# Supplementary File 1. Detailed Proteomics Methods

### Linear ion trap tandem mass spectrometry

The proteolytic digests were further analyzed in duplicate per biological replicate as described previously [3], with the exception of the reverse phase gradient portion of the 2D capillary HPLC separation, which was as follows: 5% B for 13 min, 5-20% B in 1 min, 20% B for 6 min, 20-50% B in 45 min, 50-80% B in 1 min, 80% B for 9 min, 80-5% B in 5 min and 5% B for 10 min. Solvent A was 99.5% water and 0.5% acetic acid, and solvent B was 99.5% acetonitrile and 0.5% acetic acid (v/v). The mass spectrometry data acquisition parameters were as described previously [3] without change, and utilized a Thermo LTQ linear ion trap interfaced with a Michrom Magic 2002 HPLC configured in-house for 2D capillary HPLC operation [38-40, S9, S10]. Under these data-dependent scanning conditions 1 full scan ($MS^1$) and 10 product ion scans ($MS^2$) were acquired approximately every 3 sec, over a range of 400-2000 *m/z* units in $MS^1$ mode. The mass range in $MS^2$ is set by the instrument in real time based on the precursor ion *m/z* value. The absolute scanning speed of the LTQ used in this manner is $16.6 \times 10^3$ u/sec in centroid mode. Briefly, for each of the five pre-fractions, a complete 2D capillary HPLC analysis (38, 39) consisted of a seven part step gradient (38-40, S9, S10) from the cation exchange portion of the biphasic column, followed by the reverse phase elution described above, yielding a total of 35 separate HPLC runs per technical replicate at 60 min effective acquisition time per run, or 70 runs per biological replicate, yielding a total data acquisition time of ~280 hours for all four biological replicates, generating about 23 Gigabytes of raw data.

### SEQUEST and DTASelect

SEQUEST database searching [S3] and DTASelect [S8] filtering were performed as described [3, 4, S10]. The fasta database included normal and reversed *M. flagellatus* inferred proteins (Genbank CP000284) [11] and a forward and reversed human subset of the NCBI non-redundant database, depleted of all virus sequences [3, 4] but containing most common laboratory background proteins. The concatenated 22 Mb database comprised a total of 39,703 protein sequences, 2,759 of which belonged to *M. flagellatus*. The reversed sequences were also included in the database for purposes of assessing qualitative false discovery rates. The DTASelect Version 1.9 filtering criteria were comprised of fully tryptic peptides and ΔCn/Xcorr values for different peptide charge states of 0.08/1.9 for +1, 0.08/2.2 for + 2, and 0.08/3.3 for +3. The complete Sequest search parameters can be found in either the Sequest.params file or the filter.txt files provided (http://depts.washington.edu/mhlab/m_flagellatus/). All redundant spectra detected for each sequence were retained (t = 0 in DTASelect). Two peptides unique to a particular ORF were required for positive identification, based on a comparison of observed FDRs using different settings for this parameter (see Table below). Unique means that the peptide sequence could only be found in one open reading frame entry in the concatenated database. *M. flagellatus* contains a large duplicated segment of sequence predicted to encode 143 proteins (Mfla_820 - Mfla_962 and Mfla_964 - Mfla_1106). Since the sequences of these predicted proteins are identical, no distinction could be made between the duplicate proteins, and thus they are reported together in the data tables (Supplementary File 2). The DTASelect 1.9 filter data that support the protein identifications given in the summary tables are posted on an archival website (*http://depts.washington.edu/mhlab/*). These files contain the peptide Sequest scores, search parameters, individual peptide sequences, protein coverage by detected peptides and other peptide level information [S8].

**Estimation of the qualitative random false discovery rate (FDR)**

Random false positive identifications at the protein level were assessed from the reversed protein sequences in the concatenated database. Peptide search results that passed the criteria described above and matched an entry in the decoy database were counted as false positives. The protein level qualitative FDR [S2, S5] was thus estimated to be in the range of 1.0 to 2.5% across the four biological replicates, when the number of unique peptides required for identification was two, see the table below. The estimates were calculated based on the ratio of high scoring matches to the decoy database to the total matches for both decoy (reverse) and forward *M. flagellatus* sequences. There are obvious problems in a broad genome wide study if the stringency is increased to a level that drops the protein level FDR below 0.5%. This causes an increase in false negative risk beyond acceptable bounds, potentially distorting the very patterns in the negative results that we find so intriguing, not to mention distorting the quantitation. Based on the results given in the table below, the Sequest parameters given in the paper yield a reasonable balance between false positive and false negative risk, with a requirement of n = 2 unique peptides for a protein identification.

**Table. Protein level FDR for n=1, 2, 3 and 4 non-redundant (unique) hit data for the four biological replicates**

| p | Methylmine I | FDR (%) | Methylmine II | FDR (%) | Methanol I | FDR (%) | Methanol II | FDR (%) |
|---|---|---|---|---|---|---|---|---|
| 1 | 67/1687 | 4.0 | 39/1482 | 2.6 | 79/1565 | 5.0 | 105/1707 | 6.2 |
| 2 | 16/1514 | 1.1 | 13/1362 | 1.0 | 36/1420 | 2.5 | 26/1520 | 1.7 |
| 3 | 2/1278 | 0.2 | 0/1161 | 0 | 5/1203 | 0.4 | 2/1307 | 0.2 |
| 4 | 0/1072 | 0 | 0/981 | 0 | 0/1028 | 0 | 0/1125 | 0 |

**Proteomics experimental design, data normalization and significance testing**

The overall experimental design involved two complete biological replicates as described above for each of the two nutrient conditions: methanol_1 and methanol_2 and

methylamine_1 and methylamine_2. Each methanol replicate was compared to each methylamine replicate, yielding four sets of abundance ratios. The mean of the four abundance ratios is reported for each ORF in the data tables. Quantitative protein abundance ratios were calculated using label-free approaches as is in our previous work [3, 4, S10]. This design functions for non-label quantitation in a manner analogous to stable isotope "flip" replicates for metabolic labelling proteomic studies or dye swap replicates in a transcription microarray analysis. Each biological replicate consisted of the mean value of normalized total counts and total intensity calculated for each protein observed in two technical replicate analyses of the same prep as noted above. For the intensity measurements the $MS^1$ intensities for all precursor ions subsequently confirmed by CID and uniquely mapped to a single translated protein computationally were summed [3, 40, S9, S10]. Global normalization of the data was based on the spectral counts or summed intensity observed for the most abundant biological replicate. After normalization the average summed intensity or average spectral count was calculated. Proteins detected in only one condition were noted solely as qualitatively detected. Additionally, multiple ratios are required for significance testing and so only proteins detected in three or four of the four measurements could be analyzed for significance. For spectral counting, the *G*-test was used with each ratio determination, as we have published previously [3, S10], followed by calculation of $G_{Total}$, as per the method described in Sokal and Rohlf [33]. For each value of $G_{Total}$, a *p*-value was calculated as in our previous work [3, S10]. The uncorrected *p*-value was used as an input into the R package QVALUE (*http://faculty.washington.edu/~jstorey/qvalue/*), yielding a measure of the quantitative FDR. For the summed intensity measurements, an unpaired two-sample t-test was employed [S6]. The *t*-test was used to generate a global *p*-value over all four possible comparisons. As with the *G*-test results, the *p*-values were input into the QVALUE R package using the default parameters. A significance level of $q = 0.01$ or

lower was chosen as a cut-off for both methods, based on the criterion of achieving a best balance between false positive and false negative errors [S9]. A second filter was used for significance based on the spectral counting *G*-test results. The *G*-test is insensitive to the direction of change between samples, e.g. increased or decreased protein abundance between methylamine and methanol growth conditions, which can cause problems under certain circumstances [40]. Only proteins achieving the significance level cut-off of $q = 0.01$ or lower that showed the same direction of change in all comparisons where considered significant for the *G*-test results. A *q*-value of zero in the tables means that the calculated value was less than $2.22 \times 10^{-16}$ for the work reported here. As a check on the reasonableness of the *q* cut-off, the lists of significantly changed proteins were plotted against regions of random error derived from the random scatter about zero expression change in the technical replicates [3, S9]. The scatter plots (data not shown) and correlation analyses ($R^2 = 0.999$ for biological replicates of the same condition, $R^2 = 0.993$ to 0.994 for different conditions) were performed using the standard, un-weighted linear regression routine in the R statistical programming environment on the log-transformed data (R Version 2.2.1, *http://www.r-project.org*/). All comparisons of biological replicates were treated identically in terms of the regression analyses. The consensus abundance change assignments in Supplementary File 2 were made according to the following rules. If both the spectral counting and summed intensity abundance ratios deviated in the same direction from zero by more than 0.10 $\log_2$ units, and at least one of these assignments were statistically significant, then the consensus abundance change was coded either red or green for up or down, respectively. All other cases in which protein was detected qualitatively were coded as yellow for no abundance change.

**References**

*Note that references labelled by numbers only are listed in the main text and references labelled by 'S' are listed in this file.*


S1. **Benjamini, Y., and Y. J. Hochberg**. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Royal Statist. Soc. Ser. B. 57:289-300.

S2. **Elias, J. E., F. D. Gibbons, O. D. King, F. P. Roth, and S. P. Gygi**. 2004. Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nat. Biotechnol. 22:214-219.

S3. **Eng, J. K., A. L. McCormack, and J. R. Yates JR 3rd**. 1994. An approach to correlate tandem mass spectral data of peptides with amino-acid-sequences in a protein database. J. Am. Soc. Mass Spectrom. 5:976-989.

S4. **Hendrickson, E. L., Q. Xia, T. Wang, R. J. Lamont, and M. Hackett**. 2009. Pathway analysis for intracellular *Porphyromonas gingivalis* using a strain ATCC 33277 specific database. BMC Microbiol. 9:185.

S5. **Peng, J., J. E. Elias, C. C. Thoreen, L. J. Licklider, and S. P. Gygi**. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2:43-50.

S6. **Sokal, R. R., and F. J. Rohlf**. 1995. Biometry: the principles and practice of statistics in biological research, 3rd ed. Freeman W. H. and Co., NY.

S7. **Storey, J. D., and R. Tibshirani**. 2003. Statistical significance for genome wide studies. Proc. Natl. Acad. Sci. USA 100:9440-9445.

S8. **Tabb, D. L., W. H. McDonald, and J. R. Yates 3rd**. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. 1:21-26.

S9. **Xia, Q., T. Wang, Y. Park, R. J. Lamont, and M. Hackett**. 2007. Differential quantitative proteomics of *Porphyromonas gingivalis* by linear ion trap mass spectrometry: Non-label methods comparison, *q*-values and LOWESS curve fitting. Intl. J. Mass Spectrom. 259:105-116.

S10. **Xia, Q., T. Wang, F. Taub, Y. Park, C. A. Capestany, R. J. Lamont, and M. Hackett**. 2007. Quantitative proteomics of intracellular *Porphyromonas gingivalis.* Proteomics 7:4323-4337.