THE INTRASPECIFIC GENETIC DIVERSITY OF *TRYPANOSOMA CRUZI* CELL SURFACE ANTIGENS. USING A NOVEL PCR APPROACH

G. A. Makalliwa[1], W. M. Gaunt[2] and E. Mitaki[3]

[1]*College of Health Sciences, Jomo Kenyatta University Agriculture and Technology, Nairobi, Kenya*
[2]*London School of Hygiene and Tropical Medicine, London, England, United Kingdom*
[3]*Department of Health Sciences, Kisii CBD Campus Jomo Kenyatta University Agriculture and Technology, Kisii, Kenya*
E-mail: george_makalliwa@yahoo.co.uk

Abstract

Trypanosoma cruzi the causative agent of Chagas disease has a vast collection of surface antigen genes classified into families. These families are *mucins*, t*ransialidases* and *MASPS* which constitute half of the parasites genome. The members of these families are characterized by conserved sequence regions at their 5' and 3' ends that do encode a signal peptide and GPI anchor sequence. Their central regions constitute hypervariable regions some of which are marked with repeat sequences. A simple PCR technique that would form the basis to a pilot study for capturing the intragenomic diversity of the *T.cruzi* surface antigen genes is needed as a step towards more advanced techniques.Degenerate primers for targeting mucins, trans-sialidase and MASP genes were designed from conserved nucleotide sequences at the 5' and 3' regions of the genes and used in highly stringent PCR reactions to amplify a whole library of surface antigen genes. Purified PCR products were cloned into pGEM-T Easy vectors sequenced by Sanger methodology. Generated sequences were aligned against surface antigen gene sequences from the CL-Brener genome accessed through tritrypDB database. Sequence analysis of the nucleotide sequence from PCR products were phylogenetically analysed to determine their diversity. Degenerate mucin primers were able to produce PCR products from genomic DNA of *T. cruzi* from all DTUs some of which were confirmed to mucins. The generated sequences were diverse from each other with half of the sequences showing similarities to a cosmid C71. A non-surface antigen gene, histone deacetylase, was also discovered and found to share similarity with 7 of the generated sequences. Exploration of traditional southern, western and northern blots to microarray and next generation sequencing techniques for study of parasite surface antigen are recommended.

1.0   Introduction
1.1   The History and Epidemiology of Chagas Disease

Chagas Disease, also referred to as American trypanosomiasis, was identified to be a result of infection by *Trypanosoma cruzi* in 1909 by Carlos Chagas. This anthropozoonotic disease is a threat to more than 25 million people in Latin America (WHO, 2010). *T. cruzi DNA* and organ enlargement identified in human mummies from the Atacama desert in Chile, provides strong evidence that Chagas disease has existed for more than 9000 years (Aufderheide *et al*., 2004). *T. cruzi*, infects 7.7 million people of which 12,500 die annually (Rassi Jr *et al.,* 2010). While it is the most important parasitic infection in Latin America, migration has brought about the emergence of this infection in non-endemic countries such as those in Europe which makes Chagas disease even more important. For example, 29 Latin-Americans migrants working in London are now confirmed to have *T. cruzi* infection (Chiodini and Nolder, unpublished data). It is more likely that this statistic has increased with continued immigration. With millennia of Chagas disease affecting people and its discovery 102 years ago, researchers and clinicians still find it an exciting challenge to understand and wipe out this disease. Treatment of Chagas disease is at present being managed by two drugs, nifurtimox (nitrofuran derivative) and benzidazole (which is a nitroimidazole). Although these two drugs have 100% efficacy at the acute phase of the disease, the drug efficacy drops sharply as the disease progresses (Urbina, 2001). In addition there is no vaccine available to prevent the spread or even eradicate Chagas disease.

1.2   *Trypanosoma cruzi* Biology and Lifecycle

The transmission of *T. cruzi* is propagated by a haematophagous triatomine insect vector when its parasite infested faeces make contact with open bruised skin. The vectors that implicated in transmission of the majority of infections are *Triatoma infestans* and *Rhodnius prolixus* (Miles *et al.*, 2003). To date, the reservoirs of *T .cruzi* identified are sylvatic mammals such as marsupial opossums and armadillos and domestic animals such as dogs cats and guinea pigs (Yeo *et al.,* 2005). These reservoirs form a major part of the *T. cruzi* transmission. Transmission cycle can be loosely categorized as both sylvatic and domestic. Other routes of infection by *T .cruzi* are ingestion of food contaminated by the parasite, blood transfusion, organ transplantations and transplacental infections (WHO, 2010).

Infective metacyclic trypomastigotes within the faeces and urine of an infected triatomine enter through the skin and into the body of a host such as a human through a lesion often caused by a bugs bite. Within the lesion, the parasite multiplies by binary fission intracellularly in the macrophages surrounding the lesion producing an inflammation known as Chagoma. Newly formed trypomastigotes parasites now transformed from amastigotes then go on invading other cells in various tissues such as muscles while other escape into the blood stream as non-replicative trypomastigotes (Vickerman, 1985). These trypomastigotes can then be taken up by triatomine bugs when they have a blood meal. While in the mid-gut of the bug, the trypomastigotes transform to epimastigotes which then travel to the hind-gut and transform to metacyclic trypomastigotes which are primed and ready to invade into a host when deposited into a skin lesion of a host (Vickerman, 1985).

1.3 *Trypanosoma cruzi* Genetic Molecular Markers

In an effort to characterize and distinguish different strains and isolates of *T. cruzi* and correlate the differences clinically and epidemiologically, genes or the genome of the parasite have been explored using various molecular biology techniques. First, *T. cruzi* was divided into three main strain categories of zymodemes namely; Z1, Z2 and Z3 (Miles *et al*., 1977). Z3 was further subdivided in Z3-ASAT (aspartate-aminotransferase) as a result of the discovery that a proportion of Z3 had a Z1 type ASAT isoenzyme (Miles *et al.,* 1981). Progressing further from isoenzyme analyses, other genetic markers have been analyzed and have showed complex genetic profiles of *T. cruzi*. These markers include; kinetoplast DNA restriction fragment length polymorphism (RFLP) (Bogliolo *et al*., 1996), miniexon and ribosomal RNA (Fernandes et al., 1998), karyotype variation, random amplified polymorphic DNA (RAPD) and multilocus enzyme electrophoresis (MLEE) (Tibayrenc *et al*., 1993) and multilocus sequence typing (MLST) (Yeo *et al.,* 2011). *T. cruzi* has been classified into 6 discrete typing units (DTUs), TcI, TcII, TcIII, TcIV, TcV and TcVI (Zingales et al., 2009), formerly TcI, TcIIb, TcIIc, TcIIa, TcIId and TcIIe respectively (Brisse et al., 2000). Z1 and Z2 later corresponded to the DTUs; TcI and TcIIb respectively. TcI, TcIII and TcIV are predominating within wild sylvatic mammals (Miles *et al*., 2009). TcI, associated with less severe or asymptomatic infections, dominates in northern South America while TcII, TcV and TcVI are more common in the southern South America and are linked with Chagas megasyndromes (Miles *et al.,* 2009, Ramírez *et al*., 2010). The genome sequencing project of *T. cruzi* strain, CL Brener, discovered 19607 genes (diploid) of which at least 50% comprises surface antigen genes (El-Sayed *et al.,* 2005). When compared with *Leishmania* and theAfrican trypanosome, the *T. cruzi* genome has an extensive number of surface antigen genes which needs much studying. The ability of *T. cruzi* to invade host's cells and survive the host's immune response has been hypothetically linked with its surface antigen genes existing as multi-copy gene families in which certain protein translations of specific members of a given gene family escape clearing by the host immune system. (Franzén *et al*., 2011). This hypothesis is not surprising considering it has been proven in studies investigating the multi-copy surface antigen genes of *Plasmodium falciparum (var)* and *Trypanosoma brucei* (*VSG*) (Jiang et al., 2010, Dzikowski and Deitsch, 2006, Stockdale *et al*., 2008). The VSG gene family of *T. brucei* is diverse as a result of recombination and reassortment of VSG-like pseudogenes to make up functional *VSG* genes hence the ability of the parasite to cause a chronic infection in its hosts (Thon *et al*., 1989).

Trypanosoma cruzi: Surface Antigens

The surface antigen families in *T. cruzi* share common motifs in the N-terminal, central, and C-terminal regions despite their hypervariability within their families (Buscaglia *et al.,* 2006, Frasch, 2000). The common motifs within the surface antigen families that have been most studied are those found in trans-sialidase and mucins. The *T. cruzi* trans-sialidases have conserved motifs that are also shared with bacterial sialidases. These motifs are FRIP (Phe-Arg-Ile-Pro) and a maximum of 4 copies of the Asp (Ser-X-Asp-X-Gly-X-Thr-Trp) box. Other conserved motifs present in the trans-sialidase protein family are; Val-Thr-Val-X-Asn-Val-X-Leu-Tyr-Asn-Arg which is trypanosomal specific and a GPI anchor motif. The FRIP and Asp boxes together with Val-Thr-Val-X-Asn-Val-X-Leu-Tyr-Asn-Arg motif make up the sialidase domain. The mucin protein family also has conserved motifs which are: N-terminal signal peptide and a C-terminus GPI anchor sequence. If exploited, these motifs could provide the possibility of amplifying over 50% of the *T. cruzi* genome using < 10 PCR assays. The surface antigen genes of *T. cruzi* comprises; trans-sialidase *(TS85)*, mucin (*TcMUC* and *TcSMUG*), mucin associated surface protein (*MASP*) and GP63 (*Tcgp63*) that have been implied to be located in the parasite's membrane (Buscaglia *et al*., 2006, Cuevas *et al*., 2003, Frasch, 2000). Trans-sialidase protein transfers sialic acid from the host cells' glycoconjugates to parasite mucins (Buscaglia *et al*., 2006). Mucins and GP63 are glycophosphatidylinositol (GPI) anchored glycoproteins in the parasites cell surface. Mucins are important for

protection and cell adhesion (Buscaglia *et al.,* 2006). GP63 contain metalloprotease activity which is essential for tissue and cell invasion (Cuevas *et al*., 2003).

One of the most advanced and accurate methods of identifying strains of *T. cruzi* have been developed using genome-wide microsatellite markers (Llewellyn *et al*., 2009b). With this technique, it was discovered that the genetic diversity of *T. cruzi* in individual opossums is one log higher than previously assumed (Llewellyn *et al.,* 2011). No genetic diversity studies *T. cruzi* (Guhl and Ramírez, 2011, Llewellyn *et al.,* 2009a, Llewellyn *et al*., 2011), have investigated the intergenomic diversity of surface antigen genes between 2 or more genomes let alone in a population of *T. cruzi* species. In *Plasmodium falciparum*, the study of *var* gene intraspecific diversity in Papua New Guinea revealed 185 of 895 known distinct DBLα types by sequencing the *var* genes amplified using degenerate primers (Barry et al., 2007, Bull *et al*., 2008). The genetic diversity in *var*, coupled with recombination and Darwinian selection in simple models, could explain the ability of *P. falciparum* to re-infect a malaria patient severally during their life (Barry *et al*., 2007, Bull *et al*., 2008). The PCR approach used in the amplification of *var* genes by the study (Barry *et al.,* 2007) coupled with DNA-DNA microarray and size length polymorphism indel assay will enable the determination of *T. cruzi* surface antigen genetic diversity on a population level. This enables the absence of alleles or genes between strains to be detected.

The annotated genome of TcI (Sylvio X10/1) has recently been published (Franzén *et al.,* 2011). This genome in conjunction with a well annotated TcVI genome (CL Brener) provides the necessary bioinformatic material to study the intrahost and intraspecific diversity of *T. cruzi* surface antigens. Four housekeeping genes that provide inter-DTU distinction of the *T. cruzi* lineages; metacyclin-III (Met-III), RNA-binding protein-19 (*RB19*), glutathione-dependent peroxidise II (TcGPXII) and dyhydrofolate reductase-thymidylate synthase (DHFR-TS) are single copy genes (Yeo *et al*., 2011). That will provide controls during PCR amplification of the multicopy surface antigen genes.

Aim

The aim is to amplify the diverse surface antigen genes by PCR and compare them within and between TcI DTUs in chagasic mammalian hosts, sylvatic hosts and triatomines.

2.0   Methods and Materials

2.1   Bioinformatics: Primer Design

Amino acid sequences of *T. cruzi* cell surface antigen proteins, Trans-sialidase, MASP and mucin were downloaded from Genbank at the National Center for Biotechnology Information (NCBI). Trans-sialidase, MASP and Mucin amino acid sequences were individually aligned using ClustalX2.1 applying default parameters. The accession numbers associated with the cDNA (DNA coding) sequences of the surface antigen proteins were extracted from Genbank files of the downloaded amino-acid sequences using a find and print script written in *perl*. The printed accession numbers were used as a query to search and download the cDNA sequence of the surface antigens from Genbank. Multiple alignments of the downloaded cDNA sequences for each surface antigen protein family were made using transAlign. All the alignments were visualised and manually edited using Se-Al and a graphical print out of the alignment was done using CodonCodeAligner software. Conserved motifs discovered from the cDNA sequence alignment were used to design overlapping forward and reverse degenerate primers aimed at amplifying the hypervariable regions of the surface antigen genes. Primer3 was used for designing the primers and Codon code aligner was used to determine the degeneracy of the primers.

Table 2. Details of degenerate primers designed to target surface antigen genes.

Surface Antigen gene family	Forward Primers	Primer Sequence (5'->3')	Tm [⁰C]	Reverse Primers	Primer Sequence (5'->3')	Tm [⁰C]
Trans-sialidases	Trans-4.for	ATGBMYYCMMGYGYTGYYG (19)	59.2	Trans-11.rev	DRBHRVVRDDVCVCAVABBCCMA (23)	63.5
	Trans-27.for	TVRDGRYRHYCHRYRHAVRHRYYDT (26)	62.9	Trans-15.rev	DRBHRNVRDDVCVGAVCBBCC (21)	63.6
	Trans-45.for	MWRYYRTYRCKGYRGRKYGKMT (23)	61.5	Trans-19.rev	HMVVNDRBHRVVRDDVGVCA (20)	58.9
	Trans-50.for	RYVVCVRRKRYHYMRMATKTYNHG (24	62.1			
	Trans-78.for	RRRRRRGRRRSRRRRRRRRBGYG (23)	65.4			
	Trans-single.for	ACACACAACCGTTGCCTTAACA (22)	58.4			
Mucins	TcMUC1_1.for	YTKHWSWRCDYYYTKYTKNYKMT (23)	57.4	TcMUC1_1.rev	YGRGGAKYCRRCRAGCRSMAG (21)	64.7
	13TcMUC1_14.for	TKYKKRYKYTYYYYYTMRKYT (21)	54.9	TcMUC1_10.rev	TGYRMRMCAGYGYGGAKRCDKC (22)	64.6
	15TcMUC1_14.for	TKYKKBYKATTGGTCTGTGCT(21)	57.2	TcMUC1_18.rev	AHMGYVRSTRYGGTGKRMRMCAGMRCG (28)	70.2
MASP	MASP_4.for	RYGRDTKRYCCYBHGCTDYYG (21)	63.0	MASP_4C.rev	YRYRRNCRYBRCHRYHRHVGC (21)	62.4
	MASP_46T.for	VYDTDYKBYYTYHKYRYSYYKYT (23)	59.4	MASP_4T.rev	YRYRRNCRYBRCHRYHRHVGT (21)	60.4
	MASP_46G.for	VYDTDYKBYYTYHKYRYSYYKYG (23)	61.2	MASP_8G.rev	YYRYYYRRNCRYBRCHRYHG (20)	60.0
				MASP_8A.rev	YYRYYYRRNCRYBRCHRYHA (20)	58.8

2.2 PCR Amplification of T. Cruzi Single Copy Genes and Surface Antigen Gene Families

Single copy genes, used as controls, and members of the surface antigen gene families were amplified by PCR using the following reaction cycle: a 3 minute denaturation step at 94⁰C, then 30 cycles of amplification (30 seconds at 94⁰C, 1 minute at; 60⁰C (mucin and TcGPXII), 58⁰C (MASP, transialidases and TR), 4 minutes at 72⁰C) and finally 72⁰C for 10 minutes. The PCR reaction was set up with a final volume of 50 μl comprising 5 μl (10X) NH₄ buffer (Bioline, UK), 1.5 μl (50mM) MgCl₂, 1 μl (2mM) dNTPs; 2.5 μl 10pmols each of forward and reverse primers, 0.5 μl (5U) *Taq* polymerase (Bioline, UK) and 1 μl genomic DNA (genomic DNA concentrations, see Table 1). The sizes of the PCR amplicons were separated by electrophoresis in a 0.05% EtBr stained 1% agarose gel submerged in (1X) TAE buffer and the gel observed under ultraviolet (UV) light.

2.3 Cloning of Specific Amplicons Produced using Surface Antigen Gene Degenerate Primers

Gel pieces containing discrete DNA bands were carefully cut out and the DNA within them extracted, purified and eluted in a 30 μl final volume following the Qiagen Gel extraction protocol (October, 2010) to give a high concentration of DNA. The pure DNA from the extraction was inserted into pGEM-T Easy cloning plasmid vector and the resultant recombinant vector transformed into JM109 High Efficiency Competent Cells following the Promega pGEM-T Easy Vector System protocol. The JM109 cells were pelleted and resuspended in 350 μl SOC medium of which 175 μl were plated on each of two duplicate LB agar plates containing; 0.5mM IPTG, 100 μg/ml ampicillin and 80 μg/ml X-gal to give a high number of colonies.

2.4 PCR and Restriction Digest Screening and Sequencing the Pgem-T Easy Recombinant Plasmid

100 Transformed colonies of JM109 recombinant cells per plate were individually picked and transferred to 5ml LB broth media containing 100 μg/ml ampicillin to grow overnight at 37⁰C. The recombinant plasmid vectors produced in the cloning process were extracted from the JM109 cells grown on the LB broth cultures following the Qiagen miniprep protocol. The plasmid minipreps were screened by PCR using the following reaction cycle; a 3 minute denaturation step at 94⁰C, then 30 cycles of amplification (30 seconds at 94⁰C, 1 minute at; 50⁰C, 4 minutes at 72⁰C) and finally 72⁰C for 10 minutes. The PCR reaction was set up with a final volume of 10 μl comprising 1 μl (10X) KCl buffer (New England Biolabs, UK), 1 μl (2mM) dNTPs; 0.4 μl 10pmols each of T7 and SP6 primers, 0.25 μl (5U) *Taq* polymerase (Bioline, UK) and 1 μl plasmid miniprep DNA. The sizes of the PCR amplicons were separated by electrophoresis in a 0.05% EtBr stained agarose gel submerged in (1X) TAE buffer and the gel observed under ultraviolet (UV) light. An *Eco*RI restriction digest of the plasmid minipreps was done in 10 μl reaction volumes containing; 2 μl 10X restriction enzyme buffer H, 1 μl acetylated BSA (10ug/μl), 1ul of plasmid and 0.25 μl *Eco*RI (10u/μl). The products generated from the PCR and *Eco*R1 restriction digest were analysed on a 0.05% EtBr stained agarose gel submerged in (1X) TAE buffer and the gel observed under ultraviolet (UV) light. Plasmid minipreps positive for PCR product inserts were bi-directionally sequenced (5' and 3') using T7 and SP6 primers in order generate the sequence of the PCR product inserts within the recombinant plasmid. Sequencing was done in an ABI 3730 capillary DNA sequencing machine

according to Big Dye terminator Cycle Sequencing protocol V3.1 (Applied Biosystems). Sequences generated that showed high quality base reading on both directions were carried forward for further analyses.

2.5 Sequence Analysis

The generated sequences of the recombinant plasmid inserts were assembled into contigs, the sequence assemblies edited using Codon Code Aligner software. The sequences were then searched using Basic Local Alignment Search Tool (BLAST) against DNA or protein databases for similarities at the NCBI. The sequence match with the highest E-value was recorded. Combined with DNA sequences of the genes TcMUCI, TcMUCII downloaded from Trypanosome Genome Database (www.tritrypDB.org), the sequences were aligned using ClustalX2. Tree produced from the alignment using neighbor joining were also produced and viewed using Treeview.

3.0 Results

3.1 Highly Diverse Surface Antigen Genes with Conserved 5' and 3' Ends Enable Degenerate Primer Design

The surface antigen gene families, mucins, MASPS and trans-sialidases have gene members that contain very diverse DNA sequences. In order to amplify such highly diverse genes with very minimal PCR reactions per gene family, the sequences of these gene families were aligned using ClustalX2.1 and alignment viewed using Se-Al. Alignments from the trans-sialidases showed the highest degree of sequence conservation while the MASPs had the least degree of conservation. Although not the most conserved, the mucin gene sequences had the most conserved 5' and 3' ends. From the conserved 5' and 3' ends of the surface antigen genes, degenerate primers were able to be designed (Table 1).

3.2 Mucins amplified using Degenerate Primers

Discrete banding patterns between 2000 bp and 800 bp; there provided the bands of the amplicons produced using the primer pair, 13TcMUC1_14.for- TcMUC1_18.rev, had low intensity. The most commonfeature seen in the agarose gel analysis was the banding pattern of the 900bp and 800bp bands amplified from the *T.cruzi* genomic DNA of isolates, M6421 cl6 (Tc III) and Tula cl2 (Tc VI) using primers, 13TcMUC1_14.for and TcMUC1_10.rev (arrow pointed, Figure 2) . These DNA banding profile on the agarose gel analysis was similar between the two isolates both of which also originate from *Homo sapiens* (Table 1). The 900bp DNA band had almost twice the intensity of the 800bp DNA band. PCR amplicons produced using the primer pairs, 15TcMUC1_14.for- TcMUC1_18.rev and 15TcMUC1_14.for- TcMUC1_10.rev produced two bands (just approximately > or < 600bp) from each of the 11 *T .cruzi* genomic DNA isolates covering all six DTUs (data not shown). Two more bands, one from M6421 cl6 (Tc III) and the other from Tula cl2 (Tc VI), both sized at 900 bp were also realised from PCR product amplified with the primer pair, 15TcMUC1_14.for- TcMUC1_10.rev. To verify that we were targeting a multicopy gene family, in this case *mucins*, a single copy gene, *TcGPXII* was used as a positive control. *TcGPXII* primers were used alongside *mucin* primers to amplify *TcGPXII* from 12 *T.cruzi* genomic DNA isolates. A single DNA band of approximately 500 bp in size was produced from all the isolates.
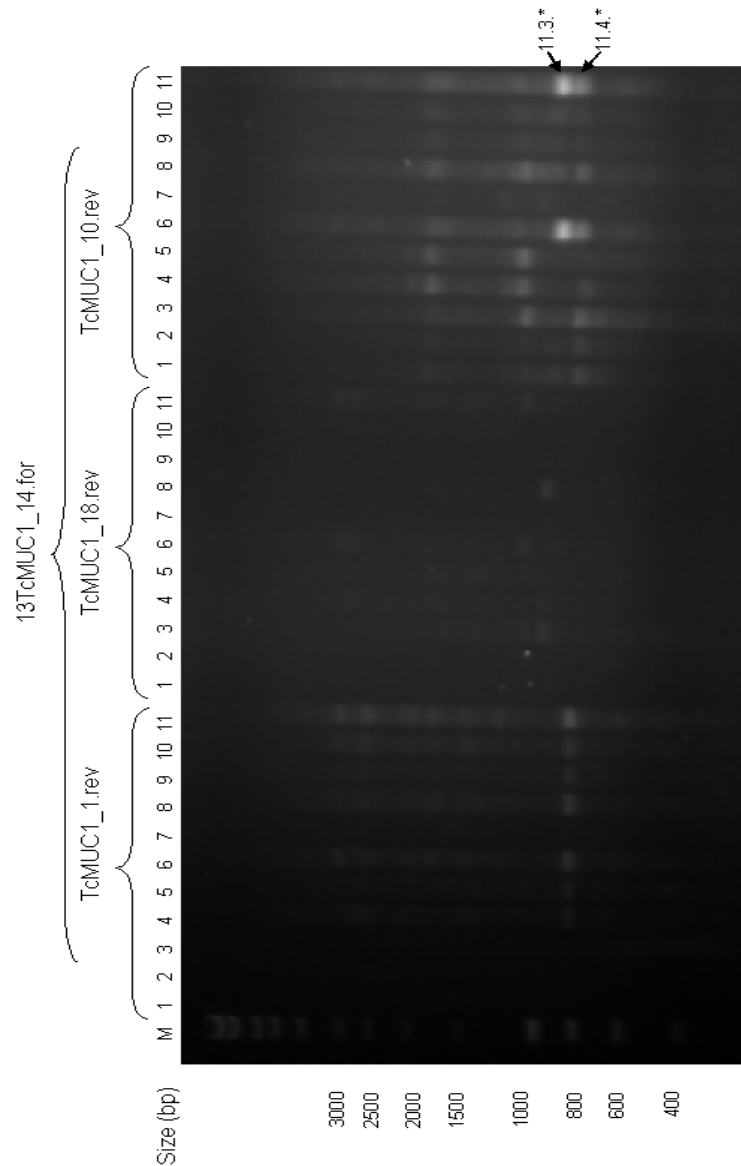
Figure 2: Agarose gel analysis of PCR products amplified with mucin primers and *TcGPXII* gene primers (positive controls) from *T. cruzi* genomic DNA. (A) Forward primer, 13TcMUC1_14.for paired with mucin reverse primers; TcMUC1_1.rev, TcMUC1_18.rev and TcMUC1_10.rev. (B) Forward primer, 15TcMUC1_14.for paired with mucin reverse primer, TcMUC1_18.rev. (C) Forward primer, 15TcMUC1_14.for, paired with mucin reverse primer, TcMUC1_10.rev. (D) *TcGPXII* forward and reverse primers. The *T. cruzi* isolates in the analysis are; (1) Sylvio X10/1, (2) JR cl4, (3) 92090802P cl1, (4) Esm cl3, (5) IVV cl4, (6) M6421 cl6, (7) Can III cl1 (8) 92-80 cl2, (9)PAH 179 cl5, (10) CL Brener, (11) Tula cl2, (12) P251 cl7 (Table 1). Arrows points at bands that were cloned and sequenced and are assigned 11.3.* and 11.4.* (sequence clone names) M: DNA Marker HyperLadder I (Bioline)

3.3 DNA Smears Produced with Trans-Sialidase and MASP Degenerate Primers

Only one transialidase degenerate primer pair, Trans-78.for- Trans-15.rev (data not shown), and four MASP degenerate primer pairs, MASP_4.for paired with MASP_4C.rev, MASP_4T.rev, MASP_8G.rev and MASP_8A.rev were able to produce amplicons. The amplicons resulted in smears on the agarose gel, a feature shared across all 12 *T. cruzi* genomic DNA isolates and not discrete bands as those shown when using *mucin* primers, described above. These results were realised after several rounds of PCR with increasing stringent conditions to a final annealing temperature of 58°C. The positive control to this set of PCR reactions was another PCR reaction using primers for a single copy gene, *TR,* against the same array of 12 *T. cruzi* genomic DNA isolates. The PCR product amplified using *TR* primers produced a band sized at approximately 1.3Kb across all the isolates.

3.4 Mucin and Trans-Sialidase Degenerate Primer Cocktails Do Not Produce PCR Products

To determine whether using a cocktail of either mucin or trans-sialidase primers in PCR reaction would generate products, a cocktail of all mucin primers and all trans-sialidase primers, separately were made. Each primer cocktail was put against all 12 *T. cruzi* genomic DNA isolates, individually, to amplify *mucin* and *trans-sialidase* gene fragments. The PCR reaction products were analysed by agarose gel electrophoresis which showed no amplified DNA product were produced in the reactions (data not shown). However with TcGPXII primers used in accompanying PCR reactions as positive control, a single band of 500 bp in size was produced across all the 12 *T. cruzi* genomic DNA isolates (data not shown).

Table3: BLAST search results showing highly similar sequences to *mucin* PCR products.

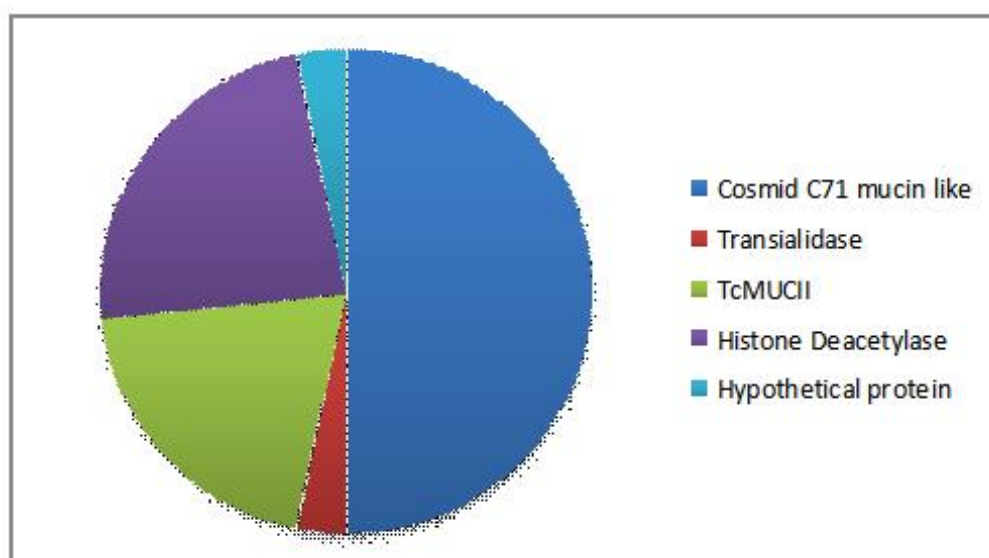| Tula Amplicon clones	Similar sequences from BLAST			
	Accession Number	% coverage	% identity	Gene Name
11.3.4	XM814886.1	85	99	*Histone Deacetylase*
11.3.05.1	AY298908.1	93	97	*Cosmid C71 mucin-like*
11.3.1	XM813917.1	78	98	*TcMUCII*
11.3.10k	AY298908.1	86	96	*Cosmid C71 mucin-like*
11.3.11	AY298908.1	86	97	*Cosmid C71 mucin-like*
11.3.12.6	XM814886.1	85	99	*Histone Deacetylase*
11.3.14	AY298908.1	87	96	*Cosmid C71 mucin-like*
11.3.15	XM816810.1	72	98	*TcMUCII*
11.3.2	AY298908.1	86	96	*Cosmid C71 mucin-like*
11.3.2.1	XM814886.1	86	99	*Histone Deacetylase*
11.3.3j	XM814886.1	86	99	*Histone Deacetylase*
11.3.4	AY298908.1	88	96	*Cosmid C71 mucin-like*
11.3.5	XM806515.1	70	92	*TcMUCII*
11.3.8	XM806515.1	75	92	*TcMUCII*
11.3.9f	XM815657.1	42	93	*Hypothetical protein*
11.4.11	AY298908.1	88	96	*Cosmid C71 mucin-like*
11.4.12	XM803891.1	70	88	*TcMUCII*
11.4.12.13.16	AY298908.1	87	97	*Cosmid C71 mucin-like*
11.4.13	AY298908.1	87	95	*Cosmid C71 mucin-like*
11.4.15	XM814886.1	86	99	*Histone Deacetylase*
11.4.16	XM798918.1	42	98	*Trans-sialidase*
11.4.3	AY298908.1	89	97	*Cosmid C71 mucin-like*
11.4.4	AY298908.1	87	96	*Cosmid C71 mucin-like*
11.4.7	AY298908.1	86	96	*Cosmid C71 mucin-like*
11.3.3	AY298908.1	86	97	*Cosmid C71 mucin-like*
11.4.8	XM813917.1	78	98	*TcMUCII*
11.1.05.7c	AY298908.1	86	97	*Cosmid C71 mucin-like*
11.3.16	AY298908.1	88	96	*Cosmid C71 mucin-like*
11.3.1a	XM814886.1	87	99	*Histone Deacetylase*
11.4.2b	AY298908.1	87	97	*Cosmid C71 mucin-like*
11.4.9	XM814886.1	86	99	*Histone Deacetylase*

Figure 4: **Results of sequence similarity search.** A proportionate representation of BLAST results using sequences of PCR amplicons (11.3.* and 11.4.*).

3.5 Positive Identification of Mucins from Sequence Analyses

BLAST search results showed that half of the sequences searched were highly identical to *cosmid C71 mucin like* gene, about a quarter were highly identical to histone deacetylaces and a fifth were highly identical to *TcMUCII* (Table 3 and Figure 4). Trans-sialidase and a hypothetical protein gene were also realized from the BLAST search results. The sequences highly similar to cosmid C71 aligned with *TcMUCI* protein sequences downloaded from the tritrypDB database and also showed variable number of amino acid repeats with a consensus sequence of T_8KP_2 in addition to varying sequences (Figure 5). The variation in the sequences is also evident in the unrooted neighbor-joining tree of the DNA sequences from mucin PCR products (11.3.* and 11.4.*) with CDS sequenced from the cosmid C71. The tree shows the Mucin product being split into two groups.
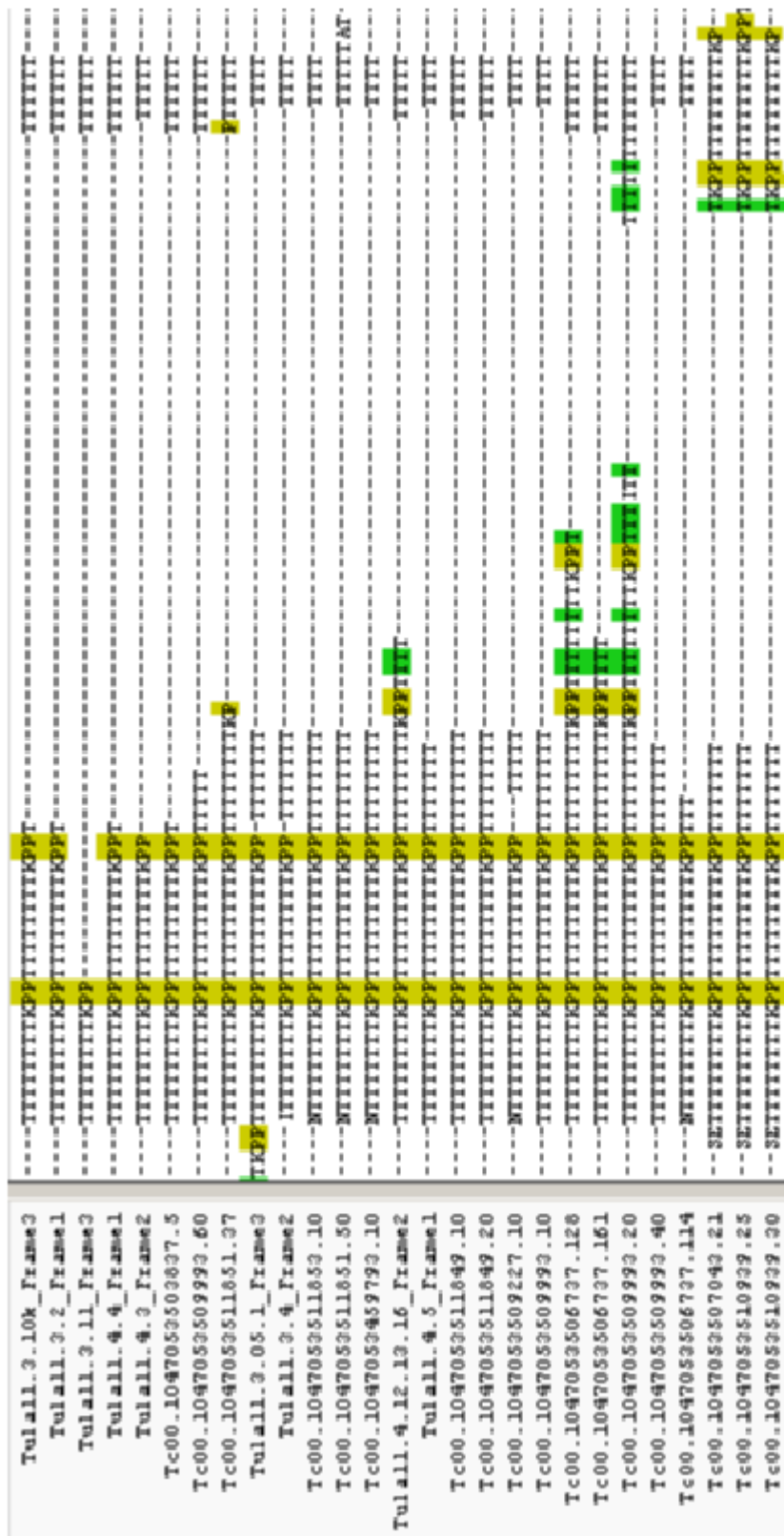
Figure 5: **TcMUCI protein alignment**. Section of sequences generated from cloned PCR products (Tula11.3.* and Tula11.4.*) and aligned with *TcMUCI* genes downloaded from tritrypDB using ClustalX. Alignment shows T_8KP_2 repeats.
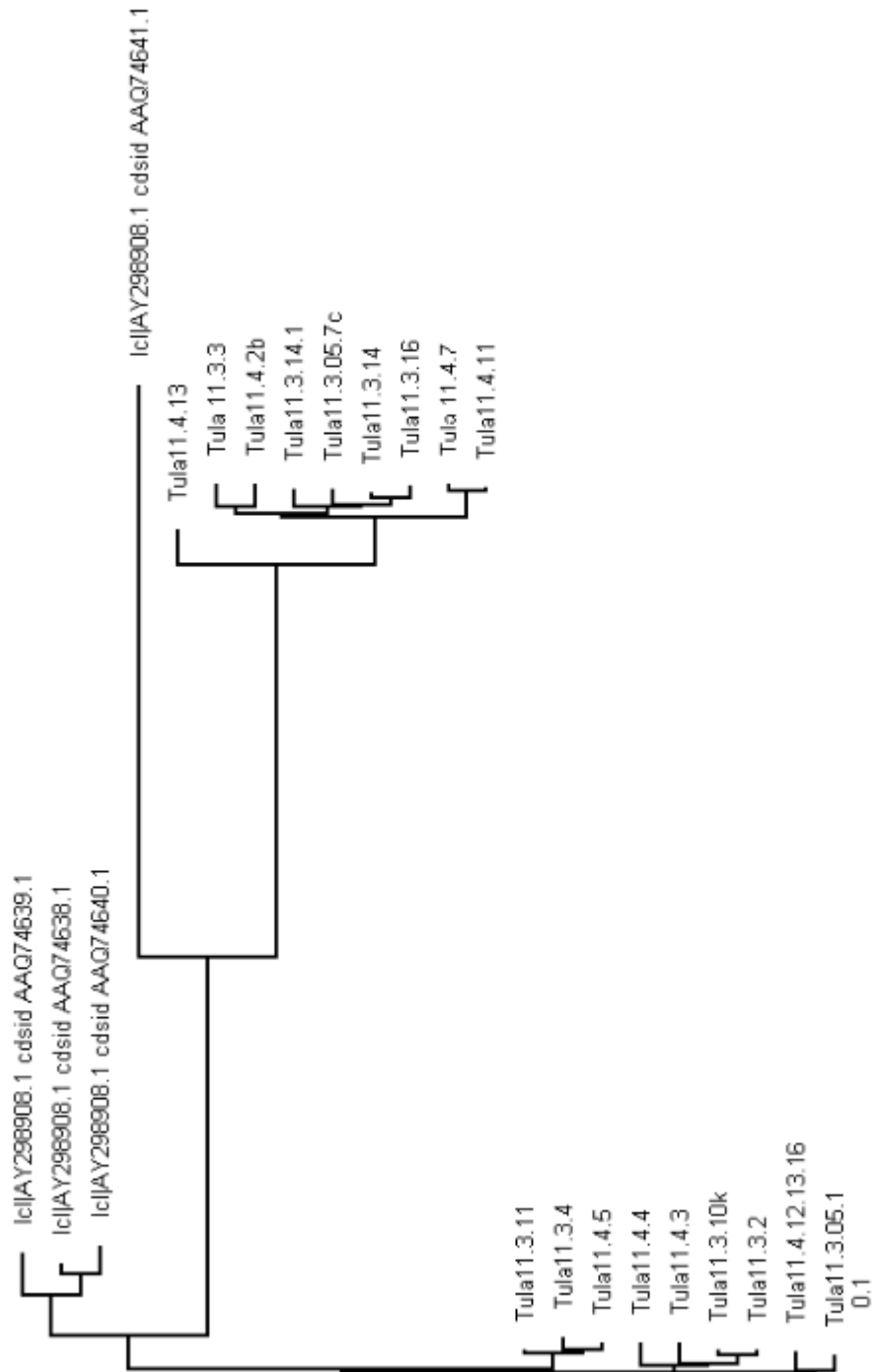
Figure 6: Unrooted Neighbour-joining tree of aligned CDS sequences from Cosmid C71 with *mucin* PCR products that showed similarity to the cosmid.

4.0   Discussion

This study has revealed genetic profiles of *trans-sialidases* and *mucins* that agree with reports by other authors. These revelations are that both *T. cruzi* surface antigen gene families, *mucins* and *trans-sialidases*, do contain members which share highly similar DNA sequences in both their 5' and 3' ends arising from observations of multiple DNA sequence alignments. In addition, genes within *trans-sialidase* family especially showed a high level of sequence homology in parts of its central regions. This observation provides further evidence that these conserved sequence regions are necessary for the effective functioning of the surface antigen proteins. The conserved 5' regions code for a signal peptide while the conserved 3' code for a GPI-anchor which are both essential for their location on the parasite's cell membrane. The conserved central regions of *trans-sialidase* codes for the sialidase domain located in the N-terminal half of the protein which transfers sialic groups from host glycoproteins to parasite mucins. A hyper variable 3' central region also identified in the multiple *trans-sialidase* sequence alignment could be thought to code for amino acid repeats

that are of variable size or sequence. The presence of highly conserved sequences at both ends of these surface antigen genes enabled degenerate primers to be designed and used to amplify whole gene fragments that make up one gene family (see Table).

The gel electrophoretic profile produced by PCR product obtained using mucin primers appeared to show the expected variable size fragments that would contribute to differently sized proteins described in previous studies (Buscaglia *et al*., 2006, Frasch, 2000). The high intense DNA bands sized at approximately 900bpand 800bp produced by amplification using the primers 13TcMUC1_14.for and TcMUC1_10.rev, against genomic DNA isolated from Tula cl2 (Tc VI) was cloned and sequenced (Figure 2). The difference in band intensity between the two bands was a very interesting finding. It may be deduced that a copy number variation between the DNA sequences in the two bands is present in the genome which could be confirmed by southern analysis in future experiments. The sequences resulting from DNA clones revealed various *mucins* (TcMUCI and TcMUCII), a transialidase and histone deacetylase. A cosmid C71 mucin-like gene was discovered and which seemed to have been missed when downloading mucins from Genbank for primer design. Further study of this cosmid C71 mucin-like sequence revealed that it was one of many cosmid libraries that were generated from inserting DNA fragments produced by restriction digest and DNA shearing into a cosmid vector (Hanke *et al.,* 1996). The cosmid C71 has been shown by hybridization analysis that they contain *TcMUCI* genes (Campo *et al.,* 2004). It could be that the sequences that showed high sequence similarity to the cosmid C71 belonged to the mucin subfamily, *TcMUCI*. *TcMUCI* genes encode a variable number of amino acid repeats with a consensus sequence of T_8KP_2 (Campo *et al.,* 2004). These repeats were evident in the protein translations of the Sanger generated sequences aligned with *TcMUCI* proteins and it provides the confirmation that the sequences highly homologous to the cosmid C71 indeed belong to *TcMUCImucin* subfamily. Another feature showed by these sequences was a split into two groups suggesting that the *TcMUC1* group has a potential to be split further concise groups. The capture of significant *TcMUCI* genes associated with cosmid C71 by PCR from genomic DNA provides an interesting indication that these genes dominate the *T.cruzi* genome more than other mucin genes. Therefore these cosmid mucin- like sequences would form the basis of nucleotide probes that could be hybridized against the T. cruzi genomes from all 6 DTU's to get a snap of their copy number variations. The discovery of PCR amplicons containing sequences similar to *histone deacetylase* was also an interesting observation. It could be explained that the histone deacetylase realized here could be a pseudogene that may have resulted from possible DNA mutations that resulted in its sequence similar to particular members of the *mucin* family. In addition, this similarity in DNA sequence to mucins may have made it possible for the degenerate primer to anneal against it.

During the course of this study, a number of limitations were experienced along the way meant that only a few sequences were generated from one DNA band out of the several bands produced during PCR. One major limitation was the inability to achieve high transformation efficiency during cloning. The low transformation efficiency during the cloning process may have been due to a low DNA insert to vector ratio in the ligation reactions that may have resulted to a high proportion of self-ligated vectors being transformed into the JM109 cells. Time was also a limiting factor that prevented the correction of technical errors and further cloning and sequencing of PCR products from other DNA bands. Another limitation in this study was that distinct DNA band sizes could not be produced using *MASP* of *trans-sialidase* primers. An explanation to this observation could be that the designed primers were able to anneal to each other or non-specifically anneal to random loci within the *T. cruzi* genome. In order to avoid these limitations in the future, the concentration of pure DNA extracted from agarose gel would be measured so as to accurately get the optimal insert to vector ratio in the ligation reactions. *Trans-sialidase*s are about 3kb in size and therefore exploiting internal conserved motifs such as those that encode the Asp box and Val-Thr-Val-X-Asn-Val-X-Leu-Tyr-Asn-Arg motif and making primers against them would be a significant step forward. *MASP* primers would also be designed after careful realigning of the sequences of genes within the family.

The results generated from this study provide areas of further research in *T.cruzi* surface antigens. This study employed a simple PCR technique using degenerated primers which has attempted to capture the diversity of surface antigen genes especially the *mucins*. This approach with substantial improvements can be carried and applied for other surface antigen gene families such as *gp63*. More approaches are still available to be exploited to capture the diversity of surface antigen families both at the DNA, mRNA and protein level. At the DNA level, advances in sequencing techniques such as 454 or pyrosequencing, SOLEXA (Illumina), ABI Solid and Helicos coupled with a steady drop in sequencing costs means that amplification and sequencing of surface antigen genes and their transcripts to capture their intragenomic and intergenomic diversity is easily feasible

(Mardis, 2008). An example of a study where pyrosequencing has been used to study the diversity of surface antigen genes is single strand sequencing of merozite surface proteins 1 and 2 (MSP1 and MSP2) in *P. falciparum* (Juliano et al., 2010). In this study, a six-fold more variants per infection than that described using recommended genotyping methods was achieved. RNA Seq (Wang et al., 2009), is a technique of sequencing mRNA transcripts that could potentially give surface antigen transcriptomic profiles of *T. cruzi* in different hosts and through the course of a chronic infection is an exciting approach to understand the mechanisms *T. cruzi* employs to evade the host immune system. One striking feature in the T. *cruzi* genome is at least one in every three functional surface antigen gene is a pseudogene. A method is therefore needed to distinguish these pseudogenes from functional genes in order to accurately determine the diversity of functional surface antigen genes. For this distinction to be achieved protein expression profiles of *T. cruzi* surface proteins and sequences of these proteins would need to be produced and referenced back against the genome. Generation of high affinity monoclonal antibodies raised against the surface antigen protein would be a great advancement in deducing the molecular structures of these proteins.

5.0   Conclusion

The findings from these experiments does show that more still needs to be done in order to fully understand the genetic diversity and the genetic dynamism of these *T. cruzi* cell surface antigens. This understanding will propel the realization of potential vaccines against American trypanosomiasis and other Kinetoplastid parasitic infections.

References

Aufderheide, A. C., Salo, W., Madden, M., Streitz, J., Buikstra, J., Guhl, F., Arriaza, B., Renier, C., Wittmers, L. E. and Fornaciari, G. (2004). A 9,000-year record of Chagas' disease. *Proceedings of the National Academy of Sciences of the United States of America,* 101, pp 2034.

Barry, A. E., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, S. M., Mcvean, G. A. V. & Day, K. P. (2007). Population genomics of the immune evasion (var) genes of *Plasmodium falciparum*. *PLoS Pathogens,* **3,** e34.

Bogliolo, A. R., Lauria-Pires, L. & Gibson, W. C. (1996). Polymorphisms in Trypanosoma cruzi: evidence of genetic recombination. *Acta Tropica,* **61,** pp 31-40.

Brisse, S., Barnabé, C. and Tibayrenc, M. (2000). Identification of six *Trypanosoma cruz*i phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *International Journal for Parasitology,* **30,** pp 35-44.

Bull, P. C., Buckee, C. O., Kyes, S., Kortok, M. M., Thathy, V., Guyah, B., Stoute, J. A., Newbold, C. I. and Marsh, K. (2008). *Plasmodium falciparum* antigenic variation. Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Molecular microbiology,* **68,** p 1519-1534.

Buscaglia, C. A., Campo, V. A., Frasch, A. C. C. and Di Noia, J. M. (2006). *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology,* **4,** pp 229-236.

Campo, V., Di Noia, J. M., Buscaglia, C. A., Aguero, F., Sanchez, D. O. and Frasch, A. C. C. (2004). Differential accumulation of mutations localized in particular domains of the mucin genes expressed in the vertebrate host stage of *Trypanosoma cruzi*. *Molecular and biochemical parasitology,* **133**, pp 81-91.

Cuevas, I. C., Cazzulo, J. J. and Sanchez, D. O. (2003). Gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infection and immunity,* **71, pp** 5739.

Dzikowski, R. and Deitsch, K. (2006). Antigenic variation by protozoan parasites: insights from *Babesia bovis*. *Molecular microbiology,* **59, pp** 364-366.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L. and Blandin, G.(2005). The genome sequence of *Trypanosoma cruzi,* etiologic agent of Chagas disease. *Science,* **309, pp** 409.

Fernandes, O., Souto, R. P., Castro, J. A., Pereira, J. B., Fernandes, N. C., Junqueira, A., Naiff, R. D., Barrett, T. V., Degrave, W. and Zingales, B. (1998). Brazilian isolates of *Trypanosoma cruzi* from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. *The American journal of tropical medicine and hygiene,* **58, pp** 807.

Franzén, O., Ochaya, S., Sherwood, E., Lewis, M. D., Llewellyn, M. S., Miles, M. A., Andersson, B. & Bates, P. A. (2011). Shotgun Sequencing Analysis of Trypanosoma cruzi I Sylvio X10/1 and Comparison with T. cruzi VI CL Brener. *PLoS Neglected Tropical Diseases,* **5**.

Frasch, A. (2000). Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitology Today,* **16, pp** 282-286.

Guhl, F. and Ramírez, J. D. (2011). *Trypanosoma cruzi* I diversity: towards the need of genetic subdivision? *Acta Tropica*.
Hanke, J., Sanchez, D., Henriksson, J., Aslund, L., Pettersson, U., Frasch, A. and Hoheisel, J. (1996). Mapping the *Trypanosoma cruzi* genome: analyses of representative cosmid libraries. *Biotechniques,* **21, pp** 686-693.

Jiang, L., López-Barragán, M. J., Jiang, H., Mu, J., Gaur, D., Zhao, K., Felsenfeld, G. and Miller, L. H. (2010). Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proceedings of the National Academy of Sciences,* **107, pp** 2224.

Juliano, J. J., Porter, K., Mwapasa, V., Sem, R., Rogers, W. O., Ariey, F., Wongsrichanalai, C., Read, A. and Meshnick, S. R. (2010). Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proceedings of the National Academy of Sciences,* **107, pp** 20138.

Llewellyn, M. S., Lewis, M. D., Acosta, N., Yeo, M., Carrasco, H. J., Segovia, M., Vargas, J., Torrico, F., Miles, M. A. and Gaunt, M. W. (2009a). *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. *PLoS Neglected Tropical Diseases,* **3,** e510.

Llewellyn, M. S., Miles, M. A., Carrasco, H. J., Lewis, M. D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V. and Valente, S. A. (2009b). Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens,* **5,** e1000410.

Llewellyn, M. S., Rivett-Carnac, J. B., Fitzpatrick, S., Lewis, M. D., Yeo, M., Gaunt, M. W. and Miles, M. A. (2011). Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a mechanism of diversifying selection. *International Journal for Parasitology*.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *TRENDS in Genetics,* **24, pp** 133-141.

Miles, M., Llewellyn, M., Lewis, M., Yeo, M., Baleela, R., Fitzpatrick, S., Gaunt, M. and Mauricio, I. (2009). The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. *Parasitology,* **136, pp** 1509-1528.

Miles, M., Povoa, M., De Souza, A., Lainson, R., Shaw, J. and Ketteridge, D. (1981). Chagas's disease in the Amazon Basin: II. The distribution of Trypanosoma cruzi zymodemes 1 and 3 in Pará State, north Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **75, pp** 667-674.

Miles, M., Toye, P., Oswald, S. C. and Godfrey, D. (1977). The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **71**, pp 217-225.

Miles, M. A., Feliciangeli, M. D. andDe Arias, A. R. 2003. American trypanosomiasis (Chagas' disease) and the role of molecular epidemiology in guiding control strategies. *Bmj,* **326, pp** 1444.

Ramírez, J. D., Guhl, F., Rendón, L. M., Rosas, F., Marin-Neto, J. A., Morillo, C. A. and Da Costa Santiago, H. (2010). Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. *PLoS Negl Trop Dis,* **4,** e899.

Rassi Jr, A., Rassi, A. & Marin-Neto, J. A. 2010. Chagas Disease. *The Lancet,* 375**,** 1388-1402.

Stockdale, C., Swiderski, M. R., Barry, J. D. & Mcculloch, R. (2008). Antigenic variation in Trypanosoma brucei: joining the DOTs. *PLoS biology,* **6,** e185.

Thon, G., Baltz, T. and Eisen, H. (1989). Antigenic diversity by the recombination of pseudogenes. *Genes & development,* **3, pp** 1247.

Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F., Skarecky, D. and Ayala, F. J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proceedings of the National Academy of Sciences,* **90, pp** 1335.

Urbina, J. A. (2001). Specific treatment of Chagas disease: current status and new developments. *Current opinion in infectious diseases,* **14, pp** 733.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *British medical bulletin,* **41, pp** 105.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* **10, pp** 57-63.

WHO. 2010. *Chagas Disease* [Online]. WHO Media Centre. Available: http://www.who.int/mediacentre/factsheets/fs340/en/index.html [Accessed 11/08/2011 2011].

Yeo, M., Acosta, N., Llewellyn, M., Sánchez, H., Adamson, S., Miles, G. A. J., López, E., González, N., Patterson, J. S. and Gaunt, M. W. (2005). Origins of Chagas disease: Didelphis species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of Trypanosoma cruzi II, including hybrids. *International Journal for Parasitology,* **35, pp** 225-233.

Yeo, M., Mauricio, I. L., Messenger, L. A., Lewis, M. D., Llewellyn, M. S., Acosta, N., Bhattacharyya, T., Diosque, P., Carrasco, H. J. and Miles, M. A. (2011). Multilocus Sequence Typing (MLST) for Lineage Assignment and High Resolution Diversity Studies in *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases,* **5,** e1049.

Zingales, B., Andrade, S. G., Briones, M. R. S., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M. & Machado, C. R. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz,* **104, PP** 1051-1054.