In-depth proteome coverage of in vitro-cultured Treponema pallidum and quantitative comparison analyses with in vivo-grown treponemes

Simon Houston, Alloysius Gomez, Andrew Geppert, Mara C. Goodyear and Caroline E. Cameron

2024

Original citation:

Houston, S., Gomez, A., Geppert, A., Goodyear, M. C., & Cameron, C. E. (2024). In-depth proteome coverage of in vitro-cultured Treponema pallidum and quantitative comparison analyses with in vivo-grown treponemes. *Journal of Proteome Research*, *23*(5), 1725–1743. https://doi.org/10.1021/acs.jproteome.3c00891

Journal of proteome research

pubs.acs.org/jpr

Article

In-Depth Proteome Coverage of In Vitro-Cultured *Treponema pallidum* and Quantitative Comparison Analyses with In Vivo-Grown Treponemes

Published as part of Journal of Proteome Research virtual special issue "Canadian Proteomics".

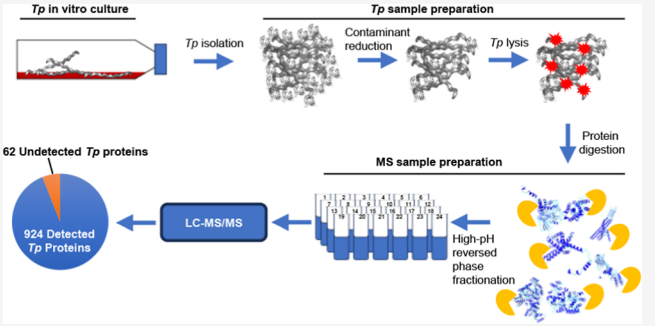
Simon Houston, Alloysius Gomez, Andrew Geppert, Mara C. Goodyear, and Caroline E. Cameron*

Cite This: *J. Proteome Res.* 2024, 23, 1725−1743

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Previous mass spectrometry (MS)-based global proteomics studies have detected a combined total of 86% of all *Treponema pallidum* proteins under infection conditions (in vivo-grown *T. pallidum*). Recently, a method was developed for the long-term culture of *T. pallidum* under in vitro conditions (in vitro-cultured *T. pallidum*). Herein, we used our previously reported optimized MS-based proteomics approach to characterize the *T. pallidum* global protein expression profile under in vitro culture conditions. These analyses provided a proteome coverage of 94%, which extends the combined *T. pallidum* proteome coverage from the previously reported 86% to a new combined total of 95%. This study provides a more complete understanding of the protein repertoire of *T. pallidum*. Further, comparison of the in vitro-expressed proteome with the previously determined in vivo-expressed proteome identifies only a few proteomic changes between the two growth conditions, reinforcing the suitability of in vitro-cultured *T. pallidum* as an alternative to rabbit-based treponemal growth. The MS proteomics data have been deposited in the MassIVE repository with the data set identifier MSV000093603 (ProteomeXchange identifier PXD047625).



KEYWORDS: *Treponema pallidum, syphilis, global proteomics, mass spectrometry (MS), protein expression profile, outer membrane proteins, vaccine candidates*

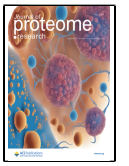
INTRODUCTION

The spirochete bacterium *Treponema pallidum* ssp. *pallidum* (hereafter *T. pallidum*) is the causative agent of syphilis, a chronic, multistage, sexually transmitted infection (STI). Syphilis remains a global health concern, as evident through rising rates of infectious and congenital syphilis across the world[1−6] and the increased risk of acquiring and transmitting other STIs, including HIV, with symptomatic syphilis infections.[7,8]

T. pallidum is a highly invasive pathogen, with the ability to disseminate throughout the host soon after infection, infect any organ or tissue, and establish persistent, lifelong infection in the absence of treatment.[9] Several unique features of *T. pallidum* contribute to its success as a pathogen, including its scarcity of surface-exposed outer membrane proteins (OMPs) and ability to undergo antigenic and phase variation of OMPs belonging to the 12-membered *T. pallidum* repeat (Tpr) protein family.[9−21]

T. pallidum is an obligate human pathogen, and its existence within the nutrient-rich environment of the human host has led to extensive genome reduction[12,22] during its centuries-long relationship with the host.[9] The *T. pallidum* genome comprises a 1.14 Mb single circular chromosome containing approximately 1000 predicted protein-coding genes.[12] Global transcriptomics studies performed on *T. pallidum*, using DNA microarray and RNA-seq methodologies, have shown that all predicted protein-coding genes are expressed at the transcript level.[23,24] Parallel global proteomics studies have resulted in less protein expression coverage,[25−27] having been hindered by the historical requirement for growth of *T. pallidum* in rabbit testes (in vivo-grown *T. pallidum*), an environment which provides relatively low numbers of treponemes (compared to other, more conventional, Gram-negative bacteria) and high numbers and concentrations of contaminating rabbit proteins. Consequently, isolation of *T. pallidum* from this complex in vivo environment diminishes the likelihood of effectively detecting low-abundance treponemal proteins using MS-based proteomics approaches.

Received: December 15, 2023
Revised: March 14, 2024
Accepted: March 22, 2024
Published: April 18, 2024



Three previous MS-based global proteomics studies,[25−27] including our prior study which used an optimized sample preparation workflow,[25] detected the expression of a combined total of 847 *T. pallidum* proteins, corresponding to 86% coverage of the *T. pallidum* proteome. These MS-based global *T. pallidum* proteomics studies have also provided insights into the expression profiles of specific protein classes of interest under infection conditions, including the detection of known/predicted OMPs, proteins being considered as vaccine candidates, and proteins with putative virulence functions.[25−27]

Since 1905, when *T. pallidum* was identified as the causative agent of syphilis,[28,29] research into this enigmatic pathogen has been hindered by the inability to continuously culture the spirochete under in vitro growth conditions. After many efforts and the publication of several research studies focused on achieving continuous culture,[30−34] in 2018, Edmondson, Hu, and Norris succeeded in establishing long-term culturing of *T. pallidum* under nonaxenic in vitro conditions in the presence of rabbit epithelial cells (in vitro-cultured *T. pallidum*).[35] This long-term culture system has now been applied successfully to the in vitro cultivation of several *T. pallidum* strains and has been shown to result in the propagation of treponemes that retain their characteristic spiral-shaped structure, motility, multiplication rate, and full infectivity in the rabbit model of infection studies.[35−37] Furthermore, a comparative transcriptomics study of in vivo-grown and in vitro-cultured *T. pallidum* using RNA-seq demonstrated that the overall global transcript profiles from both growth conditions are highly correlated; no significant differences were observed in the expression levels of 91% of *T. pallidum* genes.[24] This latter finding supports the possible replacement of in vivo-grown treponemes with its in vitro-cultured counterpart in future *T. pallidum* studies.[24]

Although the long-term in vitro culture method was established six years ago, proteomic investigations with an overall goal of in-depth characterization of the global protein expression profile of in vitro-cultured *T. pallidum* have yet to be performed. In the present study, we aimed to address this knowledge gap using an MS-based proteomics approach to achieve deep coverage of the in vitro proteome. This was investigated using a sample preparation methodology that was previously optimized for in vivo-grown *T. pallidum* that focused on the preparation of MS-compatible samples[25] (Figure 1). When combined with high-pH reversed-phase fractionation and protein detection using Orbitrap Fusion Tribrid high-resolution tandem MS, we were able to achieve 94% proteome coverage from the analyses of three in vitro-cultured *T. pallidum* biological replicate samples. This finding extends our knowledge of the global protein repertoire expressed by *T. pallidum* and provides additional insights into the expression of key treponemal proteins that have been reported to be vaccine candidates or play a role in *T. pallidum* pathogenesis. Expression- and quantitation-based comparative analyses with in vivo-grown *T. pallidum* showed that the global protein profiles derived from both growth conditions are similar. Importantly, the latter finding provides additional evidence for the suitability of the use of in vitro-cultured *T. pallidum* as an alternative to in vivo-grown *T. pallidum* in future treponemal research.

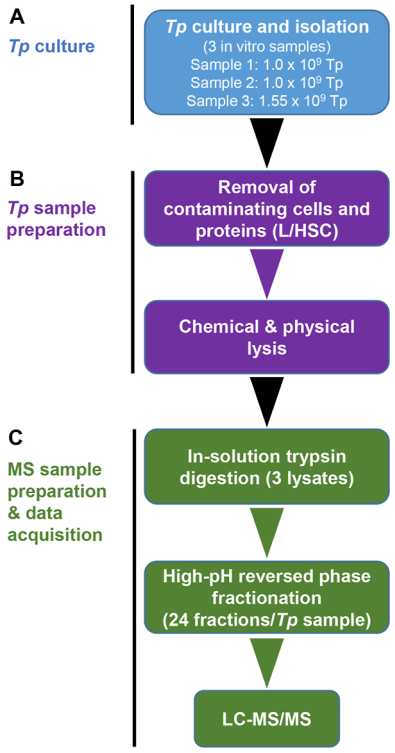


Figure 1. MS-based proteomics workflow for the global analyses of in vitro-cultured *T. pallidum*. (A) In vitro culture and isolation of three *T. pallidum* (*Tp*) samples. The number of treponemes in each in vitro-cultured *T. pallidum* sample is shown. For comparison, the number of in vivo-grown treponemes used for proteomic analyses in our previous study was $1.7-4.1 \times 10^8$.[25] (B) Two main steps involved in *Tp* lysate/sample preparation. Low- and high-speed centrifugation (L/HSC) was used to remove contaminating cells and soluble contaminants and to wash the isolated treponemes. (C) The three major steps for MS sample preparation and data acquisition. For (B,C), the optimized MS methodology developed for in vivo-grown *Tp*[25] was used to process the in vitro-grown *Tp*.

MATERIALS AND METHODS

Ethics Statement

The animal study was reviewed and approved by the local institutional review board at the University of Victoria under protocol 2020-024 (4) and was conducted in strict accordance with standard accepted principles as set forth by the Canadian Council on Animal Care, the National Institutes of Health, and the United States Department of Agriculture in a facility accredited by the Canadian Council on Animal Care and the American Association for the Accreditation of Laboratory Animal Care.

Propagation and Harvesting of *T. pallidum*

T. pallidum subsp. *pallidum* (Nichols strain) was propagated in outbred male specific pathogen-free (SPF) New Zealand white rabbits (3.0−3.5 kg, Charles River Laboratories, ON, Canada), as described elsewhere.[38] Treponemes were extracted from rabbit testes 10−12 days postinfection in *T. pallidum* culture medium 2 (TpCM-2).[35] Extractions were performed at room temperature in a microaerophilic chamber (model 2002 incubator; Coy Laboratories, Grass Lake, MI, USA) in an atmosphere of 1.5−3% O_2 and 5% CO_2, balanced with N_2. Insoluble rabbit gross cellular debris was separated and removed via two centrifugation steps at 220*g* (5 min each, room temperature) followed by two additional centrifugation steps at 400*g* (7 min each, room temperature). Treponemes in suspension were enumerated by darkfield microscopy (Nikon Eclipse E600 darkfield microscope; Nikon Canada, Mississauga,

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

ON, Canada) using a Petroff-Hauser counting chamber (Hauser Scientific, Horsham, PA).

In Vitro Culture of *T. pallidum*

Freshly harvested in vivo-grown *T. pallidum* (extracted in TpCM-2) was used for in vitro culture and subculture of *T. pallidum* in the presence of Sf1Ep (NBL-11) cottontail rabbit epithelial cells (ATCC CCL-68) (American Type Culture Collection [ATCC], Rockville, MD, USA). In vitro culturing of *T. pallidum* was performed as previously described.[35−37] Briefly, Sf1Ep cells were cultured in Eagle's minimum essential medium (EMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) in Sarstedt T75 tissue culture flasks at 37 °C in 5% CO_2 (8×10^5 cells/T75 flask). At least 3 h prior to initiating the in vitro culture of *T. pallidum*, Sf1Ep media was removed from the tissue culture flask and replaced with TpCM-2. The tissue culture plate was then transferred to a low-oxygen incubator (Thermo Scientific Forma Series 3130) to allow for pre-equilibration at 34 °C in an atmosphere of 1.5% O_2 and 5% CO_2, balanced with N_2. To initiate *T. pallidum* in vitro culturing, the pre-equilibrated Sf1Ep tissue culture was inoculated with approximately 1 mL of fresh culture containing 3×10^7 treponemes extracted in TpCM-2. For *T. pallidum* subculturing, the TpCM-2 medium (15−20 mL) was removed from the tissue flasks containing in vitro-cultured *T. pallidum* and set aside (reserved media). To dissociate the cells, the flask was rinsed with trypsin-free dissociation buffer[37] which was combined with reserved media. An additional 2 mL of trypsin-free dissociation buffer was added to the flask, which was then incubated for 30 min at 34 °C in 1.5% O_2 and 5% CO_2, balanced with N_2 (with tapping used to facilitate cell dissociation). Cells were monitored by using an Olympus CKX41 inverted microscope to ensure complete dissociation. The reserved media (5 mL) was added to the flask containing the dissociated Sf1Ep and in vitro-cultured *T. pallidum* cells, and the cells were gently rinsed. For scaling up, 2−3 mL of dissociated cells was used to inoculate new flasks containing Sf1Ep tissue cultures that had been prepared as described above. Final in vitro-cultured *T. pallidum* samples suspended in sterile 0.9% saline (0.9% w/v NaCl, pH 7.0) contained between 1.0 and 1.55×10^9 treponemes (Figure 1A).

T. pallidum Protein Sample Preparation

Protein samples for MS analyses were prepared using the same sample preparation workflow that was previously optimized for the proteomic analyses of in vivo-grown *T. pallidum*.[25] In the present in vitro study, we only used the final optimal conditions and methods for *T. pallidum* sample preparation that were identified in the previous in vivo study[25] (Figure 1B), as described as follows. Insoluble rabbit gross cellular debris was removed from in vitro-cultured *T. pallidum* via two slow-speed centrifugation steps at 220*g* (5 min each, room temperature). High-speed centrifugation steps were then performed to remove soluble contaminants in the treponemal supernatant and to wash the suspended treponemes. In this step, the *T. pallidum*-containing supernatants were centrifuged at 17,000*g* for 5 min at room temperature, after which the *T. pallidum* pellet was gently resuspended in sterile saline (0.9% w/v NaCl, pH 7.0) and centrifuged again at 17,000*g* for 5 min at room temperature. The supernatant was discarded, and chemical lysis of the *T. pallidum* pellets was performed by resuspending and incubating in lysis buffer (500 μL per sample; 50 mM ammonium bicarbonate pH 8.0, 0.9% sodium deoxycholate [Sigma-Aldrich Canada Co., Oakville, ON, Canada]) for 30 min on ice, with 30 s vortex mixing steps every 5 min. To ensure complete lysis, the samples were then subjected to physical lysis via ultrasonication at 6 °C using a Covaris ME220 focused ultrasonicator with the following parameters: 3 × 30 s at 6 W, 20% duty factor, 200 cycles per burst, and 30 s rest between the three cycles (Covaris, LLC, Woburn, MA, USA). The *T. pallidum* lysate samples were then centrifuged at 17,000*g* for 10 min at 4 °C to remove precipitated, aggregated, or insoluble proteins and cellular debris. Protein concentrations were determined by measuring the absorbance of the lysate supernatants at 280 nm (Beckman Coulter DU 730 life science UV/vis spectrophotometer; Beckman Coulter Canada, Mississauga, ON, Canada) and by performing bicinchoninic acid (BCA) assays (Thermo Scientific Pierce BCA protein assay kit; Thermo Fisher Scientific, Ottawa, ON, Canada). Treponeme cell numbers and motility were determined following high-speed centrifugation using darkfield microscopy, as described above.

In-Solution Trypsin Digestion

Lysate samples from in vitro-cultured *T. pallidum* were reduced at 37 °C for 30 min by the addition of dithiothreitol (10 mM final concentration) (Sigma-Aldrich Canada Co., Oakville, ON, Canada) and then alkylated at room temperature in the dark for 30 min by the addition of iodoacetamide (40 mM final concentration) (Fisher Scientific Canada, Edmonton, AB, Canada). In-solution tryptic digestion was performed at 37 °C for 18 h by the addition of 100 μg of trypsin (Promega, Madison, WI, USA) per mg of protein (Figure 1C). Formic acid (1% final concentration) was added to stop tryptic digestions. Solid phase extraction using hydrophilic−lipophilic-balanced (HLB) columns (Waters Corporation, Milford, MA, USA) was performed to enable cleanup of each trypsin-digested protein sample. The elution of each digested protein sample from the HLB columns was performed via the addition of 60% acetonitrile/0.1% formic acid (300 μL per sample). Samples were then reduced to dryness in a SpeedVac vacuum concentrator.

High-pH Reversed-Phase Fractionation

To reduce sample complexity, high-pH reversed-phase fractionation was used to separate each trypsin-digested *T. pallidum* sample into 24 fractions based on hydrophobicity (Figure 1C). Fractionation was performed on an Agilent 1290 HPLC system (Agilent, Santa Clara, CA, USA) equipped with an XBridge BEH300 C18 HPLC column (250 × 4.6 mm, 5 μm bead size, 300 Å pore size; Waters Corporation, Milford, MA, USA). The *T. pallidum* samples were diluted to a total volume of 0.9 mL with buffer A (10 mM ammonium hydroxide, pH 10.0) and injected onto the column with a constant flow rate set to 0.75 mL/min. Following a 5 min equilibration period in buffer A, a gradient of 5−45% buffer B (80% acetonitrile and 10 mM ammonium hydroxide, pH 10) was applied over 75 min. Fractions were collected every min for 96 min, lyophilized, rehydrated with 300 μL of 2% acetonitrile/0.1% formic acid, and concatenated into 24 fractions by combining every 24th fraction (e.g., fractions 1, 25, 49, and 73 were combined).

Liquid Chromatography−Tandem MS (LC−MS/MS)

Fractionated samples were subjected to LC−MS/MS for global protein identifications (Figure 1C). A 10 μL aliquot of each concatenated fraction was separated via online reversed-phase liquid chromatography using a Thermo Scientific EASY-nLC 1000 system with an Acclaim PepMap100 reversed-phase C18 precolumn (100 μm I.D., 2 cm length, 5 μm bead size, 100 Å pore size) (Thermo Fisher Scientific, San Jose, CA) and an

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Acclaim PepMap100 C-18 reversed-phase nanoanalytical column (75 μm I.D., 15 cm length, 3 μm bead size, 100 Å pore size) (Thermo Fisher Scientific, San Jose, CA) at a flow rate of 300 nL/min. The chromatography system was coupled online with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a Nanospray Flex NG source (Thermo Fisher Scientific). A 120 min gradient of solvent A (A; 2% acetonitrile/0.1% formic acid) and solvent B (B; 90% acetonitrile/0.1% formic acid) used to separate tryptic peptides comprised the following steps: (i) 0−100 min, gradient change from 95% A/5% B to 58% A/42% B, (ii) 100−115 min, gradient change from 58% A/42% B to 0% A/100% B, and (iii) 115−120 min, gradient held at 0% A/100% B. The Orbitrap Fusion Tribrid mass spectrometer instrument parameters (Fusion Tune 3.4 software) were as follows: nanoelectrospray ion source with spray voltage = 2.5 kV; capillary temperature = 275 °C; survey MS1 scan = m/z range 350−1800 in profile mode, resolution = 120,000 fwhm@200m/z, number of microscans = 1, and automatic maximum injection time. The lock mass for siloxane (445.12003 m/z) was used as a reference for internal calibration. Data-dependent acquisition Orbitrap survey spectra were scheduled at least every 3 s, with the software determining the maximum number of MS/MS acquisitions during this period. The automatic gain control (AGC) target value for FTMS (Fourier transform MS) was set to standard (400,000 counts) and the automatic maximum fill time to ensure optimal sensitivity and cycle time. The most intense ions with a charge state of 2−5 that exceeded 50,000 counts were selected for higher-energy collisional dissociation (HCD) ion trap MS/MS fragmentation in the ion-routing multipole. Monoisotopic Precursor Selection (MIPS) was enabled, and dynamic exclusion settings were as follows: repeat count = 2; repeat duration = 10 s; exclusion duration = 15 s with a 10 ppm mass window. The data-dependent (ddMS2) IT HCD scan settings were as follows: quadrupole isolation window = 1.6 Da; ion trap rapid scan rate; automode normal m/z range; centroid detection; 1 microscan; automatic maximum injection time; automatic gain control (AGC) target value = 10,000 counts; and stepped HCD collision energy = 28, 30, and 32%.

Protein Identification: Parameters and Data Analysis

Data analyses for protein identification and validation were performed as previously described.[25] Specifically, raw files were created using XCalibur v.4.3.73.11 software (Thermo Scientific). Tandem mass spectra were extracted, and the charge state was deconvoluted by Proteome Discoverer version 2.5 (Thermo Scientific). Deisotoping was not performed. All MS/MS samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version IseNode in Proteome Discoverer 2.5.0.400) containing a single customized database consisting of 1261 *T. pallidum* proteins (Nichols strain, NCBI reference sequence NC_021490, all proteome annotation revisions from June 17, 2013 to July 4, 2021), 895 rabbit proteins (UniProt *Oryctolagus cuniculus* proteome, UP000001811), and common contaminants [https://www.thegpm.org/crap/] (Table S1). Similar to our previous in vivo-grown *T. pallidum* global proteomics study,[25] searches for *T. pallidum* and rabbit proteins were performed simultaneously using the combined *T. pallidum*/rabbit protein database. Database search parameters were the same as those of previous in vivo-grown *T. pallidum* studies:[25,27] precursor tolerance (parent ion tolerance) = 10 ppm; MS/MS tolerance (fragment ion mass tolerance) = 0.6 Da; enzyme specificity = trypsin, with a maximum of two missed cleavages allowed; instrument type = ESI-TRAP; fixed modification = carbamidomethylation (C); and variable modifications = acetylation of the peptide N-terminus and oxidation (M). Scaffold (version Scaffold_5.1.2; Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Percolator posterior error probability calculation,[39] and protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Peptide and Protein Prophet algorithms.[40] Percolator (as a node in Proteome Discoverer) was used to generate decoy sequences (randomized *T. pallidum* and rabbit sequences from our customized *T. pallidum* database). The false discovery rate (FDR) calculated by Scaffold for confident protein identification was less than 5%. Proteins harboring similar peptides that could not be differentiated through MS/MS analysis alone were clustered together, adhering to the principles of parsimony.[41] Identification of nonidentical protein paralogs was confirmed by the detection of one or more peptides that are unique to a single paralog. Identical paralogs (full-length proteins with identical amino acid sequences, e.g., TPANIC_0117 [TprC] and TPANIC_0131 [TprD]) could not be distinguished as separate proteins. Proteome coverage was calculated based on the *T. pallidum* Nichols strain (NCBI reference sequence NC_021490, July 2021 annotation, 986 proteins); 964 proteins from predicted protein-coding genes, 15 proteins potentially encoded by "pseudo genes", and seven detected proteins that are not annotated in the July 4, 2021 proteome (pre-July 2021 proteome annotations).

Label-Free Quantification: Parameters and Data Analysis

Label-free quantification (LFQ), performed as previously described,[25] was based on peptide ion peak intensities as a relative quantitative measure. In summary, MSFragger (version 3.4), Philosopher (version 4.1.1), and IonQuant were accessed using the FragPipe proteomics pipeline (version 17.1) and used for database searching/peptide identification, downstream postprocessing, and protein quantification with FDR-controlled match-between-run (MBR) functionality, respectively.[42−44] Raw spectral files were converted to mzML format (ProteoWizard MS convert: http://proteowizard.sourceforge.net) and loaded into LFQ-MBR-configured FragPipe. The customized *T. pallidum* database, as described above and elsewhere,[25] was used for searches and peptide identifications. MSFragger database search parameters were as follows: precursor mass tolerance = −20 to 20 ppm; fragment mass (MS/MS) tolerance = 0.6 Da; enzyme specificity = trypsin, with a maximum of two missed cleavages allowed; fixed modification = carbamidomethylation (C); variable modifications = acetylation of the peptide N-terminus and oxidation (M). LFQ-MBR in MS1 Quant was performed with the following parameters: ion quant selected; match between runs enabled; protein quant = MaxLFQ; min ions = 2. Default settings for all of the other parameters were used. PeptideProphet and ProteinProphet were used to validate and filter peptides and proteins for high-confidence identifications (≥95%). Protein relative abundances were determined by ranking the MaxLFQ intensity values from the highest to the lowest intensities.

Global Functional Categorization of *T. pallidum* Proteins

All treponemal protein sequences from the customized *T. pallidum* database, as described above, were submitted to the

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

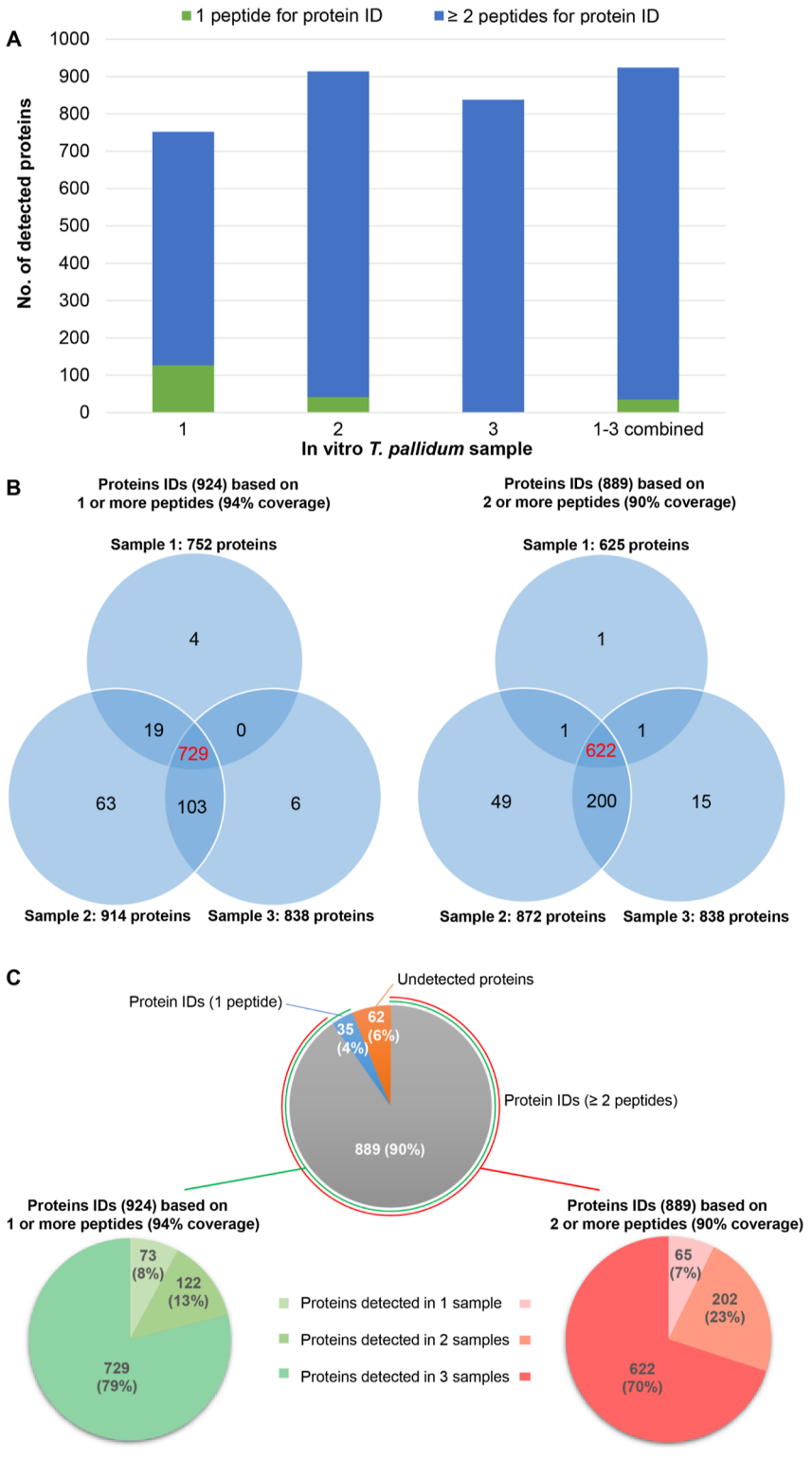


Figure 2. Deep proteome coverage of in vitro-cultured *T. pallidum*. (A) Total number of proteins detected individually and combined from the three in vitro-grown *T. pallidum* biological replicate samples. The number of proteins identified via the detection of one tryptic peptide (green) or two tryptic peptides (blue) is indicated. (B) Venn diagrams showing the number of shared and exclusive protein identifications in the three *T. pallidum* samples. The number of proteins identified in all three samples is highlighted in red text. (C) Pie charts summarizing the distribution of protein identification frequencies for the 924 *T. pallidum* proteins identified from the three samples via the detection of one or more peptides (left) and for the 889 treponemal proteins identified from the three samples via the detection of at least two peptides (right). Values in parentheses indicate the percentages of *T. pallidum* proteins found in each of the three identification frequency groups.

genome-wide functional annotation tool, eggNOG-mapper (version 2.1.9) (http://eggnog-mapper.embl.de/).[45] Searches were performed using the eggNOG 5 database,[46] default search parameters (minimum hit e-value = 0.001, minimum hit bit-score = 60, percentage identity = 40, minimum % of query coverage = 20, minimum % of subject coverage = 20), and

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Table 1. Summary of *T. pallidum* Proteins Identified Based on the Detection of One or More Peptides

Number of *T. pallidum* proteins detected in each of the 3 biological replicate samples			
In vitro sample	Number of proteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1	752 (76.3% proteome coverage)	127 (12.9% proteome coverage)	625 (63.4% proteome coverage)
2	914 (92.7% proteome coverage)	42 (4.3% proteome coverage)	872 (88.4% proteome coverage)
3	838 (85.0% proteome coverage)	0 (0% proteome coverage)	838 (85.0% proteome coverage)

Total coverage of the *T. pallidum* proteome			
In vitro samples	Number of proteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1−3 combined	924 (93.7% proteome coverage) (851 detected in ≥2 biological replicates) (729 detected in 3 biological replicates)	35 (3.5% proteome coverage)	889 (90.2% proteome coverage) (202 detected in ≥2 biological replicates) (622 detected in 3 biological replicates)

Proteome coverage of *T. pallidum*: this study and previous studies[25−27]			
In vitro samples	Number of proteins detected	1 peptide for protein ID[b]	2 or more peptides for protein ID[c]
1−3 combined and previous studies[25−27]	936 (95.0% proteome coverage)[a] (90 proteins identified only in the present study)	18 (1.8% proteome coverage)	901 (91.4% proteome coverage) (72 proteins identified only in the present study)

Detection of *T. pallidum* miniproteins of unknown function			
In vitro samples	Number of miniproteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1−3 combined	39 (57.4% miniprotein coverage) (8 miniproteins uniquely detected in the present study)	4 (5.9% miniprotein coverage)	35 (51.5% miniprotein coverage) (7 miniproteins uniquely detected in the present study)

Detection of hypothetical proteins and proteins of unknown function			
In vitro samples	Number of proteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1−3 combined	230/268 total proteins detected (85.8% coverage)	17 (6.3% coverage)	213 (79.5% coverage)
	132/166 "hypothetical proteins" (79.5% coverage)	12 (7.2% coverage)	120 (72.3% coverage)
	33/35 DUF domain proteins (94.3% coverage)	4 (11.4% coverage)	29 (82.9% coverage)
	65/67 "poorly annotated proteins" (97.0% coverage)	1 (1.5% coverage)	64 (95.5% coverage)

Detection of known or predicted OMPs			
In vitro samples	Number of proteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1−3 combined	31 (91.2% known/predicted OMP coverage) (1 OMP uniquely detected in the present study)	1 (2.94% known/predicted OMP coverage)	30 (88.24% known/predicted OMP coverage) (1 OMP uniquely detected in the present study)

Detection of putative pathogenesis-related proteins (PPRPs)			
In vitro samples	Number of proteins detected	1 peptide for protein ID	2 or more peptides for protein ID
1−3 combined	32 (94.1% coverage) (1 PPRP identified only in the present study)	2 (5.9% coverage)	30 (88.2% coverage) (1 PPRP identified only in the present study)

[a]Identified proteins, Tp0315 and Tp0619, included as a single detection (both proteins contain the single identified peptide). [b]Protein IDs based on the detection of a single peptide only in the present study (17 additional proteins identified with single peptides in the present study were also detected in previous studies). [c]Protein IDs based on the detection of two or more peptides in the present study.

default annotation options (taxonomic scope = auto adjust per query, orthology restrictions = transfer annotations from any ortholog, gene ontology evidence = transfer nonelectronic annotations, PFAM refinement = report PFAM domains from orthologs, SMART annotation = skip SMART annotation). The functional categorization of each *T. pallidum* protein was based on the "Clusters of Orthologous Genes" (COG)[47] category (https://www.ncbi.nlm.nih.gov/research/cog) assigned to each of the submitted proteins. EggNOG-mapper was also used to assign COG functional categories to the top 50 high-abundance transcripts identified previously,[24] which allowed for the direct comparison with the top 50 high-abundance proteins identified in the present study. All transcript and protein comparative analyses were performed using Microsoft Excel.

RESULTS AND DISCUSSION

In-Depth Global Proteome Profiling of In Vitro-Cultured *T. pallidum*

The ability to culture *T. pallidum* in vitro using a rabbit epithelial cell coincubation system has lessened the challenge of studying this obligate human pathogen at the genetic and transcriptomic levels.[24,48] In the present study, the major aim was to advance our knowledge of *T. pallidum* at the proteomic level by using the long-term in vitro culturing system[35−37] to grow and isolate sufficient treponemes for in-depth global protein expression profiling. This system was used in combination with a proteomics workflow previously optimized for *T. pallidum*[25] (Figure 1B,C) to analyze three in vitro-cultured *T. pallidum* biological replicate samples.

As shown in Figure 2 and Table 1, MS analyses resulted in the identification of between 752 and 914 *T. pallidum* proteins in each of the three samples via the detection of at least one peptide. In total, 13,722 *T. pallidum* peptides were detected in the three treponemal samples, corresponding to 924 *T. pallidum* proteins and representing a total proteome coverage of 94% (Figure 2A,1 and S2). When the detection of at least two peptides was used for protein identifications, a combined total of 889 *T. pallidum* proteins were identified, representing 90% total proteome coverage (Figure 2A and Table 1). A total of 3192 unique rabbit peptides were identified, which resulted in 387 rabbit protein identifications. Only two peptides were shown to be shared between *T. pallidum* and the rabbit (TPANIC_0844 [LTGMAFR] and TPANIC_0984 [ELISNASDALDK]) (Table S2). Expression of TPANIC_0844 and TPANIC_0984 was confirmed via the detection of 18−30 and 31−44 additional

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

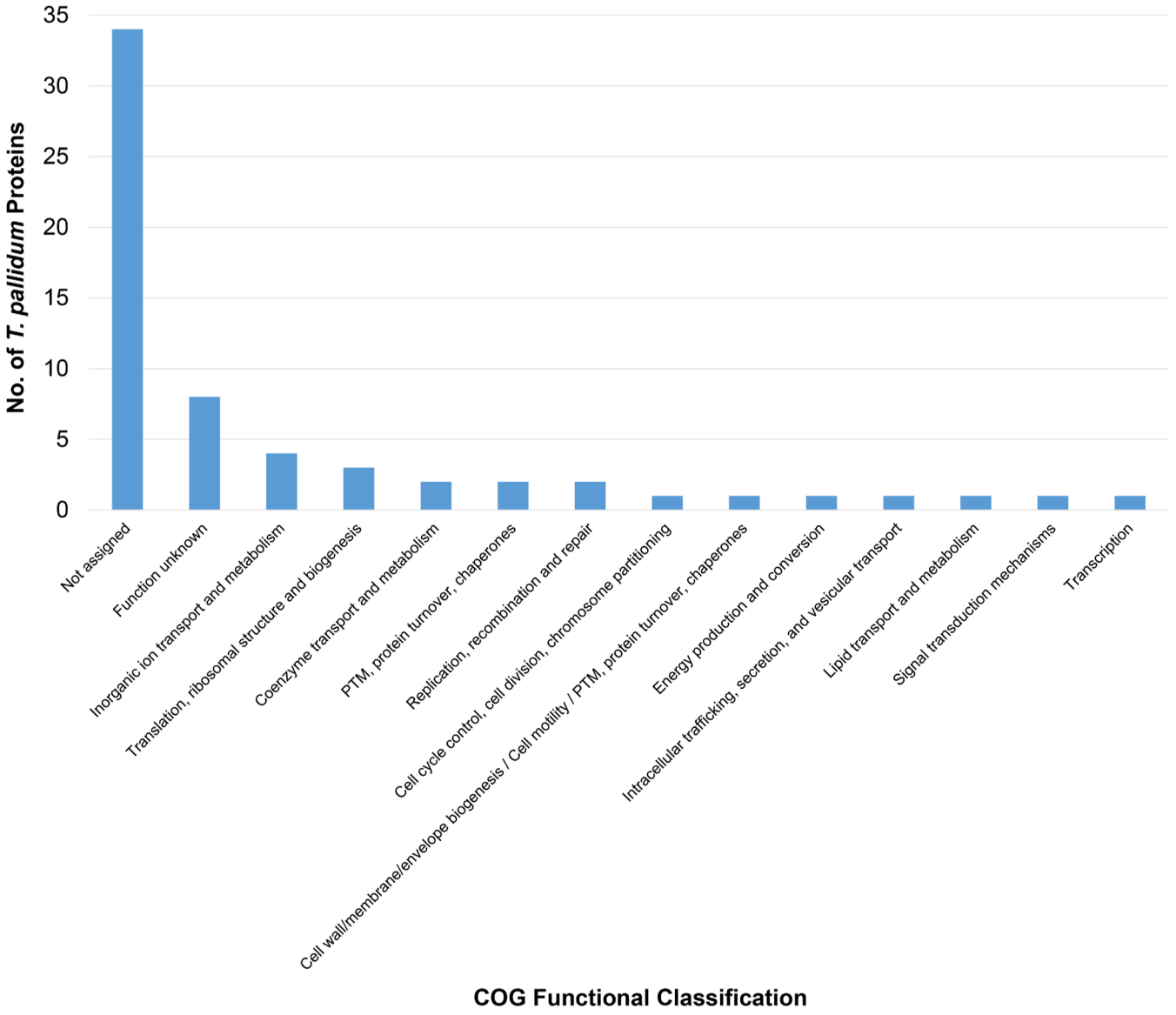


Figure 3. Functional categorization of undetected proteins from in vitro-cultured *T. pallidum* in the present study. Bar chart showing the distribution of 62 undetected *T. pallidum* proteins based on the COG functional categories. PTM: post-translational modification. Not assigned: proteins that could not be assigned COG functional categories by eggNOG-mapper.

peptides in the three samples, respectively, which are unique to these two *T. pallidum* proteins (Table S2). Detailed MS data and Scaffold peptide reports for the three in vitro-cultured *T. pallidum* samples (including rabbit peptides and proteins) are presented in Table S2. Overall, the majority of identified treponemal proteins were detected in all three biological replicate samples, with the detection of 729 treponemal proteins in all three samples (or 79% of all identifications) based on at least one peptide and 622 proteins in all three samples (or 70% of all identifications) based on at least two peptides (Figure 2B,C, Tables 1, S3, and S4). The in-depth proteome coverage obtained here provides a comprehensive snapshot of *T. pallidum* protein expression patterns during in vitro culture and allows for comparative studies with *T. pallidum* isolated from infection conditions (in vivo-grown *T. pallidum*).

Undetected Treponemal Proteins from In Vitro-Cultured *T. pallidum*

In the present study, we were unable to detect the expression of 62 *T. pallidum* proteins, which corresponds to 6% of the treponemal proteome (Figure 3 and Table S5). Similar to a previous global proteomics study performed in our laboratory that focused on in vivo-grown *T. pallidum*,[25] the majority of undetected proteins from in vitro-cultured *T. pallidum* (61%) were annotated in the proteome as proteins of unknown function, specifically "hypothetical proteins" and DUF (domain of unknown function) domain-containing proteins (Table S5). In addition, 68% of these undetected proteins were either classified as "function unknown" or could not be assigned a function using COG[45,47] analysis (Figure 3 and Table S5). In accordance with our previous findings from in vivo-grown treponemes,[25] a high proportion (60%) of the undetected proteins were also found to be miniproteins comprising less than 150 amino acids that often contain high numbers of lysine and arginine residues.[49] This results in the production of particularly

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

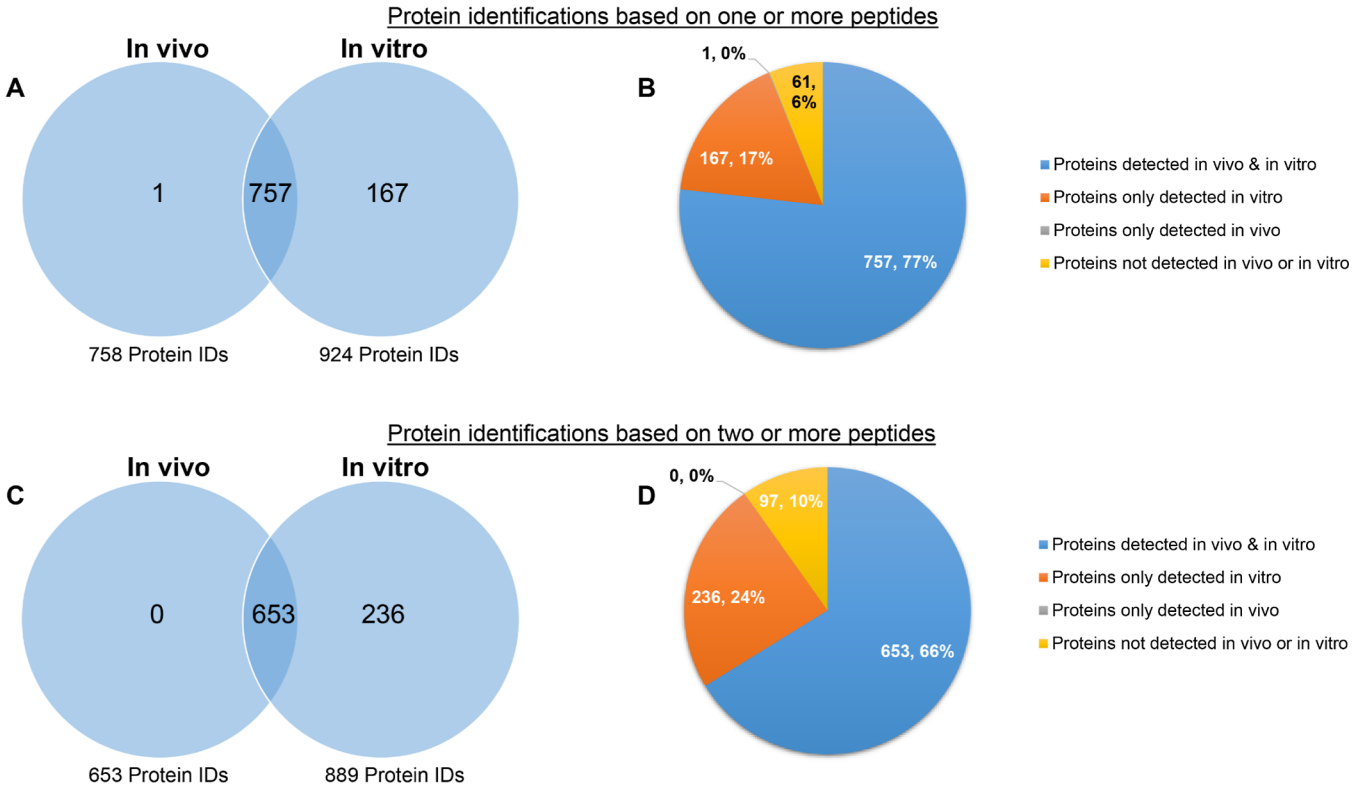


Figure 4. Comparison of global protein expression profiles between in vivo-grown and in vitro-cultured *T. pallidum*. Venn diagrams showing the number of shared and exclusive protein identifications under the two growth conditions based on the detection of (A) at least one tryptic peptide or (C) two or more tryptic peptides. Pie charts summarizing the number and percentage of protein identifications and unidentified proteins from in vivo- and in vitro-cultured *T. pallidum* based on the detection of one or more (B) or two or more (D) peptides.

small trypsin-cleaved peptides during proteomics sample preparation that are below the size detection limit of most mass spectrometers. Therefore, these physicochemical properties are likely important contributing factors that limit the detection of this protein class by MS.[25]

To determine the tendency for a subcellular compartment bias, the cellular locations of the 62 undetected *T. pallidum* proteins were predicted using the bacterial localization prediction tool PSORTb 3.0.3 (https://www.psort.org/psortb/)[50] and the subcellular location annotations from UniProt Archive (UniParc) searches (https://www.uniprot.org/). Thirty-nine of the 62 undetected proteins could not be assigned predicted cellular locations, 13 were predicted to be located within the cytoplasmic membrane, eight were predicted to be located within the cytoplasm, and two were predicted to be located within either the cytoplasm or the ribosome (Table S5). While these findings may imply a tendency toward a cytoplasmic membrane location, the absence of location predictions for most of these proteins prevents the formation of definitive conclusions into subcellular compartment bias among the undetected proteins.

Interestingly, two predicted OMPs belonging to the 12-membered *T. pallidum* repeat (Tpr) protein family[9,51−53] were also undetected (TprA/TPANIC_0009 and TprF/TPANIC_0316) (Table S5). Since members of this protein family are currently being investigated as promising vaccine candidates, confirmation of protein expression within the pathogen is an important consideration in vaccine design and may help inform vaccine candidate selection. Another undetected protein of interest was annotated as an LptF/LptG family permease (TPANIC_0884) (Table S5). This protein is necessary for the transport of LPS from the inner membrane to the outer membrane in conventional Gram-negative bacteria.[54] It is well-known that *T. pallidum* does not contain LPS on its surface,[9,12] and thus, the lack of confirmation of the expression of this essential LPS export protein may help explain the absence of this major inflammation-inducing endotoxin on the surface of *T. pallidum*.[12,13,55] However, as a limitation of this study, it is important to note that MS can only provide information with regard to detected proteins and cannot confirm that a lack of protein detection equates to a lack of protein expression. Therefore, it remains possible that these undetected proteins are expressed by *T. pallidum*, but inherent factors, including MS incompatibility, protein physiochemical properties, and low protein expression levels, prevent detection by MS.

Expression-Based Comparative Analyses of In Vivo-Grown and In Vitro-Cultured *T. pallidum*

Prior to the establishment of the in vitro culture system for *T. pallidum* growth, and since *T. pallidum* is an obligate host pathogen that has a limited host range, isolation of *T. pallidum* for experimentation has historically required growth in the testes of rabbits.[38] Consequently, to link (at the protein expression level) prior experiments conducted with in vivo-grown *T. pallidum* with current and future experiments conducted with in vitro-grown *T. pallidum*, in this study, we also compared the global protein expression profile of in vivo-grown *T. pallidum* with the profile of in vitro-grown *T. pallidum*. In both studies, the same optimized methods and parameters were used for sample preparation and MS. This methodological consistency facilitated the comparative analyses by reducing or eliminating protein expression profiling differences that arise due to experimental

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

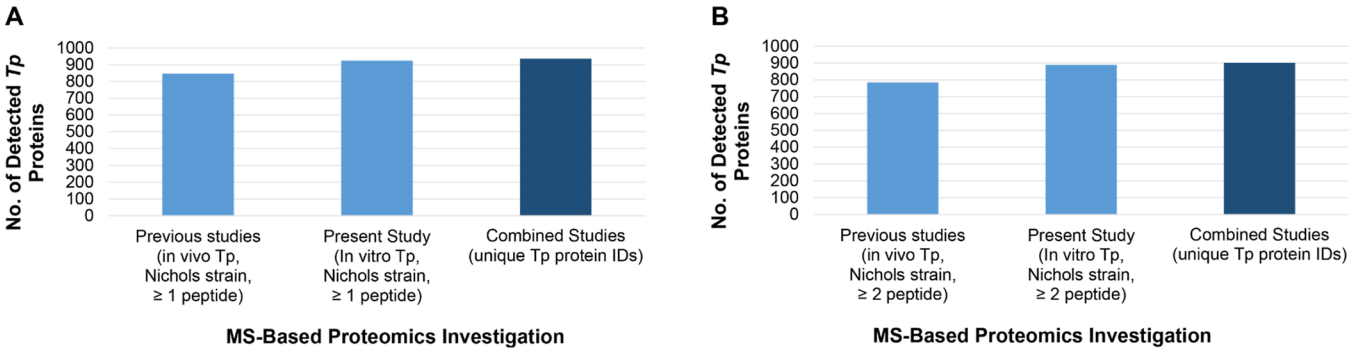


Figure 5. Combined proteome coverage of *T. pallidum*. Bar graphs showing the total number of *T. pallidum* proteins detected in the three previous combined in vivo studies and the present in vitro study (light blue bars) and the combined number of proteins detected in all studies (dark blue bars) based on the detection of (A) at least one or (B) at least two tryptic peptides.

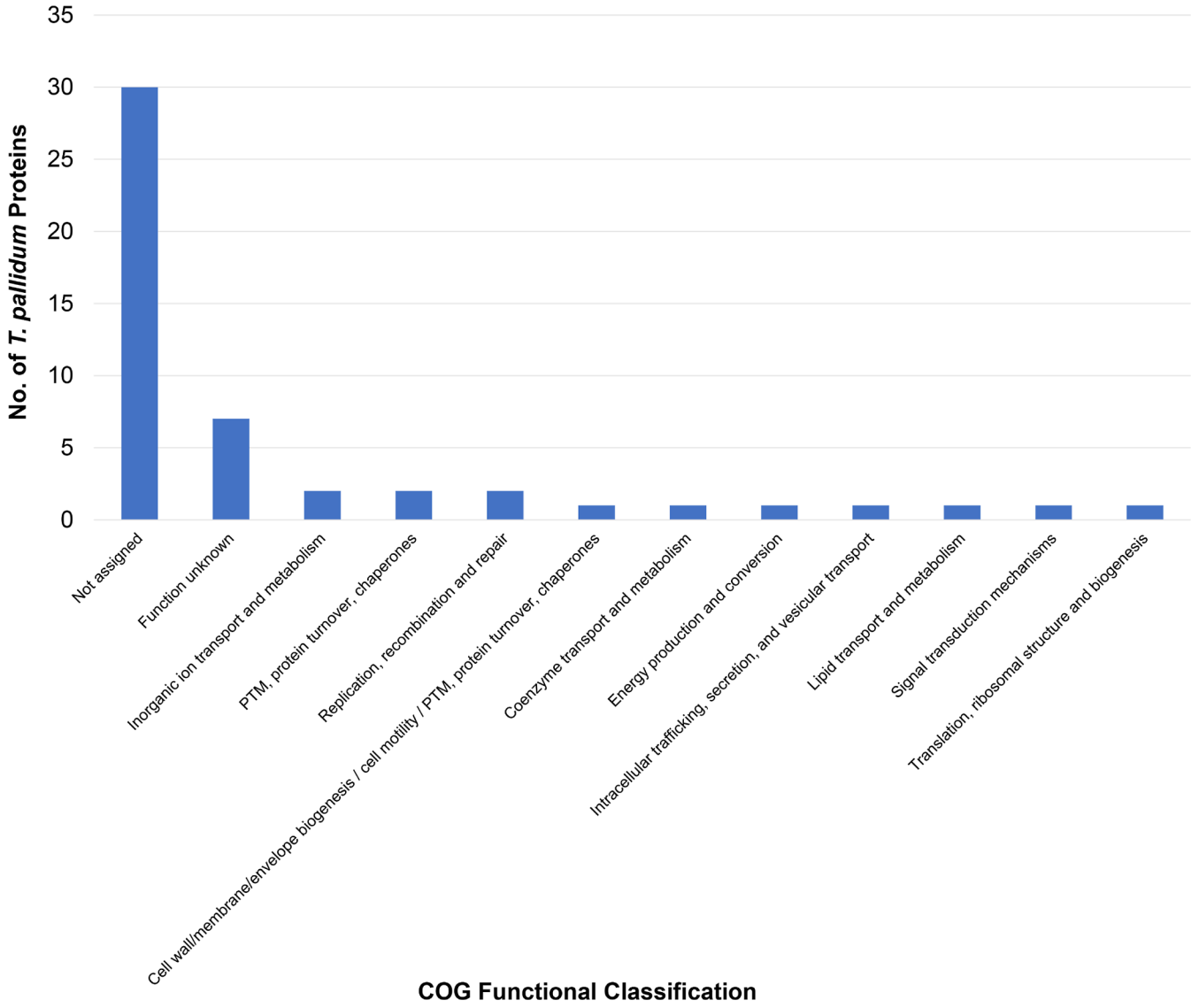


Figure 6. Functional categorization of undetected proteins from in vitro-cultured and in vivo-grown *T. pallidum*. Bar chart showing the COG functional category distribution of 50 *T. pallidum* proteins not detected in previous[25−27] or current MS-based global proteomics studies. PTM: post-translational modification. Not assigned: proteins that could not be assigned COG functional categories by eggNOG-mapper.

differences. As shown in Figure 4, only one treponemal protein was detected exclusively from in vivo-cultured *T. pallidum*. This protein, TPANIC_0849 (50S ribosomal protein L35), was detected previously in a single sample, based on the detection of a single tryptic peptide.[25] TPANIC_0849 is also a lysine-rich miniprotein (66 amino acid length), which may have contributed to the lack of in vitro detection, as previously described.[25] Seventy-seven percent of all *T. pallidum* proteins

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

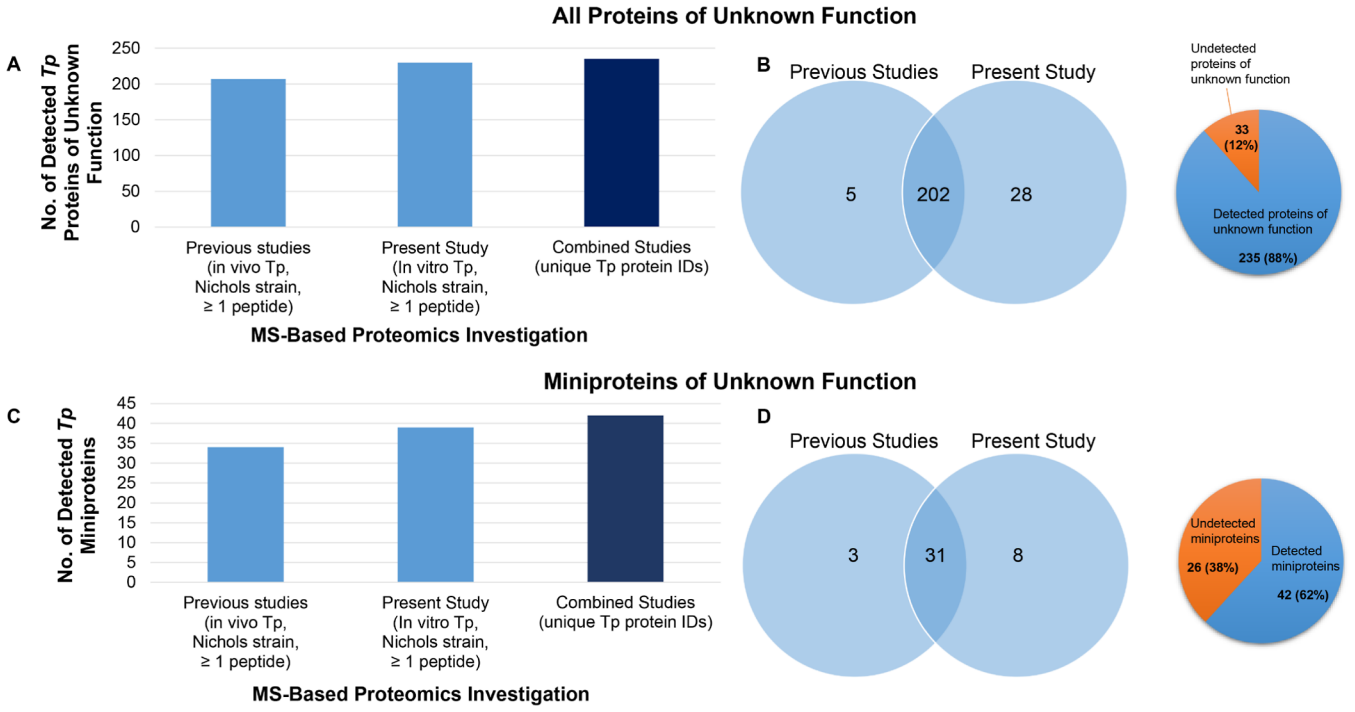


Figure 7. *Treponema pallidum* "proteins of unknown function" proteome coverage. Bar graphs depicting the number of (A) all proteins of unknown function (no size limit) and (C) miniproteins of unknown function (150 amino acids or less) detected in the three previous combined in vivo studies and the present in vitro study (light blue bars) and the combined number detected in all studies (dark blue bars). Venn diagrams summarizing the number of shared and exclusive identifications corresponding to (B) all proteins of unknown function and (D) miniproteins of unknown function detected in previous and present MS-based proteomics studies. Pie charts show (B) the combined numbers and percentages of detected and undetected proteins of unknown function (no size limit) and (D) miniproteins of unknown function.

were detected in both in vivo- and in vitro-cultured treponemes based on the detection of at least one tryptic peptide per protein (Figure 4A,B). When the detection of a minimum of two peptides was used for protein identification, 66% of all treponemal proteins were detected in both in vivo- and in vitro-cultured *T. pallidum* (Figure 4C,D). Overall, these findings demonstrate a high level of overlap between the global expression profiles of in vivo-grown and in vitro-cultured *T. pallidum*.

The proteome coverage of in vitro-cultured treponemes was 17% higher compared to that of in vivo-grown treponemes. Given the more complex environment of rabbit testes from which in vivo-grown treponemes are isolated, this proteome coverage difference is expected to be due to the carryover of higher levels of rabbit proteins in the in vivo-grown samples compared with those of the in vitro-cultured samples. These highly prevalent contaminants mask the low-abundance *T. pallidum* proteins, which decreases the probability of MS-based identification since their concentrations are below the detection limit of the MS analyzer.[56]

Combined Global Proteome Coverage of *T. pallidum*

Previously, 86% of all proteins from in vivo-grown *T. pallidum* were detected and identified in three MS-based global proteomics studies.[25−27] Here, we identified an additional 90 *T. pallidum* proteins, based on the detection of at least one peptide, which extended the combined in vivo/in vitro global proteome coverage to 936 proteins, corresponding to 95% coverage (Figure 5A, Tables 1 and S6). A combined total of 901 *T. pallidum* proteins were identified (91% combined in vivo/in vitro proteome coverage) when protein identifications were based on the detection of two or more peptides; 72 proteins were identified only in the present study using the two-peptide identification criteria (Figure 5B, Tables 1 and S6). Collectively, the combined findings from past[25−27] and present *T. pallidum* MS-based global proteomics studies have facilitated the assembling of one of the highest coverages of any bacterial proteome to date.

Undetected Proteins from In Vitro-Cultured and In Vivo-Grown *T. pallidum*

When the results from the three in vivo-grown *T. pallidum* MS-based global proteomics studies[25−27] were combined with the in vitro-cultured *T. pallidum* findings from the current study, 50 *T. pallidum* proteins remained undetected, which corresponds to 5% of the proteome (Table S7). This elusive protein category continued to exhibit a bias toward proteins of unknown function (66% annotated as hypothetical proteins or DUF domain-containing proteins) (Table S7). In agreement, 74% of the 50 undetected proteins could not be assigned functional categories or were categorized as "function unknown" using COG analysis[45,47] (Figure 6). Eight of the 50 undetected proteins corresponded to open reading frames (ORFs) that are annotated in the NCBI proteome as "pseudo" genes (noncoding ORFs). The number of positively charged miniproteins in this class of proteins also remained high at 62% (average size, 93 amino acids). Since *T. pallidum* is an obligate human pathogen, with one of the smallest genomes and minimal metabolic capabilities,[12,57] it is unlikely that the pathogen has retained ORFs that are not functionally relevant or remain unexpressed. Further, since miniproteins often have important biological functions,[58] it is possible that the majority or all of the 50 undetected proteins are expressed by *T. pallidum* and that MS-based detection is unsuccessful for the reasons outlined above.

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

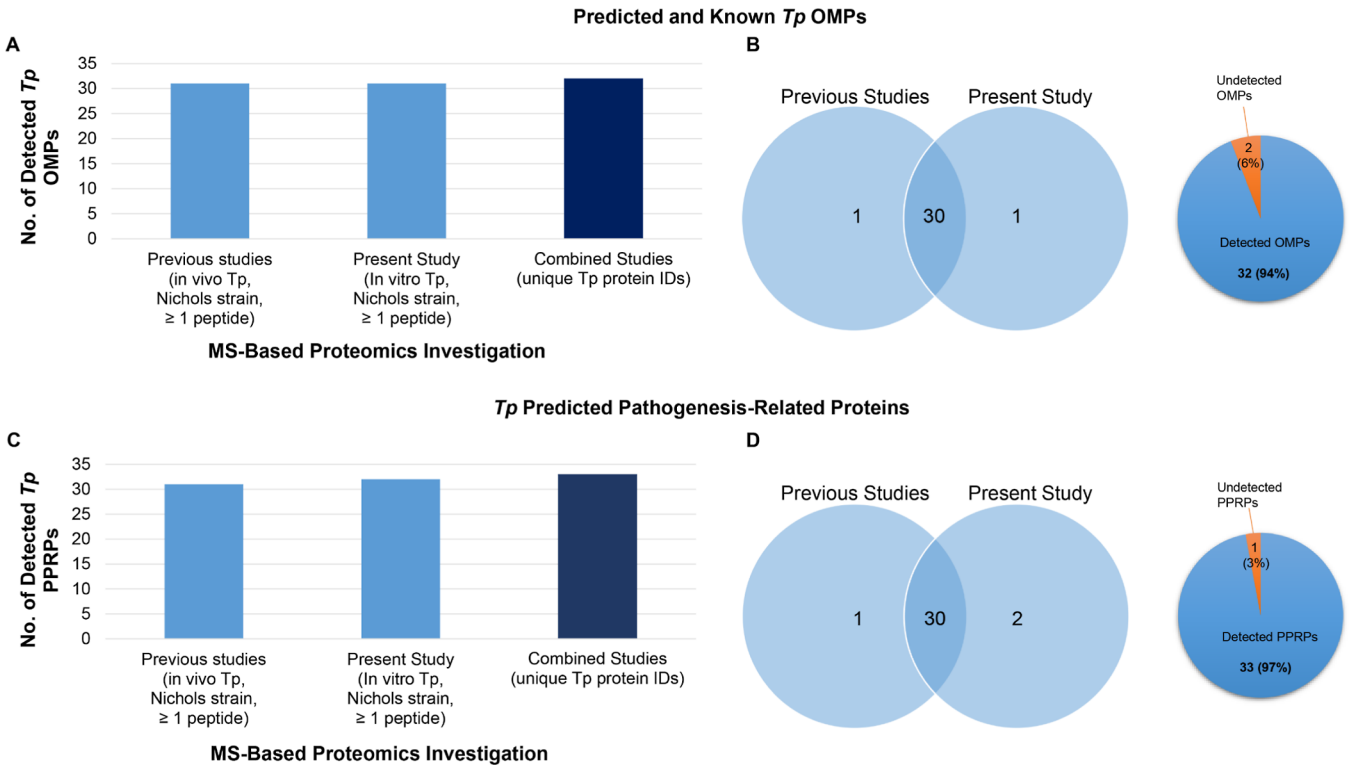


Figure 8. Proteome coverage of known/predicted OMPs and putative pathogenesis-related proteins from *T. pallidum*. Bar graphs depicting the number of (A) known/predicted OMPs and (C) putative pathogenesis-related proteins (PPRPs) detected in the three previous combined in vivo studies and the present in vitro study (light blue bars) and the combined number detected in all studies (dark blue bars). Venn diagrams summarizing the number of shared and exclusive identifications corresponding to (B) known/predicted OMPs and (D) PPRPs detected in previous and present MS-based proteomics studies. Pie charts show (B) the combined numbers and percentages of detected and undetected proteins of known and predicted OMPs and (D) PPRPs.

Quantitation-Based Comparative Analyses of In Vivo-Grown and In Vitro-Cultured *T. pallidum*

Label-free quantification (LFQ) based on peptide ion peak intensities (MaxLFQ intensity values) was used to determine the relative abundance of proteins detected in each of the three in vitro-cultured *T. pallidum* samples. This approach allowed for the relative quantification of between 600 and 799 proteins in each of the samples (Table S8). A list of the high-abundance proteins (all proteins with MaxLFQ intensity values greater than the mean average) in each of the three in vitro-cultured treponemal samples is provided in Table S9. In general, there was high similarity in relative protein abundances between in vivo-grown[25] and in vitro-cultured *T. pallidum,* as highlighted by the following findings: (1) in both growth conditions, most high-abundance proteins were assigned functions related to metabolism, homeostasis, chemotaxis and motility, and protein translation (Tables S8 and S9, Figure S1 and ref 25); (2) less than 10% of high-abundance proteins from in vivo- and in vitro-cultured treponemes were annotated in the proteome as proteins of unknown function ("hypothetical proteins" and DUF domain-containing proteins) (Table S9 and ref 25); (3) 75% of all high-abundance proteins were found in both in vitro- (mean average of three samples) and in vivo-grown *T. pallidum*[25] (Table S9); and (4) 36 of the top 50 highest abundance-proteins were found in both in vitro- and in vivo-grown *T. pallidum*[25] (Table S10 and Figure S2).

A previous global proteomics study performed within our laboratory provided insights into the protein expression profiles of three specific *T. pallidum* protein classes of interest from in vivo-grown treponemes; (1) proteins of unknown function (including miniproteins), (2) putative pathogenesis-related proteins, and (3) known/predicted OMPs.[25] The high proteome coverage of in vitro-cultured treponemes in the present study and consistency in experimental methods between the two studies enabled in vivo/in vitro quantitation-based comparative analyses of the three protein classes. In the present study, we detected 230 of the 268 proteins of unknown function from in vitro-cultured treponemes (Figure 7A,B, Tables 1 and S11). When combined with previous studies,[25−27] 88% of these functionally unannotated proteins have now been detected in *T. pallidum* (Figure 7A,B, Tables 1 and S11). However, only a limited number of proteins from this category were identified as high-abundance proteins from either growth condition (Table S9 and ref 25). Sixty-eight miniproteins, two of which have been shown to exhibit antimicrobial peptide (AMP) activities,[49] comprise a subgroup of the treponemal proteins of unknown function, which comprise 150 amino acids or less. In the current study, we were able to detect 39 of the 68 miniproteins of unknown function, which, when combined with the previous proteomics studies,[25−27] enabled the identification of 62% of this protein subgroup (Figure 7C,D, Tables 1 and S12). Although we were able to detect more miniproteins in the present study compared to the in vivo study,[25] only the same three miniproteins were detected as high-abundance proteins in both studies: TPANIC_0084, TPANIC_0777, and the predicted AMP, TPANIC_0847.[49]

In our preceding *T. pallidum* global proteomics study, we compiled a list of 34 proteins, identified from the literature,[25] which comprised experimentally confirmed and predicted *T.*

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Table 2. MS-Based Detection of Predicted/Known OMPs from *T. pallidum*

OMP locus tag	NCBI functional annotation	MS detection	References
TPANIC_0009	hypothetical protein, TprA	ND	52
TPANIC_0011	major outer sheath C-terminal domain-containing protein, TprB	present, Osbak[27]	27, 52, and 64
TPANIC_0117[a]	major outer sheath N-terminal domain-containing protein, TprC	present, Houston,[25] Osbak[27]	27, 52, 60, and 64
TPANIC_0126	hypothetical protein	present, Houston,[25] Osbak[27]	16, 27, 66, 68, and 71
TPANIC_0131[a]	major outer sheath N-terminal domain-containing protein, TprD	present, Houston,[25] Osbak[27]	27, 52, 60, and 64
TPANIC_0155	M23 family metallopeptidase	present, Houston,[25] Osbak[27]	27, 51, 52, and 61
TPANIC_0313	major outer sheath N-terminal domain-containing protein, TprE	present, Houston,[25] Osbak[27]	27 and 52
TPANIC_0316	hypothetical protein, TprF	ND	51 and 52
TPANIC_0324/325	translocation/assembly module TamB domain-containing protein	present, Osbak[27]	27, 52, and 68
TPANIC_0326[a]	outer membrane protein assembly factor BamA	present, Houston,[25] Osbak[27]	27, 51, 52, 62, 65, 68, and 71
TPANIC_0421	tetratricopeptide repeat protein	present, Houston,[25] Osbak[27]	27, 52, and 71
TPANIC_0479	DUF2715 domain-containing protein	present, Houston[25]	68
TPANIC_0483	fibronectin type III domain-containing protein	present, Houston[25]	51
TPANIC_0515	LPS-assembly protein LptD	present, Houston[25]	16, 68, and 71
TPANIC_0548	UPF0164 family protein	present, Houston[25]	16, 52, 68, and 71
TPANIC_0557	DUF1007 family protein	present, Houston[25]	51
TPANIC_0620	major outer sheath N-terminal domain-containing protein, TprI	present, Houston,[25] Osbak[27]	27, 51, 52, 59, 64, and 67
TPANIC_0621	major outer sheath N-terminal domain-containing protein, TprJ	present, Houston,[25] Osbak[27]	27 and 52
TPANIC_0698	DUF2715 domain-containing protein	present	68
TPANIC_0733	outer membrane β-barrel protein	present, Houston[25]	16, 68, and 71
TPANIC_0751[a]	vascular adhesin/metalloprotease pallilysin	present, Houston[25]	51, 70, and 72
TPANIC_0855	hypothetical protein	present, Houston,[25] Osbak[27]	27 and 52
TPANIC_0856	UPF0164 family protein	present, Houston,[25] Osbak[27]	16, 51, 68, and 71
TPANIC_0858	UPF0164 family protein	present, Houston,[25] Osbak[27]	16, 27, 52, 68, and 71
TPANIC_0859	UPF0164 family protein	present, Osbak	16, 68, and 71
TPANIC_0865	UPF0164 family protein	present, Houston,[25] Osbak[27]	16, 27, 52, 68, and 71
TPANIC_0897[a]	MSP porin, TprK	present, Houston[25]	52 and 63
TPANIC_0923	PEGA domain-containing protein	Osbak[27]	27
TPANIC_0952	α/β fold hydrolase	present, Houston[25]	51
TPANIC_0966	hypothetical protein, TolC	present, Houston[25]	16, 68, and 71
TPANIC_0967	hypothetical protein, TolC	present, Houston,[25] Osbak[27]	16, 68, and 71
TPANIC_0968	hypothetical protein, TolC	present, Houston,[25] Osbak[27]	68 and 71
TPANIC_0969	hypothetical protein, TolC	present, Houston,[25] Osbak,[27] McGill[26]	16, 27, 52, 68, and 71
TPANIC_1031	major outer sheath N-terminal domain-containing protein, TprL	present, Houston[25]	52 and 69

[a]Proteins with experimental evidence indicating *T. pallidum* surface exposure. ND: protein not detected. McGill: Proteins detected in McGill et al. study (2010). Osbak: Proteins detected in Osbak et al. study (2016). Houston: Proteins detected in Houston et al. study (2023). Present: Proteins detected in the present study (Present: protein identification based on the detection of one tryptic peptide). Bold font: Protein from *T. pallidum* identified exclusively in the present study. In addition to the NCBI functional annotations, Tpr family member names are also included where applicable.

pallidum surface-exposed OMPs.[16,26,27,51,52,59−72] In the present study, we were able to detect the protein expression of 31 known and predicted treponemal OMPs, which increased the combined OMP repertoire coverage to 94% (Figure 8A,B, and Tables 1 and 2). Similar to in vivo-grown *T. pallidum* where none of the 34 predicted/known OMPs were identified as high-abundance proteins,[25] only one member of this protein class (TPANIC_0858; relative abundance ranking = 47/600) was assigned an LFQ intensity value higher than the mean average in one of the three in vitro-cultured samples (high-abundance protein) (Table 9). The LFQ rankings for known/potential OMPs from in vitro-cultured treponemes ranged from (1) 47−568 out of 600 (sample 1); (2) 134−770 out of 799 (sample 2); and (3) 199−778 out of 799 (sample 3) (Table S8 and Figure S3). Notably, when the mean averages of the three samples were analyzed, 40% of all quantitated known/predicted OMPs from in vitro-cultured *T. pallidum* were listed in the top 50% of LFQ-ranked proteins (Table S8). However, only 22% of all quantitated known/predicted OMPs from in vivo-grown *T. pallidum* were found in the top 50%,[25] suggesting that this class of proteins have generally higher relative abundances in treponemes isolated from in vitro cultures. Furthermore, from an average of 733 LFQ-quantitated proteins, an average of 21 known/predicted OMPs were successfully quantitated in the three in vitro-cultured samples (Table S8). In contrast, an average of only 7 known/predicted OMPs from in vivo-grown treponemes were quantitated out of an average of 450 proteins in the three in vitro-cultured samples (Table S8). This latter finding further suggests that relative expression levels of known/predicted OMPs may be higher during in vitro culture as lower-abundance proteins are less amenable to semiquantification via LFQ.[73] However, a limitation of the present study is that the MS data used for LFQ analyses of in vivo-grown[25] and in vitro-cultured *T. pallidum* (present study) were acquired several months apart. Therefore, we were unable to analyze the samples from the different growth conditions in a single combined LFQ experiment due to technical and MS instrumental variation, which would reduce the reliability of the analyses. Consequently, in the absence of detailed statistical analyses, our investigations

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Table 3. MS-Based Detection of Predicted Pathogenesis-Related Proteins from *T. pallidum*

Locus tag	Functional annotation (Nichols strain NC_021490, July 2021 annotation)	Potential functions from previous studies[a,b,c]	MS detection
TPANIC_0020[c]	VWA domain-containing protein	TgMIC2 (*Toxoplasma gondii* micronemal protein 2 A/I domain)	present, Houston,[25] Osbak[27]
TPANIC_0027[a]	hemolysin family protein	putative hemolysin (HlyC)	Osbak[27]
TPANIC_0028[a]	hemolysin family protein	putative hemolysin (HlyC)	present
TPANIC_0126[c]	hypothetical protein	outer membrane protein W (*Escherichia coli*)	present, Houston,[25] Osbak[27]
TPANIC_0134[c]	hypothetical protein	bacterial sialidases/neuraminidases	present, Houston[25]
TPANIC_0225[c]	leucine-rich repeat domain-containing protein	leucine-rich repeat surface proteins	present, Houston[25]
TPANIC_0246[c]	VWA domain-containing protein	TRAP protein (*Plasmodium vivax*)	present, Osbak[27]
TPANIC_0262[b]	Crp/Fnr family transcriptional regulator	PrfA (*Listeria monocytogenes* virulence factor transcriptional regulator)	present, Houston,[25] Osbak[27]
TPANIC_0399[a]	flagellar M-ring protein FliF	type 3 (virulence-related) secretory pathway protein (FliF)	present, Houston,[25] Osbak[27]
TPANIC_0401[a]	flagellar assembly protein FliH	type 3 (virulence-related) secretory pathway protein (FliH)	present, Houston,[25] Osbak[27]
TPANIC_0402[a]	flagellar protein export ATPase FliI	type 3 (virulence-related) secretory pathway protein (FliL)	present, Houston,[25] Osbak[27]
TPANIC_0421[c]	tetratricopeptide repeat protein	PknD (*Mycobacterium tuberculosis* serine/threonine protein kinase, extracellular domain)	present, Houston,[25] Osbak[27]
TPANIC_0544[c]	SpnA family nuclease	SmcL (*Listeria ivanovii* sphingomyelinase-C)	present, Houston,[25] Osbak[27]
TPANIC_0579[c]	hypothetical protein	YenC2 (*Yersinia entomophaga* ABC toxin; BC component)	present, Osbak[27]
TPANIC_0594[c]	DUF2147 domain-containing protein	HP1028 (*Helicobacter pylori* lipocalin)	present, Houston[25]
TPANIC_0598[c]	hypothetical protein	BamB (*Moraxella catarrhalis* β barrel assembly machinery protein B)	present, Houston,[25] Osbak[27]
TPANIC_0625[c]	hypothetical protein	BamD (β barrel assembly machinery protein *Rhodothermus marinus*, *E. coli*)	present, Houston,[25] Osbak[27]
TPANIC_0649[a]	hemolysin family protein	putative hemolysin (TlyC)	present, Houston,[25] Osbak[27]
TPANIC_0714[a]	flagellar biosynthesis protein FlhA	type 3 (virulence-related) secretory pathway protein (FlhA)	present, Houston,[25] Osbak[27]
TPANIC_0715[a]	flagellar biosynthesis protein FlhB	type 3 (virulence-related) secretory pathway protein (FlhB)	present, Houston,[25] Osbak[27]
TPANIC_0733[c]	outer membrane β-barrel protein	NspA (*Neisseria* surface protein A)	present, Houston[25]
TPANIC_0783[c]	hypothetical protein	BamB (*E. coli* β barrel assembly machinery protein B)	present, Houston[25]
TPANIC_0789[c]	outer membrane lipoprotein-sorting protein	LolA (*Pseudomonas aeruginosa* outer membrane lipoprotein carrier/localization protein)	present, Houston,[25] Osbak,[27] McGill[26]
TPANIC_0854[c]	SpoIIE family protein phosphatase	bacterial sialidases/neuraminidases	present, Houston,[25] Osbak[27]
TPANIC_0862[b]	FKBP-type peptidyl-prolyl cis−trans isomerase	Mip (*Legionella pneumophila* macrophage infectivity potentiator protein)	present, Houston,[25] McGill[26]
TPANIC_0911[c]	FlhB-like flagellar biosynthesis protein	EscU (*E. coli* type 3 secretion system protein)	ND
TPANIC_0928[c]	hypothetical protein (previous proteome annotation; not annotated in July 2021)	SurA (*E. coli* chaperone)	present, Houston,[25] Osbak[27]
TPANIC_0936[a]	hemolysin family protein	putative hemolysin	present
TPANIC_0966[c]	hypothetical protein	TolC (*E. coli* outer membrane channel protein)	present, Houston[25]
TPANIC_0967[c]	hypothetical protein	TolC (*E. coli* outer membrane channel protein)	present, Houston,[25] Osbak[27]
TPANIC_0968[c]	hypothetical protein	TolC (*E. coli* outer membrane channel protein)	present, Houston,[25] Osbak[27]
TPANIC_0969[c]	hypothetical protein	TolC (*E. coli* outer membrane channel protein)	present, Houston,[25] Osbak,[27] McGill[26]
TPANIC_1033[b]	patatin-like phospholipase family protein	VipD (*L. pneumophila* phospholipase effector protein)	present, Houston,[25] Osbak[27]
TPANIC_1037[a]	hemolysin III family protein	putative hemolysin III (HlyIII)	present, Houston[25]

[a]Proteins identified as potential virulence factors in *T. pallidum* genome sequencing and comparison studies.[74−76] [b]Proteins identified as novel virulence factor candidates (proteins previously annotated with non-virulence-related functions) using whole proteome structure modeling (highest-ranking predicted virulence-related structural homologue listed).[71] [c]Proteins identified as novel virulence factor candidates (proteins of unknown function) using whole proteome structure modeling (highest-ranking predicted virulence-related structural homologue listed).[71] ND: Protein not detected. McGill: Proteins detected in McGill et al. study (2010). Osbak: Proteins detected in Osbak et al. study (2016). Houston: Proteins detected in Houston et al. study (2023). Present: Proteins detected in the present study (Present: protein identification based on the detection of one tryptic peptide). Bold font: Protein from *T. pallidum* identified exclusively in the present study.

were limited to relative quantification comparisons based solely on LFQ rankings.

Thirty-four *T. pallidum* proteins have previously been identified as potential pathogenesis-related proteins (PPRPs)

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Table 4. Proteome Annotation Errors Identified from In Vitro-Cultured *T. pallidum*

Locus tag	Functional annotation	NCBI proteome annotation errors (in Nichols strain NC_021490, July 2021 annotation)
TPANIC_0126a (in vitro)	hypothetical protein	protein deleted from proteome
TPANIC_0415 (in vitro)	hypothetical protein	protein deleted from proteome
TPANIC_0927[a] (in vitro)	hypothetical protein	protein deleted from proteome
TPANIC_0928[a]	hypothetical protein	protein deleted from proteome
TPANIC_RS02075[a]	hypothetical protein	protein deleted from proteome
TPANIC_RS04705[a]	hypothetical protein	protein deleted from proteome
TPANIC_RS05180 (in vitro)	hypothetical protein	protein deleted from proteome
TPANIC_0087 (in vitro)	DUF192 domain-containing protein	different amino acid sequence (N-terminus)
TPANIC_0246 (in vitro)	VWA domain-containing protein	incorrectly truncated N-terminus
TPANIC_0446	(*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	incorrectly truncated N-terminus
TPANIC_0535	hypothetical protein	incorrectly truncated N-terminus
TPANIC_0672 (in vitro)	glycine-tRNA ligase	incorrectly truncated N-terminus
TPANIC_0765[b]	ATP-dependent zinc metalloprotease FtsH	incorrectly truncated N-terminus
TPANIC_0907 (in vitro)	16S rRNA processing protein/ribosome maturation factor RimM	incorrectly truncated N-terminus
TPANIC_0953 (in vitro)	TraB/GumN family protein	incorrectly truncated N-terminus
TPANIC_1023 (in vitro)	recombination regulator RecX	incorrectly truncated N-terminus
TPANIC_0007	DUF3798 domain-containing protein	"pseudo" (noncoding annotation)
TPANIC_RS01255[c]	hypothetical protein	"pseudo" (noncoding annotation)
TPANIC_0533[c]	V-type ATP synthase subunit I	"pseudo" (noncoding annotation)
TPANIC_0813[c]	hypothetical	"pseudo" (noncoding annotation)
TPANIC_0897	MSP porin (TprK)	"pseudo" (noncoding annotation)
TPANIC_0993[c]	septal ring lytic transglycosylase RlpA family protein	"pseudo" (noncoding annotation)
TPANIC_RS05465/ TPANIC_0126d[a,d] (in vitro)	hypothetical	"pseudo" (noncoding annotation)
	hypothetical/major outer sheath C-terminal domain-containing	protein deleted from proteome

[a]Proteins have been re-added in the Nichols strain (March 2023 Nichols NC_021490 annotation [TPANIC_RS04705 locus tag corresponds to TPANIC_RS05630 in the March 2023 annotation which has an incorrectly truncated N-terminus based on the results from a previous study[25]]). [b]Protein has been reannotated with the correct N-terminus (March 2023 Nichols NC_021490 annotation). [c]Proteins have been reannotated as coding proteins (March 2023 Nichols NC_021490 annotation). [d]Identified protein (TPANIC_RS05465) may be TPANIC_0126d (hypothetical/major outer sheath C-terminal domain-containing protein—contains all identified peptides and, unlike Tp_RS05465, contains a start methionine). (In vitro): 11 proteome annotation errors identified from in vitro *T. pallidum* only.

based on proteome-wide tertiary structure modeling and genome sequencing.[71,74−76] In the current study, we detected the expression of 32 of these PPRPs, which increased the proteome coverage of this protein class to 97% (Figure 8C,D and Tables 1 and 3). In agreement with the findings from in vivo-grown *T. pallidum*,[25] most PPRPs from in vitro-cultured treponemes were also found to be of low relative abundance. Although we detected the in vitro expression of 32/34 PPRPs (Table 3), only four members of this group were identified as high-abundance proteins; TPANIC_0020 (structural homologue of *Toxoplasma gondii* micronemal protein 2, MIC2[71]), TPANIC_0225 (structural homologue of leucine-rich repeat surface proteins[71]), TPANIC_0789 (structural homologue of outer membrane lipoprotein carrier/localization protein, LolA[71]), and TPANIC_0862 (structural homologue of macrophage infectivity potentiator, MIP[71]) (Table S9). The correlation between in vivo and in vitro PPRP expression is further highlighted by the finding that only TPANIC_0225, TPANIC_0789, and TPANIC_0862 were identified as high-abundance proteins from in vivo-grown *T. pallidum*.[25] These findings further suggest that a low expression level of treponemal proteins that are important for *T. pallidum* pathogenesis (PPRPs) and/or exposed to the host (OMPs) is a host detection avoidance strategy that may contribute to the well-documented persistent nature of *T. pallidum*.[10,11]

Overall, our comparative analyses show that the protein expression profiles of treponemes from both in vivo and in vitro cultures are similar. These findings align with previous studies that showed a correlation in global gene expression profiles between in vivo/in vitro *T. pallidum*, as well as the retention of general treponemal characteristics related to morphology, growth, metabolism, and infectivity in *T. pallidum* from in vitro cultures.[24,35−37] As would be expected, specific differences in gene expression at the RNA level[24] and potentially at the protein level (current study) can occur in the two growth conditions. But overall, the observations in the current study reinforce the finding that in vitro-cultured *T. pallidum* broadly emulates in vivo-grown treponemes.[35]

Identification of Known and Potential Novel Annotation Errors in the *T. pallidum* Proteome

The recent *T. pallidum* global protein expression profiling study performed within our laboratory reported the identification of several annotation errors in the proteome of in vivo-grown *T. pallidum*.[25] In that study, we used a customized database (containing all *T. pallidum* proteins that had been annotated in the NCBI proteome revisions and updates from June 2013 to July 2021) for MS peptide/protein searches and identifications. This approach allowed for the identification of 12 proteome annotation errors; three proteins that were incorrectly deleted from the proteome ("deletion errors"), three proteins that were

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

annotated with prematurely truncated N-termini ("N-terminal errors"), and six proteins for which we detected expression but whose ORFs had been annotated as "pseudo" genes/noncoding ORFs ("pseudo errors").[25] In the present study, using the same customized database, we identified the same 12 proteome annotation errors as were detected with in vivo-grown *T. pallidum*, reinforcing the accuracy of these findings (Tables 4 and S13). We also identified 11 potential additional proteome annotation errors from in vitro treponemes. These included four "deletion errors", six "N-terminal errors" (including one, TPANIC_0087, for which we detected an N-terminal peptide with a different sequence), and one "pseudo error", with most of the errors found in proteins of unknown function (Tables 4 and S13, Figure S4). These findings translate into a potential proteome annotation error rate of 2.3%.

Notably, one detected N-terminal error results in what is expected to be a different starting site for the protein. In this situation, the original N-terminal sequence, as annotated in the genome, is MLARIESVLYKAFEEHVSFYLHR. Our mass spectrometry analyses detected an altered amino acid at amino acid position 9, with a switch of a leucine to a methionine, resulting in a detected peptide of MYKAFEEHVSFYLHR. Interestingly, if this alternative start site is correct, this alters the signal peptide type predicted for this protein using SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/)[77] from a cleavable Sec/SPI signal sequence to a Sec/SPII (lipoprotein) signal sequence, suggesting that this protein could have a different subcellular locale within the bacterium. Possible functional consequences of this expected difference in subcellular localization for this protein must await further studies since TPANIC_0087 is one of the treponemal proteins of unknown function (DUF192 domain-containing protein).

A limitation of these findings is the possibility that the remaining nine potential proteome errors identified in this study but not detected from in vivo-grown treponemes[25] are due to potential "artificial" conditions introduced during in vitro treponeme culture and/or possible post-translational modifications such as N-terminal proteolytic processing. However, genome sequencing of an in vitro-cultured *T. pallidum* Nichols strain revealed that in vivo-grown and in vitro-cultured *T. pallidum* genome sequences are highly similar[78] and any observed sequence differences are unlikely to account for the 11 novel genome errors identified in the present in vitro study. Another limitation is that the Nichols strain used in our previous[25] and present study was originally isolated in 1912 from the cerebrospinal fluid (CSF) of a secondary syphilis patient.[79] Due to the fact that this laboratory reference strain has been passaged continuously in rabbits since 1912, future MS-based global protein expression profile investigations using recently isolated clinical strains will be required to determine the relevance of our findings to expression profiles from clinically relevant, low-passage treponemal strains. Currently, in vitro growth of clinical *T. pallidum* strains is in its infancy. However, as clinical strains become more amenable to routine in vitro culturing, future comparative proteomics studies of clinical strains will provide useful information.

Comparison of the Top 50 Highest-Abundance *T. pallidum* Proteins and Expressed Genes from In Vitro-Grown and In Vivo-Cultured *T. pallidum*

In the comparative transcriptomics study of in vivo-grown and in vitro-cultured *T. pallidum*, the authors identified the 50 genes with the highest transcript expression levels in both growth conditions.[24] Here, we investigated potential correlations between *T. pallidum* protein and transcript abundances by comparing the top 50 highest-abundance proteins from in vivo-grown[25] and in vitro-cultured treponemes (current study) with the most highly expressed genes reported in the RNA-seq study.[24] Fourteen transcript/protein pairs from in vivo-grown *T. pallidum* and 16 transcript/protein pairs from in vitro-cultured *T. pallidum* were among the top 50 most abundant molecules in both the RNA-seq and proteomics data (Figure 9A,B and Table S14: "top 50 in vivo shared" and "top 50 in vitro shared"). In addition, 10 transcripts/proteins were found to be present in the top 50 most abundant transcripts/proteins of both in vivo-grown and in vitro-cultured treponemes for both RNA-seq and proteomics (Table S14: "10 found in all"). A functional comparison (via the assignment of COG categories) of the top 50 highest-abundance transcripts/proteins from in vivo-grown[24] and in vitro-cultured *T. pallidum* showed an overall similarity between the four groups with unknown/unassigned functions being the largest category in each group (Figure 9C).

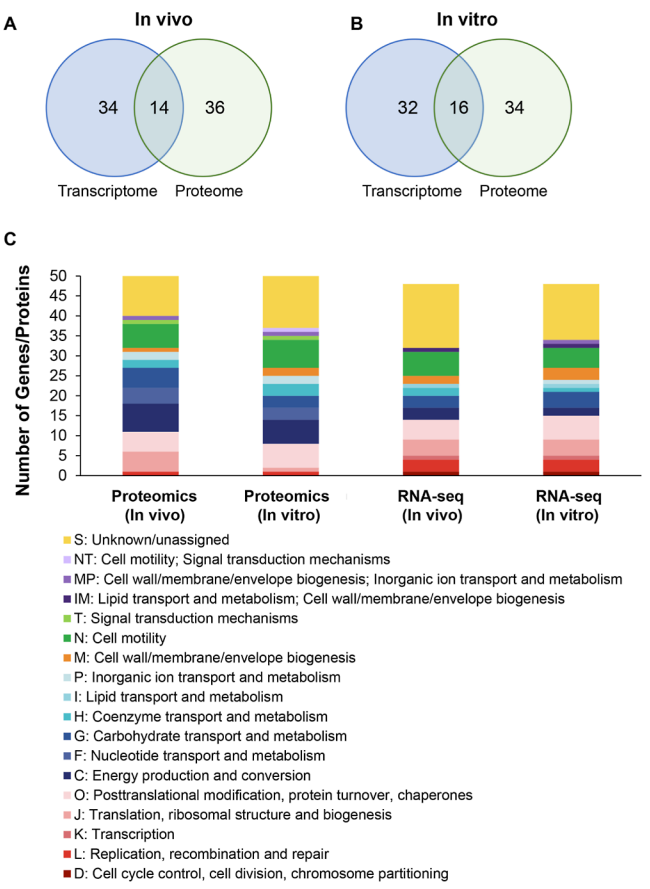


Figure 9. Comparison of the highest expressed *T. pallidum* genes and proteins from in vivo-grown and in vitro-cultured treponemes. Venn diagrams summarizing the number of shared and unique high-abundance transcripts[24] and proteins (ref 25 and present study) from (A) in vivo-grown *T. pallidum* and (B) in vitro-cultured *T. pallidum*. (C) Bar graph showing the COG functional categories assigned to the top 50 high-abundance transcripts[24] and proteins (ref 25 and present study) for in vivo-grown and in vitro-cultured *T. pallidum*. Transcripts/proteins with multiple COG categories are assigned separate categories. Two of the 50 highly expressed genes from in vivo-grown and in vitro-cultured treponemes were not protein coding[24] and were excluded from the comparisons.

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

These results agree with previous studies that have found low correlations between the transcript and protein levels in complex biological samples, including in *T. pallidum*.[24,27,80]

CONCLUSIONS

T. pallidum omics studies have mostly focused on the genomic, transcriptomic, and proteomic analyses of in vivo-grown *T. pallidum*.[12,23−27,74−76,81,82] As a result of the recent development of a long-term treponemal in vitro culture system,[35,37] in the current study we were able to report an in-depth global proteomic analysis of in vitro-cultured *T. pallidum*. By using a previously developed sample preparation protocol tailored for MS analysis of the *T. pallidum* proteome,[25] in combination with high-resolution MS, we were able to detect the expression of 94% of proteins from in vitro-cultured *T. pallidum*. This result provides a comprehensive overview of the *T. pallidum* protein expression profile by extending the total combined proteome coverage of in vivo- and in vitro-grown treponemes to 95% and provides the first in-depth global protein expression and relative quantification profiles of in vitro-cultured *T. pallidum*. In agreement with a previous in vivo/in vitro global comparative transcriptomics study,[24] our studies showed an overall similarity between in vivo- and in vitro-grown *T. pallidum*. Although known and predicted OMPs were found to be generally at a low abundance in both treponemal growth conditions, our results suggest that the relative abundances of this treponemal protein class may be higher in treponemes isolated from in vitro cultures. However, further protein quantification studies that simultaneously analyze treponemes from both growth conditions are required to determine the validity of these initial OMP expression findings. It is also important to emphasize that any differences in protein expression profiles between *T. pallidum* grown in vivo and in vitro must be considered within the constraints of mass spectrometry analyses, as described above. Future drug discovery and vaccine design strategies focused on proteins detected only under in vitro conditions will necessitate further investigation to confirm their expression during infection.

In vitro-cultured *T. pallidum* generally exhibit the same morphological, motility, and infectivity characteristics as in vivo-grown treponemes.[35−37] Combined with the proteomics analyses presented in the present study, collectively, these observations support the finding that in vitro-cultured *T. pallidum* broadly emulates in vivo-grown treponemes.[35] The overall similarities described here, together with the ethical benefits of reducing the use of rabbits for experimentation, the reduction in expense, time, and labor associated with in vivo growth of *T. pallidum*, and the reduction in host contaminants that may serve as confounders in experimentation with in vivo-grown *T. pallidum*, reinforce the concept that the *T. pallidum* in vitro culture system is an appropriate replacement for in vivo propagation of *T. pallidum*.

ASSOCIATED CONTENT

Data Availability Statement

The MS proteomics data (raw data files and Scaffold search engine files) have been deposited to the MassIVE repository (https://massive.ucsd.edu/ProteoSAFe/private-dataset.jsp?task=fbf1937f5e784187b6cf655be16a592a) with the data set identifier MSV000093603.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.3c00891.

- Functional classification of high-abundance *T. pallidum* proteins, comparative analyses of the top 50 highest-abundance proteins from in vivo-grown and in vitro-cultured *T. pallidum*, relative abundances of in vitro-cultured *T. pallidum* OMPs, and *T. pallidum* proteome annotation errors (PDF)
- Customized database (Tp_CustomDB) for mass spectrometry searches (XLSX)
- Master sheet containing detailed mass spectrometry data for the three in vitro-cultured *T. pallidum* samples (XLSX)
- *T. pallidum* protein identifications and the number of in vitro-cultured biological replicate samples containing each treponemal protein (XLSX)
- All *T. pallidum* proteins detected in each of the three in vitro-cultured *T. pallidum* samples (XLSX)
- Sixty-two *T. pallidum* proteins that were not detected in the present study (XLSX)
- *T. pallidum* proteins (Nichols Strain) detected in the present study and three previous mass spectrometry-based proteomics studies (XLSX)
- Fifty *T. pallidum* proteins that were not detected in the present or past *T. pallidum* global proteomics studies (XLSX)
- Relative abundances of proteins detected from in vitro-cultured *T. pallidum* samples 1−3 (XLSX)
- High-abundance proteins with LFQ_ intensity values (relative abundances) greater than the mean average in each of the three in vitro-cultured *T. pallidum* samples (XLSX)
- Comparison of the top 50 highest-abundance proteins from in vitro-cultured and in vivo-grown *T. pallidum* (XLSX)
- Proteins of unknown function identified in the present and previous *T. pallidum* mass spectrometry-based proteomics studies (XLSX)
- Miniproteins of unknown function identified in the present and past *T. pallidum* mass spectrometry-based proteomics studies (XLSX)
- NCBI annotation errors identified in the proteome of *T. pallidum* (PDF)
- Comparison of the highest expressed *T. pallidum* genes and proteins (XLSX)

AUTHOR INFORMATION

Corresponding Author

Caroline E. Cameron − *Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada; Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0002-8786-4359; Email: caroc@uvic.ca

Authors

Simon Houston − *Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada*

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

Alloysius Gomez − *Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada*
Andrew Geppert − *Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada*
Mara C. Goodyear − *Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.3c00891

Author Contributions

S.H., C.E.C., and A. Gomez designed the study. S.H. and A. Gomez performed the experiments. S.H., A. Geppert, M.C.G., and C.E.C. analyzed and interpreted the data. C.E.C. acquired financial support for the work. S.H. wrote the manuscript with contributions from C.E.C. and M.C.G. All authors reviewed the manuscript before submission for accuracy and intellectual content and approved the submitted version.

Funding

This work was supported by grants R37AI051334 and U19AI144133 (CEC) from the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH), as well as an award from Open Philanthropy (52345). A. Geppert is the recipient of a Canada Graduate Scholarship-Masters (CGS-M), and M.C.G. is the recipient of a Developmental Research Project Award from the parent U19AI144133 award.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the University of Victoria-Genome BC Proteomics Centre for providing MS analytical services. We also acknowledge the significant generosity and assistance provided by Drs. Steven Norris and Diane Edmondson, University of Texas Health Science Center at Houston, to establish the in vitro *T. pallidum* culture system within our laboratory and for kindly providing the Sf1Ep cottontail rabbit epithelial cells. To our knowledge, we were the first laboratory to reproduce the Norris/Edmondson in vitro culture system for *T. pallidum*, a feat that was only possible because of Dr. Norris and Dr. Edmondson's generosity with time, reagents, advice, and expertise. Their groundbreaking advancement in *T. pallidum* growth has changed the course of research performed within our laboratory and has made experiments that were previously unachievable possible.

■ REFERENCES

(1) World Health Organization. *Global Incidence and Prevalence of Selected Curable Sexually Transmitted Infections—2008*, 2012. https://apps.who.int/iris/bitstream/handle/10665/75181/9789241503839_eng.pdf;jsessionid=990C4C535FC5E3BA233AFB567DE91D43?sequence=1.
(2) Herbert, L. J.; Middleton, S. I. An estimate of syphilis incidence in Eastern Europe. *J. Glob. Health* **2012**, *2*, 010402.
(3) Savage, E. J.; Marsh, K.; Duffell, S.; Ison, C. A.; Zaman, A.; Hughes, G. Rapid increase in gonorrhoea and syphilis diagnoses in England in 2011. *Euro Surveill.* **2012**, *17*, pii=20224.
(4) Chen, G.; Cao, Y.; Yao, Y.; Li, M.; Tang, W.; Li, J.; Babu, G. R.; Jia, Y.; Huan, X.; Xu, G.; Yang, H.; Fu, G.; Li, L. Syphilis incidence among men who have sex with men in China: results from a meta-analysis. *Int. J. STD AIDS* **2017**, *28*, 170−178.
(5) Korenromp, E. L.; Rowley, J.; Alonso, M.; Mello, M. B.; Wijesooriya, N. S.; Mahiane, S. G.; Ishikawa, N.; Le, L. V.; Newman-Owiredu, M.; Nagelkerke, N.; Newman, L.; Kamb, M.; Broutet, N.; Taylor, M. M. Global burden of maternal and congenital syphilis and associated adverse birth outcomes-Estimates for 2016 and progress since 2012. *PLoS One* **2019**, *14* (2), No. e0211720.
(6) Arrieta, A. C.; Singh, J. Congenital Syphilis. *N. Engl. J. Med.* **2019**, *381*, 2157.
(7) Nusbaum, M. R. H.; Wallace, R. R.; Slatt, L. M.; Kondrad, E. C. Sexually transmitted infections and increased risk of co-infection with human immunodeficiency virus. *J. Am. Osteopath. Assoc.* **2004**, *104*, 527−535.
(8) Douglas, J. M. Penicillin treatment of syphilis: clearing away the shadow on the land. *JAMA* **2009**, *301*, 769−771.
(9) LaFond, R. E.; Lukehart, S. A. Biological basis for syphilis. *Clin. Microbiol. Rev.* **2006**, *19* (1), 29−49.
(10) Radolf, J. D. *Treponema pallidum* and the quest for outer membrane proteins. *Mol. Microbiol.* **1995**, *16*, 1067−1073.
(11) Cameron, C. E. The *T. pallidum* outer membrane and outer membrane proteins. In *Pathogenic Treponema, Molecular and Cellular Biology*; Radolf, J. D., Lukehart, S. A., Eds.; Caister Academic Press: Norfolk, England, 2005; pp 237−266.
(12) Fraser, C. M.; Norris, S. J.; Weinstock, G. M.; White, O.; Sutton, G. G.; Dodson, R.; Gwinn, M.; Hickey, E. K.; Clayton, R.; Ketchum, K. A.; Sodergren, E.; Hardham, J. M.; McLeod, M. P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J. K.; Chidambaram, M.; Utterback, T.; McDonald, L.; Artiach, P.; Bowman, C.; Cotton, M. D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Sandusky, M.; Weidman, J.; Smith, H. O.; et al. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **1998**, *281*, 375−388.
(13) Hardy, P. H.; Levin, J. Lack of endotoxin in *Borrelia hispanica* and *Treponema pallidum*. *Proc. Soc. Exp. Biol. Med.* **1983**, *174*, 47−52.
(14) Liu, J.; Howell, J. K.; Bradley, S. D.; Zheng, Y.; Zhou, Z. H.; Norris, S. J. Cellular architecture of *Treponema pallidum*: novel flagellum, periplasmic cone, and cell envelope as revealed by cryo electron tomography. *J. Mol. Biol.* **2010**, *403*, 546−561.
(15) Izard, J.; Renken, C.; Hsieh, C. E.; Desrosiers, D. C.; Dunham-Ems, S.; La Vake, C.; Gebhardt, L. L.; Limberger, R. J.; Cox, D. L.; Marko, M.; Radolf, J. D. Cryo-electron tomography elucidates the molecular architecture of *Treponema pallidum*, the syphilis spirochete. *J. Bacteriol.* **2009**, *191*, 7566−7580.
(16) Radolf, J. D.; Kumar, S. The *Treponema pallidum* outer membrane. *Curr. Top. Microbiol. Immunol.* **2017**, *415*, 1−38.
(17) Centurion-Lara, A.; Godornes, C.; Castro, C.; Van Voorhis, W. C.; Lukehart, S. A. The *tprK* gene is heterogeneous among *Treponema pallidum* strains and has multiple alleles. *Infect. Immun.* **2000**, *68*, 824−831.
(18) Centurion-Lara, A.; LaFond, R. E.; Hevner, K.; Godornes, C.; Molini, B. J.; Van Voorhis, W. C.; Lukehart, S. A. Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. *Mol. Microbiol.* **2004**, *52*, 1579−1596.
(19) LaFond, R. E.; Centurion-Lara, A.; Godornes, C.; Rompalo, A. M.; Van Voorhis, W. C.; Lukehart, S. A. Sequence diversity of *Treponema pallidum* subsp. *pallidum* tprK in human syphilis lesions and rabbit-propagated isolates. *J. Bacteriol.* **2003**, *185*, 6262−6268.
(20) Giacani, L.; Godornes, C.; Puray-Chavez, M.; Guerra-Giraldez, C.; Tompa, M.; Lukehart, S. A.; Centurion-Lara, A. TP0262 is a modulator of promoter activity of tpr Subfamily II genes of *Treponema pallidum* ssp. *pallidum*. *Mol. Microbiol.* **2009**, *72*, 1087−1099.
(21) Giacani, L.; Lukehart, S.; Centurion-Lara, A. Length of guanosine homopolymeric repeats modulates promoter activity of subfamily II tpr genes of *Treponema pallidum* ssp. *pallidum*. *FEMS Immunol. Med. Microbiol.* **2007**, *51*, 289−301.
(22) Seshadri, R.; Myers, G. S.; Tettelin, H.; Eisen, J. A.; Heidelberg, J. F.; Dodson, R. J.; Davidsen, T. M.; DeBoy, R. T.; Fouts, D. E.; Haft, D.

H.; Selengut, J.; Ren, Q.; Brinkac, L. M.; Madupu, R.; Kolonay, J.; Durkin, S. A.; Daugherty, S. C.; Shetty, J.; Shvartsbeyn, A.; Gebregeorgis, E.; Geer, K.; Tsegaye, G.; Malek, J.; Ayodeji, B.; Shatsman, S.; McLeod, M. P.; Smajs, D.; Howell, J. K.; Pal, S.; Amin, A.; Vashisth, P.; McNeill, T. Z.; Xiang, Q.; Sodergren, E.; Baca, E.; Weinstock, G. M.; Norris, S. J.; Fraser, C. M.; Paulsen, I. T. Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5646−5651.

(23) Šmajs, D.; McKevitt, M.; Howell, J. K.; Norris, S. J.; Cai, W. W.; Palzkill, T.; Weinstock, G. M. Transcriptome of *Treponema pallidum*: gene expression profile during experimental rabbit infection. *J. Bacteriol.* **2005**, *187*, 1866−1874.

(24) De Lay, B. D.; Cameron, T. A.; De Lay, N. R.; Norris, S. J.; Edmondson, D. G. Comparison of transcriptional profiles of *Treponema pallidum* during experimental infection of rabbits and in vitro culture: Highly similar, yet different. *PLOS Pathog.* **2021**, *17* (9), No. e1009949.

(25) Houston, S.; Gomez, A.; Geppert, A.; Eshghi, A.; Smith, D. S.; Waugh, S.; Hardie, D. B.; Goodlett, D. R.; Cameron, C. E. Deep proteome coverage advances knowledge of *Treponema pallidum* protein expression profiles during infection. *Sci. Rep.* **2023**, *13*, 18259.

(26) McGill, M. A.; Edmondson, D. G.; Carroll, J. A.; Cook, R. G.; Orkiszewski, R. S.; Norris, S. J. Characterization and serologic analysis of the *Treponema pallidum* proteome. *Infect. Immun.* **2010**, *78*, 2631−2643.

(27) Osbak, K. K.; Houston, S.; Lithgow, K. V.; Meehan, C. J.; Strouhal, M.; Smajs, D.; Cameron, C. E.; Van Ostade, X.; Kenyon, C. R.; Van Raemdonck, G. A. Characterizing the syphilis-causing *Treponema pallidum* ssp. *pallidum* proteome using complementary mass spectrometry. *PLoS Neglected Trop. Dis.* **2016**, *10* (9), No. e0004988.

(28) Schaudinn, F.; Hoffman, E. Vorläufiger bericht über das vorkommen für spirochaeten in syphilitischen krankheitsprodukten und bei papillomen. *Arb. Gesundh. Amt. Berlin* **1905**, *22*, 528−534.

(29) Schaudinn, F.; Hoffmann, E. Über spirochaetenbefunde im lymphdrüsensaft syphilitischer. *Dtsch. Med. Wochenschr.* **1905**, *31*, 711−714.

(30) Fieldsteel, A. H.; Cox, D. L.; Moeckli, R. A. Cultivation of virulent *Treponema pallidum* in tissue culture. *Infect. Immun.* **1981**, *32*, 908−915.

(31) Norris, S. J. In vitro cultivation of *Treponema pallidum:* independent confirmation. *Infect. Immun.* **1982**, *36*, 437−439.

(32) Cox, D. L. Culture of *Treponema pallidum*. *Methods Enzymol.* **1994**, *236*, 390−405.

(33) Norris, S. J.; Edmondson, D. G. Factors affecting the multiplication and subculture of *Treponema pallidum* subsp. *pallidum* in a tissue culture system. *Infect. Immun.* **1986**, *53*, 534−539.

(34) Norris, S. J.; Cox, D. L.; Weinstock, G. M. Biology of *Treponema pallidum*: correlation of functional activities with genome sequence data. *J. Mol. Microbiol. Biotechnol.* **2001**, *3*, 37−62.

(35) Edmondson, D. G.; Hu, B.; Norris, S. J. Long-term In vitro culture of the syphilis spirochete *Treponema pallidum* subsp. *pallidum*. *mBio* **2018**, *9* (3), No. e01153-18.

(36) Edmondson, D. G.; DeLay, B. D.; Kowis, L. E.; Norris, S. J. Parameters affecting continuous in vitro culture of *Treponema pallidum* strains. *mBio* **2021**, *12* (1), No. e03536-20.

(37) Edmondson, D. G.; Norris, S. J. In vitro cultivation of the syphilis spirochete *Treponema pallidum*. *Curr. Protoc.* **2021**, *1* (2), No. e44.

(38) Lukehart, S. A.; Marra, C. M. Isolation and laboratory maintenance of *Treponema pallidum*. *Curr. Protoc. Microbiol.* **2007**, *7*, 12A.1.1−12A.1.18.

(39) Kall, L.; Storey, J. D.; Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* **2008**, *24* (16), i42−i48.

(40) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646−4658.

(41) Searle, B. C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10* (6), 1265−1269.

(42) Yu, F.; Haynes, S. E.; Nesvizhskii, A. I. IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. *Mol. Cell. Proteomics* **2021**, *20*, 100077.

(43) da Veiga Leprevost, F.; Haynes, S. E.; Avtonomov, D. M.; Chang, H. Y.; Shanmugam, A. K.; Mellacheruvu, D.; Kong, A. T.; Nesvizhskii, A. I. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **2020**, *17*, 869−870.

(44) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **2017**, *14*, 513−520.

(45) Cantalapiedra, C. P.; Hernandez-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **2021**, *38*, 5825−5829.

(46) Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernandez-Plaza, A.; Forslund, S. K.; Cook, H.; Mende, D. R.; Letunic, I.; Rattei, T.; Jensen, L. J.; von Mering, C.; Bork, P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309−D314.

(47) Tatusov, R. L.; Koonin, E. V.; Lipman, D. J. A genomic perspective on protein families. *Science* **1997**, *278*, 631−637.

(48) Romeis, E.; Tantalo, L.; Lieberman, N.; Phung, Q.; Greninger, A.; Giacani, L. Genetic engineering of *Treponema pallidum* subsp. *pallidum*, the syphilis spirochete. *PLOS Pathog.* **2021**, *17* (7), No. e1009612.

(49) Houston, S.; Schovanek, E.; Conway, K. M. E.; Mustafa, S.; Gomez, A.; Ramaswamy, R.; Haimour, A.; Boulanger, M. J.; Reynolds, L. A.; Cameron, C. E. Identification and functional characterization of peptides with antimicrobial activity from the syphilis spirochete, *Treponema pallidum*. *Front. Microbiol.* **2022**, *13*, 888525.

(50) Yu, N. Y.; Wagner, J. R.; Laird, M. R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S. C.; Ester, M.; Foster, L. J.; Brinkman, F. S. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26* (13), 1608−1615.

(51) Cameron, C. E. Identification of a *Treponema pallidum* laminin-binding protein. *Infect. Immun.* **2003**, *71*, 2525−2533.

(52) Cox, D. L.; Luthra, A.; Dunham-Ems, S.; Desrosiers, D. C.; Salazar, J. C.; Caimano, M. J.; Radolf, J. D. Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of *Treponema pallidum*. *Infect. Immun.* **2010**, *78*, 5178−5194.

(53) Radolf, J. D. Role of outer membrane architecture in immune evasion by *Treponema pallidum* and *Borrelia burgdorferi*. *Trends Microbiol.* **1994**, *2*, 307−311.

(54) Ruiz, N.; Gronenberg, L. S.; Kahne, D.; Silhavy, T. J. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5537−5542.

(55) Belisle, J. T.; Brandt, M. E.; Radolf, J. D.; Norgard, M. V. Fatty acids of *Treponema pallidum* and *Borrelia burgdorferi* lipoproteins. *J. Bacteriol.* **1994**, *176*, 2151−2157.

(56) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198−207.

(57) Radolf, J. D.; Deka, R. K.; Anand, A.; Smajs, D.; Norgard, M. V.; Yang, X. F. *Treponema pallidum*, the syphilis spirochete: making a living as a stealth pathogen. *Nat. Rev. Microbiol.* **2016**, *14*, 744−759.

(58) Wang, F.; Xiao, J.; Pan, L.; Yang, M.; Zhang, G.; Jin, S.; Yu, J. A systematic survey of mini-proteins in bacteria and archaea. *PLoS One* **2008**, *3* (12), No. e4027.

(59) Anand, A.; LeDoyt, M.; Karanian, C.; Luthra, A.; Koszelak-Rosenblum, M.; Malkowski, M. G.; Puthenveetil, R.; Vinogradova, O.; Radolf, J. D. Bipartite topology of *Treponema pallidum* repeat proteins C/D and I: outer membrane insertion, trimerization, and porin function require a C-terminal beta-barrel domain. *J. Biol. Chem.* **2015**, *290*, 12313−12331.

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743

(60) Anand, A.; Luthra, A.; Dunham-Ems, S.; Caimano, M. J.; Karanian, C.; LeDoyt, M.; Cruz, A. R.; Salazar, J. C.; Radolf, J. D. TprC/D (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of *Treponema pallidum*, has a bipartite domain structure. *J. Bacteriol.* **2012**, *194*, 2321−2333.
(61) Cameron, C. E.; Brown, E. L.; Kuroiwa, J. M. Y.; Schnapp, L. M.; Brouwer, N. L. *Treponema pallidum* fibronectin-binding proteins. *J. Bacteriol.* **2004**, *186*, 7019−7022.
(62) Cameron, C. E.; Lukehart, S. A.; Castro, C.; Molini, B.; Godornes, C.; Van Voorhis, W. Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies pallidum Tp92. *J. Infect. Dis.* **2000**, *181*, 1401−1413.
(63) Centurion-Lara, A.; Castro, C.; Barrett, L.; Cameron, C.; Mostowfi, M.; Van Voorhis, W. C.; Lukehart, S. A. *Treponema pallidum* major sheath protein homologue TprK is a target of opsonic antibody and the protective immune response. *J. Exp. Med.* **1999**, *189*, 647−656.
(64) Centurion-Lara, A.; Giacani, L.; Godornes, C.; Molini, B. J.; Brinck Reid, T.; Lukehart, S. A. Fine analysis of genetic diversity of the *tpr* gene family among treponemal species, subspecies and strains. *PLoS Neglected Trop. Dis.* **2013**, *7* (5), No. e2222.
(65) Desrosiers, D. C.; Anand, A.; Luthra, A.; Dunham-Ems, S. M.; LeDoyt, M.; Cummings, M. A.; Eshghi, A.; Cameron, C. E.; Cruz, A. R.; Salazar, J. C.; Caimano, M. J.; Radolf, J. D. TP0326, a *Treponema pallidum* beta-barrel assembly machinery A (BamA) orthologue and rare outer membrane protein. *Mol. Microbiol.* **2011**, *80*, 1496−1515.
(66) Giacani, L.; Brandt, S. L.; Ke, W.; Reid, T. B.; Molini, B. J.; Iverson-Cabral, S.; Ciccarese, G.; Drago, F.; Lukehart, S. A.; Centurion-Lara, A. Transcription of TP0126, *Treponema pallidum* putative OmpW homolog, is regulated by the length of a homopolymeric guanosine repeat. *Infect. Immun.* **2015**, *83*, 2275−2289.
(67) Giacani, L.; Sambri, V.; Marangoni, A.; Cavrini, F.; Storni, E.; Donati, M.; Corona, S.; Lanzarini, P.; Cevenini, R. Immunological evaluation and cellular location analysis of the TprI antigen of *Treponema pallidum* subsp. *pallidum*. *Infect. Immun.* **2005**, *73*, 3817−3822.
(68) Hawley, K. L.; Montezuma-Rusca, J. M.; Delgado, K. N.; Singh, N.; Uversky, V. N.; Caimano, M. J.; Radolf, J. D.; Luthra, A. Structural modeling of the *Treponema pallidum* outer membrane protein repertoire: a road map for deconvolution of syphilis pathogenesis and development of a syphilis vaccine. *J. Bacteriol.* **2021**, *203*, No. e0008221.
(69) Haynes, A. M.; Fernandez, M.; Romeis, E.; Mitja, O.; Konda, K. A.; Vargas, S. K.; Eguiluz, M.; Caceres, C. F.; Klausner, J. D.; Giacani, L. Transcriptional and immunological analysis of the putative outer membrane protein and vaccine candidate TprL of *Treponema pallidum*. *PLoS Neglected Trop. Dis.* **2021**, *15* (1), No. e0008812.
(70) Houston, S.; Hof, R.; Honeyman, L.; Hassler, J.; Cameron, C. E. Activation and proteolytic activity of the *Treponema pallidum* metalloprotease, pallilysin. *PLOS Pathog.* **2012**, *8* (7), No. e1002822.
(71) Houston, S.; Lithgow, K. V.; Osbak, K. K.; Kenyon, C. R.; Cameron, C. E. Functional insights from proteome-wide structural modeling of *Treponema pallidum* subspecies *pallidum*, the causative agent of syphilis. *BMC Struct. Biol.* **2018**, *18*, 7.
(72) Parker, M. L.; Houston, S.; Petrosova, H.; Lithgow, K. V.; Hof, R.; Wetherell, C.; Kao, W. C.; Lin, Y. P.; Moriarty, T. J.; Ebady, R.; Cameron, C. E.; Boulanger, M. J. The structure of *Treponema pallidum* Tp0751 (pallilysin) reveals a non-canonical lipocalin fold that mediates adhesion to extracellular matrix components and interactions with host cells. *PLOS Pathog.* **2016**, *12* (9), No. e1005919.
(73) Karpievitch, Y. V.; Dabney, A. R.; Smith, R. D. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinf.* **2012**, *13* (S16), S5.
(74) Cejkova, D.; Zobanikova, M.; Chen, L.; Pospisilova, P.; Strouhal, M.; Qin, X.; Mikalova, L.; Norris, S. J.; Muzny, D. M.; Gibbs, R. A.; Fulton, L. L.; Sodergren, E.; Weinstock, G. M.; Smajs, D. Whole genome sequences of three *Treponema pallidum* ssp. *pertenue* strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. *PLoS Neglected Trop. Dis.* **2012**, *6* (1), No. e1471.
(75) Petrosova, H.; Pospisilova, P.; Strouhal, M.; Cejkova, D.; Zobanikova, M.; Mikalova, L.; Sodergren, E.; Weinstock, G. M.; Smajs, D. Resequencing of *Treponema pallidum* ssp. *pallidum* strains Nichols and SS14: correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters. *PLoS One* **2013**, *8* (9), No. e74319.
(76) Petrosova, H.; Zobanikova, M.; Cejkova, D.; Mikalova, L.; Pospisilova, P.; Strouhal, M.; Chen, L.; Qin, X.; Muzny, D. M.; Weinstock, G. M.; Smajs, D. Whole genome sequence of *Treponema pallidum* ssp. *pallidum*, strain Mexico A, suggests recombination between yaws and syphilis strains. *PLoS Neglected Trop. Dis.* **2012**, *6* (9), No. e1832.
(77) Teufel, F.; Almagro Armenteros, J. J.; Johansen, A. R.; Gislason, M. H.; Pihl, S. I.; Tsirigos, K. D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **2022**, *40*, 1023−1025.
(78) Edmondson, D. G.; De Lay, B. D.; Hanson, B. M.; Kowis, L. E.; Norris, S. J. Clonal isolates of *Treponema pallidum* subsp. *pallidum* Nichols provide evidence for the occurrence of microevolution during experimental rabbit infection and in vitro culture. *PLoS One* **2023**, *18* (3), No. e0281187.
(79) Nichols, H. J.; Hough, W. H. Demonstration of *Spirochaeta pallida* in the cerebrospinal fluid from a patient with nervous relapse following the use of Salvarsan. *JAMA* **1913**, *60*, 108−110.
(80) Maier, T.; Guell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583*, 3966−3973.
(81) Matejkova, P.; Strouhal, M.; Smajs, D.; Norris, S. J.; Palzkill, T.; Petrosino, J. F.; Sodergren, E.; Norton, J. E.; Singh, J.; Richmond, T. A.; Molla, M. N.; Albert, T. J.; Weinstock, G. M. Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. *BMC Microbiol.* **2008**, *8*, 76.
(82) Smajs, D.; Norris, S. J.; Weinstock, G. M. Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. *Infect. Genet. Evol.* **2012**, *12*, 191−202.

https://doi.org/10.1021/acs.jproteome.3c00891
J. Proteome Res. 2024, 23, 1725−1743