Genetic profiling of the isoprenoid and sterol biosynthesis pathways of *Trypanosoma cruzi*

Raúl Cosentino[1], Patricio Diosque[2], and Fernán Agüero*[1]

[1] Instituto de Investigaciones Biotecnológicas – Instituto Tecnológico de Chascomús, Universidad de San Martín – CONICET, B 1650 HMP, San Martín, Argentina

[2] Instituto de Patología Experimental, Facultad de Ciencias de la Salud, Universidad Nacional de Salta, A 4408 FVY, Salta, Argentina

Version 1 **18.Jul.2013**

Correspondent author information:
Email: fernan@unsam.edu.ar, fernan.aguero@gmail.com
Phone: +54 11 4006-1500 ext 2110
Fax: +54 11 4006-1559

Keywords: Trypanosoma cruzi, sterol biosynthesis pathway, terpenoid, isoprenoid biosynthesis pathway, mevalonate pathway, single-nucleotide polymorphisms, re-sequencing, genetic diversity, comparative genomics

Abbreviations: SBP, sterol biosynthesis pathway, SNP, single nucleotide polymorphism; DTU, Discrete Typing Unit; PMK, phosphomevalonate kinase; SMO, Sterol Methyl Oxidase

*Correspondent author

Abstract

Background. Sterols such as cholesterol, are important components of cellular membranes. But unlike mammalian cells, the main sterols found in the membranes of trypanosomes and fungi are ergosterol, and other 24-methyl sterols, which are required for growth and viability. In spite of this strict requirement, this group of organisms have evolved different strategies to produce and/or obtain sterols. *Trypanosoma cruzi* is the causative agent of Chagas Disease. In this parasite, one of the few validated targets for chemotherapeutic intervention is the sterol biosynthesis pathway. In this work we present a study of the genetic diversity observed in genes of the isoprenoid and sterol biosynthesis pathways in *T. cruzi*, and a comparative analysis of the diversity found in other trypanosomatids.

Methodology/Principal Findings. Using a number of bioinformatic strategies, we first completed a number of holes in the pathway by identifying the sequences of genes that were missing and/or were truncated in the draft *T. cruzi* genome. Based on this analysis we identified a non-orthologous homolog of the yeast ERG25 gene (sterol methyl oxidase, SMO) and propose that the orthologs of ERG25 have been lost in trypanosomes (but not in leishmanias). Next, starting from a set of 16 *T. cruzi* strains representative of six major evolutionary lineages, we have amplified and sequenced ~ 24Kbp from 18 genes of the pathway, and identified a total of 975 SNPs or fixed differences, of which 28% represent non-synonymous changes. We observed different patterns of accumulation of nucleotide changes for different genes of the pathway, from genes with a density of substitutions ranging from those close to the average (~ 2.5/100 bp) to some showing a high number of changes (11.4/100 bp, for a putative lathosterol oxidase gene). The majority of genes are under apparent purifying selection. However, two genes (TcPMK, TcSMO-like) have a ratio of non-synonymous to synonymous changes that is close to neutrality. None of the non-synonymous changes identified affect a catalytic or a ligand binding site residue. However, after mapping these changes on top of available structural data, we identified a number of changes that are in the close vicinity (7Å) of key residues, and that could therefore be functionally important. A comparative analysis of the corresponding *T. brucei* and *Leishmania* genes, obtained from available complete genomes highlights a high degree of conservation of the pathway, but with differences in the genes that are under apparent purifying selection in each case.

Conclusions/Significance. We have identified a number of genes of the sterol biosynthesis pathway that were missing from the *T. cruzi* genome assembly. Also, we have identified unequal apparent selection acting on these genes, which may provide essential information for the future of drug development studies focused on this pathway.

Introduction

Trypanosoma cruzi, a protozoan parasite of the order Kinetoplastida, is the causative agent of Chagas Disease, a neglected disease that is endemic in South America, affecting in excess of 8 million people (*Rassi, Rassi & Marin Neto*, 2010). The currently available drugs used to treat Chagas' Disease (Nifurtimox, Benznidazole) have several drawbacks including toxicity, and the fact that they are mostly effective during the acute phase of the infection.

The *T. cruzi* species has a structured population, with a predominantly clonal mode of reproduction, with infrequent genetic exchange (*Tibayrenc & Ayala*, 2002, 2013). Through the use of a number of genetic markers the population has been divided into six evolutionary lineages (*Barnabé, Brisse & Tibayrenc*, 2000; *Brisse, Barnabé & Tibayrenc*, 2000; *Zingales, Miles, Campbell et al.*, 2012). Lineages TcV and TcVI (this latter lineage includes the strain used for the first genomic sequence of *T. cruzi*, CL Brener) have a very high degree of heterozygosis. The currently favoured hypothesis suggests that these two lineages originated after one or two ancestral hybridization events (*Machado & Ayala*, 2001; *Westenberger, Barnabé, Campbell et al.*, 2005; *Flores López & Machado*, 2011). The estimated time of divergence of these lineages is 1-4 Myr, which suggests that the diversification of *T. cruzi* was linked to the origin of its blood-sucking triatomine vectors, well before the contact with humans in South America (*Flores López & Machado*, 2011). This divergence was accompanied with a diversification of phenotypic and biological properties. Indeed, several investigations suggest that the observed diversity in host preference, cell tropism and drug susceptibility might be properties of different strains and/or lineages (*Buscaglia & Di Noia*, 2003; *Vago, Andrade, Leite et al.*, 2000; *Andrade, Machado, Chiari et al.*, 1999; *Andrade, Galvão, Meirelles et al.*, 2010).

The parasite ergosterol biosynthesis pathway is one of the major routes for chemotherapeutic intervention against *T. cruzi*. Inhibitors that block sterol biosynthesis or the biosynthesis of isoprenoid precursors inhibit growth of the parasite and cause severe morphological defects (*Urbina*, 2009). Triazole derivatives that inhibit the parasite C14-α sterol demethylase are the most promising compounds, with proved curative activity in murine models of acute and chronic Chagas disease (*Urbina, Payares, Molina et al.*, 1996; *Silva, de Meirelles, Almeida et al.*, 2006; *Lepesheva, Villalta & Waterman*, 2011; *Buckner & Urbina*, 2012). And one of them (posaconazole) is undergoing a number of clinical trials. Other ergosterol biosynthesis inhibitors with good *in vitro* and *in vivo* potency, include those that target 3-hydroxy-3-methyl-glutaryl-CoA reductase, farnesyl diphosphate syntetase (*Rosso, Szajnman, Malayil et al.*, 2011), squalene synthase(*Rodrígues Poveda, González Pacanowska, Szajnman et al.*, 2012), lanosterol synthase (*Buckner, Griffin, Wilson et al.*, 2001), squalene epoxidase (*Gerpe, Odreman Nuñez, Draper et al.*, 2008; *Gerpe, Alvarez, Benítez et al.*, 2009) and 24-C sterol methyl transferase (*Magaraci, Jimenez Jimenez, Rodrigues et al.*, 2003; *Lorente, Rodrigues, Jiménez Jiménez et al.*, 2004; *Braga, Magaraci, Lorente et al.*, 2005), as well as compounds with dual mechanisms of action (ergosterol biosynthesis inhibition and free radical generation) (reviewed in *Urbina*, 2009; *Urbina & Docampo*, 2003).

The azoles, like the triazoles, are used extensively for the treatment of fungal infections with excellent results, though different resistant strains have appeared over time in different species. One of the main resistance mechanisms observed is based on a diminished affinity of the target enzyme for the compound; which is caused by specific mutations in the gene. Different point mutations were identified in several fungi species as responsible for this azole resistance (reviewed in *Lupetti, Danesi, Campa et al.*, 2002; *Morio, Loge, Besse et al.*, 2010).

In this paper we analyze the genetic diversity present in the ergosterol biosynthesis pathway of *T. cruzi* and describe the apparent selection acting on these genes.

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

Results

Filling pathway holes: genes involved in sterol biosynthesis in T. cruzi

To analyze the genetic diversity of the *T. cruzi* sterol biosynthesis pathway (SBP) we decided to sequence all enzymes of the pathway, starting from enzymes that produce the terpenoid backbone precursors, and going down to the last enzyme that produces ergosterol as a product. Therefore, as a first step we looked for *T. cruzi* genes that were mapped to the corresponding KEGG metabolic pathway maps (*Kanehisa, Goto, Sato et al.*, 2012). SBP genes in KEGG are classified in two maps: the steroid biosynthesis pathway map (TCR00100, http://www.genome.jp/kegg/pathway/tcr/tcr00100.html, and the terpenoid backbone biosynthesis pathway map (TCR00900, http://www.genome.jp/kegg/pathway/tcr/tcr00900.html). These maps contain information derived from the *T. cruzi* CL-Brener reference genome. From this analysis we were able to identify 15 genes mapped to these pathways. However, we also detected a number of holes in the pathway: enzymatic reactions with no enzyme mapped, and cases in which the enzymes available in KEGG were truncated (probably because of genome assembly problems). Therefore, before attempting to amplify and sequence the corresponding genes, we invested some effort in analyzing the existing sequence data to obtain a relevant complement of genes. As mentioned, in one case the corresponding genes from the reference genome were truncated, probably because of genome assembly problems. This was the case of the isopentenyl-diphosphate delta-isomerase gene (TcIDI1), which was cloned and sequenced by Dr. TK Smith (unpublished, GenBank accession number: AJ866772, 1071 bp). The corresponding genes in KEGG, mapped from the CL-Brener genome (Esmeraldo-like and non-Esmeraldo-like alleles) were both shorter, at 537 bp (TcCLB.408799.19), and 540 bp (TcCLB.510431.10). Therefore for this work we used the full-length TcIDI1 sequence obtained from GenBank (AJ866772).

To fill in other identified gaps, we used the *Saccharomyces cerevisiae* sterol biosynthetic pathway as a reference model. The yeast SBP has been studied extensively, and is essentially complete in pathway databases. Using the yeast genes mapped to this pathway, we looked for orthologs in *T. cruzi* by doing sequence similarity searches (BLAST) against the complete *T. cruzi* genome or using databases of orthologs compiled from complete genome data, such as the OrthoMCL database (*Chen, Mackey, Stoeckert et al.*, 2006). As a result of this strategy, we were able to map five additional genes (see Table 1), which are the orthologs of the *S. cerevisiae* genes: ERG13 (3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase), ERG7 (lanosterol synthase), ERG26 (C-3 sterol dehydrogenase), ERG3, (C-5 sterol desaturase), and ERG5 (C-22 sterol desaturase). These genes were present in the *T. cruzi* genome, but were not mapped to the corresponding metabolic maps. In all these cases except for one (ERG26), the identification of the corresponding ortholog did not present further complications. The putative *T. cruzi* ortholog of the yeast ERG26 gene (TcCLB.510873.10, C3-sterol dehydrogenase) was found in the *T. cruzi* genome database as a truncated gene of 675 bp (224 aa). This may explain why it was not mapped to the corresponding pathway map in KEGG. Reasoning that this might be a consequence of assembly problems, as in the case of the TcIDI1 gene, we set out to identify the missing portions of this gene by performing BLAST searches against a database of unassembled genomic reads (GSS or WGS reads) from the CL-Brener genome project (data provided originally by the TIGR-SBRI-KI sequencing consortium). Starting with the truncated TcCLB.510873.10 sequence as query, we identified a number of matching sequence reads (BLASTN, E-value < 10^{e-40}). These sequences were then assembled into a single contiguous sequence, which was used as a query in successive rounds of BLASTN searches, followed by reassembly. This iteration cycle was repeated until we recovered the complete (full-length) sequence of the *T. cruzi* putative C3-sterol dehydrogenase

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

gene, as judged by its alignment against the yeast ERG26 gene. This reconstructed full-length sequence of the *T. cruzi* C-3 sterol dehydrogenase gene has 1,221 bp and has been used to design primers for amplification from *T. cruzi* DNA (see Methods). The final sequence, obtained after amplification and sequencing from CL-Brener DNA, was submitted to GenBank under the accession number JN050853.

At this point, there were still two gaps present when modeling the *T. cruzi* SBP on top of the yeast pathway. These correspond to the enzymatic reactions catalyzed by the yeast genes ERG27 (3-keto sterol reductase) and ERG25 (C-4 methyl sterol oxidase). Our failure at identifying the orholog of the yeast ERG27 gene came as no surprise, as the enzymes performing C-3 ketoreduction in land plants and sterol synthesizing bacteria are still unknown (*Bouvier, Rahier & Camara*, 2005; *Desmond & Gribaldo*, 2009).

However, a number of putative homologs of ERG25 in *T. cruzi* were readily identified in sequence similarity searches. In this case, it was difficult to discern which of these were the true orthologs of the yeast ERG25 gene.

Loss of ERG25 homologues in T. cruzi and T. brucei

A BLASTP search using the yeast ERG25p as query against kinetoplastid proteomes retrieved 5 *T. cruzi* significant hits, with 4 of them having identical length (278-279 aa, See Supplementary Excel File S1), sharing a common Pfam Domain (PF04116, Fatty Acid Hydroxylase Superfamily), and a common architecture with 3-4 predicted *trans*-membrane domains (the fifth gene is a truncated copy). These hits correspond to two pairs of alleles, and are currently annotated as 'C-5 sterol desaturases' in the *T. cruzi* genome database (or 'lathosterol oxidases', both are synonymous terms). Because there is detectable similarity between ERG3 and ERG25 genes, a similar search using the yeast ERG3p as query retrieves the same set of *T. cruzi loci*. Therefore, to identify which *loci* correspond to the true orthologs of ERG25, we carried out a detailed phylogenetic analysis between these genes, in different organisms. Using the sequences of fungi, plant, insect, human, fish and trypanosomatid enzymes, we obtained a maximum-likelihood tree with a clear segregation of ERG25 and ERG3 orthologs in separate branches (Figure 1). The tree suggests that there is one *T. cruzi locus* (containing TcCLB.473111.10 and its TcIII-like allele, TcCLB.507853.10) that is the true ortholog of the yeast ERG3 gene. Interestingly, the branch containing ERG25 orthologs does not show genes from African or American trypanosomes (the only kinetoplastid genes are from Leishmanias). The other trypanosomatid genes identified in our BLAST searches are grouped together in a third branch, carrying only trypanosomatid genes. These results are consistent with reciprocal best-hits identified in BLAST searches (see Fig S1). Reciprocal or bidirectional best hits provide support for the conjecture that the genes are equivalent orthologs (*Wolf & Koonin*, 2012; *Koonin*, 2005). In these searches, there is a clear drop in the BLAST score when the query sequence is ERG3 or a ERG3 homolog. This score drop is not observed when the query sequence is ERG25 or an ERG25 homolog. Apart from this evidence, there is additional support for this separate branch from the BLAST searches, in the observed reciprocity of best hits. For example when using the *L. major* gene LmjF.36.2540 as query (grouped with ERG25 orthologs in Fig 1) the best hits are the yeast ERG25 and their orthologs. And viceversa, when doing the reciprocal BLAST search using the yeast ERG25p as query, the best *L. major* hit is again LmjF.36.2540, even though in this case the scores are lower than for those obtained for ERG3 orthologs. These bidirectional hits are not observed for the group of trypanosomatid genes grouped in the middle branch in the phylogenetic tree.

Taken together, the most parsimonious interpretation of these results is that the C-4 sterol oxidase gene has been lost in the ancestor of *T. cruzi* and *T. brucei* (but not in the ancestor of

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

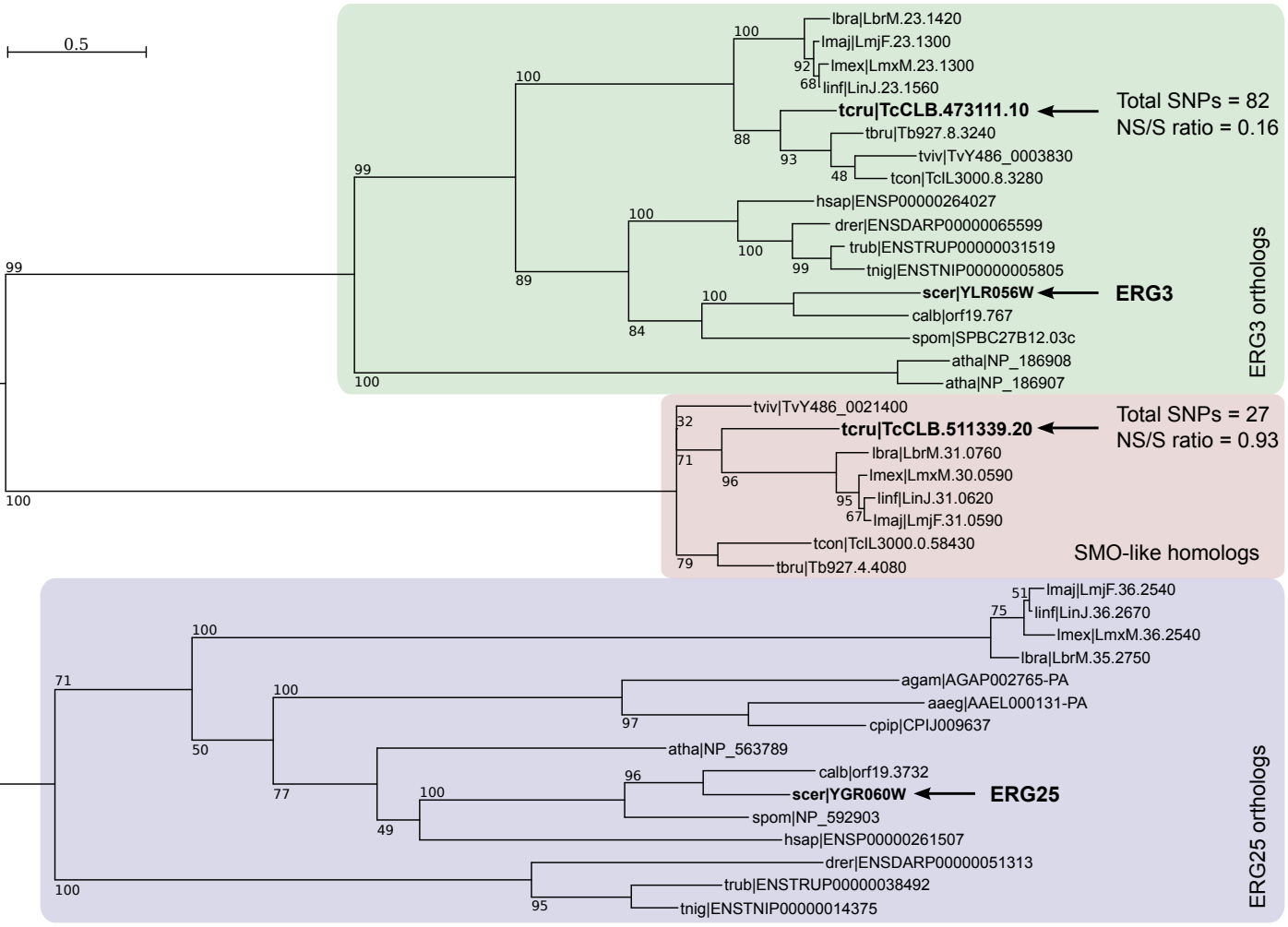

Figure 1: Phylogenetic tree of selected ERG3/ERG25 orthologs. Seleced orthologs from kinetoplastids, plants, vertebrates, invertebrates and fungi were aligned with t_coffee. A phylogenetic reconstruction was calculated using PhyML (LG model, bootstrap resampling with N=1000). For clarity, highly similar genes/alleles were not included in the final tree. Organism abbreviations are: lbra = *L. braziliensis*; lmex = *L. mexicana*; linf = *L. infantum*; lmaj = *L. major*; tcru = *T. cruzi*; tviv = *T. vivax*; tcon = *T. congolense*; tbru = *T. brucei*; agam = *Anopheles gambiae*; aaeg = *Aedes aegypti*; cpip = *Culex pipiens*; hsap = *Homo sapiens*; calb = *Candida albicans*; scer = *S. cerevisiae*; spom = *Schizosaccharomyces pombe*; atha = *A. thaliana*; drer = *Danio rerio*; trub = *Takifugu rubripes*; tnig = *Tetraodon nigroviridis*.

Leishmanias). The presence of Leishmanial and Trypanosomal proteins grouped in the middle branch of the tree suggests that these group of homologs have a different ancestral origin (see Discussion). Because these genes have significant sequence similarity to ERG25 we decided to call them Sterol Methyl Oxidase-like (SMO-like) in this work.

The Table 1 summarizes the results of our efforts to close pathway gaps. Once the *T. cruzi* SBP genes were identified, we proceeded to study the genetic diversity by re-sequencing these genes in a panel of *T. cruzi* strains.

Genetic diversity in the sterol biosynthesis pathway

To obtain sequence information from the selected genes we decided to use a methodology based on PCR amplification followed by direct sequencing. Therefore, it was important to reduce the possibility of amplification problems generated by polymorphisms that could prevent the annealing of the primers. A number of aspects were considered to reduce these risks i) we decided

S. cerevisiae					T. cruzi				
SBP genes	**Molecular function**	**EC Number**	**Pathway order**	**Ortholog group** [1]	**Current annotation** [2]	**Locus Identifier(s)** [2]	**Ortholog group** [3]	**In KEGG?** [4]	**Gene Name** [5]
ERG10 YPL028W	Acetoacetyl-CoA thiolase	2.3.1.9	1	OG4_10214	Hypothetical protein	TcCLB.511003.60		Yes	TcACAT
ERG13 YML126C	3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase	2.3.3.10	2	OG4_11016	Hypothetical protein, conserved	TcCLB.511903.40 TcCLB.511071.50		**NO**	TcHMGS
HMG 1 YML075C	HMG-CoA reductase	1.1.1.34	3	OG4_11458	3-hydroxy-3-methylglutaryl-CoA reductase, putative	TcCLB.509167.20 TcCLB.506831.40		Yes	TcHMGR
ERG12 YMR208W	Mevalonate kinase	2.7.1.36	4	OG4_11698	Mevalonate kinase, putative	TcCLB.436521.9 TcCLB.509237.10		Yes	TcMK
ERG8 YMR220W	Phosphomevalonate kinase	2.7.4.2	5	OG4_15366	Phosphomevalonate kinase-like protein, putative	TcCLB.508277.140 TcCLB.507913.20		Yes	TcPMK
ERG19 YNR043W	Mevalonate pyrophosphate decarboxylase	4.1.1.33	6	OG4_11688	Diphosphomevalonate decarboxylase, putative	TcCLB.507993.330 TcCLB.511281.40		Yes	TcMVD
ERG20 YJL167W	Farnesyl pyrophosphate synthetase	2.5.1.10	7	OG4_11009	Farnesyl pyrophosphate synthase, putative	TcCLB.511823.70 TcCLB.508323.9		Yes	TcFPPS
IDI1 YPL117C	Isopentenyl diphosphate isomerase (IPP isomerase)	5.3.3.2	7'	**OG4_12197**	Isopentenyl-diphosphate delta-isomerase, putative	TcCLB.408799.19 [6] TcCLB.510431.10 [6]	**OG4_14499**	Yes [6]	TcIDI
ERG9 YHR190W	Squalene synthase	2.5.1.21	8	OG4_13084	Farnesyl transferase, putative Squalene synthase, putative	TcCLB.507897.20 TcCLB.508369.20		Yes	TcSQS
ERG1 YGR175C	Squalene epoxidase	1.14.99.7	9	OG4_13490	Squalene monooxigenase, putative	TcCLB.509589.20 TcCLB.503999.10		Yes	TcSQLE
ERG7 YHR072W	Lanosterol synthase (oxidosqualene cyclase)	5.4.99.7	10	OG4_11381	Lanosterol synthase, putative	TcCLB.508175.70 TcCLB.506825.170		**NO**	TcOSC
ERG11 YHR007C	Lanosterol 14-α-demethylase	1.14.13.70	11	OG4_12975	Lanosterol 14-α-demethylase, putative	TcCLB.506297.260 TcCLB.510101.50		Yes	Tc14DM TcCYP51
ERG24 YNL280C	C-14 sterol reductase	1.3.1.70	12	OG4_12018	C-14 sterol reductase, putative	TcCLB.507969.60 TcCLB.507129.30		Yes	TcC14SR
ERG25 YGR060W	C-4 methyl sterol oxidase	1.14.13.72	13	**OG4_13007**	—	—	—	—	
—	—	—	—	—	C-5 sterol desaturase, putative	TcCLB.511339.20 TcCLB.511895.69[7] TcCLB.509235.20	**OG4_20087**	**NO**	TcSMO-like
ERG26 YGL001C	C-3 sterol dehydrogenase	1.1.1.170	14	OG4_12533	NAD(P)-dependent steroid dehydrogenase protein, putative	TcCLB.510873.10 [8]		**NO**	TcNSDHL
ERG27 YLR100W	3-keto sterol reductase	1.1.1.270	15	OG4_16167	—	—	—	**NO**	—
ERG6 YML008C	Delta(24)-sterol C-methyltransferase	2.1.1.41	16	OG4_13307	Sterol 24-C-methyltransferase, putative	TcCLB.504191.10 TcCLB.505683.10 TcCLB.510185.10		Yes	Tc24SMT
ERG2 YMR202W	C-8 sterol isomerase	5.-.-.-	17	OG4_14573	C-8 sterol isomerase, putative	TcCLB.510329.90		Yes	Tc8SI
ERG3 YLR056W	C-5 sterol desaturase (lathosterol oxidase)	1.3.3.-	18	OG4_12421	Lathosterol oxidase, putative	TcCLB.473111.10 TcCLB.507853.10		**NO**	TcSC5D
ERG5 YMR015C	C-22 sterol desaturase	1.14.14.-	19	OG4_14688	Cytochrome p450-like protein, putative	TcCLB.506945.190		**NO**	TcSC22D
ERG4 YGL012W	C-24(28) sterol reductase	1.3.1.71	20	OG4_16908	Sterol C-24 reductase, putative	TcCLB.506577.120 TcCLB.507709.90		Yes	TcSC24R

[1] OrthoMCL Database, version 4 (http://orthomcl.org)
[2] From the CL-Brener Genome Project (El-Sayed N, 2005; currently available at http://TriTrypDB.org)
[3] Only mentioned if it differs from the yeast putative ortholog
[4] Is gene currently mapped to the corresponding KEGG metabolic map?
[5] Nomenclature used in this work. Some of these names were already used for the corresponding trypanosomatid genes.
[6] Both alleles are truncated copies of the gene. The full-length gene is deposited in GenBank as AJ866772
[7] The TcCLB.511895.69 allele is shorter, and probably truncated.
[8] The copy annotated by the genome project is truncated due to genome assembly problems. The full-length gene is deposited in GenBank as JN050853

Table 1: The *S. cerevisiae* sterol biosynthesis pathway genes and their orthologs in *Trypanosoma cruzi*.

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

to focus our analysis on coding sequences where possible, as they are generally less variable than non-coding sequences; ii) we used the *T. cruzi* SNP database (TcSNP, http://snps.tcruzi.org) (*Ackermann, Carmona & Agüero*, 2009), to select the best regions for primer design, avoiding regions with candidate SNPs; and iii) we also limited the size of amplification products, to optimize the quality of the final sequence (see Methods). In this latter case, an optimal size was set to ~ 750–800 bp, which would allow us to get complete sequence coverage, with good quality on both strands. Depending on the size of each gene, one or more overlapping amplification products had to be analyzed (the number of amplification products per gene is listed in Table 2).

All amplification products were analyzed with a software package (PolyPhred, see Methods) that allows the identification of heterozygous peaks in the chromatograms. Therefore, for all selected genes, we identified two types of sequence polymorphims: i) allelic variation within a strain/clone (heterozygous peaks identified by Polyphred); and ii) variation between strain/clones (identified in a multiple sequence alignment of sequenced products). For every gene, we identified the position of the variant sites, and the type of change introduced (synonymous or non synonymous). The complete information for each gene is available as supplementary material (Table S2), and a table summarizing these results is included herein (Table 2). Using this strategy, we generated ~ 24 Kb of sequence, from 16 strains covering the 6 major *T. cruzi* evolutionary lineages. Overall we attained a coverage of ~ 90% of the total coding sequence for the selected genes. From this analysis we identified 975 polymorphic sites, producing 978 codon changes, which generate 692 synonymous changes (72%) and 273 non-synonymous changes (28%, see Table 2). The genes Tc24SMT, TcHMGR and TcNSDHL had the lowest SNP density (2.68, 2.74 and 2.50 every 100bp respectively), while TcSC5D had the highest SNP density (11.39 every 100bp). This density is at least two times higher than that found in other SBP genes. However, at the same time, this gene had one of the lowest ratios of non-synonymous SNPs over synonymous SNPS (NS/S ratio). Upon further investigation of this gene, we noticed a number of informative SNPs that could be exploited in a lineage typing assay. Therefore, we performed a separate re-sequencing experiment for this gene, in an expanded panel of strains, and developed two alternative typing assays based on this *locus* (published separately, *Cosentino & Agüero*, 2012).

Other genes with low NS/S ratios were TcHMGS and Tc14DM (0.15), indicating a high degree of conservation in the panel of strains analyzed. On the other extreme, the genes TcSMO-like and TcPMK had the highest NS/S ratios (0.93 and 0.94 respectively) and together with TcMVD have the highest non-synonymous SNP density (between 1.8 and 2.5 non synonymous SNPs every 100bp). The average NS/S ratio in our dataset was 0.4, and in no case was this ratio higher than 1, indicating the relatively high degree of conservation of *T. cruzi* SBP genes.

Lineage-specific and intra-lineage differences.

All the collected data on sequence diversity was also analyzed in the context of the current separation of *T. cruzi* in discrete evolutionary lineages. For this analysis we considered three types of polymorphisms or fixed differences: i) lineage specific polymorphisms (LSPs) (these are the polymorphic sites that could differentiate one lineage from all the others; ii) intra-lineage polymorphisms (ILP) (differences between strains of the same lineage); and iii) heterozygous sites (those that differ between alleles of a single diploid individual). In this comparative analysis we found that the lineage TcI showed more than one LSP every 200 bp, while in the other lineages, the LSP density was at least four times lower (data not shown). This data agrees with the hypothesis that TcI is one of the ancestral lineages *Flores López & Machado* (2011); *Zingales, Miles, Campbell et al.* (2012). Analyzing the ILP distribution, it was found that the TcIV lin-

	Polymorphisms (SNPs)						Length (bp)		
	Synonymous (S)		Non synonymous (NS)		Ratio	Total	CDS	Resequenced	
Genes	No. (%)	density	No. (%)	density	NS/S	No.	bp	bp (%)	PCR Fragments
TcACAT	35 (75.00)	3.27	11 (25.00)	1.09	0.33	48	1188	1100 (92.59)	2
TcHMGS	37 (86.05)	2.66	6 (13.95)	0.43	0.16	43	1497	1390 (92.85)	2
TcHMGR	27 (79.41)	2.18	7 (20.59)	0.56	0.26	34	1356	1240 (91.45)	2
TcMK	31 (79.49)	3.20	12 (30.77)	1.24	0.39	50*	1086	970 (89.32)	2
TcPMK	35 (53.85)	2.65	33 (50.77)	2.50	0.94	69*	1428	1320 (92.44)	2
TcMVD	32 (62.75)	3.08	19 (37.25)	1.83	0.59	51	1140	1040 (91.23)	2
TcFPPS	22 (66.67)	2.27	11 (33.33)	1.13	0.50	36*	1275	970 (76.08)	2
TcIDI	25 (78.13)	2.50	7 (21.88)	0.70	0.28	32	1071	1000 (93.37)	2
TcSQS	32 (69.57)	2.88	14 (30.43)	1.26	0.44	46	1212	1110 (91.58)	2
TcSQLE	50 (72.46)	3.07	21 (30.43)	1.29	0.42	73*	1716	1630 (94.99)	3
TcOSC	61 (66.30)	2.53	31 (33.70)	1.29	0.51	92	2706	2410 (89.06)	4
Tc14DM	47 (87.04)	3.46	7 (12.96)	0.51	0.15	54	1443	1360 (94.25)	2
Tc14SR	33 (63.46)	2.58	19 (36.54)	1.48	0.58	52	1371	1280 (93.36)	2
TcSMO-like	14 (51.85)	2.09	13 (48.15)	1.94	0.93	27	837	670 (80.05)	1
TcNSDHL	21 (70.00)	1.75	9 (30.00)	0.75	0.43	30	1221	1200 (98.28)	2
Tc24SMT	17 (77.27)	2.07	5 (22.73)	0.61	0.29	22	1050	820 (78.10)	1
Tc8SI	18 (78.26)	3.16	5 (21.74)	0.88	0.28	23	654	570 (87.16)	1
TcSC5D	70 (89.74)	9.72	11 (14.10)	1.53	0.16	82*	834	720 (86.33)	1
TcSC22D	29 (69.05)	2.09	13 (30.95)	0.94	0.45	42	1518	1390 (91.57)	2
TcSC24R	53 (75.71)	3.79	17 (24.29)	1.21	0.32	70	1467	1400 (95.43)	2
Total/Avg	692 (71.71)	2.93	273 (28.29)	1.16	0.39	975	26070	23590 (90.50)	39

Table 2: Comparative analysis of the quantity, density and type of SNPs identified in the sterol biosynthesis pathway of *Trypanosoma cruzi*. SNP counts marked with * indicate totals that do not result from the sum of synonymous + non-synonymous substitutions, in all cases because of the presence of two substitutions in the same codon.

eage presents the highest density of ILPs, with almost one ILP every 100 bp. Interestingly > 90% of these differences are restricted to the comparison of the CanIII strain (former TcIIa lineage) with the other two strains from this lineage (Dog Theis and 92122102R, data not shown). The number of differences between CanIII and these other TcIV strains was even greater than those observed between lineage TcV and TcVI (the two hybrid lineages). These results agree with previous publications that describe a considerable phylogenetic distance between CanIII and Dog Theis (*de Freitas, Augusto Pinto, Pimenta et al.*, 2006), even though they are still placed into the same evolutionary lineage based on the current typing methods. As expected, the hybrid lineages (TcV and TcVI) show the highest level of heterozygosity, with more than 3 herozygote sites every 200bp. Lineage II (former Tcruzi IIb) showed an intermediate level of heterozygosity with aproximately 3 sites every 2Kb, while the other lineages showed less than 10 sites in the 20Kb analyzed.

Potentially important non-synonymous changes.

In many cases, variations in susceptibility or resistance to a drug are associated to mutations that occur near the interaction site of a substrate or an inhibitor. To analyze the potential relevance of the non-synonymous SNPs found, we gathered relevant information from the literature on the selected targets, their PFAM domains, PROSITE motifs, and the available crystallographic structures in the Protein Data Bank (PDB, http://www.rcsb.org) for the *T. cruzi* enzyme or their orthologs (see Methods). Starting with these crystal structures, we first identified residues near the co-crystalized ligands (substrate, inhibitor or co-factor). For this we established a maximum distance limit of 7Å from each corresponding ligand. This analysis revealed a number of

potentially important changes in the Tc14DM (TcCYP51), and TcMK genes.

In the azole target gene, Tc14DM (TcCYP51), we identified only 5 non-synonymous substitutions, none of which affect key residues of the enzyme (see Figure 2). However, after mapping these substitutions on top of the available TcCYP51 structure (*Chen, Leung, Guilbert et al.*, 2010), we identified a number of potentially important substitutions that lie just next to important residues, or are within 7Å of the co-crystalized ligand. One of these is a A117S substitution that sits just next to a tyrosine (Y116) that is within 7Å of the ligand, and that has been shown to be involved in the generation of resistance to azoles in *C. albicans* and *U. necator* (*Kelly, Lamb & Kelly*, 1999; *Délye, Laigret & Corio Costet*, 1997). In trypanosomatids, residue 117 marks the start of a short helix in the structure, that is uniquely found in trypanosomatids (*Lepesheva, Park, Hargrove et al.*, 2010). Alanine at residue 117 is present in *T. brucei* and was found in strains from the TcIII lineage (former Tcruzi IIc) in homozygosis and in heterozygosis in hybrid strains, while Serine at this position was found in homozygosis in strains from lineages TcI, TcII, and TcIV (former Tcruzi I, IIa and IIb).

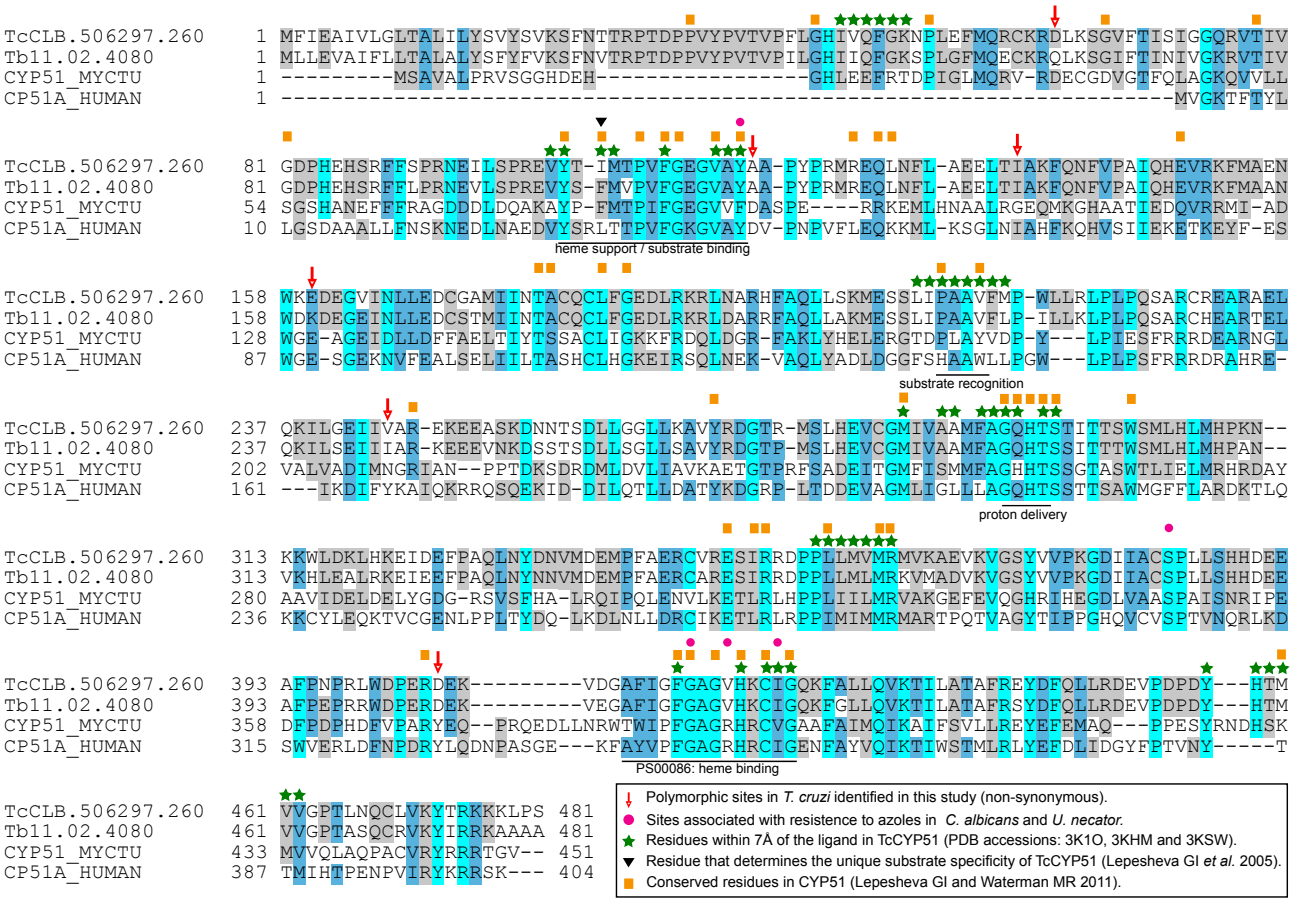

Figure 2: Alignment of *T. cruzi*, *T. brucei*, *M. tuberculosis* and human Lanosterol 14-α demethylases, showing the non-synonymous changes identified in this work (red arrows). Important residues either in Tc14DM or in the CYP51 family are noted (*Lepesheva, Zaitseva, Nes et al.*, 2006; *Lepesheva & Waterman*, 2011), as well as residues associated with resistance to azoles in *C. albicans* and *U. necator* (*Kelly, Lamb & Kelly*, 1999; *Délye, Laigret & Corio Costet*, 1997). PS00086 is the Prosite Cytochrome 450 motif (cysteine heme-iron ligand signature)

In the case of the TcMK gene, we identified 12 non-synonymous changes. One of these is a H29Y change, located in a conserved region of the sequence, close to a group of residues that are involved in substrate binding (at <7Å of the ligand (mevalonate) in the *L. major* structure (LmjF.31.0560, PDB: 2HFU) (*Sgraja, Smith & Hunter*, 2007). Histidine is encoded by the genes

from strains M5631 and X109/2, both from lineage TcIII, while tyrosine is encoded in all other strains, including M6241, also of lineage TcIII. The second possibly important change in TcMK is a G287A substitution, also located in a very conserved motif of the MK gene, that is part of the GHMP kinase family domain (Pfam PF08544). That particular motif, contains a number of highly conserved glycines, some of which are within 7Å of the ligand in the *L. major* MK. In this polymorphic position strains from lineage TcV (Sc43 and Mn) encode glycine and alanine in heterozygosis, while all the other strains encode glycine in homozygosis.

For other targets, crystallographic structures were not available, and so we resorted to analyze non-synonymous substitutions in the context of functionally important motifs or domains. In the case of the squalene epoxidase gene (TcSQLE), we identified two changes (S306G and I307L) that stand between highly conserved proline residues (at 95% and 94% identity within the squalene epoxidase domain, PF08491). The I307L substitution is apparently conservative, as these are the two most frequent residues in this position in the Pfam domain. In contrast, the S306G substitution introduces a glycine that is completely absent at this position in the 429 sequences currently available for this protein family in the Pfam database. Serine is the second most frequent residue in this position, and is present in 14% of the sequences, all from fungi (glutamine is the most frequent residue, in 50% of the sequences). The Ser residue occurs in lineage TcII in homocygosis, and in lineages TcV and TcVI in heterocygosis, and its conservation in fungi suggests that this is the ancestral character at this position. Finally in the C-14 sterol reductase gene (Tc14SR), we identified two consecutive substitutions (P208H, V209F) in strains 92122102R and Dog Theis (lineage TcIV), that fall within the sterol reductase family signature motif (PS01017). However the substitutions are conservative, because they are both described by the motif pattern.

The phosphomevalonate kinase of T. cruzi is not under strong purifying selection

The phosphomevalonate kinases (PMKs) of pathogenic bacteria, fungi, and trypanosomes are attractive targets for the design of selective inhibitors, because the same phosphorylation reaction in humans and other animals is catalyzed by a non-orthologous enzyme (*Houten & Waterham*, 2001). In the first group of organisms, phosphorylation of mevalonate is performed by orthologs of the yeast ERG8 gene, while in animals it is performed by a group of orthologs of the human PMK gene (hPMK). These two groups of enzymes differ in a number of kinetic, biophysical properties, and in the ATP-binding motifs (ERG8-like kinases contain a protein kinase motif, while orthologues of the human enzyme have a P-loop or "Walker A" motif)(*Houten & Waterham*, 2001; *Chang, Yan, Gu et al.*, 2008). Analysis of the TcPMK gene in different strains of *T. cruzi* showed that the gene has accumulated 68 changes (35 synonymous, 33 non-synonymous, see Table 2), and 1 non-sense substitution (see below) in these independently evolving lineages. With a ratio of non-synonymous to synonymous differences close to 1 (0.94), the substitutions can be considered largely neutral, and together with the TcSMO-like gene, these are the only two genes in the analyzed pathway that are not under apparent purifying selection. Moreover, an interesting substitution was observed at codon 136 (407bp) in the *T. cruzi* strain IVV (TcII). In this strain, one allele encoded a Serine (TCA), as in all other strains, whereas the second allele encoded a premature STOP codon (TAA). All ERG8-like PMKs are composed of two GHMP kinase domains: an N-terminal domain (Pfam PF00288) located in the middle of the protein (starting at residue 160 in *T. cruzi*), and a C-terminal domain (PF08544) closer to the C-termini. In *T. cruzi*, the nonsense mutation in the IVV strain is located upstream of this first domain (PF00288). Considering that the next possible translational start codon is located

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

downstream of this domain, the IVV strain is therefore probably producing a very short non-functional protein from this allele, or two truncated proteins, both devoid of this domain. Thus, the most plausible hypotesis is that the IVV strain carries only one functional PMK allele, and can be considered naturally hemizygous for the PMK gene.

Comparative analysis of genetic diversity in kinetoplastid sterol biosynthesis pathways

Although phylogenetically related, kinetoplastid parasites have evolved different adaptations to their host environments. This is particularly evident in the case of ergosterol dependency: both *T. cruzi* and *Leishmania* have an essential requirement for ergosterol and/or other 24-methylsterols, and are unable to survive by salvaging cholesterol from the host (*Roberts, McLeod, Rice et al.*, 2003). Sterol biosynthesis in *Trypanosoma brucei*, is apparently suppressed in bloodstream forms, relying instead on receptor-mediated endocytosis of host low-density lipoproteins carrying cholesterol (*Coppens & Courtoy*, 2000). Apart from differences in their strict requirement for *de novo* synthesis of sterols, there are also differences in the exact type and abundance of synthesized sterols, which may be explained by differences in the complement of genes and/or their activities or regulation. As an example, it has been described that *T. cruzi* amastigotes lack $\Delta^{5,7}$ sterols, suggesting the lack of Δ^5 desaturase activity in this stage (*Liendo, Visbal, Piras et al.*, 1999; *Roberts, McLeod, Rice et al.*, 2003). Based on these premises, we decided to investigate the accumulation of nucleotide changes (fixed differences) in genes from the sterol biosynthesis pathway in African Trypanosomes, and Leishmanias, reasoning that the selection acting on these genes could be different in each case. For this we identified the corresponding orthologs of the yeast and *T. cruzi* genes used in this work (see Table 1) in currently available kinetoplastid genomes (see Methods). The available information includes that of *T. brucei brucei* (2 strains) and *T. brucei gambiense*; and that of 4 species of Leishmanias (major, infantum, braziliensis and mexicana). Using this information we proceeded to analyze the nucleotide substitutions observed in each group. A summary of this analysis is presented in Table 3.

Because the genetic space explored in each case is different, it would be perhaps incorrect to compare the number of nucleotide changes observed for each gene between the three groups of organisms. However, it is still possible to analyze and compare the information on sequence diversity within each group (e.g. column-wise in Table 3). When looking at data in this way, a number of observations can be made: even though the genetic diversity in African trypanosomes is the lowest observed (in number of substitutions per 100 bp), the ratios of non-synonymous to synonymous changes are higher than those observed in Leishmania and *T. cruzi* in 13 of 16 cases with data. The genes accumulating less changes in African trypanosomes are the C-24 sterol methyl transferase (24SMT, ortholog of ERG6), and the HMGR gene (HMG CoA reductase). In Leishmanias, although we observe higher densities of substitutions, the ratios of Non-syn to Syn are the lowest overall, suggesting that the pathway is under purifying selection. However, interestingly the two genes with the highest NS/S ratios are the Leishmanial SMO-like and ERG25 orthologs.

Discussion

The main goal of our work was to study the genetic diversity present in the isoprenoid and sterol biosynthesis pathway (SBP) of *T. cruzi*. These pathways are validated chemotherapeutic targets for Chagas Disease, with a number of compounds currently undergoing clinical trials. Because

Yeast gene, Std symbol	Syn SNPs / 100 bp			Non-Syn SNPs / 100 bp			Total SNPs / 100 bp			Ratio Non-Syn/Syn		
	Tbr	Leish	Tcr	Tbr	Leish	Tcr	Tbr	Leish	Tcr	Tbr	Leish	Tcr
ERG10 / ACAT	NA	NA	3.27	NA	NA	1.09	NA	NA	4.36	NA	NA	0.33
ERG13 / HMGS	0.60	9.96	2.81	0.33	**2.59** *	0.43	0.93	**12.55** *	3.24	0.56	0.26	**0.15** *
HMG1 / HMGR	0.08	10.50	2.18	**0** *	3.60	0.56	0.08	14.10	2.74	**0** *	0.34	0.26
ERG12 / MK	0.10	**16.77** †	3.20	0.20	5.56	1.24	0.30	**22.32** †	5.15	**2.00** †	0.33	0.39
ERG8 / PMK	0.56	11.47	2.65	0.42	6.54	**2.50** †	0.99	18.01	5.23	0.75	0.57	**0.94** †
ERG19 / MVD	0.26	14.93	3.08	0.44	5.90	1.83	0.70	20.83	4.90	1.67	0.40	0.59
IDI1 / IDI	0.19	12.61	2.50	0.09	3.08	0.70	0.28	15.69	3.20	0.50	**0.24** *	0.28
ERG20 / FPPS	**0.63** †	**9.55** *	2.27	**0.45** †	3.67	1.13	**1.09** †	13.22	3.71	0.71	0.38	0.50
ERG9 / FDFT , SQS	0.15	10.92	2.88	0.15	5.62	1.26	0.29	16.55	4.14	1.00	0.51	0.44
ERG1 / SQLE	0.40	12.16	3.07	0.23	6.49	1.29	0.63	18.65	4.48	0.57	0.53	0.42
ERG7 / LSS, OSC	0.40	12.00	2.53	0.33	**6.98** †	1.29	0.74	18.98	3.82	0.82	0.58	0.51
ERG11 / 14DM, CYP51	0.41	10.07	3.46	0.28	2.78	0.51	0.69	12.85	3.97	0.67	0.28	**0.15** *
ERG24 / 14SR, TM7SF2	0.65	10.65	2.58	0.29	6.39	1.48	0.94	17.05	4.06	0.44	0.60	0.58
ERG25 / SMO / SC4MOL *	—	12.01	—	—	9.39	—	—	21.40	—	—	**0.78** †	—
— / SMO-like	0.27	10.55	2.09	0.27	6.55	1.94	0.54	17.09	4.03	1.00	0.62	0.93
ERG26 / NSDHL	0.08	10.16	1.75 ♮	0.16	3.66	0.75	0.25	13.83	2.50 ♮	**2.00** †	0.36	0.43
ERG27 / 3KSR	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ERG6 / 24SMT	**0** *	10.45	**2.07** *	**0** *	6.12	**0.61** *	**0** *	16.57	**2.68** *	DBZ	0.59	0.29
ERG2 / 8SI	0.30	12.80	3.16	**0.45** †	6.85	0.88	0.75	19.64	4.04	1.50	0.53	0.28
ERG3 / SC5D	0.12	12.54	**9.72** †	0.12	4.84	1.53	0.25	17.38	**11.39** †	1.00	0.39	0.16
ERG5 / SC22D	NA	12.18	2.09	NA	4.85	0.94	NA	17.03	3.02	NA	0.40	0.45
ERG4 / SC24R	NA	11.47	3.79	NA	6.44	1.21	NA	17.91	5.00	NA	0.56	0.32
Average	0.31	11.67	3.06	0.25	5.18	1.16	0.56	16.86	4.27	0.81	0.44	0.38
Std deviation	0.21	1.80	1.66	0.15	1.50	0.53	0.34	2.69	1.87	0.57	0.13	0.32

TABLE 3: Comparative analysis of the genetic diversity present in the sterol biosynthesis pathway of different kinetoplastids. The table shows the density of synonymous or non-synonymous nucleotide changes (substitutions per 100 bp) observed in the comparison of *T. brucei* (Tbr), *Leishmania* (Leish) and *T. cruzi* (Tcr) genes from different strains and/or species. The lowest (*) and highest (†) figures in each column are marked. NA = sequence not available; DBZ = Division by Zero. In some cases (♮) sequences were not available for all strains.

the *T. cruzi* SBP pathway appeared to be incomplete in metabolic pathway databases such as KEGG when we started this work, and because the annotation of the SBP genes was also incomplete, we had to perform a small-scale bioinformatics analysis to fill in the gaps in available sequence and annotation. This task was performed primarily based on the well studied *S. cerevisiae* ergosterol biosynthesis pathway. As a result of this strategy, the majority of the genes of the pathway have now been identified, with the exception of the orthologs of the yeast gene ERG27. This gene encodes a 3-keto sterol reductase. As mentioned, the gene(s) responsible for this activity in land plants and sterol synthesizing bacteria have not been identified yet (*Desmond & Gribaldo*, 2009). It is therefore highly likely that the trypanosomatid 3-keto sterol reductase is phylogenetically closer to the plant enzymes, and that once this ellusive gene is identified it will be readily identified in trypanosomatids.

We selected 21 genes from this pathway to build a genetic diversity profile from representative strains of the six major evolutionary lineages of *T. cruzi*. For this analysis we used at least 3 strains for each evolutionary lineage (DTU) therefore effectively sampling a large genetic space. Although it is certainly likely that other SNPs or fixed differences can be discovered when sequencing these *loci* from new isolates, most probably these new mutations will correspond to changes that are unique to the new isolate (e.g. introduced during the clonal expansion of this particular isolate). In our experience, when expanding our re-sequencing analysis of the TcSC5D gene to 15 additional strains, 10 new polymorphic sites were identified (see Fig S1 in *Cosentino & Agüero*, 2012), 4 of which came from a new lineage/DTU (TcBat/TcVII *Marcili, Lima, Cavazzana et al.* (2009)) that was not included in this analysis, while 4 others came from a TcIV strain

(see discussion above about intra-lineage diversity of TcIV). This is a highly polymorphic gene, so for the other genes we believe we have sampled a significant genetic space.

The strategy employed consisted in the obtention of overlapping amplification products of approximately 750bp (with 100bp of overlap) for each gene, followed by direct sequencing of amplification products in both strands. As a result, the majority of the sequenced bases were read at least twice. The primers were designed based on the CL-Brener genome sequence, and the majority of them were designed against the corresponding coding sequence to reduce the possibility of amplification problems (under the hypothesis that coding sequences are much less polymorphic than non-coding sequences). Moreover, when designing primers we avoided SNPs already identified from sequences in the public domain by checking against the TcSNP database. This strategy enabled the amplification and sequencing of all the selected gene fragments in strains from all the lineages, except for the first amplification product of the TcMK gene, that could not be amplified initially from lineage I. These strains carry a SNP that is specific for TcI and that mapped exactly at the 3'-end of the forward primer (see Table S3). We fixed this primer after recent genomic data was made available for a number of Tcruzi I strains (Sylvio X10 (*Franzén, Ochaya, Sherwood et al.*, 2011), JR cl4 (accessed trough TriTrypDB (*Aslett, Aurrecoechea, Berriman et al.*, 2010)), and TcAdriana (Westergaard G and Vazquez M, unpublished results)).

TcSC5D was the most polymorphic gene with 11 SNPs/100bp (for a total of 82 SNPs in 720bp analyzed), a SNP density that is at more than twofold larger than those observed for other re-sequenced genes. However, the TcSC5D gene is apparently under strong purifying selection, with one of the lowest ratios of NS/S SNPs in the panel (together with the Tc14DM and TcHMGCoA genes, with values of 0.16, 0.15 and 0.15 respectively (See Table 2). Tc14DM and TcHMGCoA have also the lowest non synonymous SNP density of the panel, with only 0.51 and 0.43 non synonymous SNPs every 100bp, so at least from the genetic evidence alone, these genes would be the best candidates for drug development.

The TcSMO-like genes of T. cruzi are not under strong purifying selection

As described above, when looking for the *T. cruzi* orthologs of the ERG25 (C4-methyl oxidase) gene, we identified two homologs (two pairs of allelic variants from the hybrid genome), which are members of the Fatty Acid Hydroxylase superfamily (Pfam Domain PF04116). Based on best-reciprocal BLAST hits and careful examination we concluded that these genes (TcCLB.511339.20, TcCLB.509235.20) are not orthologous to ERG25, and are probably divergent homologs of a different ancestral gene. In any case, apart from the divergence noticed when analyzing this group of homologs across different organisms (see Figs 1, and S1) in this work we also observed that these genes are also diverging within the *T. cruzi* species, as revealed by the accumulation of non-synonymous nucleotide changes in different *T. cruzi* strains/lineages. This could indicate an ongoing process of *neo*-functionalization of the encoded protein. Interestingly, when performing BLASTP searches against fungal genomes, we noticed a number of SUR2 (Sphinganine C4-hydroxylase) homologs among the significant hits (see Fig S1). In yeast, SUR2 (also a member of the FA hydroxylase superfamily) catalyzes the conversion of sphinganine to phytosphingosine in sphingolipid biosyntheis. Recently, the presence of phytosphingosine in trypanosomes was demonstrated by mass spectrometry (*Vacchina, Tripodi, Escalante et al.*, 2012). In this article the authors also show that the biosynthesis of phytosphingosine is driven by bifunctional hydroxylase/desaturase enzymes. However, the trypanosomatid genes identified as responsible for this activity are not orthologs of yeast SUR2 (see Fig 2 in *Vacchina, Tripodi, Escalante et al.*, 2012). In the reciprocal BLAST searches, using the yeast SUR2 protein as query against trypanosomatid genomes, we always retrieve the same set of ERG3 orthologs, and SMO-like genes. Therefore it

is tempting to speculate that in trypanosomes the orthologs of both SUR2 and ERG25 have been lost, and that the SMO-like genes grouped in the middle branch in Figure 1 could represent an ancestral hydroxylase/desaturase that has adjusted (or is still adjusting) to a new functional niche in these organisms (cellular localization or time/stage of expression, etc). Interestingly the expression of this gene is higher in amastigotes and trypomastigotes, the two life cycle stages that occur in the mammalian host (data from *Minning, Weatherly, Atwood et al.*, 2009).

The protein sequences encoded by these genes show the three canonical conserved histidine boxes (HxxxH, HxxHH, and HxxHH) present in all FA_hydroxylase family members. The distribution of the accumulated changes is shown in Figure S1. None of the non-synonymous changes affect these highly conserved motifs, and at least a third of these (depending on the membrane topology prediction) are predicted to be exposed (not embedded in the membrane). This is important because, as reviewed in *Sperling, Ternes, Zank et al.*, the evolution of new regioselectivities in these enzymes would not involve the active site, but adjacent sequences. However, the failure to predict a reasonable topology (see Fig S1) points to the need to do an in-depth study of the membrane topology of the protein (similar to that performed by *Diaz, Mansilla, Vila et al.*, 2002).

Conclusion

Methods

T. cruzi stocks and strains.

Strains used in this study (and the corresponding current lineage classification) were: Sylvio X10 cl1 and Dm28 (TcI); MAS1 cl1, TU18 cl93, IVV cl4 (TcII); M6241 cl6, M5631 cl5 and X109/2 (TcIII); CanIII cl1, Dog Theis and 92122102R (TcIV); Sc43 cl9 and MN cl2 (TcV); Tulahuen cl2, CL Brener and P63 cl1 (TcVI).

Oligonucleotides and gene identifiers.

For each selected gene, a number of primers were designed for PCR-amplification. Taking into account that in the direct sequencing of PCR products the chromatogram quality is optimal in the range from 50 to 700 bp, a desirable length of the amplification products was set around 750bp. This length would also maximize our ability to sequence both strands of the amplification product, with good quality. Depending on the size of each selected gene, one or more overlapping amplification products were obtained. The list of the designed primers for each gene and the size of the corresponding amplification product is shown in Table S3. In only one case we had to design a separate primer (Tc-Mev-kinase26-fw) to amplify a fragment in one lineage (TcI) because there was a SNP at the 3' end of the primer that was absent in the release 1 of the TcSNP database.

Amplification and sequencing.

Selected fragments were amplified by PCR using Taq polymerase (Invitrogen) in a Biometra T Professional Gradient 96 cycler. Amplification mixtures contained 10 pmol of each primer, PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 50 ng of genomic DNA, 200 μM dNTPs, 2.5 U Taq polymerase (Invitrogen), and water to a final volume of 25 μl. After denaturing at 94°C for 2 minutes, thermal cycling was performed for 35 cycles at 94°C for 30 seconds, followed by 30

PeerJ PrePrints | https://peerj.com/preprints/44v1/ | v1 received: 18 Jul 2013, published: 18 Jul 2013, doi: 10.7287/peerj.preprints.44v1

seconds at a temperature set to 5°C less than the melting temperature of the selected primers, followed by 72°C for 30 seconds. Reactions were finished by a 5 minute incubation at 72°C. Amplification products were checked in 1.2% agarose gels stained with ethidium bromide to verify the presence of a single amplification product. Next, an aliquot (10 μl) of the amplification reaction was treated with 1 U of Exonuclease I (Fermentas) and 10 U of Shrimp Alkaline Phosphatase (Fermentas) for 45 minutes at 37°C and then for 30 minutes at 80°C to inactivate these enzymes. Subsequently two sequencing reactions were prepared, each with one of the primers used for the amplification of the product. Sequencing was carried out in an Applied Biosystems 3130 capillary sequencer using a Big-Dye terminator cycle sequencing kit, according to the instructions of the manufacturer.

SNP identification and scoring.

Gene fragments were PCR-amplified from every strain of the panel and sequenced in both strands. Base calling of chromatograms, assembly of sequences, detection of polymorphisms and manual inspection of assembled sequences and polymorphisms was done using a software package composed of Phred (version 0.020425.c) (*Ewing, Hillier, Wendl et al.*, 1998; *Ewing & Green*, 1998), Phrap (version 0.9909329), Polyphred (version 5.04) (*Nickerson, Tobe & Taylor*, 1997) and Consed (version 15.0) (*Gordon*, 2003). Basecalling of chromatograms was done by Phred. Sequences were then assembled by Phrap. Polyphred was used to process phrap assemblies to detect polymorphic sites. All candidate SNPs identified by PolyPhred (score >70/99), including heterozygous peaks, were visualized with Consed. A few false positives, and false negatives were removed/added after this manual inspection.

Other sequences used in this study

For the comparative analysis of SBP genes in kinetoplastid genomes, we have obtained the corresponding protein sequences for each of the yeast and/or *T. cruzi* orthologs used in this study from the TriTrypDB database(*Aslett, Aurrecoechea, Berriman et al.*, 2010). A BLAST search using the corresponding SBP gene as query was used to identify the corresponding ortholog. This has been cross-checked by inspection of ortholog clusters at the OrthoMCL database (*Chen, Mackey, Stoeckert et al.*, 2006). The sequences used belong to the following species/strains: *T. brucei brucei* strains TREU927 (*Berriman, Ghedin, Hertz Fowler et al.*, 2005) and Lister 427 (George Cross, unpublished), *Trypanosoma brucei gambiense* (*Jackson, Sanders, Berry et al.*, 2010); *L. major* (*Ivens, Peacock, Worthey et al.*, 2005), *L. infantum* (*Peacock, Seeger, Harris et al.*, 2007), *L. braziliensis* (*Peacock, Seeger, Harris et al.*, 2007), and *L. mexicana* (Wellcome Trust Sanger Institute Pathogen Sequencing Unit, unpublished).

Mapping substitutions on three dimensional structures

Crystallographic structures of enzymes of the sterol biosynthetic pathway were obtained from the Protein Data Bank (PDB, `http://www.rcsb.org`). Using the molecular graphics viewer VMD (version 2.8.6) (*Humphrey, Dalke & Schulten*, 1996) together with the multiple sequence alignment plugin for VMD (*Eargle, Wright & Luthey Schulten*, 2006) we mapped sequences from different strains on top of the reference sequence from which the structure was obtained. Atomic distances from each residue to the co-crystallized ligand were measured using standard tools implemented in VMD. Structures used in this work (and their source organism, and *T. cruzi*

homolog) were: 1wl5 (human, TcACAT); 2fa3 (*Brassica juncea*, TcHMGS); 3bgl (human, TcHMGR); 2hfu (*L. major*, TcMK); 2hke (*T. brucei*, TcMVD); 1yhm (*T. cruzi*, TcFPPS); 1ezf (human, TcSQS); 1w6k (human, TcOSC); 3khm, 3k10, 3ksw (*T. cruzi*, TcCYP51).

Data deposition

The sequences reported in this paper have been deposited in the GenBank database under the following accession numbers: JN050313-JN050853, and KF290395-KF290460. Heterozygous sequence polymorphisms have been submitted as ambiguities in the sequence using standard IUPAC notation. Sequence polymorphisms identified between different strains/clones will be available in a future release of the TcSNP database (http://snps.tcruzi.org, (*Ackermann, Carmona & Agüero*, 2009)).

Acknowledgments

This work was supported by grants from the National Agency for the Promotion of Science and Technology (ANPCyT, Argentina), the National Research Council of Argentina, and the Special Programme for Research and Training in Tropical Diseases (WHO / UNDP / UNICEF / World Bank). RC is now a fellow of the National Research Council of Argentina (CONICET) and was supported during this work by a fellowship from the University of San Martín. FA and PD are members of the Research Career of CONICET.

References

Ackermann AA, Carmona SJ, Agüero F. 2009. TcSNP: a database of genetic variation in Trypanosoma cruzi. *Nucleic Acids Res* **37**:D544–D549. doi:10.1093/nar/gkn874. URL http://dx.doi.org/10.1093/nar/gkn874.

Andrade LO, Galvão LMC, Meirelles MdNSL, Chiari E, Pena SDJ, Macedo AM. 2010. Differential tissue tropism of Trypanosoma cruzi strains: an in vitro study. *Mem Inst Oswaldo Cruz* **105**:834–837.

Andrade LO, Machado CR, Chiari E, Pena SD, Macedo AM. 1999. Differential tissue distribution of diverse clones of Trypanosoma cruzi in infected mice. *Mol Biochem Parasitol* **100**:163–172.

Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ Jr, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* **38**:D457–D462. doi:10.1093/nar/gkp851. URL http://dx.doi.org/10.1093/nar/gkp851.

Barnabé C, Brisse S, Tibayrenc M. 2000. Population structure and genetic typing of Trypanosoma cruzi, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. *Parasitology* **120 (Pt 5)**:513–526.

Berriman M, Ghedin E, Hertz Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Böhme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El Sayed NM. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science* **309**:416–422. doi:10.1126/science.1112642. URL http://dx.doi.org/10.1126/science.1112642.

Bouvier F, Rahier A, Camara B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog Lipid Res* **44**:357–429. doi:10.1016/j.plipres.2005.09.003. URL http://dx.doi.org/10.1016/j.plipres.2005.09.003.

Braga MV, Magaraci F, Lorente SO, Gilbert I, de Souza W. 2005. Effects of inhibitors of Delta24(25)-sterol methyl transferase on the ultrastructure of epimastigotes of Trypanosoma cruzi. *Microsc Microanal* **11**:506–515. doi:10.1017/S143192760505035X. URL http://dx.doi.org/10.1017/S143192760505035X.

Brisse S, Barnabé C, Tibayrenc M. 2000. Identification of six Trypanosoma cruzi phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol* **30**:35–44.

Buckner FS, Griffin JH, Wilson AJ, Van Voorhis WC. 2001. Potent anti-Trypanosoma cruzi activities of oxidosqualene cyclase inhibitors. *Antimicrob Agents Chemother* **45**:1210–1215. doi:10.1128/AAC.45.4.1210-1215.2001. URL http://dx.doi.org/10.1128/AAC.45.4.1210-1215.2001.

Buckner FS, Urbina JA. 2012. Recent Developments in Sterol 14-demethylase Inhibitors for Chagas Disease. *Int J Parasitol Drugs Drug Resist* **2**:236–242. doi:10.1016/j.ijpddr.2011.12.002. URL http://dx.doi.org/10.1016/j.ijpddr.2011.12.002.

Buscaglia CA, Di Noia JM. 2003. Trypanosoma cruzi clonal diversity and the epidemiology of Chagas' disease. *Microbes Infect* **5**:419–427.

Chang Q, Yan XX, Gu SY, Liu JF, Liang DC. 2008. Crystal structure of human phosphomevalonate kinase at 1.8 A resolution. *Proteins* **73**:254–258. doi:10.1002/prot.22151. URL http://dx.doi.org/10.1002/prot.22151.

Chen CK, Leung SSF, Guilbert C, Jacobson MP, McKerrow JH, Podust LM. 2010. Structural characterization of CYP51 from Trypanosoma cruzi and Trypanosoma brucei bound to the antifungal drugs posaconazole and fluconazole. *PLoS Negl Trop Dis* **4**:e651. doi:10.1371/journal.pntd.0000651. URL http://dx.doi.org/10.1371/journal.pntd.0000651.

Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* **34**:D363–D368. doi:10.1093/nar/gkj123. URL http://dx.doi.org/10.1093/nar/gkj123.

Coppens I, Courtoy PJ. 2000. The adaptative mechanisms of Trypanosoma brucei for sterol homeostasis in its different life-cycle environments. *Annu Rev Microbiol* **54**:129–156. doi:10.1146/annurev.micro.54.1.129. URL http://dx.doi.org/10.1146/annurev.micro.54.1.129.

Cosentino RO, Agüero F. 2012. A simple strain typing assay for Trypanosoma cruzi: discrimination of major evolutionary lineages from a single amplification product. *PLoS Negl Trop Dis* **6**:e1777. doi:10.1371/journal.pntd.0001777. URL http://dx.doi.org/10.1371/journal.pntd.0001777.

de Freitas JM, Augusto Pinto L, Pimenta JR, Bastos Rodrigues L, Gonçalves VF, Teixeira SMR, Chiari E, Junqueira ACV, Fernandes O, Macedo AM, Machado CR, Pena SDJ. 2006. Ancestral genomes, sex, and the population structure of Trypanosoma cruzi. *PLoS Pathog* **2**:e24. doi:10.1371/journal.ppat.0020024. URL http://dx.doi.org/10.1371/journal.ppat.0020024.

Desmond E, Gribaldo S. 2009. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol Evol* **1**:364–381. doi:10.1093/gbe/evp036. URL http://dx.doi.org/10.1093/gbe/evp036.

Diaz AR, Mansilla MC, Vila AJ, de Mendoza D. 2002. Membrane topology of the acyl-lipid desaturase from Bacillus subtilis. *J Biol Chem* **277**:48099–48106. doi:10.1074/jbc.M208960200. URL http://dx.doi.org/10.1074/jbc.M208960200.

Délye C, Laigret F, Corio Costet MF. 1997. A mutation in the 14 alpha-demethylase gene of Uncinula necator that correlates with resistance to a sterol biosynthesis inhibitor. *Appl Environ Microbiol* **63**:2966–2970.

Eargle J, Wright D, Luthey Schulten Z. 2006. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22**:504–506. doi:10.1093/bioinformatics/bti825. URL http://dx.doi.org/10.1093/bioinformatics/bti825.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**:186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**:175–185.

Flores López CA, Machado CA. 2011. Analyses of 32 loci clarify phylogenetic relationships among Trypanosoma cruzi lineages and support a single hybridization prior to human contact. *PLoS Negl Trop Dis* **5**:e1272. doi:10.1371/journal.pntd.0001272. URL http://dx.doi.org/10.1371/journal.pntd.0001272.

Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, Andersson B. 2011. Shotgun sequencing analysis of Trypanosoma cruzi I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Negl Trop Dis* **5**:e984. doi:10.1371/journal.pntd.0000984. URL http://dx.doi.org/10.1371/journal.pntd.0000984.

Gerpe A, Alvarez G, Benítez D, Boiani L, Quiroga M, Hernández P, Sortino M, Zacchino S, González M, Cerecetto H. 2009. 5-Nitrofuranes and 5-nitrothiophenes with anti-Trypanosoma cruzi activity and ability to accumulate squalene. *Bioorg Med Chem* **17**:7500–7509. doi:10.1016/j.bmc.2009.09.013. URL http://dx.doi.org/10.1016/j.bmc.2009.09.013.

Gerpe A, Odreman Nuñez I, Draper P, Boiani L, Urbina JA, González M, Cerecetto H. 2008. Heteroallyl-containing 5-nitrofuranes as new anti-Trypanosoma cruzi agents with a dual mechanism of action. *Bioorg Med Chem* **16**:569–577. doi:10.1016/j.bmc.2007.07.031. URL http://dx.doi.org/10.1016/j.bmc.2007.07.031.

Gordon D. 2003. Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics* **Chapter 11**:Unit11.2. doi:10.1002/0471250953.bi1102s02. URL http://dx.doi.org/10.1002/0471250953.bi1102s02.

Houten SM, Waterham HR. 2001. Nonorthologous gene displacement of phosphomevalonate kinase. *Mol Genet Metab* **72**:273–276. doi:10.1006/mgme.2000.3133. URL http://dx.doi.org/10.1006/mgme.2000.3133.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* **14**:33–8, 27–8.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, 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 JC, Müller Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schäfer M, Schein J, Schwartz DC, 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 DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ. 2005. The genome of the kinetoplastid parasite, Leishmania major. *Science* **309**:436–442. doi:10.1126/science.1112680. URL http://dx.doi.org/10.1126/science.1112680.

Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, Chukualim B, Capewell P, MacLeod A, Melville SE, Gibson W, Barry JD, Berriman M, Hertz Fowler C. 2010. The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis. *PLoS Negl Trop Dis* **4**:e658. doi:10.1371/journal.pntd.0000658. URL http://dx.doi.org/10.1371/journal.pntd.0000658.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**:D109–D114. doi:10.1093/nar/gkr988. URL http://dx.doi.org/10.1093/nar/gkr988.

Kelly SL, Lamb DC, Kelly DE. 1999. Y132H substitution in Candida albicans sterol 14alpha-demethylase confers fluconazole resistance by preventing binding to haem. *FEMS Microbiol Lett* **180**:171–175.

Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* **39**:309–338. doi:10.1146/annurev.genet.39.073003.114725. URL http://dx.doi.org/10.1146/annurev.genet.39.073003.114725.

Lepesheva GI, Park HW, Hargrove TY, Vanhollebeke B, Wawrzak Z, Harp JM, Sundaramoorthy M, Nes WD, Pays E, Chaudhuri M, Villalta F, Waterman MR. 2010. Crystal structures of Trypanosoma brucei sterol 14alpha-demethylase and implications for selective treatment of human infections. *J Biol Chem* **285**:1773–1780. doi:10.1074/jbc.M109.067470. URL http://dx.doi.org/10.1074/jbc.M109.067470.

Lepesheva GI, Villalta F, Waterman MR. 2011. Targeting Trypanosoma cruzi sterol 14α-demethylase (CYP51). *Adv Parasitol* **75**:65–87. doi:10.1016/B978-0-12-385863-4.00004-6. URL http://dx.doi.org/10.1016/B978-0-12-385863-4.00004-6.

Lepesheva GI, Waterman MR. 2011. Structural basis for conservation in the CYP51 family. *Biochim Biophys Acta* **1814**:88–93. doi:10.1016/j.bbapap.2010.06.006. URL http://dx.doi.org/10.1016/j.bbapap.2010.06.006.

Lepesheva GI, Zaitseva NG, Nes WD, Zhou W, Arase M, Liu J, Hill GC, Waterman MR. 2006. CYP51 from Trypanosoma cruzi: a phyla-specific residue in the B' helix defines substrate preferences of sterol 14alpha-demethylase. *J Biol Chem* **281**:3577–3585. doi:10.1074/jbc.M510317200. URL http://dx.doi.org/10.1074/jbc.M510317200.

Liendo A, Visbal G, Piras MM, Piras R, Urbina JA. 1999. Sterol composition and biosynthesis in Trypanosoma cruzi amastigotes. *Mol Biochem Parasitol* **104**:81–91.

Lorente SO, Rodrigues JCF, Jiménez Jiménez C, Joyce Menekse M, Rodrigues C, Croft SL, Yardley V, de Luca Fradley K, Ruiz Pérez LM, Urbina J, de Souza W, González Pacanowska D, Gilbert IH. 2004. Novel azasterols as potential agents for treatment of leishmaniasis and trypanosomiasis. *Antimicrob Agents Chemother* **48**:2937–2950. doi:10.1128/AAC.48.8.2937-2950.2004. URL http://dx.doi.org/10.1128/AAC.48.8.2937-2950.2004.

Lupetti A, Danesi R, Campa M, Del Tacca M, Kelly S. 2002. Molecular basis of resistance to azole antifungals. *Trends Mol Med* **8**:76–81.

Machado CA, Ayala FJ. 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi. *Proc Natl Acad Sci U S A* **98**:7396–7401. doi:10.1073/pnas.121187198. URL http://dx.doi.org/10.1073/pnas.121187198.

Magaraci F, Jimenez Jimenez C, Rodrigues C, Rodrigues JCF, Vianna Braga M, Yardley V, de Luca Fradley K, Croft SL, Souza Wd, Ruiz Perez LM, Urbina J, Gonzalez Pacanowska D, Gilbert IH. 2003. Azasterols as inhibitors of sterol 24-methyltransferase in Leishmania species and Trypanosoma cruzi. *J Med Chem* **46**:4714–4727. doi:10.1021/jm021114j. URL http://dx.doi.org/10.1021/jm021114j.

Marcili A, Lima L, Cavazzana M, Junqueira ACV, Veludo HH, Maia Da Silva F, Campaner M, Paiva F, Nunes VLB, Teixeira MMG. 2009. A new genotype of Trypanosoma cruzi associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology* **136**:641–655. doi:10.1017/S0031182009005861. URL http://dx.doi.org/10.1017/S0031182009005861.

Minning TA, Weatherly DB, Atwood J, Orlando R, Tarleton RL. 2009. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. *BMC Genomics* **10**:370. doi:10.1186/1471-2164-10-370. URL http://dx.doi.org/10.1186/1471-2164-10-370.

Morio F, Loge C, Besse B, Hennequin C, Le Pape P. 2010. Screening for amino acid substitutions in the Candida albicans Erg11 protein of azole-susceptible and azole-resistant clinical isolates: new substitutions and a review of the literature. *Diagn Microbiol Infect Dis* **66**:373–384. doi:10.1016/j.diagmicrobio.2009.11.006. URL http://dx.doi.org/10.1016/j.diagmicrobio.2009.11.006.

Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25**:2745–2751.

Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M, Kerhornou A, Ivens A, Fraser A, Rajandream MA, 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 SL, Faulconbridge A, Jeffares D, Depledge DP, Oyola SO, Hilley JD, Brito LO, Tosi LRO, Barrell B, Cruz AK, Mottram JC, Smith DF, Berriman M. 2007. Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nat Genet* **39**:839–847. doi:10.1038/ng2053. URL http://dx.doi.org/10.1038/ng2053.

Rassi AJ, Rassi A, Marin Neto JA. 2010. Chagas disease. *Lancet* **375**:1388–1402. doi:10.1016/S0140-6736(10)60061-X. URL http://dx.doi.org/10.1016/S0140-6736(10)60061-X.

Roberts CW, McLeod R, Rice DW, Ginger M, Chance ML, Goad LJ. 2003. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol Biochem Parasitol* **126**:129–142.

Rodrígues Poveda CA, González Pacanowska D, Szajnman SH, Rodríguez JB. 2012. 2-alkylaminoethyl-1,1-bisphosphonic acids are potent inhibitors of the enzymatic activity of Trypanosoma cruzi squalene synthase. *Antimicrob Agents Chemother* **56**:4483–4486. doi:10.1128/AAC.00796-12. URL http://dx.doi.org/10.1128/AAC.00796-12.

Rosso VS, Szajnman SH, Malayil L, Galizzi M, Moreno SNJ, Docampo R, Rodriguez JB. 2011. Synthesis and biological evaluation of new 2-alkylaminoethyl-1,1-bisphosphonic acids against Trypanosoma cruzi and Toxoplasma gondii targeting farnesyl diphosphate synthase. *Bioorg Med Chem* **19**:2211–2217. doi:10.1016/j.bmc.2011.02.037. URL http://dx.doi.org/10.1016/j.bmc.2011.02.037.

Sgraja T, Smith TK, Hunter WN. 2007. Structure, substrate recognition and reactivity of Leishmania major mevalonate kinase. *BMC Struct Biol* **7**:20. doi:10.1186/1472-6807-7-20. URL http://dx.doi.org/10.1186/1472-6807-7-20.

Silva DT, de Meirelles MdNSL, Almeida D, Urbina JA, Pereira MCS. 2006. Cytoskeleton reassembly in cardiomyocytes infected by Trypanosoma cruzi is triggered by treatment with ergosterol biosynthesis inhibitors. *Int J Antimicrob Agents* **27**:530–537. doi:10.1016/j.ijantimicag.2005.12.012. URL http://dx.doi.org/10.1016/j.ijantimicag.2005.12.012.

Sperling P, Ternes P, Zank TK, Heinz E. 2003. The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids* **68**:73–95.

Tibayrenc M, Ayala FJ. 2002. The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* **18**:405–410.

Tibayrenc M, Ayala FJ. 2013. How clonal are Trypanosoma and Leishmania? *Trends Parasitol* **29**:264–269. doi:10.1016/j.pt.2013.03.007. URL http://dx.doi.org/10.1016/j.pt.2013.03.007.

Urbina JA. 2009. Ergosterol biosynthesis and drug development for Chagas disease. *Mem Inst Oswaldo Cruz* **104 Suppl 1**:311–318.

Urbina JA, Docampo R. 2003. Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol* **19**:495–501.

Urbina JA, Payares G, Molina J, Sanoja C, Liendo A, Lazardi K, Piras MM, Piras R, Perez N, Wincker P, Ryley JF. 1996. Cure of short- and long-term experimental Chagas' disease using D0870. *Science* **273**:969–971.

Vacchina P, Tripodi KEJ, Escalante AM, Uttaro AD. 2012. Characterization of bifunctional sphingolipid Δ4-desaturases/C4-hydroxylases of trypanosomatids by liquid chromatography-electrospray tandem mass spectrometry. *Mol Biochem Parasitol* **184**:29–38. doi:10.1016/j.molbiopara.2012.04.005. URL http://dx.doi.org/10.1016/j.molbiopara.2012.04.005.

Vago AR, Andrade LO, Leite AA, d'Avila Reis D, Macedo AM, Adad SJ, Tostes S, Moreira MC, Filho GB, Pena SD. 2000. Genetic characterization of Trypanosoma cruzi directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. *Am J Pathol* **156**:1805–1809.

Westenberger SJ, Barnabé C, Campbell DA, Sturm NR. 2005. Two hybridization events define the population structure of Trypanosoma cruzi. *Genetics* **171**:527–543. doi:10.1534/genetics.104.038745. URL http://dx.doi.org/10.1534/genetics.104.038745.

Wolf YI, Koonin EV. 2012. A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biol Evol* **4**:1286–1294. doi:10.1093/gbe/evs100. URL http://dx.doi.org/10.1093/gbe/evs100.

Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, Schijman AG, Llewellyn MS, Lages Silva E, Machado CR, Andrade SG, Sturm NR. 2012. The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol* **12**:240–253. doi:10.1016/j.meegid.2011.12.009. URL http://dx.doi.org/10.1016/j.meegid.2011.12.009.

Supplementary Material

Table S1: Supplementary Excel File. The file contains a summary of BLASTP searches against fungal and kinetoplastid protein databases. BLAST searches using the yeast ERG3/ERG25 protein sequences as query, were run at the TriTrypDB BLAST server against a database of Kinetoplastid proteomes from reference and draft genomes. BLAST searches using a number of putative *T. cruzi* orthologs of these yeast genes were run at the SGD Fungal BLAST Server, against a database containing a selection of fungal genomes. Each BLASTP search is shown in a separate tab in the Excel workbook. File: Table-S1.xls

Table S2: List of nucleotide changes (SNPs, fixed differences) identified for each analyzed gene. The excel file contains one spreadsheet per gene with information on the location of each SNP relative to the start codon, the PolyPhred score for the SNP, and the character state of the SNP in each strain/lineage. File: Table-S2.xls

Table S3: List of oligonucleotide primers and amplification products analyzed in this study. File: Table-S3.xls

Figure S1: Supplementary PDF Figure. Distribution of observed SNPs in the TcSMO-like genes of T. cruzi. Based on the prediction of trans-membrane spanning domains (see TMHMM probability plot at the bottom), we created two alternative representations, following *Sperling, Ternes, Zank et al.*. The distribution of synonymous and non-synonymous SNPs is shown according to these models. The representations differ in the presence/absence of the second (non-predicted) trans-membrane domain. In these two representations the location of the 3rd histidine box always lies on the opposite side of the membrane. Both topologies may be wrong and an in-depth study ((similar to the one performed by *Diaz, Mansilla, Vila et al.*, 2002) may be required to establish the correct topology of these proteins. File: Figure-S1.pdf