate isomerase: GPI;

6. ACKNOWLEDGEMENTS

I would like to thank Uppsala University for having enriched my career in so many different ways for over the last two years. Heartfelt thanks to Professor *Staffan G. Svärd* for accepting me as a project student and allowing me to learn different lab techniques. Thanks to *Elin Einarsson* and all the *Giardia*'s team members for substantial support and suggestion. So, I am deeply grateful to Uppsala University.

7. REFERENCES

1. Andersson JO, Sjogren AM, Horner DS, Murphy CA, Dyal PL, et al. (2007) A genomic survey of the fish parasite Spironucleus salmonicida indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. BMC Genomics 8: 51.
2. Brugerolle G, Lee JJ. (2002) Order Diplomonadida**.** In An Illustrated Guide to the Protozoa, 2nd edition*.* Edited by Lee JJ, Leedale GF, Bradbury P. Lawrence, Kansas, Society of Protozoologists: 1125-1135.
3. Jorgensen A, Sterud E (2006) The marine pathogenic genotype of Spironucleus barkhanus from farmed salmonids redescribed as Spironucleus salmonicida n. sp. J Eukaryot Microbiol 53: 531-541.
4. Jorgensen A, Sterud E (2004) SSU rRNA gene sequence reveals two genotypes of Spironucleus barkhanus (Diplomonadida) from farmed and wild Arctic charr Salvelinus alpinus. Dis Aquat Organ 62: 93-96.
5. Paull GC, Matthews RA (2001) Spironucleus vortens, a possible cause of hole-in-the-head disease in cichlids. Dis Aquat Organ 45: 197-202.
6. Jorgensen A, Torp K, Bjorland MA, Poppe TT (2011) Wild arctic char Salvelinus alpinus and trout Salmo trutta: hosts and reservoir of the salmonid pathogen Spironucleus salmonicida (Diplomonadida; Hexamitidae). Dis Aquat Organ 97: 57-63.
7. Roxstrom-Lindquist K, Jerlstrom-Hultqvist J, Jorgensen A, Troell K, Svard SG, et al. (2010) Large genomic differences between the morphologically indistinguishable diplomonads Spironucleus barkhanus and Spironucleus salmonicida. BMC Genomics 11: 258.
8. Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, et al. (1998) A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA 95: 229-234.
9. Ankarklev J, Jerlstrom-Hultqvist J, Ringqvist E, Troell K, Svard SG (2010) Behind the smile: cell biology and disease mechanisms of Giardia species. Nat Rev Microbiol 8: 413-422.
10. Fard MR, Jorgensen A, Sterud E, Bleiss W, Poynton SL (2007) Ultrastructure and molecular diagnosis of Spironucleus salmonis (Diplomonadida) from rainbow trout Oncorhynchus mykiss in Germany. Dis Aquat Organ 75: 37-50.
11. Adam RD (2001) Biology of Giardia lamblia. Clin Microbiol Rev 14: 447-475.

12. Lane S, Lloyd D (2002) Current trends in research into the waterborne parasite Giardia. Crit Rev Microbiol 28: 123-147.
13. Prucca CG, Rivero FD, Lujan HD (2011) Regulation of antigenic variation in Giardia lamblia. Annu Rev Microbiol 65: 611-630.
14. Rendtorff RC, Holt CJ (1954) The experimental transmission of human intestinal protozoan parasites. IV. Attempts to transmit Endamoeba coli and Giardia lamblia cysts by water. Am J Hyg 60: 327-338.
15. Buret AG, Cotton J (2011) Pathophysiological processes and clinical manifestations of Giardiasis. In *Giardia* A model organism. 1st edition. Edited by Lujan HD, Svard SG. Springer Wien, New York pp. 300-310.
16. Buret AG (2005) Immunopathology of giardiasis: the role of lymphocytes in intestinal epithelial injury and malfunction. Mem Inst Oswaldo Cruz 100 Suppl 1: 185-190.
17. Said DE, Elsamad LM, Gohar YM (2012) Validity of silver, chitosan, and curcumin nanoparticles as anti-Giardia agents. Parasitol Res 111: 545-554.
18. Ankarklev J, Svard SG, Lebbad M (2012) Allelic sequence heterozygosity in single Giardia parasites. BMC Microbiol 12: 65.
19. Jerlstrom-Hultqvist J, Franzen O, Ankarklev J, Xu F, Nohynkova E, et al. (2010) Genome analysis and comparative genomics of a Giardia intestinalis assemblage E isolate. BMC Genomics 11: 543.
20. McCaffery JM, Faubert GM, Gillin FD (1994) Giardia lamblia: traffic of a trophozoite variant surface protein and a major cyst wall epitope during growth, encystation, and antigenic switching. Exp Parasitol 79: 236-249.
21. Lauwaet T, Davids BJ, Reiner DS, Gillin FD (2007) Encystation of Giardia lamblia: a model for other parasites. Curr Opin Microbiol 10: 554-559.
22. Palm D, Weiland M, McArthur AG, Winiecka-Krusnell J, Cipriano MJ, et al. (2005) Developmental changes in the adhesive disk during Giardia differentiation. Mol Biochem Parasitol 141: 199-207.
23. Marti M, Hehl AB (2003) Encystation-specific vesicles in Giardia: a primordial Golgi or just another secretory compartment? Trends Parasitol 19: 440-446.
24. Sun CH, McCaffery JM, Reiner DS, Gillin FD (2003) Mining the Giardia lamblia genome for new cyst wall proteins. J Biol Chem 278: 21701-21708.
25. Jarroll EL, Macechko PT, Steimle PA, Bulik D, Karr CD, et al. (2001) Regulation of carbohydrate metabolism during Giardia encystment. J Eukaryot Microbiol 48: 22-26.
26. Abdul-Wahid A, Faubert GM (2004) Similarity in cyst wall protein (CWP) trafficking between encysting Giardia duodenalis trophozoites and CWP-expressing human embryonic kidney-293 cells. Biochem Biophys Res Commun 324: 1069-1080.
27. Gottig N, Elias EV, Quiroga R, Nores MJ, Solari AJ, et al. (2006) Active and passive mechanisms drive secretory granule biogenesis during differentiation of the intestinal parasite Giardia lamblia. J Biol Chem 281: 18156-18166.
28. Jenikova G, Hruz P, Andersson MK, Tejman-Yarden N, Ferreira PC, et al. (2011) Alpha1-giardin based live heterologous vaccine protects against Giardia lamblia infection in a murine model. Vaccine 29: 9529-9537.
29. Abdul-Wahid A, Faubert G (2007) Mucosal delivery of a transmission-blocking DNA vaccine encoding Giardia lamblia CWP2 by Salmonella typhimurium bactofection vehicle. Vaccine 25: 8372-8383.
30. Larocque R, Nakagaki K, Lee P, Abdul-Wahid A, Faubert GM (2003) Oral immunization of BALB/c mice with Giardia duodenalis recombinant cyst wall protein inhibits shedding of cysts. Infect Immun 71: 5662-5669.
31. Takishita K, Kolisko M, Komatsuzaki H, Yabuki A, Inagaki Y, et al. (2012) Multigene phylogenies of diverse Carpediemonas-like organisms identify the closest relatives of 'amitochondriate' diplomonads and retortamonads. Protist 163: 344-355.
32. Emelyanov VV, Goldberg AV (2011) Fermentation enzymes of Giardia intestinalis, pyruvate:ferredoxin oxidoreductase and hydrogenase, do not localize to its mitosomes. Microbiology 157: 1602-1611.

33. Jedelsky PL, Dolezal P, Rada P, Pyrih J, Smid O, et al. (2011) The minimal proteome in the reduced mitochondrion of the parasitic protist Giardia intestinalis. PLoS One 6: e17285.
34. Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N (2009) Importing mitochondrial proteins: machineries and mechanisms. Cell 138: 628-644.
35. Morf L, Spycher C, Rehrauer H, Fournier CA, Morrison HG, et al. (2010) The transcriptional response to encystation stimuli in Giardia lamblia is restricted to a small set of genes. Eukaryot Cell 9: 1566-1576.
36. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, et al. (2007) Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 317: 1921-1926.
37. Lopez AB, Sener K, Jarroll EL, van Keulen H (2003) Transcription regulation is demonstrated for five key enzymes in Giardia intestinalis cyst wall polysaccharide biosynthesis. Mol Biochem Parasitol 128: 51-57.
38. Henze K, Horner DS, Suguri S, Moore DV, Sanchez LB, et al. (2001) Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes Giardia intestinalis, Spironucleus barkhanus and Trichomonas vaginalis. Gene 281: 123-131.
39. Pietsch F, (2010) Characterization of assemblage specific genes in *Giardia intestinalis*. Degree project in biology, Supervisor Svard SG, DiVA, IBG Uppsala University 10-13
40. Vera A, Gonzalez-Montalban N, Aris A, Villaverde A (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng 96: 1101-1106.

8. APPENDIX

- Title page picture is taken from localization result (Figure4), in which WB labelled with anti P15_874 antibody.

2X loading dye preparation (10 ml)

Component	Stock Volume
Tris-HCl Ph6.8	1ml
10 % SDS	4 ml
Glycerol	3.75 ml
BPB	250 µl
DDT 1M	1 ml

Sonication Buffer preparation (10 ml)

DDT	250 µl
Protease inhibitor tablet	1
PBS	9.75ml

10% and 15% separating gel Preparation (10 ml)

Component	Stock volume (10%)	Stock volume (15%)
H2O	4 ml	3.3 ml
30% Acrylamide solution (Bio-Rad)	3.3 ml	4ml
1.5 M Tris (PH 8.8)	2.5 ml	2.5 ml
10% SDS	100 µl	100 µl
10% Ammonium persulphate	100 µl	100 µl
TEMED	4 µl	4 µl

5% Stacking gel Preparation (5 ml)

Component	Stock volume
H2O	3.4 ml
30% Acrylamide solution (Bio-Rad)	830
1.5 M Tris (PH 6.8)	630
10% SDS	50 µl
10% Ammonium persulphate	50 µl
TEMED	5 µl

Coomassie blue staining

	Pre-fix Solution (100ml)	Stain solution(200ml)	Destain Solution(200ml)
Methanol	50 ml	100 ml	10 ml
Acetic acid	10 ml	20 ml	15 ml
H2O	40 ml	80 ml	175 ml
Coomassie R-250	-	0.05 g	-

Blocking Solution

Component	Stock
5% Nonfat dry milk	2.5g
PBS	50 ml

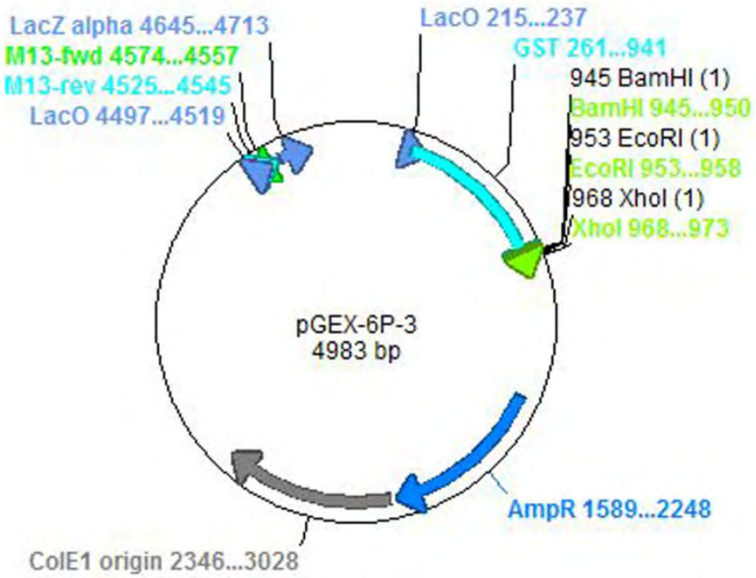


Figure 7: Schematic map of the pGEX-6P-3 vector used for transfections. The gene of interest was cloned between BamHI and Xhol sites.

Table 5. Complete list of Conserved genes between *Giardia* and *Spironucleus salmonicida*

Giardia assemblages						*Spironucleus* Salmonicida			
ISOLATE	**GENE ID**	**ANNOTATION**	**Organism with best hit in genbank**	**E-value**	**Domains or signatures**	**SSK-GENE ID**	**SSK ANNOTATION**	**SSK E-Value**	**SSK Score**
P15	GLP15_4285	Hypothetical protein				No hits found			
P15	GLP15_3399	Hypothetical protein				SS50380_12083	Hypothetical Protein	0.44	29.3
P15	GLP15_4274	Hypothetical protein				SS50380_17313	Hypothetical Protein	0.10	32
P15	GLP15_174	Hypothetical protein				SS50380_11136	Hypothetical Protein	1.40	28.1
P15	GLP15_167	Hypothetical protein				SS50380_11414	NAD-dependent histone deacetylase Sir2	0.37	29.6
P15	GLP15_965	Hypothetical protein				SS50380_16084	Conserved hypothetical protein containing transmembrane domain	0.45	28.5
P15	GLP15_967	Hypothetical protein				SS50380_14361	Tubulin tyrosine ligase	0.73	28.9
P15	GLP15_907	Hypothetical protein				SS50380_13144	Cysteine-rich protein	0.48	29.6
P15	GLP15_906	Hypothetical protein				No hits found			
P15	GLP15_1245	Hypothetical protein				No hits found			
P15	GLP15_2838	Hypothetical protein				No hits found			
P15	GLP15_3432	Hypothetical protein				No hits found			
P15	GLP15_3882	Hypothetical protein				SS50380_12047	Hypothetical Protein	1.20	28.1
P15	GLP15_2340	Hypothetical protein			Two Zinc fingers, C2H2-like	SS50380_15593	Dynein heavy chain	3.50	27.3
P15	GLP15_1876	Hypothetical protein				SS50380_15946	Protein containing Glycosyl hydrolase family 20, catalytic domain	0.093	32
P15	GLP15_4083	Hypothetical protein				No hits found			

P15	GLP15_874	Acetyltransferase	Anaerostipes caccae DSM 14662	5.00E-55	Acetyltransferase GNAT-superfamily (Acyl-CoA N-acyltransferases (Nat))	SS50380_17711	Hypothetical Protein	0.57	28.1
P15	GLP15_2691	Hypothetical protein				SS50380_10368	Hypothetical Protein	0.40	29.3
P15	GLP15_2819	Hypothetical protein				No hits found			
P15	GLP15_3604	Hypothetical protein				No hits found			
P15	GLP15_1674	Hypothetical protein				SS50380_15058	Conserved hypothetical protein	1.4	28.1
P15	GLP15_633	Hypothetical protein				SS50380_15913	H-SHIPPO 1	1.2	26.6
P15	GLP15_1891	Hypothetical protein				SS50380_17031	Protein containing Myb-like DNA-binding domain	0.42	30.0
P15	GLP15_1962	Hypothetical protein				No hits found			
P15	GLP15_2100	Hypothetical protein				No hits found			
P15	GLP15_2463	Hypothetical protein				SS50380_15593	Dynein heavy chain	0.46	30.4
P15	GLP15_1621	Hypothetical protein				SS50380_17047	Hypothetical Protein	1.8	25.4
P15	GLP15_1464	Hypothetical protein				No hits found			
P15	GLP15_2480	Hypothetical protein			Putative signalpeptide	SS50380_11284	Conserved hypothetical protein	1.3	28.5
P15	GLP15_2612	Hypothetical protein				SS50380_12363	Hypothetical Protein	5.2	25.8
P15	GLP15_2613	Hypothetical protein				SS50380_10708	Hypothetical Protein	0.72	28.5
P15	GLP15_2886	Hypothetical protein				No hits found			
P15	GLP15_2887	Hypothetical protein				SS50380_10607	Kinesin-9	0.11	32
P15	GLP15_2888	Hypothetical protein				SS50380_15593	Dynein heavy chain	0.46	30.4

P15	GLP15_3988	Hypothetical protein				No hits found			
P15	GLP15_4513	Hypothetical protein			Putative signalpeptide	No hits found			
P15	GLP15_4892	Hypothetical protein				SS50380_18459	Protein containing Meckelin domain	9.00	24.6
P15	GLP15_4996	Hypothetical protein			Reverse transcriptase (RT) catalytic domaim	SS50380_13706	Reverse transcriptase, putative	0.0004	38.5
GS	GL50581_5657	Hypothetical protein				SS50380_17575	Hypothetical protein	0.034	33.5
GS	GL50581_5676	Hypothetical protein			SMC domain	SS_JH_010	Conserved hypothetical protein	2.40	29.3
GS	GL50581_3637	Hypothetical protein				SS50380_13188	Conserved hypothetical protein	6E-11	65.9
GS	GL50581_4508	Hypothetical protein	Entamoeba dispar SAW760	4.00E-29	Methyltransferase TRM13 superfamily domain	SS50380_13668	Methyltransferase TRM13 family protein	3E-26	114
GS	GL50581_2613	Hypothetical protein	Desulfatibacillum alkenivorans	2.00E-68	Lactamase B superfamily protein	SS50380_17701	Ribonuclease BN	2.7	27.3
GS	GL50581_9818	Hypothetical protein			SMC prokaryote B domain	SS50380_10140	Histidine phosphatase superfamily protein	0.072	33.1
GS	GL50581_7	Hypothetical protein				No hits found			
GS	GL50581_62	Hypothetical protein				SS50380_13259	Hypothetical protein	9	25.4
GS	GL50581_9876	Hypothetical protein				SS50380_14564	Protein containing SMC domain Length=2145	0.084	32
GS	GL50581_9883	Hypothetical protein				SS50380_16954	Hypothetical protein	0.022	33.9
GS	GL50581_100	Hypothetical protein				No hits found			
GS	GL50581_10211	Hypothetical protein				SS50380_10251	Hypothetical protein	1.3	28.1
GS	GL50581_10581	Hypothetical protein				SS50380_14016	Conserved hypothetical protein Length=1944	1.1	28.5
GS	GL50581_1632	Rep protein, putative	*Giardia* intestinalis BRIS/92/HEPU/1541	5.00E-72	Rep protein, putative	SS50380_15122	Hypothetical protein	2.6	27.7
GS	GL50581_1633	Hypothetical protein				No hits found			

GS	GL50581_2037	Hypothetical protein	Clostridium perfringens E str. JGS1987	1.00E-27	Conserved hypothetical protein (T_den_put_tspse; putative transposase or invertase). Possible signalpeptide.	SS50380_17054	Hypothetical protein	0.74	29.6
GS	GL50581_2038	Hypothetical protein			Zinc Ring-finger domain	SS50380_14237	Hypothetical protein	0.1	31.6
GS	GL50581_11542	Hypothetical protein				No hits found			
GS	GL50581_11586	Hypothetical protein	*Giardia intestinalis* BRIS/92/HEPU/1541	4.00E-18	Replication-associated protein REP2, putative	SS50380_16933	ABC transporter	1.5	27.7
GS	GL50581_11615	Hypothetical protein			SMC_N domain	SS50380_17610	Conserved hypothetical protein	0.13	31.6
GS	GL50581_11623	Hypothetical protein				SS50380_12281	Hypothetical protein	1.6	27.3
GS	GL50581_3038	Hypothetical protein			Uncharacterized conserved protein, contains RING Zn-finger (COG5219)	SS50380_12274	Protein containing Zinc finger, C3HC4 type (RING finger)	0.0002	40.4
Relication GS	GL50581_3039	associated protein REP1,	*Giardia intestinalis* BRIS/92/HEPU/1541	3.00E-69	Rep protein, putative	SS50380_10904	Nucleolar GTP-binding protein 2	1.8	28.5
GS	GL50581_3192	Hypothetical protein	Bryantella formatexigens DSM 14469	1.00E-92	4-carboxymuconolactone decarboxylase	SS50380_15521	Conserved hypothetical protein	2.9	26.9
GS	GL50581_3319	Hypothetical protein			Zinc finger, RING-type	SS50380_15986	Protein containing %09CHY zinc finger domain	0.002	37.4
GS	GL50581_3333	Replicase-associated protein, putative	Faba bean necrotic yellows virus	7.00E-19	Viral replication N-terminal domain	SS50380_17719	Dynein heavy chain	0.57	30
GS	GL50581_3339	Hypothetical protein				SS50380_18679	Conserved hypothetical protein	7.1	25.8
GS	GL50581_3340	Rep protein, putative	*Giardia intestinalis* BRIS/92/HEPU/1541	4.00E-68	Viral Rep protein N-terminal domain, Helicase domain	SS50380_12551	26S proteasome ATPase subunit S4, putative	0.037	33.9

GS	GL50581_2039	Hypothetical protein				SS50380_15986	Protein containing %09CHY zinc finger domain	0.0004	38.5
GS	GL50581_4567	Hypothetical protein			SMC domain	SS50380_11928	Hexose transporter	0.057	32
GS	GL50581_3321	Hypothetical protein	Mannheimia succiniciproducens MBEL55E	2.00E-45	Conserved hypothetical protein	SS50380_16133	Hypothetical protein	0.24	29.3
WB	GL50803_137678	Hypothetical protein				No hits found			
WB	GL50803_4430	Hypothetical protein			Zinc finger, C3HC4 RING-type	SS50380_16229	Protein containing Zinc finger, C3HC4 type (RING finger) domain	0.011	35
WB	GL50803_21184	Hypothetical protein				No hits found			
WB	GL_UO_447 (CH991779:1123452-1124147)	Hypothetical protein			Overlaps with tRNA-Ile	No hits found			
WB	GL50803_101423	Hypothetical protein				SS50380_12195	Conserved hypothetical protein	0.24	30.8