PNAS
www.pnas.org

Supplementary Information for

Proteo-genetic analysis reveals clear hierarchy of ESX-1 secretion in *Mycobacterium marinum*

Rachel M. Cronin, Micah J. Ferrell, Clare W. Cahir, Matthew M. Champion and Patricia A. Champion.

Patricia A. Champion and Matthew M. Champion
Email: pchampio@nd.edu and mchampio@nd.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Supplementary Information Text

Supplementary Results

Proteo-genetic analysis of *M. marinum* secretion reveals distinct secretory profiles for ESX-1 substrates. As indicated in Figure 2C, the secretion of several proteins clustered with the secretion of ESX-1 substrates. EsxA and EsxB share a clade with EsxN, an EsxA-like protein secreted by the ESX-5 system (1) and a broader clade with the probable serine protease PepA (MMAR_0324) and the ESX-1 substrate PPE68 (2). In *M. marinum,* there are seven paralogous *esxN* genes, five of which yield identical proteins. It is unclear which EsxN protein(s) mirrors ESX-1 substrate secretion, which has been used before to detect novel ESX-1 substrates (2-4). EspE and EspF shared a broader clade with the probable cellulase, CelA (MMAR_0107) and EgtD, part of the L-ergothioneine pathway which protects *M. tuberculosis* against oxidative stress (5). EspA and EspC shared a broader clade with MMAR_4028, a conserved secreted protein of unknown function (6). Finally, EspB and EspJ share a larger clade with MMAR_1442 (PG/PGRS protein), FadD2 (fatty acid Co-A ligase), MmaA3 (methyl mycolic acid synthase) and Efp (translation elongation factor) (6). Bioinformatic analysis of consensus targeting signals suggested that three of the nine proteins had predicted Sec signal sequences (7). The remaining six proteins had no predictable signal sequences (Sec or Tat systems) and were not predicted lipoproteins, similar to proteins secreted by alternate secretion pathways.

Supplementary Methods

Growth of *M. marinum* strains. All *M. marinum* strains (listed in Table S1) were derived from the M strain (WT; ATCC BAA-525). Bacterial strains were maintained as previously described (4, 8, 9). Briefly, *M. marinum* strains were grown at 30°C in Middlebrook 7H9

liquid broth (Sigma-Aldrich, St. Louis, MO) with 0.5% glycerol and 0.1% Tween 80 (Amresco, Solon, OH) or on Middlebrook 7H11 agar (Sigma-Aldrich) supplemented with 0.05% glycerol and 0.5% glucose. Kanamycin (IBI Scientific, Peosta, IA) (20 µg/ml) or hygromycin (EMD Millipore, Billerica, MA) (50 µg/ml) were added to broth or agar when appropriate. Strains with integrating plasmids were grown in the absence of antibiotics. For all assays, it was estimated that 1 OD_{600} unit is equivalent to 7.7 x 10^7 cells/ml for *M. marinum*. *E. coli* DH5α (New England Biolabs, Ipswich, MA) was grown in LB (Luria-Bertani) medium (VWR) with 50µg/ml kanamycin or 200µg/ml hygromycin (Thermo Fischer, Waltham, MA) when necessary.

Nomenclature. The nomenclature used in this manuscript are aligned with those proposed by Bitter et al. (10). Briefly, components of the membrane complex of the ESX-1 secretory apparatus are named ESX-1 conserved component (Ecc). Additional ESX associated proteins (Esp) can be either substrates or components of the system. Esx proteins are ~100 AA proteins with a conserved WXG motif.

Generation of *M. marinum* strains. *M. marinum* strains were generated using allelic exchange (11) as published previously (12). Approximately 1,500bp upstream and downstream of the annotated open reading frame was amplified using PCR (see Table S2 for oligonucleotide primers). The PCR products were introduced into the p2NIL vector (Addgene plasmid number 20188; a gift from Tanya Parish (11)) by three-part FastCloning as previously described (13). Plasmid constructs were digested with PacI (NEB) and dephosphorylated using Antarctic phosphatase (NEB) with heat treatment at 80°C. The pGOAL19 vector (Addgene plasmid number 20190; a gift from Tanya Parish (11)) was digested with PacI (NEB) followed by 65°C heat treatment. pGOAL19 was ligated into p2NIL KO construct as previously described (9). Plasmids were quantified on a NanoDrop

instrument (Thermo Fisher), and 3 µg of plasmid was irradiated with 0.1 J/cm^2 UV light in a CL-1000 UV cross-linker (UVP) followed by electroporation into 500µl of electrocompetent *M. marinum* cells using a GenePulser XCell (Bio-Rad). Electrocompetent cells were prepared exactly as described previously (9, 14). Following electroporation, cells were transferred into 2 ml of 7H9 with 0.1% Tween 80 to incubate overnight at 30°C. Cells were collected by centrifugation, resuspended in 200µl of media, and plated on 7H11 agar (Sigma) supplemented with oleic acid-albumin-dextrose-catalase (OADC), 20 µg/ml kanamycin, and 60 µg/ml 5-bromo-4-chloro-3-indolyl-D-galactopyranoside (X-Gal). Merodiploids were picked, cultured, and then plated on 7H11 agar supplemented with OADC, 60 µg/ml X-Gal and 2% sucrose (Macron). White colonies were picked and cultured in 3 ml of 7H9 broth with 0.1% Tween 80. After approximately 5 days of growth, 500µl of culture was transferred to screw-cap tubes and 0.1 mm zirconia disruption beads (RPI) were added. Pellets were lysed by three 30 second pulses on a mini bead beater (Biospec Products) followed by a 10-minute centrifugation. 1µl of lysate was used in 10 µl PCR reactions. PCR products were run on TAE agarose gels stained with ethidium bromide (VWR) and imaged using a Gel Doc EZ imager (BioRad) and Image Lab software (Bio-Rad). PCR products were interpreted as WT or knockouts based on the size of amplicons.

All oligonucleotide primers (Table S2) were purchased from Integrated DNA Technologies (IDT, Coralville, IA). All plasmids used in this study are listed in Table S1. All plasmids and genetic deletions were confirmed by targeted DNA sequencing performed by the Notre Dame Genomics and Bioinformatics Facility.

Generation of complementation plasmids. All complementation plasmids were generated by amplifying each gene from *M. marinum* genomic DNA using the primers listed in Table S2. All complementation plasmids were confirmed using targeted Sanger DNA sequencing at the Notre Dame Genomics and Bioinformatics Facility.

Protein preparation and analysis. Protein secretion assays were performed as described previously (4, 8, 9). In short, *M. marinum* strains were grown in 5 ml 7H9 broth with 0.1% Tween 80, then moved to 25 ml 7H9 broth with 0.1% Tween 80. Next, *M. marinum* strains were diluted to an OD_{600} of 0.8 in Sauton's media with 0.01% Tween 80. Following 48 h of growth, *M. marinum* cells were collected by centrifugation. Supernatants were filtered through 0.2 µm Nalgene Stericups with polyethersulfone (PES) filters and 500µl of phenylmethylsulfonyl fluoride (PMSF) at a concentration of 174.2 µg/ml in isopropanol was added to each sample. Supernatants were then concentrated by ultrafiltration in a 3,000-molecular-weight-cutoff (MWCO) Amicon filter (Millipore) to isolate secreted protein fractions. On ice, *M. marinum* cells were resuspended in 500 µl of PBS with 5µl of PMSF at a concentration of 174.2 µg/ml was added to each sample. Cells were lysed using a Biospec Mini-BeadBeater-24, yielding the cell-associated protein fraction. The protein concentration was determined using the Pierce MicroBCA kit. The amount of protein loaded is indicated in figure legends. All antibodies were diluted in 5% nonfat dry milk in 1x phosphate-buffered saline (PBS) with 0.1% Tween 20 at the following concentrations: RNA polymerase subunit β (RNAP: 1:20,000, ab12087, Abcam), ESAT-6 (EsxA 1:500, HYB 076-08). The following reagents were obtained through BEI Resources, NIAID, NIH: polyclonal anti-*Mycobacterium tuberculosis* CFP-10 (gene *Rv3874*) (antiserum, rabbit 1:5,000, NR-13801) and polyclonal anti-*Mycobacterium tuberculosis* MPT-32 (gene *Rv1860*) (antiserum, rabbit; 1:30,000) (NR-13807). Horseradish peroxidase (HRP)-conjugated goat anti-mouse immunoglobulin secondary antibody (Bio-Rad) was used at 1:5,000 for detection of RNAP and ESAT-6. HRP-conjugated goat anti-rabbit IgG secondary antibody (Bio-Rad) was utilized at 1:20,000 for detection of CFP-10 and MPT-32. All proteins were detected as described previously (9).

Hemolysis Assay. Sheep red blood cell (sRBC) hemolysis assays were performed as previously described (15, 16), with the following modifications. Following three washes with PBS, *M. marinum* pellets were resuspended in 300µl of PBS, mixed with 100 µl of sRBCs, centrifuged, and incubated for 2 h at 30°C. The data shown in Figure 1B, S2 and S10 are the results of three independent biological replicates, each with three technical replicates. The error bars represents the standard deviation.

Macrophage infections. RAW 264.7 cells (ATCC TIB-71) were cultured and passaged as described previously (4). RAW 264.7 cells were seeded in 200 µl DMEM plus 10% FBS per well at 2 x 10^5 cells/ml in a clear 96-well plate (Thermo Fisher, Denmark) and allowed to grow for 24 h. Bacteria were added at an estimated MOI of 5 in Figure 1C (1 x 10^6 cells/ml) and an estimated MOI of 2.5 in Figure S10 (5 x 10^5 cells/ml) in technical triplicate and mixed. Infections continued for 2h before gentamycin (RPI Corporation) was added at 100 µg/ml. Infections were incubated for an additional 4 h before washing three times with sterile 1x PBS and adding fresh DMEM plus 10% FBS. Infections then continued to proceed for 24 h. After 24 h, medium was aspirated and 50 µl of an EthD-1 (1 µl/ml), Calcein-AM (0.25 µl/ml) (Live/Dead viability/cytotoxicity kit; Life Technologies, Carlsbad, CA) solution in 1x PBS was added. Cells were incubated for an additional 30 min at 37°C. Cells were imaged as described previously (4). Five images were taken per well, and ImageJ software was then used to quantify the number of red cells per image, exactly as described previously (17). The data shown in Figure 1C are the results of three independent biological replicates for each strain, each with three technical replicates. The data shown in Figure S10 are the results of four independent biological replicates for each strain, each with three technical replicates.

Proteomics. *Protein preparation-digestion.* Secreted protein fractions from *M. marinum* strains were generated as described above. Samples were prepared for mass-spectrometry proteomics essentially as described in (18, 19). 100µg of protein from each sample above was precipitated using a seven-fold excess of acetone, chilled at -20° C for 1 hour then centrifuged to pellet, decanted, and dried at room temperature. Pellets were resuspended in 40µl 7% SDS, 10mM TCEP (Tris-carboxyethyl phosphine), 100mM TEAB (Triethyl ammonium bicarbonate) and heated to 95°C for 5 minutes with vortexing. Proteins were alkylated with 10mM iodoacetamide in the dark for 30 minutes, then quenched and prepared for S-Trap (Protifi, NY) digestion by addition of phosphoric acid to 1.3% v/v. Acidified protein was flocculated by addition of 7 volumes of 90:10 MeOH:100 mM TEAB and loaded on to an S-Trap column and washed according to manufacturer's instructions. 2.5 µg of sequencing grade trypsin (Promega, Madison, WI) was added to each sample, with 100 µl excess 100mM TEAB and allowed to digest for 12 hours at 37° C. Peptides were collected by centrifugation, and solvent was removed by Speed-vac until dry.

iTRAQ Labeling. Protein digests were resuspended in 100 µl 500 mM TEAB with vortexing and each sample was labeled with 1 vial 8-plex iTRAQ reagent according to manufacturer's instructions. The 113, and 114 mass channels were used for matched WT and Δ*eccCb*$_1$ controls for each biological replicate. Table S3 describes the labeling key. Labeling was allowed to proceed at room temperature for 2 hours with moderate shaking. After labeling, samples were quenched by addition of 20 µl 1M glycine, mixed and allowed to rest for 15 minutes, then formic acid was added to 0.2%. Samples were pooled and water with 0.1% formic acid was added to dilute the iTRAQ solvents to <2%. Pooled labeled samples were desalted using 30 mg HLB solid phase extraction sorbent (Waters, Milford, MA) using recommended procedures. Samples were dried down and stored at -20° C prior to nano-UHPLC-MS/MS analysis.

MS-MS/MS Analysis. *Proteomics Data Analysis*: Individual sets of iTRAQ labeled samples were subjected to nano-UHPLC MS-MS/MS essentially as in (19, 20). Triplicate MS-MS/MS were acquired on a Q-Exactive HF instrument running an iTRAQ adjusted TOP 15 acquisition. RAW files were converted to .mgf using MS-Convert (Protowizard) and protein-spectral matching and iTRAQ quantification was performed using Protein Pilot with the Paragon Algorithm (21, 22) and background quantitative correction. RAW and processed files are available at MassIVE ftp://massive.ucsd.edu/MSV000088597/PDX030584. Samples were normalized using median normalization and adjusted manually as needed to correct for the individual WT samples. Data from the mutant strains was background subtracted to eliminate isobaric interference generating false-positive abundance for the deleted proteins. In some cases we identified signals for proteins whose genes were deleted which we expect is due to isobaric interference, despite our efforts to minimize this. However, these levels of detection were very low. For example, the levels of PPE68 we measured were <0.1 of the amount of PPE68 protein detected from the WT strain. I.e., PPE68 is the 600th protein on the list in order of relative abundance (See Dataset S1A). Moreover, PPE, EsxB, and EspJ proteins generate poor tryptic fragments, many of which are identical to other proteins. EspJ in particular is expected to be extremely poorly resolved using trypsin as the protease. Despite its moderate mass (~26 KDa) it produces only a single tryptic peptide with high reliability. In Bosserman et al., we did not detect EspJ in the protein fractions from the Δ*MMAR_2894* strain and reported the fold change as the limit-of-detection in the experiment.

Three or more of the tryptic peptides for PPE68 are nearly identical to those found in a PPE68 homolog, B2HM15 MMAR_5417 which could be a source of mis-assigned signal or near-isobar interference. We consider the data here (~8 fold reduction) to be more reliable, however EspJ is problematic as a proteomics target. In general the data here under

report fold-change as compared to label-free data. This is to some extent expected as isobaric tags are generally known to have a lower dynamic range than targeted and label-free quantification (23). The resulting normalized protein data were exported as CSV files and joined using R Studio (See Supplementary Methods).

Our initial *P* value cut-off used in this study was *P* = .05. The standard error used in the *P* value calculation is based on the average of the number of peptides detected. Since ESX substrates in general are small, standard errors can vary widely and the software fixes the standard deviation at 0.17 for cases in which proteins are assigned with <4 peptides which is the vast majority of ESX-related substrates (24). This is likely an overestimate of the standard deviation. We initially erred on the side of caution, and made exceptions as described in the text to this strict cutoff for cases in which the proteins are missing from a sample in absolute terms, (E.g. a deletion), or small numbers of available peptides for quantification (ESX-1 substrates).

Functional Analysis: To define the proteins whose secretion was significantly reduced by the deletion of individual genes; functional analysis was performed as follows. Using the untrimmed data that generated the volcano plots, we used a cut off of $\log_2$ fold change $\geq$ -1, with *P* values $\leq$ 0.05. The functional category of each resulting protein was defined using Mycobrowser (6). The number of proteins in each functional category was defined using the COUNTIF command in Microsoft Excel, and the resulting data was plotted on Donut plots using Prism 9.

Statistical Analyses: All statistical analyses were performed as described in the figure legends using Prism 9.

***Proteomic Data Assembly*:** R version 4.0.4 (2021-02-15) -- "Lost Library Book" in R Studio was used to join the six datasets resulting from the iTRAQ experiments as follows:

```
#Read in data for alpha set of deletion strains:
library(readxl)
Alpha_iT8_1_QEHF_R1_03A_tidy <- read_excel("Downloads/wran-
     gle03a/Alpha_iT8_1_QEHF_R1_03A_tidy.xlsx")
View(Alpha_iT8_1_QEHF_R1_03A_tidy)
library(readxl)
Alpha_iT8_A_QEHF_R1_03A_tidy <- read_excel("Downloads/wran-
     gle03a/Alpha_iT8_A_QEHF_R1_03A_tidy.xlsx")
View(Alpha_iT8_A_QEHF_R1_03A_tidy)
#Join the alpha subset of data and remove redundant columns
library(dplyr)
A_HF_joined <- full_join(Alpha_iT8_1_QEHF_R1_03A_tidy, Al-
     pha_iT8_A_QEHF_R1_03A_tidy,  by =  "Accession")
View(A_HF_joined)
A_HF_joined_1 <- A_HF_joined %>% select(-"Name.y", -"Species.y")
#Read in data for beta set of deletion strains:
library(readxl)
Beta_iT8_1_QEHF_R1_03A_tidy <- read_excel("Downloads/wran-
     gle03a/Beta_iT8_1_QEHF_R1_03A_tidy.xlsx")
View(Beta_iT8_1_QEHF_R1_03A_tidy)
library(readxl)
Beta_iT8_A_QEHF_R1_03A_tidy <- read_excel("Downloads/wran-
     gle03a/Beta_iT8_A_QEHF_R1_03A_tidy.xlsx")
View(Beta_iT8_A_QEHF_R1_03A_tidy)
#Join the beta subset of data and remove redundant columns
B_HF_joined <- full_join(Beta_iT8_1_QEHF_R1_03A_tidy,
     Beta_iT8_A_QEHF_R1_03A_tidy,  by =  "Accession")
View(B_HF_joined)
B_HF_joined_1 <- B_HF_joined %>% select(-"Name.y", -"Species.y")
View(B_HF_joined_1)
#Read in data for delta set of deletion strains:
library(readxl)
Delta_iT8_1_QEHF_R1_03A_tidy <-read_excel("Downloads/wran-
     gle03a/Delta_iT8_1_QEHF_R1_03A_tidy.xlsx")
View(Delta_iT8_1_QEHF_R1_03A_tidy)
library(readxl)
Delta_iT8_A_QEHF_R1_03A_tidy <- read_excel("Downloads/wran-
     gle03a/Delta_iT8_A_QEHF_R1_03A_tidy.xlsx")
View(Delta_iT8_A_QEHF_R1_03A_tidy)
#Join the delta subset of data and remove redundant columns
D_HF_joined <- full_join(Delta_iT8_1_QEHF_R1_03A_tidy,
     Delta_iT8_A_QEHF_R1_03A_tidy,  by =  "Accession")
View(D_HF_joined)
D_HF_joined_1 <- D_HF_joined %>% select(-"Name.y", -"Species.y")
#Join the alpha, beta and delta subset of data and remove redundant columns
AB_HF_joined<- full_join(A_HF_joined_1, B_HF_joined_1,  by =
     "Accession")
```

```
View(AB_HF_joined)
AB_HF_joined_1 <- AB_HF_joined %>% select(-"Name.x.y", -"Spe-
cies.x.y")
ABD_HF_joined<- full_join(AB_HF_joined_1, D_HF_joined_1,  by =
     "Accession")
View(ABD_HF_joined)
ABD_HF_joined_1 <- ABD_HF_joined %>% select(-"Name.x", -"Spe-
cies.x")
library(writexl)
write_xlsx(ABD_HF_joined_1,"All_KOs_03A")
```

***Generation of Volcano plots*:** The average fold-change was converted to Log_2 Fold Change, and plotted against the $-log_{10}$ of the average P value from 3 (mutant) or 2 (complemented) biological replicates (except for EspB complementation, which only had one usable complemented data set, and EspA complementation which had three biological replicates). The data for the volcano plots was untrimmed. Conversion to $-log_{10}$ P value and thresholding was performed in R as follows (Example shown is for *eccCb$_1$*/WT, Figure 3A):

#Read in data

```
library(readxl)
EccCb1_Volcano_test_ <- read_excel("Dropbox/HEFQ itraq
     data/EccCb1 Volcano test .xlsx")
View(EccCb1_Volcano_test_)
```

#Generate the volcano plot, including calculating the $-log_{10}$Pvalue and thresholding to Pval<0.05

```
library(ggplot2)
EccCb1_Volcano_test_$threshold=as.factor(EccCb1_Vol-
     cano_test_$Avg_Pval < 0.05)
g <- ggplot(data = EccCb1_Volcano_test_, aes(x=Log2_Fold_Change,
     y = -log10(Avg_Pval), color = threshold)) + geom_point(al-
     pha = 0.4, size = 4) + xlim(c(-8, 8)) + theme_bw()
g
```

#Annotate the known ESX-1 substrates, add lines, format x axis and increase text size

```
g +theme(text=element_text(size = 20))+  geom_hline(yintercept =
     -log10(0.05), linetype = "dashed") + geom_vline(xintercept
     = c(log2(0.5), log2(2)), linetype = "dashed") + anno-
     tate("text", x = -5.07, y = 3.35, label = "EspB", size =
     8)+annotate("text", x = -4.54, y = 1.25, label = "EsxA",
     size = 8) + annotate("text", x = -5.65, y = 1.8, label =
     "EspF", size = 8) + annotate("text", x=-6.1, y=1.25, label
     = "EspE", size = 8) + annotate("text", x= -4.59, y = 1.65,
     label = "EspA", size = 8) + annotate("text", x = -5.27, y =
     2.1, label = "EsxB", size = 8) + annotate("text", x = -
```

```
3.06, y = 0.65, label = "PPE68", size = 8) + anno-
tate("text", x = -4.13, y = 1.90, label = "EspK", size = 8)
+ annotate("text", x = -2.4006, y = .80, label = "EspC",
size = 8)+ annotate("text", x = -1.625, y = 2.90, label =
"Pks5", size = 8) + annotate("text", x = -1.337, y = 1.77,
label = "Pks15/1", size = 8)
```

This website aided us with this script: https://rforbiochemists.blogspot.com/2015/06/drawing-proteomic-data-volcano-plot.html, https://r-graphics.org/recipe-scatter-labels. The thresholded data used for the generation of the volcano plots is found in Data Set 2.

Generation of ESX-1 substrate heatmap: To generate the heatmap (Figure 4A), log2 fold change values from Supplementary Data Set 1 were extracted for known ESX-1 substrates (Data Set 1, Tab E). The heat map was generated using the following code in R Studio:

#Input the datafile
```
library(readxl)
ESX_1_substrates_tidy_for_heatmap <- read_excel("Dropbox/HEFQ
itraq data/ESX-1 substrates tidy for heatmap.xlsx")
View(ESX_1_substrates_tidy_for_heatmap)
```
#Rename numbered rows using the substrate Names
```
library(pheatmap)
row.names(ESX_1_substrates_tidy_for_heatmap)<- ESX_1_sub-
     strates_tidy_for_heatmap$Name
```
#Generate the heat map
```
df <- data.matrix(ESX_1_substrates_tidy_for_heatmap)
pheatmap(df[, -(1)], cutree_rows = 4)
```

Generation of M. marinum whole proteome heatmap: To generate the heatmap of the entire proteome, the dataset was trimmed to remove all “reversed” protein calls. The heatmap in Figure 2, and the readable version of the Heatmap in Figure S4 was generated from the data in Dataset S1, tab B using the following code in R Studio:

Uploading data, making the accession number for each protein the row name and omitting N/A from the dataset.
```
library(readxl)
test_case_trimmed <- read_excel("Dropbox/HEFQ itraq data/test
case.xlsx")
View(test_case)
```

```
test_case_trimmed<-as.data.frame(test_case)
row.names(test_case)<-test_case$Accession
test_case<-test_case[,-1]
df <- as.matrix(test_case)
df <- na.omit(df)
```

#Formatting and downloading the readable heatmap (Figure S4)

```
pheatmap(df, scale = "row", cluster_rows = TRUE , cluster_cols =
      TRUE,
      + clustering_distance_cols = "euclidean", clustering_dis-
      tance_rows = "euclidean", +          clustering_method =
      "complete", fontsize=9, fontsize_row=5, cellwidth = 9, +
      cellheight = 5, main = "Heatmap", filename =
      "Heatmap_rowclear_PAC.pdf")
```

#Formatting and downloading the readable heatmap (Figure 2)

```
pheatmap(df, scale = "row", cluster_rows = TRUE , cluster_cols =
TRUE,
      +          clustering_distance_cols = "euclidean", cluster-
      ing_distance_rows = "euclidean",
      +          clustering_method = "complete", show_rownames=F,
      main = "Heatmap", filename = "Heatmap_PAC.pdf")
```

Generation of Pearson Correlation, Figure 2C. Pearson correlation was performed in R Studio using the data in Dataset S1 tab C as follows:

#Uploading data and performing and formatting the correlation

```
library(corrplot)
library(ggplot2)
library(readxl)
All_Clusters_for_correlation <- read_excel("Dropbox/HEFQ itraq
     data/ All_Clusters_for_correlation.xlsx")
View(All_Clusters_for_correlation)
cc=cor(All_Clusters_for_correlation, method = "pearson")
corrplot(cc)
corrplot(cc, tl.col = "black", addrect = 4, tl.cex = 0.7)
```

This website (https://jkzorz.github.io/2019/06/11/Correlation-heatmaps.html) assisted us in developing this code.

Generation of Venn Diagrams. Venn diagrams were generated using Venny (*Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams.* https://bioinfogp.cnb.csic.es/tools/venny/index.html). The input for Venny was the untrimmed list of names of proteins whose log2 fold change ≤ -1. The resulting output was

used to generate each scaled Venn Diagram in R using R Studio, taken from “The code was R: Draw a Venn diagram with two sets, by Hanbo Chen”

as follows:

```
venn.plot <- draw.pairwise.venn(
+     area1 = 20,
+     area2 = 35,
+     cross.area = 15,
+     category = c("core", "espB"),
+     fill = c("blue", "red"),
+     lty = "blank",
+     cex = 2,
+     cat.cex = 2,
+     cat.pos = c(285, 105),
+     cat.dist = 0.09,
+     cat.just = list(c(-1, -1), c(1, 1)),
+     ext.pos = 30,
+     ext.dist = -0.05,
+     ext.length = 0.85,
+     ext.line.lwd = 2,
+     ext.line.lty = "dashed"
+ )
grid.newpage()
grid.draw(venn.plot)
```

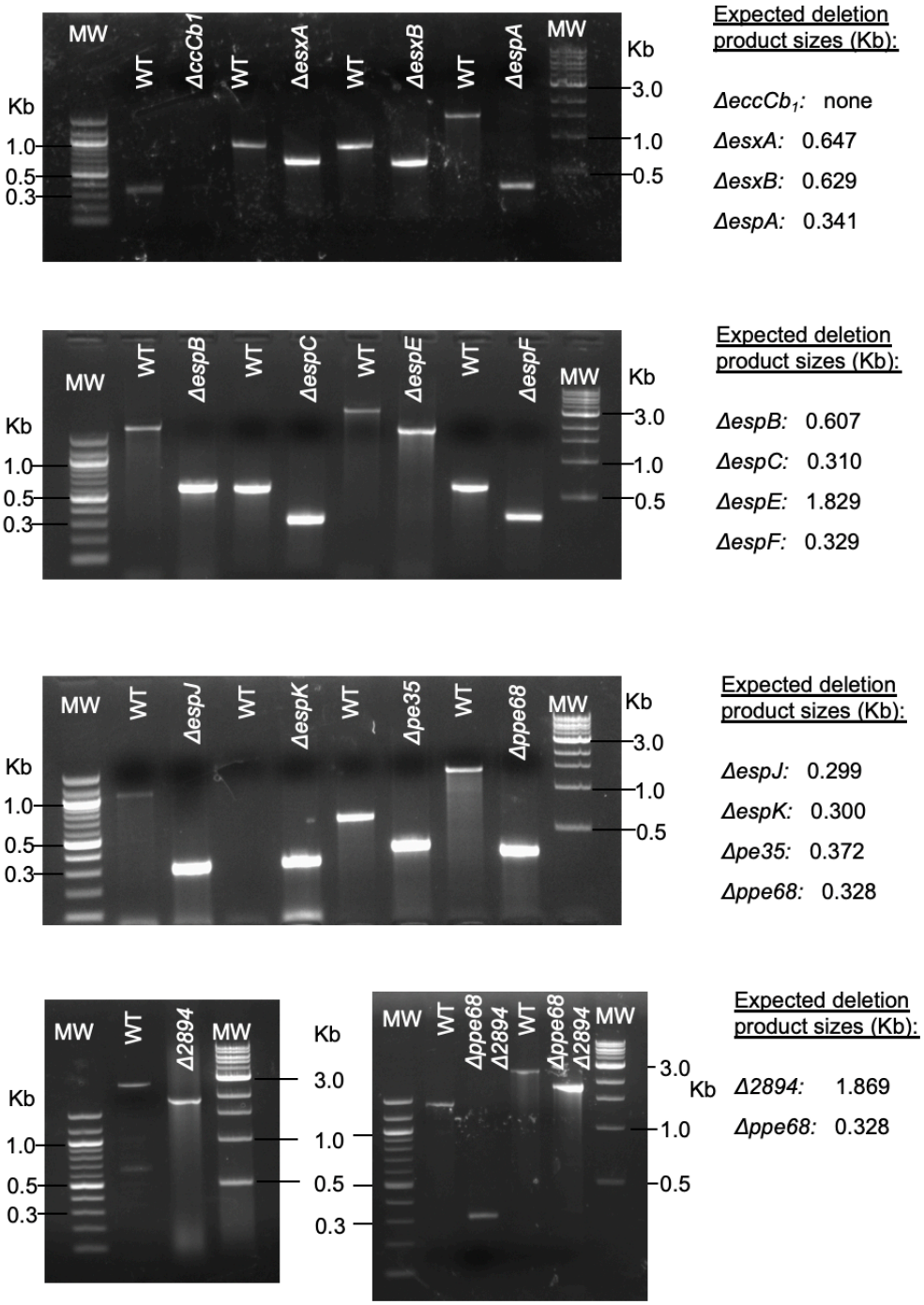


Figure S1: Confirmation of ESX-1 substrate gene deletions in *M. marinum*. Genetic deletion of ESX-1 substrate genes was confirmed using PCR on the whole cell lysates with the primers in Table S2. All deletions were further verified by targeted DNA sequencing.

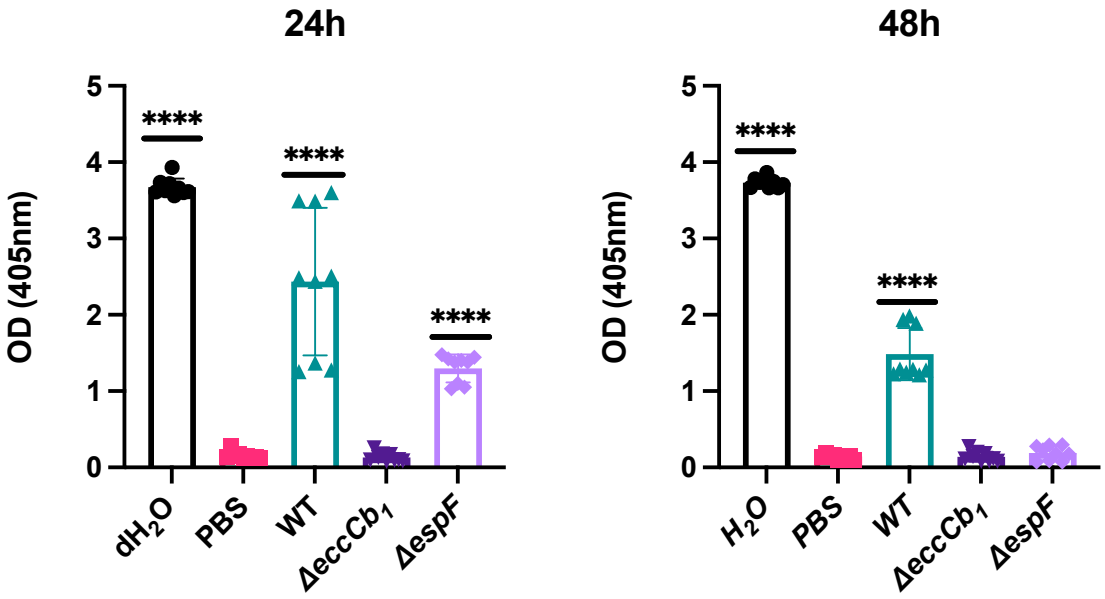


Figure S2: The requirement of EspF in hemolysis is dependent upon growth phase. Hemolysis assay performed on *M. marinum* strains. Strains were grown in 10 ml 7H9 broth for 24 or 48 hours. The same 10ml culture was used to measure hemolysis at 24 and 48 hour time points. Means of data from three biological replicates, in technical triplicate, are shown. Statistical analysis was performed using a one-way ordinary ANOVA (P<.0001), followed by a Tukey's Multiple Comparison Test. ****, P<.0001, as compared to the Δ*eccCb*$_1$ strain. The data in Figure 1B were generated after 24 hours in 7H9 broth. The data we published previously demonstrating a requirement for EspF in hemolysis was generated after 48 hours in 7H9 broth (8). Error bars represent standard deviation.

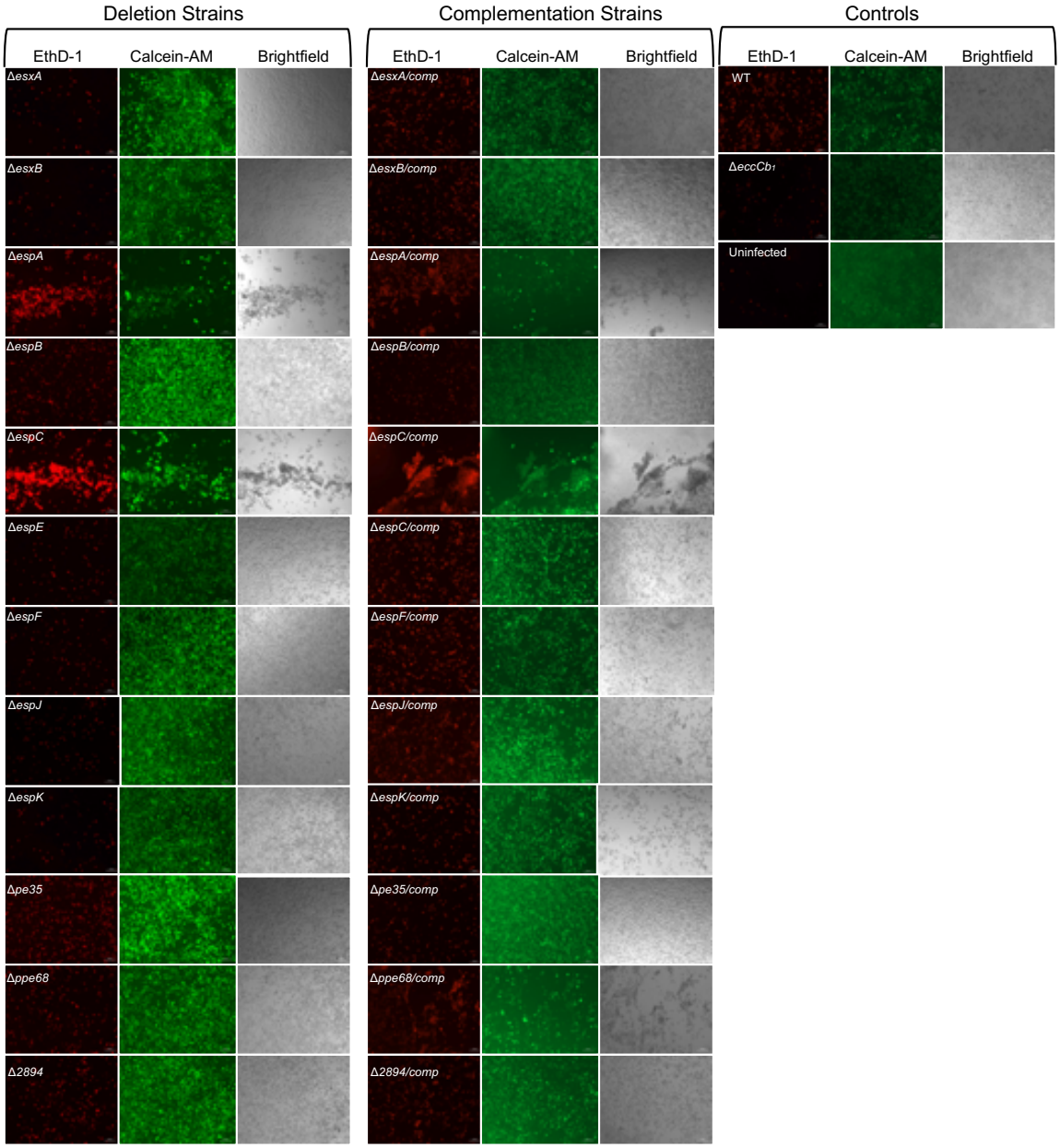


Figure S3: Images from Live/ Dead analysis of Macrophage Infections (Related to Figure 1) Representative images for RAW 264.7 macrophage infections from Figure 1C. EthD-1 and all brightfield images were adjusted with +38% brightness to aid in visualization. Scale bar represents 50μm.

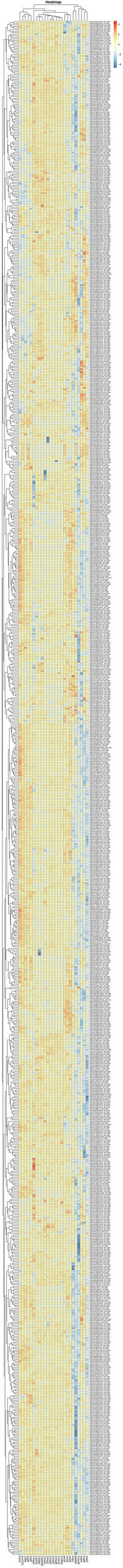

Heatmap
4
2
0
-2
-4

Figure S4: Readable Heat Map (Related to Figure 2A). Heatmap generated from the Dataset S1E, using code in the Supplementary Methods. Accession number is shown for each protein.

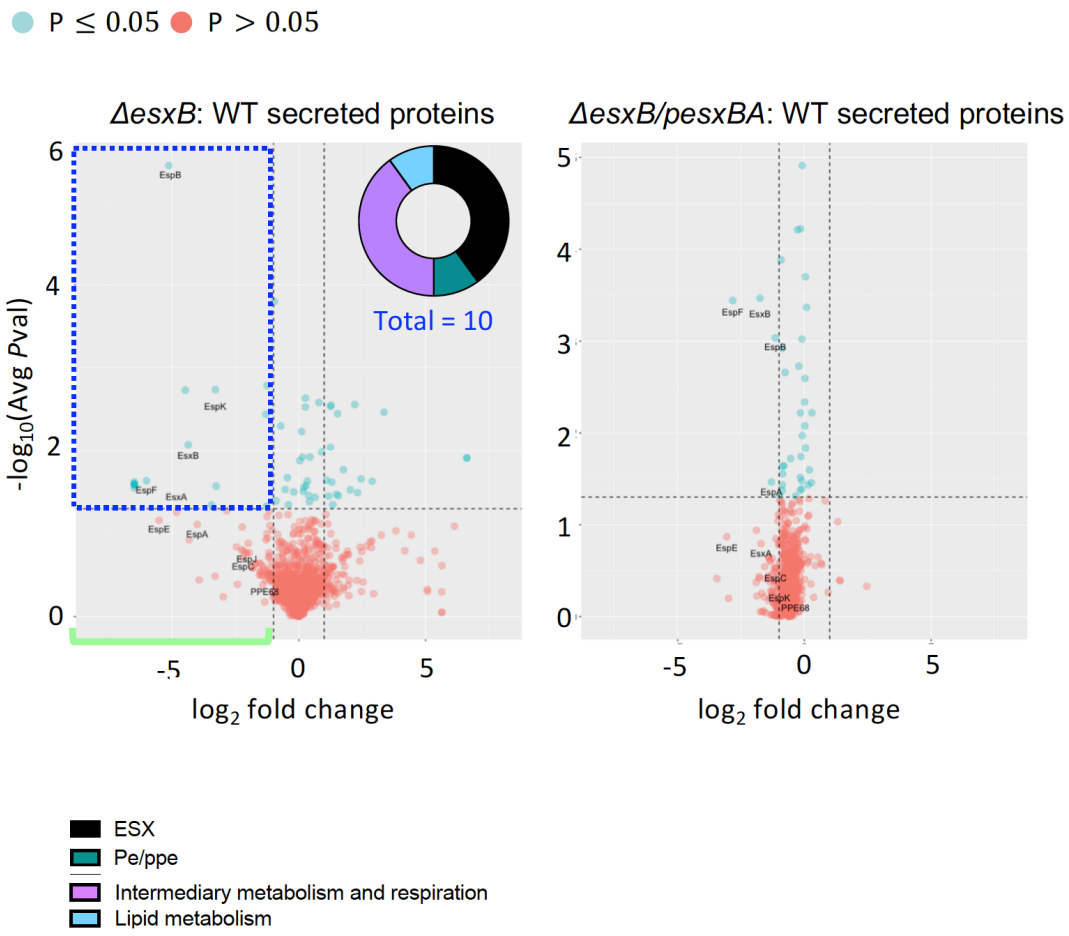


Figure S5. Analysis of the requirement of *esxB* for protein secretion (Related to Figure 3B). Volcano plots of proteins measured in the Δ*esxB* and Δ*esxB* complemented strains. The $\log_2$ fold change of secretion levels from each strain as compared to the WT strain is plotted against the $\log_{10}$ of the average *P* value. Horizontal black dashed line signifies a *P* value of.05. Blue dots have a *P* value ≤.05, red dots have a *P* value of >.05. Vertical black dashed lines signify a $\log_2$ fold change of -1. Known ESX-1 substrates are labeled. The donut plot is a functional analyses of *M. marinum* proteins in the quadrant indicated by the blue dashed lines based upon the functional annotation in Mycobrowser (6). The green bracket at the bottom left of the volcano plot signifies the proteins used in Figure 3B.

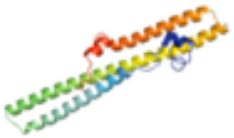
Protein Name	Predicted Functional Category: predicted product	Conserved protein in *M. tb*	High confidence structural prediction
MMAR_0431	PE/PGRS; PE-PGRS family protein	No	
MMAR_3064	Conserved hypotheticals; conserved membrane protein	No	
MMAR_3570	PE/PPE; PE-PGRS family protein	No	
MMAR_3763	PE/PPE; PE-PGRS family protein	Rv1450	
MMAR_4953	PE/PPE; PE-PGRS family protein	Rv0754	

Figure S6 Structural Predictions for non-ESX-1 Core Proteins with N-terminal Helices. Sequences for each predicted protein were obtained from Mycobrowser (6). Using these sequences, protein structures were predicted using Robetta online user interface, program RoseTTAFold (25). All proteins were modeled in their entirety, except MMAR_3763. Only the model of the first 1136 amino acids is shown, due to the size of the protein.

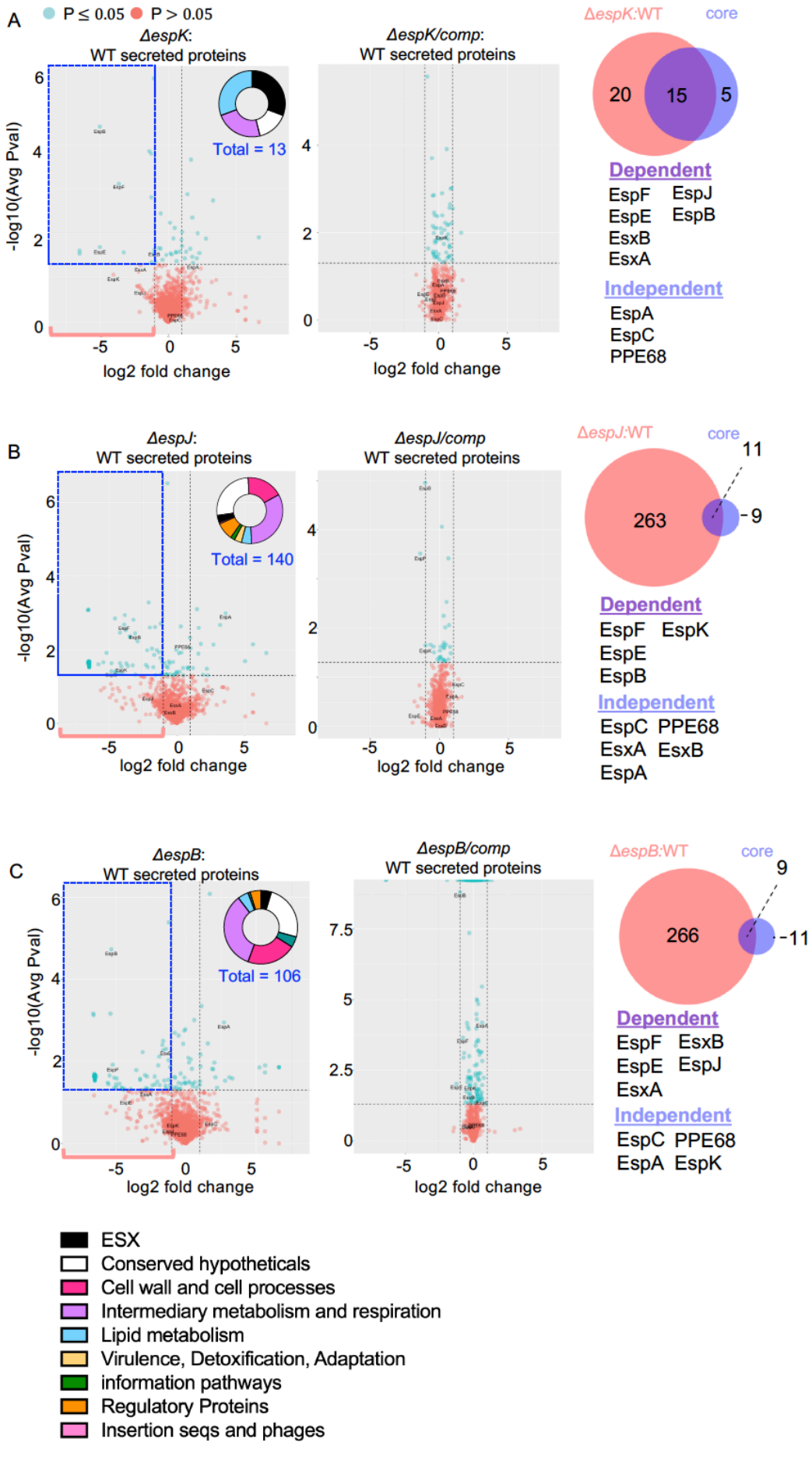

A
P ≤ 0.05
P > 0.05
ΔespK:
WT secreted proteins
ΔespK/comp:
WT secreted proteins
-log10(Avg Pval)
log2 fold change
Total = 13
ΔespK:WT
core
20
15
5
Dependent
EspF EspJ
EspE EspB
EsxB
EsxA
Independent
EspA
EspC
PPE68
B
ΔespJ:
WT secreted proteins
ΔespJ/comp
WT secreted proteins
Total = 140
ΔespJ:WT
core
263
11
9
Dependent
EspF EspK
EspE
EspB
Independent
EspC PPE68
EsxA EsxB
EspA
C
ΔespB:
WT secreted proteins
ΔespB/comp
WT secreted proteins
Total = 106
ΔespB:WT
core
266
9
11
Dependent
EspF EsxB
EspE EspJ
EsxA
Independent
EspC PPE68
EspA EspK
ESX
Conserved hypotheticals
Cell wall and cell processes
Intermediary metabolism and respiration
Lipid metabolism
Virulence, Detoxification, Adaptation
information pathways
Regulatory Proteins
Insertion seqs and phages

Figure S7. Group II Substrates: Analysis of requirement for protein secretion. Volcano plots of proteins measured in the **A.** Δ*espK* and Δ*espK* complemented strains, **B.** Δ*espJ* and Δ*espJ* complemented strains and **C.** Δ*espB* and Δ*espB* complemented strains. The $\log_2$ fold change of secretion levels from each strain as compared to the WT strain is plotted against the $\log_{10}$ of the average *P* value. Horizontal black dashed line signifies a *P* value of .05. Blue dots have a *P* value ≤.05, red dots have a *P* >.05. Vertical black dashed lines signify a $\log_2$ fold change of -1, and 1. Known ESX-1 substrates are labeled. The donut plot is a functional analyses of *M. marinum* proteins in the quadrant indicated by the blue dashed lines based upon the functional annotation in Mycobrowser (6). Venn diagrams were generated as described in the Supplemental Materials and Methods.

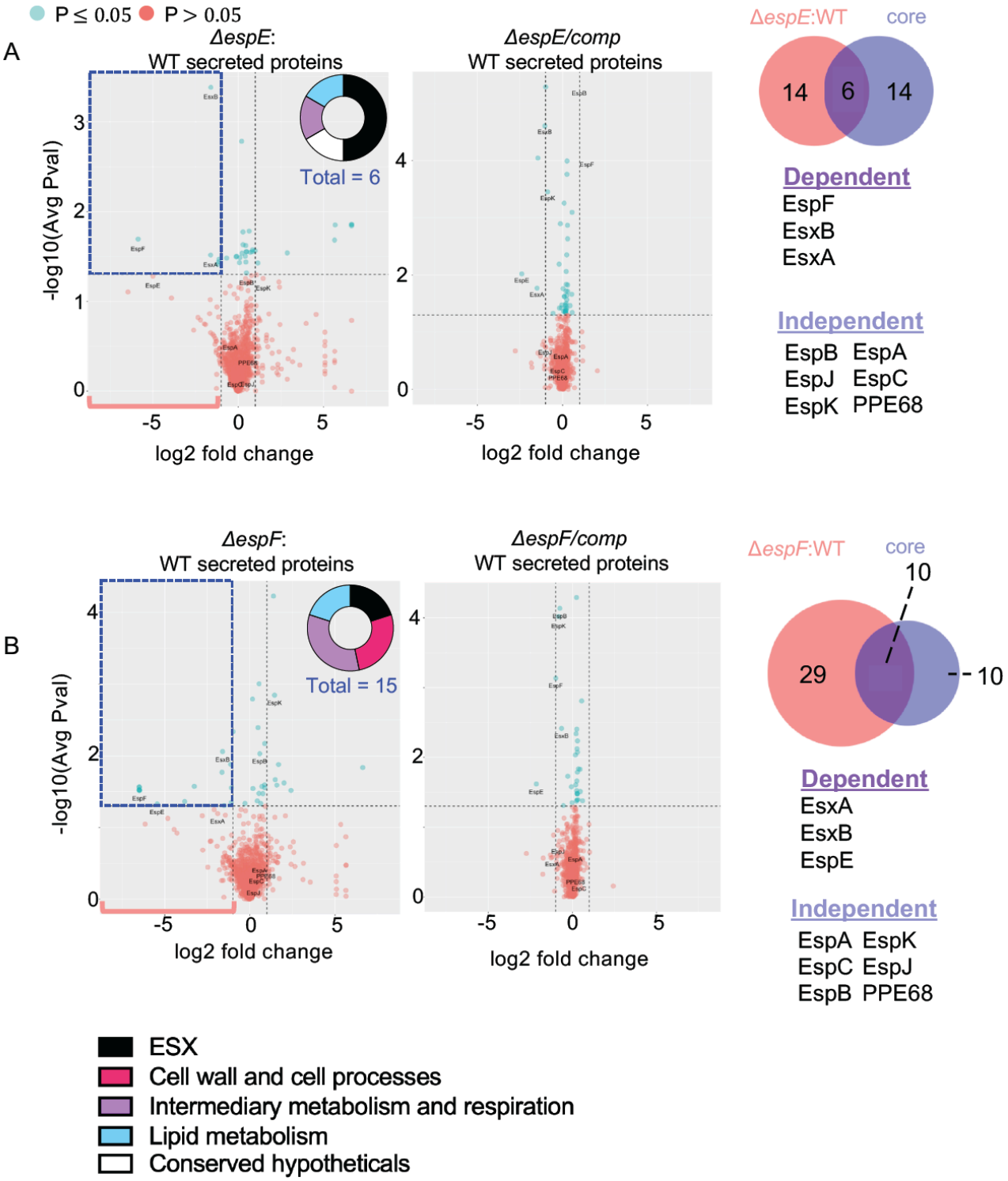

P ≤ 0.05
P > 0.05
A
ΔespE:
WT secreted proteins
ΔespE/comp
WT secreted proteins
-log10(Avg Pval)
log2 fold change
Total = 6
ΔespE:WT
core
14
6
14
Dependent
EspF
EsxB
EsxA
Independent
EspB EspA
EspJ EspC
EspK PPE68
B
ΔespF:
WT secreted proteins
ΔespF/comp
WT secreted proteins
Total = 15
ΔespF:WT
core
10
29
10
Dependent
EsxA
EsxB
EspE
Independent
EspA EspK
EspC EspJ
EspB PPE68
ESX
Cell wall and cell processes
Intermediary metabolism and respiration
Lipid metabolism
Conserved hypotheticals

Figure S8: Group III Substrates: Analysis of requirement for protein secretion. Volcano plots of proteins measured in the **A.** Δ *espE* and Δ*espE* complemented strains, **B.** Δ*espF* and Δ*espF* complemented strains. The $\log_2$ fold change of secretion levels from each strain as compared to the WT strain is plotted against the $\log_{10}$ of the average *P* value. Horizontal black dashed line signifies a *P* value of .05. Blue dots have a *P* value $\leq$.05, red dots have a *P* value of >.05. Vertical black dashed lines signify a $\log_2$ fold change of -1, and 1. Known ESX-1 substrates are labeled. The donut plot is a functional analyses of *M. marinum* proteins in the quadrant indicated by the blue dashed lines based upon the functional annotation in Mycobrowser (6). Venn diagrams were generated as described in the Supplemental Materials and Methods.

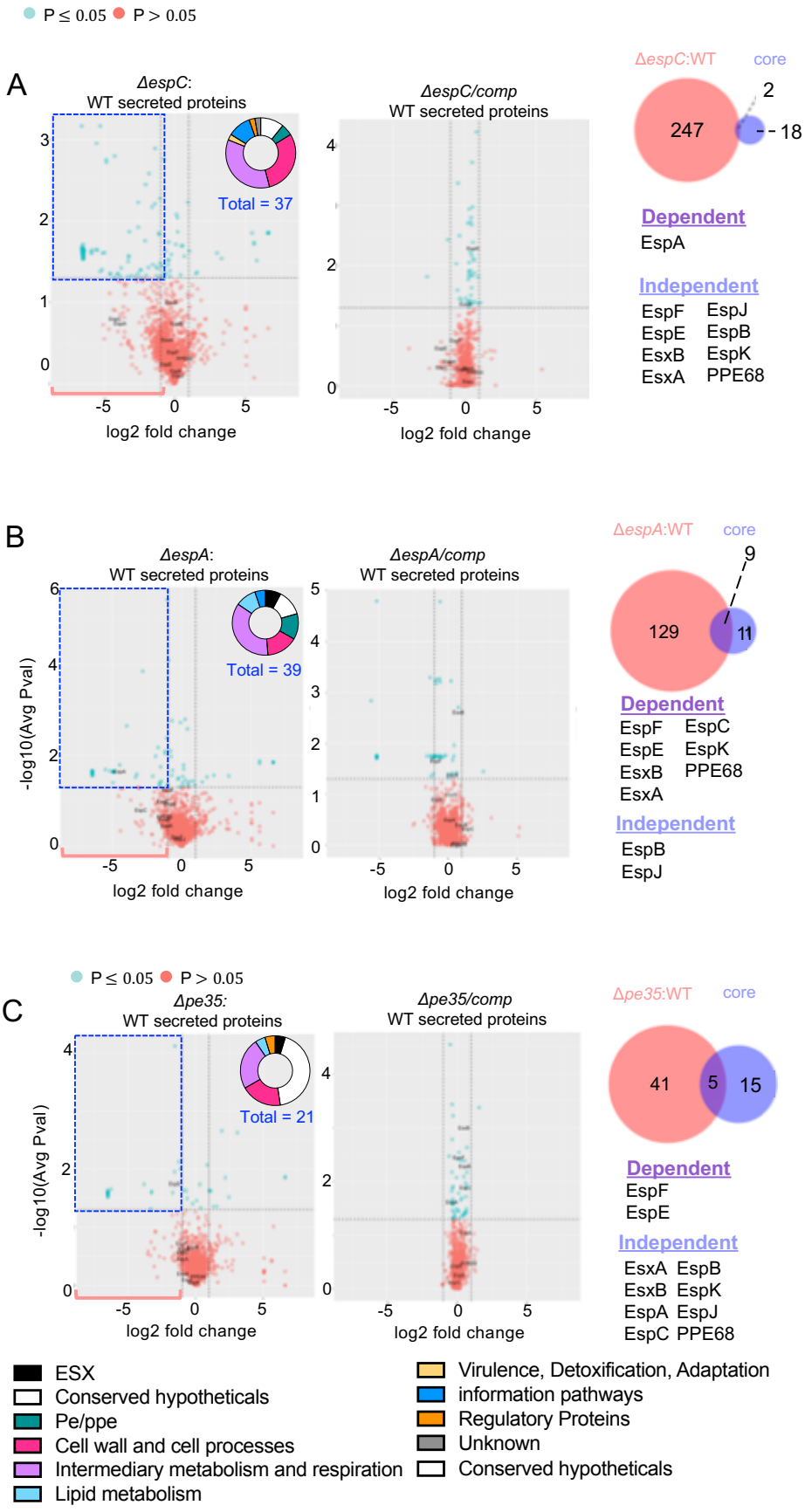

P ≤ 0.05 P > 0.05
A
ΔespC:
WT secreted proteins
Total = 37
log2 fold change
ΔespC/comp
WT secreted proteins
log2 fold change
ΔespC:WT
core
247
2
18
Dependent
EspA
Independent
EspF EspJ
EspE EspB
EsxB EspK
EsxA PPE68
B
ΔespA:
WT secreted proteins
Total = 39
-log10(Avg Pval)
log2 fold change
ΔespA/comp
WT secreted proteins
log2 fold change
ΔespA:WT
core
129
9
11
Dependent
EspF EspC
EspE EspK
EsxB PPE68
EsxA
Independent
EspB
EspJ
C
P ≤ 0.05 P > 0.05
Δpe35:
WT secreted proteins
Total = 21
-log10(Avg Pval)
log2 fold change
Δpe35/comp
WT secreted proteins
log2 fold change
Δpe35:WT
core
41
5
15
Dependent
EspF
EspE
Independent
EsxA EspB
EsxB EspK
EspA EspJ
EspC PPE68
ESX
Conserved hypotheticals
Pe/ppe
Cell wall and cell processes
Intermediary metabolism and respiration
Lipid metabolism
Virulence, Detoxification, Adaptation
information pathways
Regulatory Proteins
Unknown
Conserved hypotheticals

Figure S9: Group IV Substrates: Analysis of requirement for protein secretion. Volcano plots of proteins measured in the **A.** Δ*espC* and Δ*espC* complemented strains, **B.** Δ*espA* and Δ*espA* complemented strains, and **C.** Δ*pe35* and Δ*pe35* complemented strains. The $\log_2$ fold change of secretion levels from each strain as compared to the WT strain is plotted against the $\log_{10}$ of the average *P* value. Horizontal black dashed line signifies a *P* value of.05. Blue dots have a *P* value $\leq$.05, red dots have a *P* value of >.05. Vertical black dashed lines signify a $\log_2$ fold change of -1, and 1. Known ESX-1 substrates are labeled. The donut plot is a functional analyses of *M. marinum* proteins in the quadrant indicated by the blue dashed lines based upon the functional annotation in Mycobrowser (6). Venn diagrams were generated as described in the Supplemental Materials and Methods.

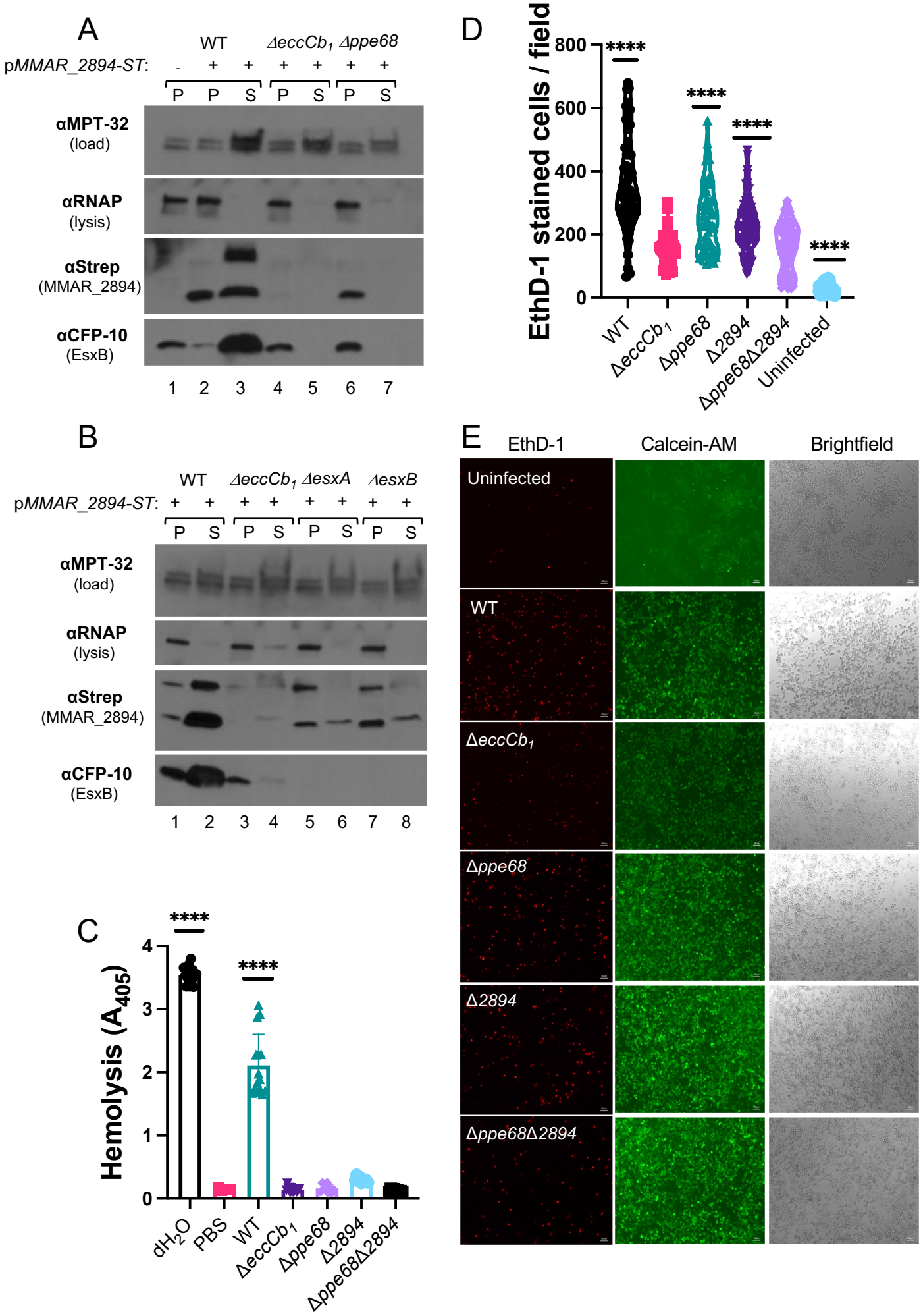

A
WT
ΔeccCb1 Δppe68
pMMAR_2894-ST:
P P S P S P S
αMPT-32
(load)
αRNAP
(lysis)
αStrep
(MMAR_2894)
αCFP-10
(EsxB)
1 2 3 4 5 6 7
B
WT
ΔeccCb1 ΔesxA ΔesxB
pMMAR_2894-ST:
P S P S P S P S
αMPT-32
(load)
αRNAP
(lysis)
αStrep
(MMAR_2894)
αCFP-10
(EsxB)
1 2 3 4 5 6 7 8
C
Hemolysis (A405)
****
****
dH2O
PBS
WT
ΔeccCb1
Δppe68
Δ2894
Δppe68Δ2894
D
EthD-1 stained cells / field
****
****
****
****
WT
ΔeccCb1
Δppe68
Δ2894
Δppe68Δ2894
Uninfected
E
EthD-1
Calcein-AM
Brightfield
Uninfected
WT
ΔeccCb1
Δppe68
Δ2894
Δppe68Δ2894

Figure S10: Supporting data assigning MMAR_2894 to Group I Substrates. A. PPE68 is required for the secretion of MMAR_2894-ST. **B.** EsxA and EsxB are dispensable for the secretion of MMAR_2894-ST. For A and B, secretion assays were performed as described in the Materials and Methods. P= cell -associated fraction, S = secreted fraction. 10µg of protein was loaded in each lane. Western blots are representative of three biological replicates. **C.** Hemolysis assay. Each data point is a technical replicate. Three biological replicates, each in technical triplicate were performed. Error bars represent standard deviation. Significance was determined using an ordinary one-way ANOVA (P< .0001), followed by a Tukey's multiple comparison test. Significance is shown relative to the Δ*eccCb*$_1$ strain (****, P< .0001). **D.** Cytolysis assay. Data are the average of at least four biological replicates, each in technical triplicate. EthD-1-stained cells were counted from 5 fields from each well, represented by dots. Statistical significance was determined using a one-way ordinary ANOVA (P<.0001) followed by a Tukey's multiple comparison test. Significance shown is compared to the Δ*eccCb*$_1$ strain. The cytolytic activity of the Δ*eccCb*$_1$ strain was not significantly different from Δ*ppe68*Δ*MMAR_2894* strain. ****, P<.0001. **E.** Representative images for RAW 264.7 macrophage infections from panel D. Scale bar represents 50µm.

Table S1: Strains and Plasmids used in this study.

Mycobacterial strains		
Name	**Genotype**	**Reference**
M. marinum M	Wild type; parent strain for all strains in this study	ATCC:BAA-545
Δ*eccCb$_1$*	Unmarked deletion of the *eccCb$_1$* gene (*MMAR_5446)*	(12)
Δ*esxA*	Unmarked deletion of the *esxA* gene (*MMAR_5450)*	This study
Δ*esxB*	Unmarked deletion of the *esxB* gene (*MMAR_5449)*	This study
Δ*espA*	Unmarked deletion of the *espA* gene (*MMAR_4166)*	This study
Δ*espB*	Unmarked deletion of the *espB* gene (*MMAR_5457)*	(8)
Δ*espC*	Unmarked deletion of the *espC* gene (*MMAR_4167)*	This study
Δ*espE*	Unmarked deletion of the *espE* gene (*MMAR_5439)*	(8)
Δ*espF*	Unmarked deletion of the *espF* gene (*MMAR_5440)*	(8)
Δ*espJ*	Unmarked deletion of the *espJ* gene (*MMAR_5453)*	(8)
Δ*espK*	Unmarked deletion of the *espK* gene (*MMAR_5455)*	(8)
Δ*pe35*	Unmarked deletion of the *pe35* gene (*MMAR_5447)*	This study
Δ*ppe68*	Unmarked deletion of the *ppe68* gene (*MMAR_5448)*	This study
Δ*MMAR_2894*	Unmarked deletion of the *MMAR_2894* gene	(4)
WT*/pMV306*	M strain with *pMV306* empty vector integrated at *attB*	This study
Δ*eccCb$_1$/pMV306*	Δ*eccCb$_1$* with *pMV306* empty vector integrated at *attB*	This study
Δ*esxA/p$_{MOP}$_esxBA*	Δ*esxA* with *p$_{MOP}$_esxBA* integrated at *attB*	This study
Δ*esxB/p$_{MOP}$_esxBA*	Δ*esxB* with *p$_{MOP}$_esxBA* integrated at *attB*	This study
Δ*espA/p$_{MOP}$_espA*	Δ*espA* with *p$_{MOP}$_espA* integrated at *attB*	This study
Δ*espB/p$_{MOP}$_espB*	Δ*espB* with *p$_{MOP}$_espB* integrated at *attB*	This study
Δ*espC/p$_{MOP}$_espC*	Δ*espC* with *p$_{MOP}$_espC* integrated at *attB*	This study
Δ*espE/pMV306_EFG$_{MM}$*	Δ*espE* with *pMV306_espEFG$_{MM}$* integrated at *attB*	(8)
Δ*espF/pMV306_EFG$_{MM}$*	Δ*espF* with *pMV306_espEFG$_{MM}$* integrated at *attB*	(8)
Δ*espJ/p$_{MOP}$_espJ*	Δ*espJ* with *p$_{MOP}$_espJ* integrated at *attB*	This study
Δ*espK/p$_{MOP}$_espK*	Δ*espK* with *p$_{MOP}$_espK* integrated at *attB*	This study
Δ*pe35/p$_{MOP}$_pe35*	Δ*pe35* with *p$_{MOP}$_pe35* integrated at *attB*	This study
Δ*ppe68/p$_{MOP}$_ppe68*	Δ*ppe68* with *p$_{MOP}$_ppe68* integrated at *attB*	This study
Δ*MMAR2894/p$_{MOP}$_2894-strep*	Δ*2894* with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	(4)
WT*/p$_{MOP}$_2894-strep*	M strain with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	(4)
Δ*eccCb$_1$/ p$_{MOP}$_2894-strep*	Δ*eccCb$_1$* with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	(4)
Δ*esxA/ p$_{MOP}$_2894-strep*	Δ*esxA* with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	This study
Δ*esxB/ p$_{MOP}$_2894-strep*	Δ*esxB* with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	This study
Δ*ppe68/ p$_{MOP}$_2894-strep*	Δ*ppe68* with *p$_{MOP}$_2894* Strep-tag integrated at *attB*	This study
Δ*ppe68*Δ*MMAR2894*	Unmarked deletion of the *ppe68* (*MMAR_5448)* and *MMAR_2894* genes	This study
Plasmids Used in This Study		
Name	**Genotype**	**Reference**
p2NIL	kan^R , amp^R , oriE; Parental vector for allelic exchange	(11)
pGOAL19	amp^R, contains the selectable GOAL cassette includes hyg^R , *lacZ*+, *sacB*, *oriE*; The GOAL cassette is integrated into the p2NIL plasmid.	(11)

p2NIL∆*eccCb*$_1$GOAL	Allelic exchange plasmid to generate the ∆*eccCb*$_1$ strain; contains *M. marinum eccCb*$_1$ flanking regions (NC 010612.1:6586249…6587725;6591053)	(12)
p2NIL∆*esxA*GOAL	Allelic exchange plasmid to generate the ∆*esxA* strain; contains *M. marinum esxA* flanking regions (NC010612.1:6589566...6591514;6591764...6593330)	This study
p2NIL∆*esxB*GOAL	Allelic exchange plasmid to generate the ∆*esxB* strain; contains *M. marinum esxB* flanking regions (NC010612.1:6589566...6591175;6591443...6593330)	This study
p2NIL∆*espA*GOAL	Allelic exchange plasmid to generate the ∆*espA* strain; contains *M. marinum espA* flanking regions (NC010612.1:5126969...5128565;5129726...5131224)	This study
p2NIL∆*espB*GOAL	Allelic exchange plasmid to generate the ∆*espB* strain; contains *M. marinum espB* flanking regions (NC010612.1:6599907...6601456;6602820...6604411)	(8)
p2NIL∆*espC*GOAL	Allelic exchange plasmid to generate the ∆*espC* strain; contains *M. marinum espC* flanking regions (NC010612.1:5128342...5129806;5130089...5131424)	This study
p2NIL∆*espE*GOAL	Allelic exchange plasmid to generate the ∆*espE* strain; contains *M. marinum espE* flanking regions (NC 010612:6577412...6579158;6580390...6582018)	(8)
p2NIL∆*espF*GOAL	Allelic exchange plasmid to generate the ∆*espF* strain; contains *M. marinum espF* flanking regions (NC 010612:6579019...6580518;6580805...6582166)	(8)
p2NIL∆*espJ*GOAL	Allelic exchange plasmid to generate the ∆*espJ* strain; contains *M. marinum espJ* flanking regions (NC010612.1:6594681...6596198;6597034...6598635)	(8)
p2NIL∆*espK*GOAL	Allelic exchange plasmid to generate the ∆*espK* strain; contains *M. marinum espK* flanking regions (NC 010612.1:658234...6596734;6600663...6599170)	(8)
p2NIL∆*pe35*GOAL	Allelic exchange plasmid to generate the ∆*pe35* strain; contains *M. marinum pe35* flanking regions (NC010612.1:6587990…6589642;6589904…6591577)	This study
p2NIL∆*ppe68*GOAL	Allelic exchange plasmid to generate the ∆*ppe68* strain; contains *M. marinum ppe68* flanking regions (NC 010612.1:6588436…6589954;6591055…6592631)	This study
p2NIL∆*2894*GOAL	Allelic exchange plasmid to generate the ∆*2894* strain; contains *M. marinum 2894* flanking regions (NC 010612.1:3500899...3499384;3498720...3497211)	(4)
pMV306Hyg	Promoterless shuttle vector, oriE, L5 integrase, hygR . attB integration.	(26)
p$_{MOP}$_*esxBA*	*M. marinum esxB* and *esxA* genes behind the pMOP promoter. *hyg*^{R} . *attB* integration.	(27)
p$_{MOP}$_*espA*	*M. marinum espA* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*espB*	*M. marinum espB* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*espC*	*M. marinum espC* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study

pMV306_*EFG*$_{MM}$	*espE*, *espF*, *espG* genes from *M. marinum* (NC 010612.1:6577497...6581787) in pMV306. *oriE*, L5 integrase, *hyg*^{R} . a*ttB* integration.	(8)
p$_{MOP}$_*espJ*	*M. marinum espJ* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*espK*	*M. marinum espK* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*pe35*	*M. marinum pe35* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*ppe68*	*M. marinum ppe68* gene behind the pMOP promoter. *hyg*^{R}. *attB* integration.	This study
p$_{MOP}$_*2894-strep*	*M. marinum 2894* gene, 3' streptactin-tagged behind the pMOP promoter. *hyg*^{R} . *attB* integration.	(4)

Table S2: Oligonucleotides used in this study.

Name	Sequence (5'→3')	Application & Reference
orb48	TGGTGTCACGCTCGTGTCCAGCACCAGACCGAAAC	Amplifying flanking regions for *eccCb*$_1$ knockout, (27)
orb49	GTGATGCGTCGCTGGCAATGAAATAATCCTAA CCGGGGCTAGGTTCGGCAGTCATGAGC	
orb50	AGCCCCGGTTAGGATTATTTC	
orb51	GCAGTCAGGCACCGTTCGTCATCCCAGTCGTCTTC	
omf166	CGTGGTGTCACGCTCGTGCGACAACCAAATGAGGATTTGTCC	Amplifying flanking regions for *esxA* knockout, this study
omf563	CCGTGACCTTAAGCCACTGCTGTTCTGTCATGTCGTATTGC	
omf564	AGCAGTGGCTTAAGGTCACGGGGATGTTTGCTTAATCCCC	
omf169	CGCAGTCAGGCACCGTGTTTCGCCTCATCGGCTCG	
omf561	TTGCGACTTAAGGGTCTTCATCTCTGCCATGCTGG	Used with omf166 and omf169, amplifying flanking regions for *esxB* knockout, this study
omf562	GAAGACCCTTAAGTCGCAAATGGGCTTCTGATTCC	
olc11	TGGTGTCACGCTCGTGCCAAATCGCAGCGCTCTAC	Amplifying flanking regions for *espA* knockout, this study
orc19	GGGGCTAGCCTTAAGTCATGGCTACTACCTGCCCAAC	
orc20	TAGCCATGACTTAAGGCTAGCCCCACCTGACC	
orc21	GCAGTCAGGCACCGTGATTTCGCCGTAGTCGTG	
omf210	CGTGGTGTCACGCTCGTCTGACGCTGGTGCTCGGATTGG	Amplifying flanking regions for *espB* knockout, (8)
omf211	GTCCTGACGCTTAAGCTGCGGCTGGCTCATGTCGG	
omf212	CAGCCGCAGCTTAAGCGTCAGGACAACAAGTAGCCGTCAGG	
omf213	ACGCAGTCAGGCACCGTGAGTACCCGGGATACCCGGGATACC	
orb42	TGGTGTCACGCTCGTAGCAACCCGACACGACACTC	Amplifying flanking regions for *espC* knockout, this study
orb43	GAATAACGACTTAAGCAGGTTCTCCGTCATTGGTG	
orb44	CTTAAGTCGTTATTCAGCTGAACAGC	
orb45	GCAGTCAGGCACCGTAAATCGTCGCGGGTGATCAG	
olc179	CGTGGTGTCACGCTCGTTCTGCCCGAGGATCGTATAG	Amplifying flanking regions for *espE* knockout, (8)
olc180	CTAGAGGAGGGTCTTAAGCTTTGGCACCATTCGCG	
olc181	ATGGTGCCAAAGCTTAAGACCCTCCTCTAGTCCGGGC	
olc182	CGCAGTCAGGCACCGTAAGATCTCTTCGGCCAGAGC	
orb34	TGGTGTCACGCTCGTAAAGTGCCCGCACTACCAAG	Amplifying flanking regions for *espF* knockout, (8)
orb35	GCAGTCAGGCACCGTCGGCAAGTTGAGCGTCATAC	
orb36	GGTAGCGAGCGGACCGGTCATATCAGCCAAAA	
orb37	ATCTTGTCCTTAAGGTTCAGTAGTCCTGTCATGG CTTAAGGACAA-GATTTTTGGCTGATATG	
omf435	CGTGGTGTCACGCTCGTACGAGACCGTCGCGATCTTGGCC	Amplifying flanking regions for *espJ* knockout, (8)
omf436	CGCTCTTAAGCAGAGGCTCAGCCATCGCAGAATCCC	
omf437	CCTCTGCTTAAGAGCGCTCCCGTCGAAAGCACCG	
omf438	ACGCAGTCAGGCACCGTGAGACCGAAGATCTGCGGCTGCTGG	
omf457	CGTGGTGTCACGCTCGTTCGATGGACCAGCAGAAGGAGTGG	Amplifying flanking regions for *espK* knockout, (8)
omf458	AAGCAGCTTAAGCGGCCTCGGAATACCCATTGC	
omf459	GGCCGCTTAAGCTGCTTGCCAACGCTTTGGCGG	
omf460	ACGCAGTCAGGCACCGTGCACAGCTCGTCTCGGACACCAAACC	
omf520	CGTGGTGTCACGCTCGTCTGCAGACGTTGGTCATGTCGG	
omf521	GACGCTCTTAAGGCCGTGTGACTTTTGTTCCATTTGC	

omf522	CACGGCCTTAAGAGCGTCATCGTGTAGAAAATCAGTCACC	Amplifying flanking regions for *pe35* knockout, this study
omf523	ACGCAGTCAGGCACCGTGGAATGAATGCTGGTGACATTTCCC	
olc198	TGGTGTCACGCTCGTCTACCTGGGCACCAAGATCG	Amplifying flanking regions for *ppe68* knockout, this study
olc199	CAGTCGTCCTTAAGGTGCCACAGCATGGTGATCAC	
olc200	CTGTGGCACCTTAAGGACGACTGGTGAGCGTCGTC	
olc201	GCAGTCAGGCACCGTTAGCGGTGACTTCGGTGGTG	
orb130	TGGTGTCACGCTCGTCCATCGCATAAAGCGCTGTC	Amplifying flanking regions for *2894* knockout, (4)
orb131	TGACATGGGATCCATTTGCACG	
orb132	AATGGATCCCATGTCAGAACGCTTGATGTAGCGGGC	
orb133	GCAGTCAGGCACCGTTGCCACCCGAATAGGCTAGG	
omf170	TCGTCAACACGAACAGACTTCCC	Verification of Δ*esxA* and Δ*esxB* strains, this study
omf171	GTCATCTGGAGGTCCGGAACC	
orc34	TTGATCACCAAGCCGAGGAC	Verification of Δ*espA* strain, this study
orc35	TTAGCCCAGAGGTGACTTCG	
omf214	TAATCGGCGAGACCGTGAACC	Verification of Δ*espB* strain, this study
omf215	GTCCAAGGCCGACTGGAACC	
olc9	GCGACCAGCAAGAAGTACTC	Verification of Δ*espC* strain, this study
olc10	GTCGAGCATGGCGAAGTTGG	
olc183	ACGGTCGGAGCACTGCTAAG	Verification of Δ*espE* strain, this study
olc184	TCGGTCTCGTCGCTCAGTTC	
olc1	TGGCAACTCACCTTGAATCG	Verification of Δ*espF* strain, this study
olc2	TACCTCGACTCCGACGACATC	
omf439	AAACATCCGATTCTGAGTCACCGGC	Verification of Δ*espJ* strain, this study
omf440	GTTGGATGTCGCCGGTAACACC	
omf461	GTACCACTGACCAGCTGACGCG	Verification of Δ*espK* strain, this study
omf462	ATAACTTACTTTCGAAAAACGCACTGGCAGCG	
omf524	GATTATTTCATTGCCAGCGACGCATCACCC	Verification of Δ*pe35* strain, this study
omf525	AGTTCAAGCGCGCGGTCAATTCG	
orc32	GCTCAGCGGGAGATCCACC	Verification of Δ*ppe68* strain, this study
orc33	CGGAGATCCGCTCGAAATTACCTG	
orb150	TCGTATGCGACGGCTGTTTG	Verification of Δ*MMAR_2894* strain, (4)
orb151	TTGGCGACGGTCTTGAACCC	
orb38	ATTCAGGAGTCCAGCATGGCAGAGATGAAGACCGATG	Amplifying the *esxBA* genes for p_{MOP}_*esxBA* plasmid construction, this study
orb39	GCCTGAGCGGTCCCGTTAAGCAAACATCCCCGTGAC	
occ86	AGGAGTCCAGCCATATGAGCAGAGCCTTCATCATCGATCCCAC-GATTAGC	Amplifying the *espA* gene for p_{MOP}_*espA* plasmid construction, this study
occ87	GCCTGAGCGGTCCCGACTAGTCTAGCCGGCGCGTCGCTCCG	
occ94	AGGAGTCCAGCCATATGACGGAGAACCTGAAAGTGCAGCC	Amplifying the *espC* gene for p_{MOP}_*espC* plasmid construction, this study
occ95	GAGCGGTCCCGACTAGTTCAGCTGAATAACGATCCGAGCGCC	

omf301	TTCAGGAGTCCAGCCATATGAGCCAGCCGCAGACCGTC	Amplifying the *espB* gene for p_{MOP}_*espB* plasmid construction, this study
omf302	CGAACTGCGGGTGGCTCCATCCCTACTTGTTGTCCTGAC-GGCGGCG	
olc7	GAGTGAGCTGATACCGGCGCGATGGGCCATGATAAG	Amplifying *M. marinum espE-espG* genes for pMV306_EFG_{MM} plasmid construction, (8)
olc8	ATTCTGCAGCTGGATAGTGCACCGAGTCCAGAGTC	
omf477	AGGAGTCCAGCCATATGGCTGAGCCTCTGGCCGTCG	Amplifying the *espJ* gene for p_{MOP}_*espJ* plasmid construction, this study
omf478	ACGCCTGAGCGGTCCCGACTAGTCAGATCGGAGCTGAGAC-CGAACCC	
omf479	AGGAGTCCAGCCATATGGGTATTCCGAGGCCGACGGG	Amplifying the *espK* gene for p_{MOP}_*espK* plasmid construction, this study
omf480	ACGCCTGAGCGGTCCCGACTAGTCAGGACGCGTCCGCCAAAGC	
okn27	AGGAGTCCAGCCATATGGAACAAAAGTCACACG	Amplifying the *pe35* gene for p_{MOP}_*pe35* plasmid construction, this study
okn28	GCCTGAGCGGTCCCGACTAGTCTACACGATGACGCTGG	
occ88	AGGAGTCCAGCCATATGCTGTGGCACGCAATGCCACC	Amplifying the *ppe68* gene for p_{MOP}_*ppe68* plasmid construction, this study
occ89	GCCTGAGCGGTCCCGACTAGTTCACCAG-TCGTCCTCGTCATCCCAGTCG	
orb74	CAGGAGTCCAGCCATATGGATCCCATGTCACACGACC	Amplifying the *MMAR_2894-ST* gene for p_{MOP}_*2894-ST* plasmid construction, (4)
orb75	TGCGGGTGGCTCCATCCCATCAAGCGTTCGACCGGG	

Table S3: LC-MS/MS Sample Key, Labeling Table

Biological Replicate	Strain	Sample ID	iTRAQ Reporter
Deletion 1	M Strain (WT) 1	a-1	113
Deletion 1	M Strain (WT) 2	a-A	113
Deletion 1	Δ*eccCb*$_1$ 1	a-1	114
Deletion 1	Δ*eccCb*$_1$ 2	a-A	114
Deletion 1	Δ*MMAR_2894*	a-1	115
Deletion 1	Δ*espA*	a-A	115
Deletion 1	Δ*espB*	a-A	116
Deletion 1	Δ*espC*	a-A	117
Deletion 1	Δ*espE*	a-A	118
Deletion 1	Δ*espF*	a-1	116
Deletion 1	Δ*espJ*	a-1	117
Deletion 1	Δ*espK*	a-1	118
Deletion 1	Δ*esxA*	a-1	119
Deletion 1	Δ*esxB*	a-1	121
Deletion 1	Δ*pe35*	a-A	119
Deletion 1	Δ*ppe68*	a-A	121
Deletion 2	M Strain (WT) 1	b-1	113
Deletion 2	M Strain (WT) 2	b-A	113
Deletion 2	Δ*eccCb*$_1$ 1	b-1	114
Deletion 2	Δ*eccCb*$_1$ 2	b-A	114
Deletion 2	Δ*MMAR_2894*	b-1	115
Deletion 2	Δ*espA*	b-1	116
Deletion 2	Δ*espB*	b-1	117
Deletion 2	Δ*espC*	b-1	118
Deletion 2	Δ*espE*	b-1	119
Deletion 2	Δ*espF*	b-1	121
Deletion 2	Δ*espJ*	b-A	115
Deletion 2	Δ*espK*	b-A	116
Deletion 2	Δ*esxA*	b-A	117
Deletion 2	Δ*esxB*	b-A	118
Deletion 2	Δ*pe35*	b-A	119
Deletion 2	Δ*ppe68*	b-A	121
Deletion 3	M Strain (WT) 1	d-1	113
Deletion 3	M Strain (WT) 2	d-A	113
Deletion 3	Δ*eccCb*$_1$ 1	d-1	114
Deletion 3	Δ*eccCb*$_1$ 2	d-A	114
Deletion 3	Δ*MMAR_2894*	d-1	115
Deletion 3	Δ*espA*	d-1	116
Deletion 3	Δ*espB*	d-1	117
Deletion 3	Δ*espC*	d-1	118

Deletion 3	Δ*espE*	d-1	119
Deletion 3	Δ*espF*	d-1	121
Deletion 3	Δ*espJ*	d-A	115
Deletion 3	Δ*espK*	d-A	116
Deletion 3	Δ*esxA*	d-A	117
Deletion 3	Δ*esxB*	d-A	118
Deletion 3	Δ*pe35*	d-A	119
Deletion 3	Δ*ppe68*	d-A	121

Table S4: Summary of Hemolysis and Cytolysis Data from Figure 1.

low / middle / high	Hemolytic Activity (Avg +/- St. Dev)*	*P* (vs WT)	*P* (vs Δ*eccCb*$_1$)	Cytotoxicity (Med, Q1, Q3) †	*P* (vs WT)	*P* (vs Δ*eccCb*$_1$)
Controls						
dH_20	3.65 +/- .115	<0.0001	<0.0001	-	-	-
PBS	.159 +/-.030	<0.0001	ns	-	-	-
Uninfected	-	-	-	39.5 27, 52.75	<0.0001	0.0009
***M. marinum* strains**						
WT	2.91 +/- .506	-	<0.0001	427 339,520	-	<0.0001
Δ*eccCb*$_1$	.143 +/- .050	<0.0001	-	112 88,129	<0.0001	-
Δ*esxA*	.077 +/- .007	<0.0001	ns	108, 88, 151	<0.0001	ns
Δ*esxA*/ comp	2.91 +/- .102	ns	<0.0001	380, 313, 428	ns	<0.0001
Δ*esxB*	.074 +/- .013	<0.0001	ns	101, 83, 120	<0.0001	ns
Δ*esxB*/ comp	1.79 +/- .200	<0.0001	<0.0001	405, 265.5, 528.5	ns	<0.0001
Δ*espA*	2.54 +/- .263	0.0247	<0.0001	368, 284, 466	ns	<0.0001
Δ*espA*/ comp	3.57 +/- .136	<0.0001	<0.0001	453, 245, 580.25	ns	<0.0001
Δ*espB*	.069 +/- .011	<0.0001	ns	284, 230, 369	<0.0001	<0.0001
Δ*espB*/ comp	2.74 +/- .520	ns	<0.0001	267, 177.25, 404.25	<0.0001	<0.0001
Δ*espC*	3.04 +/- .173	ns	<0.0001	221, 175, 311	<0.0001	<0.0001
Δ*espC*/ comp	3.38 +/- .113	0.0002	<0.0001	241.5, 139, 374	<0.0001	<0.0001
Δ*espE*	.073 +/- .013	<0.0001	ns	233 194, 304	<0.0001	<0.0001
Δ*espE*/ comp	3.00 +/- .306	ns	<0.0001	342.5 266, 437	ns	<0.0001
Δ*espF*	.942 +/- .284	<0.0001	<0.0001	190 143, 263	<0.0001	0.0053
Δ*espF*/ comp	2.98 +/- .195	ns	<0.0001	287.5 228.75, 346.75	<0.0001	<0.0001
Δ*espJ*	1.29 +/- .142	<0.0001	<0.0001	281 205, 385	<0.0001	<0.0001
Δ*espJ*/ comp	3.39 +/- .037	0.0001	<0.0001	472, 355, 604	ns	<0.0001
Δ*espK*	.918 +/- .523	<0.0001	<0.0001	153 113, 198	<0.0001	ns
Δ*espK*/ comp	3.29 +/- .029	0.0112	<0.0001	336, 242.5, 414	0.0043	<0.0001
Δ*pe35*	2.99 +/- .202	ns	<0.0001	277 218, 385	<0.0001	<0.0001
Δ*pe35*/ comp	3.14 +/- .410	ns	<0.0001	268.5 219, 406.5	<0.0001	<0.0001
Δ*ppe68*	.081 +/- .020	<0.0001	ns	214 158,352	<0.0001	<0.0001
Δ*ppe68*/ comp	3.30 +/- .049	0.0066	<0.0001	513, 338.5, 665.5	ns	<0.0001
Δ*2894*	.227 +/- .044	<0.0001	ns	293 215, 345	<0.0001	<0.0001
Δ*2894*/ comp	2.78 +/- .146	ns	<0.0001	374, 325, 447	ns	<0.0001

*Averages, Standard Deviation and *P* values from Figure 1B. ns, not significant

†Median values, Quartile 1 and Quartile 3, and *P* values from Figure 1C.

Table S5. Comparison of *M. marinum* studies

Strain	EsxA Secretion*	EsxB Secretion	Complementation?#	Study	Approach
M6 (*espJ*::Tn)	-	+	no	Gao et al 2004	western
	+	+	no	Lienard et al 2020	western
	+/-	+/-	no	Champion et al 2014	MRM MS
Δ*espJ*	+	+	yes, $espJ_{MM}$	This study	iTRAQ MS
M7 (*espK*::Tn)	-	+	yes, $espK_{MT}$	Gao et al 2004	western
	+	+	no	Lienard et al 2020	western
	+/-	+/-	no	Champion et al 2014	MRM MS
Δ*espK*	+	+	yes	This study	iTRAQ MS
M8 (*espB*::Tn)	-	+/-	yes, $espL\text{-}espB_{MM}$, $espB_{MT}$	Gao et al 2004	western
	-	+/-	no	Lienard et al 2020	western
	-	-	no	Champion et al 2014	MRM MS
Δ*espB*	+	+	Yes, $espB_{MM}$	This study	iTRAQ MS
espE::Tn	-	+/-	no	Lienard et al 2020	western
Δ*espE*	+	+	yes, $espEFG_{MM}$, $espE_{MT}$	Chirakos et al, 2020	western
	+	+	yes, $espEFG_{MM}$	This study	iTRAQ MS

*negative for secretion (-), similar to WT for secretion (+), intermediate secretion (+/-)

MM subscript, *M. marinum* genes, MT subscript, *M. tuberculosis* genes

Dataset S1 (separate file): iTRAQ Proteomics Data

A. Untrimmed Data
B. Trimmed data used for heatmap in Figure 2A
C. Trimmed data used for Pearson Correlation in Figure 2C
D. P values for Pearson Correlation in Figure 2C
E. Trimmed data used for ESX-1 heatmap in Figure 5A
F. Relative changes in the secretion of ESX-1 substrates

Dataset S2 (separate file): Untrimmed Volcano Plot Proteomics Data and Functional Analyses, organized by strain.

SI References

1. D. Bottai *et al.*, Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol* **83**, 1195-1209 (2012).
2. M. M. Champion, E. A. Williams, R. S. Pinapati, P. A. Champion, Correlation of Phenotypic Profiles Using Targeted Proteomics Identifies Mycobacterial Esx-1 Substrates. *J Proteome Res* **3**, 5151-5164 (2014).
3. P. A. Champion, M. M. Champion, P. Manzanillo, J. S. Cox, ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. *Mol Microbiol* **73**, 950-962 (2009).
4. R. E. Bosserman, K. R. Nicholson, M. M. Champion, P. A. Champion, A New ESX-1 Substrate in *Mycobacterium marinum* That Is Required for Hemolysis but Not Host Cell Lysis. *J Bacteriol* **201** (2019).
5. B. M. Cumming, K. C. Chinta, V. P. Reddy, A. J. C. Steyn, Role of Ergothioneine in Microbial Physiology and Pathogenesis. *Antioxid Redox Signal* **28**, 431-444 (2018).
6. A. Kapopoulou, J. M. Lew, S. T. Cole, The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)* **91**, 8-13 (2011).
7. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786 (2011).
8. A. E. Chirakos, K. R. Nicholson, A. Huffman, P. A. Champion, Conserved ESX-1 substrates EspE and EspF are virulence factors that regulate gene expression. *Infect Immun* 10.1128/IAI.00289-20 (2020).
9. K. G. Sanchez *et al.*, EspM Is a Conserved Transcription Factor That Regulates Gene Expression in Response to the ESX-1 System. *mBio* **11** (2020).
10. W. Bitter *et al.*, Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog* **5**, e1000507 (2009).
11. T. Parish, N. G. Stoker, Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology* **146 (Pt 8)**, 1969-1975 (2000).
12. E. A. Williams *et al.*, A Nonsense Mutation in *Mycobacterium marinum* That Is Suppressible by a Novel Mechanism. *Infect Immun* **85** (2017).
13. C. Li *et al.*, FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol* **11**, 92 (2011).
14. R. a. P. Goude, "Electroporation of Mycobacteria" in Methods in Molecular Biology*,* T. P. a. D. M. Roberts, Ed. (Springer, 2015), chap. Electroporation in Mycobacteria, pp. 117-130.
15. L. Y. Gao *et al.*, A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* **53**, 1677-1693 (2004).
16. M. M. Champion, E. A. Williams, G. M. Kennedy, P. A. Champion, Direct detection of bacterial protein secretion using whole colony proteomics. *Mol Cell Proteomics* **11**, 596-604 (2012).
17. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

18. P. A. Champion, S. A. Stanley, M. M. Champion, E. J. Brown, J. S. Cox, C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis. Science* **313**, 1632-1636 (2006).
19. L. Sun *et al.*, Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. *Sci Rep* **4**, 4365 (2014).
20. E. H. Peuchen *et al.*, Phosphorylation Dynamics Dominate the Regulated Proteome during Early *Xenopus* Development. *Sci Rep* **7**, 15647 (2017).
21. I. V. Shilov *et al.*, The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* **6**, 1638-1655 (2007).
22. W. H. Tang, I. V. Shilov, S. L. Seymour, Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J Proteome Res* **7**, 3661-3667 (2008).
23. Z. Li *et al.*, Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res* **11**, 1582-1590 (2012).
24. W. H. Tang *et al.*, Discovering known and unanticipated protein modifications using MS/MS database searching. *Anal Chem* **77**, 3931-3946 (2005).
25. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
26. C. K. Stover, G. P. Bansal, S. Langerman, M. S. Hanson, Protective immunity elicited by rBCG vaccines. *Dev Biol Stand* **82**, 163-170 (1994).
27. R. E. Bosserman *et al.*, WhiB6 regulation of ESX-1 gene expression is controlled by a negative feedback loop in *Mycobacterium marinum*. *Proc Natl Acad Sci U S A* 10.1073/pnas.1710167114 (2017).