

Syddansk Universitet

## Development of a *Trypanosoma cruzi* strain typing assay using MS2 peptide spectral libraries (Tc-STAMS2)

de Oliveira, Gilberto Santos; Kawahara, Rebeca; Fernandes, Livia Rosa; Mule, Simon Ngao; Avila, Carla Cristi; Teixeira, Marta M.G.; Larsen, Martin Røssel; Palmisano, Giuseppe

*Published in:*  
PLOS Neglected Tropical Diseases

*DOI:*  
[10.1371/journal.pntd.0006351](https://doi.org/10.1371/journal.pntd.0006351)

*Publication date:*  
2018

*Document license*  
CC BY

*Citation for published version (APA):*  
de Oliveira, G. S., Kawahara, R., Rosa-Fernandes, L., Mule, S. N., Avila, C. C., Teixeira, M. M. G., ... Palmisano, G. (2018). Development of a *Trypanosoma cruzi* strain typing assay using MS2 peptide spectral libraries (Tc-STAMS2). *PLOS Neglected Tropical Diseases*, 12(4), [e0006351]. DOI: 10.1371/journal.pntd.0006351

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

# Development of a *Trypanosoma cruzi* strain typing assay using MS2 peptide spectral libraries (Tc-STAMS2)

Gilberto Santos de Oliveira<sup>1</sup>, Rebeca Kawahara<sup>1</sup>, Livia Rosa-Fernandes<sup>2</sup>, Simon Ngao Mule<sup>1</sup>, Carla Cristi Avila<sup>1</sup>, Marta M. G. Teixeira<sup>1</sup>, Martin R. Larsen<sup>2</sup>, Giuseppe Palmisano<sup>1\*</sup>

**1** Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil, **2** Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

\* [palmisano.gp@usp.br](mailto:palmisano.gp@usp.br), [palmisano.gp@gmail.com](mailto:palmisano.gp@gmail.com)



## Abstract

### Background

Chagas disease also known as American trypanosomiasis is caused by the protozoan *Trypanosoma cruzi*. Over the last 30 years, Chagas disease has expanded from a neglected parasitic infection of the rural population to an urbanized chronic disease, becoming a potentially emergent global health problem. *T. cruzi* strains were assigned to seven genetic groups (TcI-TcVI and TcBat), named discrete typing units (DTUs), which represent a set of isolates that differ in virulence, pathogenicity and immunological features. Indeed, diverse clinical manifestations (from asymptomatic to highly severe disease) have been attempted to be related to *T. cruzi* genetic variability. Due to that, several DTU typing methods have been introduced. Each method has its own advantages and drawbacks such as high complexity and analysis time and all of them are based on genetic signatures. Recently, a novel method discriminated bacterial strains using a peptide identification-free, genome sequence-independent shotgun proteomics workflow. Here, we aimed to develop a *Trypanosoma cruzi* Strain Typing Assay using MS/MS peptide spectral libraries, named Tc-STAMS2.

### Methods/Principal findings

The Tc-STAMS2 method uses shotgun proteomics combined with spectral library search to assign and discriminate *T. cruzi* strains independently on the genome knowledge. The method is based on the construction of a library of MS/MS peptide spectra built using genotyped *T. cruzi* reference strains. For identification, the MS/MS peptide spectra of unknown *T. cruzi* cells are identified using the spectral matching algorithm SpectraST. The Tc-STAMS2 method allowed correct identification of all DTUs with high confidence. The method was robust towards different sample preparations, length of chromatographic gradients and fragmentation techniques. Moreover, a pilot inter-laboratory study showed the applicability to different MS platforms.

### OPEN ACCESS

**Citation:** de Oliveira GS, Kawahara R, Rosa-Fernandes L, Mule SN, Avila CC, Teixeira MMG, et al. (2018) Development of a *Trypanosoma cruzi* strain typing assay using MS2 peptide spectral libraries (Tc-STAMS2). PLoS Negl Trop Dis 12(4): e0006351. <https://doi.org/10.1371/journal.pntd.0006351>

**Editor:** Herbert B. Tanowitz, Albert Einstein College of Medicine, UNITED STATES

**Received:** October 14, 2017

**Accepted:** February 27, 2018

**Published:** April 2, 2018

**Copyright:** © 2018 de Oliveira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All raw data have been submitted to PRIDE archive (<https://www.ebi.ac.uk/pride/archive/>), project accession: PXD008088.

**Funding:** The Lundbeck foundation is acknowledged for the grant sponsoring the Orbitrap Velos Pro. CNPq (GP 441878/2014-8) and FAPESP (GP: 2014/06863-3), RK is supported by FAPESP (2015/02866-0). The funders had no role in study design, data collection and analysis,

decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Conclusions and significance

This is the first study that develops a MS-based platform for *T. cruzi* strain typing. Indeed, the Tc-STAMS2 method allows *T. cruzi* strain typing using MS/MS spectra as discriminatory features and allows the differentiation of TcI-TcVI DTUs. Similar to genomic-based strategies, the Tc-STAMS2 method allows identification of strains within DTUs. Its robustness towards different experimental and biological variables makes it a valuable complementary strategy to the current *T. cruzi* genotyping assays. Moreover, this method can be used to identify DTU-specific features correlated with the strain phenotype.

## Author summary

Chagas disease is one of the most important neglected diseases with an estimated number of 6–7 million infected individuals, the majority living in Central and South America. The *Trypanosoma cruzi* (*T. cruzi*) protozoan parasite is the etiological agent of Chagas disease. *T. cruzi* is highly genetically diverse and a new nomenclature assigned each strain to seven genetic groups (TcI-TcVI and Tcbat), named Discrete Typing Units (DTUs), based on their biochemical, immunological and phenotypical characteristics. *T. cruzi* DTUs have been correlated to diverse clinical outcomes highlighting the importance of molecular epidemiological screens. Despite the development of *T. cruzi* typing methods based on genetic signatures, each method presents its own advantages and challenges. The work presented here shows the application of mass spectrometry for *Trypanosoma cruzi* Strain Typing Assay using MS<sup>2</sup> peptide spectral libraries (Tc-STAMS2). The novelty of the method is based on the use of peptide fragmentation spectra as strain-specific fingerprints to classify and identify DTUs. Initially, a spectra library is generated from characterized *T. cruzi* strains. The library is subsequently inspected using MS/MS spectra from unknown strains and confidently assigned to a specific strain in an automated and computationally-driven approach. The Tc-STAMS2 method was challenged to test several variables such as sample type and preparation, instrument setup and identification platform. Tc-STAMS2 provided high confidence and robustness in *T. cruzi* strain typing. The Tc-STAMS2 method represents a proof-of-concept of a complementary strategy to the current DNA-based *T. cruzi* genotyping methods. Moreover, the method allows the identification of strain-specific features that could be related to the biology of *T. cruzi* strains and their clinical outcomes.

## Introduction

Chagas disease also known as American trypanosomiasis affects around 6–7 million people especially in Latin America [1]. The etiological agent of Chagas diseases is the protozoan *Trypanosoma cruzi* [2] that infects several mammalian hosts and is primarily transmitted through the contamination with feces of triatomine bugs. Besides, congenital, blood transfusions, transplants and ingestion of contaminated foods represent other ways of infection [3]. Chagas disease is characterized by an acute and chronic phase. The acute phase lasts a few weeks and may present mild symptoms such as fever and swelling around the site of infection [4]. The chronic phase is in general lifelong and asymptomatic. However, 20–30% of patients develop cardiac or gastrointestinal complications [3].

*T. cruzi* is highly genetically diverse. In order to standardize the nomenclature facilitating the communication among scientists, *T. cruzi* strains were divided into six (Tc I–VI) discrete

*typing units* (DTUs) plus Tcbat, a novel strain associated with bats. Each group represents a set of isolates that are genetically similar and can be identified by common immunological, biochemical, pathological and molecular markers [5]. *T. cruzi* strains characterization is extremely important to understand different epidemiological and pathological characteristics such as geographical distribution and clinical outcomes [6]. Several techniques have been introduced to improve the genotyping. In particular, the genetic diversity of *T. cruzi* was first recognized by multilocus enzyme electrophoresis (MLEE) and restriction analysis of kinetoplastid DNA minicircles [7, 8]. Currently, the methods used for typing of *T. cruzi* strains are based on polymorphism of the mini-exon gene (Spliced Leader) and the 24S $\alpha$  and 18S ribosomal RNA [9]. Other assays involve the analysis of complex electrophoretic patterns generated by restriction polymorphisms of PCR amplified genomic DNA [10–12]. These methods are able to discriminate *T. cruzi* strains, but they are work and time consuming and the interpretations of the results may be misleading. Multiplex Real-Time PCR Assay Using TaqMan Probes proved useful to determine DTUs of cultured *T. cruzi*, vector and blood samples from patients in acute infection [10, 12, 13].

Proteomics methods using mass spectrometry have emerged as a powerful strategy to discriminate bacteria and have been established as a valuable alternative to DNA-based bacterial identification. Indeed, MALDI-TOF MS generates an intact protein profile that is compared to a MS database of known species. This method is easy to perform, rapid and provides accurate results and has been introduced as routine in most hospitals [14]. However, this method is currently not able to distinguish at the strain level. Shotgun sequencing of peptides, derived from enzymatically digested proteins, using LC-MS/MS in combination with database search of *in silico* digested proteins derived from publicly available protein sequence has been proposed as a method to discriminate bacterial strains such as *Helicobacter pylori* and *Yersinia persis* [15]. This approach has great discrimination power compared to MALDI-TOF MS since it allows the identification of thousands of peptides with higher dynamic range. However, this method suffers from lack of suitable databases since many species have not been sequenced. Recently, a novel method based on MS/MS spectral matching has been shown to identify the blood meal of ticks (*Ixodes scapularis*) [16] and has also been applied to discriminate *E. coli* strains from different isolates [15].

Few studies have explored the possibility of using mass spectrometry as a diagnostic tool for Chagas disease. Using mass spectrometry, Ndao M. et al. [17] were able to identify full-length Apo1 as a serum biomarker of chronic Chagas disease patients using a SELDI-TOF approach. Despite several proteomic studies to understand the molecular features of *T. cruzi* in different biological conditions [18–22], to date, there is no report on the use of mass spectrometry for assaying various *T. cruzi* strains.

A pioneer work developed a peptide identification-free shotgun proteomics workflow to trace the vertebrate host that a tick (*Ixodes scapularis*) was feeding on [16]. This workflow used MS/MS spectral matching. The same strategy, named UNID, was used to profile bacterial strains [15]. In this study, we developed a platform to discriminate *T. cruzi* strains using MS/MS peptide spectral matching (Tc-STAMS2). The method was able to discriminate *T. cruzi* strains from different origins with high sensitivity and accuracy. The method was robust towards sample preparation and instrumental parameters, such as peptide purification, chromatographic gradient time, peptide fragmentation techniques and MS instruments. The MS/MS spectra were subjected to database search. More than 4000 proteins were identified in the combined six DTUs strains analyzed.

A total of 1096 proteins were differentially expressed between the six DTUs and multivariate analysis allowed the discrimination of *T. cruzi* strains using the quantitative MS signal. In conclusion, this study describes a mass spectrometry-based method to discriminate *T. cruzi*

strains. The Tc-STAMS2 method represents a proof-of-concept of an alternative strategy to DNA-based *T. cruzi* genotyping. Further studies are needed to show its applicability to biofluids in clinical isolates.

## Methods

### *T. cruzi* cultures

Epimastigote forms of *T. cruzi* cultivated in LIT (Liver Infusion Tryptose) supplemented with 10% fetal bovine serum [23] at 28 °C of exponential culture phase were employed in the present study. DTU classification of all *T. cruzi* strains were confirmed by sequencing [24]. Only validated DTUs strains were used in this study (Table 1).

### Cell growth challenging test

*T. cruzi* cells from the Sylvio X10 cl1 (DTU-I) strain were collected in the exponential and stationary phase and processed as described below. Two biological replicates for the stationary (St\_1 and St\_2) and exponential (Exp\_1 and Exp\_2) growth phase were analyzed.

### Sample preparation and nLC-MS/MS analysis

**Protein extraction and digestion.** Epimastigote cells ( $5 \times 10^8$ ) were washed three times in phosphate buffered saline (PBS), pH 7.2 (8,000g for 10 minutes at room temperature), re-suspended in 400 µL of lysis buffer (7M urea, 2M thiourea, 1 mM DTT and protease inhibitors (Amersham) and incubated under stirring for 30 minutes to solubilize the proteins. Proteins were reduced with 10 mM DTT (DL-Dithiothreitol–Sigma-Aldrich), alkylated with 40 mM iodoacetamide (Sigma-Aldrich), digested with trypsin (Promega) in the ratio 1:50 (µg trypsin/µg protein) in 50 mM ammonium bicarbonate solution at 37 °C overnight. The reaction was stopped with 1% formic acid (less than pH 3) and then the sample was desalted with C18 columns (StageTips). Four biological replicates were prepared for each DTU. Blind test samples A (DTU-III) and B (DTU-I) were prepared using a minimum of three biological replicates according to the protocol described above.

**Table 1. Cultured stocks representing the known *T. cruzi* lineages (DTUs) used to build the MS/MS library and to validate the method.**

Trypanosomatid	DTUs	Strains	Form	Culture médium
<i>T. cruzi</i>	DTU-I	Sylvio X10 cl1	Epimastigote	LIT—Liver Infusion-Tryptose supplemented with 10% fetal bovine serum at pH 7,2; 28 °C [23].
	DTU-I	Sylvio X10/4*		
	DTU-I	G*		
	DTU-II	Y		
	DTU-II	Esmeraldo*		
	DTU-III	M6241 cl6		
	DTU-III	3869*		
	DTU-IV	CanIII cl1		
	DTU-IV	José Júlio clone*		
	DTU-V	MN cl2		
	DTU-V	NR cl3*		
	DTU-VI	CL Brener		
	DTU-VI	CL14*		

\*Strains added to validate the Tc-STAMS2 methodology for intra-DTU strain discrimination.

<https://doi.org/10.1371/journal.pntd.0006351.t001>

## Sample preparation for the challenging tests

Peptide desalting was performed in acid (acid 1 and acid 2) and basic (basic 1 and basic 2) conditions. In particular, for the acid purification tryptic peptides were acidified with 0.1% TFA (pH 3) and loaded onto an acid-activated StageTip microcolumn before being eluted with 50% acetonitrile: 0.1% TFA. For the basic purification, tryptic peptides were dissolved in 0.1% ammonia water (pH 10) and loaded onto a base-activated StageTip before being eluted with 50% acetonitrile: 0.1% ammonia water. The eluted peptides were lyophilized and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS).

## Nano LC-MS/MS analysis

Peptides were separated by Reprosil-Pur C18-AQ column (3 $\mu$ m; Dr. Maisch GmbH, Germany) using Easy nano-LC HPLC (Proxeon, Odense, Denmark). The HPLC gradient was 0–34% B solvent (A = 0.1% formic acid; B = 90% ACN, 0.1% formic acid) in 70 min at a flow of 250 nL/min. The MS analysis was performed using the LTQ-Orbitrap Velos (Thermo Scientific, Bremen, Germany). The mass range was 400–1500 m/z at a resolution of 30,000 at 400 m/z for a target value of 1e<sup>6</sup> ions. For each MS scan, collision induced dissociation (CID) fragmentation was performed on the 20 most intense ions in the linear iontrap. The parameters for data acquisition were: activation time = 15 ms, normalized energy = 35, Q-activation = 0.25, exclusion = available with repeat count 1, exclusion duration = 30s and intensity threshold = 30,000, target ions = 2e<sup>4</sup> [25]. All raw data have been submitted to PRIDE archive (<https://www.ebi.ac.uk/pride/archive/>), project accession: PXD008088.

## Sample amounts, chromatographic gradients and MS/MS fragmentation types used for the challenging test

The robustness of the Tc-STAMS2 method was tested using different parameters: 1) sample amounts, 2) chromatographic gradients and 3) MS fragmentation techniques. For the different sample amounts, 0.5 $\mu$ g and 1 $\mu$ g, were loaded onto the analytical column before MS analysis and named low and high, respectively. The chromatographic elution time was set to 20, 70 and 130 min from 0–34% B solvent at 250nL/min. CID fragmentation was used to develop the Tc-STAMS2 method. Higher energy collision induced dissociation (HCD) was evaluated as alternative peptide fragmentation type on tryptic peptides separated on a 70min chromatographic gradient. For the HCD fragmentation, each MS scan was acquired at resolution of 30,000 FWHM followed by 7 MS/MS scan of the most intense ions with an activation time of 0.1 ms and normalized collision energy of 35. The spectral library for each DTU was developed on a LTQ-Orbitrap Velos Pro instrument (Thermo Fisher Scientific) located in the PR group, Department of Biochemistry and Molecular biology, University of Southern Denmark. All the other tests for testing and validating the method were performed on a LTQ-Orbitrap Velos Normal at the Biomass mass spectrometry facility (São Paulo, Brazil).

## Bioinformatics and statistical analyses

*MS/MS spectral library generation and spectral matching.* MS/MS spectral library generation and spectral matching were performed using the SpectraST software (version 4.8) as previously described [15, 16, 26]. In particular, the LC-MS/MS acquired spectra were converted to an open format (mzXML) by MSconvert [27], forming part of the software suite offered by TPP (Trans-Proteomic Pipeline) [28]. SpectraST (version 4.8) was used to build the reference spectral library and perform MS/MS spectral matching [26]. The reference spectral library was generated with three raw files for each one of the six DTUs and one raw file for each DTU was

compared against constructed reference library. The first step for the reference spectral library generation involves the application of a threshold at which MS/MS spectra originated from the same peptide precursor ion are combined to create the consensus spectrum. Moreover, low quality spectra are excluded from the library [26]. To determine the spectrum of similarity between the query spectrum and the reference library, SpectraST uses the Spectral Dataset Similarity (SDSS) function [16, 29]. In particular, the unique dot product SDSS, abbreviated as “score” along the text, was chosen as *T. cruzi* DTU strain discrimination function and reported [16]. The statistical confidence in the identification of the correct DTU is made by data bootstrap [16]. All the spectral matching experiments reported below had a bootstrap of 1 unless reported.

DiagnoProt software was used also to match the MS/MS spectra of unknown samples against a database of genotyped *T. cruzi* strains [21]. Default parameters for creating the spectral database were used: similarity threshold 0.70, precursor tolerance 4.50; activation type CID; minimum number peaks: 50; minimum relativity intensity: 0.01; minimum retention time 10.00; bin offset 0.40; bin size: 1.0005; minimum bin m/z 200.00; maximum bin m/z 1700.00. The spectral database was used to match the identity of each unknown sample.

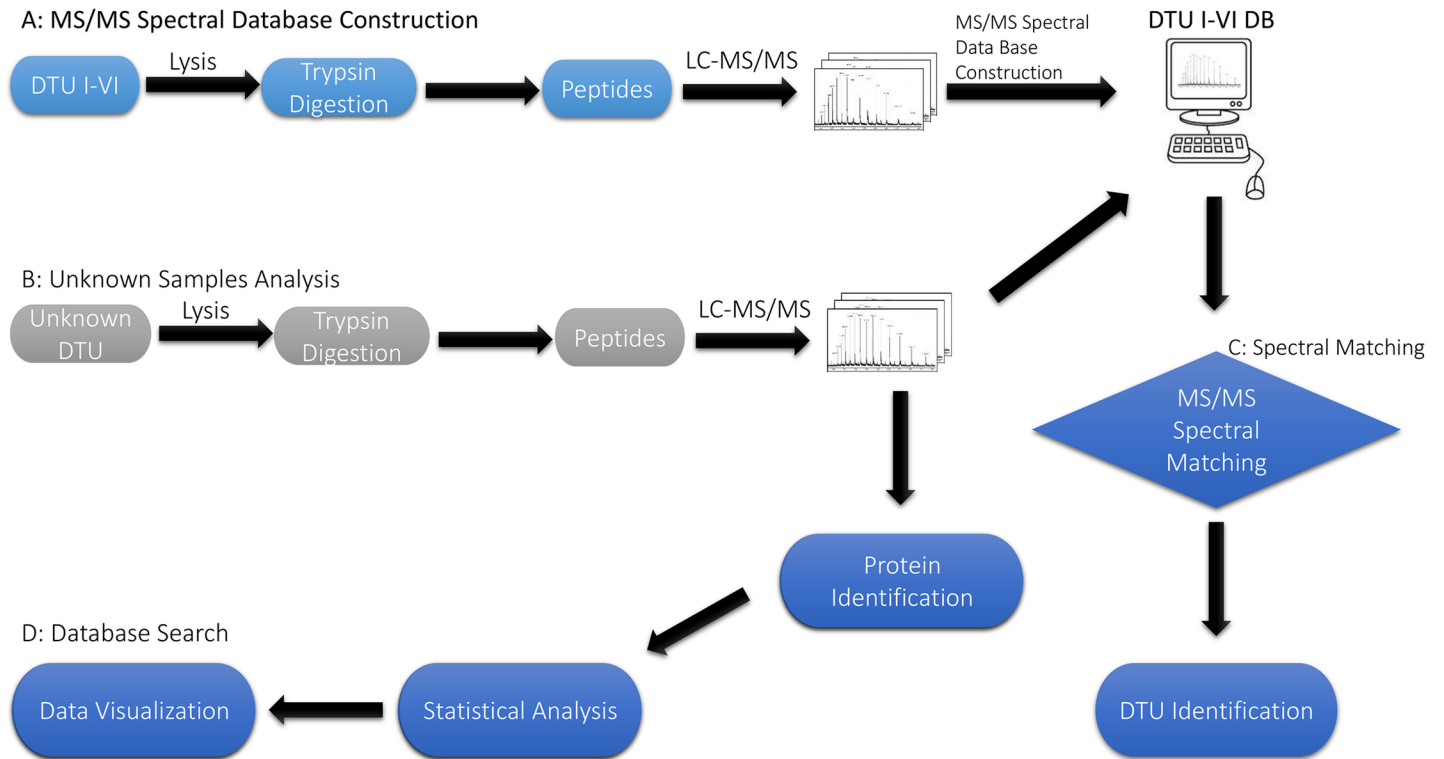
### Database searches

The raw LC-MS/MS files were analyzed using: Proteome Discoverer, MaxQuant and the Trans-Proteomic Pipeline. Proteome Discoverer v2.1 (Thermo Scientific) was used with the *T. cruzi* database using Mascot and Sequest. The searches in the database were conducted with the following parameters: precursor mass tolerance of 20 ppm; MS/MS mass tolerance 0.5 Da (CID data). Trypsin was selected as enzyme and carbamidomethyl cysteine as fixed modification. The variables modifications were oxidation of methionine and deamidation (NQ). Shared peptide sequences were grouped as grouped accessions proteins. The False Discovery Rates (FDR) was calculated using the algorithm Percolator with  $q$  equal or less than 0.01. Protein FDR was calculated in the Proteome Discoverer software and kept below 1%.

The raw files were also processed using the MaxQuant [30] version 1.2.7.429 and the MS/MS spectra were searched using the Andromeda search engine [31] against the Uniprot *T. cruzi* Protein Database (release July 11, 2017; 51,738 entries). The initial maximal allowed mass tolerance was set to 20 ppm for precursor and then set to 4.5 ppm in the main search and to 0.5 Da for fragment ions. Enzyme specificity was set to trypsin with a maximum of two missed cleavages. Carbamidomethylation of cysteine (57.02 Da) was set as a fixed modification, and oxidation of methionine (15.99 Da), deamidation (NQ) and protein N-terminal acetylation (42.01 Da) were selected as variable modifications. Bioinformatics analysis was performed using the software Perseus v.1.5.2.6 [30] available in the MaxQuant environment and reverse and contaminant entries were excluded from further analysis. Protein FDR was calculated in the MaxQuant software and kept below 1%. Label Free Quantification (LFQ) intensity values were considered to relatively compare the abundance of proteins present in the different DTUs.

Trans-Proteomic Pipeline software suite was also used to search raw files converted to mzXML [13]. The mzXML files were searched by the Comet search algorithm embedded into the TPP platform [14]. Peptide and protein FDR was estimated using the PeptideProphet and ProteinProphet algorithm embedded in the Trans-Proteomic Pipeline [32]. Identifications with less than 1% FDR were kept.

Raw data from human placental tissue [33] and *T. vivax* (Meta, BSF1 and EP1) [22] were obtained from the public MS spectra databank PRoteomics IDentifications (PRIDE) and used for the negative control test.



**Fig 1. Tc-STAMS2 workflow.** (A) Proteins were extracted from epimastigote *T. cruzi* cells, digested using trypsin and subjected to nLC-MS/MS analysis. MS/MS spectra were clustered and merged to generate a reference mass spectral library using SpectraST (<http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST>) or DiagonProt (<http://patternlabforproteomics.org/diagnoprot>) software. (B) The same steps described in panel A were used for the unknown samples. In particular, the unknown *T. cruzi* strains were lysed, proteins were extracted and digested using trypsin. Tryptic peptides were analyzed by nLC-MS/MS. (C) The MS/MS spectra from unknown *T. cruzi* strain samples were compared with the reference MS/MS spectral library in order to assign an unknown sample to a particular *T. cruzi* strain. The A, B and C steps have been used through the text in all spectral matching analyses. (D) MS/MS data were searched against the Uniprot *T. cruzi* protein database to identify and quantify proteins in each strain. Three computational platforms were used: MaxQuant (30, 31), Proteome Discoverer v2.1 (Thermo Fisher) and Trans-Proteomic Pipeline (13, 14).

<https://doi.org/10.1371/journal.pntd.0006351.g001>

## Statistical analyses

Label-free quantified peptides/proteins were analysed by the Perseus software [18]. Significantly regulated features with a p value less than 0.05 corrected with the Bonferroni post-hoc test were used to cluster the different DTUs. Hierarchical clustering of significantly regulated proteins/peptides was performed using the Z-score calculation on the log<sub>2</sub> intensity values and it was represented as a heat map. The Principal Component Analysis (PCA) was performed using the same procedure described above in Perseus software. In addition, for generation of Venn diagram we used Venn Diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/> Bioinformatics & Evolutionary Genomics).

## Results

### Tc-STAMS2 strategy allows DTUs discrimination

In this study, the combination of mass spectrometry and computational approaches was used to develop a method for *T. cruzi* DTU discrimination, named Tc-STAMS2. A schematic overview of the *T. cruzi* DTUs identification using MS/MS spectra from tryptic peptides is summarized in Fig 1. A reference spectral library was built using a total of 586513 unique MS/MS spectra of tryptic peptides derived from three raw MS files of each one of the six *T. cruzi* strains



**Table 2. Genotype discrimination based on spectral similarity searches.** Six *T. cruzi* strains (Sylvio X10 cl1, Y, M6241 cl6, CanIII cl1, MN cl2 and CL Brener) belonging to six DTUs were selected to test the Tc-STAMS2 method. The unique dot product SDSS score is reported along with the number of MS/MS spectra matches (unique/total). Genotypes identified with the highest score are highlighted in gray.

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
DTU-I	0.861 (2858/20232)	0.005 (11/13450)	0.008 (24/13419)	0.008 (26/14135)	0.016 (65/13470)	0.013 (33/14469)
DTU-II	0.010 (23/14212)	0.754 (1169/20004)	0.032 (72/14820)	0.023 (62/15126)	0.007 (22/14198)	0.020 (35/16038)
DTU-III	0.006 (18/13177)	0.020 (45/13838)	0.817 (2086/19587)	0.013 (42/13135)	0.017 (42/13689)	0.015 (35/14200)
DTU-IV	0.011 (38/14307)	0.013 (28/14213)	0.010 (32/13412)	0.849 (2648/20199)	0.004 (16/12889)	0.019 (48/14722)
DTU-V	0.008 (35/12560)	0.009 (20/12405)	0.013 (37/12771)	0.004 (16/11687)	0.863 (3086/19777)	0.011 (28/14215)
DTU-VI	0.016 (51/14230)	0.017 (27/14880)	0.013 (32/14101)	0.017 (51/14047)	0.022 (57/14849)	0.800 (1463/20238)

<https://doi.org/10.1371/journal.pntd.0006351.t002>

(Fig 1A). Each *T. cruzi* strain was processed and acquired in four biological replicates. MS/MS spectra acquired from three replicates of each DTU were used to build the reference spectral library using SpectraST software [26]. Following the construction of the reference MS/MS library, proteins from unknown *T. cruzi* strain samples were extracted and digested with trypsin before being analyzed by nLC-MS/MS (Fig 1B). A LC-MS profile of four replicates of DTU-I is reported in S1 Fig and the Pearson correlation score indicates high similarity between the different runs. The LC-MS chromatographic profile of the tryptic peptides belonging to the six *T. cruzi* strains shows high similarity between the different DTUs, S2 Fig. MS/MS spectra from different DTUs were subjected to spectral matching comparison with the library using SpectraST software [19]. The identification was made by finding the reference library with the highest similarity to the sample spectral dataset, in this case, the different DTUs, to be tested. Unique dot product SDSS score was used to provide a quantitative similarity measure between two spectral datasets (Fig 1C) [16]. MS/MS spectra were searched against the *T. cruzi* proteome database as described below (Fig 1D).

Based on the similarity scores between the MS/MS spectra from an unknown sample with the mass spectral library, the Tc-STAMS2 was able to differentiate and accurately identify the different DTUs. We observed that the score values were between 0.75 to 0.86 for true matches and close to 0 in unmatched cases, as shown in Table 2.

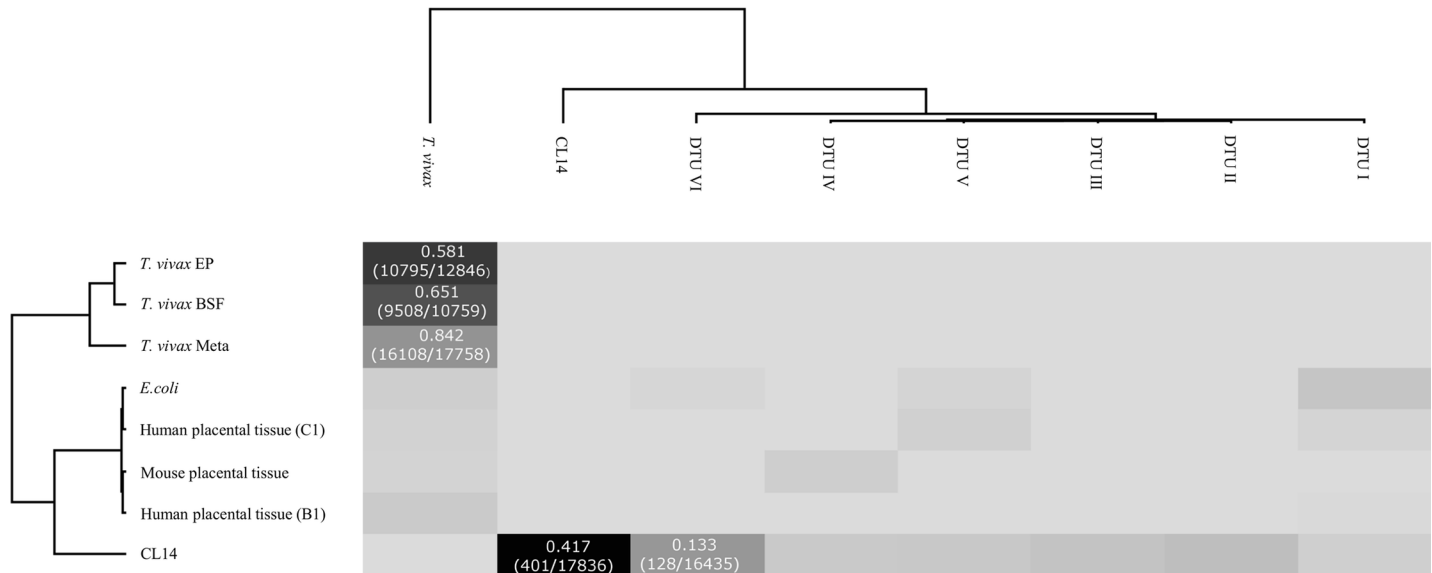
The LC-MS chromatograms obtained for each replicate and for each DTU (S1 Fig and S2 Fig) showed high similarity; however, the developed method was capable of differentiating and identifying each of them. In order to rule out the possibility that different growth phases could influence the assignment of the algorithm, Sylvio X10/1 (DTU-I) epimastigotes in the exponential and stationary phase were collected. Independently of the growth phase, the algorithm was able to assign it to the correct DTU, Table 3. This demonstrates that the identification method is not affected by the growth phase of the parasite.

Moreover, the performance of the Tc-STAMS2 spectral matching approach were tested to correctly identify MS/MS data sets from: 1) a *T. cruzi* strain that is known to belong to DTU-VI (CL14) [5] but was not included in the spectral library, 2) from a species phylogenetically related, such as *Trypanosoma vivax* and 3) from species with completely distant genome

**Table 3. MS/MS spectral matching comparison using different growth phases of Sylvio X10 cl1 (DTU-I) using SpectraST software.** St\_1—Stationary phase; St\_2—Biological replicate stationary phase; Ex—Exponential phase; Ex\_2—Biological replicate exponential phase.

Samples	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI	CL14 (DTU-VI)
A (St_1)	0.283	0.032	0.068	0.017	0.075	0.032	0.034
A (St_2)	0.288	0.035	0.067	0.015	0.068	0.029	0.030
B (Ex_1)	0.296	0.036	0.057	0.026	0.104	0.058	0.046
B (Ex_2)	0.301	0.041	0.056	0.027	0.099	0.053	0.051

<https://doi.org/10.1371/journal.pntd.0006351.t003>



**Fig 2. Tc-STAMS2 approach tested against: 1) the CL14 *T. cruzi* strain, 2) *T. vivax* dataset and 3) LC-MS/MS datasets from *E. coli* and human and mouse placental tissues.** The sensitivity of Tc-STAMS2 approach was tested for the detection of intra-DTU strains such as CL14 and CLBrenner strains belonging to DTU-VI. Moreover, the specificity of Tc-STAMS2 approach was tested for the assignment of MS/MS spectra derived from phylogenetically distant organisms such as mouse and human. In particular, the MS/MS spectral library was built using seven strains belonging to six DTUs such as Sylvio X10 c1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL Brener and CL14 (DTU-VI) strains and MS/MS data from *T. vivax* (epimastigote, metacyclic and bloodstream forms) were added to the spectral library [22]. Independent LC-MS/MS runs of the different *T. cruzi* strains, *T. vivax* life stages, human and mouse placenta tissue and *E. coli* were compared against the MS/MS spectral library using SpectraST software. Intra-DTU discrimination was achieved for CL14 and CL Brener and no assignment was made for the *E. coli*, mouse and human samples. MS/MS spectra from *T. vivax* were assigned specifically to *T. vivax* without identification of *T. cruzi*.

<https://doi.org/10.1371/journal.pntd.0006351.g002>

(ex: human, *E. coli*, mouse). For this analysis, a new MS/MS library was constructed, using the MS/MS spectra from the six DTUs, including *T. cruzi* CL14 and *T. vivax* in metacyclic stage (meta1 e meta2). Firstly, MS/MS spectra from *T. cruzi* strains CL14 were compared with the library and the similarity score matched to the CL14 (score = 0.417). Interestingly, although the similarity scores of CL14 with DTUs I to V were close to zero, the similarity between CL14 and DTU-VI was comparatively high (score = 0.133), indicating that many spectra MS/MS of CL14 are shared with DTU-VI (CL Brener), as shown in Fig 2. It should be noted that the LC-MS chromatographic profiles of the CL14 and CL Brener strains have very high similarity, S3 Fig. However, the Tc-STAMS2 was able to clearly differentiate between the two strains within the same DTU.

Subsequently, MS/MS spectra from three different life stages of *T. vivax*, Meta3 –metacyclic phase, BSF1 –bloodstream phase and EP1 –epimastigote were compared using the mass spectral library. Based on the similarity scores, the Tc-STAMS2 was able to correctly identify these samples to *T. vivax*, Fig 2.

In order to test the method with negative control, MS/MS spectra from unrelated *T. cruzi* species such as human, mouse and *E. coli* were compared to the library. For these samples, the similarity scores were close to zero, indicating that the similarity scores found between two MS/MS datasets is library-specific and not random, Fig 2.

Moreover, we also evaluated the ability of this strategy to provide correct identification from an independent sample (blind test), which was collected and processed at different days or under different conditions. In particular the different datasets were obtained in: 1) inter-laboratory studies, 2) different sample preparation strategies, 3) different LC gradients and 4) different MS fragmentation methods. To assess the robustness of the Tc-STAMS2 platform,

Samples	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI	Scores
A1			X				0.139
A2			X				0.137
B	X						0.246
B/acid 1	X						0.32
B/acid 2	X						0.309
B/basic 1	X						0.294
B/basic 2	X						0.312
B (20min CID) High	X						0.203
B (70min CID) High	X						0.267
B (130min CID) High	X						0.297
B (70min HCD) Low	X						0.113
B (20min CID) Low	X						0.21
B (70min CID) Low	X						0.279
B (130min CID) Low	X						0.292
B (70min HCD) Low	X						0.104

**Fig 3. Tc-STAMS2 was tested for its robustness towards technical and experimental variations.** Initially, the Tc-STAMS2 approach was tested for inter-laboratory comparison. Two unknown *T. cruzi* strains (A and B) were processed as described in the step B of Fig 1 and acquired using the EasynLC coupled to LTQ-Orbitrap Velos mass spectrometer located in the CEFAP mass spectrometry facility at the University of Sao Paulo, Sao Paulo, Brazil. The MS/MS spectral library was built using Sylvio X10 cl1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL Brener (DTU-VI) and acquired in the PR group, Odense, Denmark using a similar LC-MS/MS setup (EasynLC coupled to LTQ-Orbitrap Velos). A1 and A2 indicate a biological duplicate of *T. cruzi* M6241 cl6 (DTU-III). B is the *T. cruzi* Sylvio X10 cl1 (DTU-I). Different sample preparation strategies were used to test the robustness of the Tc-STAMS2 approach such as changing the pH for peptide desalting. B/acid refers to peptides derived from sample B were purified using acidic conditions (0.1% TFA). B/basic refers to peptides derived from sample B were purified using basic conditions (0.1% ammonia). Moreover, different analytical parameters were changed in order to test the robustness of the Tc-STAMS2 approach such as the MS/MS fragmentation type, CID—Collision-Induced Dissociation and HCD—Higher-energy collisional dissociation. Different sample amounts were loaded onto the nano LC column. High and Low indicate 1 and 0.5 ug, respectively. The Tc-STAMS2 approach was robust towards different analytical and experimental challenges.

<https://doi.org/10.1371/journal.pntd.0006351.g003>

another batch of *T. cruzi* strains were processed and acquired using similar chromatographic and MS conditions as described in Fig 1 in an inter-laboratory study perspective. Indeed, the mass spectra library was built with data acquired in the PR group in Odense, Denmark and the blind samples were acquired in the CEFAP mass spectrometry facility in São Paulo, Brazil. Although the instrument type and conditions were similar, a different chromatographic profile was obtained, S4 Fig. However, the biological duplicate unknown samples A1 and A2 from DTU-III matched correctly to DTU-III, Fig 3. The unknown sample B from the DTU-I also showed higher similarity scores with the MS/MS spectra library from DTU-I. The sample B was evaluated on different parameters: 1) sample preparation conditions such as acid and basic peptide desalting, 2) different chromatographic gradient such as 20 min, 70 min, 130 min, 3) different fragmentation methods, such as CID or HCD with a chromatographic gradient of 70 min and 4) different sample amount injected into the LC column. Even considering all these technical sources of variation, the similarity scores continued to match correctly to DTU-I showing the robustness of the Tc-STAMS2 towards different experimental conditions.

In addition, another MS/MS spectral library search software platform, DiagnoProt, was used instead of SpectraST [21]. DiagnoProt was able to differentiate the different DTUs and to associate the CL14 strain with the DTU -VI group as shown for the SpectraST software, Table 4. Due to that, two different spectral library search software could be implemented in the Tc-STAMS2 pipeline and used to identify *T. cruzi* strains, Tables 2 and 4. Furthermore, we tested the Tc-STAMS2 method using 13 strains belonging to six DTUs such as Sylvio X10 cl1, Sylvio X10/4 and G strains for DTU-I, Y and Esmeraldo strains for DTU-II, M6241 cl6

**Table 4. Spectral matching of different *T. cruzi* strains using the DiagnoProt software.** Seven *T. cruzi* strains (Sylvio X10 cl1, Y, M6241 cl6, CanIII cl1, MN cl2, CL Brener and CL14) belonging to six DTUs were selected to test the Tc-STAMS2 method. The database was constructed with six strains (Sylvio X10 cl1, Y, M6241 cl6, CanIII cl1, MN cl2 and CL Brener) and it was used for comparison of the different strains including CL14.

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
DTU-I	0.174	0.078	0.082	0.089	0.079	0.094
DTU-II	0.092	0.172	0.093	0.105	0.088	0.104
DTU-III	0.081	0.075	0.162	0.079	0.088	0.092
DTU-IV	0.095	0.084	0.082	0.155	0.079	0.089
DTU-V	0.070	0.065	0.077	0.063	0.155	0.095
DTU-VI	0.088	0.084	0.089	0.077	0.100	0.162
CL14	0.215	0.237	0.246	0.214	0.262	0.360

<https://doi.org/10.1371/journal.pntd.0006351.t004>

and 3869 strains for DTU-III, CanIII cl1 and José Júlio strains for DTU-IV, MN cl2 and NR cl3 strains for DTU-V and CL Brener and CL14 for DTU-VI. The Tc-STAMS2 method was able to correctly discriminate intra-DTU strains with high accuracy (Fig 4). Indeed, each strain matched correctly the corresponding one included into the database. Moreover, we assembled a MS/MS spectral library using two biological replicates of the 13 *T. cruzi* strains associated to six DTUs (Fig 5). One independent replicate of each strain was used to match the spectral library using the Tc-STAMS2 method. Each strain matched correctly to the corresponding DTU with high accuracy (Fig 5).

### Clustering analysis using peptide and protein identification

Database search was also performed to evaluate the similarity among the DTUs using peptide and protein identification results. Three database search platforms were used (MaxQuant, TPP and Proteome Discoverer). From four replicates of each DTU more than 7000 peptides and 4000 proteins were identified (S5 Fig, S1 Table, S2 Table and S3 Table). DTU-I and DTU-VI had the highest number of identifications due to the protein database used for the search. Indeed, Sylvio (DTU-I) and CL Brener (DTU-VI) are the two *T. cruzi* strains whose genome has been sequenced and their proteome annotated and deposited in the Uniprot database. Interestingly, only 30% of the MS/MS spectra were assigned, leaving behind a wealth of information for *T. cruzi* strain discrimination (S4 Table).

Analysis of variance (ANOVA  $p < 0.05$  followed by Benjamin-Hochberg FDR correction) was applied for the log2-transformed protein or peptide intensities previously identified using

Cepas	DTU-I			DTU-II		DTU-III		DTU-IV		DTU-V		DTU-VI	
	Tc Sylvio X10/4	Tc Sylvio X10 CL1	Tc G mucurae	Tc Y	Tc Esmeraldo	TCC Tc 3869	Tc M6241 CL6	Tc CanIII CL1	Tc Jose Julio	Tc MN CL2	Tc NR CL3	Tc CL14	Tc CLBrener
Tc Sylvio X10/4	0,553	0,004	0,076	0	0,009	0	0,035	0,001	0,015	0	0,027	0,001	0,001
Tc Sylvio X10 CL1	0	0,779	0,001	0,014	0,006	0,014	0	0,013	0	0,021	0	0,012	0,017
Tc G mucurae	0,148	0,001	0,584	0	0,01	0	0,057	0	0,044	0,001	0,031	0	0,001
Tc Y	0	0,019	0	0,507	0,004	0,055	0	0,037	0	0,018	0	0,048	0,041
Tc Esmeraldo	0,001	0,001	0	0	0,928	0	0	0,001	0	0,001	0	0,001	0,002
TCC Tc 3869	0	0,014	0	0,029	0,002	0,688	0	0,017	0	0,032	0	0,038	0,019
Tc M6241 CL6	0,024	0	0,025	0	0,005	0,001	0,52	0	0,048	0	0,025	0	0
Tc CanIII CL1	0	0,028	0	0,018	0,003	0,02	0	0,722	0,001	0,009	0	0,033	0,038
Tc Jose Julio	0,113	0,001	0,066	0,001	0,023	0	0,09	0,018	0,613	0,001	0,055	0	0,001
Tc MN CL2	0	0,016	0	0,01	0,006	0,023	0	0,006	0	0,763	0	0,022	0,017
Tc NR CL3	0,118	0	0,046	0	0,04	0	0,073	0	0,024	0,002	0,616	0,001	0,002
Tc CL14	0,001	0,017	0	0,037	0,007	0,043	0	0,034	0	0,038	0	0,426	0,111
Tc CLBrener	0,001	0,035	0	0,026	0,006	0,019	0	0,018	0	0,042	0	0,09	0,557

**Fig 4. *T. cruzi* strain discrimination based on spectral similarity searches.** Fourteen *T. cruzi* strains (Sylvio X10 cl1, Sylvio X10/4, G, Y, Esmeraldo, M6241 cl6, 3869, CanIII cl1, José Júlio, MN cl2, NR cl3, CL Brener and CL14) belonging to six DTUs were selected to test the Tc-STAMS2 method. The MS/MS spectral library was built using two biological replicates for each strain and one independent replicate was used to search against the library. The unique dot product SDSS score is reported.

<https://doi.org/10.1371/journal.pntd.0006351.g004>

Strains	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
<i>Tc Sylvio X10/4</i>	0,468	0,007	0,014	0,005	0,011	0
<i>Tc Sylvio X10 cl1</i>	0,491	0,009	0,013	0,011	0,019	0,025
<i>Tc G mucurae</i>	0,473	0,007	0,019	0,012	0,012	0
<i>Tc Y</i>	0,012	0,095	0,043	0,035	0,015	0,079
<i>Tc Esmeraldo</i>	0,001	0,915	0	0,001	0,001	0,002
<i>Tc TCC 3869</i>	0,009	0,007	0,582	0,014	0,029	0,048
<i>Tc M6241 cl6</i>	0,039	0,005	0,266	0,018	0,014	0
<i>Tc CanIII cl1</i>	0,02	0,006	0,016	0,642	0,009	0,056
<i>Tc Jose Julio</i>	0,125	0,021	0,042	0,275	0,026	0
<i>Tc MN cl2</i>	0,011	0,008	0,019	0,005	0,685	0,034
<i>Tc NR cl3</i>	0,117	0,037	0,037	0,009	0,255	0
<i>Tc CL14</i>	0,01	0,011	0,029	0,028	0,03	0,483
<i>Tc CLBrenner</i>	0,018	0,01	0,013	0,016	0,032	0,562

**Fig 5. Genotype discrimination based on spectral similarity searches.** Fourteen *T. cruzi* strains (Sylvio X10 cl1, Sylvio X10/4, G, Y, Esmeraldo, M6241 cl6, 3869, CanIII cl1, José Julio, MN cl2, NR cl3, CL Brener and CL14) belonging to six DTUs were selected to test the Tc-STAMS2 method. In order to build the MS/MS spectral library, each strain was assigned to the corresponding DTU. Two biological replicates for each strain were used to build the library and one independent replicate was used to match against the library. The unique dot product SDSS score.

<https://doi.org/10.1371/journal.pntd.0006351.g005>

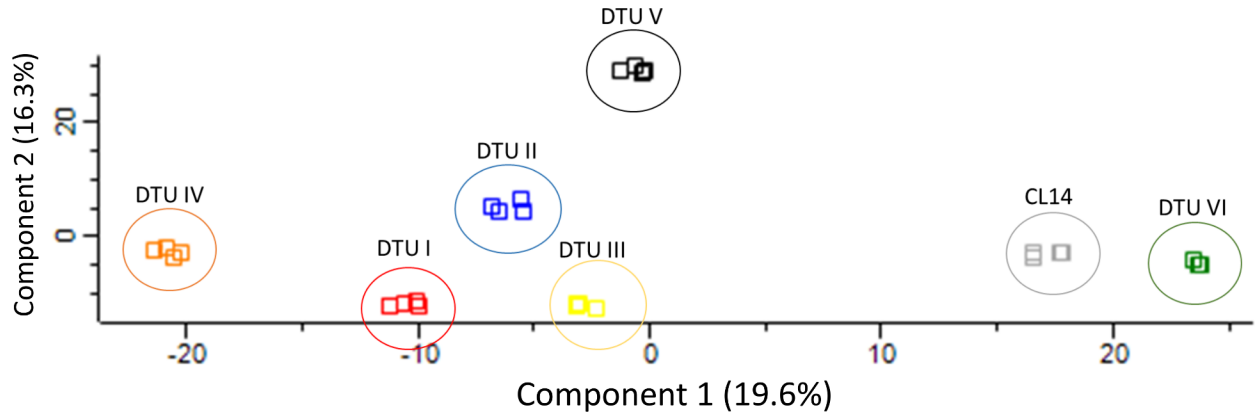
MaxQuant. A total of 1096 proteins and 6130 peptides showed significant difference in abundance among the six DTUs (S5 Table). The differentially expressed peptides and proteins were subjected to clustering analysis and visualized as heat-maps, S6A and S6B Fig. CL14 clustered together with CLBrenner (DTU-VI), indicating high similarity in the protein and peptide expression profile.

In addition, principal component analysis (PCA), which was applied in the differential expressed proteins, were able to discriminate the six DTUs showing a DTU-specific quantitative proteome repertoire (Fig 6). Interestingly, CL14 and CLBrenner, strains belonging to the DTU-VI, were also found close to each other, confirming the Euclidean clustering result obtained previously.

Additionally, to check the proteomic data reproducibility, a correlation analysis (R squared) between biological replicates from each DTU or between the different DTUs was performed using the log2-normalized intensities. As shown in S7 Fig, high correlation values were observed among replicates (R squared > 0.9). Interestingly, when we compared different DTUs (S8 Fig), the R squared dropped to 0.6–0.7, but high correlation was observed between CL14 and DTU-VI (R squared = 0.83), confirming once more the similarity between these two DTUs (S8 Fig).

## Discussion

Many studies have employed MS-based techniques to identify organisms, such as bacteria [34]. Önder et al. pioneered the use of MS/MS spectral libraries to precisely identify which animal the tick *Ixodes scapularis* was fed even if the feeding occurred months earlier [16]. Other strategies combined genetic information in conjunction with MS-based peptide identification for the correct assignment of microorganisms [35] [36, 37]. More recently, Shao et al. demonstrated that it is possible to identify *E. coli* strains with only the MS/MS fragmentation spectra, with no need for peptide identification (Shao et al. 2015).



**Fig 6. Principal component analysis of seven *T. cruzi* strains belonging to six DTUs based on differentially regulated proteins.** LC-MS/MS data acquired for Sylvio X10 cl1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL14 and CL Brener (DTU-VI) strains were searched against Uniprot *T. cruzi* Protein Database (release July 11, 2017; 51,738 entries) using MaxQuant computational platform as described in step C of Fig 1. Quantified proteins were analysed by ANOVA with Benjamin-Hochberg correction. 1096 differentially expressed proteins ( $q\text{value} < 0.05$ ) were used to perform the principal component analysis (PCA). DTU I—Red, DTU II—Blue, DTU III—Yellow, DTU IV—Orange, DTU V—Black, CLBrener—Green and CL14—Grey. All DTUs were separated and two strains belonging to DTU-VI (CL Brener and CL14) clustered together.

<https://doi.org/10.1371/journal.pntd.0006351.g006>

In the present study, we describe a genome-free, MS/MS spectral-matching methodology designed to identify different *T. cruzi* DTUs, named Tc-STAMS2. Firstly, a peptide MS/MS spectral library using three replicates from each of the six DTUs was built. The fourth replicate was used to test the ability of the method to differentiate each DTUs using the reference library. As shown in Tables 2 and 4 and Figs 5 and 6, this approach was able to differentiate strains belonging to similar and different DTUs. A unique assignment to the correct DTU was achieved. The method was tested using a dataset of peptide MS/MS spectra obtained from different growth conditions of *T. cruzi* and no bias was detected.

The next step was to test the MS/MS spectral library against phylogenetically related species such as *T. vivax* and distant organisms such as *E. coli*, *mus musculus* and *homo sapiens* (Fig 2). For such a test, another spectral library database was built with the same DTUs and MS/MS spectra from the CL14 strain and *T. vivax*. Interestingly, it is clear that even with thousands of fragmentation spectra, the scores obtained when comparing samples of human organism or *E. coli* with the library were close to zero, demonstrating the specificity of the method in identifying only samples of the species/strain, whose MS/MS spectrum is present in the library.

We also showed that the spectral matching method is robust even with inter and intra-laboratory source of variations using similar MS instruments, but from different laboratory and even performing changes in the sample preparation, chromatography and fragmentation method, it was still possible to correctly identify samples from DTU-III and DTU-I, as shown in Fig 3.

As expected, we also observed that the similarity score is dependent on the number of MS/MS spectra acquired. The longer the gradient time used, the greater the separation capacity of peptides prior to MS/MS analysis and the larger the number of acquired spectra, resulting in higher scores for the same sample when compared to the library. Using different fragmentation methods we also observed that CID provides scores higher than HCD when using 70 min gradient time. Although HCD provides high resolution MS/MS spectra CID provides faster MS/MS sequencing, thus generating more spectra that can be match with the spectral library.

In addition, fragmentation spectra were also subjected to database search analysis for peptide and protein identification using the MaxQuant software, TPP and Proteome Discoverer.

In general, the number of identifications of proteins and peptides were reproducible and consistent among different database search software. The larger number of proteins and peptides were observed for DTU-I and VI. Although the number of annotated proteins in UniProt database is greater for DTUs I and VI, the number of proteins identified to DTUs V, III and VI, was not significantly smaller. Due to that, a MS-based proteomic approach can be used to quantitatively compare the protein expression of different DTUs, even with differences in genome annotation among them and use these information's to identify DTU-specific pathways correlated with the strain phenotype.

Moreover, it was possible to cluster together DTU-VI and CL14 using the differential expressed proteins or peptides given by ANOVA test, **S6 Fig**. This result validates which is already known in the literature, where the CL14 belong to DTU-VI [5] and also validates the results obtained with the spectral matching, where the similarity score between CL14 and DTU-VI was higher compared to the other DTUs. Moreover, PCA analysis was also able to discriminate the six DTUs and determine the similarity between the DTU-VI and CL14, as seen by the close proximity between these two groups.

Clustering analysis using protein identification in different DTUs was also performed by Telleria et al., [38] however only 261 (experiment 1) and 172 (experiment 2) 2DE protein spots were considered for this analysis. In our study, the clustering analysis was performed using 1096 proteins differentially expressed among DTUs, increasing the number of features to build a robust *T. cruzi* clustering (**S6 Fig**).

The Tc-STAMS2 strategy presented here is robust, accurate, easy to perform and completely automated. Dworzanski, J. P. [37] and Shao, W. et al. [15] used a similar methodology for identification of bacteria, however this is the first time that spectral matching is applied to discriminate different *T. cruzi* DTUs.

The transmission of Chagas disease through blood transfusion and organ transplant, besides the triatomine vector, poses several public health challenges. Due to that, novel methods identify and characterize DTUs can offer opportunities to understand the DTUs diversity, link with their phenotype and provide another tool for molecular epidemiology. Recently, MALDI-based strategy was developed for the direct identification of trypanosomatids based on the MS profile, named DIT-MALDI TOF [39]. However, this method does not have the resolution to discriminate DTUs using the MS profile. In this study we present a *T. cruzi* strain typing based on MS/MS spectral matching. The Tc-STAMS2 method is robust, sensitive and powerful and it is based on the identification of peptides from their MS/MS spectra. The method could be used to complement other already established methods. The proposed method can help in the research of epidemiology to identify *T. cruzi* strains only with the use of fragmentation spectra without the need for genomic data. In this study, we used the epimastigote form of *T. cruzi* to implement the Tc-STAMS2 method. However, quantitative proteomic analysis of *T. cruzi* epimastigote and trypomastigote life stages have shown distinct protein expression profile [40, 41]. Due to that, more studies are needed to confirm the specificity of the Tc-STAMS2 method using other *T. cruzi* stages. Since MS/MS spectra are available to the research community further data mining is possible in order to improve our understanding on the biology of the different *T. cruzi* strains. The methodology shown in this study will provide a complementary tool to the current nucleic-based testing and have the possibility to be extended to other parasitic diseases.

## Supporting information

**S1 Fig. LC-MS chromatographic profile of tryptic peptides belonging to the four replicates from Sylvio X10 c11 (DTU-I).** The Pearson correlation score is reported on the right side of

the chromatogram and was calculated based on the quantified proteins in each replicate compared to the first replicate.

(TIF)

**S2 Fig. LC-MS chromatographic profile of peptides belonging to the six *T. cruzi* strains.**

(TIF)

**S3 Fig. LC-MS chromatographic profile of peptides belonging to the CLBrenner and CL14 *T. cruzi* strains.** Three replicates for each strain are reported.

(TIF)

**S4 Fig. LC-MS chromatographic profile of peptides belonging to the Sylvio X10 cl1 (DTU-I) acquired in the PR group, Odense, Denmark and in the CEFAP mass spectrometry facility at USP, Sao Paulo, Brazil (blind test).**

(TIF)

**S5 Fig.** A. Peptide identifications using MaxQuant, TPP and Proteome Discoverer software for the Sylvio X10 cl1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL Brener (DTU-VI). B. Protein identifications using MaxQuant, TPP and Proteome Discoverer software for the Sylvio X10 cl1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL Brener (DTU-VI).

(TIF)

**S6 Fig.** A) proteins and B) peptides differentially regulated between the different DTUs with  $q < 0.05$  were hierarchically clustered based on Euclidean distances using Perseus software.

(PPTX)

**S7 Fig. Multi scatter plot obtained comparing the quantified proteins of the six DTUs with four technical replicates each using all quantified proteins.** A—Sylvio X10 cl1 (DTU-I), B—Y (DTU-II), C—M6241 cl6 (DTU-III), D—CanIII cl1 (DTU-IV), E—MN cl2 (DTU-V), F—CL Brener (DTU-VI) *T. cruzi* strains.

(TIF)

**S8 Fig. Multi scatter plot obtained comparing the quantified peptides of the Sylvio X10 cl1 (DTU-I), Y (DTU-II), M6241 cl6 (DTU-III), CanIII cl1 (DTU-IV), MN cl2 (DTU-V), CL Brener and CL14 (DTU-VI) *T. cruzi* strains.**

(TIF)

**S1 Table. Number of protein IDs using MaxQuant computational platform.** Column BV with the header “Type” indicates the DTUs where each protein was identified.

(XLSX)

**S2 Table. Number of protein IDs using Trans-Proteomic Pipeline computational platform.** Column K with the header “Type” indicates the DTUs where each protein was identified.

(XLSX)

**S3 Table. Number of protein IDs using Proteome Discoverer computational platform.** Column AB with the header “Type” indicates the DTUs where each protein was identified.

(XLSX)

**S4 Table. Number of PSMs and unannotated MS/MS spectra.** Protein accessions are organized based on the DTU where they were identified.

(XLSX)



**S5 Table. Regulated proteins and peptides among the six DTUs.** Protein accessions are organized based on the DTU where they were identified. (XLSX)

## Acknowledgments

Marta Campaner is acknowledged for providing genotyped *T. cruzi* strains. Prof. Sirlei Daffre, Prof. João Marcelo Pereira Alves and Prof. Claudio R. Marinho are acknowledged for the fruitful discussion. The Biomass facility at CEFAP-USP is acknowledged for acquiring some of the LC-MS/MS runs. Prof. Graziella Eliza Rosein and Prof. Paolo di Mascio (IQ-USP) are acknowledged for access to the Orbitrap Fusion Tribrid Lumos (Thermo Fisher) instrument.

## Author Contributions

**Conceptualization:** Giuseppe Palmisano.

**Data curation:** Gilberto Santos de Oliveira, Rebeca Kawahara, Giuseppe Palmisano.

**Formal analysis:** Gilberto Santos de Oliveira, Giuseppe Palmisano.

**Funding acquisition:** Giuseppe Palmisano.

**Investigation:** Giuseppe Palmisano.

**Methodology:** Gilberto Santos de Oliveira, Rebeca Kawahara, Livia Rosa-Fernandes, Simon Ngao Mule, Giuseppe Palmisano.

**Project administration:** Giuseppe Palmisano.

**Resources:** Marta M. G. Teixeira, Martin R. Larsen.

**Software:** Gilberto Santos de Oliveira.

**Supervision:** Giuseppe Palmisano.

**Validation:** Gilberto Santos de Oliveira, Livia Rosa-Fernandes, Simon Ngao Mule, Carla Cristi Avila, Marta M. G. Teixeira, Martin R. Larsen, Giuseppe Palmisano.

**Visualization:** Giuseppe Palmisano.

**Writing – original draft:** Gilberto Santos de Oliveira, Rebeca Kawahara, Livia Rosa-Fernandes, Giuseppe Palmisano.

**Writing – review & editing:** Gilberto Santos de Oliveira, Rebeca Kawahara, Livia Rosa-Fernandes, Carla Cristi Avila, Marta M. G. Teixeira, Martin R. Larsen, Giuseppe Palmisano.

## References

1. WHO. Chagas disease (American trypanosomiasis). <http://www.who.int/mediacentre/factsheets/fs340/en/>. 2017.
2. WHO. Chagas disease (american trypanosomiasis). World Health Organization; 2015.
3. Rassi A AR Jr, José Antonio Marin-Neto. Chagas disease. *Lancet*. 2010; 375:1388–402. [https://doi.org/10.1016/S0140-6736\(10\)60061-X](https://doi.org/10.1016/S0140-6736(10)60061-X) PMID: 20399979
4. Pinto AY, Valente SA, Valente Vda C, Ferreira Junior AG, Coura JR. [Acute phase of Chagas disease in the Brazilian Amazon region: study of 233 cases from Para, Amapa and Maranhao observed between 1988 and 2005]. *Revista da Sociedade Brasileira de Medicina Tropical*. 2008; 41(6):602–14. Epub 2009/01/15. PMID: 19142440.
5. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz*. 2009; 104(7):1051–4. PMID: 20027478.

6. Zingales B. *Trypanosoma cruzi*: um parasita, dois parasitas ou vários parasitas da doença de Chagas? Revista da Biologia. 2011; 6b:44–8.
7. Henriksson J, Aslund L, Macina RA, Franke de Cazzulo BM, Cazzulo JJ, Frascch AC, et al. Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi*. Molecular and biochemical parasitology. 1990; 42(2):213–23. PMID: [2270104](#).
8. Macedo AM, Melo MN, Gomes RF, Pena SD. DNA fingerprints: a tool for identification and determination of the relationships between species and strains of Leishmania. Molecular and biochemical parasitology. 1992; 53(1–2):63–70. PMID: [1501645](#).
9. Brisse S, Verhoef J, Tibayrenc M. Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. International journal for parasitology. 2001; 31(11):1218–26. PMID: [11513891](#).
10. Cosentino RO, Aguero F. A simple strain typing assay for *Trypanosoma cruzi*: discrimination of major evolutionary lineages from a single amplification product. PLoS Neglected Tropical Diseases. 2012; 6(7):e1777. <https://doi.org/10.1371/journal.pntd.0001777> PMID: [22860154](#); PubMed Central PMCID: PMC3409129.
11. Tibayrenc M, Neubauer K, Barnabe C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proceedings of the National Academy of Sciences. 1993; 90(4):1335–9. PMID: [8433991](#); PubMed Central PMCID: PMC45867.
12. Messenger LA, Miles MA, Bern C. Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. Expert review of anti-infective therapy. 2015; 13(8):995–1029. Epub 2015/07/15. <https://doi.org/10.1586/14787210.2015.1056158> PMID: [26162928](#); PubMed Central PMCID: PMC4784490.
13. Pedrioli PG. Trans-proteomic pipeline: a pipeline for proteomic analysis. Methods Mol Biol. 2010; 604:213–38. [https://doi.org/10.1007/978-1-60761-444-9\\_15](https://doi.org/10.1007/978-1-60761-444-9_15) PMID: [20013374](#).
14. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013; 13(1):22–4. <https://doi.org/10.1002/pmic.201200439> PMID: [23148064](#).
15. Shao W, Zhang M, Lam H, Lau SC. A peptide identification-free, genome sequence-independent shotgun proteomics workflow for strain-level bacterial differentiation. Scientific reports. 2015; 5:14337. <https://doi.org/10.1038/srep14337> PMID: [26395646](#); PubMed Central PMCID: PMC4585814.
16. Onder O, Shao W, Kemps BD, Lam H, Brisson D. Identifying sources of tick blood meals using unidentified tandem mass spectral libraries. Nature communications. 2013; 4:1746. <https://doi.org/10.1038/ncomms2730> PMID: [23612287](#); PubMed Central PMCID: PMC3635114.
17. Ndao M. Diagnosis of parasitic diseases: old and new approaches. Hindawi Publishing Corporation. 2009;15. <https://doi.org/10.1155/2009/278246> PMID: [20069111](#)
18. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016; 13(9):731–40. <https://doi.org/10.1038/nmeth.3901> PMID: [27348712](#).
19. Lam H, Deutsch EW, Eddes JS, Eng JK, Stein SE, Aebersold R. Building consensus spectral libraries for peptide identification in proteomics. Nat Methods. 2008; 5(10):873–5. <https://doi.org/10.1038/nmeth.1254> PMID: [18806791](#); PubMed Central PMCID: PMCPMC2637392.
20. Perez-Morales D, Lanz-Mendoza H, Hurtado G, Martinez-Espinosa R, Espinoza B. Proteomic analysis of *Trypanosoma cruzi* epimastigotes subjected to heat shock. Journal of biomedicine & biotechnology. 2012; 2012:902803. <https://doi.org/10.1155/2012/902803> PMID: [22287837](#); PubMed Central PMCID: PMC3263753.
21. Silva ARF, Lima DB, Leyva A, Duran R, Batthyany C, Aquino PF, et al. DiagnoProt: a tool for discovery of new molecules by mass spectrometry. Bioinformatics. 2017. <https://doi.org/10.1093/bioinformatics/btx093> PMID: [28186229](#).
22. Jackson AP, Goyard S, Xia D, Foth BJ, Sanders M, Wastling JM, et al. Global Gene Expression Profiling through the Complete Life Cycle of *Trypanosoma vivax*. PLoS Negl Trop Dis. 2015; 9(8):e0003975. <https://doi.org/10.1371/journal.pntd.0003975> PMID: [26266535](#); PubMed Central PMCID: PMCPMC4534299.
23. Camargo E. Growth and differentiation in *Trypanosoma cruzi*. i. origin of metacyclic trypanosomes in liquid media (1964). Revista do Instituto de Medicina Tropical de São Paulo. 1964; 12:93–100.
24. Lima L, Espinosa-Alvarez O, Pinto CM, Cavazzana M Jr., Pavan AC, Carranza JC, et al. New insights into the evolution of the *Trypanosoma cruzi* clade provided by a new trypanosome species tightly linked to Neotropical Pteronotus bats and related to an Australian lineage of trypanosomes. Parasites & vectors. 2015; 8:657. Epub 2015/12/25. <https://doi.org/10.1186/s13071-015-1255-x> PMID: [26701154](#); PubMed Central PMCID: PMC4690318.

25. Palmisano G, Parker BL, Engholm-Keller K, Lendal SE, Kulej K, Schulz M, et al. A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Molecular Cell Proteomics*. 2012; 11(11):1191–202. <https://doi.org/10.1074/mcp.M112.017509> PMID: 22843994; PubMed Central PMCID: PMC3494194.
26. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, et al. Development and validation of a spectral library searching method for peptide identification from ms/ms. *Proteomics*. 2007; 7(5):655–67. <https://doi.org/10.1002/pmic.200600625> PMID: 17295354.
27. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008; 24(21):2534–6. <https://doi.org/10.1093/bioinformatics/btn323> PMID: 18606607; PubMed Central PMCID: PMC2732273.
28. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, et al. A guided tour of the transproteomic pipeline. *Proteomics*. 2010; 10(6):1150–9. <https://doi.org/10.1002/pmic.200900375> PMID: 20101611; PubMed Central PMCID: PMC3017125.
29. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrom*. 1994; 5(9):859–66. [https://doi.org/10.1016/1044-0305\(94\)87009-8](https://doi.org/10.1016/1044-0305(94)87009-8) PMID: 24222034.
30. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*. 2008; 26(12):1367–72. <https://doi.org/10.1038/nbt.1511> PMID: 19029910.
31. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*. 2011; 10(4):1794–805. <https://doi.org/10.1021/pr101065j> PMID: 21254760.
32. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74(20):5383–92. PMID: 12403597.
33. Lee HJ, Jeong SK, Na K, Lee MJ, Lee SH, Lim JS, et al. Comprehensive genome-wide proteomic analysis of human placental tissue for the Chromosome-Centric Human Proteome Project. *Journal of proteome research*. 2013; 12(6):2458–66. <https://doi.org/10.1021/pr301040g> PMID: 23362793.
34. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat Genet*. 2003;33 Suppl:311–23. <https://doi.org/10.1038/ng1106> PMID: 12610541.
35. Jabbour RE, Deshpande SV, Wade MM, Stanford MF, Wick CH, Zulich AW, et al. Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. *Applied and environmental microbiology*. 2010; 76(11):3637–44. <https://doi.org/10.1128/AEM.00055-10> PMID: 20363779; PubMed Central PMCID: PMC2876470.
36. Dworzanski JP, Snyder AP, Chen R, Zhang H, Wishart D, Li L. Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring. *Anal Chem*. 2004; 76(8):2355–66. <https://doi.org/10.1021/ac0349781> PMID: 15080748.
37. Dworzanski JP, Deshpande SV, Chen R, Jabbour RE, Snyder AP, Wick CH, et al. Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *Journal of proteome research*. 2006; 5(1):76–87. <https://doi.org/10.1021/pr050294t> PMID: 16396497.
38. Telleria J, Biron DG, Brizard JP, Demetree E, Seveno M, Barnabe C, et al. Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(47):20411–6. <https://doi.org/10.1073/pnas.1015496107> PMID: 21059959; PubMed Central PMCID: PMC2996636.
39. Avila CC, Almeida FG, Palmisano G. Direct identification of trypanosomatids by matrix-assisted laser desorption ionization-time of flight mass spectrometry (DIT MALDI-TOF MS). *Journal of mass spectrometry: JMS*. 2016; 51(8):549–57. Epub 2016/09/24. <https://doi.org/10.1002/jms.3763> PMID: 27659938.
40. Parodi-Talice A, Monteiro-Goes V, Arrambide N, Avila AR, Duran R, Correa A, et al. Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. *Journal of mass spectrometry: JMS*. 2007; 42(11):1422–32. Epub 2007/10/26. <https://doi.org/10.1002/jms.1267> PMID: 17960573.
41. Alves MJ, Kawahara R, Viner R, Colli W, Mattos EC, Thaysen-Andersen M, et al. Comprehensive glycoproteomic profiling of the epimastigote and trypomastigote stages of *Trypanosoma cruzi*. *Journal of proteomics*. 2017; 151:182–92. Epub 2016/06/19. <https://doi.org/10.1016/j.jprot.2016.05.034> PMID: 27318177.