Investigation of genes that may contribute to disease tropism in *Leishmania* species

Walide I. F. Saad

Wellcome Centre for Molecular Parasitology
Glasgow Biomedical Research Centre
University of Glasgow
Glasgow
G12 8TA

This thesis is presented in submission for the degree of Master of Science (MSc) in the Faculty of Veterinary Medicine.

Submitted September 2007

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

SUMMARY

Leishmania parasites cause a wide spectrum of diseases known collectively as the leishmaniases. The three main forms of the disease are cutaneous (CL), mucocutaneous (MCL) and visceral (VL) leishmaniasis and particular species are usually associated with particular disease tropism. A better understanding of the mechanisms of disease would be beneficial in identifying potential drug targets and vaccine candidates. Genome projects for three species have now been completed: *L. major* (CL), *L. infantum* (VL) and *L. braziliensis* (MCL) and analysis of the highly syntenic genomes has revealed a small subset of species-specific genes that are hypothesised to contribute in some way to the tropism of that species. Four *L. infantum*-specific genes were investigated in this project (LinJ28.0330, LinJ36.2750, LinJ20.1200 and LinJ36.2050). All four genes, which are pseudogenes (or absent) in *L. major* and *L. braziliensis*, were found to be intact in *L. donovani*, another species that can cause VL. All genes except LinJ20.1200 are expressed in all three main life cycle stages, procyclic promastigote, metacyclic promastigote and amastigote. LinJ20.1200 appears to be promastigote-specific. LinJ36.2050 encodes a protein that is a putative orthologue of yeast *SEC14*. A LinJ36.2050 null mutant (Δ*SEC14*) was generated. No defect in growth or *in vitro* infectivity of macrophages phenotype was observed with the Δ*SEC14* cells. They were unable to establish cutaneous infection in mice and at the time of writing, it is not yet known if Δ*SEC14* clones can be recovered from the spleens of hamsters at 4 months post-infection. Further detailed analysis is required to determine if the Δ*SEC14* cells have a phenotype at the cellular level.

DECLARATION

I declare that the work presented in this thesis is my own work except where otherwise stated.

Walide Saad

September 2007

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof Jeremy Mottram who gave me the opportunity to do this MSc at the University of Glasgow. I would also like to thank Dr Jim Hilley for his support and advice. Thank you also to the people in the Mottram Lab who helped me throughout my time in Glasgow.

I have to say a big thank you to my parents and my wife for their support, encouragement throughout my study since we arrived in Scotland.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

1.1 *Leishmania* and Leishmaniasis

Leishmaniasis represents a complex system of diseases with clinical and epidemiological diversity. It is caused by the kinetoplastid parasite *Leishmania*, and is endemic in 88 countries (22 in the "New World" (the Americas) and 66 in the "Old World" (Africa, Asia and southern Europe)). 16 of these are developed countries, 72 are developing countries (13 of them are among the least developed) (Desjeux, 1996). There are thirty known species of *Leishmania* and the majority are of them are human-infective (Banuls et al., 2007). They are grouped into either the *Leishmania* or *Viannia* subgenus, the latter being solely "New World". The *Leishmania* subgenus is generally "Old World" with the exception of *Leishmania (L.) mexicana* complex (Table 1.1). *Leishmania* species can also be grouped according to the general disease forms associated with that species (see Section 1.1.2).

1.1.1 *Leishmania* Life Cycle

Leishmaniasis is transmitted by blood-feeding female sand flies, as extracellular, flagellated promastigotes that differentiate into biologically distinct metacyclic forms prior to inoculation into the mammalian host. Transformation into proliferating intracellular, aflagellated amastigotes then occurs within the phagolysosome of host mononuclear phagocytes, which can facilitate the dissemination of parasites to different tissue site (Desjeux, 1996).

Parasites of the *Leishmania* species are obligate intra macrophage protozoan parasites with two life-cycle stages. Promastigotes are flagellated forms found in sandflies, and amastigotes are non-flagellated forms, which replicate in macrophage phagosomes in mammalian hosts (Lipoldova and Demant, 2006) (Figure 1.1).

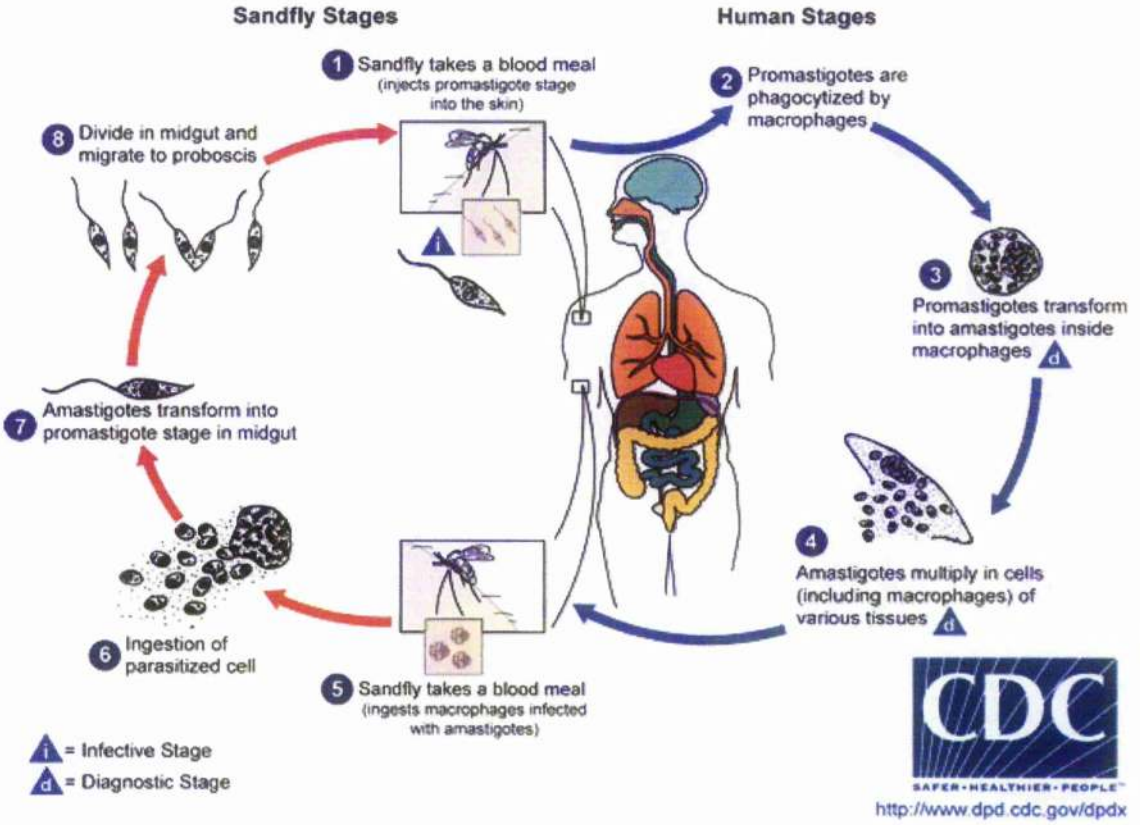


Figure 1.1: The life cycle of *Leishmania* (picture taken from the Centre for Disease Control, USA website).

1.1.2 Species Description and Disease Forms

Depending on which species is initiating the infection and on the immunological status of the host, *Leishmania* causes a wide spectrum of diseases. Clinical forms of leishmaniasis range from cutaneous (CL) and mucocutaneous (MCL) to diffuse cutaneous (DCL) and visceral (VL) and these are described in more detail below.

Three species have been identified as causing the visceral form of the disease and several other distinct *Leishmania* species cause the cutaneous and mucocutaneous forms (McMahon-Pratt and Alexander, 2004). More than 90% of the VL cases in the world are reported in Bangladesh, Brazil, India and Sudan and more than 90% of the CL cases occur in Afghanistan, Iran, Saudi Arabia and Syrian Arab Republic in the Old World and Brazil and Peru in the New World. This places an estimated 350 million people at risk (Desjeux, 1996).

Table 1.1: Different *Leishmania* species and associated diseases

Disease form	New World Species	Old World Species
Cutaneous	***L. (L.) mexicana* complex** *L. (L.) mexicana* *L. (L.) amazonensis* *L. (L.) pifanoi* *L. (L.) venezuelensis* ***L. (Viannia)* subgenus** *L. (V.) braziliensis* *L. (V.) peruviana* *L. (V.) lansoni* *L. (V.) naiff* *L. (V.) panamensis* *L. (V.) guyanensis*	***L (L) major* complex** *L. (L.) major* *L. (L.) tropica* *L. (L.) aethiopica*
Mucocutaneous	*L. (V.) braziliensis*	
Diffuse cutaneous	*L. (L.) amazonensis* *L. (L.) pifanoi*	*L. (L.) aethiopica*
Visceral	*L. (L.) chagasi**	***L. (L.) donovani* complex** *L. (L.) infantum* *L. (L.) donovani*

* - Note that *L. (L.) chagasi* is widely considered to be a synonym of *L. (L.) infantum* as no evidence of ancient lineage can be found (Mauricio et al., 2000). Rather it is held that the species has been introduced from dogs/rodents by European settlers.

1.1.3 Visceral Leishmaniasis

Also known as kala-azar, this form of the disease is caused by species of the *L. donovani* complex and is the most severe form. The disease is characterised by fever, enlargement of the spleen (splenomegaly) and liver (hepatomegaly), anaemia and severe weight loss. This form of the disease is usually fatal if left untreated. Serious epidemics of VL have occurred in India and East Africa in recent years, while the number of cases of VL is currently increasing in Brazil (Desjeux, 1996).

Figure 1.2: Sudanese woman with visceral Leishmaniasis (Taken from Herwaldt (1999)).

1.1.4 Cutaneous Leishmaniasis

Cutaneous Leishmaniasis in the Old World is due to *L. major* which produces self-healing lesions as the parasites remain at the sites of infection, the sand fly bite (Figure 1.3). The mechanisms involved in cutaneous disease caused by for example, *L. major* and *L. mexicana,* are limited. Cutaneous infection can remain sub-clinical or become clinically apparent after a variable incubation period that averages several weeks. Lesions can form at the site of infection and are usually self-healing (Herwaldt, 1999).

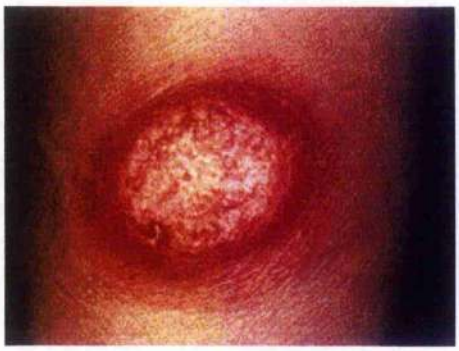


Figure 1.3: Patient suffering from cutaneous leishmaniasis (Taken from Herwaldt (1999)).

1.1.5 Mucocutaneous Leishmaniasis

Mucocutaneous Leishmaniasis is caused by parasites of the *Viannia* subgenus, for example *L. braziliensis*. Clinical symptoms can indicate extensive destruction of the oral-nasal and pharyngeal cavities resulting in severe disfigurement (Figure 1.4). Typically, mucosal disease becomes evident because of chronic nasal symptoms, in context of a hyperactive immune response. The disease is not normally self-limiting and requires treatment, however, diagnosis is difficult, even when clinically active. MCL is more difficult to treat than CL. Currently, the best treatment options are pentavalent antimony drugs (Herwaldt, 1999).

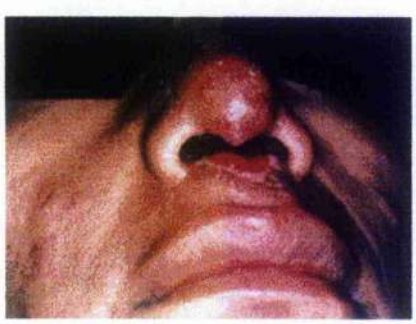


Figure 1.4: Mucocutaneous Leishmaniasis sufferer with perforation of the nasal septum. (Taken from Herwaldt (1999)).

1.1.6 Factors that May Give Rise to Different Disease Tropism

The different disease tropisms caused by different *Leishmania* species are thought to be determined by a number of factors (McMahon-Pratt and Alexander, 2004):

- Immune status of the host
- Genetic differences between *Leishmania* species

The extent to which these parasites are able to resist the host's innate and acquired immunity will directly contribute to their establishment in the host and consequently to disease severity (Campos-Ponce et al., 2005). For example, although species of the *L. donovani* complex generally cause visceral disease, there have been reports of cases of cutaneous infection by these species (McMahon-Pratt and Alexander, 2004; Siriwardana et al., 2007).

1.2 Modes of Infection

1.2.1 Host-Parasite Interaction

The macrophage is the immune-effector cell that, upon activation, is normally able to kill foreign pathogens. This cell, however, is the main mammalian host cell for all *Leishmania*.

1.2.1.1 Uptake by Macrophages (Phagocytosis)

Upon first exposure to the mammalian host during sand fly feeding, *Leishmania* promastigotes encounter host macrophages which can bind to (via *Leishmania* surface molecules – see Section 1.2.1.2) and internalise the parasite cells in the same way other pathogens are taken up. *Leishmania* spp are remarkably resistant to macrophages and can survive, differentiate and proliferate in the parasitophorous vacuole of the macrophage. Other host cells can take up *Leishmania* promastigotes,

also (see Section 1.2.1.4), but the macrophage is the cell type in which the parasites can differentiate and proliferate within the mammalian host.

1.2.1.2 Surface Molecules & Parasite Interaction with Host Cells

There are two major families of promastigote surface molecules: -

- GP63:- The major surface metallopeptidase, found in large numbers on the promastigote surface, but not the amastigote surface
- Phosphoglycans: - Includes glycolipids such as lipophosphoglycan (LPG), phosphoglycosylated proteins and proteophosphoglycans (Handman and Bullen, 2002).

These have been proposed to be required for binding to macrophages, the initial step for establishment of infection (Kelleher et al., 1995). They have also been implicated in the resistance of promastigotes to complement-mediated lysis prior to uptake by macrophages (Ilgoutz and McConville, 2001).

1.2.1.3 Amastigote Invasion of Macrophages

L. major LPG has been found to be involved in initial survival of promastigotes in the host prior to and following cell invasion by promastigotes but are less important to amastigotes, where LPGs are strongly downregulated (Spath et al., 2003; Ilgoutz and McConville, 2001). It is believed that infected macrophages in mammalian hosts eventually rupture, releasing amastigotes which then go on to infect more macrophages, however, little is known about the mechanism and molecules involved in amastigote invasion and survival in these cells (Handman and Bullen, 2002).

1.2.1.4 Uptake by Other Cells (Trojan Horses)

As already discussed the mammalian stage amastigotes exist intracellularly, and the preferred host cell is the macrophage. *Leishmania* promastigotes are quickly killed in the extracellular tissue environment and can only survive if they enter phagocytic cells. Often, polymorphonuclear neutrophil (PMN) granulocytes, rather than macrophages are the first cells *Leishmania* will encounter (van Zandbergen et al., 2004). *L. major* has been shown to be capable of entering and surviving as promastigotes within PMN. This causes apoptosis of the PMN, which is then taken up by macrophages, together with the surviving *L. major*. Thus, the parasite appears to be using the PMN as an intermediate host cell for up to 42 hours post-infection, prior to uptake of infected PMN by macrophages (van Zandbergen et al., 2004).

1.3 Genomics

1.3.1 Genome Projects

The Wellcome Trust Sanger Institute has sequenced the *L. major* genome to completion and more than 8000 genes have been annotated in the ~33.6Mb genome (Ivens et al., 2005). The genomes of *L. infantum* and *L. braziliensis* were the next genome projects to be completed (Peacock et al., 2007) and *L. mexicana* will be the fourth, although this project is in the early stages at the time of writing (http://www.sanger.ac.uk/sequencing/Leishmania/mexicana/).

The number of chromosomes and genes in each species is summarised in Table 1.2. *L. braziliensis* has 35 chromosomes compared with 36 in both *L. major* and *L. infantum* and this is due to the apparent fusion of chromosomes 20 and 34 in *L. braziliensis* (Peacock et al., 2007).

Table 1.2: Genomes of three *Leishmania* species in summary

***Leishmania* species**	***L. infantum***	***L. major***	***L. braziliensis***
Number of chromosomes	36	36	35
Number of genes	8154	8298	8153

1.3.1.1 Comparative Gene Content Analysis of *Leishmania*

The organisation of the genes in all three species of *Leishmania* sequenced to date has shown to have a high degree of synteny i.e. the positions of genes relative to one another is highly conserved in all species studied (Peacock et al., 2007). An example of this synteny is shown in Figure 1.5.

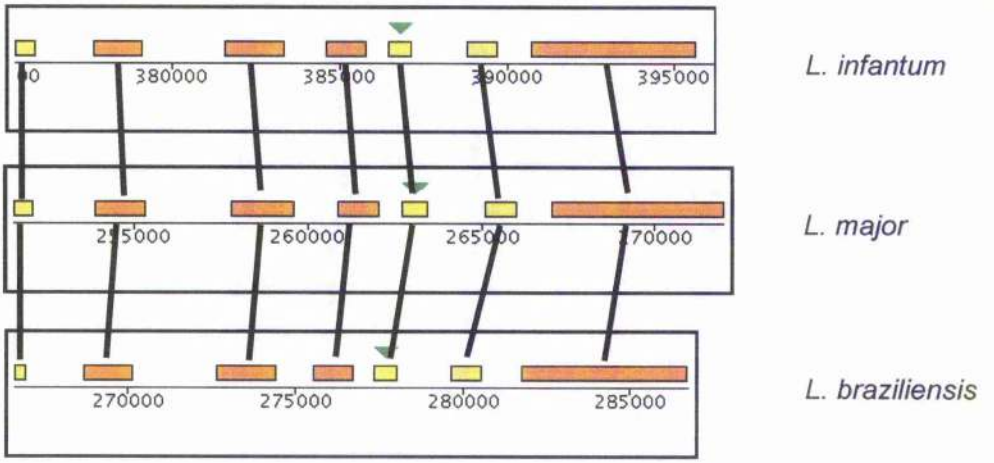


Figure 1.5: An example of the high degree of synteny between the three species of *Leishmania* with genomes sequenced to date. The example shown is the genomic locus of a putative RNA Binding Protein (LinJ32.1150) (green arrow) and the genes surrounding it.

This has made it possible to directly compare the genes found in each species and identifies those that are specific to each species. Peacock et al (2007) have reported that the number of species-specific genes is remarkably small (Figure 1.6).

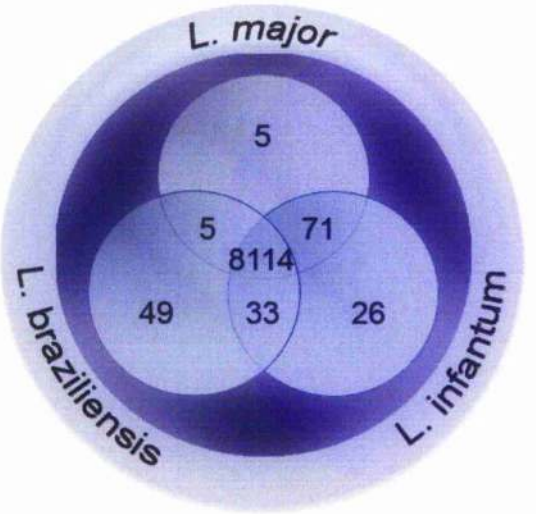


Figure 1.6: Venn Diagram showing the number of genes that are unique to each species as well as those genes that are common to two or more species. The majority of these genes are hypothetical.

Despite the vastly different disease forms, a surprisingly small number of genes have been found to be unique to each species: 26 *L. infantum*-specific genes, 5 *L. major*-specific genes, and 49 *L. braziliensis*-specific genes (Figure 1.6). The hypothesis underlying the program of study, of which this project is a part, is that at least some of these genes will contribute to the molecular mechanisms giving rise to the disease tropism of a particular species.

1.3.2 Genes Differentially Distributed Between Three *Leishmania* Species

To date the multi-copy A2 gene (LinJ22.0680) is the only gene shown to be involved in *L. donovani* visceral disease tropism (Zhang and Matlashewski, 1997; Zhang and Matlashewski, 2001; Zhang et al., 2003). The gene is not expressed in *L. major* and detailed characterisation of the *L. donovani* and *L. major* A2 loci revealed that the multiple amino acid repeat sequences found in the *L. donovani* protein are absent from *L. major*; the gene is truncated containing only a single amino acid repeat region. The A2 gene has been shown to be necessary for survival of *L. donovani* in visceral organs and loss of several copies of the A2 array results in compromised BALB/c mouse infections (Zhang and Matlashewski, 2001). Interestingly, heterologous expression of the *L. donovani* A2 gene in *L. major* prevented the cutaneous infection being established in both resistant and susceptible mice. This indicates that disease tropism can be influenced by just a single gene. The full mechanism of visceralisation will almost certainly require additional genes. The genes that are unique to *L. infantum* are listed in Table 1.3.

Table 1.3: The unique genes of *L. infantum* (Peacock et al., 2007). The genes that will be examined in this study are highlighted in red.

Product	*L. infantum*	Product	*L. infantum*
Hypothetical protein	LinJ02.0670	Multidrug resistance protein	LinJ30.1840
Hypothetical protein	LinJ08.0140	n-acyl-l-amino acid amidohydrolase	LinJ31.1490
Tuzin	LinJ08.0750	Hypothetical protein	LinJ31.2630
Hypothetical protein	LinJ10.1430	p-nitrophenyl-phosphatase	LinJ31.3030
Phosphatidylinositol 3-kinase	LinJ14.0020	Hypothetical protein	LinJ31.3160
Hypothetical protein	LinJ15.0890	Hypothetical protein	LinJ32.1900
Hypothetical protein	LinJ20.1210	Hypothetical protein	LinJ33.2710
Methylenetrahydrofolate dehydrogenase	LinJ22.0330	Hypothetical protein	LinJ34.3170
Hypothetical protein	LinJ22.0410	**Phosphatidylinositol/ phosphatidylcholine SEC14 cytosolic factor**	**LinJ36.2050**
Hypothetical protein	LinJ22.0680	Hypothetical protein	LinJ36.2060
Hypothetical protein	LinJ24.1430	Glyceraldehyde 3-phosphate	LinJ36.4900
Glutathionyl-spermidine synthase	LinJ25.2500	**Hypothetical protein**	**LinJ36.2750**
Hypothetical protein	**LinJ28.0330**	**Endo-1,4-β-Xylanase**	**LinJ20.1200**

1.3.3. *L. infantum*-Specific Genes of Interest

There are four specific genes of interest that were analysed during this project, they are:-

- LinJ28.0330 – encodes a hypothetical protein
- LinJ36.2750 – encodes a hypothetical protein
- LinJ20.1200 – encodes a putative endo-1,4-β-xylanase precursor protein
- LinJ36.2050 – encodes putative SEC14 cytosolic factor

These four genes were chosen from the 27 unique genes in *L. infantum* for further study in this project. The two hypothetical genes were selected at random (although LinJ36.2750 was chosen because it has a homologous domain (TRF4) that may provide a handle for further study (see Section 1.3.3.2). LinJ36.2050 was selected as it is potentially involved in trafficking, a process that could influence host-parasite interactions and therefore potentially, disease tropism (see Section 1.3.3.4). LinJ20.1200 was selected as a poor candidate for involvement in disease tropism that could quickly be eliminated from the overall study (see Section 1.3.3.3).

1.3.3.1. LinJ28.0330

This gene encodes a hypothetical protein and according to GeneDB it is intact in *L. infantum*, whereas the gene in the equivalent position in both *L. major* and *L. braziliensis* contains frameshifts and/or introduced stop codons (Figure 1.7). Presumably, the latter two species, by contrast with *L. infantum*, have lost a biologically relevant function for this protein, if indeed it is expressed in *L. infantum*. There are no recognisable domains in the primary amino acid sequence to yield any clues to the potential role of this protein.

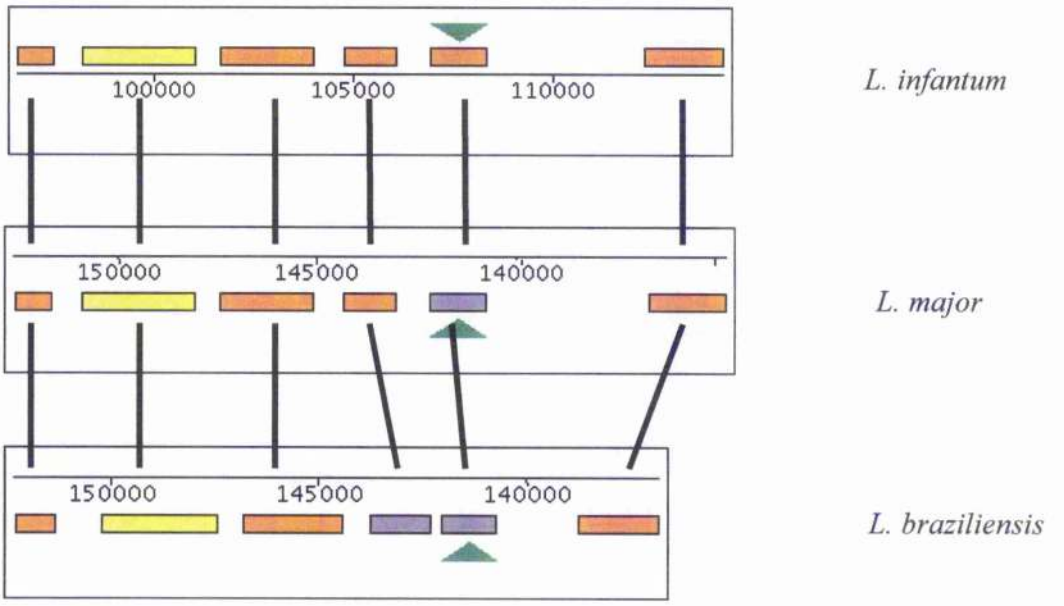


Figure 1.7: LinJ28.0330 Locus of chromosome 28 (in all three species) showing the pseudogene orthologues (grey) in both *L. major* and *L. braziliensis*. This also highlights the high degree of synteny of each species.

1.3.3.2. LinJ36.2750

This gene also encodes a hypothetical protein and is present in *L. infantum* but is apparently missing from both *L. major* and *L. braziliensis* (Figure 1.8). Similarly to LinJ28.0330, if LinJ36.2750 is expressed in *L. infantum*, then the other species are likely to have lost a biological requirement for the gene and it has undergone mutation such that the gene is no longer intact.

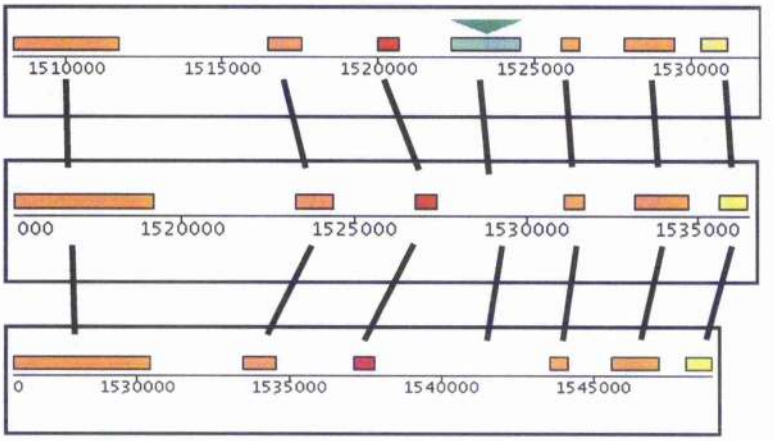


Figure 1.8: LinJ36.2750 locus on chromosome 36 in *L.* infantum and L. *major* and chromosome 35 of *L.* braziliensis. This schematic demonstrates that the gene is apparently missing from both *L. major* and *L. braziliensis*.

On closer inspection, remnants of the gene can be identified in the gap region where the gene ought to be in *L. major*. There are also remnants of the gene in *L. braziliensis* but these are even more divergent than the *L. major* sequence suggesting that the gene was lost long ago in evolutionary terms.

The encoded protein, although hypothetical, contains a domain that has homology to TRF4, a domain related to the TRF4 proteins of the yeast *Saccharomyces cerevisiae*. These form part of a nuclear poly adenylation complex that can mediate the polyadenylation and degradation of ribosomal RNA (rRNA) and/or small nucleolar RNA (snoRNA) precursors (LaCava et al., 2005). Reciprocal BLAST analysis with the *S. cerevisiae* TRF4 amino acid sequence (GeneDB Omniblast against predicted proteins in *L. infantum*, *L. major* and *L. braziliensis*) has revealed several genes encoding TRF4 domains. LinJ36.2750 was the third highest hit.

1.3.3.3. Endo-1,4-β-Xylanase

This gene encodes a protein with predicted function based on homology with other characterised proteins and is present in *L. infantum*. It appears to be a pseudogene in *L. major* and *L. braziliensis* (Figure 1.9). Xylanases are enzymes that degrade the polysaccharide xylan, a major component of some plant cell walls. Xylanases are also present in fungi for the degradation of plant matter as part of the plant pathogenesis mechanism or to assimilate usable nutrients (Brito et al., 2006). Some species of sand flies (e.g. *Phlebotomus papatasi*) are known to consume plant matter between blood meals for sustenance (Schlein and Jacobson, 2002) and one of the major components of the cell walls of one species of plant, *Proposis farcta,* consumed by these sand flies is xylan (Schlein and Jacobson, 1995; Pasiecznik et al., 2001). One might therefore predict that this *L. infantum*-specific enzyme will be preferentially expressed in the promastigote (sand fly) life-cycle stage, because it is during this stage that the cells are most likely to encounter plant material, in the midgut of the sandfly.

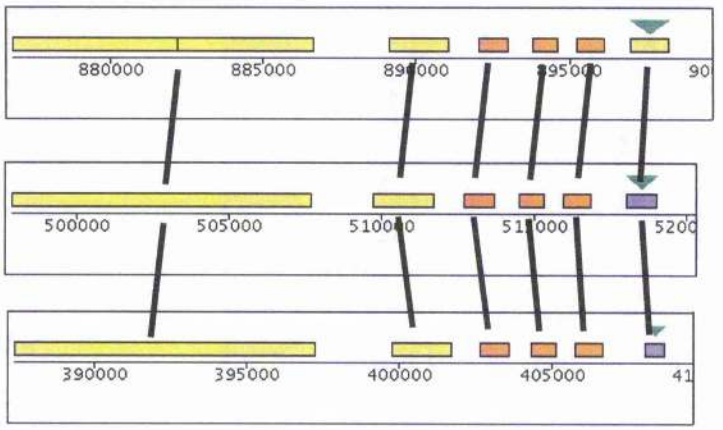


Figure 1.9: The genomic locus of chromosome 20 containing LinJ20.1200 (endo-1,4-β-xylanase precursor) in *L. infantum* compared with the locus in *L. major* and *L. braziliensis*. The gene has become a pseudogene in the latter two species but is intact in *L. infantum*.

1.3.3.4. LinJ36.2050 - SEC14 Cytosolic Factor

LinJ36.2050 has been annotated on GeneDB as encoding SEC14 cytosolic factor. This gene is not annotated in *L. major* and only remnants of the gene have been identified in *L. braziliensis* (Figure 1.10).

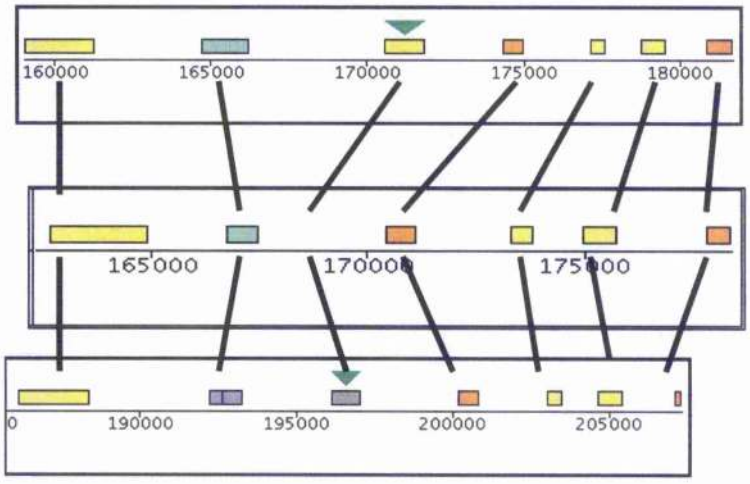


Figure 1.10: The genomic locus of the SEC14 gene in three species of *Leishmania*. The gene is apparently intact in *L. infantum* but is "missing" from *L. major* and remnants of the gene have been identified in *L. braziliensis*.

The annotation of the gene in GeneDB as SEC14 Cytosolic Factor indicates that it is a possible orthologue of the SEC14 gene of the budding yeast *Saccharomyces cerevisiae*. The encoded protein (Sec14p) is the main phosphatidylinositol transfer protein (PITP) in this organism and plays an essential role in protein transport from the trans-Golgi network (TGN) to the plasma membrane. This was originally demonstrated in yeast temperature sensitive mutants that exhibit a block in secretory vesicle formation at the TGN at non-permissive temperatures (Bankaitis et al., 1989). To date over 500 SEC14-like proteins all within eukaryotic species have been identified with roles in processes such as vesicle trafficking and phospholipid metabolism (Mousley et al., 2007; Phillips et al., 2006). At the most basic level of activity, these PITPs can extract a single phosphatidylinositol molecule (PI) from membrane and transfer it to another by an unknown mechanism (Figure 1.11). Yeast Sec14p can also extract and transfer phosphatidylcholine (PC) molecules (Hsuan and Cockcroft, 2001). Five other SEC14 homologues (Sfh1-5) have been identified in

yeast and most have retained their ability to transfer PI but not PC (Phillips et al., 2006).

The *L. infantum SEC14* gene encodes a protein with approximately 16% identity and 26% similarity to yeast Sec14p. If this protein is involved in the secretory pathway as it is in yeast, then it could potentially influence molecules that are either secreted or found on the surface of the parasite and so contribute to host-parasite interactions.

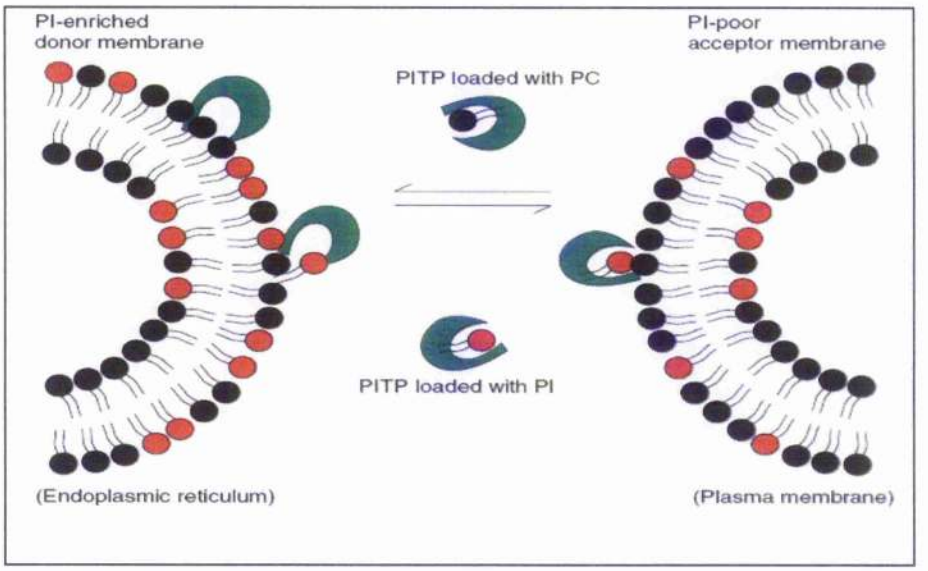


Figure 1.11: The phosphatidylinositol/phosphatidylcholine transfer activity of PITPs such as yeast Sec14p. Taken from (Hsuan and Cockcroft, 2001).

1.3.4 Project Aim

The aim of this project is to test the hypothesis that some of the *L. infantum*-specific genes contribute to the disease tropism of that species. Four of the 26 *L. infantum*-specific genes (Figure 1.6) were selected to study in more detail. These are highlighted in red in Table 1.3.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Molecular Methods

2.1.1 Bacterial Strains

Two strains of bacteria were used during this study as described in Table 2.1.

Table 2.1: Bacterial strains

Strain	Uses	Reference
XL1-Blue cells	Supercompetent cells for high-efficiency transformation	Stratagene
XL10-Gold	Ultracompetent cells for high-efficiency transformation	Stratagene

2.1.2 Bacterial Culture Methods

Approximately 10ml of Luria-Bertani (LB) broth containing appropriate antibiotics in a sterile glass test tube was inoculated with a single colony using a sterilised toothpick. Cultures were grown overnight at 37°C in a rotary incubator. Where medium-scale plasmid preparation was required, 1-2ml of an overnight culture (grown from a single colony) was used to inoculate 25-50ml of fresh LB broth.

2.1.3 Long Term Storage of Bacteria

Cells were plated on Luria-Bertani (LB)-agar plates with appropriate antibiotics and incubated overnight at 37°C. A single colony was used to inoculate 5ml of LB broth with antibiotics and cells were grown overnight at 37°C. 1.0ml of the overnight culture was mixed with 1.0ml 2% peptone, 40% glycerol and cells were stored at -80°C.

2.1.4 Preparation of Competent Bacteria Using $CaCL_2$

A single colony of XL1 *E. coli* strain was used to inoculate a 5ml LB plus appropriate antibiotic and shake overnight at 37°C. 25ml LB plus antibiotic was inoculated with overnight culture. Cells were centrifuged for 10 minutes at 4°C. Then pellet was suspended in 0.5 original volume cold, sterile 50 mM $CaCl_2$ (12.5ml) and resuspended gently by pipetting up and down a few times with a wide bore pipet. Cells were left on ice for 30 minutes and centrifuged as before at 4°C. The pellet was resuspended in 0.1 original volume of $CaCl_2$ and split into 50-200µl aliquots. 50µl was used for each transformation.

2.1.5 Plasmid Purification

Several different kits were used to purify plasmid depending on requirements:

- Qiagen Qiaprep Spin Miniprep Kit was used for small scale plasmid purification.
- Qiagen Hispeed Plasmid Midi Kit was used for larger-scale plasmid purification.

The quality and quantity of the purified DNA was assessed spectrophotometrically by recording absorbance at 260nm and 280nm wavelength.

DNA concentration was calculated as follows: A_{260} x 50 x dilution factor (generally 50 fold dilutions of mini-prep DNA so the concentration in ng/µl would be calculated by A_{260} x 50 x 50.

2.1.6 Polymerase Chain Reaction (PCR)

PCR was used to amplify DNA fragments for subsequent cloning and analysis or for the colony-screening of transformed bacteria. For general PCR analysis a 20µl reaction was set up with the non-proofreading Taq DNA Polymerase (NEB) as shown in the Table 2.2.

Table 2.2: Contents of a Standard PCR Reaction

Component	**Volume/ 20µl reaction**
10x PCR mix (See Section 2.4)	2µl
Primer A (0.1µg/ml)	2µl
Primer B (0.1µg/ml)	2µl
Template DNA	0.5µl
MilliQ H_2O	13µl
Taq DNA Poylmerase	0.5µl

Where high sequence fidelity was required, the polymerase was replaced with one of several proof-reading DNA polymerases available (Table 2.3) according to

manufacturer's instructions. Wherever possible, the reaction buffer supplied with the proof-reading enzyme was used instead of 10x PCR Buffer

Table 2.3: Proof-reading DNA Polymerases Used in this Study

Enzyme	Manufacturer
Pfu Turbo	Stratagene
Deep Vent	NEB
Phusion	NEB (Finnzymes)
Pfx50	Invitrogen

Products generated by high-fidelity PCR that were to be cloned into T-vectors such as pGEM-T Easy (Promega) or pCR2.1 TOPO (Invitrogen) required the addition of A-overhangs. This was accomplished by adding 0.5-1.0µl of Taq DNA Polymerase (NEB) directly to the PCR reactions at the end of the final cycle and incubating for a further 10-20 minutes at 72°C prior to electrophoresis and gel-extraction.

2.1.7 Restriction Digests

Restriction enzymes were obtained from New England Biolabs and used according to manufacturer's instructions with the buffers supplied/recommended. Generally, 1-5µg of plasmid DNA was digested in a single 20µl reaction for 1-2 hours at 37°C. Larger-scale digests were scaled up in volume and/or digest-time.

2.1.8 DNA Gel Electrophoresis

Gels were prepared by dissolving 0.8-1.2 %(w/v) agarose in 0.5 x TBE buffer. Prior to pouring the gels, 2µl of SYBR Safe DNA Gel Stain (Invitrogen) was added per 100ml of molten gel. DNA samples were mixed with 10 x DNA loading buffer and were electrophoresed at 60-170V until the dye in the loading buffer had migrated approximately two thirds of the length of the gel. Gels were viewed under UV illumination and images obtained by using a Bio-Rad gel documentation system. If required, DNA bands were cut from the gels for purification and subsequent cloning.

2.1.9 Purification of DNA from Agarose Gels

DNA cut from agarose gels was purified using the Qiaquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions.

2.1.10 Phosphatase Treatment of Digested Plasmids

Plasmids digested with a single restriction enzyme were treated with calf intestinal alkaline phosphatase or CIAP (NEB) prior to ligation. This removes the 5' terminal phosphate in order to reduce the chances of plasmid re-ligation, rather than ligation with the added DNA fragment of interest. This was done as follows:

Plasmid digests were incubated at 65°C for 20 minutes to heat-inactivate the restriction enzymes (where possible). 1µl of CIAP was added directly to the heat-inactivated restriction digest mix and incubated at 37°C for 30-60 minutes. The digests were run out on DNA agarose gels and appropriate bands were excised form the gel for subsequent use.

2.1.11 DNA Ligation

Ligation of DNA fragments with available T vectors such as pCR2.1 TOPO TA (Invitrogen) and pGEM-T or pGEM-T Easy (Promega) was carried out according to manufacturer's instructions. Other ligations were carried out by adding various volumes of plasmid and insert in a total volume of 10µl containing 1x T4 DNA ligase buffer and 0.5µl T4 DNA ligase (New England Biolabs). Reactions were either incubated at room temperature for 30-120 minutes or overnight at 16°C (4°C for pGEM-T cloning).

2.1.12 Transformation of Competent Bacteria

Chemically competent cells made from the XL1-blue strain by the $CaCl_2$ method were used for routine transformations. For lower efficiency ligations, commercially available XL1-Blue Supercompetent Cells and XL10-Gold Ultracompetent cells (Stratagene) were used. Cells were mixed with 5µl of ligation reaction and transformed by heat shock for 30 seconds (chemically-competent bacteria) or according to manufacturer's protocol for cells obtained from Stratagene.

2.1.13 DNA Sequencing

The DNA Sequencing Facility at the University of Dundee was used for all sequence analysis (http://www.dnaseq.co.uk). Plasmid concentration was assessed by A_{260}/A_{280} absorbance. 250-300ng of high-quality plasmid DNA in 15µl was required per sequencing reaction. "Standard" primers such as M13 Forward (-20), M13 Reverse, T7, T3 and SP6 were provided by the DNA Sequencing Facility. Gene-specific primers were prepared by diluting primers to 3.2pmoles/µl in water and

these were sent to the DNA Sequencing Facility along with the plasmids to be sequenced.

Sequences were analysed using Vector-NTI Advance 10 software package (Invitrogen).

2.2 *Leishmania infantum* Methods

2.2.1 Cell Lines

The following *Leishmania* cell lines were used in this study (Table 2.4).

Table 2.4: *Leishmania* Species and Strains

Species	Strain	Notes
L. infantum	JPCM5	Genome project strain
L. donovani	BPK206/0 (Clone 10)	Field Isolate from bone marrow of Nepalese VL patient, 2003. Provided by Lesley McCaig, University of Glasgow

2.2.2 *Leishmania* Tissue Culture

L. infantum and *L. donovani* promastigotes were grown at 25°C in HOMEM medium (Gibco BRL) supplemented with 10% heat-inactivated foetal calf serum (FCS) and 1% Penicillin-Streptomycin (Sigma). Cells were inoculated into fresh medium at approximately $5x10^5$ cells/ml and were sub-passaged into fresh medium when cultures reached late-log or early stationary phase.

2.2.2.1 *Leishmania* Cell Counts

Cells were diluted and fixed by adding 10µl of cells to 10µl of PBS/ 1% formaldehyde. 10µl of fixed cells were loaded into each chamber of an improved Neubauer haemocytometer and counted under microscopy.

2.2.2.2 Harvesting *Leishmania*

Cells in culture were generally harvested by centrifugation at 1200xg for 5 minutes followed by two washes in PBS (both followed by centrifugation as before). Cells were used immediately or frozen at -80°C. For RNA extraction, 1 x 10^9 cells were resuspended in 1ml of Trizol (Invitrogen) immediately following the first centrifugation (i.e. before any PBS washes).

2.2.2.3 Harvesting *L. infantum* Amastigotes for RNA Collection

To purify amastigotes from infected hamsters, the buffers and the sephadex were prepared and swelled overnight at 4°C before the purification.

Table 2.5: Component for purification (PSGEMKA)

Component	Volume
NaCl	15.75g
Glucose	15g
EDTA	0.285g
$MgCl2.6H_2O$	3.04g
KCl	1.11g
Albumin (Bovine)	0.30g
Buffer X (0.02M $Na_2HPO_4.2H2O$)	108ml
Buffer Y (0.02M $NaH_2PO_4.2H2O$)	34.5ml
H_2O	1357ml

2.5g of Sephadex was swelled overnight in 100ml of PSGEMKA buffer. Hamster spleen was removed using standard asceptic techniques. The spleen and was cut into

small fragments which were broken up in the tissue grinder with the gradual addition of 400mls of PSGEMKA until no solid pieces of lesion remained. 100µl of crude extract was removed for counting amastigotes. 50mg of Saponin was resuspended in 5-10 ml PSGEMKA and was mixed with the 400ml crude amastigote suspension. The suspension was split in to 8x50ml falcon tubes and centrifuged for 10 minutes at 3.400rpm in a swing bucket rotor (approximately 1500xg). Supernatant was removed with 25ml pipette and the pellets washed two more times and resuspended in a final volume of 100ml. 100µl of this suspension was collected for counting (pre-column). The amastigote purification column was prepared and PSGEMKA used to remove air bubbles before the Sephadex solution was poured in the column. The column was washed with 100ml PSGEMKA. Then the amastigote suspension was added to the column. The amastigotes flow through the column, while the tissue debris and host cells remain in the column. 50ml of PSGEMKA was added to flush amastigotes through and the flow through was collected. 100µl of the flow through was used to count cell numbers (post column). Purified amastigotes were centrifuged and washed twice with PBS and resuspended at 5 x 10^8 cells per ml of Trizol.

2.2.3 Long Term Storage of *L. infantum* Cell Lines

0.5ml of a log phase culture was diluted with an equal volume of fresh medium containing 10% DMSO. Samples were stored overnight at -20°C and stored overnight at -80°C and then transferred to liquid nitrogen for long-term storage.

2.2.4 Transfection of *L. infantum*

Cells were harvested and washed before being resuspended at $2x10^8$ cells/ml in cytomix (Reference Ngo etal 1998, Robinson and Beverley 2003) and kept on ice. $1x10^8$ cells in 500µl were transferred to a pre-chilled 0.4cm electroporation cuvette. Target DNA (either an episome or a linearised knockout cassette) or water (control transfection) was added to each cuvette and cells were electroporated with 2 pulses, 10 seconds apart (1.50kV and 25µF capacitance) in a Gene Pulser II Electroporation System (Bio-Rad). Cells were allowed to recover on ice for 10 minutes before being transferred to 10ml fresh medium and immediately split into 2 flasks (5ml each). Control transfection cultures were not split in this way but remained in 10ml of medium. Flasks were incubated overnight at 25°C overnight. Cells were then plated on 96 well plates by serial dilution in fresh medium containing the appropriate antibiotics (see Table 2.6).

To obtain clones, 4ml of each overnight culture was diluted into 20ml HOMEM/FCS + antibiotics (Dilution 1). 2ml of Dilution 1 was added to 22ml of HOMEM/FCS + antibiotics (Dilution 2) and 2ml of Dilution 2 was then added to 22ml of HOMEM/FCS + antibiotics (Dilution 3). 200µl of Dilution 1 was added to every well on a 96 well plate. The same was done for Dilutions 2 and 3, giving 3 plates for each overnight culture. To the remaining 1ml in each overnight flask, 9ml of fresh HOMEM/FCS + antibiotics was added – a transfected population. Cells growing up in this culture are unlikely to be clonal so further serial dilution is generally required. For the control (no DNA) transfections, 1ml of these cells were added to 9ml of fresh HOMEM/FCS + antibiotics. This dilution was necessary to prevent the cells reaching stationary phase before antibiotic selection was complete.

2.2.4.1 Antibiotics Used for Selection of *Leishmania* Transfectants.

The antibiotics used in this study are detailed in Table 2.6.

Table 2.6: Antibiotics Used for Selection of *Leishmania infantum*

Resistance Marker Gene	Antibiotic	Reference
SAT	Nourseothricin	Hans Knoll Inst., Germany
BLE	Phleomycin	Zeocin (Invitrogen)
NEO	G418 (analogue of neomycin)	Merck

2.3 DNA Manipulation

2.3.1 Isolation of Genomic DNA from *L.infantum*

Genomic DNA was generally isolated from 10ml of late log phase or stationary phase cultures using the DNEasy Blood and Tissue Kit (Qiagen) according to manufacturer's protocol for cultured animal cells. For larger scale DNA preparation, larger volumes or cells were used with multiple DNEasy preps and the isolated genomic DNA was pooled.

2.3.2 Precipitation of DNA

DNA precipitation was used to concentrate DNA samples that were too dilute for practical analysis or experimentation. Generally, isopropanol precipitation was used as follows: 0.7 volumes of isopropanol was added to DNA and incubated at room temperature for 15 minutes. DNA was pelleted by centrifugation at 12,000xg for 15 minutes. Pellets were washed in 70% ethanol and centrifuged again. Pellets were

allowed to air-dry for 5 minutes before resuspending in MilliQ water or appropriate buffer.

Concentration and quality of DNA was assessed by spectrophotometer (A_{260}/A_{280}) and/or by DNA gel electrophoresis.

2.3.3 Southern Blotting of DNA Fragments

3-5µg genomic DNA was digested overnight with appropriate restriction enzymes and digested products were electrophoresed on a 0.7% agarose gel. The gel was incubated with 0.3µg/ml ethidium bromide and analysed under UV to allow a scale image of the gel to be obtained. The gel was washed in distilled water for a few minutes before incubation for 30 minutes in depurination solution. The gel was then rinsed in distilled water and incubated for 30 minutes in denaturation solution. After another wash in water the gel washed for 10 minutes in 20xSSC .DNA was transferred to nylon membrane overnight. The transferred DNA fragment on nylon membrane was attached to the membrane by using UV for 20 seconds. The membrane was added to 5ml of Church-Gilbert solution and incubated at 100°C. The probe was radiolabeled and purified by using the Megaprime DNA Labeling Kit (Amersham). The probe was added to the Church-Gilbert solution and Southern blot membrane in the hybridisation tube and incubated overnight at 65°C. The blot was washed 3 times for 15 minutes in 2 x SSC/0.1% SDS followed by 3 x 15 minutes washes with 0.1 x SSC/0.1%SDS. The membrane was exposed to autoradiography film for a few days and developed.

2.3.4 FACS Analysis to Assess DNA Content

1ml of log phase promastigotes were pelleted by slow-speed centrifugation (1000xg for 5 minutes) and resuspended in 70% methanol/PBS overnight to fix. Fixed cells were washed with PBS followed by slow-speed centrifugation. Cells were resuspended in 1ml of PBS containing 10μg/ml propidium iodide (Sigma-Aldrich) and 10μg/ml RNase A and incubated at 37°C for 1 hour protected from the light. Cells were then immediately analysed in FACScalibur System (BD Biosciences) for DNA content.

2.4 Buffers and Reagents

10x PCR Mix: 45mM Tris-HCl (pH8.8), 11mM Ammonium phosphate, 4.5 mM $MgCl_2$, 6.7mM β-mercaptoethanol, 4.4mM EDTA (pH8.0), 1mM dATP, 1mM dCTP, 1mM dGTP, 1mM dTTP and 113μg/ml BSA. Store at -20°C.

Ampicillin: Stock solution is 100mg/ml in distilled water. Storage at -20C. Used at 1μg/ml.

Church-Gilbert's solution: 340mM Na_2HPO_4, 158mM NaH_2PO_4, 240Mm SDS, 1mM EDTA .Heated gently to solubilise components. Store at room temperature.

Cytomix: 120mM KCl2, 0.15mM CaCl2, 10mM K2HPO4, 25mM HEPES, 2mM EDTA and MgCl2; pH 7.6

DNA loading buffer (10x): 25% Ficoll, 0.42% bromophenol blue. Store at room temperature.

Depurination solution: 0.250M HCl. Store at room temperature.

Denaturation solution: 1.5M NaCl, 0.5M NaOH. Store at room temperature.

Ethidium bromide: 10mg/ml stock solution in distilled water. Store at room temperature.

Luria-Bertani (LB) broth: LB broth plus 0.8% (w/v) agar. Autoclaved to sterilise and store at room temperature.

Neutralisation solution: 1.5M NaCl, 0.5 M Tris-HCl pH7.0, store at room temperature.

PBS: 10mM phosphate buffer, 2.7Mm KCl, 137Mm NaCl, Ph 7.4. Autoclaved when required for tissue culture. Store at room temperature.

SSC (20 x): 3M NaCl, 0.3M Tri-sodium citrate pH7.0. Store at room temperature.

TBE (10x): 0.9M Tris-HCl, 0.9M boric acid, 25 mM EDTA. Store at room temperature.

2.5 Oligos Used in This Study

All oligos used in this study are listed in Table 2.7.

Table 2.7: Details of all oligos used in this study.

Primer	Sequence	Description
OL1984	GCACCCGGGATGGCGGCAACTCATCTTAC	Forward primer for 36.2050 (SEC14) ORF in *L. infantum*
OL1985	GCAGGATCCTCACTTCGGCAAACCGTTCTTTC	Reverse primer for 36.2050 (SEC14) ORF in *L. infantum*
OL1986	GCACCCGGGATGTCCGCCATCATCATCAC	Forward primer for 20.100 (Xylanase) ORF in *L. infantum*
OL1987	GCAGGATCCTCAGTGACGAGGAAGAAAGC	Reverse primer for 20.100 (Xylanase) ORF in *L. infantum*
OL1997	GCAAAGCTTCTCTAACTCTTTCTCTATCGCTG	Forward PCR primer for amplifying SEC14 5' flanking region of *L. infantum*
OL1998	GCAGTCGACGGTGAGTGACACGTCCTTTGCG	Reverse PCR primer for amplifying SEC14 5' flanking region of *L. infantum*
OL1999	GCACCCGGGGCTACGGATGCCCGTGTGCTGG	Forward PCR primer for amplifying SEC14 3' flanking region of *L. infantum*
OL2000	GCAAGATCTAAAAGGGAAAAGGAGTCCATCAC	Reverse PCR primer for amplifying SEC14 3' flanking region of *L. infantum*
OL2051	GCACCCGGGATGGCCGATGTGCAGCTCTCC	Forward primer for 28.0330 ORF in *L. infantum*
OL2052	GCAGGATCCGTCACATATCCATCAAGATTTCG	Reverse primer for 28.0330 ORF in *L .infantum*
OL2053	GCACCCGGGATGGGGCGAATCGACTCCTC	Forward primer for 36.2750 ORF in *L. infantum*
OL2054	GCAGGATCCTCAAGAGTTAGTCGGCAGCCGAG	Reverse primer for 36.2750 ORF in *L. infantum*
OL2095	GGCTCCAGCGGGCAAACTAC	Forward sequencing primer for xylanase *L. infantum*
OL2096	GACGTTCACCTTACCCTGCTG	Reverse sequencing primer for xylanase *L. infantum*
OL2097	GACAACCACTACCGCACCGAG	Forward sequencing primer for xylanase *L. infantum*
OL2098	GGCGTTATGAGGATGCGGGAAG	Forward sequencing primer for 28.0330 *L. infantum*
OL2099	CTTCCCCGCATCCTCATAACGCC	Reverse sequencing primer for 28.0330 *L. infantum*
OL2100	CATCTGCCTCGCCTTTAACG	Forward sequencing primer for 28.0330 *L. infantum*
OL2101	CGTTAAAGGCGAGGCAGATG	Reverse sequencing primer for 28.0330 *L. infantum*
OL2102	CATGAGGAAGCGGCACGGGAC	Forward sequencing primer for 36.2750 *L. infantum*
OL2103	CTCCCGTGCCGCTTCCTCATG	Reverse sequencing primer for 36.2750 *L. infantum*
OL2104	CACGCATCGCCCGCATCG	Forward sequencing primer for 36.2750 *L. infantum*
OL2105	CGATGCGGGCGATGCGTG	Reverse sequencing primer for 36.2750 *L. infantum*
OL2106	GGAGCACGATGAGCAGCACG	Forward sequencing primer for 36.2750 *L. infantum*
OL2107	CGTGCTGCTCATCGTGCTCC	Reverse sequencing primer for 36.2750 *L .infantum*

OL2112	ATGGGTCCAATCGACTCCTC	Forward primer for 36.2750 equivalent pseudogene in *L. major*
OL2113	GAACAGATCCACAGCGCCAC	Reverse primer for 36.2750 equivalent pseudogene in *L. major*
OL2114	GTCCGGTGTACATTGACCAAG	Forward primer for 28.0330 equivalent pseudogene in *L. major*
OL2115	TCACATATCCATCAAGATTTCG	Reverse primer for 28.0330 equivalent pseudogene in *L. major*
OL2116	ATGTCCACCATCATCATCAC	Forward primer for xylanase equivalent pseudogene in *L. major*
OL2117	CGGCTGCCTCAGTGACGAG	Reverse primer for xylanase equivalent pseudogene in *L. major*
OL2285	CAAAGCGCTCCCGTAGCAT	Reverse primer for PCR and sequencing SEC14 in *L. infantum*
OL2300	GGACGCCAACTTTTACCATCTTTC	Forward PCR primer for 5' integration check at SEC14 locus of *L. infantum*
OL2301	AGAAGCAAGGGCACTGGGCG	Reverse PCR primer for 3' integration check at SEC14 locus of *L. infantum*
OL2403	GACGGTGGAGCAGGACAAC	Forward sequencing primer for SEC 14 ORF of *L. infantum*
OL2404	GTTGTCCTGCTCCACCGTC	Reverse sequencing primer for SEC 14 ORF of *L. infantum*

CHAPTER 3 - RESULTS

3.1. Cloning and Sequencing of Four *L. infantum*–Specific Genes

With the recent publication of the genomes of *L. infantum* and *L. braziliensis*, together with the comparative analysis of the gene content compared with *L. major*, a small subset of genes have been identified as being specific to *L. infantum* (Peacock et al., 2007). More than two thirds of all genes in *Leishmania* encode hypothetical proteins and roughly the same proportion of hypothetical genes is represented in the subset of *L. infantum* specific genes. It is therefore important not to neglect the hypothetical genes when further characterisation is made.

To this end, the four *L.* infantum genes selected for this study were:

- LinJ28.0330 (hypothetical)
- LinJ36.2750 (hypothetical, with some domain homology (TRF4))
- LinJ20.1200 (putative endo-1,4-β-xylanase precursor)
- LinJ36.2050 (putative SEC14 cytosolic factor)

The analysis of these genes consists of three parts:

1. Sequence analysis of GeneDB sequence for each gene and encoded protein;
2. Amplification, cloning and sequencing of each gene from *L. infantum* to confirm GeneDB data as well as to ensure PCR products can be used for further experiments and

3. Amplification, cloning and sequencing of each gene from *L. donovani* to determine if genes are intact in another species that can cause visceral leishmaniasis.

3.1.1 Bioinformatic Analysis of the Encoded Proteins of Each Gene

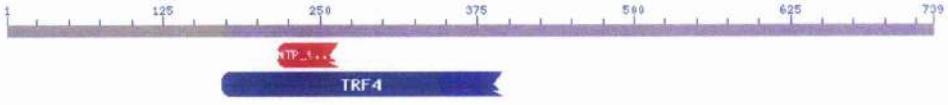
3.1.1.1 LinJ28.0330 Protein

Protein BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/) with the amino acid sequence of LinJ28.0330 protein against the non-redundant protein sequences database revealed no significant homology to any other protein known. Moreover, checks against the Conserved Domain Database (NCBI) revealed no significant homology to any known protein domain, so nothing can be surmised about the potential role of this protein in *L. infantum* without further study.

3.1.1.2 LinJ36.2750 Protein

The amino acid sequence of LinJ36.2750 was also checked against the non-redundant protein sequences database and the Conserved Domain Database. The encoded protein, although hypothetical, contains a domain that has homology to Topoisomerase-Related Function 4 (TRF4), a domain related to the TRF4 proteins of the yeast *Saccharomyces cerevisiae* (Figure 3.1.1A). Reciprocal BLAST analysis with the *S. cerevisiae* TRF4 amino acid sequence against predicted proteins in *L. infantum*, *L. major* and *L. braziliensis* (GeneDB Omniblast) revealed several genes encoding TRF4 domains, of which LinJ36.2750 was the third highest hit (Figure 3.1.1B). This provides a good basis on which to plan future experiments, beyond the scope of this thesis.

A



ref|XP_686065.2|	PREDICTED: similar to topoisomerase-related ...	72.0	1e-10
gb|AAH76872.1|	LOC445836 protein [Xenopus laevis]	71.6	2e-10
emb|CAN88001.1|	novel protein similar to vertebrate polymeras...	71.2	2e-10
gb|AAI33102.1|	LOC568678 protein [Danio rerio]	70.9	3e-10
ref|XP_697115.1|	PREDICTED: similar to PAP-associated domain-...	70.5	5e-10
ref|XP_001083145.1|	PREDICTED: similar to PAP associated doma...	69.7	6e-10
sp|Q8NDF8|PAPD5_HUMAN	PAP-associated domain-containing protei...	69.7	7e-10
ref|XP_535307.2|	PREDICTED: similar to PAP associated domain ...	69.7	7e-10
gb|AAD45198.1|AF089896_1	topoisomerase-related function protein	69.3	9e-10
ref|XP_001371537.1|	PREDICTED: similar to POLS protein [Monodelp	69.3	9e-10
gb|EAW82749.1|	PAP associated domain containing 5, isoform CR...	69.3	1e-09
gb|EAW82750.1|	PAP associated domain containing 5, isoform CR...	69.3	1e-09
ref|XP_001256516.1|	PREDICTED: similar to PAP-associated doma...	69.3	1e-09
ref|XP_001501238.1|	PREDICTED: similar to LAK-1 [Equus caballus]	68.9	1e-09
sp|Q68ED3|PAD5_MOUSE	PAP-associated domain-containing protein...	68.2	2e-09
ref|XP_134422.8|	PREDICTED: similar to PAP-associated domain-...	68.2	2e-09
gb|AAH80314.1|	Papd5 protein [Mus musculus]	68.2	2e-09
gb|EAW82747.1|	PAP associated domain containing 5, isoform CR...	68.2	2e-09
ref|XP_001062703.1|	PREDICTED: similar to DNA polymerase sigma [	67.8	3e-09

B

Sequences producing High-scoring Segment Pairs:		High Score	Sum Probability P(N)	N
LinJ07_V3.0780	|||topoisomerase-related function protein-...	292	2.6e-24	2
LinJ26_V3.0480	|||DNA polymerase sigma-like protein|Leish...	136	6.6e-12	3
LinJ36_V3.4190	**|||hypothetical protein|Leishmania infantu...**	**98**	**1.3e-07**	**2**
LinJ04_V3.0490	|||hypothetical protein, conserved|Leishma...	102	6.5e-06	2
LinJ32_V3.2600	|||hypothetical protein, conserved|Leishma...	112	4.8e-05	1
LinJ14_V3.1260	|||poly(A) polymerase, putative|Leishmania...	102	0.0016	1
LinJ29_V3.2710	|||poly(A) polymerase, putative|Leishmania...	99	0.0021	1
LinJ21_V3.2110	|||hypothetical protein, conserved|Leishma...	99	0.0027	2
LinJ20_V3.0670	|||hypothetical protein, conserved|Leishma...	84	0.068	1
LinJ28_V3.0840	|||hypothetical protein, conserved|Leishma...	68	0.72	2
LinJ26_V3.0380	|RET2||RNA editing 3' terminal uridylyl tr...	59	0.84	2
LinJ14_V3.1020	|3'TUTASE||RNA editing 3' terminal uridyly...	77	0.98	2
LinJ24_V3.0250	|||hypothetical protein, conserved|Leishma...	74	0.9998	1
LinJ31_V3.0230	|||hypothetical protein, conserved|Leishma...	61	0.9999	1

Figure 3.1.2: Analysis of LinJ36.2750 sequence. A) Protein BLAST analysis of LinJ36.2750 was carried out (http://www.ncbi.nlm.nih.gov/BLAST/) showing the known proteins with significant homology. B) Reciprocal BLAST with *S. cerevisiae* TRF4. Note that Linj36.2750 has been renamed Linj36 V3.4190 (shown in bold typeface) in the latest release of the *L. infantum* genome.

3.1.1.3 LinJ20.1200 (Endo-1,4-β-xylanase precursor)

BLASTp analysis of the amino acid sequence of this *L. infantum*-specific protein reveals that the most similar proteins are bacterial (Figure 3.1.2A). The proteins with highest homology are found in *Solibacter usitatus* and *Acidobacteria bacterium*, raising the possibility that the genes were acquired by horizontal gene transfer as has been seen previously with the *L. major*-specific gene encoding PFP1 (Eschenlauer et al., 2006). Reciprocal BLAST of the *Solibacter usitatus* amino acid sequence against predicted protein databases in GeneDB for *L. infantum*, *L. major* and *L. braziliensis* has identified LinJ20.1200 as the top hit in *L. infantum* (with a P/N value of 5.4e-37) but has no significant homology to proteins in either of the other two species (Figure 3.1.2B).

A

Sequences producing significant alignments:	Score (Bits)	E Value
ref|YP_823348.1| putative esterase [Solibacter usitatus Ellin...	149	5e-34
ref|YP_592320.1| putative esterase [Acidobacteria bacterium E...	145	5e-33
ref|ZP_01718065.1| predicted xylanase [Algoriphagus sp. PR1] ...	134	1e-29
gb|EDO10802.1| hypothetical protein BACOVA_03435 [Bacteroides...	129	4e-28
ref|YP_825710.1| putative esterase [Solibacter usitatus Ellin...	127	1e-27
ref|YP_824333.1| putative esterase [Solibacter usitatus Ellin...	127	1e-27
emb|CAJ19109.1| putative acetyl xylan esterase [unidentified mic	125	6e-27
ref|YP_824260.1| putative esterase [Solibacter usitatus Ellin...	124	1e-26
ref|ZP_02025871.1| hypothetical protein RUBVEN_01126 [Eubacte...	120	1e-25
ref|YP_824257.1| putative esterase [Solibacter usitatus Ellin...	119	3e-25
ref|YP_590954.1| putative esterase [Acidobacteria bacterium E...	119	6e-25
ref|YP_001297384.1| glycoside hydrolase family 43 [Bacteroide...	114	2e-23
ref|ZP_01720999.1| Possible xylan degradation enzyme (alpha/b...	112	4e-23
ref|YP_001196208.1| putative esterase [Flavobacterium johnson...	112	5e-23
ref|ZP_01718168.1| putative glycosyl hydrolase exoenzyme [Alg...	111	7e-23
ref|ZP_01718170.1| predicted xylanase [Algoriphagus sp. PR1] ...	110	3e-22
ref|ZP_01120522.1| putative glycosyl hydrolase exoenzyme [Rob...	108	6e-22
ref|YP_001303364.1| putative xylanase [Parabacteroides distas...	103	2e-20
ref|YP_824332.1| putative esterase [Solibacter usitatus Ellin...	103	2e-20
ref|ZP_01882216.1| putative esterase [Pedobacter sp. BAL39] >...	103	2e-20
emb|CAJ19122.1| acetyl xylan esterase [unidentified microorganis	100	1e-19

B

Summary for: L. brasiliensis predicted proteins [wublastp], for query: UNKNOWN-QUERY

Name:	LbrM23.0510	**Score:**	77	**(P/N):**	0.16	**N:**	1
Name:	LbrM22.0760	**Score:**	83	**(P/N):**	0.22	**N:**	1
Name:	LbrM35.5610	**Score:**	78	**(P/N):**	0.29	**N:**	1
Name:	sequence orphan	**Score:**	67	**(P/N):**	0.94	**N:**	1
Name:	LbrM26.1220	**Score:**	71	**(P/N):**	0.96	**N:**	1

Summary for: L. infantum predicted proteins [wublastp], for query: UNKNOWN-QUERY

Name:	LinJ20.1200	**Score:**	382	**(P/N):**	5.4e-37	**N:**	1
Name:	LinJ28.0850	**Score:**	68	**(P/N):**	0.38	**N:**	1
Name:	LinJ32.2890	**Score:**	77	**(P/N):**	0.43	**N:**	1
Name:	LinJ23.0520	**Score:**	72	**(P/N):**	0.58	**N:**	1
Name:	LinJ08.0490	**Score:**	69	**(P/N):**	0.76	**N:**	1

Summary for: L. major predicted proteins [wublastp], for query: UNKNOWN-QUERY

Name:	LmjF16.0630	**Score:**	80	**(P/N):**	0.24	**N:**	1
Name:	LmjF06.0470	**Score:**	82	**(P/N):**	0.27	**N:**	1
Name:	LmjF23.0490	**Score:**	74	**(P/N):**	0.32	**N:**	1
Name:	LmjF28.0825	**Score:**	68	**(P/N):**	0.38	**N:**	1
Name:	LmjF35.5000	**Score:**	71	**(P/N):**	0.57	**N:**	1

Figure 3.1.2: Analysis of LinJ20.1200 sequence. A Protein BLAST analysis of LinJ20.1200 (xylanase) was carried out (http://www.ncbi.nlm.nih.gov/BLAST/) showing the known proteins with significant homology. B Reciprocal BLAST (Omniblast) analysis of the *Solibacter usitatus* amino acid sequence against the *L. braziliensis*, *L. infantum* and *L. major* predicted protein databases in GeneDB.

3.1.1.4 LinJ36.2050 (SEC14 Cytosolic Factor)

This gene has been identified as encoding a putative orthologue of the yeast Sec14p protein, which is involved in membrane restructuring and secretory vesicle formation at the trans-Golgi network (TGN). Interestingly, when the yeast amino acid sequence is reciprocally blasted against the predicted proteins database for *L.*

infantum, *L. major and L. braziliensis*, the protein encoded by LinJ36.2050 is actually the second most significant hit (Figure 3.1.3).

Summary for: *L. infantum* predicted proteins [wublastp], for query: UNKNOWN-QUERY

Name:	LinJ35.3630	**Score:**	116	**(P/N):**	1.7e-13	**N:**	2
Name:	LinJ36.2050	**Score:**	176	**(P/N):**	1.3e-12	**N:**	1
Name:	LinJ15.0640	**Score:**	117	**(P/N):**	9.2e-10	**N:**	2
Name:	LinJ30.2020	**Score:**	99	**(P/N):**	1.4e-08	**N:**	2
Name:	LinJ36.3340	**Score:**	121	**(P/N):**	6.4e-06	**N:**	1

Summary for: *L. major* predicted proteins [wublastp], for query: UNKNOWN -QUERY

Name:	LmjF35.3560	**Score:**	117	**(P/N):**	6.4e-14	**N:**	2
Name:	LmjF15.0610	**Score:**	120	**(P/N):**	2.2e-10	**N:**	3
Name:	LmjF36.3430	**Score:**	134	**(P/N):**	2.3e-07	**N:**	1
Name:	LmjF30.1680	**Score:**	90	**(P/N):**	2.7e-07	**N:**	2
Name:	LmjF31.2050	**Score:**	72	**(P/N):**	0.0060	**N:**	3

Summary for: *L. brasiliensis* predicted proteins [wublastp], for query: UNKNOWN-QUERY

Name:	LbrM32_V2.1420	**Score:**	160	**(P/N):**	4.5e-12	**N:**	2
Name:	LbrM34_V2.3490	**Score:**	111	**(P/N):**	1.1e-11	**N:**	2
Name:	LbrM15_V2.0650	**Score:**	132	**(P/N):**	1.1e-10	**N:**	2
Name:	LbrM35_V2.3660	**Score:**	142	**(P/N):**	2.8e-08	**N:**	1
Name:	LbrM31_V2.2310	**Score:**	91	**(P/N):**	3.9e-06	**N:**	3

Figure 3.1.3: OmniBLAST output from analysis of the amino acid sequence of yeast Sec14p against the predicted protein databases for all three *Leishmania* species in GeneDB.

The top hit in *L. infantum* and *L. major* is LinJ35.3630 and LmjF35.3560 (which are orthologues of each other). Closer analysis of the proteins has revealed that the domain structure of these two proteins is, perhaps unsurprisingly, more similar to yeast Sec14p than the *L. infantum*–specific gene (LinJ36.2050) under investigation (Figure 3.1.4). The domains identified in this analysis are the CRAL_TRIO domain

and the SEC14 domain. These domains are lipid binding domains and they are thought to be part of the same domain (Saito et al., 2007) even though they are annotated as separate domains in the NCBI Conserved Domain Database. This is because not all SEC14 domain containing proteins possess the CRAL_TRIO portion (Saito et al., 2007). This seems to be true of the *L. infantum*–specific LinJ36.2050, which has a "CAP_GLY" domain instead. These domains have been shown in other systems to be involved in associating CAP_GLY-containing proteins with elements of the cytoskeleton, notably microtubules (Mishima et al., 2007). Interestingly, the only known proteins with this domain structure are found in *L. infantum* and its orthologue in *T. cruzi*. The gene is also missing from *T. brucei*.

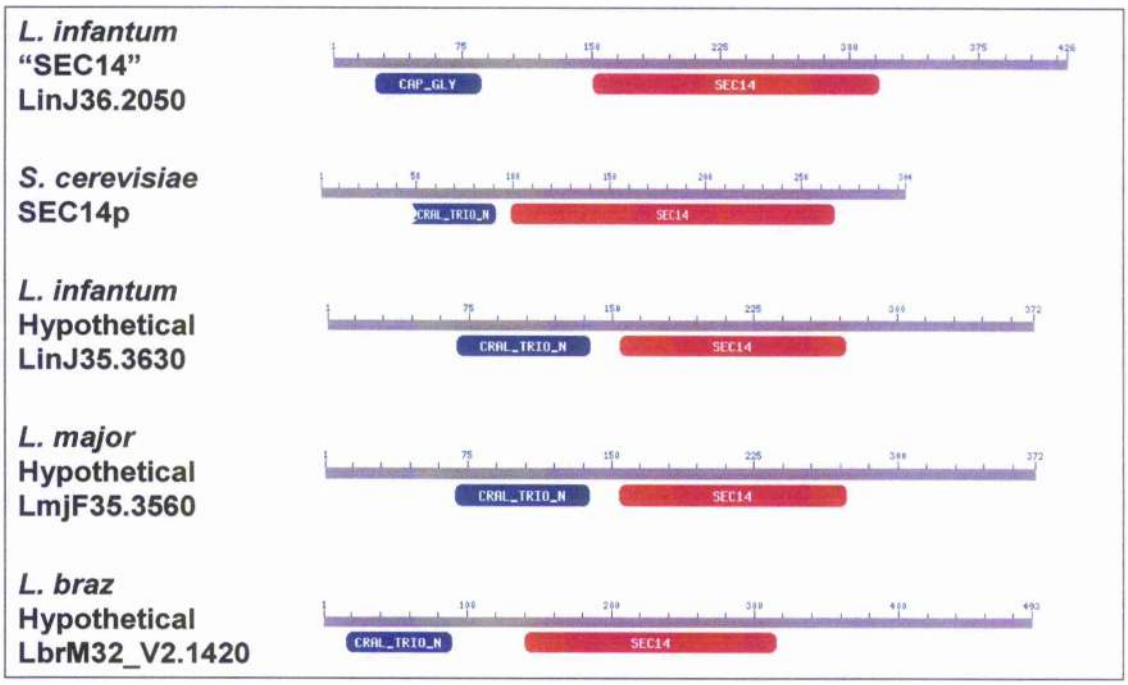


Figure 3.1.4: Domain structure of putative Sec14p orthologues in *Leishmania*. Protein BLAST analysis was carried out using the NCBI Blast server. (http://www.ncbi.nlm.nih.gov/BLAST/).

3.1.2 Cloning and Sequencing of the Four Genes from *L. infantum*

In order to confirm that the sequence held on GeneDB for each of the four genes mentioned above, the open reading frames must be isolated and characterised by sequencing. To this end, primers were designed to each of the four genes in order to amplify the complete open reading frames (ORFs) (Table 3.1.1). *Xma*I and *Bam*HI restriction sites were engineered onto the 5' and 3' ends respectively, of the primers and therefore the PCR products, to allow eventual cloning and expression from the *Leishmania* episomal vector pXG.

Table 3.1.1: Designed primers for PCR for four genes from *L. infantum* (restriction sites are underlined)

GeneDB Designation	Designed Primers	Primer Sequence
LinJ28.0330	OL2051	GCACCCGGGATGGCCGATGTGCAGCTCTCC
	OL2052	GCAGGATCCGTCACATATCCATCAAGATTTCG
LinJ36.2750	OL2053	GCACCCGGGATGGGGCGAATCGACTCCTC
	OL2054	GCAGGATCCTCAAGAGTTAGTCGGCAGCCGAG
LinJ20.1200 (Xylanase)	OL1986	GCACCCGGGATGTCCGCCATCATCATCAC
	OL1987	GCAGGATCCTCAGTGACGAGGAAGAAAGC
LinJ36.2050 (SEC14)	OL1984	GCACCCGGGATGGCGGCAACTCATCTTAC
	OL1985	GCAGGATCCTCACTTCGGCAAACCGTTCTTTC

3.1.2.1 PCR of Four *L. infantum*-Specific Genes

Polymerase chain reaction (PCR) was carried out to amplify the fragments. The primers were initially tested at a range of annealing temperatures to determine the

optimal conditions. This was done with a variety of proof-reading enzymes in either the supplied PCR buffer or in the 10x PCR buffer commonly used in the Mottram laboratory. The fragments were amplified and cloned by using T-Vector such as pGEM-T Easy (Promega) or pCR2.1 TOPO TA (Invitrogen). This required the addition of A-overhangs by the inclusion of Taq DNA polymerase at the end of the PCR with a 10 minutes incubation step at 72°C, as the proof-reading enzymes do not leave overhanging A-residues. Reactions were electrophoresed on gels to visualise and allow purification of the appropriate DNA fragment by gel extraction (Figure 3.1.5).

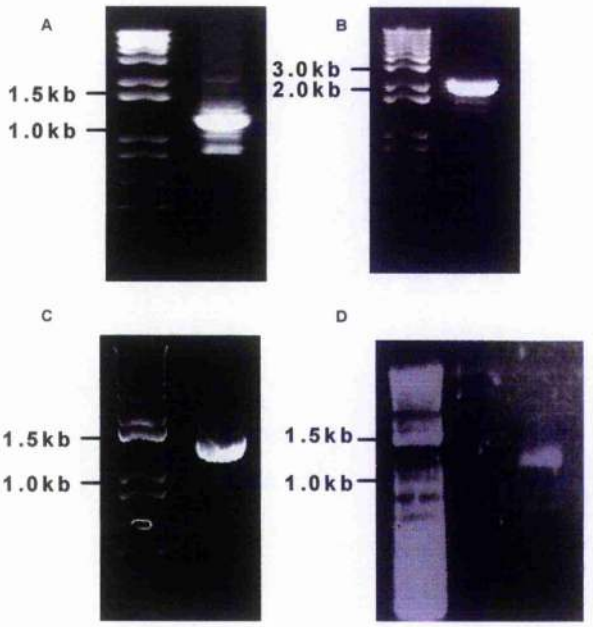


Figure 3.1.5: PCR Amplification for each of the four genes under investigation. A) LinJ28.0330 ORF PCR product (predicted size is ~1.4kb). B) LinJ36.2750 ORF PCR product (predicted size is ~2.2kb). C) LinJ20.1200 (Xylanase) ORF PCR product (predicted size is ~1.3kb). D) LinJ36.2050 (SEC14) ORF PCR product (predicted size is ~1.3kb).

Once PCR products were cloned into a T-vector, they were sent to the University of Dundee Sequencing Service to be sequenced using a variety of primers both internal to the gene and external, within the plasmid backbone.

3.1.3 Sequence Analysis

Sequence fragments for each gene were assembled into contigs and analysed using the ContigExpress Software (Vector-NTI Advance v10, Invitrogen). Contiguous sequence of each open reading frame was aligned with the published sequence on GeneDB using Align-X software (Vector-NTI, Invitrogen) and these are shown in Figure 3.1.6. Note that only a single PCR fragment was cloned and sequenced for each ORF, so errors may have been introduced by PCR. Independent clones would need to be sequenced to confirm this.

A LinJ28.0330 (Hypothetical) Alignment

```
                                1                                                50
Cloned 28.0330 Seq      (1) ATGGCCGATGTGCAGCTCTCCAACTCCAGCAATGTCAAGGCTGTCTATGC
       LinJ28.0330      (1) ATGGCCGATGTGCAGCTCTCCAACTCCAGCAATGTCAAGGCTGTCTATGC
                               51                                               100
Cloned 28.0330 Seq     (51) AGGACACAGCTTGAGCCTCTTCATGGACCTCCGCCCGCCGGCGGAGGTGC
       LinJ28.0330     (51) AGGACACAGCTTGAGCCTCTTCATGGACCTCCGCCCGCCGGCGGAGGTGC
                              101                                               150
Cloned 28.0330 Seq    (101) TGGCGACGTGCAGCGGCTCCGTTGTCGAGGACCTGCCCGTGACACTAAGT
       LinJ28.0330    (101) TGGCGACGTGCAGCGGCTCCGTTGTCGAGGACCTGCCCGTGACACTAAGT
                              151                                               200
Cloned 28.0330 Seq    (151) GTCATGCTGCGTCGCATCCATAGCGAAAGCGAGACCCCGCTTTCGTTTCG
       LinJ28.0330    (151) GTCATGCTGCGTCGCATCCATAGCGAAAGCGAGACCCCGCTTTCGTTTCG
                              201                                               250
Cloned 28.0330 Seq    (201) TTGTTCTTCCTCGCATCGACTTCTCACCTACGTGCAGTGCGAAGACGGTG
       LinJ28.0330    (201) TTGTTCTTCCTCGCATCGACTTCTCACCTACGTGCAGTGCGAAGACGGTG
                              251                                               300
Cloned 28.0330 Seq    (251) CTGGCGGCGATGCTGTCCAGCTGGGAATGAGCGGTAACAACCTGCTCAAG
       LinJ28.0330    (251) CTGGCGGCGATGCTGTCCAGCTGGGAATGAGCGGTAACAACCTGCTCAAG
                              301                                               350
Cloned 28.0330 Seq    (301) GTCACTATAGACCTTGCTACCTTCCCGGACCGGTTGCAGCGACTAGCAGC
       LinJ28.0330    (301) GTCACTATAGACCTTGCTACCTTCCCGGACCGGTTGCAGCGACTAGCAGC
                              351                                               400
Cloned 28.0330 Seq    (351) GGCCGATGTCGCTGCAGCTGAGCATGCCTTGCTTGATCAGTCGTGGGTGA
       LinJ28.0330    (351) GGCCGATGTCGCTGCAGCTGAGCATGCCTTGCTTGATCAGTCGTGGGTGA
                              401                                               450
Cloned 28.0330 Seq    (401) TCGGCTCGCGGTTCCGCCTCTCAATGCAGCTGCGTACCAGAGCGCGTGCC
       LinJ28.0330    (401) TCGGCTCGCGGTTCCGCCTCTCAATGCAGCTGCGTACCAGAGCGCGTGCC
                              451                                               500
Cloned 28.0330 Seq    (451) GAGTGGCCACCCCTCGTCGCCCCGCAGTGGAGCGAGTCCAGTGACGGCTG
       LinJ28.0330    (451) GAGTGGCCACCCCTCGTCGCCCCGCAGTGGAGCGAGTCCAGTGACGGCTG
                              501                                               550
Cloned 28.0330 Seq    (501) CCTTGTCCCGACACCTTGCAGCCATCACATGGACAATCATGACAGCGCGG
       LinJ28.0330    (501) CCTTGTCCCGACACCTTGCAGCCATCACATGGACAATCATGACAGCGCGG
                              551                                               600
Cloned 28.0330 Seq    (551) GCGTTATGAGGATGCGGGAAGAGTTCGCCGCGAAGGGCGTATCCGCGTCG
       LinJ28.0330    (551) GCGTTATGAGGATGCGGGAAGAGTTCGCCGCGAAGGGCGTATCCGCGTCG
```

```
                                601                                     650
Cloned 28.0330 Seq    (601) CCCAGTATCGTTCGGCGGAATCGGCCACTGTCTCTATACGACAAGTCAGC
       LinJ28.0330    (601) CCCAGTATCGTTCGGCGGAATCGGCCACTGTCTCTATACGACAAGTCAGC
                                651                                     700
Cloned 28.0330 Seq    (651) ACCGCGCTCTCAGTTGAAGGGTCCGGTGTACATTGACCAAGAGCTGTTCC
       LinJ28.0330    (651) ACCGCGCTCTCAGTTGAAGGGTCCGGTGTACATTGACCAAGAGCTGTTCC
                                701                                     750
Cloned 28.0330 Seq    (701) CGATCGTTCGTATTTTGTTTGTTCCTCAACTCTACACGGAGGCCATCCCG
       LinJ28.0330    (701) CGATCGTTCGTATTTTGTTTGTTCCTCAACTCTACACGGAGGCCATCCCG
                                751                                     800
Cloned 28.0330 Seq    (751) ACGGCAACGGCGGTGTTGCGGAGCATGACGGTGCAAGTCGGAGTGGAGCG
       LinJ28.0330    (751) ACGGCAACGGCGGTGTTGCGGAGCATGACGGTGCAAGTCGGAGTGGAGCG
                                801                                     850
Cloned 28.0330 Seq    (801) TTGCCGCGAGAGCCGCGGCGGCAAGCCGCTTCGTCGCGCCCTGCGAGACC
       LinJ28.0330    (801) TTGCCGCGAGAGCCGCGGCGGCAAGCCGCTTCGTCGCGCCCTGCGAGACC
                                851                                     900
Cloned 28.0330 Seq    (851) GCAGCACTAAAGACAACGTCGCCGTGCTCGATCGCGTGCCGGCAAGTAGC
       LinJ28.0330    (851) GCAGCACTAAAGACAACGTCGCCGTGCTCGATCGCGTGCCGGCAAGTAGC
                                901                                     950
Cloned 28.0330 Seq    (901) AGCACGTTAGATGAGGAGAGCAAAGCCCTACTGCAACCGTACGAGGCCAT
       LinJ28.0330    (901) AGCACGTTAGATGAGGAGAGCAAAGCCCTACTGCAACCGTACGAGGCCAT
                                951                                    1000
Cloned 28.0330 Seq    (951) CTGCCTCGCCTTTAACGCCCATGTTCGTCAGCCCTCAATGACGTCGGTCA
       LinJ28.0330    (951) CTGCCTCGCCTTTAACGCCCATGTTCGTCAGCCCTCAATGACGTCGGTCA
                               1001                                    1050
Cloned 28.0330 Seq   (1001) CTGCCTACGACGTGGCACACGCGAGGCGACTTGTTGTGCAGGACTACACC
       LinJ28.0330   (1001) CTGCCTACGACGTGGCACACGCGAGGCGACTTGTTGTGCAGGACTACACC
                               1051                                    1100
Cloned 28.0330 Seq   (1051) TTTGAGGTGATGGTCACGGCAGCGAGGATGAAAAGCGAAACCGTGGAGAC
       LinJ28.0330   (1051) TTTGAGGTGATGGTCACGGCAGCGAGGATGAAAAGCGAAACCGTGGAGAC
                               1101                                    1150
Cloned 28.0330 Seq   (1101) CGTGCAGCGCGCTATTCAGGCCTATCAGTACGAGAGCGGCGATGCAGATG
       LinJ28.0330   (1101) CGTGCAGCGCGCTATTCAGGCCTATCAGTACGAGAGCGGCGATGCAGATG
                               1151                                    1200
Cloned 28.0330 Seq   (1151) TGCTGGGCATGAACCTCGATGAAGCCGTCAGCTCCGTCGAAGAGCGGTTT
       LinJ28.0330   (1151) TGCTGGGCATGAACCTCGATGAAGCCGTCAGCTCCGTCGAAGAGCGGTTT
                               1201                                    1250
Cloned 28.0330 Seq   (1201) GGCTACATGAAGGACGCGCCGTCGGTGGACTTGATGAGCTTTCACCATGT
       LinJ28.0330   (1201) GGCTACATGAAGGACGCGCCGTCGGTGGACTTGATGAGCTTTCACCATGT
                               1251                                    1300
Cloned 28.0330 Seq   (1251) TGTTTGGGAGGCGATGCGAGCATGCATGGTGGCCGACGTCACTTCCAGTG
       LinJ28.0330   (1251) TGTTTGGGAGGCGATGCGAGCATGCATGGTGGCCGACGTCACTTCCAGTG
                               1301                                    1350
Cloned 28.0330 Seq   (1301) CACTTGCAAACGAACTGGAGGACGAGCTCGACCGCTATGAGGACCGCATC
       LinJ28.0330   (1301) CACTTGCAAACGAACTGGAGGACGAGCTCGACCGCTATGAGGACCGCATC
                               1351            1374
Cloned 28.0330 Seq   (1351) AACGAAATCTTGATGGATATGTGA
       LinJ28.0330   (1351) AACGAAATCTTGATGGATATGTGA
```

B LinJ36.2750 (Hypothetical) Alignment

```
                                  1                                      50
Cloned 36.2750 Seq      (1) ATGGGGCGAATCGACTCCTCTTGTGTCGCTGCCACGGTGACGCCTAGCAA
       LinJ36.2750      (1) ATGGGGCGAATCGACTCCTCTTGTGTCGCTGCCACGGTGACGCCTAGCAA
                                 51                                     100
Cloned 36.2750 Seq     (51) GGTCGGCGCATCGTCGTCTTGTCCCGGCACGGCGACAAGCACTGGCAAGA
       LinJ36.2750     (51) GGTCGGCGCATCGTCGTCTTGTCCCGGCACGGCGACAAGCACTGGCAAGA
                                101                                     150
Cloned 36.2750 Seq    (101) TAGACGAGCCTGCTTCTCGCGCCGCTGCCCCGCAGGGCCGGGCGCAGCGA
       LinJ36.2750    (101) TAGACGAGCCTGCTTCTCGCGCCGCTGCCCCGCAGGGCCGGGCGCAGCGA
                                151                                     200
Cloned 36.2750 Seq    (151) TGTGGCCGAGCGCGACCCGCCCAGATGCACAACAGCCTCAGTGCGGGGAT
       LinJ36.2750    (151) TGTGGCCGAGCGCGACCCGCCCAGATGCACAACAGCCTCAGTGCGGGGAT
                                201                                     250
Cloned 36.2750 Seq    (201) GTGTGATAACCAACAGCAGCACGCCGCAGGCGTTTGGAAGTTGATTGCAG
       LinJ36.2750    (201) GTGTGATAACCAACAGCAGCACGCCGCAGGCGTTTGGAAGTTGATTGCAG
                                251                                     300
Cloned 36.2750 Seq    (251) GGCCGCAGAGCGCCTCCACGACTCCGCCGCCGCTCGTGATGGGTGACGGC
       LinJ36.2750    (251) GGCCGCAGAGCGCCTCCACGACTCCGCCGCCGCTCGTGATGGGTGACGGC
                                301                                     350
Cloned 36.2750 Seq    (301) ACAGGCGGCGAGGATCATGCGCAGGCTGCCCACACCGGCGTCCGGTGGAC
       LinJ36.2750    (301) ACAGGCGGCGAGGATCATGCGCAGGCTGCCCACACCGGCGTCCGGTGGAC
```

```
                                351                                 400
Cloned 36.2750 Seq    (351) TAAGCACGCGTGGGAAAAGGGCAACGACTGCCCTGCGGGTAGCGCGTACG
       LinJ36.2750    (351) TAAGCACGCGTGGGAAAAGGGCAACGACTGCCCTGCGGGTAGCGCGTACG
                                401                                 450
Cloned 36.2750 Seq    (401) GGTCTTTTGTAGGCGCGCCGCTGCGCAGGCTGCCGCCAACGCCTCCGTTC
       LinJ36.2750    (401) GGTCTTTTGTAGGCGCGCCGCTGCGCAGGCTGCCGCCAACGCCTCCGTTC
                                451                                 500
Cloned 36.2750 Seq    (451) CCTGGTGTGCGTCTTCCCGCCTGGTGTCGGTCCGCGTCGGCGACGGCGTA
       LinJ36.2750    (451) CCTGGTGTGCGTCTTCCCGCCTGGTGTCGGTCCGCGTCGGCGACGGCGTA
                                501                                 550
Cloned 36.2750 Seq    (501) TCCGCTCTCCGAGGATGGCCTCACAAACGAGCTGCTGGACTTCTTCTATT
       LinJ36.2750    (501) TCCGCTCTCCGAGGATGGCCTCACAAACGAGCTGCTGGACTTCTTCTATT
                                551                                 600
Cloned 36.2750 Seq    (551) ATCTGCAGTTGACCTCCCATGAGGAAGCGGCACGGGAGCGGCTGCTTGGC
       LinJ36.2750    (551) ATCTGCAGTTGACCTCCCATGAGGAAGCGGCACGGGAGCGGCTGCTTGGC
                                601                                 650
Cloned 36.2750 Seq    (601) TACGTGCAGGCGTGCGTGGCGAAGTTGTGGGGCCCCTGTAAACCGGGGAG
       LinJ36.2750    (601) TACGTGCAGGCGTGCGTGGCGAAGTTGTGGGGCCCCTGTAAACCGGGGAG
                                651                                 700
Cloned 36.2750 Seq    (651) CGAAGCCGAGGGAACAGCACAGGTGATGCTGTACGGCAGCTACGCTCTAG
       LinJ36.2750    (651) CGAAGCCGAGGGAACAGCACAGGTGATGCTGTACGGCAGCTACGCTCTAG
                                701                                 750
Cloned 36.2750 Seq    (701) GGCTTTCTCTTCCAAGCAGCGACATTGATCTTGCCCTGACTTTCCCGGCC
       LinJ36.2750    (701) GGCTTTCTCTTCCAAGCAGCGACATTGATCTTGCCCTGACTTTCCCGGCC
                                751                                 800
Cloned 36.2750 Seq    (751) GAGGAGCAGGTGGACGTCGCGACCATTGCAGTTGAGGAAAGGAGTGGGGA
       LinJ36.2750    (751) GAGGAGCAGGTGGACGTCGCGACCATTGCAGTTGAGGAAAGGAGTGGGGA
                                801                                 850
Cloned 36.2750 Seq    (801) TCGCGCTCTCCTGGCGCCGGCGGTCTCGAAGAAACGCCAGGCGCTGCACC
       LinJ36.2750    (801) TCGCGCTCTCCTGGCGCCGGCGGTCTCGAAGAAACGCCAGGCGCTGCACC
                                851                                 900
Cloned 36.2750 Seq    (851) TTGAGCGGCTCCACGATCTTGCCGAGCAGCTGCGGAACTCTGCGACGTTT
       LinJ36.2750    (851) TTGAGCGGCTCCACGATCTTGCCGAGCAGCTGCGGAACTCTGCGACGTTT
                                901                                 950
Cloned 36.2750 Seq    (901) CCCGAGTTGGAGGTAGAGGTGTATGATCAGTGCCGTGTGCCGCGCATTCA
       LinJ36.2750    (901) CCCGAGTTGGAGGTAGAGGTGTATGATCAGTGCCGTGTGCCGCGCATTCA
                                951                                1000
Cloned 36.2750 Seq    (951) TCTACGAGACAGGACTTGTGGTGGCGTGTCGTGTGACATCAACAGCTCCT
       LinJ36.2750    (951) TCTACGAGACAGGACTTGTGGTGGCGTGTCGTGTGACATCAACAGCTCCT
                               1001                                1050
Cloned 36.2750 Seq   (1001) TCGCCTCAGCACGCATCGCCCGCATCGTGGCGCGGCAGCGGCTATGGCTA
       LinJ36.2750   (1001) TCGCCTCAGCACGCATCGCCCGCATCGTGGCGCGGCAGCGGCTATGGCTA
                               1051                                1100
Cloned 36.2750 Seq   (1051) CAGGACTCCCCGCTCGCGGCGTTCCTCGTGCGCGTCACCAAGGCAGCCGT
       LinJ36.2750   (1051) CAGGACTCCCCGCTCGCGGCGTTCCTCGTGCGCGTCACCAAGGCAGCCGT
                               1101                                1150
Cloned 36.2750 Seq   (1101) GAAGCAGTGAGGCCTCCACGAAGTGTTCTGGGGCGGCGTCGCCTCCACTG
       LinJ36.2750   (1101) GAAGCAGTGGGGCCTCCACGAAGTGTTCTGGGGCGGCGTCGCCTCCACT
                               1151                                1200
Cloned 36.2750 Seq   (1151) CGCTGTACTGCCTAGTGCTCCGCTTCCTCGCCCAGATGGAACAGCTTTGC
       LinJ36.2750   (1151) CGCTGTACTGCCTAGTGCTCCGCTTCCTCGCCCAGATGGAACAGCTTTGC
                               1201                                1250
Cloned 36.2750 Seq   (1201) CGACACGCGCAGGTGGAGGAGAATTTGTCTCCACCGCGCTACGCGGAGGC
       LinJ36.2750   (1201) CGACACGCGCAGGTGGAGGAGAATTTGTCTCCACCGCGCTACGCGGAGGC
                               1251                                1300
Cloned 36.2750 Seq   (1251) CGCCGCCGCGCATGTCGCAACGCAGCTCTCTTCGTTGGCTTGTTCGTCAC
       LinJ36.2750   (1251) CGCCGCCGCGCATGTCGCAACGCAGCTCTCTTCGTTGGCTTGTTCGTCAC
                               1301                                1350
Cloned 36.2750 Seq   (1301) CGTTGTCTTCACCACAACTCTCCATGCCACACCCAGCCAACGAGTCTTTC
       LinJ36.2750   (1301) CGTTGTCTTCACCACAACTCTCCATGCCACACCCAGCCAACGAGTCTTTC
                               1351                                1400
Cloned 36.2750 Seq   (1351) GCAGCGTTTACCCCAGTTCCAGCCGGCACGGCTTCCACTGCTCGCGGCCT
       LinJ36.2750   (1351) GCAGCGTTTACCCCAGTTCCAGCCGGCACGGCTTCCACTGCTCGCGGCCT
                               1401                                1450
Cloned 36.2750 Seq   (1401) CAACAGGCTAAGCGCGGCGGCAGCCGTTCCTTCATATGTACCCAGCTGGG
       LinJ36.2750   (1401) CAACAGGCTAAGCGCGGCGGCAGCCGTTCCTTCATATGTACCCAGCTGGG
                               1451                                1500
Cloned 36.2750 Seq   (1451) TTGTGAATAGCAGGAAGCACGACTTGAACAAGTCGTCGTCCTTGTGCGAT
       LinJ36.2750   (1451) TTGTGAATAGCAGGAAGCACGACTTGAACAAGTCGTCGTCCTTGTGCGAT
                               1501                                1550
Cloned 36.2750 Seq   (1501) ATGGACAAGAACAGCGCGCTGTGCCCTGTCAGTTTGACCGCTTGCTCCCT
       LinJ36.2750   (1501) ATGGACAAGAACAGCGCGCTGTGCCCTGTCAGTTTGACCGCTTGCTCCCT
                               1551                                1600
Cloned 36.2750 Seq   (1551) TTCCGATTGGAGCACGATGAGCAGCACGTGCGACGAGGATAGGGAGGCGG
       LinJ36.2750   (1551) TTCCGATTGGAGCACGATGAGCAGCACGTGCGACGAGGATAGGGAGGCGG
```

```
                          1601                                        1650
Cloned 36.2750 Seq  (1601) ACGACGAGGTGGATGGCTACACGCAGGCGGGTCTCACGCGCACCTCCGGC
       LinJ36.2750  (1601) ACGACGAGGTGGATGGCTACACGCAGGCGGGTCTCACGCGCACCTCCGGC
                          1651                                        1700
Cloned 36.2750 Seq  (1651) GCCACCACCACTGCGACTACGCCGATCTCTCGCACCGCGGCGAGCACACC
       LinJ36.2750  (1651) GCCACCACCACTGCGACTACGCCGATCTCTCGCACCGCGGCGAGCACACC
                          1701                                        1750
Cloned 36.2750 Seq  (1701) GATGCCGGCTGCAAGCGAAGCGCATGAGGGAGGCACAGCGACCGCATCGA
       LinJ36.2750  (1701) GATGCCGGCTGCAAGCGAAGCGCATGAGGGAGGCACAGCGACCGCATCGA
                          1751                                        1800
Cloned 36.2750 Seq  (1751) GTGCCTTGGCACCACCAGTGAGCGTCAATTCTGCTGCGCTTCGCGAGGGC
       LinJ36.2750  (1751) GTGCCTTGGCACCACCAGTGAGCGTCAATTCTGCTGCGCTTCGCGAGGGC
                          1801                                        1850
Cloned 36.2750 Seq  (1801) GTGTTTCCCAGTGCGGATGGACTTACTCAGATCGCGCGATCACGGTATGG
       LinJ36.2750  (1801) GTGTTTCCCAGTGCGGATGGACTTACTCAGATCGCGCGATCACGGTATGG
                          1851                                        1900
Cloned 36.2750 Seq  (1851) CGCGTCGCCAGCGCGTCTACTGCTGAAGCTGTGGAAGTTCCTTTCTGCCG
       LinJ36.2750  (1851) CGCGTCGCCAGCGCGTCTACTGCTGAAGCTGTGGAAGTTTCTTTCTGCCG
                          1901                                        1950
Cloned 36.2750 Seq  (1901) ACGCTTTTGCGAACGGGTATCAGGTAGCAGATGCCTTTGGAGATGACGCG
       LinJ36.2750  (1901) ACGCTTTTGCGAACGGGTATCAGGTAGCAGATGCCTTTGGAGATGACGCG
                          1951                                        2000
Cloned 36.2750 Seq  (1951) GTGTGGTGCGACTGCGGCGAGGCGGGCGACTCGGCCGCGCTACCAAAGCA
       LinJ36.2750  (1951) GTGTGGTGCGACTGCGGCGAGGCGGGCGACTCGGCCGCGCTACCAAAGCA
                          2001                                        2050
Cloned 36.2750 Seq  (2001) GCTTCCGCTCATGGCGTCGCTTGAACTGGCCACGATTGCGGGCATGAGCA
       LinJ36.2750  (2001) GCTTCCGCTCATGGCGTCGCTTGAACTGGCCACGATTGCGGGCATGAGCA
                          2051                                        2100
Cloned 36.2750 Seq  (2051) GCGCCGACCTCTCCGCTGCATCTTTTCGGCTTCCTGAGCTACTGGCACTC
       LinJ36.2750  (2051) GCGCCGACCTCTCCGCTGCATCTTTTCGGCTTCCTGAGCTACTGGCACTC
                          2101                                        2150
Cloned 36.2750 Seq  (2101) TTTCGCCACTCGAGCACGTCACTGGAGAGCATGCTCCGGTATCAGCGCTA
       LinJ36.2750  (2101) TTTCGCCACTCGAGCACGTCACTGGAGAGCATGCTCCGGTATCAGCGCTA
                          2151                                        2200
Cloned 36.2750 Seq  (2151) CCCGCGGCGCACCGTGCCAACGATGCTCTCTACCATCTTTGTCGACCCTC
       LinJ36.2750  (2151) CCCGCGGCGCACCGTGCCAACGATGCTCTCTACCATCTTTGTCGACCCTC
                          2201        2220
Cloned 36.2750 Seq  (2201) GGCTGCCGACTAACTCTTGA
       LinJ36.2750  (2201) GGCTGCCGACTAACTCTTGA
```

C LinJ20.1200 (Xylanase) Alignment

```
                          1                                           50
       LinJ20.1200     (1) ATGTCCGCCATCATCATCACACCCGTTGCTCCTGCGACGACGCCGCCCAC
Cloned Linf Xyl Seq    (1) ATGTCCGCCATCATCATCACACCCGTTGCTCCTGCGACGACGCCGCCCAC
                          51                                          100
       LinJ20.1200     (51) TCGCACCCCTGCAATGCATCTGCATAGCTTTAGCCAACCCACTCAAGCAC
Cloned Linf Xyl Seq    (51) TCGCACCCCTGCAATGCATCTGCATAGCTTTAGCCAACCCACTCAAGCAC
                         101                                          150
       LinJ20.1200    (101) AGCAGCAGCAGCAGCAGCTCCTTTTCCGGCCGCGCTGCGAGGAGGGGGTG
Cloned Linf Xyl Seq   (101) AGCAGCAGCAGCAGCAGCTCCTTTTCCGGCCGCGCTGCGAGGAGGGGGTG
                         151                                          200
       LinJ20.1200    (151) ACATATAGCGAAGACAACAACGGCACCGTGCACTACCGTTTCTACCTTCC
Cloned Linf Xyl Seq   (151) ACATATAGCGAAGACAACAACGGCACCGTGCACTACCGTTTCTACCTTCC
                         201                                          250
       LinJ20.1200    (201) CCATGCCAGCTCTGTTGTTGTCGCTCCTGTGAAGGTTTGCCTGGTGGACA
Cloned Linf Xyl Seq   (201) CCATGCCAGCTCTGTTGTTGTCGCTCCTGTGAAGGTTTGCCTGGTGGACA
                         251                                          300
       LinJ20.1200    (251) GCGATGGTGGCACGGTCCCTGTCGCCTCCATCGGCGCCGCCGCACCCATG
Cloned Linf Xyl Seq   (251) GCGATGGTGGCACGGTCCCTGTCGCCTCCATCGGCGCCGCCGCACCCATG
                         301                                          350
       LinJ20.1200    (301) ACAAAGCACCAAGACGGAGTGTGGGTCGGAACCGTGTCCGCACCAGTCGG
Cloned Linf Xyl Seq   (301) ACAAAGCACCAAGACGGAGTGTGGGTCGGAACCGTGTCCGCACCAGTCGG
                         351                                          400
       LinJ20.1200    (351) ACTGCAGTGTGTCGTCCTCATGGTGGACGGCAACCCTGTGCTCACGCCCC
Cloned Linf Xyl Seq   (351) ACTGCAGTGTGTCGTCCTCATGGTGGACGGCAACCCTGTGCTCACGCCCC
                         401                                          450
       LinJ20.1200    (401) ACCTCAGCATCGGGTGCCTGCATGGGCTCCAGCGGGCAAACTACATCGAT
Cloned Linf Xyl Seq   (401) ACCTCAGCATCGGGTGCCTGCATGGGCTCCAGCGGGCAAACTACATCGAT
                         451                                          500
       LinJ20.1200    (451) GTCCCGCCACCGAACCCAAATCGATGTGTCTACGCCATGCGGCCCTCCGT
Cloned Linf Xyl Seq   (451) GTCCCGCCACCGAACCCAAATCGATGTGTCTACGCCATGCGGCCCTCCGT
```

```
                                 501                                               550
           LinJ20.1200   (501)  CGAGCACGGGATGGTGGCCCACAACTACTTGACGTCCTACACAATGGACA
 Cloned Linf Xyl Seq     (501)  CGAGCACGGGATGGTGGCCCACAACTACTTGACGTCCTACACAATGGACA
                                 551                                               600
           LinJ20.1200   (551)  CAACCGAGGAAGTTCTTATCTACGTGCCGCCCTCGTATCACAAGGCGAGC
 Cloned Linf Xyl Seq     (551)  CAACCGAGGAAGTTCTTATCTACGTGCCGCCCTCGTATCACAAGGCGAGC
                                 601                                               650
           LinJ20.1200   (601)  AGCGCAACACGTCGGTACCCCGTGCTCTACCTCCTGCACGACGACCGCGA
 Cloned Linf Xyl Seq     (601)  AGCGCAACACGTCGGTACCCCGTGCTCTACCTCCTGCACGACGACCGCGA
                                 651                                               700
           LinJ20.1200   (651)  GTACCCAATGAACTGCGTACAGCAGGGTAAGGTGAACGTCATCGCCGACA
 Cloned Linf Xyl Seq     (651)  GTACCCAATGAACTGCGTACAGCAGGGTAAGGTGAACGTCATCGCCGACA
                                 701                                               750
           LinJ20.1200   (701)  ACCTCATCGCCGACGGCAAGATGACGGAGATGATCATCGTAATGAAGAGT
 Cloned Linf Xyl Seq     (701)  ACCTCATCGCCGACGGCAAGATGACGGAGATGATCATCGTAATGAAGAGT
                                 751                                               800
           LinJ20.1200   (751)  AGTGTGAGCGCACGCGCCAATGGTGAATGCATTCCATGCGATGCGGCCAA
 Cloned Linf Xyl Seq     (751)  AGTGTGAGCGCACGCACCAATGGTGAATGCATTCCATGCGATGCGGCCAA
                                 801                                               850
           LinJ20.1200   (801)  GCTCTGCGAAGACCTGACGGAGGACATCATTCCGTACGTCGACAACCACT
 Cloned Linf Xyl Seq     (801)  GCTCTGCGAAGACCTGACGGAGGACATCATTCCGTACGTCGACAACCACT
                                 851                                               900
           LinJ20.1200   (851)  ACCGCACCGAGGCAGATCGCGACAACCGCGCCATCGCCGGTCTCTACATG
 Cloned Linf Xyl Seq     (851)  ACCGCACCGAGGCAGATCGCGACAACCGCGCCATCGCCGGTCTCTACATG
                                 901                                               950
           LinJ20.1200   (901)  GGCTCCATACAAGCCAGCAGGCTCTGCATAACACGCCACGATCTCTTCGC
 Cloned Linf Xyl Seq     (901)  GGCTCCATACAAGCCAGCAGGCTCTGCATAACACGCCACGATCTCTTCGC
                                 951                                              1000
           LinJ20.1200   (951)  CTACGCCGGCATGTTCTCCGGCTTCCTGAGGAGCAACTGGAACGGCATCA
 Cloned Linf Xyl Seq     (951)  CTACGCCGGCATGTTCTCCGGCTTCCTGAGGAGCAACTGGAACGGCATCA
                                 1001                                             1050
           LinJ20.1200  (1001)  GTACGGACAGCGACCACATCGAAGCCCTCCGCCGCGATCCCGTAGCCTTC
 Cloned Linf Xyl Seq    (1001)  GTACGGACAGCGACCACATCGAAGCCCTCCGCCGCGATCCCGTAGCCTTC
                                 1051                                             1100
           LinJ20.1200  (1051)  CAGGCCGCCATGAAAGTGCTCTTCCGTTGCATTGGCGACGACAACACCCA
 Cloned Linf Xyl Seq    (1051)  CAGGCCGCCATGAAAGTGCTCTTCCGTTGCATTGGCGACGACAACACCCA
                                 1101                                             1150
           LinJ20.1200  (1101)  CCGCGCCGCGTTCGAGGCGGACGACGCTCTGCTGGCGGAGCTGGGCGTGG
 Cloned Linf Xyl Seq    (1101)  CCGCGCCGCGTTCGAGGCGGACGACGCTCTGCTGGCGGAGCTGGGCGTGG
                                 1151                                             1200
           LinJ20.1200  (1151)  CGTGCGAGCGGCGCATCTACGCGGGCTCACATAGCTGGCAGGTATGGCGC
 Cloned Linf Xyl Seq    (1151)  CGTGCGAGCGGCGCATCTACGCGGGCTCACATAGCTGGCAGGTATGGCGC
                                 1201                                             1250
           LinJ20.1200  (1201)  CAGGCCGCAGCCGATTTCCTGCCAATGCTATTCAAGGACCTAAGCTTTCT
 Cloned Linf Xyl Seq    (1201)  CAGGCCGCAGCCGATTTCCTGCCAATGCTATTCAAGGACCTAAGCTTTCT
                                 1251      1263
           LinJ20.1200  (1251)  TCCTCGTCACTGA
 Cloned Linf Xyl Seq    (1251)  TCCTCGTCACTGA
```

D LinJ36.2050 (SEC14) Alignment

```
                               1                                                  50
Cloned SEC14 Seq        (1)  ATGGCGGCAACTCATCTTACCTTTGATGATGCGAAGGTGGGCATGCGCGT
     LinJ36.2050        (1)  ATGGCGGCAACTCATCTTACCTTTGATGATGCGAAGGTGGGCATGCGCGT
                               51                                                100
Cloned SEC14 Seq       (51)  CCAGGATTACTGGGGCTGCTGCGGCACGCTACGTTGGATGGGGAAGCTCG
     LinJ36.2050       (51)  CCAGGATTACTGGGGCTGCTGCGGCACGCTACGTTGGATGGGGAAGCTCG
                               101                                               150
Cloned SEC14 Seq      (101)  AGAAGAACAATTCGCCTAACAAGGAAACAGGCAAGTTCTTCGGTATCGAG
     LinJ36.2050      (101)  AGAAGAACAATTCGCCTAACAAGGAAACAGGCAAGTTCTTCGGTATCGAG
                               151                                               200
Cloned SEC14 Seq      (151)  TACGACGACGAGAGTGACAATCCGCTGCGCAGCAATGGCACGTGGAACGG
     LinJ36.2050      (151)  TACGACGACGAGAGTGACAATCCGCTGCGCAGCAATGGCACGTGGAACGG
                               201                                               250
Cloned SEC14 Seq      (201)  CTGCAAGTACTTTGAGTGCGGGCCGCGTAAGGGCCGTCTTGTGAAGGTAG
     LinJ36.2050      (201)  CTGCAAGTACTTTGAGTGCGGGCCGCGTAAGGGCCGTCTTGTGAAGGTAG
                               251                                               300
Cloned SEC14 Seq      (251)  GCCAAGTTTACGCCGAGATCAACACCGAGCGGGTGGCGATGCTACGGGAG
     LinJ36.2050      (251)  GCCAAGTTTACGCCGAGATCAACACCGAGCGGGTGGCGATGCTACGGGAG
```

```
                          301                                                 350
Cloned SEC14 Seq    (301) CGCTTTGGGGAGCGCGTCGCGACATGGCACGACTTTGAGCTGGTGAAGTT
     LinJ36.2050    (301) CGCTTTGGGGAGCGCGTCGCGACATGGCACGACTTTGAGCTGGTGAAGTT
                          351                                                 400
Cloned SEC14 Seq    (351) CTGCATTGCGCGACAGTTCGATATGGAAAAGGTCTATGAGATGCTGGAGA
     LinJ36.2050    (351) CTGCATTGCGCGACAGTTCGATATGGAAAAGGTCTATGAGATGCTGGAGA
                          401                                                 450
Cloned SEC14 Seq    (401) GGCACCTGCAGTGGCGCGGGAGATTCCAGCCGTGCGCCGACGAGTACTTC
     LinJ36.2050    (401) GGCACCTGCAGTGGCGCGGGAGATTCCAGCCGTGCGCCGACGAGTACTTC
                          451                                                 500
Cloned SEC14 Seq    (451) CCTCAGACGATTCGTGAGGATTACCCGTGCGGCTACACAGGCACTACAGA
     LinJ36.2050    (451) CCTCAGACGATTCGTGAGGATTACCCGTGCGGCTACACAGGCACTACAGA
                          501                                                 550
Cloned SEC14 Seq    (501) TTACGACGAGAACCTCATTTACTGCGAACGCCCCGGCAATGCTGGCCACT
     LinJ36.2050    (501) TTACGACGAGAACCTCATTTACTGCGAACGCCCCGGCAATGCTGGCCACT
                          551                                                 600
Cloned SEC14 Seq    (551) GCCAACCGTCCGAGTTTGTGCGCAAGTACACGCTACCGGTAATTGCGCGG
     LinJ36.2050    (551) GCCAACCGTCCGAGTTTGTGCGCAAGTACACGCTACCGGTAATTGCGCGG
                          601                                                 650
Cloned SEC14 Seq    (601) TGGCACGCGTGCGCTATCGAGATGGGCATTGCGCGCATGCGCGCCACGAA
     LinJ36.2050    (601) TGGCACGCGTGCGCTATCGAGATGGGCATTGCGCGCATGCGCGCCACGAA
                          651                                                 700
Cloned SEC14 Seq    (651) CTACCGCTCCAAGCGAGTGTGCTGCATTGTGGATCTGTTGAACGTAAAAG
     LinJ36.2050    (651) CTACCGCTCCAAGCGAGTGTGCTGCATTGTGGATCTGTTGAACGTAAAAG
                          701                                                 750
Cloned SEC14 Seq    (701) CCATGTCGCGTTCGATGATCGGCTTTGCGCAGACGCTGGCGACGGTGGAG
     LinJ36.2050    (701) CCATGTCGCGTTCGATGATCGGCTTTGCGCAGACGCTGGCGACGGTGGAG
                          751                                                 800
Cloned SEC14 Seq    (751) CAGGACAACTACCCTGAGAACCTAGGGTGCGTCTTTATAGTGAACTGCCC
     LinJ36.2050    (751) CAGGACAACTACCCTGAGAACCTAGGGTGCGTCTTTATAGTGAACTGCCC
                          801                                                 850
Cloned SEC14 Seq    (801) CATGTTTTTCTGCTTTGCCTGGAAGCTGCTGAAGATCTTCATCGACGAAC
     LinJ36.2050    (801) CATGTTTTTCTGCTTTGCCTGGAAGCTGCTGAAGATCTTCATCGACGAAC
                          851                                                 900
Cloned SEC14 Seq    (851) GCACCAATAAAAAGATTAATTTCTGCGCTCCAAACAAGGCGGTGGAAGCG
     LinJ36.2050    (851) GCACCAATAAAAAGATTAATTTCTGCGCTCCAAACAAGGCGGTGGAAGCG
                          901                                                 950
Cloned SEC14 Seq    (901) ATGCTGCCGGTGATGCGGAAGGAGGACATACCGAACTTCTGCGGCGGGAC
     LinJ36.2050    (901) ATGCTGCCGGTGATGCGGAAGGAGGACATACCGAACTTCTGCGGCGGGAC
                          951                                                1000
Cloned SEC14 Seq    (951) CAGCAACAAGTGGATGGAAACGGCCAACGGCATCATCGGCTCGACAAACC
     LinJ36.2050    (951) CAGCAACAAGTGGATGGAAACGGCCAACGGCATCATCGGCTCGACAAACC
                         1001                                                1050
Cloned SEC14 Seq   (1001) CGAAAAAGGTCTACAGGGGCGAGGACTACAGCCCGCCCAGCATGACGAGC
     LinJ36.2050   (1001) CGAAAAAGGTCTACAGGGGCGAGGACTACAGCCCGCCCAGCATGACGAGC
                         1051                                                1100
Cloned SEC14 Seq   (1051) GAAGAGCTGAATGAGACGCAGTTGCGAGCCGATAGTGAGAGCCCCCACAG
     LinJ36.2050   (1051) GAAGAGCTGAATGAGACGCAGTTGCGAGCCGATAGTGAGAGCCCCCACAG
                         1101                                                1150
Cloned SEC14 Seq   (1101) GAGCCTCCGCGAAGGTGAGGCACCGACGACCACGCGCTTCATCGCGCCCG
     LinJ36.2050   (1101) GAGCCTCCGCGAAGGTGAGGCACCGACGACCACGCGCTTCATCGCGCCCG
                         1151                                                1200
Cloned SEC14 Seq   (1151) ACGCGGTGTCCTTGAGCTTCACGCCCACCAGTAGCTCCAAATCGCCTGAG
     LinJ36.2050   (1151) ACGCGGTGTCCTTGAGCTTCACGCCCACCAGTAGCTCCAAATCGCCTGAG
                         1201                                                1250
Cloned SEC14 Seq   (1201) GAGCCCAAGATGACGTCTGGGGCGCAGAGCAGCAGCGACACGGCAAGCGG
     LinJ36.2050   (1201) GAGCCCAAGATGACGTCTGGGGCGCAGAGCAGCAGCGACACGGCAAGCGG
                         1251                1281
Cloned SEC14 Seq   (1251) CAAGAAGCGAAAGAACGGTTTGCCGAAGTGA
     LinJ36.2050   (1251) CAAGAAGCGAAAGAACGGTTTGCCGAAGTGA
```

Figure 3.1.6: Alignment of the published sequence of genes with the sequence of cloned ORFs A) LinJ20.1200 (xylanase); B) LinJ36.2750 (hypothetical); C) LinJ28.0330 (hypothetical) and D) LinJ36.2050 (*SEC14*) on GeneDB compared with the sequence of the cloned ORF PCR product. Differences are highlighted by red arrowheads.

3.1.2 Cloning and Sequencing of the Four Genes from *L. donovani*

In order to confirm that these four genes are could be involved in disease tropism we had to confirm that the genes were intact in another species of *Leishmania* that causes the visceral form of the disease. The *L. donovani* strain BPK206/0 (Clone 10), is a field isolate from bone marrow biopsy of a 12 year old Nepalese patient in 2003. These cells were grown up in culture and genomic DNA was isolated for use as template in the PCR reactions. *L. donovani* ORFs for all four genes were amplified by PCR, using the same primers used to amplify the ORFs from *L. infantum* (see Table 3.1.1) together with the proof-reading polymerase Phusion (NEB). Two independent PCR reactions were carried out for each gene and A-overhangs were added to all PCR products. Both independent PCR products for each gene were gel-extracted and cloned into T-vectors before being sent sequencing in complete, double stranded form (the University of Dundee Sequencing Service). Note - the data presented in this section was obtained with the help and assistance of Elaine Brown, technician in the Mottram lab.

3.1.3 *L. donovani* ORF Sequence Analysis

Sequences were analysed using ContigExpress (Vector-NTI Advance 10, Invitrogen) and contiguous sequences were translated into protein sequences in Vector-NTI. These protein sequences were compared with the published *L. infantum* protein sequences using Align-X software (Vector-NTI Advance 10, Invitrogen) and these are shown in Figure 3.1.7.

A *L. infantum* versus *L. donovani* 28.0330 Proteins

```
                          1                                                  50
L. inf 28.0330      (1)   MADVQLSNSSNVKAVYAGHSLSLFMDLRPPAEVLATCSGSVVEDLPVTLS
L. don 28.0330      (1)   MADVQLSNSSNVKAVYAGHSLSLFMDLRPPAEVLATCSGSVVEDLPVTLS
     Consensus      (1)   MADVQLSNSSNVKAVYAGHSLSLFMDLRPPAEVLATCSGSVVEDLPVTLS
                          51                                                 100
L. inf 28.0330     (51)   VMLRRIHSESETPLSFRCSSSHRLLTYVQCEDGAGGDAVQLGMSGNNLLK
L. don 28.0330     (51)   VMLRRIHSESETPLSFRCSSSHRLLTYVQCEDGAGGDAVQLGMSGNNLLK
     Consensus     (51)   VMLRRIHSESETPLSFRCSSSHRLLTYVQCEDGAGGDAVQLGMSGNNLLK
                         101                                                 150
L. inf 28.0330    (101)   VTIDLATFPDRLQRLAAADVAAAEHALLDQSWVIGSRFRLSMQLRTRARA
L. don 28.0330    (101)   VTIDLATFPDRLQRLAAADVAAAEHALLDQSWVIGSRFRLSMQLRTRARA
     Consensus    (101)   VTIDLATFPDRLQRLAAADVAAAEHALLDQSWVIGSRFRLSMQLRTRARA
                         151                                                 200
L. inf 28.0330    (151)   EWPPLVAPQWSESSDGCLVPTPCSHHMDNHDSAGVMRMREEFAAKGVSAS
L. don 28.0330    (151)   EWPPLVAPQWSESSDGCLVPTPCSHHMGNHDSAGVMRMREEFAAKGVSAS
     Consensus    (151)   EWPPLVAPQWSESSDGCLVPTPCSHHM NHDSAGVMRMREEFAAKGVSAS
                         201                                                 250
L. inf 28.0330    (201)   PSIVRRNRPLSLYDKSAPRSQLKGPVYIDQELFPIVRILFVPQLYTEAIP
L. don 28.0330    (201)   PSIVRRNRPLSLYDKSAPRSQLKGPVYIDQELFPIVRILFVPQLYTEAIP
     Consensus    (201)   PSIVRRNRPLSLYDKSAPRSQLKGPVYIDQELFPIVRILFVPQLYTEAIP
                         251                                                 300
L. inf 28.0330    (251)   TATAVLRSMTVQVGVERCRESRGGKPLRRALRDRSTKDNVAVLDRVPASS
L. don 28.0330    (251)   TATAVLRSMTVQVGVERCRESRGGKPLRRALRDRSTKDNVAVLDRVPASS
     Consensus    (251)   TATAVLRSMTVQVGVERCRESRGGKPLRRALRDRSTKDNVAVLDRVPASS
                         301                                                 350
L. inf 28.0330    (301)   STLDEESKALLQPYEAICLAFNAHVRQPSMTSVTAYDVAHARRLVVQDYT
L. don 28.0330    (301)   STLDEESKALLQPYEAICLAFNAHVRQPSMTSVTAYDVAHARRLVVQDYT
     Consensus    (301)   STLDEESKALLQPYEAICLAFNAHVRQPSMTSVTAYDVAHARRLVVQDYT
                         351                                                 400
L. inf 28.0330    (351)   FEVMVTAARMKSETVETVQRAIQAYQYESGDADVLGMNLDEAVSSVEERF
L. don 28.0330    (351)   FEVMVTAARMKSETVETVQRAIQAYQYESGDADVLGMNLDEAVSSVEERF
     Consensus    (351)   FEVMVTAARMKSETVETVQRAIQAYQYESGDADVLGMNLDEAVSSVEERF
                         401                                                 450
L. inf 28.0330    (401)   GYMKDAPSVDLMSFHHVVWEAMRACMVADVTSSALANELEDELDRYEDRI
L. don 28.0330    (401)   GYMKDAPSVDLMSFHHVVWEAMRACMVADVTSSALANELEDELDRYEDRI
     Consensus    (401)   GYMKDAPSVDLMSFHHVVWEAMRACMVADVTSSALANELEDELDRYEDRI
                         451
L. inf 28.0330    (451)   NEILMDM-
L. don 28.0330    (451)   NEILMDM-
     Consensus    (451)   NEILMDM
```

B *L. infantum* versus *L. donovani* 36.2750 Proteins

```
                          1                                                  50
L. inf 36.2750      (1)   MGRIDSSCVAATVTPSKVGASSSCPGTATSTGKIDEPASRAAAPQGRAQR
L. don 36.2750      (1)   MGRIDSSCVAATVTPSKVGASSSCPGTATSTGKIDEPASRAAAPQGRAQR
 Consensus          (1)   MGRIDSSCVAATVTPSKVGASSSCPGTATSTGKIDEPASRAAAPQGRAQR
                          51                                                 100
L. inf 36.2750     (51)   CGRARPAQMHNSLSAGMCDNQQQHAAGVWKLIAGPQSASTTPPPLVMGDG
L. don 36.2750     (51)   CGRARPAQMHNGLSAGMCDNQQQHAAGVWQLIAGPQSASTTPPPLVMGDG
 Consensus         (51)   CGRARPAQMHN LSAGMCDNQQQHAAGVW LIAGPQSASTTPPPLVMGDG
                         101                                                 150
L. inf 36.2750    (101)   TGGEDHAQAAHTGVRWTKHAWEKGNDCPAGSAYGSFVGAPLRRLPPTPPF
L. don 36.2750    (101)   TGGEDHAQAAHAGVRWTKHSWEKGNDCPAGSAYGSFVGAPLRRLPPTPPF
 Consensus        (101)   TGGEDHAQAAH GVRWTKHAWEKGNDCPAGSAYGSFVGAPLRRLPPTPPF
                         151                                                 200
L. inf 36.2750    (151)   PGVRLPAWCRSASATAYPLSEDGLTNELLDFFYYLQLTSHEEAARERLLG
L. don 36.2750    (151)   PGVRLPAWCRSASATAYPLSEDGLTNELLDFFYYLQLTSHEEAARERLLG
 Consensus        (151)   PGVRLPAWCRSASATAYPLSEDGLTNELLDFFYYLQLTSHEEAARERLLG
                         201                                                 250
L. inf 36.2750    (201)   YVQACVAKLWGPCKPGSEAEGTAQVMLYGSYALGLSLPSSDIDLALTFPA
L. don 36.2750    (201)   YVQACVAKLWGPCKPGSEAEGTAQVMLYGSYALGLSLPSSDIDLALTFPA
 Consensus        (201)   YVQACVAKLWGPCKPGSEAEGTAQVMLYGSYALGLSLPSSDIDLALTFPA
                         251                                                 300
L. inf 36.2750    (251)   EEQVDVATIAVEERSGDRALLAPAVSKKRQALHLERLHDLAEQLRNSATF
L. don 36.2750    (251)   EEQVDVATIAVEERSGDRALLAPAVSKKRQALHLERLHDLAEQLRNSATF
 Consensus        (251)   EEQVDVATIAVEERSGDRALLAPAVSKKRQALHLERLHDLAEQLRNSATF
```

```
                    301                                                 350
L. inf 36.2750  (301) PELEVEVYDQCRVPRIHLRDRTCGGVSCDINSSFASARIARIVARQRLWL
L. don 36.2750  (301) PELEVEVYDQCRVPRIHLRDRTCGGVSCDINSSFASARIARIVARQRLWL
 Consensus      (301) PELEVEVYDQCRVPRIHLRDRTCGGVSCDINSSFASARIARIVARQRLWL
                    351                                                 400
L. inf 36.2750  (351) QDSPLAAFLVRVTKAAVKQWGLHEVFWGGVASTALYCLVLRFLAQMEQLC
L. don 36.2750  (351) QDSPLAAFLVRVTKAAVKQWGLHEVFWGGVASTALYCLVLRFLAQMEQLC
 Consensus      (351) QDSPLAAFLVRVTKAAVKQWGLHEVFWGGVASTALYCLVLRFLAQMEQLC
                    401                                                 450
L. inf 36.2750  (401) RHAQVEENLSPPRYAEAAAAHVATQLSSLACSSPLSSPQLSMPHPANESF
L. don 36.2750  (401) RHAQVEENLSPPRYEEAAAAHVATQLSSLACSSPLSSPQLSMPHPANESF
 Consensus      (401) RHAQVEENLSPPRY EAAAAHVATQLSSLACSSPLSSPQLSMPHPANESF
                    451                                                 500
L. inf 36.2750  (451) AAFTPVPAGTASTARGLNRLSAAAAVPSYVPSWVVNSRKHDLNKSSSLCD
L. don 36.2750  (451) AAFTPVPAGTASTARGLNRLSAAAAVPSYVPSWVVNSRKHGLSKSSSLCD
 Consensus      (451) AAFTPVPAGTASTARGLNRLSAAAAVPSYVPSWVVNSRKH L KSSSLCD
                    501                                                 550
L. inf 36.2750  (501) MDKNSALCPVSLTACSLSDWSTMSSTCDEDREADDEVDGYTQAGLTRTSG
L. don 36.2750  (501) MDKNSALCPVSVTACSLSDWSTMSSTCDEDREADDEVDGYTQAGLTRTSG
 Consensus      (501) MDKNSALCPVSLTACSLSDWSTMSSTCDEDREADDEVDGYTQAGLTRTSG
                    551                                                 600
L. inf 36.2750  (551) ATTTATTPISRTAASTPMPAASEAHEGGTATASSALAPPVSVNSAALREG
L. don 36.2750  (551) ATTTATTPISRTAASTPMPAASEAHEGGTATASSALAPPVSVNSAALREG
 Consensus      (551) ATTTATTPISRTAASTPMPAASEAHEGGTATASSALAPPVSVNSAALREG
                    601                                                 650
L. inf 36.2750  (601) VFPSADGLTQIARSRYGASPARLLLKLWKFLSADAFANGYQVADAFGDDA
L. don 36.2750  (601) VFPSADGLTHIARSRYGASPARLLLKLWKFLSADAFANGYQVADAFGDDT
 Consensus      (601) VFPSADGLT IARSRYGASPARLLLKLWKFLSADAFANGYQVADAFGDD
                    651                                                 700
L. inf 36.2750  (651) VWCDCGEAGDSAALPKQLPLMASLELATIAGMSSADLSAASFRLPELLAL
L. don 36.2750  (651) VWCDCGEAGDSAALPKQLPLMASLELATIAGMSSADLSAASFRLPELMAL
 Consensus      (651) VWCDCGEAGDSAALPKQLPLMASLELATIAGMSSADLSAASFRLPELLAL
                    701                                 740
L. inf 36.2750  (701) FRHSSTSLESMLRYQRYPRRTVPTMLSTIFVDPRLPTNS-
L. don 36.2750  (701) FRHSSTSLENMLRYQRYPRRTVPTMLSTIFVDPRLPTNS-
 Consensus      (701) FRHSSTSLE MLRYQRYPRRTVPTMLSTIFVDPRLPTNS
```

C *L. infantum* versus *L. donovani* 20.1200 (Xylanase) Proteins

```
                    1                                                   50
L. inf Xyl      (1) MSAIIITPVAPATTPPTRTPAMHLHSFSQPTQAQQQQQQLLFRPRCEEGV
L. don Xyl      (1) MSAIIITPAAPATTPPTRTPAMHLHSFSQPTQAQQQQQ-LLFRPRCEEGV
 Consensus      (1) MSAIIITP APATTPPTRTPAMHLHSFSQPTQAQQQQQ LLFRPRCEEGV
                    51                                                  100
L. inf Xyl     (51) TYSEDNNGTVHYRFYLPHASSVVVAPVKVCLVDSDGGTVPVASIGAAAPM
L. don Xyl     (50) TYSEDNNSTVHYRFYLPHASSVVVAPVKVCLVDSDGGTVPVASIGAAAPM
 Consensus     (51) TYSEDNN TVHYRFYLPHASSVVVAPVKVCLVDSDGGTVPVASIGAAAPM
                    101                                                 150
L. inf Xyl    (101) TKHQDGVWVGTVSAPVGLQCVVLMVDGNPVLTPHLSIGCLHGLQRANYID
L. don Xyl    (100) TKHQDGVWVGTVSAPVGLQCVVLMVDGNPVLTPHLSIGCLHGLQRANYID
 Consensus    (101) TKHQDGVWVGTVSAPVGLQCVVLMVDGNPVLTPHLSIGCLHGLQRANYID
                    151                                                 200
L. inf Xyl    (151) VPPPNPNRCVYAMRPSVEHGMVAHNYLTSYTMDTTEEVLIYVPPSYHKAS
L. don Xyl    (150) VPPPNPNRCVYAMRPSVEHGMVAHNYLTSYTMDTTEEVLIYVPPSYHKAS
 Consensus    (151) VPPPNPNRCVYAMRPSVEHGMVAHNYLTSYTMDTTEEVLIYVPPSYHKAS
                    201                                                 250
L. inf Xyl    (201) SATRRYPVLYLLHDDREYPMNCVQQGKVNVIADNLIADGKMTEMIIVMKS
L. don Xyl    (200) SATRRYPVLYLLHDDREYPMNCVQQGKVNVIADNLIADGKMTEMIIVMKS
 Consensus    (201) SATRRYPVLYLLHDDREYPMNCVQQGKVNVIADNLIADGKMTEMIIVMKS
                    251                                                 300
L. don Xyl    (250) SVSARANGECIPCDAAKLCEDLTEDIIPYVDNHYRTEADRDNRAIAGLYM
L. inf Xyl    (251) SVSARANGECIPCDAAKLCEDLTEDIIPYVDNHYRTEADRDNRAIAGLYM
 Consensus    (251) SVSARANGECIPCDAAKLCEDLTEDIIPYVDNHYRTEADRDNRAIAGLYM
                    301                                                 350
L. inf Xyl    (301) GSIQASRLCITRHDLFAYAGMFSGFLRSNWNGISTDSDHIEALRRDPVAF
L. don Xyl    (300) GSIQASRLCITRHDLFAYAGMFSGFLRSNWNGISTDSDHIEALRRDPVAF
 Consensus    (301) GSIQASRLCITRHDLFAYAGMFSGFLRSNWNGISTDSDHIEALRRDPVAF
                    351                                                 400
L. inf Xyl    (351) QAAMKVLFRCIGDDNTHRAAFEADDALLAELGVACERRIYAGSHSWQVWR
L. don Xyl    (350) QAAMKVLFRCIGDDNTHRAAFEADDALLAELGVACERRIYAGSHSWQVWR
 Consensus    (351) QAAMKVLFRCIGDDNTHRAAFEADDALLAELGVACERRIYAGSHSWQVWR
                    401             421
L. inf Xyl    (401) QAAADFLPMLFKDLSFLPRH-
L. don Xyl    (400) QAAADFLPMLFKDLSFLPRH-
 Consensus    (401) QAAADFLPMLFKDLSFLPRH
```

D *L. infantum* versus *L. donovani* 36.2050 (SEC14) Proteins

```
                    1                                                  50
L. inf SEC14      (1) MAATHLTFDDAKVGMRVQDYWGCCGTLRWMGKLEKNNSPNKETGKFFGIE
L. don SEC14      (1) MAATHLTFDDAKVGMRVQDYWGCCGTLRWMGKLEKDNSPNKETGKFFGIE
 Consensus        (1) MAATHLTFDDAKVGMRVQDYWGCCGTLRWMGKLEK NSPNKETGKFFGIE
                    51                                                 100
L. inf SEC14     (51) YDDESDNPLRSNGTWNGCKYFECGPRKGRLVKVGQVYAEINTERVAMLRE
L. don SEC14     (51) YDDESDNPLRSNGTWNGRKYFECGPRKGRLVKVGQVYAEINTERVAMLRE
 Consensus       (51) YDDESDNPLRSNGTWNG KYFECGPRKGRLVKVGQVYAEINTERVAMLRE
                   101                                                 150
L. inf SEC14    (101) RFGERVATWHDFELVKFCIARQFDMEKVYEMLERHLQWRGRFQPCADEYF
L. don SEC14    (101) RFGERVATWHDFELVKFCIARQFDMEKVYEMLERHLQWRGRFQPCVDEYF
 Consensus      (101) RFGERVATWHDFELVKFCIARQFDMEKVYEMLERHLQWRGRFQPC DEYF
                   151                                                 200
L. inf SEC14    (151) PQTIREDYPCGYTGTTDYDENLIYCERPGNAGHCQPSEFVRKYTLPVIAR
L. don SEC14    (151) PQTIREDYPCGYTGTTDYDENLIYCERPGNAGHCQPSEFVRKYTLPVIAR
 Consensus      (151) PQTIREDYPCGYTGTTDYDENLIYCERPGNAGHCQPSEFVRKYTLPVIAR
                   201                                                 250
L. inf SEC14    (201) WHACAIEMGIARMRATNYRSKRVCCIVDLLNVKAMSRSMIGFAQTLATVE
L. don SEC14    (201) WHACAIEMGIARMRATNYRSKRVCCIVDLLNVKAMSRSMIGFAQTLATVE
 Consensus      (201) WHACAIEMGIARMRATNYRSKRVCCIVDLLNVKAMSRSMIGFAQTLATVE
                   251                                                 300
L. inf SEC14    (251) QDNYPENLGCVFIVNCPMFFCFAWKLLKIFIDERTNKKINFCAPNKAVEA
L. don SEC14    (251) QDNYPENLGCVFIVNCPMFFCFAWKLLKIFIDERTNKKINFCAPNKAVEA
 Consensus      (251) QDNYPENLGCVFIVNCPMFFCFAWKLLKIFIDERTNKKINFCAPNKAVEA
                   301                                                 350
L. inf SEC14    (301) MLPVMRKEDIPNFCGGTSNKWMETANGIIGSTNPKKVYRGEDYSPPSMTS
L. don SEC14    (301) MLPVMRKEDIPNFCGGTSNKWMETANGIIGSTNPKKVYRGEDYSPPSMKS
 Consensus      (301) MLPVMRKEDIPNFCGGTSNKWMETANGIIGSTNPKKVYRGEDYSPPSM S
                   351                                                 400
L. inf SEC14    (351) EELNETQLRADSESPHRSLREGEAPTTTRFIAPDAVSLSFTPTSSSKSPE
L. don SEC14    (351) EELNETQSRADSESPHRSLREGEAPTTTRFIAPDAVSLSFTPTSSSKSPE
 Consensus      (351) EELNETQ RADSESPHRSLREGEAPTTTRFIAPDAVSLSFTPTSSSKSPE
                   401                 427
L. inf SEC14    (401) EPKMTSGAQSSSDTASGKKRKNGLPK-
L. don SEC14    (401) EPKMTSEAQSSSDTASGKKRKNGLPK-
 Consensus      (401) EPKMTS AQSSSDTASGKKRKNGLPK
```

Figure 3.1.7: Alignments of the four *L. infantum* proteins with *L. donovani* orthologues. A) 28.0330 alignment: A single amino acid difference between these proteins. B) 36.2750 alignment: 11 amino acid differences between *L. infantum* and *L. donovani* proteins. C) 20.1220 (xylanase) alignment: 3 amino acid differences. D) 36.2050 alignment: 6 amino acid differences.

Notably, all four genes are intact in *L. donovani*, another species that can cause visceral Leishmaniasis. This is an important finding, as it adds support to the hypothesis that the genes could be involved in influencing disease tropism. The primary sequence of each gene is also well conserved between the species.

3.1.4 Analysis of Equivalent *L. major* Loci

Attempts were made to amplify and clone fragments of DNA from *L. major*, equivalent to the positions where the *L. infantum*-specific genes exist to confirm that the *L. major* sequence held in GeneDB is correct and that the genes have indeed been lost rather than simply misassembled. Unfortunately insufficient time was available to repeat this and it will be carried out in the future.

3.2 Expression Profile in *L. infantum*

This work was carried out in collaboration with Dr. Jim Hilley of the University of Glasgow. Life cycle specific cDNA was prepare previously by Dr. Jim Hilley and sufficient amounts were provided for these experiments. Briefly, procyclic promastigotes, metacyclic promastigotes (purified by the peanut agglutination method) and amastigotes (extracted and purified from the spleen of an infected hamster) were lysed and stored at -80°C in Trizol (Invitrogen). Total RNA was extracted from Trizol according to manufacturer's instructions. A quantity of total RNA was used to prepare first strand cDNA by SuperscriptIII Reverse Transcriptase (Invitrogen) according to manufacturer's protocol.

3.2.1 RT-PCR of *L. infantum*-Specific Genes

Nested PCR was used as a way of increasing the specificity and sensitivity of the polymerase chain reaction. In the first round of PCR the primer combinations (see Table 3.2.1) were used with procyclic promastigote, metacyclic promastigote and amastigote first strand cDNA at an annealing temperature of 54°C. The forward primer in all reactions was OL1760, which is a primer that anneals to the splice leader sequence of processed mRNA molecules. The reverse primer in all cases was a gene-

specific primer. This is an added layer of control over non-specific PCR from contaminating genomic DNA since the splice leader sequence is not found immediately upstream of genes in genomic DNA, only in processed messenger RNA, and thus in cDNA. On completion, the first round reactions were diluted ten-fold with water and 1µl of these dilutions was used as a template in second round PCRs with different primer combinations (Table 3.2.1), under the same conditions as the first round PCR. The forward primer in the second round of PCR was, again OL1760, and was used with a gene-specific primer that anneals to a region within the amplified fragment in the first round PCR. This method has the added benefit of allowing the splice acceptor site to be identified and thus to confirm that the GeneDB sequence data held for each gene is annotated correctly i.e. that the correct start codon has been identified.

Table 3.2.1: Primer combinations used for nested PCR with cDNA to confirm expression for the four *L. infantum* genes and sequences of the primers used.

Gene	**First round reaction**	**Second round reaction**
LinJ28.0330	OL1760 + OL2101	OL1760 + OL2099
LinJ36.2750	OL1760 + OL2105	OL1760 + OL2103
LinJ20.1200 (Xylanase)	OL1760 + OL1987	OL1760 + OL2096
LinJ36.2050 (SEC14)	OL1760 + OL1985	OL1760 + OL2285
Primer	**Sequence**	
OL1760	AACTAACGCTATATAAGTATCAGTTTCTGTACTTTATTG	
OL1985	GCAGGATCCTCACTTCGGCAAACCGTTCTTTC	
OL1987	GCAGGATCCTCAGTGACGAGGAAGAAAGC	
OL2096	GACGTTCACCTTACCCTGCTG	
OL2099	CTTCCCGCATCCTCATAACGCC	
OL2101	CGTTAAAGGCGAGGCAGATG	
OL2103	CTCCCGTGCCGCTTCCTCATG	
OL2105	CGATGCGGGCGATGCGTG	
OL2285	CAAAGCGCTCCCGTAGCAT	

3.2.2 RT-PCR Results

The reactions were electrophoresed on an agarose gel and DNA fragments in the correct size range were detected (Figure 3.2.1). The exact size of the bands is unpredictable since the position of the splice acceptor site relative to the start of the gene is unknown. The sizes shown in the schematic of each locus (Figure 3.2.1A-D) refer to the size of the fragment from the start of the gene to the reverse primer used in the second round of PCR and amplicons should be similar to or larger than these since they will correspond to these fragments in addition to any upstream sequence between the splice acceptor site and the start codon.

Bands in the appropriate size range and of approximately equal intensity were produced from cDNA from each life cycle stage for all genes except LinJ20.1200, which appears to be strongly downregulated in amastigote forms (Figure 3.2.1 A-D).

In order to clarify these expression profiles in more detail, it would be necessary to carry out Quantitative PCR for each gene, relative to a house-keeping control gene, and this will be carried out in the future. Furthermore, the identity of the amplified fragments must also be confirmed by cloning and sequencing of the PCR products. This was done for LinJ36.2050, but has not yet been done for the other three genes. The identity of LinJ36.2050 (*SEC14*) has been confirmed in each life cycle stage and the splice acceptor and start codon have been identified. The annotated sequence on GeneDB has been confirmed (Figure 3.2.2).

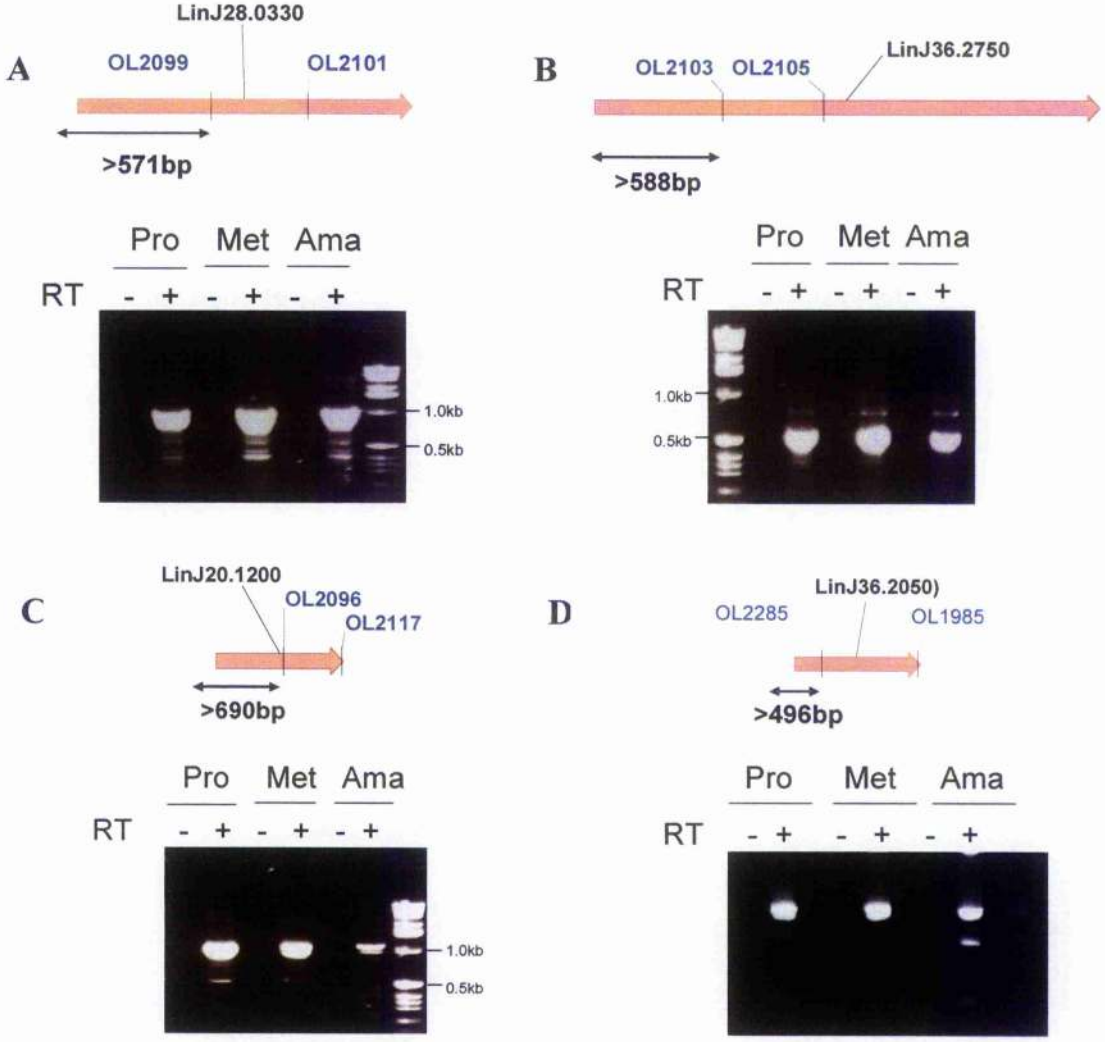


Figure 3.2.1: RT-PCR Analysis to determine expression profile of the four genes in different life cycle stages. Note the exact size of the bands is variable since the position of the splice acceptor site relative to the start of the gene is unknown. The band sizes shown refer to the size of the fragment from start of the gene to the reverse primer. Reaction conditions - Taq DNA polymerase was used with 10x PCR buffer:

94°C, 5 minutes	x 1
94°C, 1 minute 54°C, 4 minutes 72°C, 1 minute	x 25 cycles
72°C, 4 minutes	x 1

AACTAACGCTATATAAGTATCAGTTTCTGTACTTTATTGCGCGCGTGTGCGAGTGTC
TGTACGCCATTGAGTGCAGCTGTCGTGCTCGTACTGGGGACGCCCTTGTCCCATTTT
TTCCATATCGCTGCTCTGCCTTTAATTGCAGTGGCGTTGACTTACAATAAGTGCACA
AGCACGCACGCAAAGGACGTGTCACTCACCATGGCGGCAACTCATCTTACCTTTGAT
GATGCGAAGGTGGGCATGCGCGTCCAGGATTACTGGGGCTGCTGCGGCACGCTACGT
TGGATGGGGAAGCTCGAGAAGAACAATTCGCCTAACAAGGAAACAGGCAAGTTCTTC
GGTATCGAGTACGACGACGAGAGTGACAATCCGCTGCGCAGCAATGGCACGTGGAAC
GGCTGCAAGTACTTTGAGTGCGGGCCGCGTAAGGGCCGTCTTGTGAAGGTAGGCCAA
GTTTACGCCGAGATCAACACCGAGCGGGTGGCGATGCTACGGGAGCGCTTTG

Figure 3.2.2: The sequencing of the RT-PCR fragments for LinJ36.2050. The sequences were identical for PCR products for each life cycle stage. The sequence highlighted in yellow represents the position of the spliced leader primer, **OL1760** while the sequence highlighted in blue is OL2285. The start codon of the gene is highlighted in red (**ATG**). This is the first start codon downstream of the spliced leader sequence.

The presence of transcripts of all genes except LinJ20.1200 (xylanase) in all three life cycle stages indicates that these genes are expressed and they no longer need to be considered as hypothetical. Rather they are considered to be expressed genes of unknown function. Furthermore, the presence of amastigote transcripts means that a potential role for encoded proteins in the visceralisation mechanism is still a possibility. By contrast, LinJ20.1200 appears to be downregulated in amastigote stages meaning that it is less likely to be involved in a visceralisation mechanism. The quantification of transcripts at each life cycle stage by real-time quantitative PCR analysis will be carried out in the future.

3.3 Knockout of *SEC14* in *L. infantum* (JPCM5)

3.3.1 Preparation of *SEC14* Knockout Constructs

Leishmania is an organism that is amenable to the removal of genes through targeted gene disruption or replacement (Cruz et al., 1991; Cruz et al., 1993). Because *Leishmania* is diploid, there are two allelic copies of each gene and both must be replaced by sequential rounds of targeted gene replacement using two antibiotic-selectable markers (Figure 3.3.2) as has been done previously within the Mottram lab (Hilley et al., 2000; Denise et al., 2006).

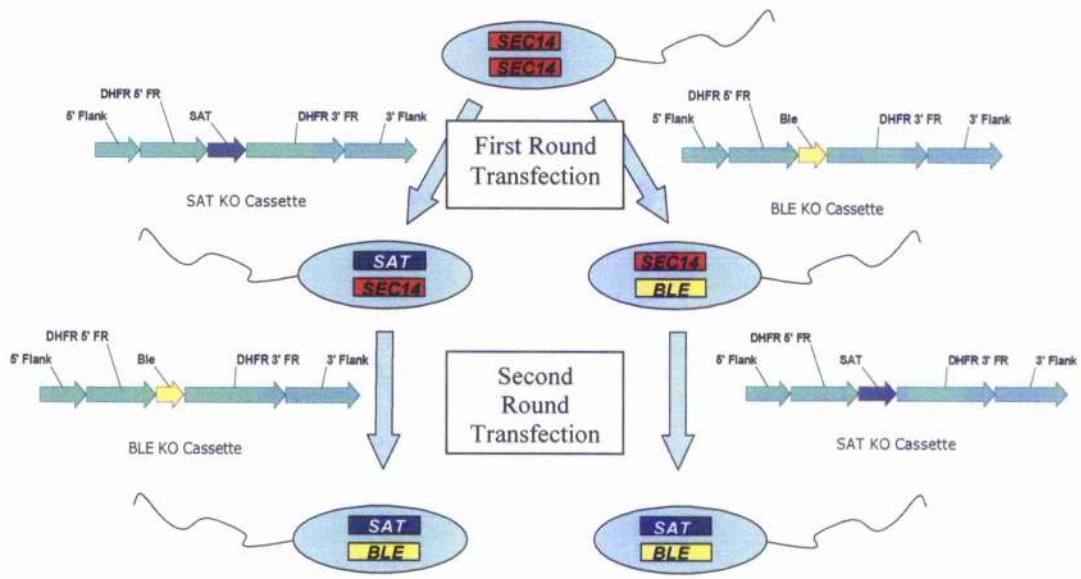


Figure 3.3.2: Overview of the SEC14 knockout process. The diagram shows the first allele being knocked out independently with each KO cassette (SAT and BLE) during the first round of transfection. The second round of transfection utilises the opposite KO cassette to the first round to replace the remaining copy of the gene.

In order to knock out both alleles of *SEC14* in *L. infantum*, two knockout constructs were produced. The starting point for these constructs was pGL1028, a knockout construct used previously to knock out a *Leishmania major* serine peptidase inhibitor protein, ISP1 in the Mottram lab (unpublished data). The first construct, pGL1534 (Figure 3.3.3) was made by dropping out the 5' (*Hin*dIII and *Sal*I) and 3' (*Xma*I and *Bgl*II) flanking regions of pGL1028 in turn and replacing them with 5' and 3' flanking regions of the *L. infantum* SEC14 gene. These flanking regions were produced by PCR and were first cloned into a T-vector before being subcloned into the knockout construct. Correct insertion of the 5' and 3' flanking regions was confirmed by PCR and sequencing. The second knockout construct was obtained by dropping out the SAT antibiotic resistance marker (using *Spe*I and *Bam*HI) and replacing it with the BLE antibiotic resistance marker. The BLE gene was obtained by digesting an independent knockout construct containing the BLE gene (pGL1422) with the same restriction enzymes (*Spe*I and *Bam*HI) and subcloning into pGL1534 to yield pGL1535 (Figure 3.3.3).

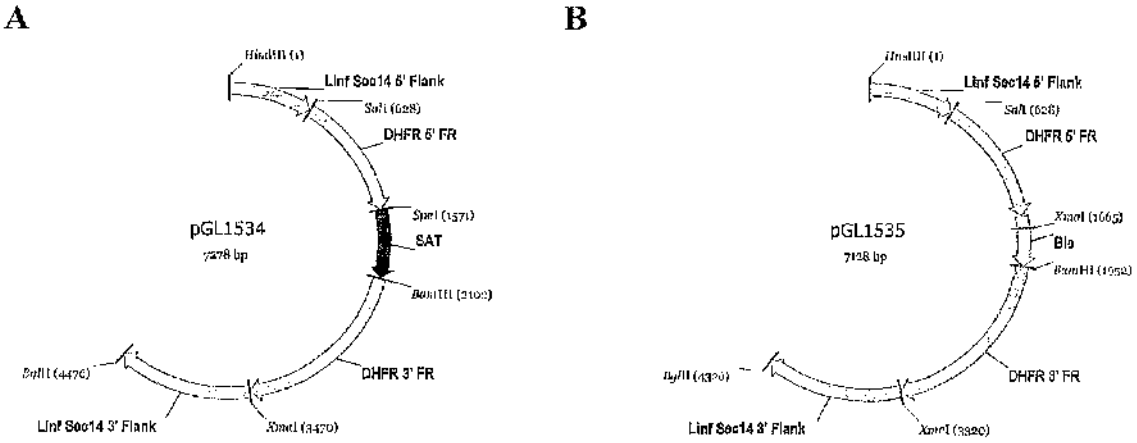


Figure 3.3.3: Schematic of constructs containing *SAT* and *BLE* genes flanked by *SEC14* 5` and 3` flanking regions. The constructs pGL1534 (A) and pGL1535 (B) were prepared by removing the existing 3` (*Xma*I and *Bgl*II) and 5` (*Hin*dIII and *Sal*I) flanks of a construct containing the SAT antibiotic resistance gene and replacing them with 3`and 5` *SEC14* flanking regions. Once obtained, plGL1534 was used to obtain pGL1535 by dropping out the antibiotic resistance gene (*Spe*I and *Bam*HI) and replacing it with a BLE-resistance gene from a similar KO construct used previously in the Mottram lab (pGL1442). This approach was taken because of an internal *Xma*I site within the BLE gene interfering with the cloning of the 3' flanking sequence.

3.3.2 Preparation for Targeted Gene Replacement of *SEC14*

In order to transfect efficiently, approximately 10µg of each knockout cassette was required. Medium-scale production of pGL1534 and pGL1535 was performed from 50ml of *E. coli* overnight cultures using the Qiagen HiSpeed Midi Kit. 40-60µg of each plasmid was digested overnight with 15µl each of *Hind*III and *Bgl*II in a 500µl reaction volume. Digests were electrophoresed on 0.7% agarose TBE gels. The

knockout cassette, which was approximately 4.4kb was extracted from the gel using multiple Qiagen Gel Extraction Kit columns. The eluted DNA from each column was pooled and ethanol-precipitated prior to transfection with wild type *L. infantum* cells.

3.3.2.1 First Round Transfection to Knockout SEC14 First Allele

The first round of transfection was carried out using the method of Beverley and Robinson (2003). Wild type *L. infantum* JPCM5 cells were transfected with either the SAT resistance cassette (from pGL1534) or the BLE resistance cassette (from pGL1535) and cultures were split immediately following transfection (S1 and S2). After several weeks of selection at 75μg/ml nourseothricin, resistant cells were growing and dividing well in both cultures, whereas the cells in the control transfections had died. The former were grown up and stabilates were made (WCMP 5599 and WCMP 5600). Additionally, BLE-resistant cells were also recovering from the first-round transfections with the BLE cassette, however these succumbed to contamination when growing up to make stabilates so were discarded and attention was focused on the SAT-resistant cells. Genomic DNA was isolated from transfected cells after a single sub-passage and subject to analysis for integration using PCR.

3.3.2.2 Check for Correct Integration by PCR

Primers were designed to anneal upstream (OL2300) and downstream (OL2301) of the 5` and 3` flanking regions of the knockout cassettes. These were used with

primers that hybridise within the knockout cassette itself to check for integration (Figure 3.3.4).

PCR using primers OL2301 and OL1031 has produced bands of the expected size (approximately 2.0kb) with genomic DNA obtained from both S1 and S2 cultures (Figure 3.3.4 C, Lanes 2 and 3). The additional bands visible in these reactions are presumably due to non-specific primer hybridisation as they are also present in the wild type control lane (Figure 3.3.4C, Lane 1). Primers OL2300 and OL1294 also gave PCR products of the predicted size of 0.7kb (Figure 3.3.4, Lanes 4 and 5) strongly suggesting successful integration.

These putative *SEC14/SAT* heterozygote cells were used in another round of transfection in an attempt to remove the second allele copy of the *SEC14* gene.

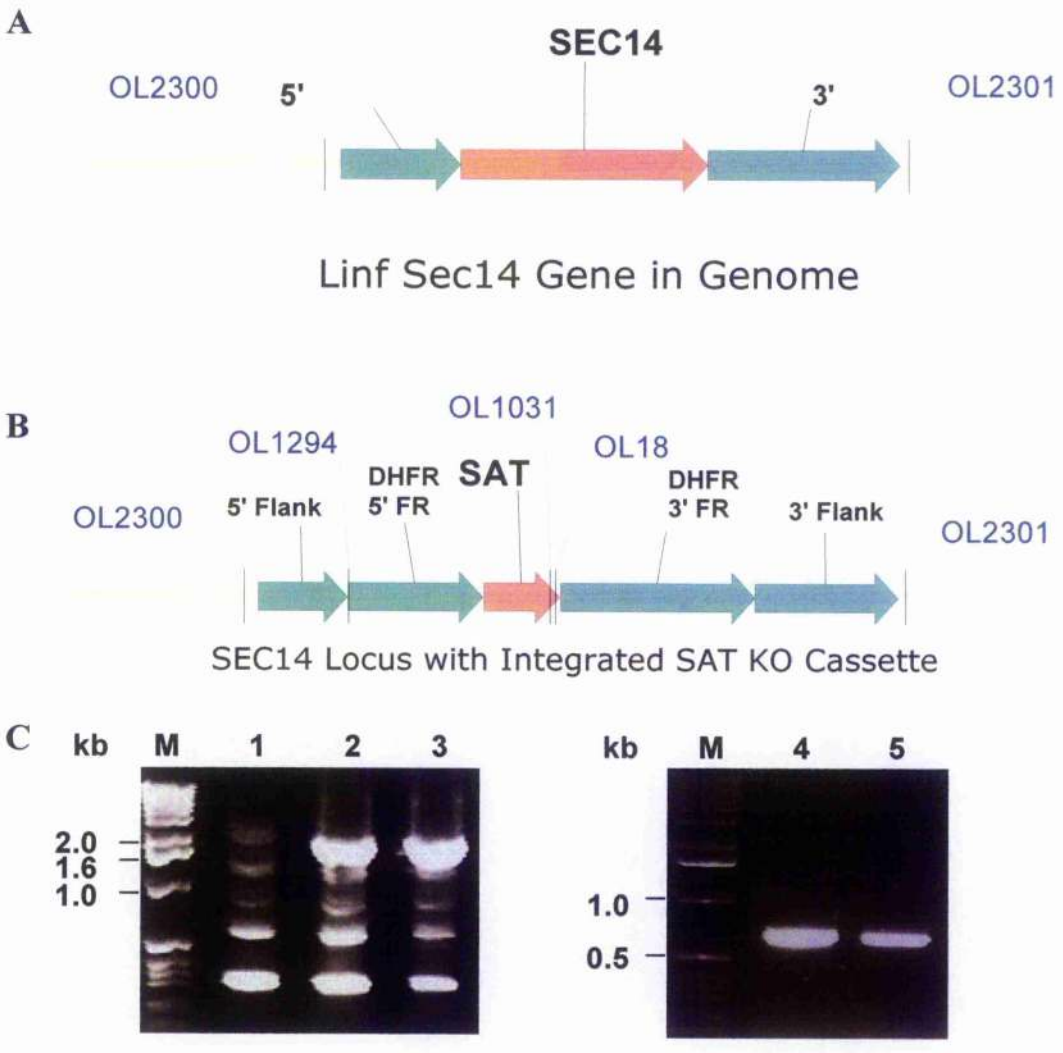


Figure 3.3.4: PCR analysis of clones resulting from first round transfection. SAT resistant clones were isolated and grown from stationary phase. 2µl of each lysate was used as DNA template in 20µl PCR reactions under the conditions described below:

95°C, 5 minutes	x1
95°C, 1 minute	x30
54°C, 1 minute	
72°C, 2 minutes	
72°C, 10minutes	x1

A) A map of the *L. infantum Sec14* gene in genome locus. B) Map of the *L.infantum SEC14* locus showing primer positions used in this PCR. C) Agarose gel showing the expected bands. Gel layout is as follows: Lane 1: Wild type OL2301+1031; Lane 2: S1-3`Flank region OL2301+1031; Lane 3: S2-3`Flank region OL2301+1031 (expected size is 2.0kb); Lane 4: S1-5`Flank region OL2300+1294; Lane 5: S2-5`Flank region OL2300+1294 (expected size is 0.7kb); Lane M DNA Size markers (1kb DNA Ladder (Invitrogen)). There are some multiple bands which are probably due to non-specific primer hybridisation.

3.3.2.3 Second Round Transfection to Completely Remove SEC14

The nourseothricin-resistant SEC14 +/- cells were transfected as before with the BLE knockout cassette. Immediately following transfection, the cultures were split (2x5ml cultures for each, except control transfections) and recovered overnight in the absence of antibiotics. 1ml of each overnight culture was added to 9ml of fresh HOMEM plus 10% FCS (plus penicillin/streptomycin) and 10% conditioned medium (CM) with 75μg/ml nourseothricin and 10μg/ml phleomycin. The remaining 4ml of the overnight cultures was added to 20ml of fresh HOMEM/FCS/CM plus antibiotics, as above (Dilution A). 2ml of Dilution A was added to a further 22ml of fresh HOMEM/FCS/CM plus antibiotics (Dilution B). Finally 2ml of Dilution B was added to another 22ml of fresh HOMEM/FCS/CM + antibiotics. Each dilution was plated out on 96 well plates in order to obtain clones.

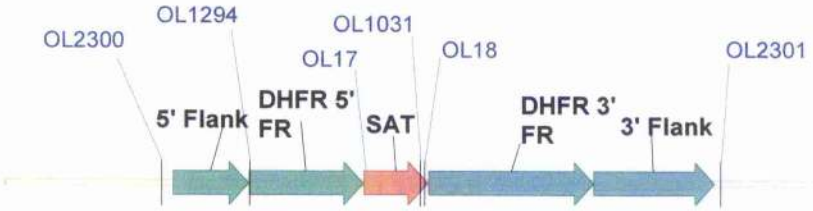
Three clones were obtained from the plate containing Dilution B (ΔSEC14-1, ΔSEC14-2 and ΔSEC14-3). These were grown up in fresh HOMEM plus 10% FCS and nourseothricin/phleomycin. These cultures sub-passaged and the remainder was

used to prepare stabilates (WCMP 5813, 5814 and 5815) and to isolate genomic DNA to test for integration by PCR (Figure 3.3.5).

PCRs on genomic DNA from wild type and ΔSEC14-1, 2 and 3 were carried out using a range of diagnostic primers to test for the correct integration of both KO cassettes (see Table 3.3.1 for details of primers as well as predicted fragment sizes). These PCRs shown in Figure 3.3.5 (D-K) strongly suggest that both KO cassettes have integrated correctly. Reactions D, G and H (Figure 3.3.5) are testing for the correct integration of the 5' portion of the KO cassettes. Reactions E, F, I and J are testing integration at the 3' end of the KO cassettes. Reaction K tests for the presence of a complete *SEC14* open reading frame. In most cases, PCR products corresponding to the predicted sizes (see Table 3.3.1) are present in all three ΔSEC14 clones but absent in the wild type negative control lanes. There are additional bands in many of the reactions but these are likely to be non-specific PCR products that are, in all cases, also present in the wild type lanes. Integration is best shown by the apparent loss of the *SEC14* ORFs shown in Figure 3.5.5K as well as the 5' integration of the SAT cassette (Figure 3.3.5D) and the 5' integration of the BLE cassette (Figure 3.3.5G).

It is likely that both copies of *SEC14* have been replaced with the two knockout cassettes and that SEC14 has been lost. This can be confirmed by Southern analysis.

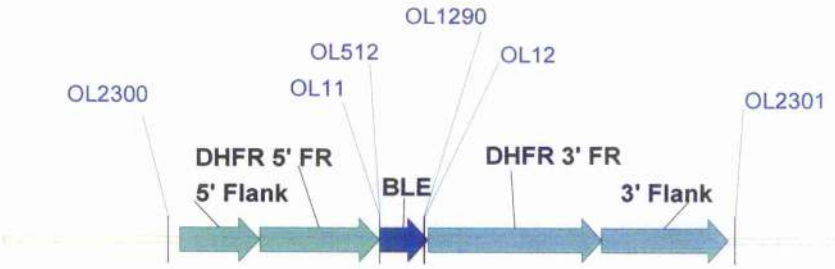
A



SEC14 Locus with Integrated SAT KO Cassette

6840 bp

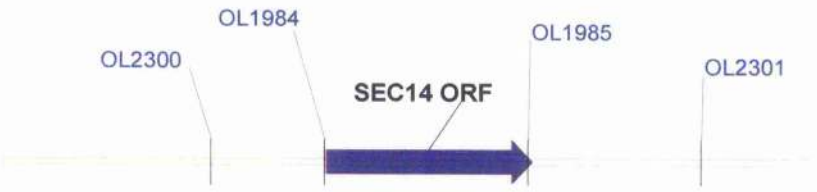
B



SEC14 Locus with Integrated BLE KO Cassette

6690 bp

C



Intact SEC14 Locus

5280 bp

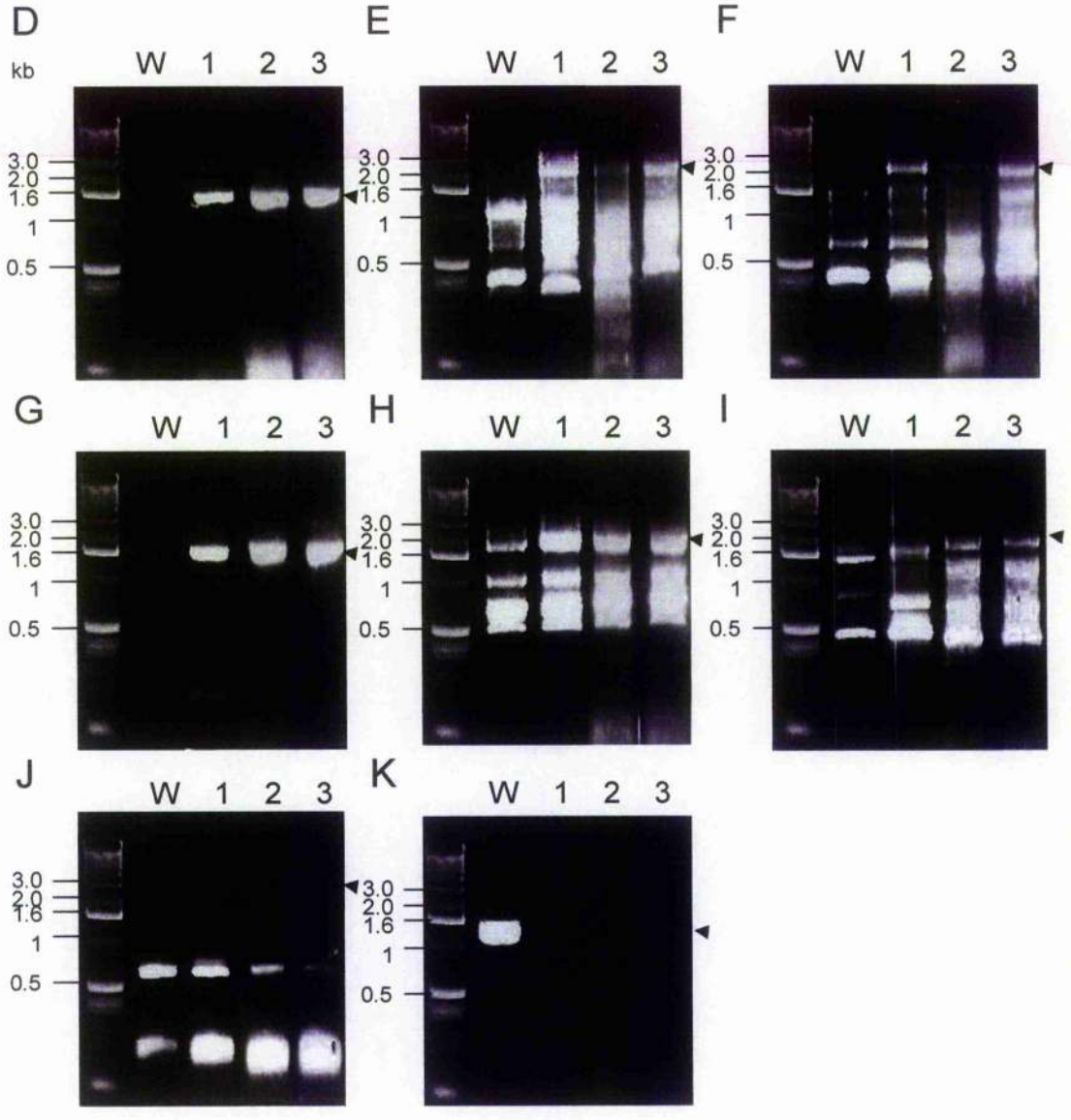


Figure 3.3.5: PCR tests to check for integration of both the SAT and BLE KO Cassettes at the *SEC14* Locus. A) The schematic of the *SEC14* locus with the SAT KO Cassette integrated. B) The schematic of the *SEC14* locus with the BLE KO Cassette integrated. C) The schematic of the *SEC14* locus with *SEC14* gene in place. The predicted fragment positions are marked by a black arrowhead in each case. Predicted fragment sizes and a summary of the above PCR experiments (D-K) are shown in Table 3.3.1. Lane W: Wild Type; Lane 1: Δ*SEC14* Clone 1; Lane 2: ΔSEC14 Clone 2; Lane 3: ΔSEC14 Clone 3

Table 3.3.1: Test PCRs to check for the integration of both the SAT and BLE KO cassettes in all three Δ*SEC14* clones, compared with wild type.

Figure 3.3.5 Ref	Primer Pair	KO Cassette Tested	Predicted Fragment Size (kb)	Positive Δ*SEC14* Clones
D	OL2300 OL17	SAT 5'	1.7	1, 2 and 3
E	OL2301 OL18	SAT 3'	2.5	1, 2 and 3
F	OL2301 OL1031	SAT 3'	2.5	1, 2 and 3
G	OL2300 OL11	BLE 5'	1.7	1, 2 and 3
H	OL2300 OL1290	BLE 5'	2.0	Possibly 1, 2 and 3
I	OL2301 OL12	BLE 3'	2.5	2 and 3
J	OL2301 OL512	BLE 3'	2.8	None
K	OL1984 OL1985	WT *SEC14* ORF	1.3	WT only

3.3.3.2 Southern blot analysis of *SEC14* mutants

In order to confirm the PCR data, Southern analysis was used to examine the *SEC14* locus to confirm integration of the knockout cassettes and the subsequent loss of the *SEC14* ORF. The sequence of the *SEC14* locus (the ORF and 10kb each of 5' and 3' flanking sequence) was obtained from the *L. infantum* database in GeneDB (http://www.genedb.org/genedb/linfantum/) and used in Vector NTi software (Invitrogen) to generate a restriction map of the locus (Figure 3.3.6).

Genomic DNA was extracted from a 30ml culture of stationary-phase wild type *L. infantum* (JPCM5) and the three putative Δ*SEC14* clones. Genomic DNA (3μg per digest) was digested overnight with either *Xho*I or HindIII. Digests were run out on a medium sized 0.7% agarose TBE gel and the gel was treated and blotted overnight by capillary transfer.

The blot was blocked and probed using a radiolabeled fragment corresponding to the *SEC14* 3' flanking region (Figure 3.3.6A). The blot was exposed to autoradiography film overnight and then for 3 days and is shown in Figure 3.3.6B.

The predicted fragment sizes for the various alleles are listed in Table 3.3.2.

Table 3.3.2: Predicted and Approximate Observed Fragment Size for Each Allele

	Digest			
	***Hind*III**		***Xho*I**	
Allele	**Predicted**	**Observed**	**Predicted**	**Observed**
Wild Type	3.2kb	~ 3.2kb	2.3kb	~2.3kb
SAT KO	4.8kb	~ 4.8kb	3.4kb	~ 3.4kb
BLE KO	4.7kb	~4.8kb	3.3kb	~3.3kb

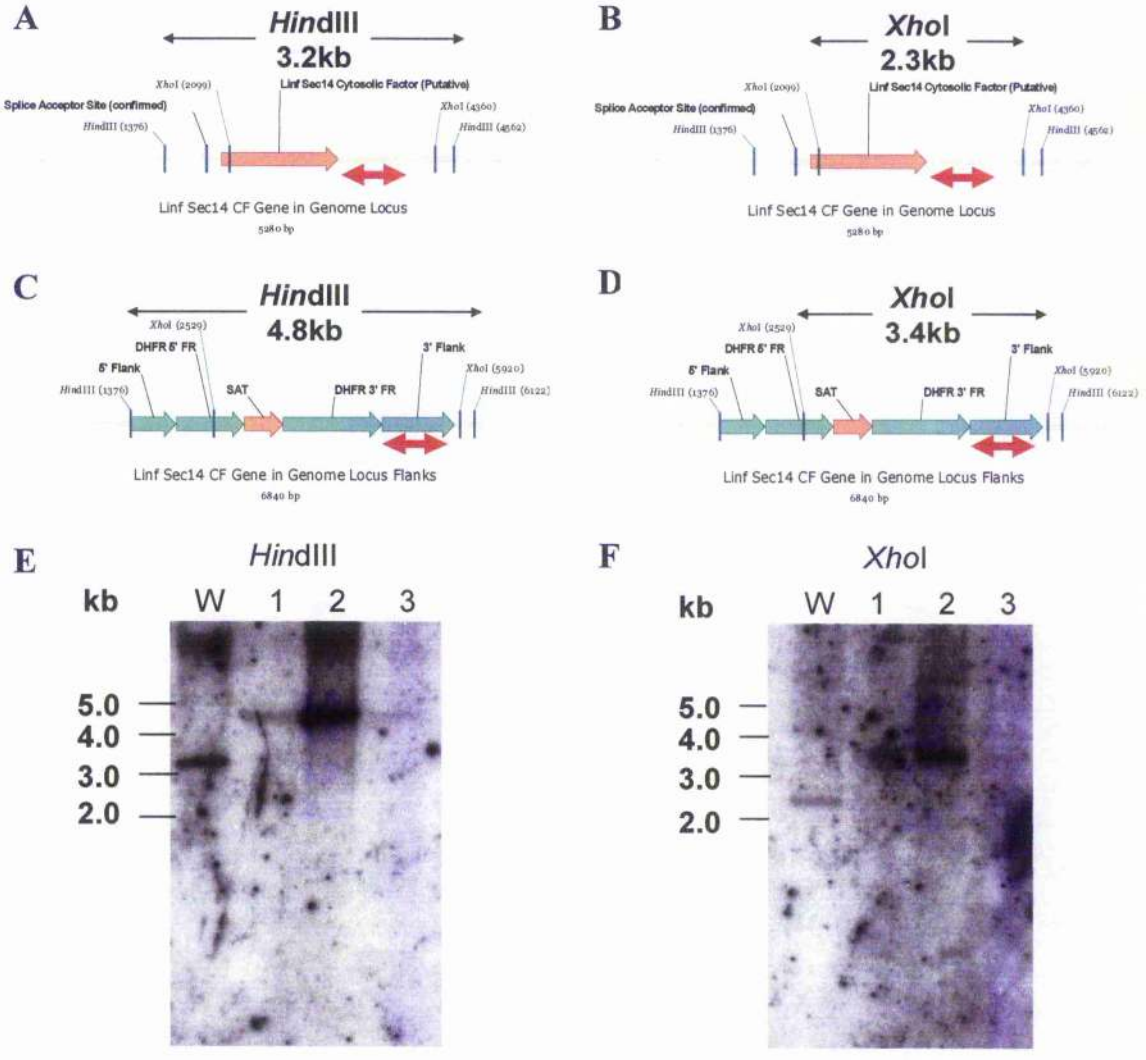


Figure 3.3.6: Southern analysis of Δ*SEC14* mutants. The schematic structure of the *SEC14* locus is containing the wild type gene showing the fragments expected when digesting with *Hind*III (A) and *Xho*I (B). The fragment sizes with the SAT KO cassette integrated correctly is also shown with predicted fragment sizes for *Hind*III (C) and *Xho*I (D). Genomic DNA from wild type *L. infantum* as well as each of the three ΔSEC14 clones was digested with *Hind*III or *Xho*I . Digests were run out on 0.7% agarose gels and the DNA was blotted onto Hibond-N membrane. A probe corresponding to part of the 3' flanking region (indicated by double headed red arrow

in A-D) was radiolabeled with 32P-CTP (Amersham) and the probe was hybridised with the membrane overnight at 65°C. Blots were washed with high stringency washing solution (0.1x SSC, 0.1% SDS) at 65°C then exposed to autoradiography film for 3 days. The *Hin*dIII and the *Xho*I Southern blots are shown in (E) and (F) respectively. The arrangement of lanes is the same for each blot:

Lane W: Wild type gDNA

Lane 1: Clone1 gDNA

Lane 2: Clone2 gDNA

Lane 3: Clone3 gDNA

The observed band sizes appear to match the predicted fragment sizes. The signal in the Clone 3 lane is very weak, although it was visible on the original autoradiogram. This is likely to be due to uneven loading of DNA between lanes as well as loss of resolution during image scanning. There is only a small difference in size between the SAT and BLE cassettes meaning that the bands are unlikely to be distinguishable from one another with these combinations of probe and restriction enzymes, hence the observed single band. When taken together with the PCR data above, this strongly indicates that the knockout cassettes have integrated at the correct position and successfully replaced both copies of *SEC14*, which has been lost from the three Δ*SEC14* clones.

Due to the well-documented plasticity of the *Leishmania* genome (Cruz et al., 1993; Smith et al., 2007) it is necessary to confirm that the ploidy of the three Δ*SEC14* clones is unaffected. This was done using Flow Cytometry (see Section 3.3.3.3).

3.3.3.3 DNA Content Analysis of Δ*SEC14* Mutants by FACS

Mid log-phase cells were methanol-fixed and stained with propidium iodide for FACS analysis to determine the DNA content of all three *L. infantum* Δ*SEC14* clones with respect to the DNA content of wild type *L. infantum*. The peaks corresponding to cells in G1 and G2 phase of the cell cycle in the mutant cells occur in the same position as those of wild type cells indicating that the DNA content of the mutants is normal (Figure 3.3.7). This suggests that *SEC14* is not an essential gene to promastigotes grown in culture.

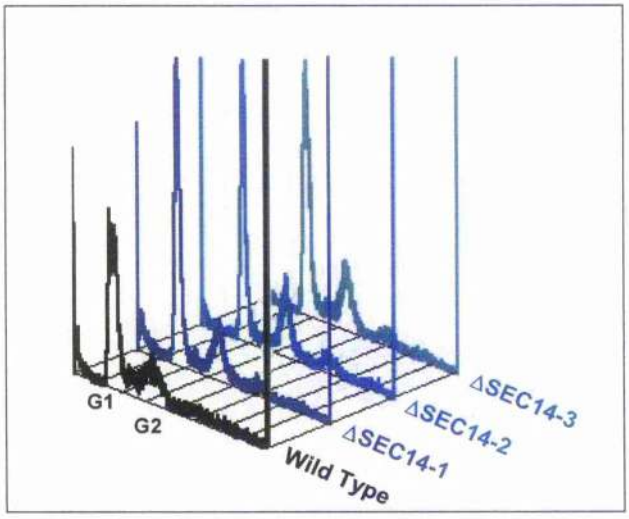


Figure 3.3.7: Analysis of the DNA content of wild type and Δ*SEC14* cells by Flow Cytometry. Wild type *L. infantum* cells together with 3 Δ*SEC14* knockout clones were propidium iodide stained and analysed by Flow Cytometry to determine if there were any abnormalities in the DNA content.

The *SEC14* knockout clones were then analysed further to determine if loss of *SEC14* results in a detectable phenotype in terms of growth and *in vitro* infectivity to macrophages.

3.4 Phenotype analysis of *SEC14* Null Mutants

3.4.1 Growth Curves

To determine the growth rate of ΔSEC14 promastigotes in culture, four 10ml replicate cultures were set up at 5 x 10^5 cells/ml for wild type and three clones (16 cultures in total). Cells were prepared for counting at 24h intervals for seven days by formaldehyde fixation and were counted using a haemocytometer. One observation was that Δ*SEC14*-1 consistently took longer to recover from the initial lag-phase but did recover and reached stationary phase at about the same time as the other knockout clones as well as wild type *L. infantum*. Overall, no significant variation between wild type and three mutants was noted (Figure 3.4.1).

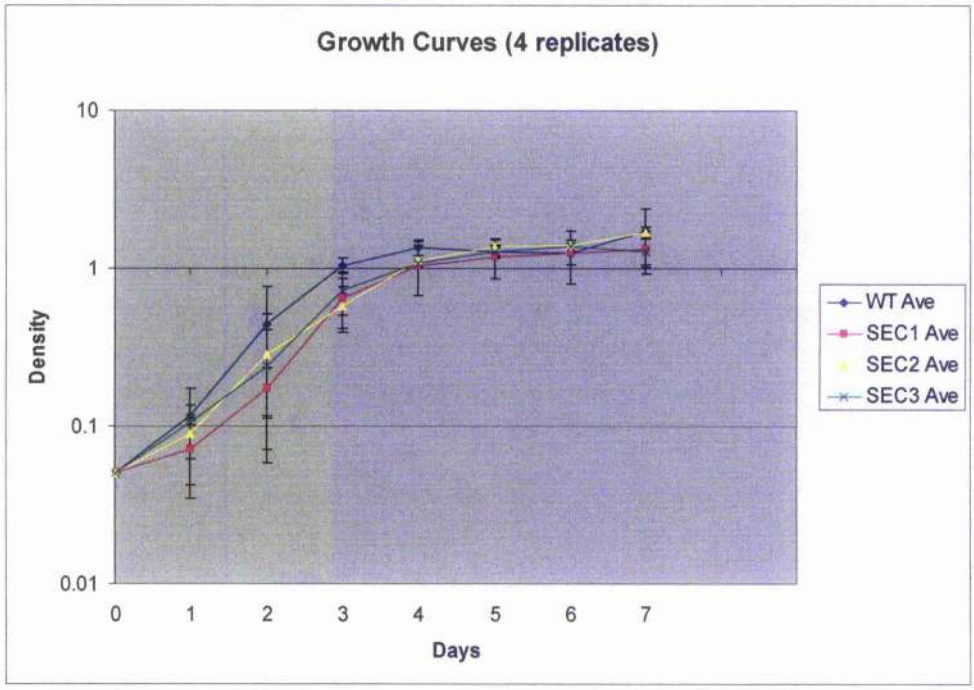


Figure 3.4.1: Growth curves for Δ*SEC14* clones 1-3 compared with wild type *L. infantum*. All growth cultures were set up at a density of 5 x 10^5 cells/ml. Cells were incubated at 25°C and counted at 24 hour intervals for seven days.

3.4.2 *In Vitro* Macrophage Infections with Δ*SEC14* Promastigotes

Intra-peritoneal macrophages were harvested from a single immunocompromised CD1/ICR1 mouse. These were counted and resuspended at a density of 5 x 10^5 cells/ml in RPMI medium + 10% FCS and allowed to adhere for 24 hours to the surface of 16-well chamber slides with covers (Labtek). After adhering, macrophages were washed in fresh RPMI/FCS and exposed to stationary phase wild type *L. infantum* and Δ*SEC14* 1-3 (four replicate wells for each cell line). After 24 hours, slides were washed with RPMI to remove free promastigotes and infection was allowed to proceed for 7 days. Three timepoints were taken (1 day, 4 days and 7

days post infection) at which time, macrophages and promastigotes were methanol-fixed. Once all timepoints had been taken and all wells were methanol-fixed, slides were stained with Giemsa for 10 minutes before rinsing with distilled water. 100 macrophages were counted for each well and the number of infected macrophages was recorded (Figure 3.4.2). The infection rate at day 1 and day 4 for wild type cells remained fairly constant at approximately 10% of macrophages infected with one or more *L. infantum* cells per macrophage. Giemsa staining of the day 7 samples revealed extensive damage to macrophages, which is likely to have been inflicted during preparation of the slides and were unfortunately too degraded to get meaningful data. This experiment will be repeated at a later date.

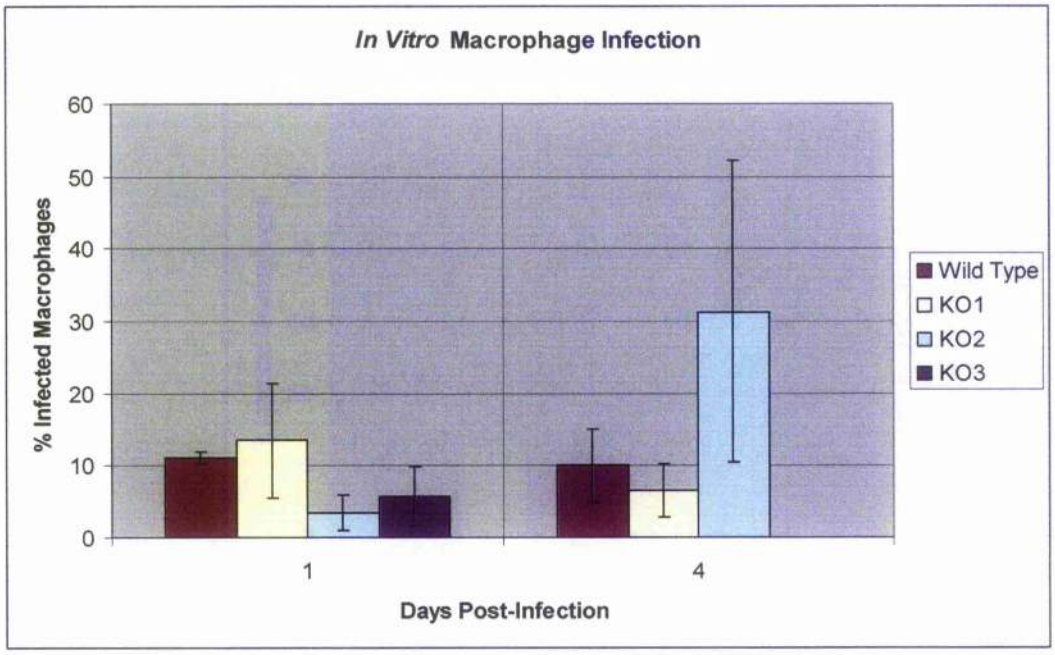


Figure 3.4.2: *In vitro* macrophage infection data. Intra-peritoneal macrophages from CD1/ICR1 mouse were infected with wild type and ΔSEC14 1-3 clones at a ratio of 8 promastigotes to1 macrophage as described in the text. Macrophages were fixed at 1 day, 4 days and 7 days post-infection and Giemsa-stained to allow

visualisation of infected macrophages. Note that the Day 4 time point for KO3 was not suitable for counting due to sample degradation and is not included in this Figure. Day 7 timepoints were also degraded and could not be accurately counted. Error bars are +/- standard deviations.

The infection rate at day 1 and day 4 for wild type cells remained fairly constant at approximately 10% of macrophages infected with one or more *L. infantum* cells per macrophage. The apparent reduction of the rate of macrophage infectivity by clone 2 falls within experimental error making it impossible to draw conclusions. Giemsa staining of the day 7 samples revealed extensive damage to macrophages, which is likely to have been inflicted during preparation of the slides and were unfortunately too degraded to get meaningful data. This experiment will be repeated at a later date. Alternative sources for macrophages will also be used in future experiments as mose peritoneal exudate-derived macrophages may not be the best source visceralising species of *Leishmania*.

3.4.3 Hamster Infections

The next stage was to test whether Δ*SEC14* promastigotes are capable of establishing infection in animals. The animal model for visceral infection is the Golden Syrian hamster. Three *L. infantum* cell lines, wild type and Δ*SEC14* clones 2 and 3 were each inoculated into two hamsters. Inoculations were carried out by intraperitoneal (i.p.) injection with 400µl of stationary phase promastigotes in PBS at a density of 2.5 x 10^7 cells/ml (a total of 10^7 stationary phase promastigotes per hamster). Spleens were extracted from one hamster for each cell line at 4 months post-injection and a small section of tissue from each was transferred to fresh HOMEM + 20%

FCS in order to determine if amastigotes could be recovered and differentiated to promastigotes. At 2 weeks post inoculation of the HOMEM culture medium, no promastigotes could be detected, however, these cultures will be incubated for longer to determine if promastigotes can eventually be recovered. At the time of writing, spleens of hamsters infected with wild type *L. infantum* and Δ*SEC14*-3 (also at 4 months post-inoculation) were extracted and used to inoculate culture medium as before. It remains to be seen whether promastigotes can be recovered from these tissue samples.

3.4.4 Mice Infections

In order to test the virulence of one of the Δ*SEC14* clones (Δ*SEC14*-1) and to see if the loss of *SEC14* results in cutaneous infection rather than visceral infection, six mice were inoculated in the right footpad with 20µl of stationary phase Δ*SEC14*-1 promastigotes at a density of 2.5 $x10^7$ ml of stationary phase promastigotes (a total of $5x10^5$ promastigotes per footpad). As a negative control for infection, a further six mice were footpad-inoculated with 40µl stationary phase wild type *L. infantum* at density of $2.5x10^7$ cells/ml. At 5 months post-inoculation, no lesions were observed at the inoculation site for either cell line in any of the mice and the experiment was terminated.

Loss of *SEC14* does not appear to cause a switch in tropism from visceral to cutaneous disease.

CHAPTER 4 - DISCUSSION

The aim of this project was to begin the analysis of four genes that are specific to *L. infantum* and that may have a role in the visceralisation mechanism. Analysis of the published sequence has revealed little about a potential role for LinJ28.0330, however the other hypothetical protein, LinJ36.2750 has a domain that has homology to the *S. cerevisiae* topoisomerase related function 4 protein (Trf4p). This protein is a component of the yeast nuclear exosome complex. The protein has been shown to have poly(A) RNA polymerase activity and is involved in the polyadenylation of tRNA and rRNA as well as small nucleolar RNAs (snoRNAs), which facilitates their degradation by the nuclear exosome complex (Haracska et al., 2005; LaCava et al., 2005). Reciprocal BLASTp searches of the *L. infantum* predicted proteins database have revealed several *Leishmania* genes that encode proteins with TRF4 domains and LinJ36.2750 is the third highest hit. The first two predicted proteins, LinJ07.0750 and LinJ26.0480 (known as LinJ07_V3.0780 and LinJ26_V3.0480 respectively in the most recent release of the *L. infantum* database in GeneDB (www.genedb.org), are also found in *L. major and L. braziliensis*.

Analysis of the amino acid sequence of LinJ36.2050 (SEC14) has shown a domain structure quite unlike any other known SEC14-like protein identified to date. Instead of the N-terminal CRAL_TRIO domain, which is an extension of the lipid binding SEC14 domain found in yeast Sec14p (Saito et al., 2007), the *L. infantum* protein has a CAP_GLY domain (Cytoskeleton Associated Protein). This domain has been identified and structurally characterised in proteins such as the mammalian cytoplasmic linker protein 170 (CLIP-170), where it mediates binding at the + end of

the microtubule (the end that is moving away from the microtubule organising centre (MTOC) (Mishima et al., 2007). This combination of CAP_GLY and SEC14 lipid binding domains might indicate a potential role in trafficking of vesicles within the cell. To date over 500 SEC14-like proteins all within eukaryotic species have been identified with roles in processes such as vesicle trafficking and phospholipid metabolism (Mousley et al., 2007; Phillips et al., 2006). Only two proteins with the CAP_GLY domain and the SEC14 domain have been identified to date: LinJ36.2050 and the orthologue of this gene in *Trypanosoma cruzi*. The reason why *T. cruzi* would require this protein is unclear, however it is important to note that *T. cruzi* also has an intracellular stage to its life cycle within mammalian hosts, so perhaps this is where the influence on disease tropism is taking place.

In order to determine if the GeneDB sequences were correct, the four genes were amplified by PCR, cloned and sequenced. The sequences for all four genes were a good match for the published sequences in GeneDB and where differences were observed, these could be explained by PCR-introduced mutations. To confirm this, an independent PCR clone for each ORF that had sequence variation, namely LinJ36.2750 and LinJ20.1200, should be sequenced and this will be carried out as part of follow up work on this project. These differences result in amino acid changes but not in frameshifts or the introduction of stop codons. If these genes were to have a potential role in the visceralisation mechanism, one would predict that the genes will be intact in other *Leishmania* species that cause VL. The four genes were therefore amplified and cloned from a strain of *L. donovani* that was available within the lab, BPK206/0. Cloning and analysis of two independent PCR products for each gene has revealed that the genes are indeed intact in *L. donovani* and that

they are well conserved between the species, indicating that they are likely have a biologically-relevant function to both species. It remains to be determined if this function forms part of the visceralisation mechanism. It is a formal albeit unlikely possibility that although these genes have been reported as pseudogenes or indeed missing for *L. major* genome project strain (MHOM/IL/80/Friedlin) that they may be intact in other strains of *L. major*, which would almost certainly rule out a role in visceralisation. Future work will involve analysing the loci of these genes in another strain of *L. major* that is available within the Mottram lab (MRHO/IR/75/ER, a strain isolated in Iran and chosen by the WHO as the vaccine challenge strain (Davoudi et al., 2005)) to confirm or otherwise, that these genes are also pseudogenes in *L. major* IR-75.

All four genes were shown to be expressed in procyclic and metacyclic promastigote stages of the life cycle and only LinJ20.1200 (endo-1,4-β-xylanase precursor) appeared to be downregulated in the amastigote stage. This strongly suggests that the requirement for an enzyme capable of degrading components of plant cell walls is not required by amastigote forms of *L. infantum*. This is in keeping with the hypothesis that this protein would most likely be required in the insect (promastigote) stages of the life cycle since it has been documented that the insect host, the female sand fly feeds on plant material between blood meals (Brito et al., 2006). In this respect it is noteworthy that a number genes encoding enzymes suitable for metabolising plant sugars have been identified in the *Leishmania* genomes, but are absent in *T. brucei* (Opperdoes and Coombs, 2007). This correlates with the feeding habits of tsetse, which are blood-sucking insects and do not feed on plant material. Whether the parasite utilises LinJ20.1200 for nutrition or for some

other purpose within the sand fly midgut and/or mouthparts remains to be determined.

LinJ36.2050 (*SEC14*) was confirmed as being expressed in the amastigote life cycle stage and was therefore selected as the main focus of the remainder of the project. *Leishmania infantum* Δ*SEC14* mutants were generated by two rounds of homologous recombination with drug selectable markers using established procedures (Hilley et al., 2000; Denise et al., 2006). As *SEC14* null mutants could be isolated, this indicates that *SEC14* is not essential to promastigotes growing in culture and indeed, growth curves revealed that cells are capable of growing and dividing in culture at a similar rate to wild type *L. infantum*. Furthermore, *in vitro* mouse peritoneal macrophage infections with all three Δ*SEC14* clones together with wild type has shown that mutant promastigotes can infect macrophages to about the same degree as wild type cells. This experiment must be repeated in order to determine if Δ*SEC14* cells can persist and proliferate as amastigotes in macrophages. The infection rate at day 1 and day 4 for wild type cells remained fairly constant at approximately 10% of macrophages infected with one or more *L. infantum* cells per macrophage. Any variation in this, for example, the apparent reduction in infection rates of clones 2 and 3, falls within the experimental error range and are therefore inconclusive. The extensive damage to macrophages of the day 7 timepoint means that it is not known whether amastigotes can persist and proliferate. This experiment will be repeated in the future, probably using alternative sources for macrophages. BALB/c mouse bone-marrow macrophages have been used successfully in *L. infantum in vitro* infection studies by collaborators (D.F. Smith, personal communication).

Δ*SEC14*-1 and wild type stationary phase promastigotes were inoculated into the footpads of 6 BALB/c mice each. No sign of cutaneous infection could be seen in any of these mice after 5 months, suggesting that the loss of *SEC14* does not cause a switch in tropism from visceral to cutaneous disease. BALB/c mice are not good mammalian hosts for *L. infantum*, so it is unsurprising that no infection was observed with either wild type or Δ*SEC14*-1 parasites. Whilst dogs provide an excellent experimental model for visceral disease with the genome strain JPC *L. infantum* (Poot et al., 2005), hamsters can also provide a suitable infection model (Denise Hubert et al., 2006). As it was not possible to carry out experiments on dogs in Glasgow, the Δ*SEC14*-2 and 3 mutants were each inoculated into 2 hamsters. One hamster containing Δ*SEC14*-2 was culled and a segment of spleen was removed to inoculate promastigote growth medium. At the time of writing (2 weeks after culture inoculation), no promastigotes could be detected in the growth medium, however it has been found within this laboratory that differentiation of *L. infantum* from spleen segments, can often take several weeks. Spleen samples have now been extracted from hamsters infected with wild type *L. infantum* and Δ*SEC14*-3 and transferred to growth medium. It remains to be seen whether promastigotes can be recovered from these tissue samples and until then, nothing can be concluded about the ability of Δ*SEC14* cells to infect hamsters.

To date, only one *L. donovani*–specific protein has been implicated in disease tropism – the A2 protein (Charest and Matlashewski, 1994). It has been shown that heterologous expression of the A2 gene in *L. major*, which lacks a functional copy of this gene, leads to an increased ability of *L. major* to survive within the spleen of

BALB/c mice as well as reduced ability to establish cutaneous infection (Zhang and Matlashewski, 2001; Zhang et al., 2003). This body of work indicates that disease tropism in *Leishmania*, although likely to involve numerous different proteins, can be influenced by the presence or absence of a single gene and is validation of the approach taken in this project. Similar to the A2 gene experiments, heterologous expression of the four *L. infantum*–specific genes in *L. major* will be carried out in future work.

By contrast with yeast Sec14p, which is essential, *L. infantum* LinJ36.2050 is not essential which raises the possibility that the encoded protein is not the true orthologue of the yeast protein. This is in agreement with unusual domain structure observed in *L. infantum* SEC14 and the discovery that there is another protein with a slightly higher degree of homology, as well as a more similar domain structure to yeast Sec14p (LinJ35.3630), that is also found in both *L. major* and *L. braziliensis* (LmjF35.3560 and LbrM32_V2.1420 respectively, see above). Yeast Sec14p is involved in the budding of secretory vesicles from the trans-Golgi network (TGN) to the plasma membrane (Bankaitis et al., 1989). One might predict that a protein involved in such a central process as secretory vesicle budding would be more likely to exist in all species of *Leishmania* rather than just *L. infantum*. It is possible that *L. infantum* has a subclass of LinJ36.2050-containing secretory vesicles that contain *L. infantum* specific virulence factors that influence disease progression.

Further experiments that will be carried out in time will involve targeted gene replacement of the two hypothetical genes, LinJ28.0330 and LinJ36.2750 and subsequent phenotype analysis of mutant lines. These genes as well as LinJ36.2050

will also be expressed heterologously in *L. major* to determine if the disease tropism of that species (CL) can be influenced. More detailed analysis of the role of LinJ36.2050 is ongoing and involves the localisation of the protein in *L. infantum* promastigotes by expression of tagged versions of the protein, as well as expression of recombinant protein for antibody production. The original *S. cerevisiae* temperature sensitive mutant has been provided by Professor Vytas Bankaitis of the University of North Carolina at Chapel Hill and yeast complementation studies will be used to determine if LinJ36.2050 and/or the putative true orthologue of Sec14p, LinJ35.3630 can rescue the yeast mutant phenotype.

While host immunity undoubtedly plays a part in tropism, it is likely that the parasites themselves have some form of control over disease progression. The ongoing project of which the presented work is a small part, aims to determine whether parasite factors indeed contribute to the disease tropism of *L. infantum* and to begin to unravel the roles of these species-specific genes.

REFERENCES

Bankaitis,V.A., Malehorn,D.E., Emr,S.D., and Greene,R. (1989). The *Saccharomyces cerevisiae* SEC14 gene encodes a cytosolic factor that is required for transport of secretory proteins from the yeast Golgi complex. J. Cell Biol. *108*, 1271-1281.

Banuls,A.L., Hide,M., and Prugnolle,F. (2007). *Leishmania* and the Leishmaniases: A Parasite Genetic Update and Advances in Taxonomy, Epidemiology and Pathogenicity in Humans. In Advances in Parasitology, J.R.Baker, ed. Academic Press), pp. 1-458.

Brito,N., Espino,J.J., and Gonzalez,C. (2006). The Endo-beta-1,4-Xylanase Xyn11A Is Required for Virulence in *Botrytis cinerea*. Molecular Plant-Microbe Interactions *19*, 25-32.

Campos-Ponce,M., Ponce,C., Ponce,E., and Maingon,R.D.C. (2005). *Leishmania chagasi/infantum*: further investigations on Leishmania tropisms in atypical cutaneous and visceral leishmaniasis foci in Central America. Experimental Parasitology *109*, 209-219.

Charest,H. and Matlashewski,G. (1994). Developmental gene expression in *Leishmania donovani*: differential cloning and analysis of an amastigote-stage-specific gene. Mol. Cell. Biol. *14*, 2975-2984.

Cruz,A., Coburn,C.M., and Beverley,S.M. (1991). Double Targeted Gene Replacement for Creating Null Mutants. PNAS *88*, 7170-7174.

Cruz,A.K., Titus,R., and Beverley,S.M. (1993). Plasticity in Chromosome Number and Testing of Essential Genes in *Leishmania* by Targeting. PNAS *90*, 1599-1603.

Davoudi,N., Tate,C.A., Warburton,C., Murray,A., Mahboudi,F., and McMaster,W.R. (2005). Development of a recombinant *Leishmania major* strain sensitive to ganciclovir and 5-fluorocytosine for use as a live vaccine challenge in clinical trials. Vaccine *23*, 1170-1177.

Denise Hubert, Poot Jacqueline, Jimenez Maribel, Ambit Audrey, Herrmann Daland C., Vermuelen Arno N., Coombs Graham H., and Mottram Jeremy C. (2006). Studies on the CPA cysteine peptidase in the *Leishmania infantum* genome strain JPCM5. BMC Molecular Biology *7*.

Desjeux,P. (1996). Leishmaniasis : Public health aspects and control. Clinics in Dermatology *14*, 417-423.

Eschenlauer,S.C.P., Coombs,G.H., and Mottram,J.C. (2006). PFPI-like genes are expressed in *Leishmania major* but are pseudogenes in other *Leishmania* species. FEMS Microbiology Letters *260*, 47-54.

Handman,E. and Bullen,D.V.R. (2002). Interaction of *Leishmania* with the host macrophage. Trends in Parasitology *18*, 332-334.

Haracska,L., Johnson,R.E., Prakash,L., and Prakash,S. (2005). Trf4 and Trf5 Proteins of *Saccharomyces cerevisiae* Exhibit Poly(A) RNA Polymerase Activity but No DNA Polymerase Activity. Mol. Cell. Biol. *25*, 10183-10189.

Herwaldt,B.L. (1999). Leishmaniasis. The Lancet *354*, 1191-1199.

Hilley,J.D., Zawadzki,J.L., McConville,M.J., Coombs,G.H., and Mottram,J.C. (2000). *Leishmania mexicana* Mutants Lacking Glycosylphosphatidylinositol (GPI):Protein Transamidase Provide Insights into the Biosynthesis and Functions of GPI-anchored Proteins. Mol. Biol. Cell *11*, 1183-1195.

Hsuan,J. and Cockcroft,S. (2001). The PITP family of phosphatidylinositol transfer proteins. Genome Biology *2*.

Ilgoutz,S.C. and McConville,M.J. (2001). Function and assembly of the *Leishmania* surface coat. International Journal for Parasitology *31*, 899-908.

Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R., Anupama,A., Apostolou,Z., Attipoe,P., Bason,N., Bauser,C., Beck,A., Beverley,S.M., Bianchettin,G., Borzym,K., Bothe,G., Bruschi,C.V., Collins,M., Cadag,E., Ciarloni,L., Clayton,C., Coulson,R.M.R., Cronin,A., Cruz,A.K., Davies,R.M., De Gaudenzi,J., Dobson,D.E., Duesterhoeft,A., Fazelina,G., Fosker,N., Frasch,A.C., Fraser,A., Fuchs,M., Gabel,C., Goble,A., Goffeau,A., Harris,D., Hertz-Fowler,C., Hilbert,H., Horn,D., Huang,Y., Klages,S., Knights,A., Kube,M., Larke,N., Litvin,L., Lord,A., Louie,T., Marra,M., Masuy,D., Matthews,K., Michaeli,S., Mottram,J.C., Muller-Auer,S., Munden,H., Nelson,S., Norbertczak,H., Oliver,K., O'Neil,S., Pentony,M., Pohl,T.M., Price,C., Purnelle,B., Quail,M.A., Rabbinowitsch,E., Reinhardt,R., Rieger,M., Rinta,J., Robben,J., Robertson,L., Ruiz,J.C., Rutter,S., Saunders,D., Schafer,M., Schein,J., Schwartz,D.C., Seeger,K., Seyler,A., Sharp,S., Shin,H., Sivam,D., Squares,R., Squares,S., Tosato,V., Vogt,C., Volckaert,G., Wambutt,R., Warren,T., Wedler,H., Woodward,J., Zhou,S., Zimmermann,W., Smith,D.F., Blackwell,J.M., Stuart,K.D., Barrell,B., and Myler,P.J. (2005). The Genome of the Kinetoplastid Parasite, *Leishmania major*. Science *309*, 436-442.

Kelleher,M., Moody,S.F., Mirabile,P., Osborn,A.H., Bacic,A., and Handman,E. (1995). Lipophosphoglycan blocks attachment of *Leishmania major* amastigotes to macrophages. Infect. Immun. *63*, 43-50.

LaCava,J., Houseley,J., Saveanu,C., Petfalski,E., Thompson,E., Jacquier,A., and Tollervey,D. (2005). RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. Cell *121*, 713-724.

Lipoldova,M. and Demant,P. (2006). Genetic susceptibility to infectious disease: lessons from mouse models of leishmaniasis. Nat Rev Genet *7*, 294-305.

Mauricio,I.L., Stothard,J.R., and Miles,M.A. (2000). The Strange Case of *Leishmania chagasi*. Parasitology Today *16*, 188-189.

McMahon-Pratt,D. and Alexander,J. (2004). Does the *Leishmania major* paradigm of pathogenesis and protection hold for New World cutaneous leishmaniases or the visceral disease? Immunological Reviews *201*, 206-224.

Mishima,M., Maesaki,R., Kasa,M., Watanabe,T., Fukata,M., Kaibuchi,K., and Hakoshima,T. (2007). Structural basis for tubulin recognition by cytoplasmic linker protein 170 and its autoinhibition. PNAS *104*, 10346-10351.

Mousley,C.J., Tyeryar,K.R., Vincent-Pope,P., and Bankaitis,V.A. (2007). The Sec14-superfamily and the regulatory interface between phospholipid metabolism and membrane trafficking. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids *1771*, 727-736.

Opperdoes,F.R. and Coombs,G.H. (2007). Metabolism of *Leishmania*: proven and predicted. Trends in Parasitology *23*, 149-158.

Pasiecznik N.M, Felker P, Harris P.J.C., Harsh L.N., Cruz G., Tewari J.C., Cadoret K., and Maldonado L.J. The *Prosopis juliflora - Prosopis pallida* Complex: A Monograph. 1-179. 2001. Coventry, UK, HDRA.
Ref Type: Serial (Book,Monograph)

Peacock,C.S., Seeger,K., Harris,D., Murphy,L., Ruiz,J.C., Quail,M.A., Peters,N., Adlem,E., Tivey,A., Aslett,M., Kerhornou,A., Ivens,A., Fraser,A., Rajandream,M.A., Carver,T., Norbertczak,H., Chillingworth,T., Hance,Z., Jagels,K., Moule,S., Ormond,D., Rutter,S., Squares,R., Whitehead,S., Rabbinowitsch,E., Arrowsmith,C., White,B., Thurston,S., Bringaud,F., Baldauf,S.L., Faulconbridge,A., Jeffares,D., Depledge,D.P., Oyola,S.O., Hilley,J.D., Brito,L.O., Tosi,L.R.O., Barrell,B., Cruz,A.K., Mottram,J.C., Smith,D.F., and Berriman,M. (2007). Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat Genet *39*, 839-847.

Phillips,S.E., Vincent,P., Rizzieri,K.E., Schaaf,G., Bankaitis,V.A., and Gaucher,E.A. (2006). The Diverse Biological Functions of Phosphatidylinositol Transfer Proteins in Eukaryotes. Critical Reviews in Biochemistry and Molecular Biology *41*, 21-49.

Poot,J., Rogers,M.E., Bates,P.A., and Vermeulen,A. (2005). Detailed analysis of an experimental challenge model for *Leishmania infantum* (JPC strain) in dogs. Veterinary Parasitology *130*, 41-53.

Saito,K., Tautz,L., and Mustelin,T. (2007). The lipid-binding SEC14 domain. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids *1771*, 719-726.

Schlein,Y. and Jacobson,R.L. (1995). Assessment of plant tissue feeding by sand flies (Diptera: Psychodidae) and mosquitoes (Diptera: Culicidae). Journal of Medical Entomology *32*, 882-887.

Schlein,Y. and Jacobson,R.L. (2002). Linkage between susceptibility of *Phlebotomus papatasi* to *Leishmania major* and hunger tolerance. Parasitology *125*, 343-348.

Siriwardana H.V.Yamuna .D., Noyes Harry A., Beeching Nicholas J., Chance Michael L., and Bates,P.A. (2007). *Leishmania donovani* and Cutaneous Leishmaniasis, Sri Lanka. Emerging Infectious Diseases *13*, 476-478.

Smith,D.F., Peacock,C.S., and Cruz,A.K. (2007). Comparative genomics: From genotype to disease phenotype in the leishmaniases. International Journal for Parasitology *37*, 1173-1186.

Spath,G.F., Garraway,L.A., Turco,S.J., and Beverley,S.M. (2003). The role(s) of lipophosphoglycan (LPG) in the establishment of *Leishmania major* infections in mammalian hosts. PNAS *100*, 9536-9541.

van Zandbergen,G., Klinger,M., Mueller,A., Dannenberg,S., Gebert,A., Solbach,W., and Laskay,T. (2004). Cutting Edge: Neutrophil Granulocyte Serves as a Vector for *Leishmania* Entry into Macrophages. J Immunol *173*, 6521-6525.

Zhang,W.W. and Matlashewski,G. (1997). Loss of virulence in *Leishmania donovani* deficient in an amastigote-specific protein, A2. PNAS *94*, 8807-8811.

Zhang,W.W. and Matlashewski,G. (2001). Characterization of the A2-A2rel gene cluster in *Leishmania donovani*: involvement of A2 in visceralization during infection. Molecular Microbiology *39*, 935-948.

Zhang,W.W., Mendez,S., Ghosh,A., Myler,P., Ivens,A., Clos,J., Sacks,D.L., and Matlashewski,G. (2003). Comparison of the A2 Gene Locus in *Leishmania donovani* and *Leishmania major* and Its Control over Cutaneous Infection. J. Biol. Chem. *278*, 35508-35515.