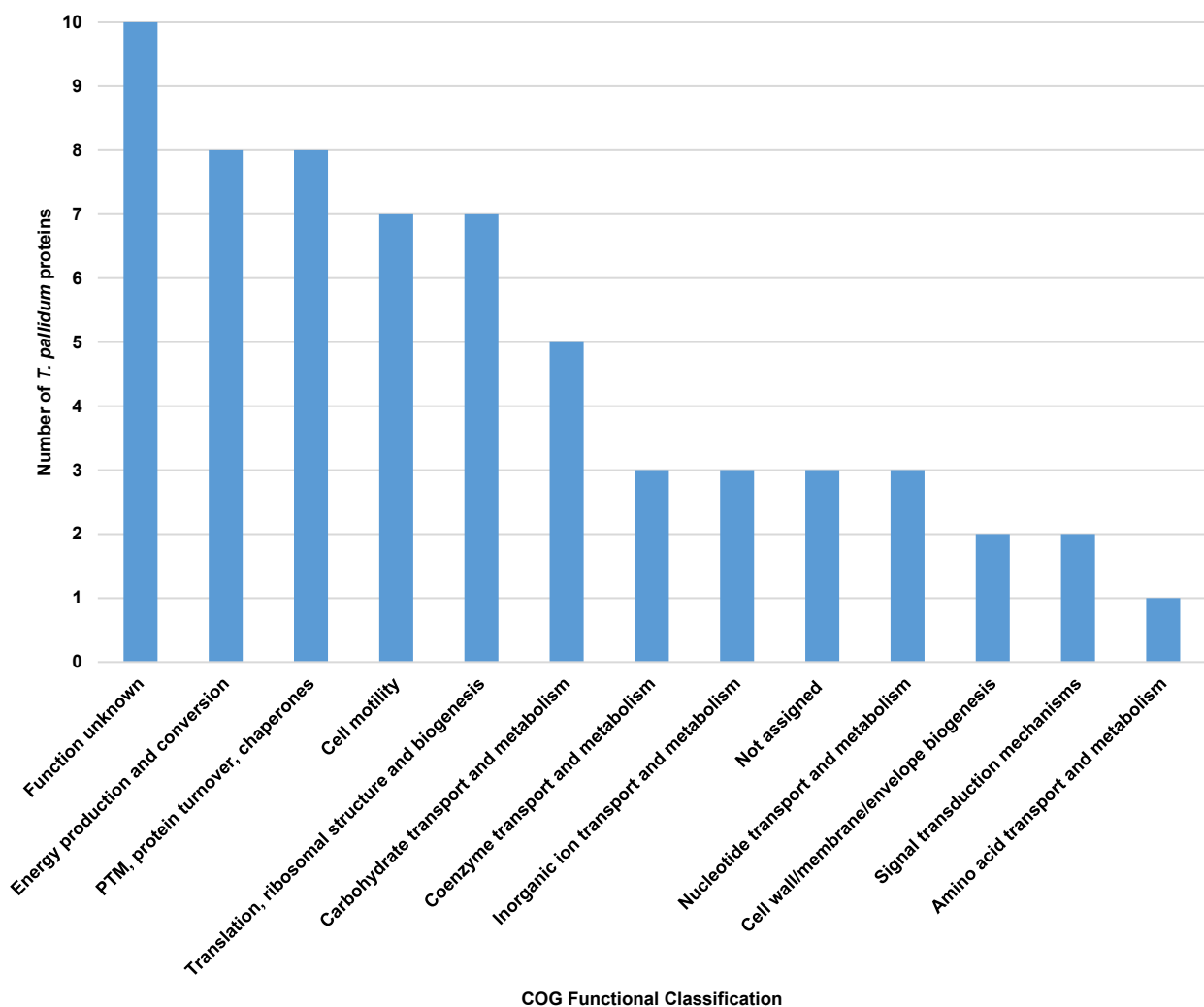# Deep Proteome Coverage Advances Knowledge of *Treponema pallidum* Protein Expression Profiles During Infection

Simon Houston[1], Alloysius Gomez[1], Andrew Geppert[1], Azad Eshghi[2], Derek S. Smith[2], Sean Waugh[1], Darryl B. Hardie[2], David R. Goodlett[1,2], and Caroline E. Cameron[1,3]*

Supplementary Information (Figs. S1-S7)
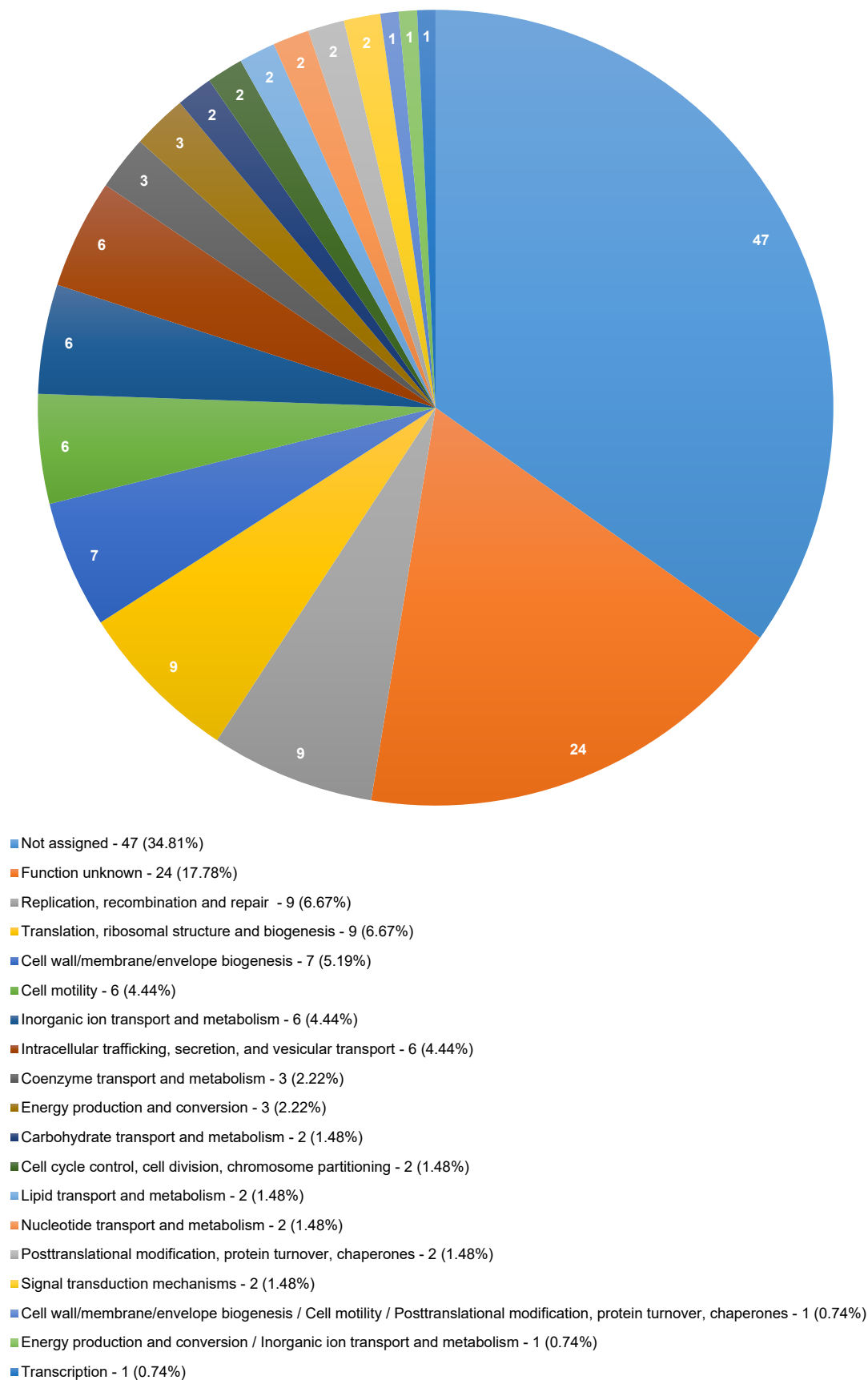
Supplementary Figure S1

Supplementary Figure S1. Functional classification of *T. pallidum* proteins detected in more than nine samples. Bar chart showing the distribution of 62 *T. pallidum* proteins detected in 9 - 12 of the in vivo-grown biological replicate samples, based on COG functional classifications. COG functional classifications were generated by the eggNOG-mapper tool. PTM – post-translational modification. Not assigned – proteins where eggNOG-mapper was unable to assign functional classifications.

Supplementary Figure S2

COG Functional Classification
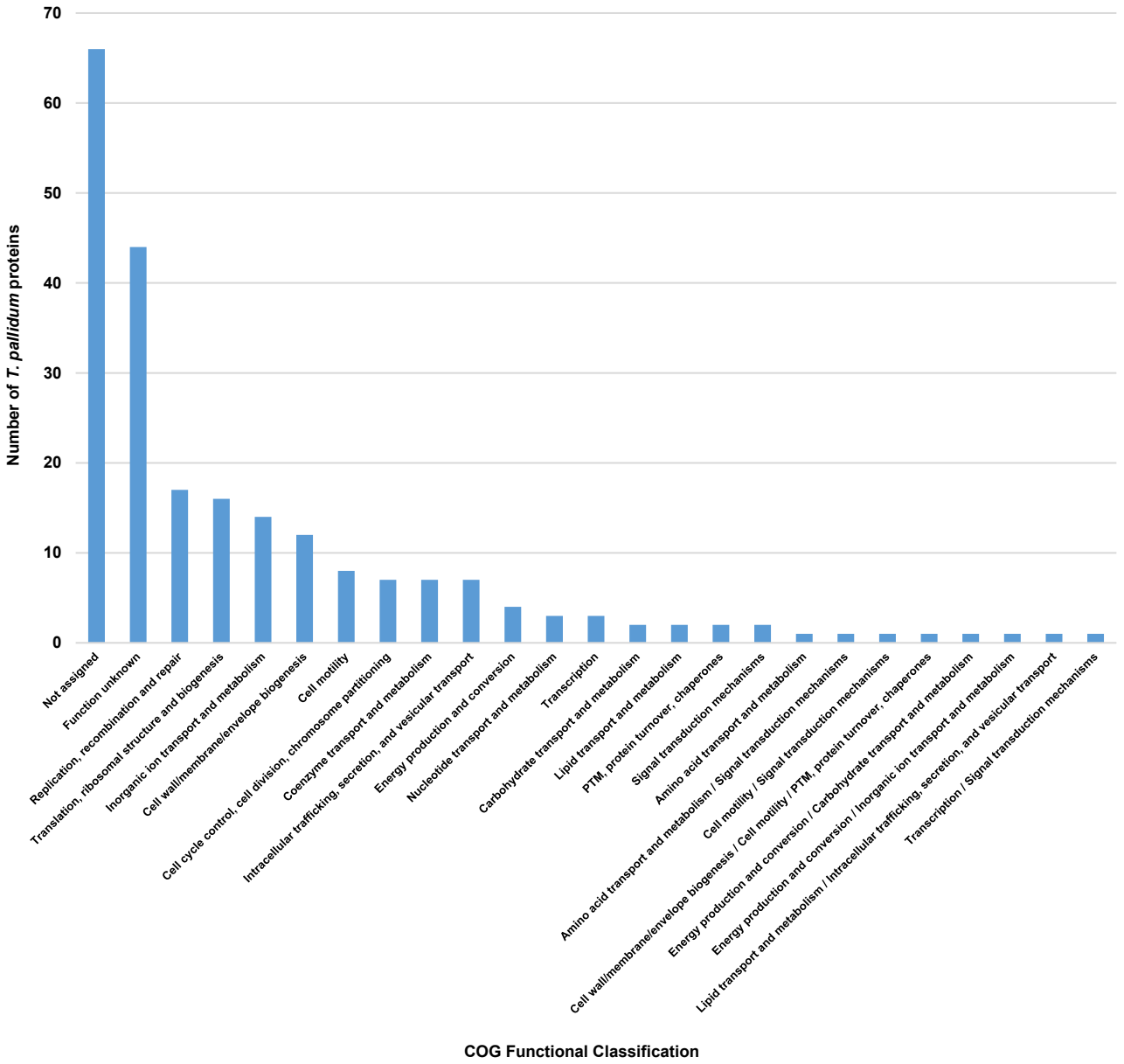
Supplementary Figure S2. Functional classification of undetected *T. pallidum* proteins in the present study. Bar chart showing the distribution of 224 undetected *T. pallidum* proteins based on COG (Clusters of Orthologous Genes) functional classifications generated by the eggNOG-mapper tool. PTM – post-translational modification. Not assigned – proteins where eggNOG-mapper was unable to assign functional classifications.

Supplementary Figure S3

■ Not assigned - 47 (34.81%)

■ Function unknown - 24 (17.78%)

■ Replication, recombination and repair - 9 (6.67%)

■ Translation, ribosomal structure and biogenesis - 9 (6.67%)

■ Cell wall/membrane/envelope biogenesis - 7 (5.19%)

■ Cell motility - 6 (4.44%)

■ Inorganic ion transport and metabolism - 6 (4.44%)

■ Intracellular trafficking, secretion, and vesicular transport - 6 (4.44%)

■ Coenzyme transport and metabolism - 3 (2.22%)

■ Energy production and conversion - 3 (2.22%)

■ Carbohydrate transport and metabolism - 2 (1.48%)

■ Cell cycle control, cell division, chromosome partitioning - 2 (1.48%)

■ Lipid transport and metabolism - 2 (1.48%)

■ Nucleotide transport and metabolism - 2 (1.48%)

■ Posttranslational modification, protein turnover, chaperones - 2 (1.48%)

■ Signal transduction mechanisms - 2 (1.48%)

■ Cell wall/membrane/envelope biogenesis / Cell motility / Posttranslational modification, protein turnover, chaperones - 1 (0.74%)

■ Energy production and conversion / Inorganic ion transport and metabolism - 1 (0.74%)

■ Transcription - 1 (0.74%)

Supplementary Figure S3. Functional classification of undetected treponemal proteins from past and present global proteomics studies. Pie chart depicting the COG-based functional classification distribution of 135 *T. pallidum* proteins that have not been detected in the present or previous [24,25] global proteomics studies from in vivo-grown *T. pallidum*. COG functional classifications were generated by the eggNOG-mapper tool. The number of proteins comprising each COG classification category are shown in the pie chart and in the figure key. The figures shown in parentheses (figure key) correspond to the percentage of undetected proteins in each of the COG functional categories (out of a total of 135 proteins). Not assigned – proteins where eggNOG-mapper was unable to assign functional classifications.

Supplementary Figure S4
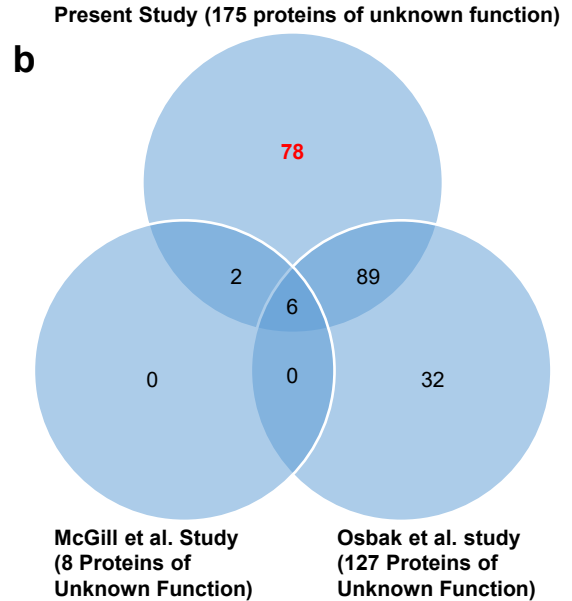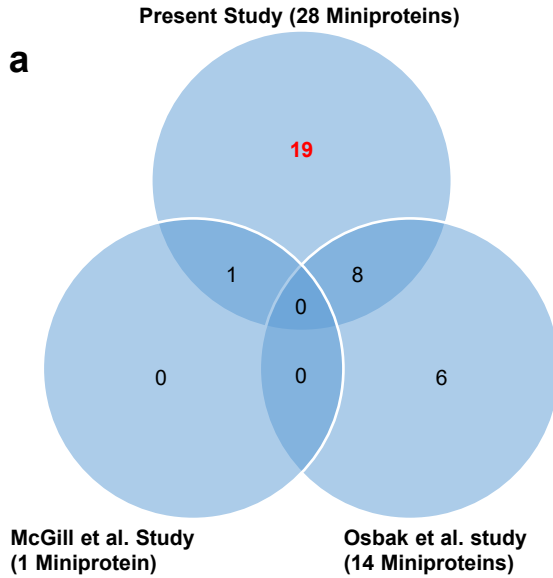
Supplementary Figure S4. Proteome coverage corresponding to *T. pallidum* proteins of unknown function. a, Venn diagram showing the total number of shared and exclusive identifications corresponding to miniproteins (≤ 150 amino acids) of unknown function (68 in the proteome) from in vivo-grown *T. pallidum* in the three mass spectrometry-based proteomics investigations. The total number of *T. pallidum* miniproteins identified exclusively in the current study (based on the detection of at least one tryptic peptide per identification) is highlighted in red text. **b,** Same as **a**, with the exception that the data relates to the identification of all (no size limitation) *T. pallidum* proteins of unknown function (from a total of 264 proteins identified in the proteome).

Supplementary Figure S5

a

TPANIC_0928

MKSFYFGRGALASTGGCVVLAVAALMFVFLPALSGGGHGHVRLGSWGGVALTNAPDSLFSQQYSALAQTAERYNLYPKDAQEREGMDRRVLRAAFRAA
VVQLAGTEQTKQSGFTLSEELLDREVLSFYADADGTYSPQRYQGTPEHVRLSQRKKMRDSLLSDQYLYHLFGKETGRGGLKLNSRELRFVQDMAKKERSFRY
VVLGEERFPAERVSAYGKEHAHLFTLHHLSLLTYSSEEDARRTARAL EKR ELSFEDAVATGSTK VGTDATGKMTHSYRSDVNEFFPDAQDLDTVLRTAVGALS
PVVKIERGFAIVRTDAEPSAPDFSDAATRERVFAHMARVERGTIERFLEEEAHTFSTRAKQGGFTAAAQSLNLDVHTS **RS** FPINFGNVDVLPALPRQSDPPLAR
IAYDEKFFSTAFALLPGQVSDPLILDSSVVLLQLHEEKSVDESVLATTGDSYAHHVRTWYPGYPLALLNTAKMPWAQESVVDTILAHPSFKDTFDNIFRR

b

TPANIC_0446

MNQRDERAARQPEEK VDSSAGVSPCNSPYGSGLLDVPLK LRPRASHYRARSLVIGGKEHVRALPLGGDAPIPIQTMWKEPLIGADLQSIVDRLLELEQ
LGCDVVRFAVPDRESAELFVALCARTRMPLVADIHFDYRLALRCMDGPVAKVRINPGNIGVRERVRAVVEKARATGTALRIGVNTGSLPGVVKRAVA
ARYADGMQSVNARAEALVQTAFAEAAYLDQLHFDRVVLSLKASTVAETVRANELFAQQSDIPLHLGVTEAGPLVSGIVKSTLAFSQLLSRNIGATVRV
SLSDSMEHEVLAAREILAECGKRAGGVRLVSCPRCGRIGFDVHAFVRRWQKELFSLKKDITVAVMGCVVNGPGEGKHADLGISGAEDSVIFFKRGKIV
RRIQVRDLCADERTRIIDAAFKEELSSL

c

TPANIC_0535

MSAAWVGNMDKGVMVR LAEVEDAAAVLVEK AQEQAQR MRESAEAAGETLVKEAEGSVVEAQEQAWHEACARIDAQRREACGQLERELSVLEEDVE
RFEAFLDGFFFG

d

TPANIC_0765

MCFFAAPCIPPQRTSLSCAVRLSHSLSTFHLLFVYHGPACPRALQKGALTEMNTRYK QSDDSSDPFGFFK FSPRPQKGPSSSRERPPRRNSRKVLSLVLLALC
ALLALANHFL FSRALRIIPFSEFKDRIASGEIVKVVVGSPYFVGYTSARPAPSARGFSLLSEREAPTYHAIGVLSDSFLSMLDERQVVYSIKPRERNYLIETFQYLFP
LLILFFVWRFFFKRMASNVSGLGSSIFSAGHARSAAVEEGKVTTRFADVAGVDEAKEELMEVVDFLKFPKKYTEIGGKIPRGVLLVGPPGTGKTLLARAVAGE
ASVPFFRISGSDFIEMFVGIGASRVRDLFKQAREKAPGIIFIDELDAIGKSRLNAIHSNDEREQTLNQLLVEMDGFDNTTGLILLAATNRPDVLDPALLRPGRFD
RQVCVDRPDLKGREAILRIHAQNVKLAPEVDLKAVARITGGYSGADLANVVNEAALLAVRSGRAQVIETDLDEAVEKTMIGLQKKSRVIREEERRIIAYHETG
HALAGTFTKGADKVHKITIIPRGTSALGYTFHIPEDDRHIVTEQQLLAEVDVLLSGRAAEFVAFGEVSTGAGNDISRATDIVRKMITDYGMSEKFQNVALTRR
GTGYLAEPQLAREYSECTQQYVDEEVARVLAERYRAVVALLTEKKELLEYIATRLLERETIERDEFEEVIRCEHDLEGLKKRLSQQQEGASAPAQGQSPPS

Supplementary Figure S5. *T. pallidum* proteome annotation errors. a, Amino acid sequence of TPANIC_0928. This protein was incorrectly deleted from the NCBI *T. pallidum* proteome (Nichols strain NC_021490, July 2021 annotation) and replaced with two shorter proteins, one homologous to the N-terminus of TPANIC_0928 (TPANIC_RS05505, homologous sequence shown in green text) and one homologous to the C-terminus of TPANIC_0928 (TPANIC_RS05510, homologous sequence shown in blue text). We detected a TPANIC_0928 peptide (ELSFEDAVATGSTK) (underlined) in three *T. pallidum* biological replicate samples that is not present in either TPANIC_RS05505 or TPANIC_RS05510. Residues in black text are not present in either TPANIC_RS05505 or TPANIC_RS05510 but present in TPANIC_0928. Amino acid sequences of **b,** TPANIC_0446, **c,** TPANIC_0535, and **d,** TPANIC_0765 were annotated with incorrect N-terminal truncations in the NCBI *T. pallidum* proteome. Peptides from previous NCBI annotated versions (pre July 2021) of each of the three proteins were detected in the present study (underlined). Blue arrows indicate the start sites of the incorrectly truncated protein versions in the NCBI July 2021 proteome annotation. The extended N-termini corresponding to the previously annotated versions of each protein is highlighted in red font.
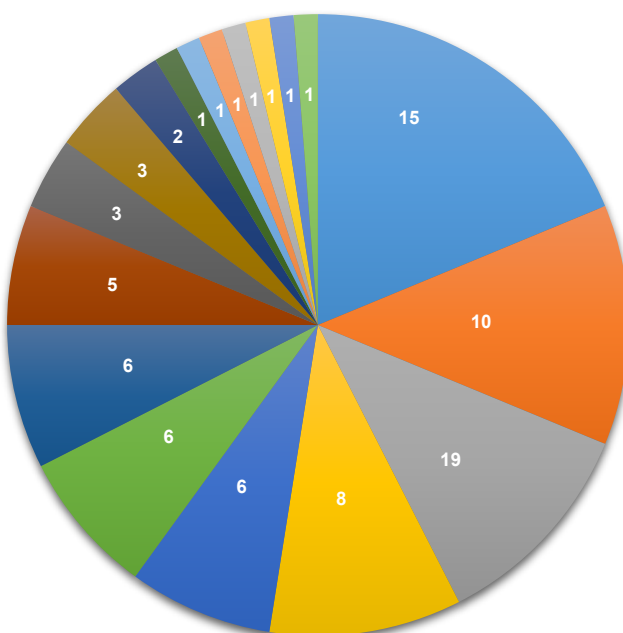
Supplementary Figure S6

Sample 10
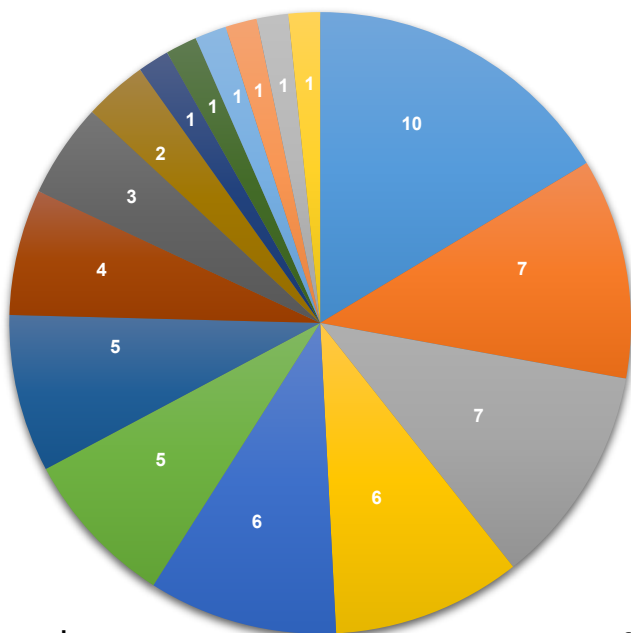
- Energy production and conversion – 10 (16.39%)
- Function unknown – 7 (11.48%)
- Translation, ribosomal structure and biogenesis – 7 (11.48%)
- Cell motility – 6 (9.84%)
- PTM, protein turnover, chaperones – 6 (9.84%)
- Carbohydrate transport and metabolism – 5 (8.20%)
- Not assigned – 5 (8.20%)
- Nucleotide transport and metabolism – 4 (6.56%)
- Coenzyme transport and metabolism – 3 (4.92%)
- Inorganic ion transport and metabolism – 2 (3.28%)
- Cell wall/membrane/envelope biogenesis / Inorganic ion transport & Metabolism – 1 (1.64%)
- Cell motility / Signal transduction mechanisms – 1 (1.64%)
- Cell wall/membrane/envelope biogenesis – 1 (1.64%)
- Replication, recombination and repair – 1 (1.64%)
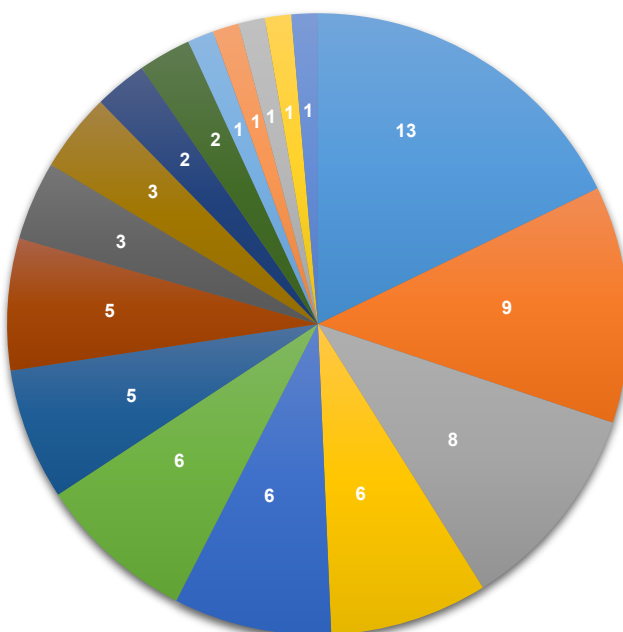- Signal transduction mechanisms – 1 (1.64%)
- Transcription – 1 (1.64%)

Sample 11

- Translation, ribosomal structure and biogenesis – 15 (18.75%)
- Function unknown – 10 (12.50%)
- Post-translational modification, protein turnover, chaperones – 9 (11.25%)
- Energy production and conversion – 8 (10.00%)
- Carbohydrate transport and metabolism – 6 (7.50%)
- Cell motility – 6 (7.50%)
- Not assigned – 6 (7.50%)
- Nucleotide transport and metabolism – 5 (6.25%)
- Coenzyme transport and metabolism – 3 (3.75%)
- Inorganic ion transport and metabolism – 3 (3.75%)
- Cell wall/membrane/envelope biogenesis – 2 (2.5%)
- Amino acid transport and metabolism – 1 (1.25%)
- Cell motility / Signal transduction mechanisms – 1 (1.25%)
- Cell wall/membrane/envelope biogenesis / Inorganic ion transport & metabolism – 1 (1.25%)
- Energy production and conversion / PTM, protein turnover, chaperones – 1 (1.25%)
- Replication, recombination and repair – 1 (1.25%)
- Signal transduction mechanisms – 1 (1.25%)
- Transcription – 1 (1.25%)

Sample 12

- Translation, ribosomal structure and biogenesis – 13 (17.81%)
- Energy production and conversion – 9 (12.33%)
- Posttranslational modification, protein turnover, chaperones – 8 (10.96%)
- Function unknown – 6 (8.22%)
- Nucleotide transport and metabolism – 6 (8.22%)
- Not assigned – 6 (8.22%)
- Carbohydrate transport and metabolism – 5 (6.85%)
- Cell motility – 5 (6.85%)
- Cell wall/membrane/envelope biogenesis – 3 (4.11%)
- Inorganic ion transport and metabolism – 3 (4.11%)
- Coenzyme transport and metabolism – 2 (2.74%)
- Signal transduction mechanisms – 2 (2.74%)
- Cell wall/membrane/envelope biogenesis / Inorganic ion transport & metabolism – 1 (1.37%)
- Amino acid transport and metabolism – 1 (1.37%)
- Cell motility / Signal transduction mechanisms – 1 (1.37%)
- Energy production and conversion / PTM, protein turnover, chaperones – 1 (1.37%)
- Replication, recombination and repair – 1 (1.37%)

Supplementary Figure S6. Functional classification of high-abundant *T. pallidum* proteins. a-c, Pie chart depicting the COG-based functional classification distribution of 61 high-abundant (LFQ intensity values greater than the mean average) *T. pallidum* proteins detected in sample 10 (**a**), 80 detected in sample 11 (**b**), and 73 detected in sample 12 (**c**). COG functional classifications were generated by the eggNOG-mapper tool. The number of proteins comprising each COG classification category are shown in the pie chart and in the figure key. The numbers shown in parentheses (figure key) correspond to the percentage of detected, high-abundant proteins in each of the COG functional categories. PTM – post-translational modification. Not assigned – proteins where eggNOG-mapper was unable to assign functional classifications.

Supplementary Figure S7

a

b

c
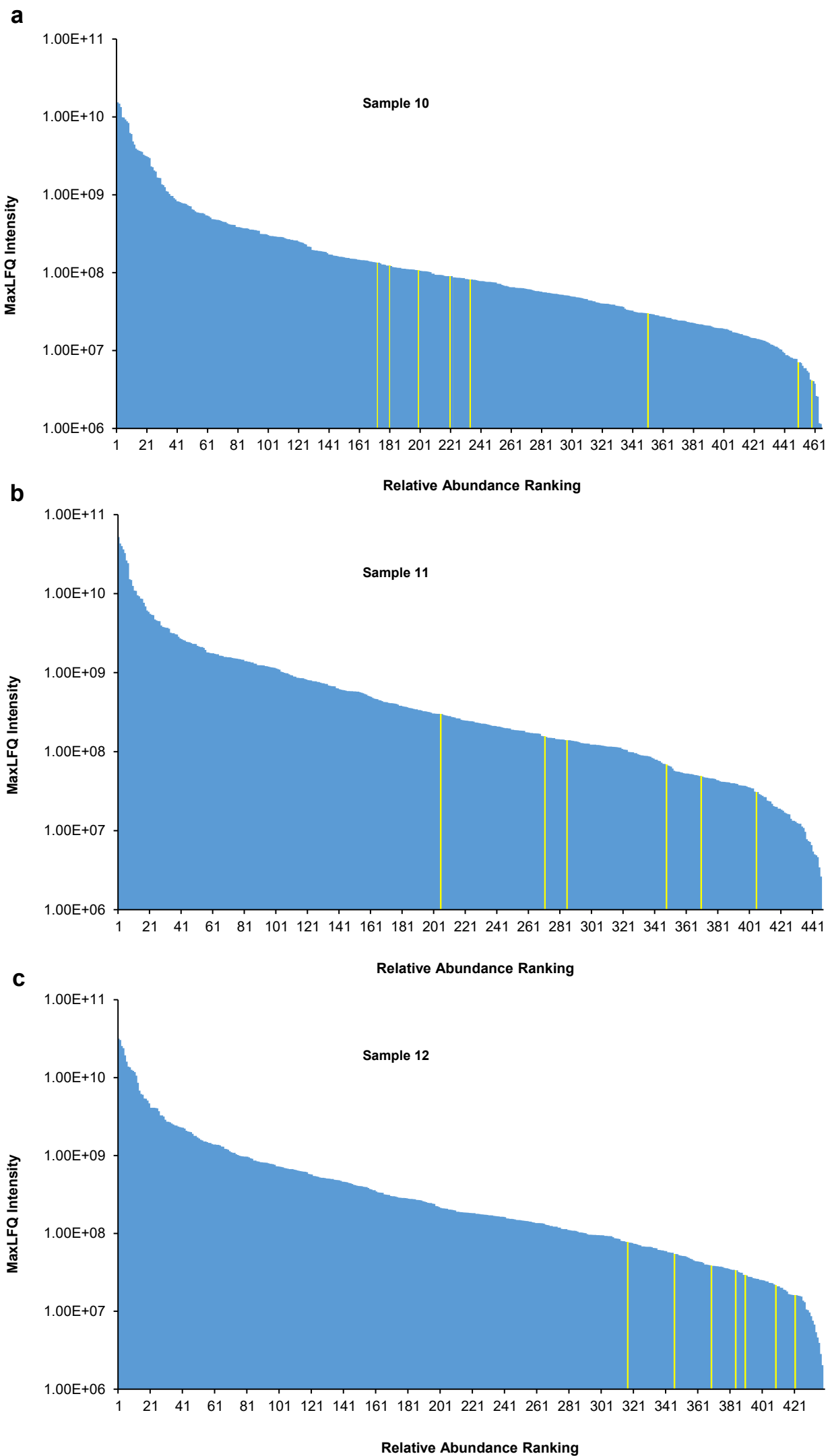
Supplementary Figure S7. Relative abundances of *T. pallidum* OMPs. a-c, Graph showing the relative abundances of 465 treponemal proteins (ranked from highest to lowest relative abundance [x-axis]), including eight known/predicted OMPs (yellow highlighting), that were semi-quantified based on peptide ion peak intensities (MaxLFQ intensity values [y-axis]) in sample 10 (**a**), 446 treponemal proteins (including six known/predicted OMPs) in sample 11 (**b**), and 438 treponemal proteins (including seven known/predicted OMPs) in sample 12 (**c**). Higher MaxLFQ intensity corresponds to higher protein abundance.