

DEVELOPMENT OF CELL METABOLOMIC PLATFORMS AND APPLICATIONS

THE DEVELOPMENT OF CELLULAR METABOLOMIC
PLATFORMS AND THEIR APPLICATION

By FAN FEI, B.Sc.

A thesis submitted to the School of Graduate Studies in Partial
Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University © Copyright by Fan Fei, December 2015

Doctor of Philosophy (2015)

McMaster University

Chemistry

Hamilton, Ontario

TITLE: The development of cellular metabolomic platforms
and their application

AUTHOR: Fan Fei, B.Sc. (McMaster University)

SUPERVISORS: Professors Brian E. McCarry,
Dawn M. E. Bowdish

NUMBER OF PAGES: xxx, 209

Abstract

In this thesis, an analytical platform was designed and applied to various *in vitro* bacterial and eukaryotic cell cultures. An extraction and an analytical protocol were developed for comprehensive and simultaneous analysis of both lipid and polar metabolites for intra- and extracellular metabolomics using HILIC-LC-TOF-MS. This analytical platform was applied to four diverse research questions such as the effect of oxygen environment on growth, the interplay between gene expression and metabolism, metabolic changes that occur with age, and PAH toxicity. Specifically: (i) the effect of oxygen on the growth, physiology and metabolism of the Gram positive *Streptococcus intermedius* were investigated by comprehensive intra- and extracellular metabolomes and transcriptome. (ii) Metabolic insights into the role of the multipartite genome of the Gram negative bacteria *Sinorhizobium meliloti* and its metabolic preferences in a nutritionally complex environment. (iii) Age-associated metabolic dysregulation in murine bone marrow-derived macrophages during bacterial lipopolysaccharide-induced inflammation. (iv) Comprehensive intracellular metabolomic profiles of *Sinorhizobium meliloti* to sub-lethal exposure of individual or mixtures of polycyclic aromatic hydrocarbon revealed additive and dose-dependent effects. This thesis has demonstrated the versatility of the designed analytical platform and its use for diverse research in cell biology.

Acknowledgements

This thesis could not have been completed without the technical and emotional support from my supervisors, committee members, lab mates, friends and family.

Foremost, I would like to thank Dr. Richard Lee, who was the first person to bring the term “bioanalytical” to my attention. I would also like to thank Dr. Philip Britz-McKibbin who provided me with a summer research position in his lab, and was where I first felt a passion towards bioanalytical research. Most importantly, I would like to thank my supervisors Drs. Brian E McCarry and Dawn M. E. Bowdish. Brian took me on as a graduate student, taught me alumina and anion exchange chromatography hand by hand, and led me into the field of bioanalytical chemistry and metabolomics. I would like to thank Dawn who was my long-term collaborator and adopted me to her lab since 2013. Since then, she welcomed me into her lab, not only financially supported my research and study, and also probably let me become one of the “best biologists” among chemists. Dawn’s optimism is contagious. Additionally, I would also like to thank my other collaborators: Michelle Mendonca, George diCenzo, Drs. Michael Surette and Turlough Finan. I would also like to thank Centre for Microbial Chemical Biology (CMCB), Biointerface (BI) Institute, and McMaster Regional Centre of Mass Spectrometry (MRCMS) for their support and instrumentation.

The insightful discussions among my committee members, Drs. Turlough Finan, Philip Britz-McKibbin, M. Kirk Green, became extremely valuable as my thesis progressed. I would like to express my gratitude to Kirk especially for his knowledge on GC, LC, MS, and constant reassurance during the confusing times. I would like to Dr. Mark McDermott for going through some of my manuscripts with me.

I would like to thank my lab mates from both McCarry and Bowdish labs. I would like to thank: Dessi and Avee for showing me how to isolate bone marrows and culture macrophages; Chris for showing me bits and pieces of everything; Kyle for the beautiful microscopy images; Mike and other friendly faces in Dawn’s lab. Most importantly, I would like to thank my lab mates from McCarry lab. Nothing was possible without them. In the good and bad times, they were always there to cheer and support me. Thank you to Uwayemi Sofowote for giving me your desk when I first joined the lab. David Bowman, Ashraf Ibrahim, Sujana Fernando, Kenneth Chalcraft, Vi Dang, Jonathan Bloomfield, Paul Hodgson, Elna and Roger Luckham, YOU ARE THE BEST!

Lastly, I would like to thank my long-time friends. I would also like to express my gratitude to my family.

Preface

This thesis contains results and publications from five years of research at McMaster University by the author. The focus of this thesis is to design a comprehensive cellular metabolomic platform and applied it to four different research applications in cell biology. Chapter three was in collaboration with Dr. Michael Surette and Michelle Mendonca from Department of Biochemistry and Biomedical Sciences at McMaster University. Chapter four was in collaboration with Dr. Turlough Finan and George diCenzo from Department of Biology at McMaster University. Chapter five was in collaboration with Keith M. Lee from Dr. Dawn M. E. Bowdish lab. Lastly, Chapter six was in collaboration with Elna D. Luckham from Dr. Brian E. McCarry lab.

Table of Contents

Abstract	iii
Acknowledgements	iv
Preface	v
Table of Contents	vi
List of Figures	ix
List of Tables	xxi
List of Abbreviations and Symbols	xxiii

Chapter 1: Introduction to metabolomics, and its application on cell biology	1
1.1 Metabolomics	1
1.1.1 Introduction to metabolomics	1
1.1.2 Comprehensive and targeted metabolomics.....	2
1.1.3 Implication of metabolomics to cell research.....	3
1.1.4 Metabolomic workflow	4
1.2 Sample preparation and extraction	4
1.2.2 Metabolism quenching	4
1.2.3 Metabolite extraction	5
1.3 Analytical instrumentation	7
1.3.2 GC-MS	7
1.3.3 LC-MS	8
1.3.4 CE-MS	10
1.3.5 Mass analyzers for mass spectrometry.....	10
1.4 Data processing, statistical and chemometric analyses	11
1.4.2 Software for metabolomic analysis	11
1.4.3 Quality Assurance	12
1.4.4 Univariate and multivariate analyses.....	12
1.5 Challenges in cell metabolomics.....	15
1.6 Thesis objectives	16
1.7 References	22
Chapter 2: Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS.....	34
Supplementary Information.....	57
Chapter 3: Metabolic and Transcriptomic Profiling of <i>Streptococcus intermedius</i> during Aerobic and Anaerobic Growth	72
Supplementary Information.....	99
Chapter 4: Metabolic insights into the role of the multipartite genome of <i>Sinorhizobium meliloti</i> and its metabolic preferences in a nutritionally complex environment	112
Supplementary information.....	136

Chapter 5: Age-associated metabolic dysregulation in macrophages during inflammation	147
Supplementary information.....	174
Chapter 6: Comprehensive Metabolomic Analysis of <i>Sinorhizobium meliloti</i> responses to sub-lethal PAH exposures	183
Chapter 7: General discussion, future directions and conclusion	206
7.1 General discussion	206
7.2 Implications of metabolomics.....	206
7.3 Future directions of research.....	208
7.4 General conclusion.....	209
7.5 References	209

List of Figures

Chapter 1

Figure 1 The central dogma of biology and the “omic cascade”. Metabolites are downstream biochemical end products of epigenetic process and post-translational modifications of genes and proteins. There is a stronger correlation between metabolic profiles and phenotypes compared to genomics, transcriptomics and proteomics.18

Figure 2 The growth of metabolomics since 1998. a) The number of publications containing “metabolo*” in 1998-2014 as per Web of Science. b) The number of metabolomic publications in 1998-2014 was classified according to the analytical platform used. “Metabolo* LC”, “metabolo* GC”, “metabolo* CE” and “metabolo* NMR” were used as key words in Web of Science. “Metabolo*” was used as the key words to include terms such as metabolomics, metabolomic, and metabolome.....19

Figure 3 Metabolomic workflow. Cell metabolomics consists of five sequential steps: (i) cell culture growth or stimulation, (ii) quenching metabolic activity and metabolite extraction, (iii) data acquisition using MS-based or nuclear magnetic resonance (NMR) spectroscopy techniques, (iv) statistical and chemometric analysis, (v) data interpretation linking metabolomics to biological process or identify biomarkers20

Figure 4 Conceptual coverage of metabolites using chromatography-MS coupled techniques21

Chapter 2

Figure 1 The extraction efficiencies of endogenous metabolites in 2×10^9 *S. meliloti* cells in the first extraction (a), and 2nd to the 7th extraction (b) (only 2nd, 3rd, and 7th extractions are shown). The extraction efficiencies at each extraction step48

Figure 2 (a) Base peak chromatogram of the HILIC separation on ten lipid classes including PGs, lyso-PG, phosphatidic acids (PAs), PSs, triacylglycerols (TAG), PEs, ceramides (CEs), PCs, sphingomyelins (SMs) and lyso-PCs. (b) Base peak chromatogram of $[M+H]^+$ ions of PEs.....49

Figure 3 The 2125 endogenous metabolite features detected in 3×10^5 murine macrophage extracts after data reduction. The radius of the data markers (filled circle) reflected the relative abundances of the metabolite features over a dynamic range of 5×10^4 . Small metabolites (90-400 m/z) in yellow, PGs in green, PEs in red, PCs in purple and lyso-PCs in blue were highlighted. The HILIC gradient was labelled with a black line with reference to percentage acetonitrile50

Figure 4 The metabolome coverage of *S. meliloti* of 1059 endogenous metabolite features found in at least one set of BD polar, BD nonpolar or MeOH/EtOH/H₂O extracted samples. Extracts were performed in sextuplicate and analyzed independently by HILIC-TOF-MS in both ESI+ and ESI- modes. The ionization responses of the metabolite features were normalized using internal standards. (a) OPLS-DA score plot comparing the endogenous metabolome coverage of *S. meliloti* extracts attained from three different extraction methods with $R^2X(\text{cum})=0.926$, $R^2Y(\text{cum})=0.986$, and $Q^2(\text{cum})=0.936$. (b) The quantity of metabolite features that were uniquely identified and shared between BD polar, BD nonpolar and MeOH/EtOH/H₂O extracted samples were listed in the Venn diagram with their estimated percentage share of the total detectable metabolome. (c) The normalized ionization responses of identified metabolites with varying polarity. The ionization responses from MeOH/EtOH/H₂O extracts was set to one as references. Polar metabolites were mostly extracted in BD polar, and lipids were seen exclusively in BD nonpolar; however, all metabolites were detected in MeOH/EtOH/H₂O. Error bars corresponded to two standard deviations. UDP, uridine diphosphate; GlcNac, N-acetylglucosamine51

Figure 5 Among 228 shared metabolite features in BD polar, BD nonpolar and MeOH/EtOH/H₂O extracts, (a) 14.6% features had lower extraction efficiencies in MeOH/EtOH/H₂O extracts than at least one of the BD extracts, and (b) 15.8% features had increased extraction efficiencies in MeOH/EtOH/H₂O extracts than all BD extracts. Each error bar corresponds to two standard deviations calculated from six independent extractions. The ionization responses were normalized to the IS and also the total summed ionization response for both BD polar and BD nonpolar. Each metabolite feature is represented by the ESI mode used for its detection (in brackets), its m/z value and retention time52

Figure S1 The schematic for normalizing the ionization efficiency of endogenous metabolite features using L-lysine-¹³C₆-¹⁵N₂, phe-phe, L-phenylalanine-d₈ and gly-phe as internal standards. The metabolite features were normalized with internal standard that eluted closest to their retention time59

Figure S2 The total ion chromatograms of endogenous (black) and exogenous (grey) metabolomic profiles from murine macrophages cultured in RPMI-1460. MeOH/EtOH/H₂O (2:2:1) was used to extract the endogenous metabolites. For exogenous metabolome, 20 μ L of cell supernatant were extracted with 80 μ L of MeOH/EtOH once and then diluted 2 fold for LC analysis. (A) The murine macrophage cells were not washed with phosphate buffered saline (PBS) prior to extraction. (B) The murine macrophage was washed once with PBS after harvest and then extracted61

Figure S3 2^3 full factorial design of experiment for HILIC chromatography optimization. A HILIC gradient with acetonitrile (solvent A) and 10 mM ammonium acetate in water adjusted to pH3 with formic acid (solvent B) was used. *S. meliloti* extracts (2×10^9 cells) were separated on 50 mm $\times$ 2.1 mm Kinetex 2.6 μ m HILIC column and detected using a Bruker MicrOTOF II. (A) The HILIC gradient was optimized with initial hold time (X1), starting acetonitrile percentage (X2), and the rate of change of acetonitrile percentage (X3) in order to maximize the number of metabolite features with $k' < 0.7$ (Y1) and minimize the time required for all features to elute during the gradient (Y2). (B) Two 2^3 factorial design was used. The axial points were repeated in triplicates, and the center point was repeated in sextuplicate. Mathematical modelled regressions based on the 2^3 full factorial designs are shown in (C) and (D) as 3D surface plots for Y1 and Y2 respectively64

Figure S4 The normalized ionization responses of *S. meliloti* endogenous metabolites found in BD polar, MeOH/EtOH and MeOH/EtOH/H₂O extracts. Error bars corresponded to two standard deviations based on a triplicate independent extractions normalized to the IS65

Figure S5 The extracted ion chromatogram (EIC) of (a) [M+H]⁺ or (b) [M-H]⁻ ion of gly-phe added post-column following HILIC separation of endogenous murine macrophage MeOH/EtOH/H₂O extract (grey) or a 60%v/v ACN blank (black). Ion suppressions (black bold lines) were found at 0.5-1.2 min (k_{app}' : -0.3-0.7) and 6.9-8.0 min (k_{app}' : 8.9-10.4) in ESI⁺ mode; 0-1.2 min (k_{app}' : -0.3-0.7), 2.9-3.5 min (k_{app}' : 3.2-4.0), and 6.9-8.0 min (k_{app}' : 8.9-10.4) in ESI⁻ mode. The gly-phe was prepared to examine whether the matrix is able to suppress ionization signals, not for quantification or quality control purposes.....66

Figure S6 The retention deviation of ca. 260,000 metabolite features detected in 137 *S. meliloti* MeOH/EtOH/H₂O extracts and pooled samples in ESI⁺ mode was between ± 6 seconds.67

Chapter 3

Figure 1 (a) The growth curves of *S. intermedius* cultured in either aerobic (in orange) or anaerobic (in blue) environment. The doubling time in aerobic condition was 54 min and the doubling time in anaerobic condition was 41 min. The optical density of the cells was measured every hour in triplicate. OPLS-DA score plots summarizing (b) 1885 metabolite features found in the intracellular extracts between aerobically and anaerobically cultured *S. intermedius*, (c) 3382 metabolite features found in extracellular medium of aerobically and anaerobically cultured *S. intermedius* and Todd Hewitt growth media, and (d) gene expression differences including 1815 transcripts obtained from RNA-seq of *S. intermedius* grown in either aerobic or anaerobic conditions. The samples belonging to the same treatment were highlighted with circles86

Figure 2 A heatmap of 28 pathways affected by aerobic or anaerobic growth conditions, based on identified metabolites in the intracellular and extracellular metabolomes of *S. intermedius*. The p value of each pathway was computed by pathway analysis using MetaboAnalyst 2.0 using Gram-positive *Staphylococcus aureus* as the model organism. Rows were metabolite pathways; columns were comparisons between treatments including aerobic and anaerobic growth conditions and the Todd Hewitt media blank. Comparisons were based on either intracellular *S. intermedius* cell extracts of 79 identified metabolites (exclude phospholipids) or extracellular supernatant of 92 identified metabolites. The color key indicates the $-\log_{10}$ of p values for pathway significance (refer to the color scale). The energy production pathways were calculated mostly based on C5, C6 monosaccharide and disaccharide abundances. A list of identified metabolites used for the analyses was included in the electronic supplementary material ESM 2. The citrate cycle (tricarboxylic acid (TCA) cycle) in *S. intermedius* is incomplete: however intermediates from the partial pathway are differentially produced.87

Figure 3 The effect of the presence of oxygen on gene expression of *S. intermedius*. Anaerobic gene expression was compared to aerobic and visualized with down-regulated, upregulated and unaffected genes coloured in orange, yellow and green respectively. Genes were classified into pathways and assembled into 4 groups based on response, namely, anaerobic response, mixed response, aerobic response and unaffected pathways. CRISPR: clustered regularly interspaced short palindromic repeats; SAM: S-adenosyl methionine; PPP: pentose phosphate pathway; CHP: conserved hypothetical protein; GlcNAc: N-acetylglucosamine; FMN: flavin mononucleotide; UDP: uridine diphosphate; NAD: nicotinamide adenine dinucleotide.....88

Figure 4 Overview of (a) glycolysis and mixed acid fermentation pathways and (b) NAD⁺/NADH cycling pathway of *S. intermedius*, which were affected by aerobic and anaerobic growth conditions. Genes that were up-regulated in anaerobic conditions were indicated in blue; genes that were down-regulated in anaerobic condition were indicated in red; genes which expression was not statistically significant were indicated in black (Student's t test, $p > 0.05$). PTS, phosphotransferase system; *manL*, putative phosphotransferase system, mannose-specific EIIAB; *manM*, PTS system, mannose-specific IIC component; *manN*, PTS system, mannose-specific IID component; *glcK*, glucokinase putative; *fba*, fructose biphosphatealdolase; *pyK*, pyruvate kinase; *pgm*, putative phosphoglucomutase/phosphomannomutase; *glgA*, glycogen synthase, ADP-glucose type; *glgB*, glycogen branching enzyme; *glgC*, glucose-1-phosphate adenylyltransferase; *glgD*, glucose-1-phosphate adenylyltransferase, GlgD subunit; *ldh*, L-lactate dehydrogenase; *pfl*, formate C-acetyltransferase; *pta*, phosphate acetyl/butaryltransferase; *ackA*, acetate kinase; *adhE*, bifunctional acetaldehyde-CoA/alcohol dehydrogenase; *adhP*, alcohol dehydrogenase; *aldB*, alpha-acetolactate

decarboxylase; *ppc*, phosphoenolpyruvate carboxylase; *mleS*, malate dehydrogenase; *nox*, NADH oxidase; *ahpC*, alkyl hydroperoxide reductase subunit C; *ahpF*, alkyl hydroperoxide reductase subunit F89

Figure 5 The arginine deiminase and pyrimidine de novo synthesis pathways. Pathway was constructed based on the BioCyc database for *S. intermedius* B196 and JTH08. The metabolite names were written in black and the gene names were written in green. The fold changes of metabolite expression were indicated in color scaled boxes for *S. intermedius* grown in aerobic (A) and anaerobic (AN) conditions and the Todd Hewitt media blank (TH). The endo-metabolome was colored in grey and the exo-metabolome was light green, fold changes in gene expressions were indicated by numerical values. The undetected metabolites were indicated with a black filled box. The fold changes in metabolite or gene levels were calculated respective to aerobic growth conditions for either endo- or exo-extracts, where an increase was shown in blue and a decrease was shown in red. n=7 except for intracellular cell extract in aerobic conditions and Todd Hewitt media. *p<0.05, **p<0.005, ***p<0.000191

Figure 6 The heatmap of 15 statistically significantly affected phospholipids (p<0.05) found in the intracellular metabolome of *S. intermedius* cultured in either aerobic or anaerobic conditions. The phospholipids were listed according to their alkyl chain length and saturation. The relative abundances of phospholipids were illustrated using a color scale, with blue indicating high abundances and red indicating low abundances. Note the chain length and saturation reported is cumulative for both fatty acid side chains.92

Figure S1 Global metabolomic differences between aerobically and anaerobically cultured *S. intermedius* shown by volcano plots, comparing (a) intracellular metabolome profiles of aerobic and anaerobic conditions, (b) extracellular metabolome profiles of aerobic and anaerobic conditions, (c) extracellular metabolome profiles of aerobic condition and Todd Hewitt media blank, and (d) extracellular metabolome profiles of anaerobic condition and Todd Hewitt media blank. Significant metabolite features with p<0.05, and greater than 1.5 fold changes were indicated with solid circles, others were labelled with open circles. Identified metabolite features were colored in red, putative lipids were indicated in blue, and the unknowns were colored in grey. Some known significant metabolite features were labelled with their chemical names. Some metabolites appear multiple times in one volcano plot due to detection in both ESI- and ESI+ modes or as adduct ions102

Figure S2 Overview of transcriptomic results for *S. intermedius* growth under aerobic and anaerobic conditions using CummeRbund. The scatter plot in (a) depicts the expression of all genes (1815 genes) under aerobic and anaerobic conditions. The theoretical correlation for equivalent expression under the two conditions (Blue) overlaps with the actual

correlation. Scatter plot analysis of significant genes (625 genes) with p-value of 10 ^{-4.5} is shown in (b). The lines correlating to 2 fold and 4 fold changes in gene expression are shown. There are few genes upregulated above 4 fold, with the majority of genes being upregulated under aerobic conditions. The volcano plot in (c) highlights statistically significant genes in red. Figures were generated using CummeRbund	103
Figure S3 The growth curves (CFU/mL over time) of <i>S. intermedius</i> cultured in either aerobic (black) or anaerobic (blue) environment.....	104
Figure S4 Variation in expression of genes involved in carbohydrate metabolism in the presence/absence of oxygen. Genes with statistically significant fold changes above 2 were classified either as an “Anaerobic response” or an “Aerobic response”, based on the condition wherein they are upregulated. Data was FPKM values generated from Cufflinks	105
Figure S5 Amino acid metabolism of strain <i>S. intermedius</i> B196, constructed based on the BioCyc database for <i>S. intermedius</i> B196 and JTH08. The metabolite names were written in black and the gene names were written in green. The fold changes in metabolite expression were indicated in color scaled boxes for <i>S. intermedius</i> grown in aerobic (A) and anaerobic (AN) conditions and the Todd Hewitt medium control (THY). The intracellular metabolome was colored in grey and the extracellular metabolome was in light green; fold changes in gene expressions were indicated by numerical values. The undetected metabolites were indicated with a black filled box. The fold changes in metabolite or gene levels were calculated respective to aerobic growth conditions for either intra- or extra-cellular extracts, an increase was shown in blue and a decrease was shown in red. n=7 except for intracellular cell extract in aerobic conditions and Todd Hewitt media. *p<0.05, **p<0.005, ***p<0.0001	106
Figure S6 The purine metabolism pathway, constructed based on the BioCyc database for <i>S. intermedius</i> B196 and JTH08, as discussed in Fig. S5	107
Figure S7 The pyrimidine salvage pathway, constructed based on the BioCyc database for <i>S. intermedius</i> B196 and JTH08, as discussed in Fig. S5	108
Figure S8 Genes involved in pyrimidine metabolism were affected by the presence/absence of oxygen. The response can be divided into “Anaerobic” or “Aerobic” based on the condition wherein they are up-regulated. Data was obtained using Cufflinks for RNAseq analysis	109
Figure S9 Induction of genes involved in oxidative stress under aerobic conditions. The heatmap includes genes that were induced aerobically above 2 fold. The up-regulated genes are involved in pathways including oxidative stress, iron metabolism and iron sulfur cluster assembly. The data is based on FPKM values generated from Cufflinks .	110

Chapter 4

Figure 1 The growth curves of wild-type, Δ pSymA, Δ pSymB and Δ pSymAB *S. meliloti* strains. a, the growth of *S. meliloti* and its mutants were examined in M9 minimal medium and b, LBmc medium. The small solid circles represent OD₆₀₀ readings from a fixed sample of each strain throughout the entire growth curve. Samples were removed from culture tubes throughout growth and prepared for metabolomic analysis. The large solid circles with a black border indicate the average OD₆₀₀ readings of the quintuplicate or sextuplicate samples collected for intracellular and extracellular metabolomic analyses. Black—wild type; blue— Δ pSymA; green— Δ pSymB; orange— Δ pSymAB.126

Figure 2 Statistical analyses of the intracellular metabolome of wild-type, Δ pSymA, Δ pSymB, and Δ pSymAB *S. meliloti* at various growth phases including exponential phase (M1-4) and stationary phase (M5). a, the OPLS-DA score plot showing the global relationship between the four *S. meliloti* strains at various growth phases based on 1474 detected metabolite features. The progression of growth is indicated by the arrow direction. The predictive statistic for the model fell below 0.4 ($Q^2=0.256$) due to similar metabolic profiles between strains and growth phases. b, heatmap of 1237 significant metabolite features (between subject two-way ANOVA with Bonferroni correction, $p<0.01$) plotted based on log₂(fold change) with respect to the average metabolite levels in M1WT. c, heatmap of 41 significant phospholipids (between subject two-way ANOVA with Bonferroni correction, $p<0.01$) based on log₂(fold change) with respect to the average metabolite levels in M1WT. M1A sample c (M1A-c) was removed as an outlier. d, hierarchical clustering analysis of the total detectable features (1474 features) for all four strains at each growth stage. Δ pSymAB. Black—wildtype; blue— Δ pSymA; green— Δ pSymB; orange— Δ pSymAB. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— Δ pSymA; B— Δ pSymB; AB— Δ pSymAB127

Figure 3 Intracellular levels of selected metabolites (a-t). Intracellular data were obtained from wild-type, Δ pSymA, Δ pSymB, and Δ pSymAB *S. meliloti* strains cultured in minimal M9 medium at 5 time points. n=5-6. Error bar represents one standard deviation from the average. black—wild type; blue— Δ pSymA; green— Δ pSymB; orange— Δ pSymAB. M1-4—exponential phase; M5—stationary phase129

Figure 4 Statistical analyses of extracellular metabolome of wild-type, Δ pSymA, Δ pSymB, and Δ pSymAB *S. meliloti* at various growth phases including mid-exponential phase (L1), early stationary phase (L2) and late stationary phase (L3) cultured in LBmc. a, the OPLS-DA score plot showing the global relationship between the four strains of *S. meliloti* at various growth phases based on 2008 detected metabolite features. The progression of growth is indicated by the arrow direction. b, heatmap of the 820

significant metabolite features (between subject two-way ANOVA with Bonferroni correction, $p < 0.01$) plotted based on $\log_2(\text{fold change})$ with respect to the average metabolite levels in LBmc. c, hierarchical clustering analysis of the total detectable features (2008 features) for all four strains at each growth stage. LBmc—red; black—wild type; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. L1—circle; L2—diamond; L3—square. WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$; LBmc—LBmc blank following 0 hr and 45 hr incubation at 30°C 130

Figure 5 Extracellular data were obtained from various growth phases of wild-type, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, and $\Delta p\text{SymAB}$ *S. meliloti* strains cultured in LBmc medium (a-z). Error bar represents one standard deviation from the average. black—wildtype; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. L1—mid-exponential phase; L2—early stationary phase; L3—late stationary phase, LBmc—average of the readouts of 0hr and 45 hr LBmc blank at 30°C..... 131

Figure S1 The OPLS-DA plot of the intracellular metabolome of wild-type, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, and $\Delta p\text{SymAB}$ *S. meliloti* strains cultured in M9-sucrose medium at multiple times points on the growth curve, as well as pooled samples. Metabolome variance between strains and during growth was summarized by x- and y-axes. Black—wild type; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$, pooled samples—purple. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$ 139

Figure S2 The OPLS-DA plot of the extracellular metabolome of wild-type, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, and $\Delta p\text{SymAB}$ *S. meliloti* strains cultured in LBmc at multiple times points on the growth curve, as well as pooled samples. Metabolome variance between strains and during growth was summarized by x- and y-axes. Black—wild type; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. L1—circle; L2—diamond; L3—square. L1LBmc—fresh LBmc medium; L3LBmc—LBmc medium incubated at 30°C for 45hr; WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$ 140

Figure S3 The OPLS-DA score plots of the intracellular metabolome of **a** wild-type, **b** $\Delta p\text{SymA}$, **c** $\Delta p\text{SymB}$, and **d** $\Delta p\text{SymAB}$ *S. meliloti* at various growth phases including exponential phase (M1-4) and stationary phase (M5). The progression of growth is indicated by the arrow direction. The metabolome variance for each strain during growth was summarized by x- and y-axes. $\Delta p\text{SymAB}$. Black—wild type; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$ 141

Figure S4 The intracellular metabolic phenotypes of $\Delta p\text{SymA}$, wild type, $\Delta p\text{SymB}$ and $\Delta p\text{SymAB}$ strains at exponential phase (**a** M1, **b** M2, **c** M3, **d** M4), and **e** stationary phase

(M5) were summarized using OPLS-DA plots. Metabolome variance for each strain during growth was summarized by x- and y-axes. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$, pooled samples—purple. M1—circle; M2—diamond; M3—triangle; M4—square; M5—star. WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$142

Figure S5 The top 100 intracellular metabolites with the most significant $p_{overall}$ computed using between subjects two-way ANOVA ($p_{overall} = \min(p_{time}, p_{strain}, p_{interaction})$). The metabolites were labelled in the order of m/z, retention time (min), and electrospray ionization mode. M1-4—exponential phase; M5—stationary phase.....143

Figure S6 The OPLS-DA plots of the extracellular metabolome of wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* cultured in LBmc at various growth phases. The extracellular metabolic phenotypes of **a** wild type, **b** $\Delta pSymA$, **c** $\Delta pSymB$ and **d** $\Delta pSymAB$ at mid-exponential phase (L1), early stationary phase (L2) and late stationary phase (L3) were summarized using OPLS-DA plots. The OPLS-DA score plots of the extracellular metabolome of wildtype, $\Delta pSymA$, $\Delta pSymB$, $\Delta pSymAB$ and LBmc medium were compared at **e** L1, **f** L2, and **g** L3. For a-d, metabolome variance for each strain during growth was summarized by x- and y-axes. For e-g, metabolome variance between strains at each growth phase was summarized by x- and y-axes. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$. L1—circle; L2—diamond; L3—square. WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$; LBmc—LBmc blank medium incubated for 0 or 45 hrs at 30°C144

Figure S7 The top 100 extracellular metabolites with the most significant $p_{overall}$ computed using between subjects two-way ANOVA ($p_{overall} = \min(p_{time}, p_{strain}, p_{interaction})$). The metabolites were labelled in the order of m/z, retention time (min), and electrospray ionization mode. L1—mid exponential phase; L2—early stationary phase; L3—late stationary phase.....145

Chapter 5

Figure 1 A) The experimental outline. B) The experimental workflow for analyzing macrophage extracts. From one macrophage culture, the sample extract was analyzed separately with HILIC-TOF-MS and GC-qMS with distinct sample preparation, data acquisition, data processing, data analysis, and quality assurance.162

Figure 2 The comprehensive analyses of bone marrow-derived macrophage extracts acquired using HILIC-TOF-MS. The ionization responses of 2125 intracellular metabolite features were normalized using IS. Extracts were performed in sextuplicate with three biological replicates and two culture replicates. (A) OPLS-DA score plot comparing the metabolic profiles of control (0 hr), 4 hr and 16 hr LPS stimulated macrophage extracts

from both young and old mice. (B) Heat map visualization of the intracellular metabolite changes of macrophages of young and old mice in response to LPS stimulation. The 920 significant metabolite features ($p < 0.05$ one-way ANOVA, fold change greater than 1.5 compared to control) are represented in rows, and the experimental conditions were listed in columns. The heat map is plotted based on $\log_2(\text{fold change})$ with respect to the average levels of each metabolite feature in the control of macrophages from young mice using Euclidean distance and complete-linkage clustering. OPLS-DA score plot comparing the metabolic profiles of control (0 hr), 4 hr, and 16 hr LPS stimulated macrophage extracts as well as “recovery” and “tolerance” treated macrophage extracts from (C) young or (D) old mice. Controls 1 and 2 obtained at the beginning ($t = 0$ hr) and the end of the experiments ($t = 22$ hrs) were considered as a single group. Samples belonging to the same treatment group were highlighted by open circles. 164

Figure 3 The relative intercellular levels and gene expression of selected metabolites and genes. The intracellular levels of adenine (A), adenosine (B), arginine (C), and ornithine (D) were acquired at control (0hr), 4 hr and 16 hr of LPS stimulation as well as at the “recovery” and “tolerance” conditions for macrophages of both young and old mice. The gene expressions of *Arg1* (E), *iNOS* (F), and *Slc7a2* (G) were acquired at control (0hr), 4 hr and 16 hr of LPS stimulation for macrophages of young and old mice. The urea pathway was illustrated in (H). Metabolomics and gene expression data of macrophages from young mice were labelled in shades of blue; those from old mice were labelled in shades of red. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ 165

Figure 4 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates in bone marrow-derived macrophages from young mice in response to (A) 16 hr LPS stimulation, (B) “recovery”, and (C) “tolerance” as compared to unstimulated macrophages. The network of metabolite interactions was built based on the BioCyc database and pathway published from Jha *et al* (1). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2(\text{fold changes})$ of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple. 167

Figure 5 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates in bone marrow-derived macrophages from old mice in response to (A) 16 hr LPS stimulation, (B) “recovery”, and (C) “tolerance” as compared to unstimulated macrophages. The network of metabolite interactions was built based on the

BioCyc database and pathway published from Jha *et al.* (1). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2(\text{fold changes})$ of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple. 169

Figure S1 The GC-EI-MS chromatogram of arginine standard..... 176

Figure S2 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates during 4 hrs of LPS stimulation in bone marrow-derived macrophages from (A) young mice or (B) old mice as compared to corresponding unstimulated macrophages. The network of metabolite interactions was built based on the BioCyc database and pathway published from Jha *et al.* (Jha *et al.* 2015). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2(\text{fold changes})$ of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple. 178

Figure S3 The relative abundances of the 25 targeted intracellular metabolites from macrophages of (A) young and (B) old mice from five different experimental conditions as acquired by GC-MS. The samples were performed in sextuplicate with three biological replicates and two culture replicates. The error bar represents one standard deviation. The relative abundances of some metabolites were amplified by 5, 10, 50, 100, or 1000 fold to scale to the figure. G1P, glucose-1-phosphate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; G6P, glucose-6-phosphate; GABA, γ -aminobutyric acid..... 180

Figure S4 The gene expression of tolerance and non-tolerance associated genes identified by Foster *et al.* (Foster *et al.* 2007). The expressions of these genes were acquired from unstimulated (0 hr), 4 hrs and 16 hrs of LPS stimulation, and LPS tolerant macrophages from either young or old mice..... 181

Chapter 6

Figure 1 Optimization for coal tar clean-up using alumina chromatography. The heat map illustrated the extraction efficiency of individual PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs in A2 sub-fractions, A2, A3, and A4 fractions of experiment (a) C1, (b) C2,

(c) C3 and (d) C4. The extraction efficiencies of PAHs and their derivatives in each fraction and sub-fraction were calculated by first normalizing the peak area of each metabolite with internal standards and then dividing the relative abundance of each individual PAHs at (sub-)fractions with its total recoverable sum in all A1-A4 fractions. Analytes in the PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs classes were ordered based on molecular weight.....196

Figure 2 Optimization for coal tar clean-up using alumina chromatography. The chromatographic distributions of PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs in A2 sub-fractions, A2, A3, and A4 fractions for experiment (a) C1, (b) C2, (c) C3 and (d) C4.....197

Figure 3 Elution profiles of PAH, Me-PAH, NPAH, Me-NPAH, PASH in A2 sub-fractions, A3 and A4 fractions in the optimized coal tar clean-up using alumina chromatography (experiment C5). (a) The heat map illustrate the elution pattern and extraction efficiencies of individual PAH or PAH derivatives in each fraction. Analytes in the PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs classes were ordered based on molecular weight. (b) The extractable concentration of each PAH groups at each fraction were illustrated. (c) The PAH concentration of A2-2 sub-fraction used for *S. meliloti* metabolomic study to investigate metabolic effects of sub-lethal PAH exposure.....198

Figure 4 PCA score plot of *S. meliloti* polar metabolome to illustrate the reproducibility of the experiment. The pooled samples, colored in purple, were clustered in the center of the score plot indicating minimal technical variance. The polar metabolome consisted of 2005 metabolic features after data reduction200

Figure 5 Comprehensive metabolomics of *S. meliloti* exposed to 0.14 mg/L and 1.4 mg/L fluorene or low molecular weight PAHs mixture. Exposure and metabolomic experiments were performed in sextuplicate. (a) Changes of metabolic profiles with regard to different PAH exposures were visualized in the OPLS-DA score plot ($R^2X=0.93$, $R^2Y=0.827$, $Q^2=0.409$). 2005 metabolic features were included in the analysis. No metabolic differences were observed between control samples and *S. meliloti* exposed to 0.14 mg/L fluorene. (b) Heat map illustrating individual metabolic feature changes with regards to the various PAH exposure. The 926 significantly expressed metabolic features (fold change >1.5 and $p<0.05$ with Student's t test between PAH exposed samples and control) were represented in rows, and various experimental conditions were listed in columns. The heat map was plotted based on the $\log_2$ (fold change) with respect to the average levels of each metabolic features in the control sample using Euclidean distance and complete-linkage clustering.201

List of Tables

Chapter 1

Table 1 Characterizations of common mass analyzers used for metabolomic studies11

Table 2 Parametric and non-parametric univariate statistical tests for datasets following normal distribution or far from normal-curve13

Chapter 2

Table 1 The shared 228 metabolite features in BD polar, BD nonpolar and MeOH/EtOH/H₂O extracts were compared between different extraction methods using unpaired heteroscedastic Student's t test with $p < 0.05$. Among the shared features, 15.8% showed greater extraction efficiency (↑), 14.6% showed lower extraction efficiency (↓) and 59.6% showed no difference (≈) between MeOH/EtOH/H₂O and BD fractions. “>”, greater extraction efficiency; “<”, less extraction efficiency; “=”, equal extraction efficiency.....53

Table S1. The mass accuracies (ppm) of selected metabolites in different biological matrix ranging from Gram-positive *Streptococcus intermedius*, Gram-negative *Sinorhizobium meliloti*, to murine macrophages before and after internal calibration using endogenous sodium formate. The mass accuracies after the internal calibration with sodium formate were in bold.68

Table S2. Metabolite features identified based on the accurate m/z values and retention time attained from authentic standards.69

Chapter 3

Table S1 Summary of metabolomic and transcriptomic data111

Chapter 4

Table S1 Intracellular and extracellular metabolome data summary.....137

Table S2 Biological variation of intracellular and extracellular metabolic profiles of the wild type, ΔpSymA, ΔpSymB, ΔpSymAB at various growth phases137

Table S3 Summary table of validation parameters (R^2X , R^2Y , Q^2) of the OPLS-DA model for the intracellular metabolome in Figures S3, S4. Model 1-4 corresponds to Figure S3 a-d, and model 5-9 corresponds to Figure S4 a-e138

Table S4 Summary table of validation parameters (R2X, R2Y, Q2) of the OPLS-DA model for the extracellular metabolome in Figure S6. Model 1-7 corresponds to Figure S6 a-g	138
---	-----

Chapter 5

Table S1 Target Metabolites for GC-qMS analysis.....	175
Table S2 Single-ion-monitoring (SIM) method for GC-qMS	176
Table S3 Coefficient of variance of macrophage metabolites from young or old mice. .	177
Table S4 Summary of total number of metabolite features extracted from LC-MS data after XCMS, CAMERA and data reduction	177

Chapter 6

Table 1 Alumina chromatographic optimization conditions for coal tar separation	186
Table 2 A2 toluene sub-fractions collected in optimized experiment C5.....	187
Table 3 Qualifying and quantifying ions for PAH, PASH and NPAH, and their methylated derivatives using GC-MS. The molecular ions were used to quantify their levels.	188
Table 4 The OD ₆₀₀ of control, 0.14 mg/L fluorene, 1.4 mg/L fluorene, 0.14 mg/L PAH mixture, and 1.4 mg/L PAH mixtures exposed <i>S. meliloti</i> culture at the time of the harvest	189
Table 5 The concentrations of low molecular weight PAH mixtures used for <i>S. meliloti</i> metabolomics study. The sample was collected and diluted from C5 A2-2 fraction	189

List of Abbreviations and Symbols

Chapter 1

AMDIS	automated mass spectrometry deconvolution and identification system
ANOVA	analysis of variance
APCI	atmospheric pressure chemical ionization
BSTFA	bis(trimethyl)silyl trifluoroacetamide
CE	capillary electrophoresis
CI	chemical ionization
CZE	capillary zone electrophoresis
DESI	desorption electrospray ionization
EI	electron ionization
EOF	electroosmotic flow
ESI	electrospray ionization
FDR	false discovery rate
FT-ICR	Fourier transform ion cyclotron resonance
FWMH	full width at half maximum
GC	gas chromatography
HCA	hierarchical cluster analysis
HILIC	hydrophilic interaction liquid chromatography
HMDB	human metabolome database
IMS-MS	ion mobility mass spectrometry
LC	liquid chromatography
LPS	lipopolysaccharides
m/z	mass-to-charge
MALDI	matrix assisted laser desorption ionization
MEKC	micellar electrokinetic chromatography
MRM	multiple reaction monitoring
MS	mass spectrometry
MSI	metabolomics standards initiative
MTBSFA	N-methyl-N-(tert-butyl)dimethylsilyl trifluoroacetamide
MTSFA	N-methyl-N-(trimethyl)silyl trifluoroacetamide
MVA	multivariate analysis
NMR	nuclear magnetic resonance
NP	normal phase

OPLS-DA	orthogonal partial least-square discriminative analysis
PAH	polycyclic aromatic hydrocarbon
PARAFAC	parallel factor analysis
PBS	phosphate buffer saline
PCA	principle component analysis
PLS-DA	partial least-square discriminative analysis
Q^2	prediction statistic
QC	quality control
RP	reverse phase
SIM	single ion monitoring
SPE	solid phase extraction
SPME	solid phase micro-extraction
TCA	citric acid cycle
TMCS	trimethylchlorosilane
TMS	trimethylsilyl group
TOF	time-of-flight
UHPLC	ultra-high pressure liquid chromatography

Chapter 2

ACN	acetonitrile
AMP	adenosine monophosphate
BD	Bligh and Dyer
CE	ceramides
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
EtOH	ethanol
FBS	fetal bovine serum
HILIC	hydrophilic interaction liquid chromatography
IS	internal standards for peak intensity normalization
LC	liquid chromatography
LCM	L929-cell conditioned medium
MeOH	methanol
MS	mass spectrometry
OPLS-DA	orthogonal partial least-square discriminative analysis
P/S	penicillin-streptomycin
PA	phosphatidic acid
PBS	phosphate buffer saline

PC	phosphatidylcholine
PCA	principle component analysis
PE	phosphatidylethanolamine
PG	phosphatidylglycerols
PS	phosphatidylserines
QC	quality control
RP	reverse phase
RS	recovery standards for recovery determination
SM	sphingomyelins
snthr	signal-to-noise threshold
TAG	triacylglycerols
TIC	total ion chromatogram
TOF	time-of-flight

Chapter 3

A	aerobic
<i>ackA</i>	acetate kinase
<i>acrB</i>	ornithine carbamoyltransferase
<i>adhE</i>	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<i>adhP</i>	alcohol dehydrogenase
ADI	arginine deiminase pathway
<i>ahpC</i>	alkyl hydroperoxide reductase, subunit C
<i>ahpCF</i>	alkyl-hydroperoxidase system
<i>ahpF</i>	alkyl hydroperoxide reductase, subunit F
<i>aldB</i>	aldehyde dehydrogenase
AN	anaerobic
<i>arcA</i>	arginine deiminase
<i>arcC</i>	carbamate kinase
<i>arcD</i>	arginine/ornithine antiporter
<i>bta</i>	bacteriocin accessory protein
CFU	colony forming units
CHP	conserved hypothetical protein
CRISPR	clustered regularly interspaced short palindromic repeats
<i>dps</i>	peroxide resistance protein
<i>fabH</i>	beta-ketoacyl-acyl carrier protein synthase III
<i>fabM/phaB</i>	enoyl-CoA hydratase protein
<i>fba</i>	fructose-1,6-bisphosphate aldolase
FMN	flavin mononucleotide

<i>glcK</i>	glucokinase
GlcNAc	N-acetylglucosamine
<i>glgA</i>	glycogen synthase
<i>glgABCD</i>	glycogen biosynthesis operon
<i>glgB</i>	glycogen branching enzyme
<i>glgC</i>	glucose-1-phosphate adenylyltransferase
<i>glgD</i>	glucose-1-phosphate adenylyltransferase, GlgD subunit
k_{app}	apparent retention factor
LC	liquid chromatography
<i>ldh</i>	lactate dehydrogenase
<i>manM</i>	PTS system, mannose-specific IIC component
<i>manN</i>	PTS system, mannose-specific IID component
<i>mleS</i>	malate dehydrogenase
NAD	nicotinamide adenine dinucleotide
<i>nox</i>	NADH oxidase
OPLS-DA	orthogonal partial least-square discriminative analysis
PC	phosphatidylcholine
PCA	principle component analysis
PE	phosphatidylethanolamine
pfl	formate C-acetyltransferase
PG	phosphatidylglycerols
<i>pgm</i>	putative phosphoglucomutase/phosphomannomutase
PL	phospholipids
<i>ppc</i>	phosphoenolpyruvate carboxylase
PPP	pentose phosphate pathway
<i>pta</i>	phosphate acetyl/butaryltransferase
PTS	phosphotransferase system
<i>pulA2</i>	pullulanase
<i>pyK</i>	pyruvate kinase
<i>pyrBCDEF</i>	pyrimidine synthesis genes
Q^2	prediction statistic
QC	quality control
rRNA	ribosomal RNA
SAM	S-adenosyl methionine
SMG	<i>Streptococcus Anginosus/Milleri</i> Group
<i>sodA</i>	superoxide dismutase
THY	Todd Hewitt medium
UDP	uridine diphosphate

Chapter 4

CPS	capsular polysaccharide
EPS	exopolysaccharide
EPS I	succinoglycan
EPS II	galactoglucan
GlcNAc	N-acetyl-glucosamine
gly-phe	glycine-phenylalanine
HCA	hierarchical cluster analysis
IS	internal standards for peak intensity normalization
L1	mid-exponential phase
L2	early stationary phase
L3	late stationary phase
LC	liquid chromatography
LPS	lipopolysaccharide
M1-4	exponential phase
M5	stationary phase
Mb	megabase
MS	mass spectrometry
MS/MS	tandem MS
OPLS-DA	orthogonal partial least-squares discriminative analysis
PA	phosphatidic acid
PC	phosphatidylcholine
PCA	principle component analysis
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
phe-phe	diphenylalanine
Q2	prediction statistic
RS	recovery standards for recovery determination

Chapter 5

2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
CAT	cationic amino acid transport
EI	electron ionization
ESI	electrospray ionization
F6P	fructose-6-phosphate
FBS	fetal bovine serum

G6P	glucose-6-phosphate
GABA	γ -aminobutyric acid
GC-MS	gas chromatography-mass spectrometry
GlcN6P	glucosamine-6-phosphate
HIF-1 α	hypoxia-inducible factor-1 α
IL-1 β	interleukin-1 β
IS	internal standards for peak area normalization
LC-MS	liquid chromatography-mass spectrometry
LPS	lipopolysaccharide
m/z	mass-to-charge
M1	macrophage with pro-inflammatory phenotype
M2	macrophage with anti-inflammatory phenotype
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide
OPLS-DA	orthogonal partial least-square discriminative analysis
PBS	phosphate buffer saline
PC	phosphatidylcholine
PCA	principle component analysis
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
RS	recovery standards for recovery determination
SIM	single-ion-monitor
SIRS	systemic inflammatory response syndrome
Slc7a2	cationic amino acid transport subunit
TCA	citric acid cycle
TMCS	chlorotrimethylsilane
UDP-G	uridine diphosphate-glucose
UDP-GlcNAc	uridine diphosphate N-acetylglucosamine

Chapter 6

EPA	Environmental Protection Agency
ESI	electrospray ionization
GC-MS	gas chromatography-mass spectrometry
gly-phe	glycine-phenylalanine
IS	internal standards for peak area normalization
LC-MS	liquid chromatography-mass spectrometry
LC ₅₀	lethal concentration of chemical kill 50% of population
LD ₅₀	median lethal dose
m/z	mass-to-charge

MW	molecular weight
NPAH	nitrogen containing polycyclic aromatic hydrocarbon
OPLS-DA	orthogonal partial least-square discriminative analysis
PAC	polycyclic aromatic compounds
PAH	polycyclic aromatic hydrocarbon
PASH	polycyclic aromatic sulfur heterocycles
PCA	principle component analysis
phe-phe	diphenylalanine
RS	recovery standards for recovery determination
χ	electronegativity

Chapter 7

LPS	lipopolysaccharides
NF- κ B	nuclear factor- κ B
NP	normal phase
PAH	polycyclic aromatic hydrocarbon
RP	reverse phase

Chapter 1: Introduction to metabolomics, and its application on cell biology

1.1 Metabolomics

1.1.1 Introduction to metabolomics

Biological systems such as cells, tissues, organisms and humans are composed of genes, transcripts, proteins, and metabolites in highly complex and interactive networks. Very little is known regarding the molecular engineering of a biological system. How does it adapt to extrinsic environmental fluctuations such as changes in oxygen levels, nutrient levels and exposure to toxins? How do intrinsic changes such as genetic or enzymatic deletions and mutations, or aging affect the basic functions of a biological system? Systems biology including genomics, transcriptomics, proteomics and metabolomics, has been developed to provide insights into how a biological system works (Fig. 1). Metabolomic analysis of primary and secondary metabolites, the downstream product of the “omics cascade”, has been a new field of study in the last decade (Fig. 2). As the endpoint of gene and enzymatic expression and cellular activity, metabolomics provides a holistic phenotypic representation of an organism and provides insightful understanding in the fields of functional genomics¹, metabolic engineering, bio-marker discovery for diseases², environmental stresses³⁻⁵, cell communication mechanisms or quorum sensing⁶, and industrial biotechnological processes⁷.

Interest in analyzing metabolites in a biological system can be traced back to the 1940s when mass spectrometry (MS) based methods were used in the fields of medicine, clinical chemistry and plant biology^{8,9}. The development of chromatographic separation techniques in the 1960s also made studying individual metabolites possible. In 1971, Linus Pauling *et al.* conducted the first comprehensive metabolic profiling of human urine and breath vapor using gas-liquid partition chromatography, in which over 250 biological compounds were detected¹⁰. This marked the beginning of metabolomic research.

Metabolomics is the study of metabolites, small molecules with molecular weight often below 1500 Da, in a biological system¹¹. Metabolites are intermediates or products of metabolic processes in a biological system. The change in metabolites levels in a biological system can be regarded as the ultimate response to genetic or environmental changes¹². The term “metabolomics” was coined in 1998 by Steve Oliver and Douglas Kell¹³. Soon after, in 1999, Jeremy Nicholson, John Lindon and Elaine Holmes coined the term “metabonomics”¹⁴. In 2001, Oliver Fiehn defined “metabolomics” as “a comprehensive and quantitative analysis of all metabolites...of a biological system”¹⁵.

The term “metabolic profiling” was suggested by Christopher Clarke and John Haselden in 2008¹⁶. “Metabolomics”, “metabonomics” and “metabolic profiling” are used interchangeably; however, “metabolomics” is more commonly used in the current literature and will be used throughout this thesis.

1.1.2 Comprehensive and targeted metabolomics

The field of metabolomics can be broadly divided into targeted and comprehensive metabolomics. The term “metabolomics” is often referred to as comprehensive metabolomics, where qualitative and semi-quantitative analysis of all low molecular weight metabolite in a biological sample at a given time is performed. The number of metabolites in the plant species *Arabidopsis* has been estimated to be over 5000, and the number of metabolites in bacteria is expected to be an order of magnitude less than in plants¹⁷. Comprehensive metabolomics involves analyzing various chemical classes with wide ranges of polarity, solubility, and volatility. Comprehensive metabolomic includes and is not limited to the analyses of amino acids, organic acids, sugars, sugar alcohols, lipids, fatty acids and those metabolites undergo post-modifications like phosphorylation or methylation. The concentration range of metabolites can be over 9 orders of magnitude. Often, multiple orthogonal analytical techniques are required in order to expand metabolites coverage in a biological system, and many of the detected metabolic features will remain unidentified.

Comprehensive metabolomics are often used to provide a real-time, global survey of the biological state of an organism as a result of intrinsic or extrinsic perturbation. For example, *Sinorhizobium meliloti* cells grown on different carbon sources can be distinguished based on their metabolic profile¹⁸; yeast mutants can be differentiated and classified based on their metabolic profiles in the spent media¹; silent mutations in *Saccharomyces cerevisiae* that do not result in any overt phenotypical changes can be distinguished from changes in their intracellular metabolic profiles¹⁹. More important, one of the greatest implications of comprehensive metabolomics has been the discovery of biomarkers that aid in clinical diagnosis or provide insight in the homeostasis of a biological system by identifying perturbed metabolites and metabolic pathway(s). The term “biomarker” will be used loosely in this thesis to refer to those metabolites that are significantly changed as a result of experimental conditions. Diagnostic markers for liver injury²⁰, Alzheimer’s disease²¹, anaphylaxis²², cancer²³, cardiovascular diseases²⁴, inborn errors of metabolism²⁵ and exposures to environmental toxins²⁶ have been found using comprehensive metabolomics. Metabolon®, a life sciences company focussing on metabolomics research, has successfully implemented two diagnostic pipelines for prediabetes and diabetes using metabolic biomarkers. In addition to other “omics”

techniques, metabolomics has enhanced our understanding of the pathology and physiology of many biological systems.

It is challenging to perform comprehensive metabolomics due to the vast chemical diversity of metabolites and their wide range of concentrations. Conversely, a common approach is to analyze dozens to hundreds of specific metabolites that have shared functionalities or properties or present in common metabolic pathways. This is referred to as targeted metabolomics. The analysis of targeted metabolomics often requires the quantification of target metabolites and uses a single analytical method. Because only a small group of metabolites is analyzed, sample preparations and extracts can be tailored towards those metabolites, thus avoiding common drawbacks such as matrix effects. Unlike comprehensive metabolomics, which is hypothesis-generating, specific hypothesis and research aims and previous knowledge of the system are required for targeted metabolomics. Gordon *et al.* analyzed short-chain fatty acids from gut microbiota to understand the microbial contribution to host obesity²⁷. Li *et al.* targeted amine- and phenol-containing metabolites in bronchoalveolar lavage fluids to understand the pathophysiology of asthma²⁸. Moreover, targeted metabolomics is used for fluxomics using isotopically labelled substrate²⁹, which has been used to trace glucose and glutamine metabolism in pro-inflammatory macrophages³⁰. Targeted metabolomics is commonly practiced in clinical diagnosis, especially for metabolic diseases such as inborn errors of metabolism. Profiling of acylcarnitine or amino acids from blood spots is used to diagnose fatty acid oxidation disorder and diseases due to defective amino acid metabolism³¹. Targeted metabolomics is widely used not only in research labs but also in clinical and industrial sectors.

1.1.3 Implication of metabolomics to cell research

Metabolomic studies of animal models and human subjects can be influenced by experimental confounders such as diet, gut microbiota, age, gender, genetic background, and environmental factors³². Metabolomic studies of cellular organisms (i.e. bacteria, fungi, eukaryotic cell-lines) can provide insights into metabolic regulation as a result of intrinsic and extrinsic stressors, and yet maintain consistent experimental conditions across samples. Moreover, cell culture studies are easier to control and manipulate, less expensive, easier to interpret and require less turn-over time than animal studies. The interest in cellular metabolomics research started in the late 1960s^{33,34}. Since the 1990s, the complete genome of many microbes has been sequenced^{35,36}, and the biology of these model microorganisms is much better understood. Since then, cells have been used extensively in molecular mechanistic studies such as gene functions, diseases, pharmacokinetics, and toxicity researches. The continuing development in the field of metabolomics allows the characterization of both intracellular and extracellular

metabolites that are part of biochemical reactions of a whole organism, thus revealing connections between metabolic pathways in a living cell. Cell metabolomics has led to the identification of metabolic bottleneck in glycolysis³⁷ and in the citric acid (TCA) cycle of pro-inflammatory macrophages³⁰; It has also led to the discovery of bioactive new and novel microbial natural products³⁸. The development and potential of cell and microbial metabolomics has been reviewed by Zhang *et al.* and Mashego *et al.*^{39,40}.

1.1.4 Metabolomic workflow

The fundamental work-flow for either comprehensive or targeted metabolomics for cellular organism is essentially the same. Cell metabolomics consists of five sequential steps: (i) cell culture growth or stimulation, (ii) quenching metabolic activity and metabolite extraction, (iii) data acquisition using MS-based or nuclear magnetic resonance (NMR) spectroscopy techniques, (iv) statistical and chemometric analysis, (v) data interpretation linking metabolomics to biological process or identify biomarkers (Fig. 3).

1.2 Sample preparation and extraction

A comprehensive understanding of *in vivo* cellular regulation and metabolic networks in different environments, growth conditions or genetic perturbations requires the analyses of both intracellular and extracellular metabolites. The intracellular metabolome includes all the metabolites that make up the cell such as metabolites found inside a cell and subcellular compartments. Metabolites that are bound to the cell surface are often quantified as part of the intracellular metabolome. Extracellular metabolites are those metabolites that are found outside of a cell in cell-free supernatant, including metabolic end-products expelled from the cell and substrates used by the cell as nutrients. For microorganisms under exponential growth, intracellular metabolomes are often acquired in the log growth phase and extracellular metabolomes are taken at the stationary phase. The concentration of metabolites very rapidly reflects the changes of extracellular environment, and half-life of intracellular metabolites is on the order of a second or less⁴¹. Therefore, a combination of fast sampling, quenching and extraction processes is required to ensure accurate representation of both the intra- and extracellular metabolomes of a cellular organism.

1.2.2 Metabolism quenching

Sample preparation for cell metabolomic analysis can be performed by two approaches, sequential or simultaneous quenching and extraction processes^{40,42}. When performed simultaneously, the quenching solution also serves as the extraction solution, for which intracellular and extracellular metabolites are measured together. In most cases, sequential quenching and extraction were used to prepare cell extracts, for which

quenching is followed by separating a cell pellet from supernatant by centrifugation or filtration, after which intracellular and extracellular metabolites are extracted and measured separately. The advantage and disadvantage of sequential and simultaneous sample processing was reviewed by Mashego *et al.*⁴⁰. In this thesis, we will focus on reviewing techniques involved in sequential quenching and extraction processes for cell metabolomics.

The first step for cell metabolomics is to rapidly quench the enzymatic and cellular activity, thus avoiding changes in the metabolites levels that are not part of the experimental setup. Rapid quenching of cellular metabolic activity is commonly achieved by addition of organic solvents (i.e. methanol, ethanol), instant change of sample temperature to low (< -20°C) or high (>80°C) temperature, or by applying extreme pH with the addition of alkali (KOH, NaOH) or acidic solution (HClO₄, HCl, trichloroacetic acid)^{40,42,43}. Quenching buffer (i.e. HEPES, ammonium carbonate) is included in some of the above quenching processes to maintain ionic strength. Freezing (usually below -20°C) is the most common and easiest quenching method. It is important that the quenching procedure should instantly stop any cellular metabolic activity while imposing no cell membrane damage and loss of intracellular metabolites due to leakage^{40,42}. The sample integrity should be preserved during the quenching process and not result changes in the chemical and physical properties of metabolites and their levels^{40,42}. Metabolic quenching with cold methanol prior to extraction is one of the most frequently used methods for microorganisms, but can result greater than 60% intracellular metabolite loss for bacteria⁴⁴. Quenching with cold isotonic phosphate buffer saline (PBS) was shown to be sufficient to halt cellular activity and ATP metabolism⁴⁴. Similar metabolic results can be obtained by quenching with liquid nitrogen (-196°C), which is commonly applied to plant and animal cells, though damages to the cellular envelopes are reported⁴². The sample preparation and storage should be maintained at low temperature to avoid uncontrolled cell metabolism. The impact of quenching on the cellular metabolome has been widely studied^{44–48} and reviewed^{40,42,43}.

1.2.3 Metabolite extraction

Metabolite extraction is a key step in cell metabolomics. Prior to extracting intracellular metabolites from cells, the cells are washed with saline or PBS buffer with matching ionic strength. A washing step is needed to remove the extracellular matrix which adheres to the cell surface to reduce ionization, improve analytical sensitivity, and remove contamination from extracellular metabolites^{49,50}. The method used for extraction can directly influence the metabolome coverage and influence the choice of the analytical technique for data acquisition. The extraction technique aims to (i) efficiently recover metabolites from sample while maintaining sample integrity, (ii) remove interferences

and matrix such as salts and proteins, (iii) make extracts compatible with analytical instrumentation, and (iv) if necessary, concentrate trace metabolites⁴². Liquid extractions with organic solvents are commonly used for protein precipitation and extracting intracellular metabolites. A large scope of polar and lipid metabolites can be extracted from polar solvents such as methanol, ethanol, and methanol/ethanol-water mixtures⁵¹. Lipophilic metabolites (i.e. lipids) can be extracted using non-polar solvents such as chloroform, hexane, diethyl ether, and ethyl acetate. Biphasic separation of polar and lipophilic metabolites can be obtained using the Bligh and Dyer extraction with various ratios of methanol, chloroform and water⁵². The choice of extraction solvent can largely bias the classes of metabolites extracted and also the levels of those metabolites. Usually cold solvents ($< 0^{\circ}\text{C}$) are used for the extraction to maintain sample integrity. Though boiling hot extraction solvent is also used, poor recovery was noted for thermally labile metabolites⁴⁸. The extraction processes are usually carried out with vortex mixing, but other extraction process such as pressurized liquid extraction, sonication, bead beating, or microwaving can enhance extraction efficiency^{53,54}. After completing the extraction process, when needed, the organic solvent can be evaporated under vacuum or a gentle stream of nitrogen, and the remaining residue can be concentrated and reconstituted in solvents compatible with the analytical technique. Solid phase extraction (SPE) and solid-phase micro-extraction (SPME) can also be used for sample concentration, but it is more commonly used for targeted metabolomic analysis for select class(es) of metabolites. The common extraction protocols for targeted and comprehensive metabolomics have been summarized and reviewed^{40,43,50,55}.

For adherent mammalian cells, trypsinization is often used to detach the cells from the growth surface prior to extraction⁵⁶. However, trypsinization alters cellular integrity, and changes intracellular metabolic profiles^{47,57}. Teng *et al.* have suggested a rapid and simultaneous cell quenching and extraction process with cold methanol on PBS washed cells, and the cells are detached from the growth surface with the help of scraping⁴⁷. Modification of this technique has been applied to endothelial-like colon adenocarcinoma cell (SW480), epithelial cell (RPTE/TERT1), and macrophages^{51,58,59}.

Cell growth media, especially rich media, consists of complex mixtures of amino acids, vitamins, salts, and biological mixtures such as fetal bovine serum or yeast extracts. The high salt content (i.e. 75 mM in M9, 139 mM in RPMI-1640) can cause severe ionization suppression, especially for liquid chromatography (LC)-MS, and thus affect detection and quantification of extracellular metabolites. Therefore, extracellular metabolites from supernatants were extracted with the “dilute-and-shoot” strategy¹. Typical dilution factor between 1:1 and 1:10 were made with the supernatant and organic solvent (i.e. methanol, ethanol) mixture. In other cases, supernatants were dried and

reconstituted in MS-compatible solvent or derivatized directly for gas chromatography (GC)-MS analysis⁶⁰, which is an analytical technique less susceptible to ion-suppression.

1.3 Analytical instrumentation

Traditionally, cellular metabolites were quantified using enzyme-based assays and thin layer chromatography⁶¹. The maturation of MS techniques since the 1990s⁶², development of soft ionization techniques such as electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) in the 1980s^{63,64}, and the production of ever more sensitive MS instruments since the 2000s, have influenced the establishment and growth in the field of metabolomics. Though many metabolomic studies were conducted by NMR and direct MS analysis, i.e. direct infusion, matrix assisted laser desorption ionization (MALDI) MS, and desorption electrospray ionization (DESI) MS^{65,66}, this thesis will be focused on reviewing chromatography coupled MS techniques for cell metabolomics.

No single analytical platform is able to fully analyze the entire intracellular or extracellular metabolome. MS cannot discriminate between isobaric metabolites with identical monoisotopic mass values. Chromatographic separation prior to MS enhances the data quality for metabolomic research by reducing number of metabolites simultaneously entering the MS, thus reducing matrix effects and ionization suppression. Coupling chromatographic separation with MS can increase metabolome coverage and also provide a secondary dimension for metabolite identification. Currently, three chromatographic techniques are predominant in the MS-based metabolomics, i.e. GC, LC and capillary electrophoresis (CE) (Fig. 2). Multidimensional separations such as two-dimensional GC×GC and LC×LC have been used to achieve greater peak capacity and separation of complex biological mixtures. Because of the enormous diversities and concentrations of metabolites inside and outside of a cell, it is impossible to cover all metabolites with a single platform (Fig. 4). Applications and advantages of each chromatography MS coupled techniques for metabolomics are reviewed here.

1.3.2 GC-MS

GC-MS is suited for the analyses of volatile and derivatized non-volatile metabolites with high separation efficiency and resolution via the use of capillary GC. Metabolite analyzed with GC-MS needs to be thermally stable and volatile⁶⁷. Esters and short-chain alcohols, hydrocarbons and lipids can be analyzed directly with GC-MS⁶⁸. Non-volatile metabolites such as sugar organic acids, amino acids, nucleic acid must be derivatized before GC-MS analysis⁶⁹. Silylation with N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA), N-methyl-N-t-butyltrimethylsilyltrifluoroacetamide (MTBSTFA), bistrimethylsilyltrifluoroacetamide (BSTFA) etc. are the most common derivatization for

GC-MS, where hydrogens from polar groups (i.e. alcohol, phenol, carboxyl, amine, amide, and hydroxyl) are replaced with the less polar trimethylsilyl group (TMS). Amines and tertiary hydroxyl groups in amino acids and nucleic acids are less reactive, thus trimethylchlorosilane (TMCS) is needed as a catalyst during the derivatization reaction. Schummer *et al.* has compared effectiveness of various silylation methods (i.e. MTBSTFA, BSTFA) for derivatizing different classes of polar metabolites⁷⁰. Silylation can be applied to a wide diversity of metabolites and there is a large number of silylation reagents to choose from. The reaction is moisture sensitive and cannot be performed in protic solvents (i.e. methanol, ethanol, water). Other derivatization methods for polar metabolites for GC-MS analysis include alkylation (methylation), esterification and acylation, but they are used less extensively than silylation⁷¹.

Compared to LC, the stationary phase of GC has significantly less contribution in influencing metabolite separation. Metabolites are predominately separated on a GC column based on their boiling points. However, the choice of stationary phases, film thickness and the length of capillary GC column depends on the polarity and the volatility of metabolites of interest⁴³. Metabolites can be ionized by electron ionization (EI) or chemical ionization (CI) when analyzed with a GC-MS platform, with EI being the more frequently used technique. The reproducible mass spectra generated by EI allowed compound identification by matching to public mass spectra libraries such as the National Institute of Standards and Technologies⁷² and Metabolomics FiehnLib⁷³. Both of these spectral libraries incorporate retention indices, which take advantage of the highly reproducible retention time of GC, to allow identification of structural isomers with similar mass spectra. Freeware like the automated mass spectrometry deconvolution and identification system (AMDIS)⁷⁴ and Metab (a R package)⁷⁵ has automated the pipeline for metabolomic analysis using GC-MS. Over 100 metabolites can be separated and identified in a single GC-MS run⁶⁹. GC-MS has been used to profile microorganisms and mammalian cells in food science, toxicology studies, disease research and functional genomics studies for characterizing mutations^{60,68,76–80}. GC×GC-MS has been used for comprehensive metabolomic analysis of yeast cells⁸¹. With the development of automated, high-throughput data processing software including commercial software such as ChromaTOF (LECO), HyperChrom (Thermo Fisher Scientific) and open source software such as parallel factor analysis (PARAFAC)⁸², GC×GC-MS with enhanced separation capacity is expected to be a prominent metabolomics platform for the analysis of complex biological samples.

1.3.3 LC-MS

LC-MS has been the most frequently used technique for both targeted and comprehensive metabolomics. It is used for identification and quantification of polar,

non-polar, ionic and neutral metabolites and requires little to no derivatization. The choice of chromatographic column in LC separation strongly determines on metabolic coverage (Fig. 4). Both normal-phase (NP) and reverse-phase (RP) columns have been used in metabolomics. RP columns such as C₁₈ and C₈ are the most utilized in metabolomics for analyzing less-polar metabolites such as lipids and peptides. NP columns, especially hydrophilic interaction liquid chromatography (HILIC) column is ideal for the analysis of polar ionic compounds such as amino acids and nucleic acids which cannot be retained on RP column⁸³. Bare silica-HILIC is also able to retain phospholipids or polar lipids via an adsorption mechanism^{51,84}. Phospholipids can be separated on a silica-HILIC column based on their polar head group. Zwitterionic HILIC columns such as ZIC[®]-HILIC and ZIC[®]-cHILIC have enhancing resolving power and sensitivity towards nucleic acid and peptide metabolites compared to regular silica HILIC⁸⁵. LC columns packed with sub-3- μ m core-shell particles are the primary choice for metabolomic analysis with enhanced chromatographic resolution and efficiency⁸⁶. Similarly, enhanced separation can also be achieved with ultra-high pressure liquid chromatography (UHPLC) operating at high pressures (400-1000 bar) with sub-2- μ m particle columns⁸⁷.

Many comprehensive metabolomics studies using LC-MS are biased to either polar or non-polar metabolites by limiting the analyses to only RP or NP (HILIC) column. A few studies that reflect true a comprehensive metabolome have analyzed the polar fraction using NP(HILIC) and the nonpolar fraction using RP columns⁸⁸⁻⁹². More than 100 metabolites in *Escherichia coli* have been quantified by using both RP and HILIC LC-MS platforms⁹²; thousands of metabolic features (peaks with unique retention time and m/z value) were measured in *Plasmodium falciparum* using RP and aqueous NP LC-MS techniques⁹⁰. Two-dimensional-LC-MS or LC $\times$ LC-MS has also been adapted to the metabolomic analysis of *E. coli* and *Saccharomyces cerevisiae*⁹³. Moreover, a tandem LC-MS approach by coupling orthogonal RP and NP in one run has been applied to metabolomic studies of serum and *S. cerevisiae*^{94,95}. The LC separation of metabolites prior to MS ionization significantly decreases ionization suppression and thus enhances sensitivity.

Electrospray ionization (ESI) is the predominant ionization technique for LC-MS, although atmospheric pressure chemical ionization (APCI) is occasionally used for LC-MS metabolomic studies. Metabolomics often require that data be acquired in both positive and negative ESI modes. Ion suppression has been a major challenge for LC-ESI-MS metabolomic analysis as a result of high salt content⁹⁶. APCI also experiences ion suppression but to a lesser extent. Separation of metabolites using various LC technique, sample dilution and sample clean-up is critical in minimizing ion suppression.

Identification of metabolites in LC-MS is based on matching accurate mass or tandem MS/MS spectra with public databases such as the human metabolome database (HMDB)⁹⁷ and METLIN⁹⁸. However, unlike GC-MS, the LC-MS and LC-MS/MS spectra are less reproducible between LC-MS systems⁹⁹, thus in house spectra libraries have been constructed by several research groups.

1.3.4 CE-MS

CE-MS is used for the analysis of charged metabolites. CE has better separation efficiency than LC due to the flat profile of electroosmotic flow (EOF) as compared to the rounded laminar flow in LC. Capillary zone electrophoresis (CZE) is the dominant CE-MS technique for metabolomic analysis¹⁰⁰. Charged metabolites such as amino acids, organic acids and nucleic acid are separated based on charge and size, and neutral metabolites eluted with EOF without separation. Other CE methods such as micellar electrokinetic chromatography (MEKC)¹⁰¹ require the use of surfactant, but are able to separate both charged and neutral metabolites. The metabolite separation in CE occurs in electrolyte solution, and non-polar metabolites are insoluble in these solutions. CE is usually coupled to ESI-MS with the addition of a sheath-flow interface to enhance ionization, though dilution from sheath liquid can reduce sensitivity. Sheathless CE-ESI interface can also be found with micro- or nano-ESI. On-line sample pre-concentration via dynamic pH junction¹⁰² and multisection injection¹⁰³ capability of CE-MS allows better sensitivity and high throughput for metabolomic analysis.

1.3.5 Mass analyzers for mass spectrometry

The mass analyzer is a major component of the mass spectrometer where gas phase ions generated at the ionization source are separated based on their m/z values. Modern mass analyzers including triple quadrupole, single quadrupole, ion trap, orbitrap, time-of-flight (TOF), and Fourier transform ion cyclotron resonance (FT-ICR) are routinely used for metabolomic studies (Table 1). Triple quadrupole and single quadrupole are nominal mass analyzers commonly used for targeted metabolomics. Multiple reaction monitoring (MRM) and single ion monitoring (SIM), respectively, can enhance sensitivity and specificity of target metabolites in a complex biological matrix. Orbitrap, TOF, and FT-ICR are predominant in comprehensive metabolomic studies due to their ability to scan through a large range of m/z values with high mass accuracy and fast acquisition rate. Coupling a quadrupole with orbitrap and TOF allows acquisition of MS/MS spectra for metabolite identification. All of the above mass analyzers coupled to GC, LC and even CE platforms have been commercialized and utilized widely for metabolomic studies.

Ion-mobility MS (IMS-MS) separates gas-phase ions based on their differential mobility in a buffer gas under the influence of a weak electric field¹⁰⁴. The separation is

generally based on the shape and size of the metabolite as opposed to the charge-to-mass ratio separation in MS. Since its commercialization in 2006, it has been applied to the separation of peptides¹⁰⁵, lipids¹⁰⁶, nucleotides¹⁰⁷ and enantiomers¹⁰⁸. It allows an additional dimension of separation to the traditional chromatography and MS system (more commonly with LC-TOF system), and therefore, increases the peak capacity and improves the separation of metabolites that cannot otherwise be separated with traditional methods.

Table 1 Characterizations of common mass analyzers used for metabolomic studies

(adapted from Dass *et al.* and CHROMacademy^{109,110})

characteristic	Mass analyzers				
	Single/triple quadrupole	Ion trap	Orbitrap	TOF	FT-ICR
Mass range (m/z)	Up to 4000	Up to 4000	Up to 4000	Over 100,000	4000
Mass resolution (FWMH) ^a	Nominal mass	10,000	150,000	60,000	Up to millions
Dynamic range	10 ⁷	10 ² -10 ⁶	10 ² -10 ⁴	10 ² -10 ⁴	10 ⁴
Acquisition rate (Hz)	1-20	1-30	10	10-500	1

^a FWHM, full width at half maximum

1.4 Data processing, statistical and chemometric analyses

1.4.2 Software for metabolomic analysis

The automation of the data processing procedure is important for metabolomic studies. There are numerous commercialized (i.e. MakerLynx, ChromAlign, MarkerView) and freeware (i.e. XCMS^{111,112}, MZmine^{113,114}, MetaboAnalyst^{115,116}) packages dedicated to comprehensive metabolomic analyses. All of these software packages use open exchange data formats such as netCDF and mzXML for automated guided data filtering, peak picking, peak integration and peak alignment. XCMS online¹¹⁷ and MetaboAnalyst¹¹⁵ are online based packages that offer built-in statistical tools including univariate and multivariate statistics as well as data visualization tools such as heat maps and hierarchical cluster analysis (HCA). XCMS2 also is integrated with METLIN and offers automated metabolite identification with accurate mass and MS/MS spectral matching¹¹². XCMS can be used for data acquired in either low or high mass resolution. Chromatographic peaks are detected with a second-derivative Gaussian filter, thus removing dubious peaks with poor peak shape. XCMS also offers the flexibility of removing metabolic features that appear less frequently in samples which have undergone

the same treatment for additional quality assurance. XCMS, MZmine and metaboAnalyst have been adapted to data acquired using either LC-MS or GC-MS.

1.4.3 Quality Assurance

Without analytical standards, quality control and data validation, interpretation of metabolomic studies will be superficial and unreliable. The experimental setup needs to include method controls (i.e. cells which undergo no treatment), and where possible, positive and negative controls. Inadequate sample size can also bias the experimental outcome. For cell metabolomics, a biological sample size of 6-10 is recommended for better statistical reliability. Isotopically labelled standards (i.e. recovery standards) need to be spiked into the sample prior to extraction to monitor recovery efficiency. Similarly, different isotopically labelled standards (i.e. internal standards) need to be spiked into the extracted samples prior to data acquisition to correct for injection error and instrumentation variability via peak intensity normalization. The analysis of solvent blanks during data acquisition is also important to ensure that there is no sample carry-over. Moreover, post-column addition or serial dilutions of a pooled quality control sample provide information on the chromatographic region or metabolites that suffer suppression. Periodic analysis of pooled biological quality control (QC) samples that are composed of all samples need to be performed during long-term data acquisition to ensure the reproducibility of analytical instrumentation. The variability of the QC sample must be significantly below that of the actual samples. Dunn *et al.* have suggested excluding those metabolic features with more than 20-30% variance in the QC samples due to poor and unreliable instrumental reproducibility⁸⁸. For targeted metabolomics, calibration curves of targeted metabolites are often required. Moreover, during data analysis, low quality data such as those metabolic features with missing data, low intensity or high variance across experimental conditions can be removed. Mass calibration with internal standards (e.g. sodium formate) prior to data conversion and processing can improve the accuracy in the automated metabolic feature alignment in open source software (i.e. XCMS, MZmine).

1.4.4 Univariate and multivariate analyses

The aim of metabolomics is to identify metabolites whose concentrations are significantly different as a result of various experimental conditions. Statistical analysis (i.e. univariate and multivariate) is needed in order to discover and validate those metabolites that were significantly altered by treatments.

Univariate analysis is applied in parallel on all detected features. A priori or post-hoc power analysis can be conducted prior or after the research study, respectively, for sample size determination. Power analysis has been extended to multivariate datasets seen in

metabolomic analysis¹¹⁸. Univariate statistical significant test (or hypothesis testing) can be divided into two groups, parametric and non-parametric, based on the variable distribution (Table 2)¹¹⁹. Normality, homogeneity and independence of variables are assumed in parametric statistical tests such as Student's t test and ANOVA. For sample sizes greater than 50, the Kolmogorov-Smirnov test is used to statistically evaluate normality¹¹⁹. Levene's and Bartlett tests can be used to statistically evaluate homoscedasticity, where $p < 0.05$ indicates heteroscedasticity in sample variance¹¹⁹. Finally, samples in metabolomic studies are considered independent. Parametric tests are more powerful than non-parametric tests; with non-parametric tests results show more type II error (false negative). Metabolomic datasets are not entirely normal or homogeneously distributed¹¹⁹. Vinaixa *et al.* have shown a minor discrepancy (<4%) between parametric and non-parametric tests on four comprehensive LC-MS metabolomic datasets¹¹⁹. Though non-normally distributed data can be transformed to normal via logarithmic transformation, data transformation should be handled carefully as it may alter data integrity and hinder data interpretation.

The greatest limitation of univariate statistical tests for metabolomic studies and biomarker discoveries is type I error (false positive)^{119,120}. Data reduction can reduce the probability of type I error. Bonferroni¹²⁰ and Benjamini-Hochberg¹²¹ corrections can also reduce the probability of getting type I error, though, Bonferroni correction is often thought to be too conserved for metabolomic studies. However, the ultimate strategies for eliminating type I error in metabolomic research are to integrate the significantly changed metabolites in metabolic pathways or to other omic studies, or validate those metabolites with further biological studies.

Table 2 Parametric and non-parametric univariate statistical tests for datasets following normal distribution or far from normal-curve

(modified from Vinaixa *et al.*¹¹⁹)

	Normal distribution (Parametric)	Far from normal-curve (non- parametric)
	Compare Means ^a	Compare Medians ^b
Compare two unpaired groups	Unpaired Student's t test	Mann-Whitney
Compare two paired groups	Paired Student's t test	Wilcoxon signed-rank
Compared more than two unmatched group	One-way ANOVA	Kruskal-Wallis
Compared more than two matched group	Two-way ANOVA	Friedman

^aMean is the average as defined by $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$;

^bMedian is the middle value that separate the higher from the lower half of a data sample

Multivariate analyses (MVA) correlate the experimental conditions to the changes of hundreds or even thousands metabolic features simultaneously. MVA can be divided into supervised and unsupervised approaches. Principal component analysis (PCA) is an unsupervised statistical approach where samples are not classified based on treatment groups. Partial-least square-discriminative analysis (PLS-DA) and orthogonal partial-least square-discriminative analysis (OPLS-DA) are unsupervised statistical approaches where samples are categorized and annotated with corresponding experimental treatments. Supervised PLS-DA and OPLS-DA provide better discrimination between treatments than the unsupervised PCA approach, and changes in metabolic features that do not correlate to changes in experimental setup are excluded in the supervised MVA approach¹²². PCA, PLS-DA and OPLS-DA projects the original dataset on a lower dimensional space, where the variability of a large metabolomic dataset can be explained by few components (or factors). OPLS-DA is largely equivalent to PLS-DA, except the components are explained in the axial directions (for PLS-DA, the components are described by a combination of x and y axes).

All PCA, PLS-DA and OPLS-DA are susceptible to large dynamic range of metabolic features and biased to those with greater abundancies. Pretreating metabolomic dataset with various scaling methods such as centering, autoscaling, pareto scaling, log transformation etc. can reduce the bias of significant metabolites to those metabolites with high abundances¹²³. Of these, autoscaling and pareto scaling are the two most common scaling methods used for metabolomics studies. Moreover, the tendency of overfitting could also be a major problem for supervised MVA. Therefore, internal validation or permutation testing, where the metabolomic dataset is artificially split into training and validation sets, is much needed. The validation dataset is projected through a model build previously with the training set, where the fitness of model build with training set is evaluated with R^2X and R^2Y , and the internal validation is evaluated by the prediction statistic (Q^2)¹²⁰. Both R^2 and Q^2 followed an upward trend from 0 to 1. For an over fit model, R^2 approaches 1, and Q^2 falls toward 0¹²⁰. A prediction statistic (Q^2) of 0.4-0.7 is indicative of a robust model, i.e. true differences exist between the compared groups, and Q^2 above 0.7 indicates that the model is highly robust^{120,124}. SIMCA-P, software for MVA, has built-in internal validation for PLS-DA and OPLS-DA.

Both univariate and multivariate statistical approaches are equally important in metabolomic research, and provide complimentary information for discovering significantly changed metabolites^{125,126}. The results from univariate and multivariate analyses do not necessarily coincide, and the data should be interpreted within each approach.

1.5 Challenges in cell metabolomics

The vast diversity and concentration of metabolites inside and outside of the cell makes cell metabolomics very challenging. Growth conditions, sample preparation (i.e. quenching and extraction), separation and detection can all concurrently affect the true representation of cell metabolome and thus the biological interpretation. Subtle differences between cell culture and growth conditions have made the cross comparison between studies difficult. Addition of serum of animal origin or yeast extract in the rich media has severely complicated the biological matrix. Moreover, the high salt levels in the media also lead to severe ion suppression during sample analysis, especially for LC-MS systems. Very few metabolomic studies have addressed the challenges of matrix effect and ion suppression.

The sample preparation (i.e. quenching, extraction) required for cell metabolomics must be standardized. Comparative studies on the sample preparation protocol for mammalian inherent cells are extremely limited. Issues regarding cell leakage and metabolites alteration during cell quenching need to be addressed. A quick and easy sample preparation that recovers the majority of metabolites and yet minimizes matrix effects with minimal sample manipulation is always in demand for high throughput and reproducible metabolomic analysis. Automatization using robotic sample preparation is gaining popularity, especially for the on-line derivitization step required for GC-MS analyses. Metabolite integrity needs to be preserved during sample preparation in order to ensure a true correlation between metabolic profiles and experimental setups.

Analytical instrumentation for sample acquisition is improving continuously with better separation efficiency, better sensitivity, and faster acquisition. However, there is no universal analytical platform that is suitable for analyzing all metabolites. Comprehensive metabolomic will continue to rely on a combination of multiple analytical platforms in order to cover the majority of the metabolites.

Identification of unknown metabolites has always been the bottleneck for metabolomic studies. This process relies heavily on matching tandem MS/MS spectra of the unknown to standards. Moreover, the ion intensity of the unknown metabolites will ultimately affect the quality of the MS/MS spectrum, and making identification of minor metabolites even more difficult. Current publicly available spectra libraries (i.e HMDB, METLIN) are focused on human or mouse metabolites, and the reference library is still incomplete. Metabolite identification is essential to integrate metabolomics with other “-omics” studies. However, metabolic pathways for many cell organisms still remain unknown or at the putative level.

Misuse of statistical tools can lead to completely erroneous results and biomarkers. Bias and inadequate sample size, excessive false discovery rate, and overfitting with lack of cross validation can all lead to apparently persuasive and yet dubious biomarkers or significantly changed metabolites. Such issues have been addressed by Broadhurst and Kell¹²⁰. Ultimately, biological models need to be built according to metabolomic findings and validated.

Metabolomics needs to be integrated with other omics approaches i.e. genomics, transcriptomics and proteomics, to provide a comprehensive, cross-validated overview of biological systems. Databases like MetaCyc and BioCyc have integrated metabolic pathways with proteomic and genomics, respectively, for over 3000 organisms¹²⁷. Standard reporting for metabolomics has been implemented through the metabolomics standards initiative (MSI) to facilitate experimental replication and data exchange^{128–131}. The initiatives standardized the metabolomic practices and greatly assist the integration of metabolomics to system biology study.

1.6 Thesis objectives

The overall goal of this thesis includes (i) developing an extraction and analytical protocol that is suitable for the comprehensive analysis of both polar and non-polar metabolites in either intra- and extra-cellular media; and (ii) applying this methodology to the study of various *in vitro* cell organisms in areas of growth environment, gene functions, gerontology, and toxicology research. During the course of this research, a workflow for treating complex metabolomic data set was developed.

Chapter 2 summarizes a small scale cell extracting protocol using MeOH/EtOH/H₂O (2:2:1) with bead beating for simultaneous extraction of both polar and lipid cellular metabolites. A complimentary LC-MS protocol using silica-HILIC was developed that enables separation and analysis of both polar and lipid metabolites in a single run. This methodology was subsequently applied to the following four projects in Chapters 3-6.

Chapter 3 provides an integrated overview of *Streptococcus intermedius* in anaerobic and anaerobic environments using a comprehensive metabolomic and transcriptomic approach. The effect of oxygen on the growth, physiology and metabolism of *S. intermedius* provides insight into understanding its pathogenic association with anaerobes in polymicrobial infections.

Chapter 4 analyzes the differential metabolomic behaviour between three mutant strains of *S. meliloti*, each with either pSymA, pSymB or both pSymA and pSymB megaplasmid(s) removed, and compared to the wildtype strain. pSymA and pSymB are essential in establishing symbiosis with the legume host and important for bacterial

fitness in the rhizosphere. However, their involvement in cellular function in the absence of the legume environment was poorly characterized, and this study allowed further understanding of the metabolic functions of these megaplasmids.

Chapter 5 characterizes the age-associated metabolic dysregulation in murine bone marrow-derived macrophages during inflammation. By comprehensively analyzing the intracellular metabolome and also targeting at the core metabolites of bone marrow-derived macrophages from young (6-8 wk) and old (18-22 mo) mice following bacterial lipopolysaccharide (LPS) stimulation and tolerance, we were able to provide more insight into metabolic regulation and bottlenecks that may hinder macrophage function with age.

Chapter 6 surveyed the metabolic changes in *Sinorhizobium meliloti* as a result of sub-lethal polycyclic aromatic hydrocarbon (PAH) mixture or fluorene (a PAH) exposure. From the metabolic profiles, we can conclude the effect of PAHs are additive and induce dose dependent responses in *S. meliloti*.

A general discussion is presented in Chapter 7, focusing on the implication of comprehensive and targeted metabolomic research in advancing cellular biology.

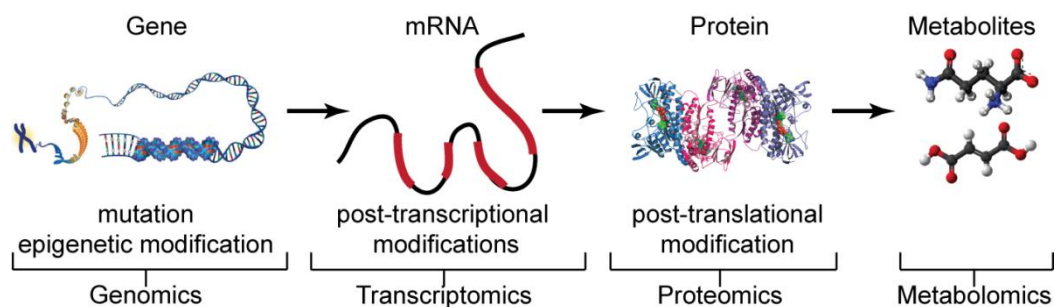


Figure 1 The central dogma of biology and the “omic cascade”. Metabolites are downstream biochemical end products of epigenetic process and post-translational modifications of genes and proteins. There is a stronger correlation between metabolic profiles and phenotypes compared to genomics, transcriptomics and proteomics.

(figure is adapted from Patti et al.¹³²)

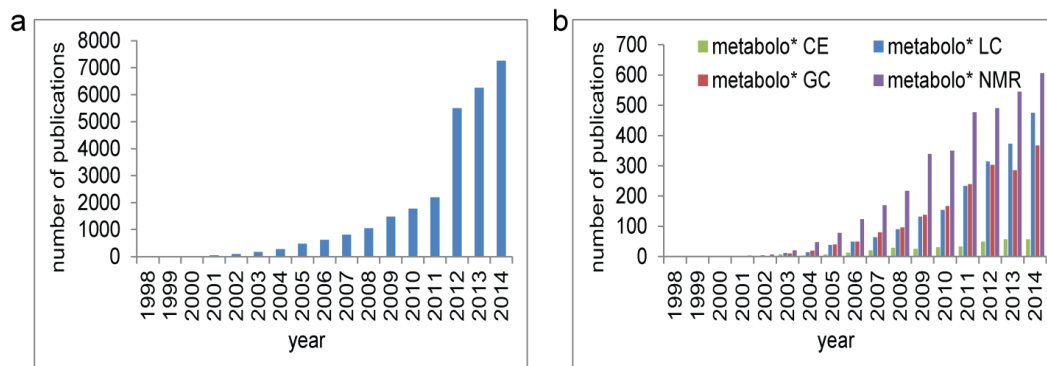


Figure 2 The growth of metabolomics since 1998. a) The number of publications containing “metabolo*” in 1998-2014 as per Web of Science. b) The number of metabolomic publications in 1998-2014 was classified according to the analytical platform used. “Metabolo* LC”, “metabolo* GC”, “metabolo* CE” and “metabolo* NMR” were used as key words in Web of Science. “Metabolo*” was used as the key words to include terms such as metabolomics, metabolomic, and metabolome.

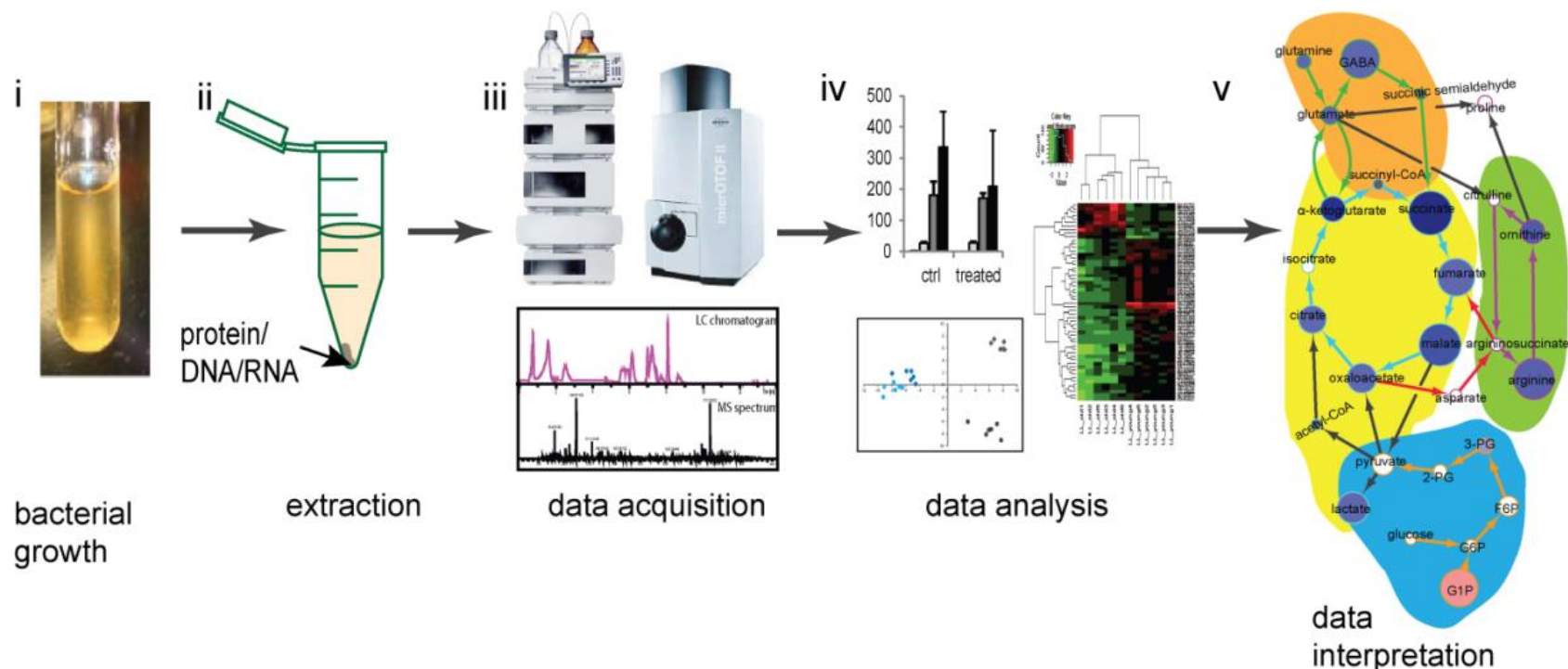


Figure 3 Metabolomic workflow. Cell metabolomics consists of five sequential steps: (i) cell culture growth or stimulation, (ii) quenching metabolic activity and metabolite extraction, (iii) data acquisition using MS-based spectroscopy techniques to generate chromatograms and MS spectra, (iv) statistical and chemometric analysis including univariate and multivariate analyses, (v) data interpretation linking metabolomics to biological process using metabolic network or identify biomarkers

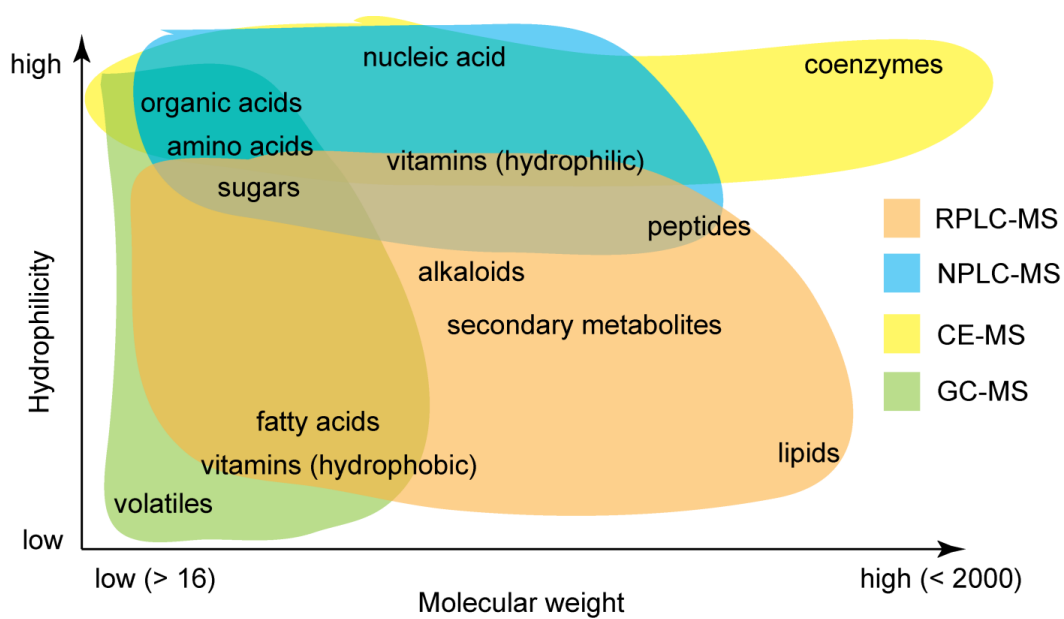


Figure 4 Conceptual coverage of metabolites using chromatography-MS coupled techniques

(figure adapted from Kusano et al. ¹³³)

1.7 References

1. Allen, J. *et al.* High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21**, 692–696 (2003).
2. Kaddurah-Daouk, R., Kristal, B. S. & Weinshilboum, R. M. Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* **48**, 653–683 (2008).
3. Brown, S. A. E., McKelvie, J. R., Simpson, A. J. & Simpson, M. J. ¹H NMR metabolomics of earthworm exposure to sub-lethal concentrations of phenanthrene in soil. *Environ. Pollut.* **158**, 2117–2123 (2010).
4. Vandenbrouck, T., Jones, O. A. H., Dom, N., Griffin, J. L. & De Coen, W. Mixtures of similarly acting compounds in *Daphnia magna*: From gene to metabolite and beyond. *Environ. Int.* **36**, 254–268 (2010).
5. Allen, J. *et al.* Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl. Environ. Microb.* **70**, 6157–6165 (2004).
6. Peiris, D. *et al.* Metabolite profiles of interacting mycelial fronts differ for pairings of the wood decay basidiomycete fungus, *Stereum hirsutum* with its competitors *Coprinus micaceus* and *Coprinus disseminatus*. *Metabolomics* **4**, 52–62 (2007).
7. Chellapandi, P., Sivaramakrishnan, S. & Viswanathan, M. B. Systems biotechnology: an emerging trend in metabolic engineering of industrial microorganisms. *J. Comput. Sci. Syst. Biol.* **03**, 43–49 (2010).
8. Elvehjem, C. A., Madden, R. J., Strong, F. M. & Woolley, D. W. Relation of nicotinic acid and nicotinic acid amide to canine black tongue. *J. Am. Chem. Soc.* **59**, 1767–1768 (1937).
9. Blodgett, K. B. Monomolecular films of fatty acids on glass. *J. Am. Chem. Soc.* **56**, 495–495 (1934).
10. Pauling, L., Robinson, A. B., Teranishit, R. O. Y. & Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Nat. Acad. Sci.* **68**, 2374–2376 (1971).
11. Bruce, S. J. *et al.* Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal. Chem.* **81**, 3285–3296 (2009).

12. Fiehn, O. Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–71 (2002).
13. Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378 (1998).
14. Nicholson, J. K., Lindon, J. C. & Holmes, E. “Metabonomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).
15. Fiehn, O. Combining genomics , metabolome analysis , and biochemical modelling to understand metabolic networks. *Comp. Funct. Genom.* **2**, 155–168 (2001).
16. Clarke, C. J. & Haselden, J. N. Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicol. Pathol.* **36**, 140–147 (2008).
17. Hall, R. *et al.* Plant metabolomics: the missing link in functional genomics strategies. *Plant Cell* **14**, 1437–1440 (2002).
18. Barsch, A., Patschkowski, T. & Niehaus, K. Comprehensive metabolite profiling of *Sinorhizobium meliloti* using gas chromatography-mass spectrometry. *Funct. Integr. Genomics* **4**, 219–230 (2004).
19. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**, 45–50 (2001).
20. Soga, T. *et al.* Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Biol. Chem.* **281**, 16768–16776 (2006).
21. Sato, Y. *et al.* Identification of a new plasma biomarker of Alzheimer’s disease using metabolomics technology. *J. Lipid Res.* **53**, 567–576 (2012).
22. Chalcraft, K. R., Kong, J., Waserman, S., Jordana, M. & McCarry, B. E. Comprehensive metabolomic analysis of peanut-induced anaphylaxis in a murine model. *Metabolomics* **10**, 452–460 (2013).
23. Lewis, G. D., Asnani, A. & Gerszten, R. E. Application of metabolomics to cardiovascular biomarker and pathway discovery. *J. Am. Coll. Cardiol.* **52**, 117–123 (2008).

24. Nam, H., Chung, B. C., Kim, Y., Lee, K. & Lee, D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. *Bioinformatics* **25**, 3151–3157 (2009).
25. Shlomi, T., Cabili, M. N. & Ruppin, E. Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol. Syst. Biol.* **5**, 263 (2009).
26. McKelvie, J. R., Yuk, J., Xu, Y., Simpson, A. J. & Simpson, M. J. ¹H NMR and GC/MS metabolomics of earthworm responses to sub-lethal DDT and endosulfan exposure. *Metabolomics* **5**, 84–94 (2009).
27. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
28. Peng, J., St. Laurent, C. D., Befus, a. D., Zhou, R. & Li, L. Metabolomic profiling of bronchoalveolar lavage fluids by isotope labeling liquid chromatography mass spectrometry: a promising approach to studying experimental asthma. *Metabolomics* **10**, 1305–1317 (2014).
29. Winter, G. & Krömer, J. O. Fluxomics - connecting 'omics analysis and phenotypes. *Environ. Microbiol.* **15**, 1901–1916 (2013).
30. Jha, A. K. *et al.* Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. *Immunity* **42**, 419–430 (2015).
31. Rashed, M. S., Ozand, P. T., Bucknall, M. P. & Little, D. Diagnosis of inborn errors of metabolism from blood spots by acylcarnitines and amino acids profiling using automated electrospray tandem mass spectrometry. *Pediatr. Res.* **38**, 324–331 (1995).
32. Patterson, A. D. & Idle, J. R. A metabolomic perspective of small molecule toxicity. *General, Applied and System Toxicology* (2009).
33. Harrison, D. E. F. & Maitra, P. K. Control of respiration and metabolism in growing *Klebsiella aerogenes*. *Biochem. J.* **112**, 647–656 (1969).
34. Gancedo, J. M. & Ganceno, C. Concentrations of intermediary metabolites in yeast. *Biochimie* **55**, 205–211 (1973).
35. Fraser, C. M., Eisen, J. A. & Salzberg, S. L. Microbial genome sequencing. *Nature* **406**, 799–803 (2000).

36. Kyrpides, N. C. *et al.* Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS biology* **12**, e1001920 (2014).
37. Theobald, U., Mailinger, W., Reuss, M. & Rizzi, M. in vivo analysis of glucose induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Anal. Biochem.* **214**, 31–37 (1993).
38. Krug, D. & Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **31**, 768–783 (2014).
39. Zhang, A., Sun, H., Xu, H., Qiu, S. & Wang, X. Cell metabolomics. *OMICS* **17**, 495–501 (2013).
40. Mashego, M. R. *et al.* Microbial metabolomics: past, present and future methodologies. *Biotechnol. Lett* **29**, 1–16 (2007).
41. de Koning, W. & van Dam, K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal. Biochem.* **204**, 118–123 (1992).
42. Álvarez-Sánchez, B., Priego-Capote, F. & Castro, M. D. L. De Metabolomics analysis II. Preparation of biological samples prior to detection. *Trac-Trend Anal. Chem.* **29**, 120–127 (2010).
43. Villas-Bôas, S. G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* **24**, 613–646 (2005).
44. Bolten, C. J., Kiefer, P., Letisse, F., Portais, J.-C. & Wittmann, C. Sampling for metabolome analysis of microorganisms. *Anal. Chem.* **79**, 3843–3849 (2007).
45. Wittmann, C., Krömer, J. O., Kiefer, P., Binz, T. & Heinzle, E. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal. Biochem.* **327**, 135–139 (2004).
46. Winder, C. L. *et al.* Global metabolic profiling of *Escherichia coli* cultures : an evaluation of methods for quenching and extraction of intracellular metabolites. *Anal. Chem.* **80**, 2939–2948 (2008).
47. Teng, Q., Huang, W., Collette, T. W., Ekman, D. R. & Tan, C. A direct cell quenching method for cell-culture based metabolomics. *Metabolomics* **5**, 199–208 (2009).

48. Maharjan, R. P. & Ferenci, T. Global metabolite analysis : the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal. Biochem.* **313**, 145–154 (2003).
49. Wu, Y. & Li, L. Development of isotope labeling liquid chromatography-mass spectrometry for metabolic profiling of bacterial cells and its application for bacterial differentiation. *Anal. Chem.* **85**, 5755–5763 (2013).
50. Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem* **403**, 1523–1548 (2012).
51. Fei, F., Bowdish, D. M. E. & McCarry, B. E. Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal. Bioanal. Chem.* **406**, 3723–3733 (2014).
52. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Phys.* **37**, 911–917 (1959).
53. Smith, R. M. Before the injection - modern methods of sample preparation for separation techniques. *J. Chromatogr. A* **1000**, 3–27 (2003).
54. Geier, F. M., Want, E. J., Leroi, A. M. & Bundy, J. G. Cross-platform comparison of *Caenorhabditis elegans* tissue - extraction strategies for comprehensive metabolome coverage. *Anal. Chem.* **83**, 3730–3736 (2011).
55. Keseler, I. M. *et al.* EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**, D334–D337 (2005).
56. Lane, A. N. & Fan, T. W.-M. Quantification and identification of isotopomer distributions of metabolites in crude cell extracts using ¹H TOCSY. *Metabolomics* **3**, 79–86 (2007).
57. Bi, H. *et al.* Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells. *Anal. Bioanal. Chem* **405**, 5279–5289 (2013).
58. Dettmer, K. *et al.* Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Anal. Bioanal. Chem.* **399**, 1127–1139 (2011).

59. Ellis, J. K. *et al.* Metabolic response to low-level toxicant exposure in a novel renal tubule epithelial cell system. *Mol. BioSyst.* **7**, 247–257 (2011).
60. Villas-Bôas, S. G., Noel, S., Lane, G. a, Attwood, G. & Cookson, A. Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Anal. Chem.* **349**, 297–305 (2006).
61. Iwata, T. & Yamasaki, K. Enzymatic determination and thin-layer chromatography of bile acids in blood. *J. Biochem.* **56**, 424–431 (1964).
62. Beynon, J. H. & Morgan, R. P. The development of mass spectroscopy: an historical account. *Int. J. Mass Spectrom. Ion Phys* **27**, 1–30 (1978).
63. Yamashitat, M. & Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **434**, 4451–4459 (1984).
64. Carroll, D. I., Dzidic, I., Stillwell, R. N., Horning, M. G. & Horning, E. C. Subpicogram detection system for gas phase analysis based upon atmospheric pressure ionization (API) mass spectrometry. *Anal. Chem.* **46**, 706–710 (1974).
65. Bedair, M. & Sumner, L. W. Current and emerging mass-spectrometry technologies for metabolomics. *Trac-Trend Anal. Chem.* **27**, 238–250 (2008).
66. Lei, Z., Huhman, D. V. & Sumner, L. W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **286**, 25435–25442 (2011).
67. Fang, M. *et al.* Thermal degradation of small molecules: a global metabolomic investigation. *Anal. Chem.* (2015).doi:10.1021/acs.analchem.5b03003
68. Wyllie, S. G., Brophy, J. J., Sarafis, V. & Hobbs, M. Volatile components of the fruit of Pistacia Lentiscus. *J. Sci. Food Agric.* **55**, 1325–1326 (1990).
69. Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N. & Willmitzer, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography - mass spectrometry. *Plant J.* **23**, 131–142 (2000).
70. Schummer, C., Delhomme, O., Appenzeller, B. M. R., Wennig, R. & Millet, M. Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta* **77**, 1473–1482 (2009).
71. Wells, R. J. Recent advances in non-silylation derivatization techniques for gas chromatography. *J. Chromatogr. A* **843**, 1–18 (1999).

72. Babushok, V. I. *et al.* Development of a database of gas chromatographic retention properties of organic compounds. *J. Chromatogr. A* **1157**, 414–421 (2007).
73. Kind, T. *et al.* FiehnLib : mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **81**, 10038–10048 (2009).
74. Davies, A. N. The new automated mass spectrometry deconvolution and identification system (AMDIS). *Spectrosc. Eur.* **3**, 24–27 (1998).
75. Aggio, R., Villas-Bôas, S. G. & Ruggiero, K. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* **27**, 2316–2318 (2011).
76. Koek, M. M., Muilwijk, B., van der Werf, M. J. & Hankemeier, T. Microbial metabolomics with gas chromatography/mass spectrometry. *Anal. Chem.* **78**, 1272–1281 (2006).
77. Rodríguez-Prados, J.-C. *et al.* Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* **185**, 605–614 (2010).
78. Keum, Y. S., Seo, J. S., Li, Q. X. & Kim, J. H. Comparative metabolomic analysis of *Sinorhizobium* sp. C4 during the degradation of phenanthrene. *Appl. Microbiol. Biotechnol.* **80**, 863–872 (2008).
79. Basconcillo, L. S., Zaheer, R., Finan, T. M. & McCarry, B. E. Cyclopropane fatty acyl synthase in *Sinorhizobium meliloti*. *Microbiology* **155**, 373–385 (2009).
80. Wong, H. S., Maker, G. L., Trengove, R. D. & O’Handley, R. M. Gas chromatography-mass spectrometry-based metabolite profiling of *Salmonella enterica* serovar typhimurium differentiates between biofilm and planktonic phenotypes. *Appl. Environ. Microbiol.* **81**, 2660–2666 (2015).
81. Humston, E. M., Dombek, K. M., Hoggard, J. C., Young, E. T. & Synovec, R. E. Time-dependent profiling of metabolites from Snf1 mutant and wild type yeast cells. *Anal. Chem.* **80**, 8002–8011 (2008).
82. Hoggard, J. C. & Synovec, R. E. Parallel factor analysis (PARAFAC) of target analytes in GC×GC-TOFMS data : automated selection of a model with an appropriate number of factors. *Anal. Chem.* **79**, 1611–1619 (2007).

83. Tang, D., Zou, L., Yin, X. & Ong, C. N. HILIC-MS for metabolomics: an attractive and complementary approach to RPLC-MS. *Mass Spectrom. Rev.* 1–27 (2014).doi:10.1002/mas
84. Schwalbe-Herrmann, M., Willmann, J. & Leibfritz, D. Separation of phospholipid classes by hydrophilic interaction chromatography detected by electrospray ionization mass spectrometry. *J. Chromatogr. A* **1217**, 5179–5183 (2010).
85. Palma, S. Di, Boersema, P. J., Heck, A. J. R. & Mohammed, S. Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis. *Anal. Chem.* **83**, 3440–3447 (2011).
86. Gritti, F. & Guiochon, G. Facts and legends about columns packed with sub-3- μ m core - shell particles. *LC GC North America* **30**, 586 (2012).
87. Wilson, I. D. *et al.* High resolution “ultra performance” liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies research articles. *J. Proteome Res.* **4**, 591–598 (2005).
88. Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **6**, 1060–1083 (2011).
89. García-Cañaveras, J. C., Donato, M. T., Castell, J. V. & Lahoz, A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. *J. Proteome Res.* **10**, 4825–4834 (2011).
90. Sana, T. R. *et al.* Global mass spectrometry based metabolomics profiling of erythrocytes infected with *Plasmodium falciparum*. *PloS one* **8**, e60840 (2013).
91. Urayama, S., Zou, W., Brooks, K. & Tolstikov, V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Commun. Mass Spectrom.* **24**, 613–620 (2010).
92. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).

93. Fairchild, J. N., Horvath, K., Gooding, J. R., Campagna, S. R. & Guiochon, G. Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. *J. Chromatogr. A* **1217**, 8161–8166 (2010).
94. Pyke, J. S. *et al.* A tandem liquid chromatography–mass spectrometry (LC–MS) method for profiling small molecules in complex samples. *Metabolomics* 1–11 (2015).doi:10.1007/s11306-015-0806-7
95. Chalcraft, K. R. & McCarry, B. E. Tandem LC columns for the simultaneous retention of polar and nonpolar molecules in comprehensive metabolomics analysis. *J. Sep. Sci.* **35**, 3478–3485 (2013).
96. Jessome, L. L. & Volmer, D. A. Ion suppression : a major concern in mass spectrometry. *LCGC North America* **24**, 498–510 (2006).
97. Wishart, D. S. *et al.* HMDB: the human metabolome database. *Nucleic Acids Res.* **35**, D521–D526 (2007).
98. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther Drug Monit.* **27**, 747–751 (2005).
99. Halket, J. M. *et al.* Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **56**, 219–243 (2005).
100. Jorgenson, J. W. & Lukacs, K. D. Capillary zone electrophoresis. *Science* **222**, 266–272 (1983).
101. Otsuka, K. & Terabe, S. Micellar electrokinetic chromatography. *Mol. Biotechnol.* **9**, 253–271 (1998).
102. Britz-mckibbin, P. & Chen, D. D. Y. Selective focusing of catecholamines and weakly acidic compounds by capillary electrophoresis using a dynamic pH junction. *Anal. Chem.* **72**, 1242–1252 (2000).
103. Kuehnbaum, N. L., Kormendi, A. & Britz-mckibbin, P. Multisegment injection-capillary electrophoresis-mass spectrometry: A high-throughput platform for metabolomics with high data fidelity. *Anal. Chem.* **85**, 10664–10669 (2013).
104. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill Jr, H. H. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **43**, 1–22 (2008).

105. Thalassinou, K. *et al.* Characterization of phosphorylated peptides using traveling wave-based and drift cell ion mobility mass spectrometry. *Anal. Chem.* **81**, 248–254 (2009).
106. Kliman, M., May, J. C. & McLean, J. A. Lipid analysis and lipidomics by structurally selective ion mobility-mass spectrometry. *Biochim. Biophys. Acta* **1811**, 935–945 (2011).
107. Woods, A. S. *et al.* Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. *Anal. Chem.* **76**, 2187–2195 (2004).
108. Dwivedi, P. *et al.* Gas-phase chiral separations by ion mobility spectrometry. *Anal. Chem.* **78**, 8200–8206 (2006).
109. Dass, C. *principles and practice of biological mass spectrometry*. (J. Wiley & Sons: New York, 2001).
110. Schug, K. & Taylor, T. The CHROMacademy essential guide: optimizing LC-MS and LC-MS/MS. *LCGC's CHROMacademy* (2013).
111. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
112. Benton, H. P., Wong, D. M., Trauger, S. A. & Siuzdak, G. XCMS2 : processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal. Chem.* **80**, 6382–6389 (2008).
113. Katajamaa, M., Miettinen, J. & Orešič, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–636 (2006).
114. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics* **11**, 395 (2010).
115. Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D. S. MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **40**, W127–W133 (2012).
116. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257 (2015).

117. Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. E. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).
118. Ferreira, J. a & Zwinderman, A. Approximate sample size calculations with microarray data: an illustration. *Stat. Appl. Genet. Mol. Biol.* **5**, Article25 (2006).
119. Vinaixa, M. *et al.* A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. *Metabolites* **2**, 775–795 (2012).
120. Broadhurst, D. I. & Kell, D. B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171–196 (2006).
121. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
122. Bylesjö, M. *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometrics* **20**, 341–351 (2006).
123. van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142–156 (2006).
124. Jones, O. A. H., Spurgeon, D. J., Svendsen, C. & Griffin, J. L. A metabolomics based approach to assessing the toxicity of the polycyclic aromatic hydrocarbon pyrene to the earthworm *Lumbricus rubellus*. *Chemosphere* **71**, 601–609 (2008).
125. Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. a. & Hendriks, M. M. W. B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**, 361–374 (2014).
126. Grove, H. *et al.* Combination of statistical approaches for analysis of 2-DE data gives complementary results. *J. Proteome Res.* **7**, 5119–5124 (2008).
127. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **42**, D459–D471 (2014).
128. Morrison, N. *et al.* Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics* **3**, 203–210 (2007).

129. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
130. Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241 (2007).
131. Creek, D. J. *et al.* Metabolite identification : are you sure ? And how do your peers gauge your confidence ? *Metabolomics* **10**, 350–353 (2014).
132. Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: the apogee of the omic trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).
133. Kusano, M., Fukushima, A., Redestig, H. & Saito, K. Metabolomic approaches toward understanding nitrogen metabolism in plants. *J. Exp. Bot.* **62**, 1439–1453 (2011).

Chapter 2: Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS

Fan Fei^{1,2}, *Dawn M. E. Bowdish*^{2*}, *Brian E. McCarry*¹

The work in this chapter has been published in: Fei, F., Bowdish, D. M. E. & McCarry, B. E. Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal. Bioanal. Chem.* **406**, 3723–3733 (2014). “This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited” (Springer Permission Request).

Author Contributions

BEM and FF conceived and designed the study. FF performed the experiment and wrote the manuscript. FF and DMEB edited the manuscript.

Author Affiliations:

¹ Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Canada L8S4M1

² Department of Pathology and Molecular Medicine, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada L8N3Z5

*Corresponding author: Email: bowdish@mcmaster.ca; Phone: (+1)905-525-9140 ext22313; Fax: (+1) 905-522-6750

Abstract

The comprehensive metabolomic analyses using eukaryotic and prokaryotic cells are an effective way to identify biomarkers or biochemical pathways which can then be used to characterize disease states, differences between cell lines, or inducers of cellular stress responses. One of the most commonly used extraction methods for comprehensive metabolomics is the Bligh and Dyer method (BD) which separates the metabolome into polar and nonpolar fractions. These fractions are then typically analyzed separately using hydrophilic interaction liquid chromatography (HILIC) and reversed-phase (RP) LC respectively. However, this method has low sample throughput and can also be biased to either polar or nonpolar metabolites. Here we introduce a MeOH/EtOH/H₂O extraction paired with HILIC-TOF-MS for comprehensive and simultaneous detection of both polar and nonpolar metabolites that is compatible for wide array of cellular species cultured in different growth medium. This method has been shown to be capable of separating polar metabolites by a HILIC mechanism and classes of lipids by an adsorption-like mechanism. Furthermore, this method is scalable and offers a substantial increase in sample throughput compared to BD with comparable extraction efficiency. This method was able to cover 92.2% of the detectable metabolome of Gram-negative bacterium *S. meliloti*, as compared to 91.6% of the metabolome by a combination of BD polar (59.4%) and BD nonpolar (53.9%) fractions. This single extraction-HILIC approach was successfully used to characterize the endometabolism of Gram-negative and Gram-positive bacteria as well as mammalian macrophages.

Introduction

Cellular metabolomics is an important part of systems biology as it reflects the phenotype of cells and monitors cellular activities in a perturbed system [1, 2]. The metabolomic profile of any organism represents a snapshot of its physiological state and reflects the overall contributions from genomic, transcriptomic, proteomic, and other environmental factors [3]. *In vitro* studies with cell cultures are convenient, fast, cost effective and more controllable compared to animal studies using sera, tissues or body fluids, yet are still able to provide insight into biological functions [4]. Therefore, cellular metabolomics is capable of providing an understanding of the global biochemical behaviour of a biological system.

Comprehensive cellular metabolomics facilitates observation-driven, hypothesis generation experiments by examining the entire detectable metabolome of a cellular system in order to discover new biochemical phenomena [5]. Comprehensive metabolomics involves the analyses of diverse chemical classes including but not limited to sugars, nucleotides, and organic acids with varying polarity, solubility and volatility. The endogenous cellular concentration of many metabolites can span over 12 orders of magnitude (from mM to fM) [6, 7]. Expanding the diverse detectable metabolome over a broad range of concentration, physical and chemical properties is challenging yet crucial since multiple pathways can often be influenced by a single external variable. In particular, central carbon metabolism and lipid metabolism are both involved in the inflammatory response of murine macrophages undergoing lipopolysaccharide stimulation [8–11]. The quality of the results often relies on the extent of metabolome coverage in order to unveil the biochemical changes entirely.

Liquid chromatography combined with mass spectrometry equipped with an electrospray ionization source (LC-ESI-MS) is the predominant analytical methodology used for comprehensive metabolomics [12–14]. The dynamic range of time-of-flight (TOF) MS is often greater than three orders of magnitude [15] and affords untargeted detection of relatively abundant metabolites with little sample pre-treatment. The detected metabolite features, including ion source fragments, adducts and isotopic ions, are defined by a unique combination of m/z and retention time value. The relative abundances of the collective list of metabolite features are used to evaluate and compare different treatments [16]. Many comprehensive metabolomics reports to date are focused on either the polar or lipid fractions of the metabolome. The few studies that reflect true comprehensive metabolome have analyzed the polar fraction using hydrophilic interaction liquid chromatography (HILIC) and the lipid nonpolar fraction using reversed-phase (RP) LC with C18 or C8 columns [17–20]. The polar and lipid fractions are conventionally extracted from a biological sample using the Bligh and Dyer (BD) method [21]. There is also single extraction-dual separation LC-MS for the analysis of a single extract containing both polar and nonpolar metabolites separately on HILIC and RPLC [16]. Both methods are able to ensure global coverage of the metabolome; however, these methodologies are time consuming in sample preparation and analysis.

Here, we present a reproducible, high throughput, single extraction-HILIC approach for comprehensive cellular metabolomic analysis. This approach is able to simultaneously analyse a broad range of polar and nonpolar metabolites, which is ideal for a large-scale hypothesis generation study with two-fold faster data acquisition and analysis time when compared to the conventional methods. The metabolome coverage of the scalable extraction was compared to both the polar and the nonpolar fractions of BD method. This approach has been successfully applied to three different cell types: the Gram-positive bacterium *Streptococcus intermedius*, the Gram-negative bacterium *Sinorhizobium meliloti*, and mammalian macrophages. The overall metabolome coverage observed with this single extraction-HILIC approach is equivalent to the BD method with separate analyses of polar and nonpolar fractions.

Experimental

Chemicals

HPLC grade methanol (MeOH), ethanol (EtOH), acetonitrile (ACN), chloroform (CHCl₃), and water (H₂O) were purchased from Caledon laboratories (Georgetown, ON, Canada). Ammonium acetate and formic acid were purchased from Fisher Scientific Company (Ottawa, ON, Canada). The 2.0 mm steel chrome ball bearings were purchased from Bearing & Oil Seals Specialists Inc. (Hamilton, ON, Canada). The isotopically labelled standards for recovery determination (RS) and for peak intensity normalization (IS) were purchased from Cambridge Isotope Laboratories (Andover, MA, USA). Lipid standards were purchased from Avanti® Polar Lipids, Inc. (Alabaster, AL, USA), and other chemical standards for LC-MS were purchased from Sigma Aldrich (St. Louis, MO, USA) and Biolog Inc. (Hayward, CA, USA). The full list of metabolite standards can be found in the Supporting Information.

Cell Culture and Collection

Detailed growth conditions for the Gram-negative bacterium, *Sinorhizobium meliloti*, the Gram-positive bacterium, *Streptococcus intermedius*, and murine macrophages are included in the Supporting Information S1. Detailed discussion regarding quenching methods for cellular metabolism and detachment protocol for adherent cells can be found in the Supplementary information and Fig. S2. The entire harvesting and extraction process was performed on ice or in a 4°C cold room.

Suspension bacterial cell culture

Cells were cultured in either 5 mL autoclaved test tubes or in sterile 96-well plates (Corning® Costar®, NY, USA). The suspension cell cultures were centrifuged at 9500 × g (13000 rpm) at 4°C in a Beckman Coulter Allegra X-22R centrifuge for 3 minutes and the supernatants were carefully aspirated with micropipette and discarded. The cell pellet was re-suspended in 1 mL cold saline solution (0.85% NaCl) or phosphate buffered saline (PBS). The mixture was centrifuged at 9500 × g for 3 minutes, and the wash solvent was aspirated and discarded. Extraction solvent of 100 µL (1:2:1 MeOH/CHCl₃/H₂O, 1:1

MeOH/EtOH or 2:2:1 MeOH/EtOH/H₂O) containing RS was added to the washed cell pellet which was then extracted immediately.

Adherent macrophage cell culture

For adherent macrophage cultures in 6-well tissue culture plates (Falcon[®], NY, USA), the growth medium was aspirated carefully with a micropipette. The cells were quickly washed with 1 mL cold saline or PBS. After removing the wash solvent via aspiration, 200 μ L of extract solvent (1:1 MeOH/H₂O for BD, 1:1 MeOH/EtOH or 2:2:1 MeOH/EtOH/H₂O) containing RS were added to each well. Cells were detached from the culture plate using a cell lifter in the presence of the extraction solvent, and the cell mixture was transferred into a 1.5 mL microtube (Diamed, Mississauga, ON, Canada). Only for Bligh and Dyer extraction, 200 μ L volume of CHCl₃ was also added to the microtube and the mixture was extracted immediately. The volume of the extraction solvent should be adjusted to the size of the well if other types of microtiter plates were used.

Intracellular Metabolite Extraction

The extraction protocol was optimized using *Sinorhizobium meliloti* (2×10^9 cells), and applied to *Streptococcus intermedius* (1×10^9 cells) as well as murine macrophages (3×10^5 cells). RS were prepared such that the final concentrations of L-methionine-d₃, L-tryptophan-d₅, and L-lysine-¹³C₆-¹⁵N₂ were 27 μ M, 24 μ M and 20 μ M, respectively, in the final reconstituted cell extracts. The final extracts were blown down to dryness under a gentle stream of nitrogen gas and reconstituted in 50 μ L of 60%v/v ACN/H₂O containing L-phenylalanine-d₈ (25 μ M), gly-phe (16 μ M), phe-phe (6 μ M) as well as cytidine-ribose-¹³C₅ (20 μ M) as IS. To minimize variability, sample extractions, addition of IS, and preparation of quality control (QC) samples were prepared in the same day using the same batch of solvents, IS and RS. The final endogenous cellular extracts were stored at -80°C freezer until HILIC-ESI-TOF-MS analyses.

Bligh and Dyer Extraction

The extraction process was based on the method of Bligh and Dyer [21] to achieve a biphasic separation using 1:2:1 MeOH/CHCl₃/H₂O as extraction solvent. The top MeOH/H₂O fraction contained polar metabolites and the bottom CHCl₃ fraction contained nonpolar metabolites (i.e. lipids). The denatured proteins and other macromolecules, such as DNA and RNA, were precipitated and suspended at the interface between polar and nonpolar fractions. The cell mixture was vigorously mixed by vortex for 2 minutes in the presence of two 2.0 mm ball bearings. After removal of the bearings, the cell mixture was centrifuged at 9500 $\times$ g for 3 minutes. The polar fraction was collected and the protein film and the lipid fraction were re-extracted twice with 50 μ L of cold MeOH/H₂O (1:1) with the same ball bearings. The combined polar fraction was back-extracted twice with 50 μ L of cold CHCl₃, and the non-polar CHCl₃ fractions were collected and combined with remaining CHCl₃ fraction from the previous polar

extraction. In total, 150 μL of polar and nonpolar fractions of bacterial extracts or 300 μL from adherent cell cultures were collected.

MeOH/EtOH/H₂O Extraction

This extraction protocol was based on a plasma extraction procedure adopted from Bruce [22]. An extraction solvent of 2:2:1 MeOH/EtOH/H₂O or 1:1 MeOH/EtOH was used to generate a single fraction which contained a mixture of polar and non-polar lipid metabolites. Proteins and other macromolecules were precipitated. Prior to extraction, 10 μL of RS were added to the cell mix. The cell mixture was mixed rigorously by vortex for 2 minutes in the presence of two 2.0 mm ball bearings. After removal of the bearings, the mixture was centrifuged at $9500 \times g$ for 3 minutes. The supernatant was collected and the precipitated pellet (containing DNA, RNA, and proteins) was re-extracted twice with 50 μL of the cold corresponding extraction solvent with beadbeating. A total of 150 μL cell extracts were collected for suspended cell cultures or 300 μL were collected for adherent cell culture.

LC-MS Analysis

The endogenous cellular extracts were analyzed using an Agilent Technologies 1200 RR Series II liquid chromatograph (LC) coupled to a Bruker MicrOTOF II Mass Spectrometer. An injection of 2 μL was separated on a 50 mm $\times$ 2.1 mm Kinetex 2.6 μm HILIC column of pore size of 100 Å (Phenomenex, CA, USA). The mobile phases were HPLC grade acetonitrile (A) and 10 mM ammonium acetate in HPLC grade water adjusted to pH 3 with formic acid (B) at a flow rate of 200 $\mu\text{L}/\text{min}$. The column temperature was maintained at 40 °C, and the auto sampler storage tray was set at 4°C. The mobile phase gradient eluted isocratically with 95% ACN for 0.5 min followed by a gradient to 35% ACN over 12 min. The gradient was maintained at 35% ACN for 0.5 min and increased to 95% ACN over 1 min. The gradient was then followed by a 10 min re-equilibration prior to the next injection. The total time for the HILIC gradient was 24 min. Positive ionization mode (ESI+) and negative ionization mode (ESI-) were performed in separate runs. Details of the optimization of the HILIC gradient method using 2³ full factorial design can be found in Supplementary information and Fig. S3.

The parameters chosen for ESI conditions were as follows: 4.0 bar nebulizer pressure; -500 V endplate offset; -3800 V or 4500 V capillary voltage; 8.0 L/min drying gas flow rate; 250°C dry gas temperature. The data were acquired in profile mode from 50 to 1000 m/z at a scan rate of 1.0 Hz (computed using a rolling average value of 2). The mass accuracy was adjusted by internal calibration using endogenous sodium formate clusters in both ESI+ and ESI- with Bruker's data analysis software.

Each of the extraction methods was repeated in sextuplicate and all of the samples were analyzed in random order. A QC pooled sample was prepared by combining equal volumes of all samples, and divided into individual aliquots after thorough mixing. A fresh QC sample and a day's worth of samples (ca. 20 samples) were thawed and run each day for a multiday experiment. The pooled sample was injected five times at the

beginning of the analysis to condition the column, and also injected after every five samples. Methanol blank and a standard mixture containing all recovery and internal standards were also run after every ten samples. Post-column addition with gly-phe was performed on all different matrices for different cell types in both ESI+ and ESI- modes for ion suppression studies.

Data Analysis

Raw data obtained from LC-ESI-mircOTOF-MS were converted to the .mzXML file format using BrukerCompassxport (<http://www.bdal.com/navi/meta/home.html>) after internal calibration. The .mzXML files were then processed with XCMS [23, 24] and CAMERA [25] in R Project (version 2.12.2). A tabulated metabolite feature list with aligned retention time and m/z values was exported in .csv format. For XCMS, the centWave algorithm [24] was used for peak picking with a resolution of 30 ppm, and a signal-to-noise threshold (snthr) set to 10. Features that appeared in less than 80% of the samples which underwent the same extraction method were removed (minfrac = 0.8). The isotopic ions, in source fragments, and adducts were identified using CAMERA.

Features with apparent retention factors k_{app} lower than 0.7 were removed because these were not retained and experienced great ion suppression. The isotopic ions, ions associated with IS, RS and sodium formate clusters were removed. Metabolite features detected in the biological samples were compared to those in the IS and RS solution in 60%v/v ACN/H₂O, and the duplicated ions that were associated with the background noise were removed. The peak areas of all metabolite features were normalized using the peak area of IS according to their retention time (Fig. S1). In source fragments and adducts were treated as separate metabolite features. Metabolite features with peak areas under 2000 were excluded. Features with greater than 30% variance in QC samples were removed. Integration of non-Gaussian or coeluting peaks using XCMS may generate inconsistent results. IS and RS with greater than 10% variance, and significantly differentiated metabolite features with greater than 20% variance were re-evaluated with manual integration using Bruker DataAnalysis 4.0.

The processed data sets were used as an input for SIMCA-P+ 11 software (Umetrics, Kinnelon, NJ). Pareto scaling was applied prior to principal component analysis (PCA) and orthogonal partial least-squares discriminative analysis (OPLS-DA). OPLS-DA was used to differentiate metabolite profiles between different extraction methods. The model validation parameter Q^2 (the fraction of variations of X and Y matrix explained by the model, X matrix was metabolite features, Y matrix was the treatment groups) values above 0.4 were indicative of a robust model, i.e. true differences between the comparing groups, and Q^2 between 0.7-1.0 indicated the model was highly robust [26]. R^2X (R^2Y) indicated the fraction in which X (Y) matrix was explained by the model. Two-tailed, unpaired heteroscedastic Student's t tests with $p < 0.05$ were computed in Microsoft Excel 2010 and used to identify metabolite features that were significantly differentially expressed in each extraction method. Metabolite features were identified by

matching the m/z and retention value to the available authentic standards. Figures were created in Adobe Illustrator CS5.

Results and Discussion

Comparison of Extraction Solvents

There have been extensive reviews that compare various different extraction strategies for cellular metabolic analyses [27–29]. Here, we have focused on comparing the extraction efficiency of both polar and non-polar metabolites with bi-phasic solvent MeOH/H₂O/CHCl₃ (1:1:2) in the BD method [21] to two other extraction solvents, MeOH/EtOH (1:1) and MeOH/EtOH/H₂O (2:2:1). BD extraction is commonly used for comprehensive metabolomics to ensure full coverage of both polar and nonpolar metabolites by running both polar and nonpolar phases separately on HILIC and RPLC. Similar to the BD method, MeOH/EtOH and MeOH/EtOH/H₂O were able to extract both polar and nonpolar metabolites but with all metabolites present in one single phase. Therefore, the latter two extraction strategies, when compared to BD, were able to minimize sample handling and also shorten the analysis time by a factor of two while still maintaining comprehensive coverage of both polar and nonpolar metabolites.

The scalable extraction method was tested on a 100 μ L *Sinorhizobium meliloti* (2×10^9) cell culture grown in M9 growth medium in a 96 microtiter plate. A beadbeating technique was adopted instead of vortex mixing to ensure full cell disruption. Sonication, though allowing complete cell disruption, was not used in order to avoid overheating which may degrade thermally labile metabolites. All three extraction solvents were able to precipitate protein, DNA and RNA as well as providing good recovery for a wide range of both hydrophilic and hydrophobic metabolites.

Based on the methionine- d_3 , the recoveries for the polar fraction of BD (BD polar), MeOH/EtOH, and MeOH/EtOH/H₂O after three extraction procedures were $77\% \pm 2\%$, $59\% \pm 5\%$ and $79\% \pm 2\%$, respectively, for a sextuplicate experiment. A much lower recovery was obtained for MeOH/EtOH when compared to the two other extraction methods. MeOH/EtOH failed to maintain a compact protein pellet during the extraction process; therefore, an extra centrifugation was required for MeOH/EtOH extraction to remove particulates in the sample extract. Moreover, based on the selected 24 endogenous metabolites in *S. meliloti*, the extra centrifugation step in MeOH/EtOH could also have caused lower extraction efficiencies when compared to MeOH/EtOH/H₂O solvent (Fig. S4). Minimal sample handling and short-time extraction procedures are critical for large-scale metabolomic studies in order to achieve greater reproducibility, sensitivity and to prevent metabolite modification and degradation with time [14, 30]. In terms of ease of performance, the MeOH/EtOH/H₂O extraction procedure outperformed MeOH/EtOH and the two-staged extraction procedure of BD for comprehensive polar and nonpolar metabolite analyses. Hence, the MeOH/EtOH/H₂O was further optimized to achieve better extraction efficiencies and robustness.

The optimal numbers of extraction processes required in order to reach a minimal of 95% extraction efficiency of the selected endogenous metabolites was determined by extracting a 100 μ L *S. meliloti* cell culture using MeOH/EtOH/H₂O seven times. Each extraction was performed in sextuplicate. The percentages of recoveries of 20 endogenous metabolites from *S. meliloti* at each extraction step were calculated by dividing the relative abundance of each individual metabolite at each step with its sum in all seven extractions (Fig. 1). Most metabolites showed greater than 80% recovery upon the first extraction. Among them, N-acetyl-aspartic acid, adenosine monophosphate (AMP), methylhistidine and acetylcarnitine were entirely (100%) recovered in the first extraction. The second extraction was able to recover the remaining 10-15% for most of the metabolites. Metabolites such as γ -aminobutyric acid, adenine, adenosine, and proline were persistent and were still detected after the seventh extraction. Therefore, a minimum of two MeOH/EtOH/H₂O extraction steps were required to ensure at least 95% extraction efficiency for the major endogenous metabolites. We recommend extracting cells three times to ensure great extraction efficiency and reproducibility.

The dissolution solvent has significant impact on peak shapes in HILIC chromatography [31]. The samples were concentrated in order to improve the detection limit. Sample volume was reduced to 50 μ L by drying with gentle stream of nitrogen, and the remaining solvent was primarily composed of water as it was the least volatile solvent in the MeOH/EtOH/H₂O extracts. Water is not an appropriate dissolution solvent for HILIC gradients with a high percentage of ACN because it causes peak broadening and, consequently, reduces sensitivity [31]. Therefore, samples were dried completely to remove all residual water and reconstituted in a solvent mix low in water to also minimize irreproducibility due to inconsistent sample volumes. We have adopted the use of 60%v/v ACN/H₂O to ensure adequate peak shape and sensitivity while still allowing full dissolution of the highly polar metabolites.

HILIC/MS for Simultaneous Detection of Both Polar and Lipid Metabolites

HILIC is typically used to separate polar compounds via hydrophilic partitioning mechanism. In 2010, HILIC was reported to be able to retain lipids, especially phospholipids, according to the polarity of the lipid heads [32, 33]. Therefore, since HILIC can simultaneously separate polar and lipid metabolites, it was selected as the chromatographic method for high throughput comprehensive metabolomic analyses. RPLC is often used to retain and separate nonpolar analytes [34]. Separation of polar compounds can also be achieved with RPLC with an ion-pairing agent in the mobile phase [35]; however, the ion-pairing reagents often lead to contamination in the MS instrument [36], and therefore was not preferred.

Unlike RPLC, small changes in pH and buffer ionic strength can often cause large retention deviations in HILIC [37]. To improve reproducibility and minimize retention deviation for better retention time alignment using XCMS, consistent preparation of the

mobile phase was critical. For large-scale comprehensive metabolomic analyses, all samples should be run using the same batch of mobile phases. The IS and RS spiked in each biological samples should also be used to correct retention time drift of metabolites when assigning metabolite identification based on the retention time of authentic standards. HILIC separations are less tolerant of fast gradients and require a longer equilibrium time compared to RPLC. Though the starting gradient at 98% ACN was able to retain a greater amount of metabolite features, it required much longer equilibration time (more than 10 min) than starting at 95% ACN (8-10 min). Running blanks and pooled samples at the beginning of the HILIC sequence is critical in order to condition the column to minimize variation in peak shape, retention time and ionization response.

HILIC is able to retain phospholipids or other polar lipids via an adsorption mechanism [32]. Silica-HILIC was chosen specifically for optimized phospholipid separation at low buffer strength instead of other commonly used zwitterionic or diol HILIC [32, 38]. The low buffer ionic strength of 10 mM ammonium acetate in mobile phase B allowed secondary interactions between the HILIC stationary phase and the polar lipid head groups via hydrogen bonding and electrostatic interaction. Therefore, our optimized HILIC gradient was able to separate phospholipids by classes based on their polar head group. Though each lipid class eluted within a very narrow time window (often within 1 min), there were still separations within each lipid class based on hydrophobicity (carbon chain length) and unsaturation (number of C=C bonds) (Fig. 2). The lipid class separation achieved with HILIC in combination with the sub 5 ppm mass accuracy attained with internal calibration was able to accurately identify lipid metabolites without running copious authentic standards. Isomers between phosphatidylcholine (PC) and phosphatidylethanolamine (PE) could be accurately identified because PCs and PEs were chromatographically separated. Fig. 3 summarizes 2125 metabolite features (after data reduction) detected in the intracellular extracts of murine macrophage. Different classes of phospholipids including phosphatidylglycerol, PCs, PEs and lyso-PCs were detected along with small polar metabolites such as nucleosides, amino acids, and organic acids. The ability of HILIC to separate both polar and lipid compounds combined with our extraction methodology allowed simultaneous analyses of both polar and lipid metabolites for enhanced sample throughput.

Mass accuracy can be significantly improved by the usage of sodium formate as an internal calibrant which was formed by the endogenous sodium ions in the cell and the formic acid in mobile phase B. The presence of sodium formate adducts with retention times at 7.2 min was used for internal mass calibration in both ESI+ and ESI- modes which dramatically improved mass accuracies for all three cell types grown in different biological media (Table S1). The confidence of metabolite identification was improved significantly with the sub-5 ppm mass accuracy attained after internal calibration with sodium formate.

The endogenous sodium formate also caused minor ion suppression regardless of the biological matrix of interest (Fig. S5). The IS, phenylalanine- d_8 , eluted in the ion

suppression region and was used to normalize peak areas of metabolite features eluting in the region to correct for varying degree of ion suppression in different samples.

The current method was applied to the comprehensive metabolomic analyses of *S. meliloti*, *S. intermedius* and murine macrophages. Over a continuous seven-day injection series of *S. meliloti* extracts, the retention time deviation was less than 7 seconds in a 24-min LC run with approximately 260,000 metabolite features in 137 samples (1900 features per sample over 137 samples in ESI+) analyzed by XCMS using the centWave method (Fig. S6). The peak area deviations of IS were all below 10%. The QC and sample data were analyzed with PCA with QC samples clustered tightly in the center of the score plot indicating that instrumental variability was minimal. The optimized HILIC-TOF-MS method was highly robust and reproducible.

Metabolome Coverage from the MeOH/EtOH/H₂O Extraction Compared to the two fractions of Bligh and Dyer Method

Untargeted comprehensive large-scale metabolomics demands that the experimental method have high sample throughput, high robustness to sustain long LC sequence, and excellent metabolome coverage. We propose using a MeOH/EtOH/H₂O extraction in combination with HILIC-TOF-MS to encompass both polar and nonpolar metabolites in a single analysis. Compared to the conventionally used BD method in which polar and nonpolar fractions are analyzed separately, the proposed method doubles the throughput and minimizes the sample handling time with comparable reproducibility. The metabolome coverage of the proposed method was compared to both of the polar and nonpolar fractions obtained using BD methods.

Traditionally, the BD polar fraction was run using HILIC, and the BD nonpolar fraction was run using RPLC [18, 19, 39]. However, in order to directly comparing the extraction efficiency of nonpolar metabolites, the BD nonpolar fraction was also run using the same optimized HILIC method as used for the MeOH/EtOH/H₂O extracted samples and BD polar extracts. Evaluating all three extract samples using the same LC method has also allowed us to compare metabolite features that were found in common between all three extracts. However, more features were expected when analyzing the BD nonpolar fraction with RPLC in comparison to HILIC. Triacylglycerols, diacylglycerols and fatty acids, which were commonly analyzed with RPLC, cannot be retained using HILIC, and were eluting in the unquantifiable dead volume with retention time below k_{app} 0.7. Gram-negative bacteria *S. meliloti* was used to compare the metabolome coverage and extraction efficiency of MeOH/EtOH/H₂O to BD polar and BD nonpolar extracts.

The XCMS centWave algorithm in combination with CAMERA has deconvoluted a total of 3378 metabolite features. All those features were present in at least one of the MeOH/EtOH/H₂O, BD polar and BD nonpolar extracts. There were more features detected in the ESI- mode (1900) compared to ESI+ mode (1478). Metabolite features from solvent contamination, instrumentation noise and spiked IS and RS that were shared in the extracted samples and the standard mixture containing IS and RS were removed

(unpaired heteroscedastic t test, $p > 0.05$ between all extracts and standard mixtures). Any features that were eluted in the dead volume with $k'_{app} < 0.7$ were removed because they were unquantifiable due to severe ion suppression. The isotopic ions annotated by CAMERA were also removed along with ions associated with sodium formate clusters. After data reduction, a final list of 1059 metabolite features was attained. The data reduction process was important to reduce the quantity of redundant data and false positives during statistical analyses.

Multivariate analysis using OPLS-DA revealed that all three types of extracts had unique metabolome profiles (Fig. 4a). The model was robust with $Q^2(\text{cum})=0.936$ and describes nearly all variables with $R^2X(\text{cum})=0.926$ and $R^2Y(\text{cum})=0.986$. All extracts from the MeOH/EtOH/H₂O extraction were clustered in between the BD polar and BD nonpolar extracts, indicating shared metabolome profiles between MeOH/EtOH/H₂O extracts and BD polar extracts as well as nonpolar extracts. Since MeOH/EtOH/H₂O extracts were not centered in the OPLS-DA score plot, these extracts contained some unique metabolite features that were absent in the BD polar and BD nonpolar extracts.

Among 1059 detectable endogenous metabolite features of *S. meliloti*, 59.4%, 53.9% and 92.2% were detected in BD polar, BD nonpolar and MeOH/EtOH/H₂O extracts, respectively (Fig. 4b). Of the metabolites we are able to detect using this method, 7.8% were not detected by the MeOH/EtOH/H₂O method, 2.9% were only detected in BD polar fraction, 4.5% were only detected in BD nonpolar fraction, and 0.3% were detected in both BD polar and nonpolar fractions. There were 34.7% of the features shared between MeOH/EtOH/H₂O and BD polar that were undetected in BD nonpolar, among them polar metabolites such as amino acids, organic acids, sugar phosphates, nucleotides and nucleosides were detected. There were 27.6% of the features shared between MeOH/EtOH/H₂O and BD nonpolar that were not detected in BD polar, lipids such as phosphatidylglycerols (PGs), PEs, PCs and phosphatidylserines (PS) were among those that were identified. There were 8.4% of the metabolite features that could only be detected in MeOH/EtOH/H₂O extracts, and they ranged in polarity and m/z values. A few of the identified metabolites and their relative responses are shown in Fig. 4c, and the ionization responses of all metabolites in MeOH/EtOH/H₂O were normalized to one as a reference. The MeOH/EtOH/H₂O method had equivalent recoveries for vast majority of those endogenous polar and lipid metabolites as compared to the separately analyzed BD polar or BD nonpolar extracts.

There were also 21.5% of the features shared between MeOH/EtOH/H₂O, BD polar and BD nonpolar extracts. Among those shared features, 59.6% of the shared metabolite features were equally extracted using MeOH/EtOH/H₂O in comparison to the most pronounced BD fractions based on unpaired heteroscedastic Student's t test of $p < 0.05$ (Table 1). In Fig. 5, the ionization responses of some selected shared metabolite features were normalized against IS and the sum of ionization response of BD polar and BD nonpolar which was normalized to one. The normalization was done under the assumption that BD polar and BD nonpolar in combination have a net 100% recovery of all metabolites. MeOH/EtOH/H₂O had lower extraction efficiency in 14.6% of the shared

features than at least one of the BD fractions (Fig. 5a). However, 15.8% of the shared features had a greater recovery in MeOH/EtOH/H₂O than both of the BD fractions (Fig. 5b). Many of those metabolite features were partially recovered in both BD polar and BD nonpolar, but were more efficiently recovered using MeOH/EtOH/H₂O. If a conventional comprehensive metabolomic method was used with each of the BD polar and nonpolar fractions run on either HILIC or RP, then features that were partially extracted and present in both fractions would be considered as different metabolites, and would be quantified separately and result in bias during multivariate analyses. Moreover, recovering these metabolites in full using MeOH/EtOH/H₂O, results in higher injected concentrations, and facilitates their detection. Therefore, the 8.4% of the metabolite features detected exclusively in the MeOH/EtOH/H₂O method would likely be low abundant metabolites in the cells which were under detection limit when partially recovered in either of the BD fractions, but detectable when more efficiently recovered in the MeOH/EtOH/H₂O extraction.

Based on all these results, MeOH/EtOH/H₂O in combination with HILIC-TOF-MS provides a very robust, high throughput and comprehensive approach for cellular metabolomic analyses. The metabolomic coverage of MeOH/EtOH/H₂O was comparable to the combined coverage of BD polar and BD nonpolar yet was twice as efficient in terms of data acquisition speed. The method was not biased towards neither of the polar or nonpolar metabolites.

Conclusion

The complex biological matrices and different culturing techniques required for the growth of many cellular organisms present a great challenge for comprehensive analyses. Large-scale comprehensive metabolomic analyses involving hundreds of samples often lead to time consuming, labour intensive sample preparation, extraction, and data acquisition. The proposed comprehensive metabolomic protocol using MeOH/EtOH/H₂O extraction paired with hydrophilic interaction liquid chromatography (HILIC)-TOF-MS analysis was able to expand metabolome coverage to both polar and lipid metabolites with high reproducibility and robustness with two-fold faster data acquisition throughput than the conventional Bligh and Dyer method coupled to reversed-phase (RP) and HILIC-MS. This scalable extraction method was applicable to Gram-positive and -negative bacteria with rigid cell walls as well as mammalian cells; it was applicable to both suspension and adherent cell cultures that were grown in either rich or minimal media and had minimal ion suppression despite the complex biological matrix. This comprehensive metabolomic method was developed as a qualitative initial screening tool for finding biomarker or pathway differences between treatments for a hypothesis generating study. However, extension to quantitative targeted metabolomic analyses should be followed for high impact study.

Acknowledgement

This work was dedicated to late Prof. Brian McCarry (1946-2013). Funding for this work was provided by NSERC and CIHR. Work in the Bowdish laboratory is supported in part by the Institute of Infectious Disease Research and the McMaster Immunology Research Centre. The authors would like to thank Dr. Mark McDermott, Dr. Kenneth Chalcraft, Dr. M. Kirk Green, and Roger Luckham for critical reading of the manuscript. The authors would like to thank the Center for Microbial Chemical Biology at McMaster for access to the LC-MS. *Streptococcus intermedius* and murine macrophages used in this paper were kindly cultured and provided by Michelle Pinto and Keith Lee.

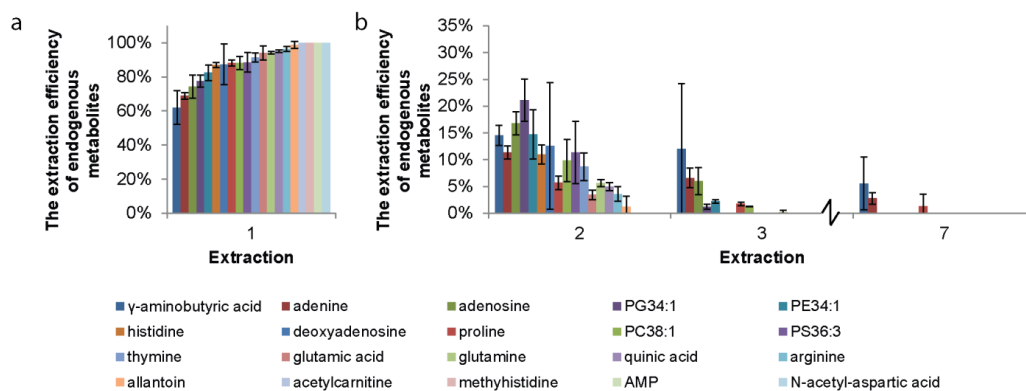


Figure 1 The extraction efficiencies of endogenous metabolites in 2×10^9 *S. meliloti* cells in the first extraction (a), and 2nd to the 7th extraction (b) (only 2nd, 3rd, and 7th extractions are shown). The extraction efficiencies at each extraction step

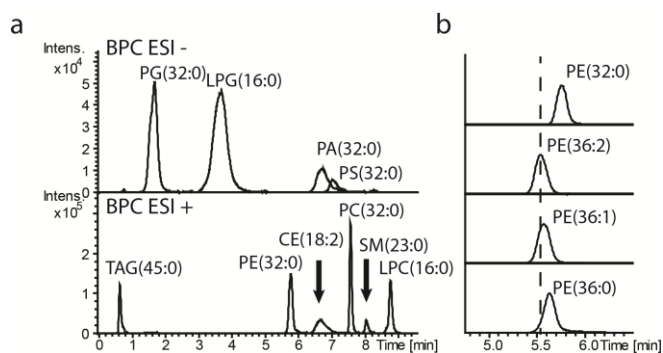


Figure 2 (a) Base peak chromatogram of the HILIC separation on ten lipid classes including PGs, lyso-PG, phosphatidic acids (PAs), PSs, triacylglycerols (TAG), PEs, ceramides (CEs), PCs, sphingomyelins (SMs) and lyso-PCs. (b) Base peak chromatogram of $[M+H]^+$ ions of PEs

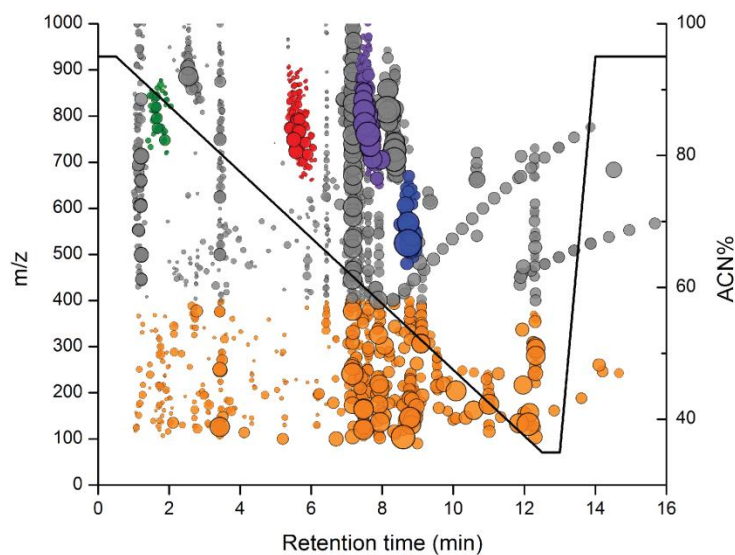


Figure 3 The 2125 endogenous metabolite features detected in 3×10^5 murine macrophage extracts after data reduction. The radius of the data markers (filled circle) reflected the relative abundances of the metabolite features over a dynamic range of 5×10^4 . Small metabolites (90-400 m/z) in yellow, PGs in green, PEs in red, PCs in purple and lyso-PCs in blue were highlighted. The HILIC gradient was labelled with a black line with reference to percentage acetonitrile

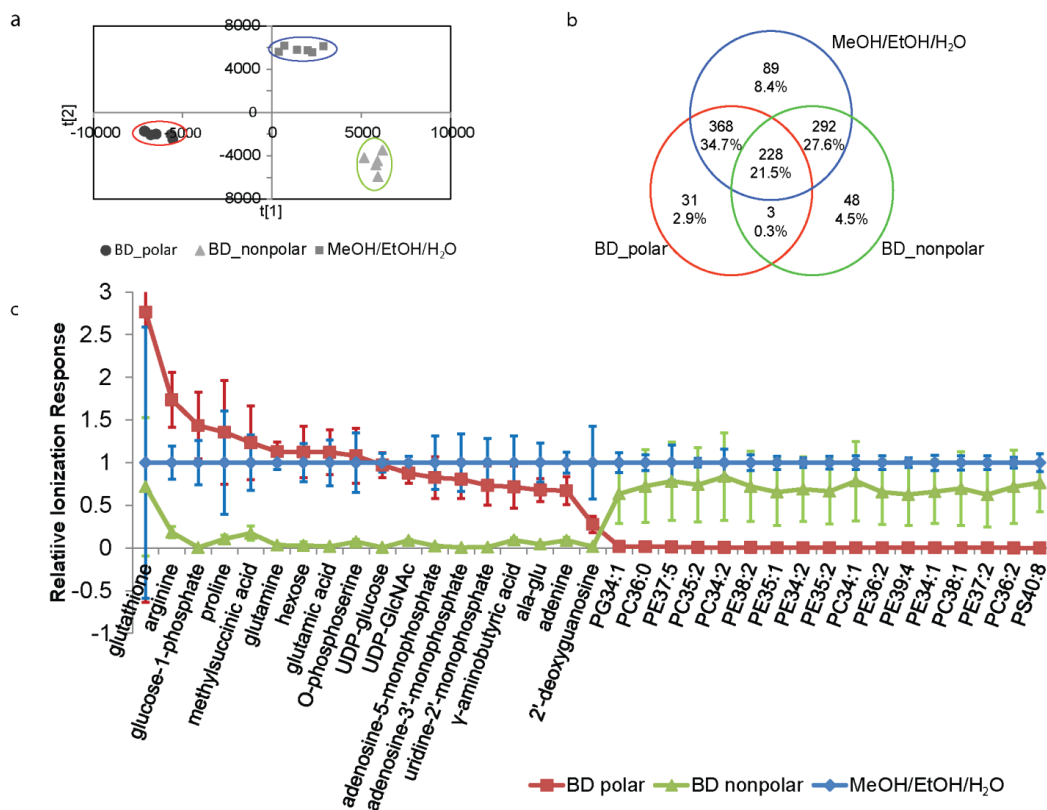


Figure 4 The metabolome coverage of *S. meliloti* of 1059 endogenous metabolite features found in at least one set of BD polar, BD nonpolar or MeOH/EtOH/H₂O extracted samples. Extracts were performed in sextuplicate and analyzed independently by HILIC-TOF-MS in both ESI+ and ESI- modes. The ionization responses of the metabolite features were normalized using internal standards. (a) OPLS-DA score plot comparing the endogenous metabolome coverage of *S. meliloti* extracts attained from three different extraction methods with $R^2X(\text{cum})=0.926$, $R^2Y(\text{cum})=0.986$, and $Q^2(\text{cum})=0.936$. (b) The quantity of metabolite features that were uniquely identified and shared between BD polar, BD nonpolar and MeOH/EtOH/H₂O extracted samples were listed in the Venn diagram with their estimated percentage share of the total detectable metabolome. (c) The normalized ionization responses of identified metabolites with varying polarity. The ionization responses from MeOH/EtOH/H₂O extracts was set to one as references. Polar metabolites were mostly extracted in BD polar, and lipids were seen exclusively in BD nonpolar; however, all metabolites were detected in MeOH/EtOH/H₂O. Error bars corresponded to two standard deviations. UDP, uridine diphosphate; GlcNac, N-acetylglucosamine

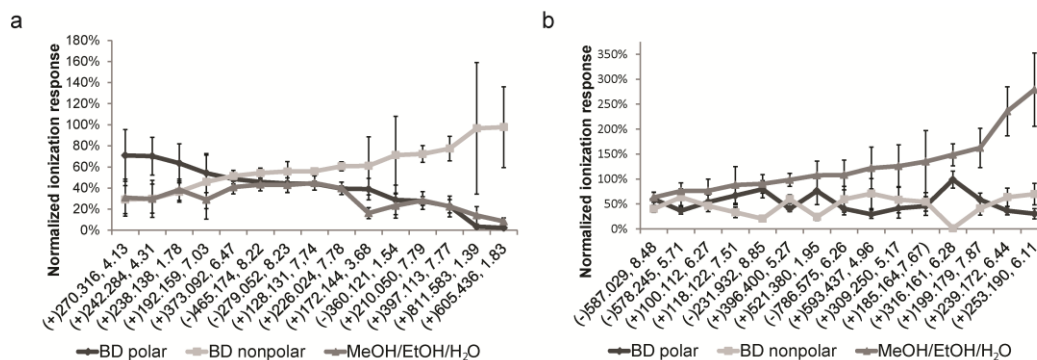


Figure 5 Among 228 shared metabolite features in BD polar, BD nonpolar and MeOH/EtOH/H₂O extracts, (a) 14.6% features had lower extraction efficiencies in MeOH/EtOH/H₂O extracts than at least one of the BD extracts, and (b) 15.8% features had increased extraction efficiencies in MeOH/EtOH/H₂O extracts than all BD extracts. Each error bar corresponds to two standard deviations calculated from six independent extractions. The ionization responses were normalized to the IS and also the total summed ionization response for both BD polar and BD nonpolar. Each metabolite feature is represented by the ESI mode used for its detection (in brackets), its m/z value and retention time

Table 1 The shared 228 metabolite features in BD polar, BD nonpolar and MeOH/EtOH/H₂O extracts were compared between different extraction methods using unpaired heteroscedastic Student's t test with $p < 0.05$. Among the shared features, 15.8% showed greater extraction efficiency ($\uparrow$), 14.6% showed lower extraction efficiency ($\downarrow$) and 59.6% showed no difference ($\approx$) between MeOH/EtOH/H₂O and BD fractions. " $>$ ", greater extraction efficiency; " $<$ ", less extraction efficiency; " $=$ ", equal extraction efficiency.

		Features detected	Percentage features of 228 shared	
Extraction efficiency of MeOH/EtOH/H ₂ O compared to BD	$\uparrow$	36	15.8%	MeOH/EtOH/H ₂ O $>$ BD polar, BD nonpolar
		5	2.2%	BD polar $>$ MeOH/EtOH/H ₂ O $>$ BD nonpolar
		6	2.6%	BD polar $<$ MeOH/EtOH/H ₂ O $<$ BD nonpolar
	$\downarrow$	2	0.9%	MeOH/EtOH/H ₂ O $<$ BD polar, BD nonpolar
		7	3.1%	MeOH/EtOH/H ₂ O = BD nonpolar $<$ BD polar
		36	15.8%	MeOH/EtOH/H ₂ O = BD polar $<$ BD nonpolar
		29	12.7%	MeOH/EtOH/H ₂ O = BD polar = BD nonpolar
	$\approx$	39	17.1%	MeOH/EtOH/H ₂ O = BD nonpolar $>$ BD polar
		68	29.8%	MeOH/EtOH/H ₂ O = BD polar $>$ BD nonpolar

^a Metabolite features seen with greater ionization responses in either BD polar and BD nonpolar fractions in comparison to those from MeOH/EtOH/H₂O extracts were considered to have lower extraction efficiencies in MeOH/EtOH/H₂O, and vice versa.

References

1. Oliver SG, Winson MK, Kell DB, Baganz F (1998) *Trends Biotechnol* 16:373–378.
2. Fiehn O (2002) *Plant Mol Biol* 48:155–171.
3. Patti GJ, Yanes O, Siuzdak G (2012) *Nat Rev Mol Cell Biol* 13:263–269.
4. Čuperlović-Culf M, Barnett DA, Culf AS, Chute I (2010) *Drug Discov Today* 15:610–621.
5. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) *Trends Biotechnol* 22:245–252.
6. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) *Nat Chem Biol* 5:593–599.
7. Lei Z, Huhman D V., Sumner LW (2011) *J Biol Chem* 286:25435–25442.
8. Tannahill GM, Curtis AM, Adamik J, Palsson-McDermott EM, McGettrick F, Goel G, Frezza C, Bernard NJ, Kelly B, Foley NH, Zheng L, Gardet A, Tong Z, Jany SS, Corr SC, Haneklaus M, Caffrey BE, Pierce K, Walmsley S, Beasley FC, Cummins E, Nizet V, Whyte M, Taylor CT, Lin H, Masters SL, Gottlieb E, Kelly VP, Clish C, Auron PE, Xavier RJ, O'Neill LAJ (2013) *Nature* 496:238–242.
9. Kita Y, Takahashi T, Uozumi N, Nallan L, Gelb MH, Shimizu T (2005) *Biochem Bioph Res Co* 330:898–906.
10. Verhoeckx KCM, Bijlsma S, Jespersen S, Ramaker R, Verheij ER, Witkamp RF, van der Greef J, Rodenburg RJT (2004) *Int Immunopharmacol* 4:1499–1514.
11. Schilling JD, Machkovech HM, He L, Sidhu R, Fujiwara H, Weber K, Ory DS, Schaffer JE (2013) *J Biol Chem* 288:2923–2932.
12. Theodoridis G, Gika HG, Wilson ID (2008) *Trends Anal Chem* 27:251–260.
13. Gummer J, Banazis M, Maker G, Solomon P, Oliver R, Trengove R (2009) *Australian Biochemist* 40:5–16.
14. Vuckovic D (2012) *Anal Bioanal Chem* 403:1523–1548.
15. Andrews GL, Simons BL, Young JB, Hawkridge AM, Muddiman DC (2011) *Anal Chem* 83:5442–5446.

16. Ivanisevic J, Zhu Z-J, Plate L, Tautenhahn R, Chen S, O'Brien PJ, Johnson CH, Marletta MA, Patti GJ, Siuzdak G (2013) *Anal Chem* 85:6876–6884.
17. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R, The Human Serum Metabolome (HUSERMET) Consortium (2011) *Nat Protoc* 6:1060–1083.
18. García-ca JC, Donato MT, Castell J V., Lahoz A (2011) *J Proteome Res* 10:4825–4834.
19. Sana TR, Gordon DB, Fischer SM, Tichy SE, Kitagawa N, Lai C, Gosnell WL, Chang SP (2013) *PloS one* 8:e60840.
20. Urayama S, Zou W, Brooks K, Tolstikov V (2010) *Rapid Commun Mass Sp* 24:613–620.
21. Bligh EG, Dyer WJ (1959) A Rapid Method of Total Lipid Extraction and Purification. *Can J Biochem Phys* 37:911–917.
22. Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S, Guy PA (2009) *Anal Chem* 81:3285–3296.
23. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) *Anal Chem* 78:779–787.
24. Smith CA (2008) LC / MS Preprocessing and Analysis with XCMS. 1–13.
25. Kuhl C, Tautenhahn R, Neumann S (2010) LC-MS Peak Annotation and Identification with CAMERA. October 1–14.
26. Jones OAH, Spurgeon DJ, Svendsen C, Griffin JL (2008) *Chemosphere* 71:601–609.
27. Villas-Bôas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J (2005) *Mass Spectrom Rev* 24:613–646.
28. Maharjan RP, Ferenci T (2006) *Anal Biochem* 313:145–154.
29. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, Goodacre R (2008) *Anal Chem* 80:2939–2948.
30. Tulipani S, Llorach R, Urpi-Sarda M, Andres-Lacueva C (2013) *Anal Chem* 85:341–348.

31. Ruta J, Rudaz S, McCalley D V., Veuthey J-L, Guillarme D (2010) *J Chromatogr A* 1217:8230–8240.
32. Zheng L, T'Kind R, Decuypere S, von Freyend SJ, Coombs GH, Watson DG (2010) *Rapid Commun Mass Sp* 24:2074–2082.
33. Schwalbe-Herrmann M, Willmann J, Leibfritz D (2010) *J.chromotogr A* 1217:5179–5183.
34. Chalcraft KR, McCarry BE (2013) *J Sep Sci* 35:3478–3485.
35. Lu W, Bennett BD, Rabinowitz JD (2008) *J Chromatogr B* 871:236–242.
36. Guan F, Ishii A, Seno H, Watanabe-Suzuki K, Kumazawa T, Suzuki O (1999) *J Chromatogr B* 731:155–165.
37. Guo Y, Gaiki S (2005) *J Chromatogr A* 1074:71–80.
38. Zhu C, Dane A, Spijksma G, Wang M, van der Greef J, Luo G, Hankemeier T, Vreeken RJ (2012) *J Chromatogr A* 1220:26–34.
39. Bayle M-L, Wopereis S, Bouwman J, van Ommen B, Scalbert A, Pujos-Guillot E (2012) *Metabolomics* 8:1114–1129.

Supplementary Information

Cell culture conditions

Cell Culture for *Sinorhizobium meliloti*

S. meliloti RmP100 (wild type strain) were streaked onto LB agar plates and incubated at 30°C for four days. A single colony from the agar plate (less than 2 weeks old) was inoculated into 5 mL LB_{mc} rich medium at 30°C for 24 hours. The culture was centrifuged, then washed with 1 mL of sterile saline solution (0.85% NaCl), and re-suspended in M9 culture. The re-suspended *S. meliloti* culture was inoculated to M9 at initial OD₆₀₀ of 0.05 and incubated at 30°C for 24 hours till reaching mid-log phase (OD₆₀₀ = 1.0).

The LB broth (10 g/L DifcoBactoTryptone, 5 g/L Difco Yeast Extract, and 5 g/L NaCl) was adjusted with 1M NaOH to pH 7.0, and autoclaved. The LB_{mc} broth was prepared by adding 250 µL of sterile 1 M MgSO₄ and 500 µL of sterile 0.5 M CaCl₂ to 100 mL of sterilized LB broth to reach a the final concentration of 2.5 mM of both MgSO₄ and CaCl₂.

The M9 medium was prepared by adding 100 mL 5x Difco M9-salts, 0.5 mL MgSO₄ (1 M), 0.25 mL CaCl₂ (0.5 M), 5 µL CoCl₂ (1 mg/mL), 50 µL biotin (1 mg/mL), 5 mL glucose (1.5 M) and 394.2 mL autoclaved distilled water to reach a final volume of 500 mL. The 5x Difco M9-salts were composed of 33.9 g/L NaHPO₄, 15 g/L NaH₂PO₄, 2.5 g/L NaCl, and 5 g/L NH₄Cl. All solutions used were autoclaved separately prior to use. The glucose solution was sterilized by filtration through a 0.45 µmSupor membrane filter (Acrodisc). The final M9 medium was composed of 20% v/v 5x Difco M9-salts, 1 mM MgSO₄, 0.25 mM CaCl₂, 0.01 µg/mL CoCl₂, 1 µg/mL biotin, and 15 mM glucose.

Cell Culture for *Streptococcus intermedius*

S. intermedius B196 was streaked on Todd Hewitt agar supplemented with yeast extract and incubated for 3 days at 37°C at 5% CO₂. A single Colony was used to inoculate Todd Hewitt broth supplemented with yeast extract (THY). Seven replicates were incubated at 37°C. Overnight cultures were diluted tenfold in THY and grown to 0.8 OD₆₀₀.

Cell Culture for Murine Macrophages

Bone marrow progenitors were isolated from the leg bones of young (6-8 weeks) C57BL/6 female mice. The progenitor cells were cultured and differentiated for 7 days at 37°C in 150 mm Petri dishes (Fisherbrand) in 25 mL Roswell Park Memorial Institute 1640 medium (RPMI-1640) supplemented with 1% penicillin-streptomycin (P/S), 1% L-glutamine, 10% fetal bovine serum (FBS), and 15% L929-cell conditioned medium

(LCM). The cell culture was changed every 2-3 days. At day 8, 3×10^5 fully differentiated macrophages were extracted for metabolomic analyses.

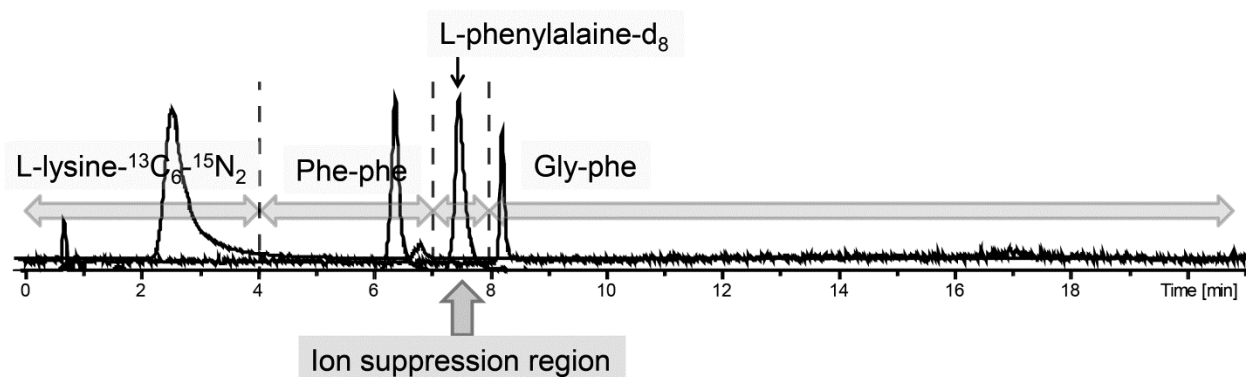


Figure S1 The schematic for normalizing the ionization efficiency of endogenous metabolite features using L-lysine-¹³C₆-¹⁵N₂, phe-phe, L-phenylalanine-d₈ and gly-phe as internal standards. The metabolite features were normalized with internal standard that eluted closest to their retention time

Extraction of extracellular medium

Directly from the culture plate, 20 μ L of the extracellular medium was extracted with 80 μ L of MeOH/EtOH (1:1) containing RS. The solution mixtures were mixed by vortex for 2 min and centrifuged at 9500x g for 3 min. The clear supernatants were collected and diluted 2-fold in 60%v/v ACN/H₂O for MS analyses. The extracellular extracts were analyzed with the same LC-MS method as the endogenous cellular extracts.

Cell Washing Removes Interferences from Extracellular Medium

The sample preparation protocol was developed to quench, harvest and extract suspension and adherent cell cultures. Most prokaryotic and eukaryotic cell lines require rich growth medium to facilitate growth. However, after aspirating the cellular medium from the cell pellet, the extracellular medium that still remained at the cell surfaces can cause significant matrix effects. This extracellular fluid contained nutrients from the growth medium as well as the exogenous metabolites released from the cell. Those metabolites, if included in the endogenous cellular extracts, can suppress MS signals and also cause bias to the true endogenous metabolite concentrations [1]. In Fig. S2A, the total ion chromatogram (TIC) of the extracellular medium of murine macrophages cultured in RPMI-1460 medium qualitatively resembled the TIC of the respective endogenous cellular extracts when the cells were not washed with PBS or saline. The largely abundant nutrients and exogenous metabolites in the growth medium masked the lesser abundant endogenous cellular metabolites. With one PBS or saline wash, the TICs of endogenous cellular extracts were completely different than those from the extracellular medium (Fig. S2B), which indicated the removal of most extracellular metabolites traces in the endogenous cellular extracts. PBS or saline was selected for cell washing in order to maintain the ionic strength balance inside and outside the cells to avoid leakage [2]. A one-time wash of the cells with PBS or saline solution was able to remove metabolite traces from extracellular medium on the cell surface.

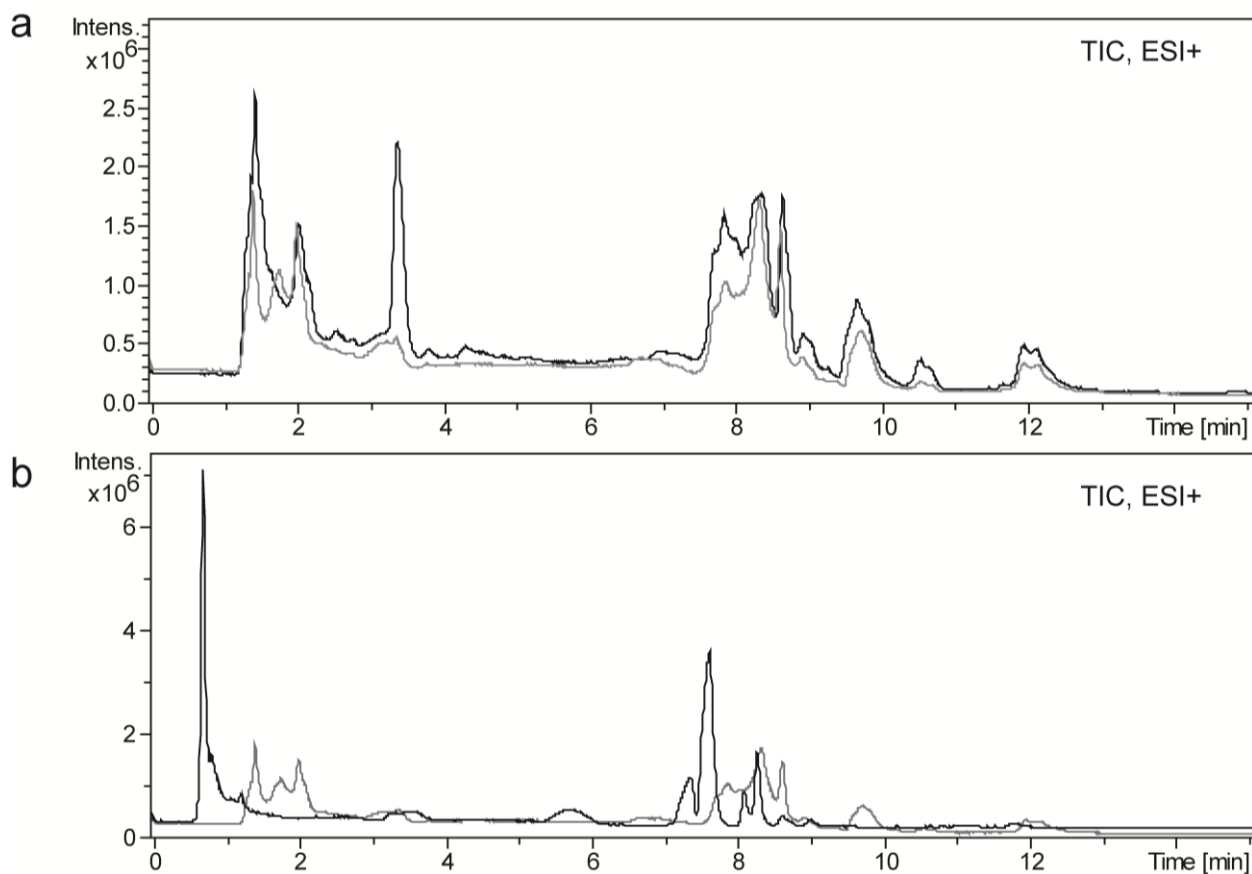


Figure S2 The total ion chromatograms of endogenous (black) and exogenous (grey) metabolomic profiles from murine macrophages cultured in RPMI-1460. MeOH/EtOH/H₂O (2:2:1) was used to extract the endogenous metabolites. For exogenous metabolome, 20 μ L of cell supernatant were extracted with 80 μ L of MeOH/EtOH once and then diluted 2 fold for LC analysis. (A) The murine macrophage cells were not washed with phosphate buffered saline (PBS) prior to extraction. (B) The murine macrophage was washed once with PBS after harvest and then extracted

Metabolism Quenching

Ice-cold PBS or saline was used to halt cellular metabolism. Typical quenching solutions such as methanol were not used to avoid cell leakage [2]. Liquid nitrogen flash freezing was not used to avoid the impact of cold shock on cellular metabolism [3]. Incubating cells at 5°C has shown greater reduction in cellular activity when compared to incubating in its physiological temperature [4]. Cold isotonic saline (0.9 w/v NaCl, 0.5°C) was reported to halt ATP to ADP and AMP conversion without damaging cells [5]. In addition, the cells were extracted with cold extraction solvent immediately following PBS/saline washes to minimize further metabolic disturbances. The entire sample preparation procedure was performed on ice or in a cold room to avoid metabolite modifications and degradation.

Cell Detachment for Adherent Cell Culture

Harvesting cells from suspension cultures for extraction was straightforward. The cells were centrifuged, the extracellular medium was aspirated, and the cell pellet was re-suspended and washed with PBS or saline, and then extracted using an extraction solvent.

For adherent cell cultures, the cell harvest was more challenging. Typically, adherent cells were detached from their growth surface via trypsinization and then centrifuged to form a cell pellet [6]. However, trypsinization altered the integrity of the cells and their extracellular environment [7,8] which consequently led to the metabolic alternations that were not specific. Furthermore, the salts in trypsin buffer and ethylenediaminetetraacetic acid (EDTA) was not ESI-MS compatible due to its high salt concentration. To address this issue, the adherent cells were washed and quenched with cold PBS/saline, and detached by physical scraping in the presence of an extraction solvent. The organic component of the extraction solvent was able to induce cell leakage and cell death, and the cell mixture was collected and extracted. The cell detachment procedure required less than 30 seconds per sample to lift the cells compared to the much more time consuming trypsinization procedure which typically requires several minutes of incubation. Also, Bi *et al.* have shown the scraping method results in a greater metabolite recovery compared to the trypsinization method due to unrecoverable metabolite leakage experienced by the latter method [8].

Further Optimization of HILIC Method

The HILIC gradient was optimized using a DoE orthogonal factorial design approach to maximize metabolome coverage (Fig. S3). The DoE approach was able to simultaneously optimize many experimental parameters with a significantly reduced number of required experiments [9]. A HILIC gradient with acetonitrile (solvent A) and 10 mM ammonium acetate in water adjusted to pH3 with formic acid (solvent B) was used. The optimization was based on a generic HILIC gradient starting with an initial hold at a high percentage of A, then decreased linearly to 35% A, held isocratically at 35% A, and then followed with equilibrium back to the initial percentage of A. Three

experimental factors were selected and optimized with the following experimental ranges: initial hold time at 95% A was adjusted between 0.5 to 2.0 min, the initial percentage of A was adjusted from 80% to 98%, and the gradient from initial high percentage of A to 35% A was set to decrease with a rate from 4% to 12% of A per min. The goal of this optimization was to minimize the number of un-retained metabolite features with $k_{app}' < 0.7$ and also to minimize the time required for all features to elute within the gradient. The LC method was designed to ensure the elution of all features during the gradient prior to the isocratic hold at 35% A in order to minimize band broadening effects.

The hold time (X1) did not reduce the number of un-retrained features (Y1), but significantly reduced the time required to complete metabolite elution (Y2). It was clear at 98% initial ACN with a 0.5 min hold followed by a drop in gradient to 35% ACN at 4% ACN per min provided that most optimum separation, for which we observed the least amount of un-retained features (y1) and required the least amount of time (y2) to completely elute all metabolites. However, the 4% gradient slope did not take advantage the separation capacity of the entire chromatography, and the features eluted ca. 3 min prior to the end of the gradient. Moreover, 98% initial ACN required longer than 10 min equilibration time. Lastly, the optimized gradient was set to be 95% ACN with 0.5 min hold, followed by a drop in gradient to 35%ACN at a rate of 5% ACN per min. The finalized HILIC-LC gradient is shown in Fig. 3.

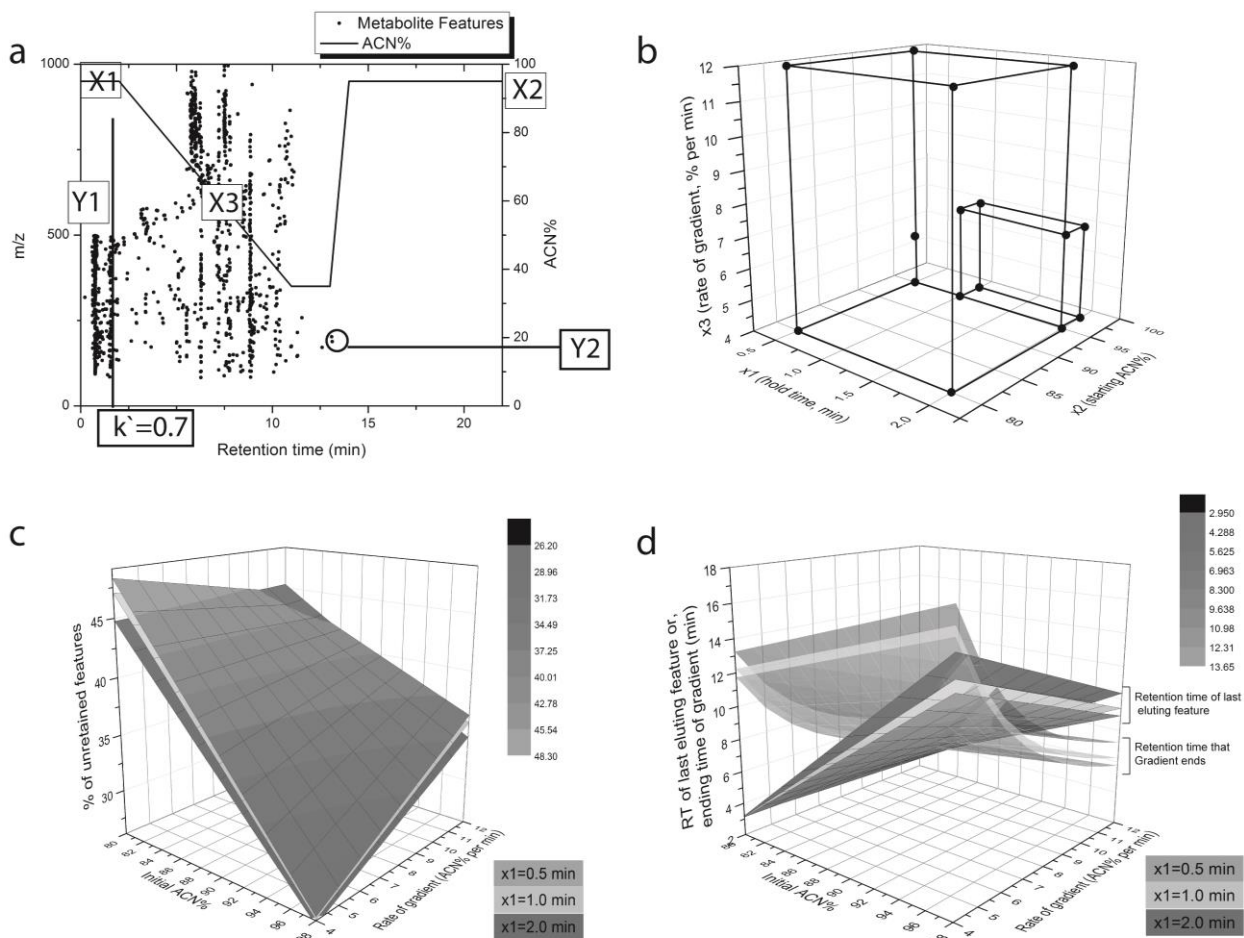


Figure S3 2^3 full factorial design of experiment for HILIC chromatography optimization. A HILIC gradient with acetonitrile (solvent A) and 10 mM ammonium acetate in water adjusted to pH3 with formic acid (solvent B) was used. *S. meliloti* extracts (2×10^9 cells) were separated on 50 mm $\times$ 2.1 mm Kinetex 2.6 μ m HILIC column and detected using a Bruker MicrOTOF II. (A) The HILIC gradient was optimized with initial hold time (X1), starting acetonitrile percentage (X2), and the rate of change of acetonitrile percentage (X3) in order to maximize the number of metabolite features with $k' < 0.7$ (Y1) and minimize the time required for all features to elute during the gradient (Y2). (B) Two 2^3 factorial design was used. The axial points were repeated in triplicates, and the center point was repeated in sextuplicate. Mathematical modelled regressions based on the 2^3 full factorial designs are shown in (C) and (D) as 3D surface plots for Y1 and Y2 respectively

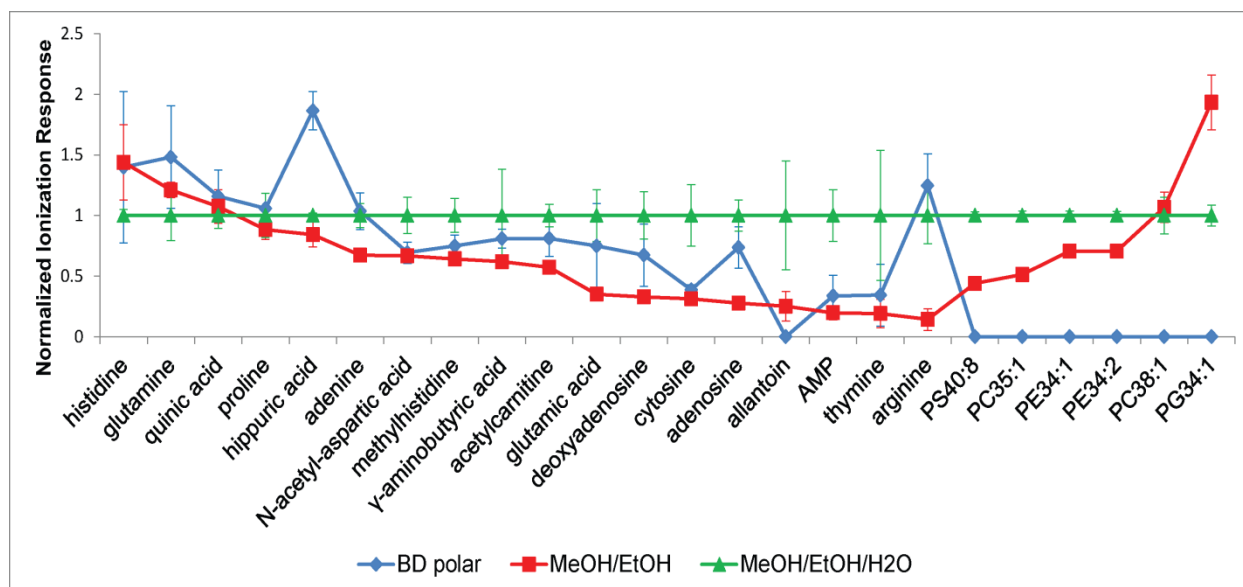


Figure S4 The normalized ionization responses of *S. meliloti* endogenous metabolites found in BD polar, MeOH/EtOH and MeOH/EtOH/H₂O extracts. Error bars correspond to two standard deviations based on a triplicate independent extractions normalized to the IS

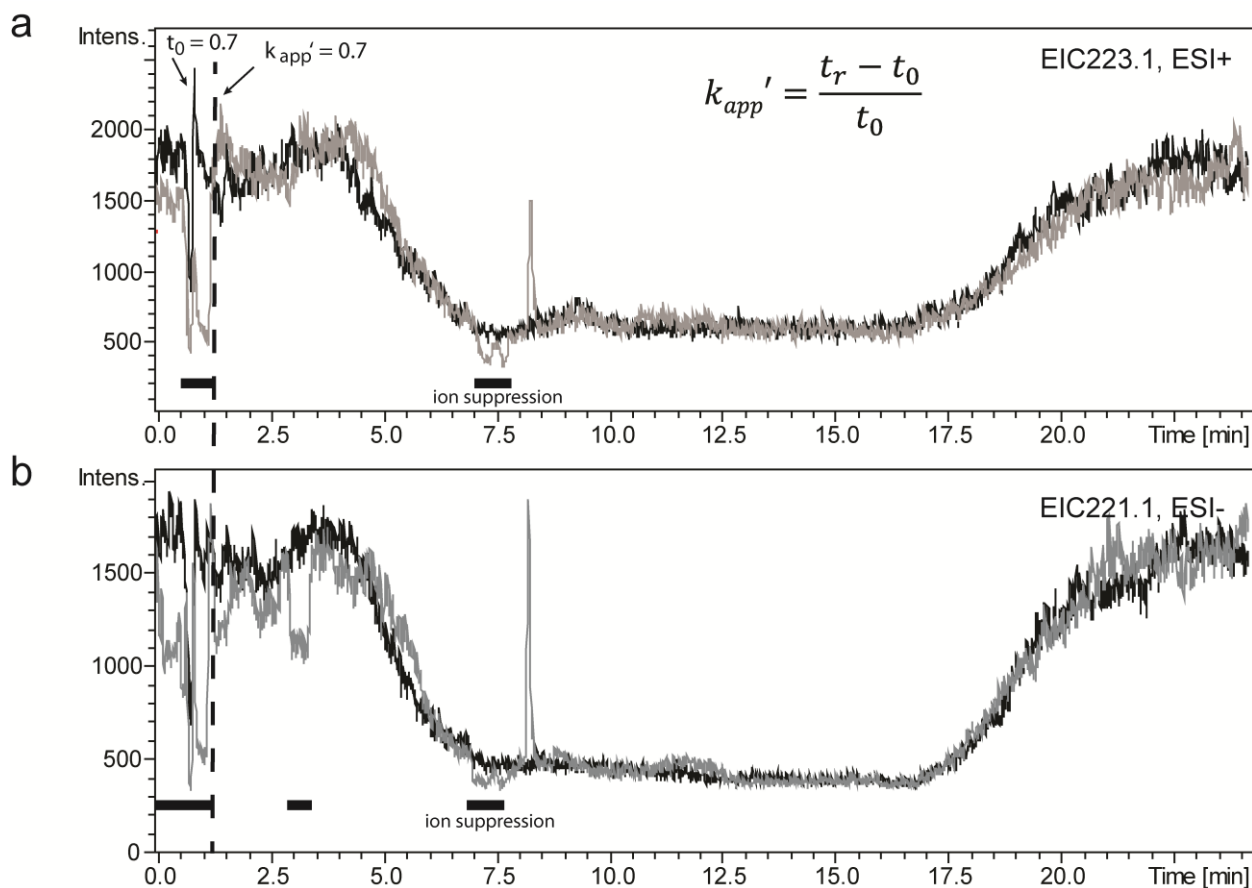


Figure S5 The extracted ion chromatogram (EIC) of (a) $[M+H]^+$ or (b) $[M-H]^-$ ion of glyphe added post-column following HILIC separation of endogenous murine macrophage MeOH/EtOH/H₂O extract (grey) or a 60%v/v ACN blank (black). Ion suppressions (black bold lines) were found at 0.5-1.2 min ($k_{app}': -0.3-0.7$) and 6.9-8.0 min ($k_{app}': 8.9-10.4$) in ESI+ mode; 0-1.2 min ($k_{app}': -0.3-0.7$), 2.9-3.5 min ($k_{app}': 3.2-4.0$), and 6.9-8.0 min ($k_{app}': 8.9-10.4$) in ESI- mode. The gly-phe was prepared to examine whether the matrix is able to suppress ionization signals, not for quantification or quality control purposes

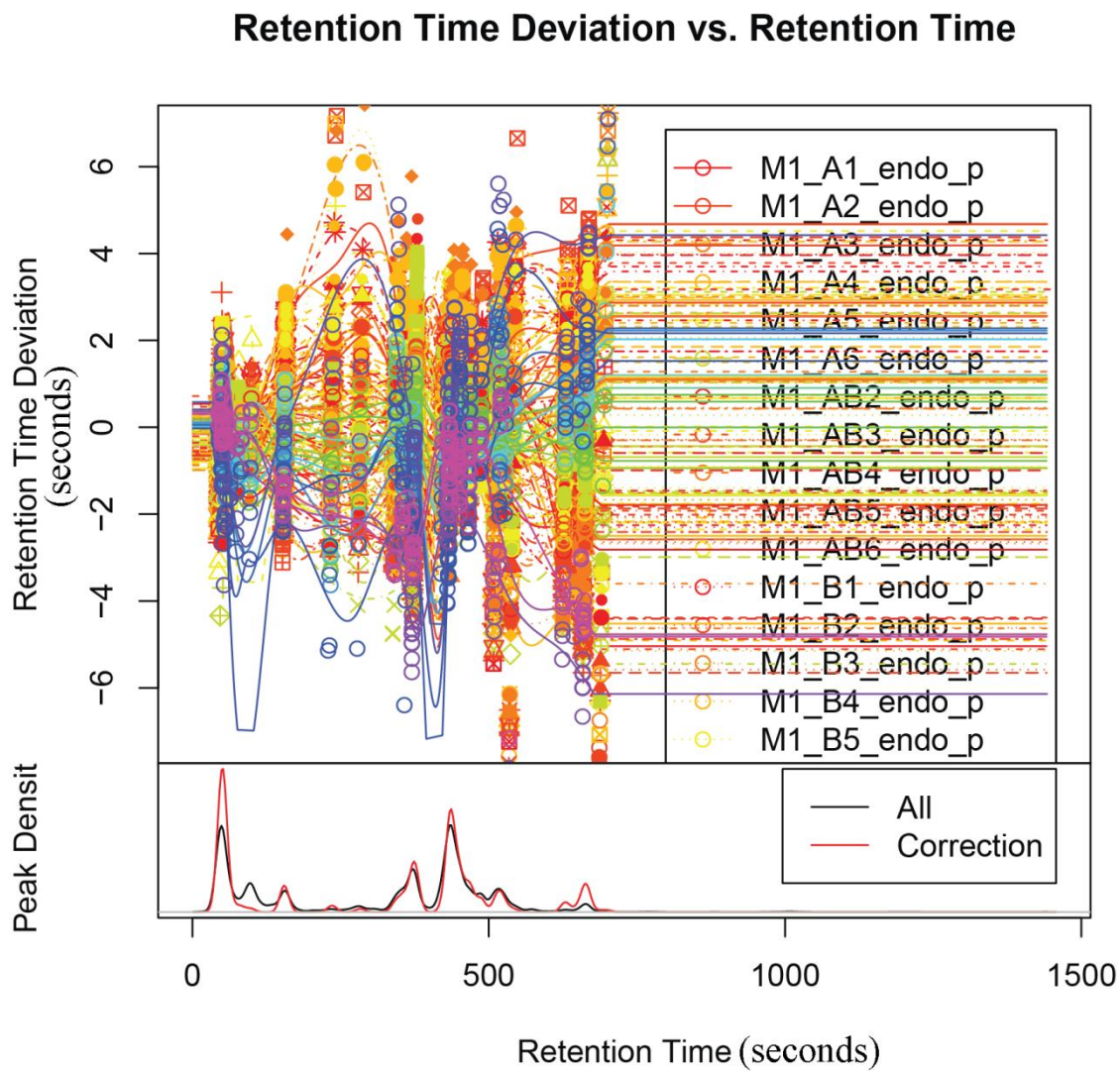


Figure S6 The retention deviation of ca. 260,000 metabolite features detected in 137 *S. meliloti* MeOH/EtOH/H₂O extracts and pooled samples in ESI⁺ mode was between ± 6 seconds.

(Figure exported directly from XCMS)

Table S1. The mass accuracies (ppm) of selected metabolites in different biological matrix ranging from Gram-positive *Streptococcus intermedius*, Gram-negative *Sinorhizobium meliloti*, to murine macrophages before and after internal calibration using endogenous sodium formate. The mass accuracies after the internal calibration with sodium formate were in bold.

			Mass Accuracy (ppm) (before/ after internal calibration)		
	RT (min)	M+H/ M+NH4*	murine macrophages	<i>Streptococcus intermedius</i> (Gram-neg.)	<i>Sinorhizobium meliloti</i> (Gram-pos.)
adenine	2.0	136.0623	75.6/ 10.5	65.1/ 10.5	48.4/ 10.1
glutamic acid	8.8	148.0610	82.5/ 16.8	42.0/ 1.6	56.8/ 2.1
methionine-d ₃	7.9	153.0777	67.1/ 1.8	56.3/ 0.9	42.9/ 2.4
phenylalanine-d ₈	7.5	174.1370	65.6/ 1.0	60.3/ 4.2	44.7/ 2.2
tryptophan-d ₅	7.4	210.1291	45.9/ 17.8	52.3/ 5.4	42.3/ 1.5
glycyl-phenylalanine	8.2	223.1083	64.8/ 2.2	59.4/ 2.8	46.7/ 3.3
phenylalanine-phenylalanine	6.4	313.1552	53.9/ 5.3	51.4/ 2.1	40.2/ 2.9
PC(34:2)	7.6	758.5694	44.1/ 1.1	37.6/ 4.1	31.6/ 0.8
PG(36:2)	1.8	792.5749*	51.0/ 6.2	48.7/ 8.4	38.5/ 8.6

Table S2. Metabolite features identified based on the accurate m/z values and retention time attained from authentic standards.

	ESI modes	m/z	retention time (min)
2'-deoxyguanosine	neg	266.088	1.93
acetylcarnitine	pos	204.104	8.39
adenine	pos	136.076	2.05
adenosine	pos	268.127	1.75
ala-glu	neg	217.083	8.97
allantoin	pos	159.090	9.14
AMP	pos	348.068	10.04
arginine	pos	175.132	10.70
cytosine	pos	112.052	2.84
deoxyadenosine	pos	252.110	1.74
glucose-1-phosphate	pos	261.039	9.75
glutamic acid	pos	148.074	8.68
glutamine	neg	145.079	8.73
glutathione	neg	306.076	8.22
hexose	neg	179.056	1.56
histidine	pos	156.090	11.10
methylhistidine	pos	170.073	9.09
methylsuccinic acid	neg	131.031	1.57
N-acetyl-aspartic acid	neg	174.094	4.28
O-phosphoserine	pos	186.016	8.86
proline	pos	116.083	8.24
quinic acid	pos	193.119	7.21
thymine	pos	127.053	1.47
UDP-GlcNAc	neg	606.075	8.61
UDP-glucose	neg	565.048	8.47
uridine-2'-monophosphate	pos	325.044	9.38
γ -aminobutyric acid	pos	104.071	8.46
PC34:1	pos	760.581	7.54
PC34:2	pos	758.565	7.53
PC35:1	pos	774.634	6.12
PC35:2	pos	772.581	7.52
PC36:0	pos	790.612	7.50
PC36:2	pos	786.598	7.50
PC38:1	pos	800.600	7.40
PE34:1	pos	718.579	5.73
PE34:2	pos	716.557	5.69

PE35:1	pos	732.553	6.01
PE35:2	pos	730.535	5.97
PE36:2	pos	744.553	5.77
PE37:2	pos	758.567	5.95
PE37:5	pos	752.521	5.98
PE38:2	pos	770.526	6.01
PE39:4	pos	782.532	5.77
PG34:1	neg	747.547	1.67
PS40:8	neg	830.492	5.98

References

1. Wu Y, Li L (2013) *Anal Chem* 85:5755–5763.
2. Bolten CJ, Kiefer P, Letisse F, Portais J-C, Wittmann C (2007) *Anal Chem* 79:3843–3849.
3. Wittmann C, Krömer JO, Kiefer P, Binz T, Heinzle E (2004) *Anal Biochem* 327:135–139.
4. Widdows J (1973) *Neth J Sea Res* 7:387–398.
5. Dietmair S, Timmins NE, Gray PP, Nielsen LK, Krömer JO (2010) *Anal Biochem* 404:155–164.
6. Lane AN, Fan TW-M (2007) *Metabolomics* 3:79–86.
7. Teng Q, Huang W, Collette TW, Ekman DR, Tan C (2009) *Metabolomics* 5:199–208.
8. Bi H, Krausz KW, Manna SK, Li F, Johnson CH, Gonzalez FJ (2013) *Anal Bioanal Chem* 405:5279–5289.
9. Gruendling T, Guilhaus M, Barner-Kowollik C (2009) *Macromol Rapid Comm* 30:589–597.

Chapter 3: Metabolic and Transcriptomic Profiling of *Streptococcus intermedius* during Aerobic and Anaerobic Growth

Fan Fei^{1,2*}, *Michelle L. Mendonca*^{3*}, *Brian E. McCarry*¹, *Dawn M. E. Bowdish*²,
Michael G. Surette^{3,4†}

The work in this chapter has been accepted by Metabolomics.

Author Contributions

BEM and MGS conceived the study. BEM, MGS, FF and MLM designed the study. MLM culture the cell, performed and analyzed the RNA-seq experiments. FF performed and analyzed the metabolomic experiments. FF and MLM wrote the manuscript. FF, MLM, DMEB and MGS edited the manuscript.

Author Affiliations:

¹ Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Canada L8S4M1

² Department of Pathology and Molecular Medicine, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada L8N3Z5

³ Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Canada L8S4K1

⁴ Department of Medicine, Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Canada L8S4K1

*Both authors contributed equally to this work.

†Corresponding author: Email: surette@mcmaster.ca; Phone: (+1)905-525-9140 ext21964; Fax: 905-522-3454

Abstract

Streptococcus intermedius, *S. constellatus*, and *S. anginosus* comprise the *Streptococcus* Anginosus/Milleri Group (SMG). They are facultative anaerobic bacteria that asymptotically colonize the upper respiratory, gastrointestinal and urogenital tracts. They are also common pathogens in pyogenic invasive infections, as well as pulmonary and urinary tract infections. Most SMG infections are polymicrobial and associated with co-infecting obligate anaerobic bacteria. To better understand the effect of oxygen on the growth and physiology of these organisms, we compared the global metabolomic and transcriptomic profiles of *S. intermedius* strain B196 under aerobic and anaerobic conditions. The largest transcriptional changes were associated with induction of oxidative stress response genes under aerobic conditions. Modest changes in expression of genes associated with primary metabolism were observed under the two conditions. Intracellular and extracellular metabolites were measured using HILIC-LCMS. Differences in the abundance of specific metabolites were correlated with observed transcription changes in genes associated with their metabolism, implying that metabolism is primarily regulated at the transcriptional level. Rather than a large shift in primary metabolism under anaerobic conditions our results suggest a modest tuning of metabolism to support the accelerated growth rate of *S. intermedius* strain B196 in the absence of oxygen. For example, under anaerobic conditions, purine metabolism, pyrimidine *de novo* synthesis and pyrimidine salvage pathways were up-regulated at metabolic and transcriptional levels. This study provides a better understanding of differences between *S. intermedius* anaerobic and aerobic metabolism. The results reflect the organism's predilection for anaerobic growth consistent with its pathogenic association with anaerobes in polymicrobial infections.

Introduction

The *Streptococcus* Milleri/Anginosus Group (SMG) is comprised of three distinct but closely related species of facultative anaerobic Gram-positive bacteria (*S. anginosus*, *S. constellatus*, and *S. intermedius*) (Gossling 1988). The SMG are often considered as commensal human microbiota and can be found asymptotically colonizing the oral cavity, upper respiratory tract, urogenital tract and gastrointestinal tract in healthy individuals (Gossling 1988; Whiley *et al.* 1992). However, the SMG are also recognized pathogens in pyogenic infections including soft tissue abscesses, pleural empyema, brain and liver abscesses, and respiratory infections (Ruoff 1988; Whiley *et al.* 1992; Coman *et al.* 1995; Shinzato and Saito 1995; Laupland *et al.* 2006; Ripley *et al.* 2006; Parkins *et al.* 2008; Sibley *et al.* 2008; Siegman-Igra *et al.* 2012; Asam and Spellerberg 2014). Phenotypic heterogeneity in this group can make their identification challenging and recent studies suggest that the SMG are under appreciated pathogens with incidence rates for pyogenic infections comparable to Group A and Group B *Streptococcus* combined (Laupland *et al.* 2006; Siegman-Igra *et al.* 2012). The SMG have been primarily associated with adults with respect to both carriage and infection, however they may be underestimated in pediatric disease (Lee *et al.* 2010).

Most infections associated with the SMG are polymicrobial with a significant burden of obligate anaerobic bacteria present in the infection site. This has been observed in lower airway infections (Shinzato and Saito 1995; Parkins *et al.* 2008; Sibley *et al.* 2008; Filkins *et al.* 2012), pleural empyema (Hocken and Dussek 1985; Van der Auwera 1985; Wong *et al.* 1995; Sibley *et al.* 2012) and abscesses (Gossling 1988; Shinzato and Saito 1994; Hirai *et al.* 2005; Sibley *et al.* 2012). Understanding how SMG adapts to aerobic and anaerobic environments may provide insight into the mechanisms used by *S. intermedius* for survival and persistence in the host during colonization and disease progression.

In this study we examined the *in vitro* growth of *S. intermedius* strain B196 in aerobic (5% CO₂) and anaerobic (90% N₂, 5% CO₂, 5% H₂) conditions using growth kinetics, transcriptomics (RNA-seq), and both intracellular and extracellular metabolomics. *S. intermedius* B196 exhibited an increased growth rate under anaerobic conditions (doubling time of 41 min vs. 54 min). We observed increased metabolite pools expected to support the more rapid growth anaerobically. These include pathways involved in DNA synthesis such as purine metabolism, pyrimidine *de novo* synthesis and pyrimidine salvage pathways. The changes in metabolism were accompanied by modest changes in gene expression for the corresponding pathways. There was also a marked increase in expression of genes associated with defense against oxidative stress under

aerobic growth conditions. These data reveal that *S. intermedius* shows shifts in metabolism consistent with increased growth kinetics under anaerobic conditions *in vitro*. This preference for an oxygen-depleted environment could contribute to growth with obligate anaerobes and disease progression in the host.

Materials and Methods

Chemicals

HPLC grade methanol (MeOH), ethanol (EtOH), acetonitrile (ACN), and water (H₂O) were purchased from Caledon laboratories (Georgetown, ON, Canada). Ammonium acetate and formic acid were purchased from Fisher Scientific Company (Ottawa, ON, Canada). 2.0mm steel chrome ball bearings were purchased from Bearing & Oil Seals Specialists Inc. (Hamilton, ON, Canada). The isotopically labelled standards for recovery determination (RS) and for peak intensity normalization (IS) were purchased from Cambridge Isotope Laboratories (Andover, MA, USA). Lipid standards were purchased from Avanti® Polar Lipids, Inc. (Alabaster, AL, USA), and other chemical standards for LC-MS were purchased from Sigma-Aldrich (St. Louis, MO, USA) and Biolog Inc. (Hayward, CA, USA).

Bacterial Strain, Media and Growth Conditions

S. intermedius B196 is an invasive isolate from the hip abscess of a cystic fibrosis patient. A complete genome sequence is available for this strain (Olson *et al.* 2013). *S. intermedius* was grown on Todd Hewitt agar supplemented with 0.5% yeast extract (THY) at 37°C in a 5% CO₂ incubator for 3 days. A single colony was inoculated into 5 mL THY broth for overnight static growth under the above conditions. For growth kinetics, overnight cultures were inoculated into 5 mL THY broth at an initial OD_{600nm} of 0.05 and cultured aerobically (5% CO₂) and anaerobically (90% N₂, 5% CO₂, 5% H₂) with optical density as well as colony forming units (CFU) recorded every hour (Supplementary information). For RNA-seq and metabolomics, the same overnight cultures were inoculated into THY broth at an initial OD_{600nm} of 0.1 and grown under either aerobic and anaerobic conditions with samples collected at mid-exponential phase (OD_{600nm} = 0.7).

Strand-specific RNA-seq

Three biological replicates were prepared for *S. intermedius* under aerobic and anaerobic conditions. A 2 mL culture from each replicate at OD_{600nm} = 0.7 was centrifuged. The cell pellets were collected and stored in RNeasy Protect bacteria reagent (Qiagen, Venlo, Netherlands) for later use at -80°C. Total cellular RNA was isolated and purified using TRIzol (Invitrogen, Carlsbad, CA, USA) and RNeasy Mini Kit (Qiagen, Venlo,

Netherlands). Ribosomal RNA (rRNA) was depleted using Ribo-Zero rRNA removal Kit for bacteria (Epicentre, Madison, WI, USA). cDNA was prepared using the Superscript III first strand cDNA synthesis kit (Invitrogen, Carlsbad, CA, USA). Strand specific RNA sequencing libraries were prepared using the dUTP approach (Parkhomchuk *et al.* 2009). The NEB-Next library preparation modules for Illumina were used for library preparation with a separate index used per biological replicate. The libraries were submitted to the McMaster Genomics Facility (McMaster University, Hamilton, Canada) for quality control (QC) and sequencing using standard Illumina protocol (Illumina HiSeq1000, San Diego). QC included assessment of fragment size using BioAnalyzer and routine qPCR quantification to quantify the amount of cDNA. The libraries were converted to FastQ format using Illumina's Casava software (version 1.8.2) with no index mismatches during demultiplexing. Approximately 20 million reads were obtained per sample with 16-17 million reads per biological replicate mapping back to the genome of *S. intermedius* B196 with high stringency. A detailed procedure can be found in Supplementary information.

RNA-seq Data Analysis

The transcriptomic data were aligned, assembled, analyzed and graphed using the Bowtie2: Tophat2: Cufflinks: Cuffdiff: CummeRbund pipeline (Trapnell *et al.* 2012). A total of 1815 genes were analyzed. Differential gene expression analysis was done using Fragments per kb per million (FPKM) values generated using the pipeline. This value takes into consideration the number of reads mapping back to each gene and normalizes that to the total number of reads generated from the sequencing run. Statistical analysis was included in the Cuffdiff analysis. Genes were classified into pathways based on BioCyc database (Caspi *et al.* 2014). The transcriptomic (RNA-seq) data are summarized in Table S1 and the dataset is included in electronic supplementary material ESM 1.

Extraction Protocol for Intracellular and Extracellular Metabolites

The extraction procedures for intracellular and extracellular metabolomics were based upon previously published work (Fei *et al.* 2014). Cells from a 2 mL cell culture at 0.7 OD_{600nm} were pelleted by centrifugation at 4°C, washed once with 1mL PBS for intracellular metabolomic analysis; from the same culture, 20 µL of the culture supernatant after centrifugation was collected for extracellular metabolomic analysis. Prior to extraction, 10 µL RS consisting of 770 µM L-methionine-d₃ and 378 µM L-tryptophan-d₅ were added to the cell pellet and THY supernatant. For intracellular metabolomics, the cell pellet was extracted with 100 µL cold methanol/ethanol/water (MeOH/EtOH/H₂O, 2:2:1) and two 2.0 mm chrome steel beads using the Powerlyzer24 (MO BIO Laboratories Inc., Carlsbad) for 2 min. The cell extract supernatant was collected after centrifugation at 9500 × g for 3 min. The cell debris (consisting of

precipitated protein and particulates) was extracted with 50 μ L MeOH/EtOH/H₂O two more times, under the same condition. For extracellular metabolomics, 20 μ L THY supernatant was extracted with 80 μ L MeOH/EtOH (1:1). The solution mixtures were vortex mixed for 2 min and centrifuged at 9500x g for 3 min. The clear supernatants were collected and diluted with 100 μ L 60%v/v ACN/H₂O. IS with 252 μ M L-phenylalanine-d₈, 151 μ M glycine-phenylalanine, and 88 μ M diphenylalanine were added to the total 150 μ L cell extracts and 200 μ L supernatant extracts. Five separate controls for THY medium were also extracted as above. Both intra- and extracellular extracts were stored in -80°C during extraction processes and before LC-MS analyses. Five biological replicates were collected for aerobic and anaerobic growth conditions and LC-MS was performed in sextuplicate for each sample.

HILIC-TOF-MS Analysis for Intracellular and Extracellular Metabolites

The HILIC-TOF-MS method and parameters were based upon previously published work (Fei *et al.* 2014). The intracellular and extracellular extracts were analyzed in two separate batches using an Agilent Technologies 1200 RR Series II liquid chromatograph (LC) coupled to a Bruker MicrOTOF II Mass Spectrometer. A 2 μ L injection was separated on a 50 mm $\times$ 2.1 mm Kinetex 2.6 μ m HILIC column of pore size of 100 Å (Phenomenex, CA, USA). The column temperature was maintained at 40 °C, and the auto sampler storage tray was set at 4°C. The mobile phases were acetonitrile (A) and 10 mM ammonium acetate in HPLC grade water adjusted to pH 3 with formic acid (B). The flow rate was kept at 0.2 mL/min during a 24-min run with the following gradient: 95 % A for 0.5 min to 35 % A at 12.5 min with an extra 0.5 min hold, then to 95% A at 14 min. The column was equilibrated at 95% A for 10 min before the next injection.

The extracts were analyzed in both ESI+ and ESI- modes. The samples were acquired in random order. A quality control pooled sample was prepared by combining 5 μ L extracts from all samples in either the intracellular or extracellular extract batches. The pooled samples were injected seven times at the beginning of each analysis and also after every 5 samples. MeOH/EtOH/H₂O blank and a standard mixture containing IS and RS were also run after every 10 samples.

LCMS Data Analysis and Metabolite Identification

The data processing and analysis were modified from a previously published protocol (Fei *et al.* 2014). Post-acquisition internal calibration using intracellular sodium formate clusters in both ESI+ and ESI- were performed with Bruker's DataAnalysis 4.0 SP4. The LC-MS data files were converted to .mzXML format using Bruker CompassXport. The metabolic features were extracted and aligned using open source

XCMS with centWave algorithm (Smith *et al.* 2006); adducts, isotopic ions, and in-source fragments were identified using CAMERA (Kuhl *et al.* 2010).

To get the final metabolite feature list, metabolite features with apparent retention factor (k_{app}) lower than 0.7 were removed. Isotopic ions, features resulting from IS, RS, and sodium formate clusters were also removed. The peak area of all metabolite features were normalized to IS. Features with greater than 20% variance in the pooled sample were removed to get the final metabolite feature list.

Metabolite features were identified based on accurate mass and retention time of authentic standards or compound analogs (for lipid identification only) with two identification points (Creek *et al.* 2014). There were 105 metabolites identified from 1885 intracellular metabolic features, 66 metabolites were identified with level 1 metabolomics standard initiative (MSI)(Sumner *et al.* 2007), 10 metabolites were putatively annotated with level 2 MSI, and 29 metabolites were assigned to compound classes with level 3 MSI. There were 26 phospholipids (PLs) identified to two phosphatidylcholines (PCs), one phosphatidylethanolamine (PEs), 20 phosphatidylglycerols (PGs), and three lyso-PGs in the intracellular metabolome based on the accurate mass and retention time of lipid analogs (Zheng *et al.* 2010; Fei *et al.* 2014). Similarly, there were 116 metabolites identified from 3382 extracellular metabolite features, with 82, 8 and 2 metabolites identified with level 1, 2, and 3 MSI, respectively. Both intra- and extracellular metabolomic data were summarized in Table S1 and the datasets are included in electronic supplementary material ESM 2 and 3, respectively.

Multivariate Statistical Analyses

Both transcriptomic and metabolomic data were subjected to principal component analysis (PCA) and orthogonal partial least-squares discriminative analysis (OPLS-DA) after pareto scaling using SIMCA-P+ 12.0.1 (Umetrics, Kinnelon, NJ). The metabolomic data were also analyzed using MetaboAnalyst 3.0 for pathway analysis based on the *Staphylococcus aureus* metabolic pathway (Xia *et al.* 2015) (electronic supplementary material ESM 4) . Intracellular and extracellular metabolite features and genes were assessed by univariate analyses such as Student's t test (two-tailed, unpaired heteroscedastic) and one-way ANOVA using Microsoft Excel 2010 and MetaboAnalyst 2.0, respectively. Metabolic features and genes with p value less than 0.05 (from Student's t test or one-way ANOVA) and fold change greater than 1.5 between conditions were considered significantly differentiated. The pathway maps were constructed based on BioCyc data for *Streptococcus intermedius* strains B196 and JTH08.

Results and Discussion

***S. intermedius* metabolism is affected by aerobic or anaerobic growth conditions**

The intracellular and extracellular metabolomes and the transcriptome were used to characterize and differentiate the responses of *S. intermedius* to anaerobic and aerobic growth conditions. Comprehensive multivariate OPLS-DA analyses were conducted on the 1885 intracellular and 3382 extracellular metabolites of *S. intermedius*. There were robust metabolomic differences between aerobic and anaerobic growth environments (Fig. 1b-c) with prediction statistic (Q^2) above 0.85 (Broadhurst and Kell 2006). The extracellular metabolomic profiles of *S. intermedius* grown in aerobic or anaerobic conditions were distinctly different from each other, and the THY medium control, implying distinct nutrient consumption and metabolite release under these growth conditions. Twenty-eight major metabolic pathways were found using MetaboAnalyst based on identified metabolites (Fig. 2).

The presence of oxygen affected both the intracellular and extracellular metabolome of *S. intermedius*. Over 37.7% of the intracellular metabolite features (710/1885 metabolite features) were differentially produced when comparing aerobic and anaerobic growth environments. Among these, 327 features were more abundant under anaerobic condition, and 383 features were less abundant (Fig. S1a). Similarly, 38.6% (1307/3382 features) of the extracellular metabolite features were significantly different among aerobic, anaerobic supernatants and THY medium controls. Although the majority of changes occurred between THY and growth conditions, 3.4% (114/3382) of the features were differentially expressed between aerobic and anaerobic supernatants (28 increased and 86 decreased under anaerobic growth conditions (Fig. S1b-d).

The RNA-seq dataset of *S. intermedius* B196 captured the transcription of 1815 genes based on the current annotation (Olson *et al.* 2013). These genes were classified to 58 gene pathways. The transcription profiles under aerobic and anaerobic growth were distinct as shown in OPLS-DA score plot (Fig. 1d). There were 625 genes that significantly affected with p values $\leq 10^{-4.5}$ (the Cuffdiff threshold cut-off) (Trapnell *et al.* 2012). Of these, 297 genes had greater than 2-fold changes in gene expression (Fig. S2). Figure 3 illustrates the percentage of genes in each pathway that were up-regulated or down-regulated with a change greater than 2-fold. The pathways can be divided into four groups based on the responses seen: (1) Aerobic Response only, where genes in pathway were up-regulated only in the presence of oxygen (2) Anaerobic Response only, where genes were up-regulated in the absence of oxygen, (3) Mixed Response, which includes a subset of genes up-regulated and a subset down-regulated under each condition, and (4) An unaffected group, where there was no oxygen dependent response. The data confirms that aerobic and anaerobic growth conditions can lead to global metabolic and transcriptional changes in *S. intermedius*.

Adaptations affecting growth

S. intermedius exhibited differential growth kinetics in the presence and absence of oxygen. Under aerobic conditions, *S. intermedius* exhibited an extended lag phase and slower overall growth rate compared to anaerobic growth conditions (Fig. 1a, S3). The doubling time during the logarithmic phase was 41 minutes under anaerobic conditions and 54 minutes under aerobic conditions based on CFUs (Fig. S3). Consistent with the increased growth rate under anaerobic conditions, pathways associated with central carbon metabolism, the arginine deaminase pathway, pyrimidine and purine metabolism were found to be increased under anaerobic conditions using both transcriptomics and metabolomics.

Central carbon metabolism of S. intermedius was up-regulated during anaerobic growth

Many facultative anaerobes such as *E. coli*, can adapt to different oxygen environments by switching from aerobic to anaerobic respiration or fermentative metabolism under oxygen-deficient conditions (Trotter *et al.* 2011). In aerobic respiration, oxygen is used as the terminal electron acceptor in the electron transport chain, which generates the proton gradient across the cell membrane and allows ATP to be generated by the cell. Lactic acid bacteria, including SMG, lack heme, the main component of the cytochromes in the electron transport chain. To compensate, lactic acid bacteria generate NAD^+ and acidic byproducts (lactate, acetate and formate) via mixed acid fermentation (Crow and Pritchard 1977). This occurs under both aerobic and anaerobic conditions. The acidic byproducts are exported and generate a proton gradient across the cell membrane, allowing ATP synthesis. Here, the effect of oxygen on glycolysis and mixed acid fermentation in *S. intermedius* were analyzed by transcriptomics and metabolomics (Fig. 4).

Genes associated with glycolysis including glucokinase (*glcK*, 2.87-fold) and fructose 1,6-bisphosphate aldolase (*fba*, 2.71-fold), were up-regulated under anaerobic conditions (Fig. 4a). Additionally, expression of the glycogen biosynthesis operon (*glgABCD*) was also increased by more than 2-fold under anaerobic growth, implying that glucose utilization exceeds energy requirements and is therefore stored in the form of glycogen. There was also a decreased expression of genes involved in the synthesis of acetoin, namely acetolactate synthase (SIR_RS12085, 5.32-fold) and aldehyde dehydrogenase (*aldB*, 6.13-fold) as has been found in other lactic acid bacteria such as *Lactococcus lactis* (Bassit *et al.* 1993).

The presence of oxygen in the growth environment is known to affect genes associated with carbohydrate uptake in non-SMG streptococci (Ahn *et al.* 2007; Ahn *et al.*

2009). The utilization of C5 and C6 sugars (e.g. ribose, glucose, trehalose) from the extracellular medium was equivalent between growth conditions. However, eight genes involved in fructose, ascorbate, glucose, mannose and N-acetylgalactosamine transport were up-regulated under anaerobic conditions (Fig. S4). Conversely, the expression of seven genes involved in the uptake of trehalose, lactose, starch and glycerol as well as six putative carbohydrate uptake genes were down-regulated during anaerobic growth. These data suggest that regulation of carbohydrate uptake is also a feature of *S. intermedius* metabolism when grown under different oxygen levels.

The NAD^+/NADH cycling pathway and carbohydrate metabolism are inextricably linked in *S. intermedius* (Fig. 4b). Under anaerobic conditions, lactate dehydrogenase (*ldh*, 3.30-fold) and malate dehydrogenase (*mleS*, 2.31-fold) were up-regulated with glycolysis genes to allow regeneration of NAD^+ (Fig. 4b, S4). However, under aerobic conditions, *nox* and *ahpCF* are up-regulated while *ldh* and *mleS* are down-regulated, implying a change in the mechanism of NAD^+/NADH cycling. Along with a lower expression of glycolysis genes under aerobic conditions, these two features may contribute to slower growth.

Up-regulation of arginine deiminase pathway in an anaerobic growth environment could lead to increase in de novo synthesis of pyrimidine

In host-pathogen interactions, the ability of bacteria to compete for nutrients with host cells is essential for bacterial colonization and pathogenesis. Amino acids have been used as the primary carbon source by bacteria in rich media (Prüb *et al.* 1994; Sezonov *et al.* 2007). However, we observed minimal net change in amino acid concentrations while comparing the growth conditions to the original THY medium aside from arginine (Fig. S5). Thus oxygen had little impact on the metabolism of other amino acids.

Arginine is known to be required for optimal SMG growth (Rogers *et al.* 1987) and was consumed during *S. intermedius* growth in our experiments. The arginine deiminase (ADI) pathway is used for energy production and also feeds into *de novo* synthesis of pyrimidine via carbamoyl phosphate (Zúñiga *et al.* 2002; Gruening *et al.* 2006; Cusumano and Caparon 2015). There are four enzymes involved in the arginine deiminase (ADI) pathway: arginine deiminase (*arcA*), ornithine carbamoyltransferase (*arcB*), carbamate kinase (*arcC*) and arginine/ornithine antiporter (*arcD*) (Gupta *et al.* 2013) (Fig. 5). The production of ammonia and ATP from carbamoyl phosphate via ArcC provides energy and protection against acid stress (Marquis *et al.* 1987; Cotter and Hill 2003). The gene expression of *arcC* was similar under both growth conditions. Therefore, the energy generation or acid stress resistance provided by carbamoyl phosphate and ADI is comparable under both conditions. On the other hand, the expression of *arcA*, *arcB* and

arcD, which lead to the synthesis of carbamoyl phosphate, were all increased significantly under anaerobic growth conditions. Additionally, the intracellular concentrations of arginine and ornithine (the carbamoyl phosphate by-product) were also elevated in the anaerobic conditions. Overall, this implies that under anaerobic condition, there is a greater conversion of arginine to carbamoyl phosphate for *de novo* synthesis of pyrimidine via the ADI pathway. Though the overall consumption of arginine was slightly greater in aerobic conditions, the arginine was preferentially metabolized to citrulline and exported, which resulted a three-fold greater extracellular citrulline concentration. Thus, up-regulation of the ADI pathway under anaerobic conditions contributed to the increased production of carbamoyl phosphate and consequently, the up-regulation of pyrimidine *de novo* synthesis. Many pathogens use the consumption of arginine as a means to impair the host's ability to produce nitric oxide (Stadelmann *et al.* 2013; Cusumano *et al.* 2014) and this may be a strategy used by *S. intermedius* in infections.

Anaerobic growth conditions enhance pyrimidine and purine metabolism in S. intermedius

The intracellular concentrations of cytosine, cytidine, adenosine, adenine, guanosine, and uridine were all higher during anaerobic growth, while their extracellular abundances were much lower implying greater cellular uptake in comparison to the aerobic culture (Fig. S6, S7). In agreement with the metabolic data, the salvage and *de novo* nucleotide synthesis pathways were also both up-regulated anaerobically at the transcription level

The pyrimidine *de novo* synthesis pathway was elevated during anaerobic growth. This was indicated by the increase of intracellular levels of orotate, a pathway intermediate, and the increased expression of pyrimidine synthesis genes (*pyrBCDEF*) (Fig. 5, S8). Pathways for production of carbamoyl phosphate (ADI pathway), aspartate (aspartate aminotransferase), and bicarbonate (*carAB*), precursors for pyrimidine *de novo* synthesis, were all up-regulated anaerobically. Moreover, the purine metabolism pathways were also elevated anaerobically. Intracellular adenine level had shown great disparity according to the aerobic and anaerobic growth condition. Compared to aerobic condition, more than three-fold increase had been observed for intracellular adenine when *S. intermedius* was anaerobically cultured. Though the adenine level in the spend media was reduced in both growth conditions, but a greater reduction was noted in the anaerobic growth. It may suggest greater influx of adenine under anaerobic growth. The genes involved in the inter-conversion between nucleosides and nucleotides were also up-regulated anaerobically to adjust to the high demand of intracellular metabolites (Fig. S6,

S7). The enhanced expression of nucleoside and nucleotide metabolism genes may contribute to the increased growth rate of *S. intermedius* under anaerobic conditions.

Adaptation to oxidative stress

Protection against oxidative stress from both internally produced and exogenous reactive oxygen species is important for the streptococci (Higuchi *et al.* 2000; Jakubovics *et al.* 2002). Oxidative stress can cause damage to iron-sulfur cluster containing proteins as well as DNA (Imlay 2013). Genes involved in oxidative stress response were the most differentially expressed genes in our study. Under aerobic conditions, NADH oxidase (*nox*, 7.38-fold) was up-regulated in comparison to anaerobic conditions, as was the alkyl-hydroperoxidase system (*ahpCF*, 25.52- and 28.35-fold), the peroxide resistance protein (*dps*; 5.74-fold) and superoxide dismutase (*sodA*; 4.96-fold). Dps removes free iron from the cell, preventing the generation of peroxides and SodA degrades superoxides while generating hydrogen peroxide which can then be reduced to water by *ahpCF* via NAD^+/NADH cycling pathway. These adaptations to oxidative stress have been observed in *S. mutans* (Higuchi 1984; Higuchi *et al.* 2000; Ahn *et al.* 2007).

Redox balance is also integral to in the cell is in part maintained through NAD^+/NADH cycling pathways oxidizing NADH to NAD^+ for glycolysis (Fig. 4b). We also observed that under aerobic conditions, genes involved in iron-sulfur cluster and iron metabolism were up-regulated in *S. intermedius* (Fig. S9) as well as genes in several DNA-repair pathways including competence (uptake of extracellular DNA), RNA metabolism, and DNA modification and DNA repair enzymes.

Other

Oxygen had minor effects on expression of virulence genes

Despite the fact that the SMG is associated with anaerobic infections (e.g. abscesses), only a minority of genes associated with virulence are differentially regulated under oxygen varying growth conditions. Under anaerobic conditions, genes in oxidative stress pathway such as sialidase (*nanA*, 2.00-fold), pullulanase (*pulA2*, 2.17-fold), and a putative membrane toxin regulator (2.57-fold) were up-regulated (Fig. 3). *nanA* and *pulA2* are associated with binding to host surfaces in streptococci (Hytönen *et al.* 2006; Brittan *et al.* 2012). On the other hand, some potential virulence genes such as proteases (Fig. 3), the bacteriocin accessory protein (*bta*, 4.00-fold) and a metalloβ-lactamase family protein (SIR_RS10820, 2.11-fold) were down-regulated anaerobically. The expression of CRISPR system (*cas1*, 2.53-fold; *cas2*, 2.01-fold; *csn2*, 2.70-fold) and nucleases (*rnc*, 2.01 fold; SIR_RS13205, 2.33 fold; *hsdR*, 2.93 fold) were increased anaerobically. These are involved in resistance to uptake of foreign genetic elements and

phage infection (Marraffini and Sontheimer 2010; Midon *et al.* 2011; Sapranaukas *et al.* 2011; Derré-Bobillot *et al.* 2013). Overall, the expression of virulence associated genes was not strongly regulated by presence or absence of oxygen and these genes may be regulated by additional host specific signals.

Cellular membrane composition re-modelled to adapt to different environmental conditions

The major phospholipid (PL) class detected in *S. intermedius* was phosphatidyl glycerol (PG) and lyso-PG. There were 13 PGs and 2 lyso-PGs that varied significantly between aerobic and anaerobic growth conditions (Fig. 6). The cellular membrane of *S. intermedius* was largely composed of saturated short-chain PLs during aerobic growth. Consistent with the observed lipid profiles, the transcriptomic analysis found enoyl-CoA hydratase protein *fabM* (also referred to as *phaB*, 2.85-fold) and beta-ketoacyl-acyl carrier protein synthase III (*fabH*, 2.50-fold), associated with the synthesis of unsaturated and branched chain lipids respectively, were down-regulated under anaerobic conditions (Choi *et al.* 2000; Marrakchi *et al.* 2002; Fozo and Quivey 2004). It is unknown whether these metabolic and transcriptomic changes could affect the cellular membrane rigidity and permeability in *S. intermedius*. Increased levels of unsaturated fatty acids have been reported in *E. faecalis* during aerobic growth (Portela *et al.* 2014).

Concluding Remarks

This study examined the global physiologic, metabolic and transcriptomic adaptations of *S. intermedius* grown in aerobic and anaerobic environments. Our study demonstrates that while *S. intermedius* is able to adapt to either condition, the anaerobic growth condition is favored with a 24% faster growth rate which also correlated with the up-regulation of the central carbon metabolism, the arginine deaminase pathway and the nucleotide *de novo* synthesis/salvage pathways. The largest transcriptional responses we observed were related to oxidative stress response under aerobic conditions. Overall, as a facultative anaerobe, *S. intermedius* is able to grow under varying oxygen tensions and may facilitate its colonization of distinct mucosal surfaces within the human host (upper respiratory, gastrointestinal and urogenital tracts). Moreover, its accelerated growth and adaption to anaerobic conditions may reflect its propensity for polymicrobial pyogenic infections with anaerobic bacteria. This adaptability allows *S. intermedius* to coexist in complex polymicrobial environments, both as a commensal and a pathogen.

Electronic Supplementary Materials:

Electronic Supplementary datasets of RNAseq, intracellular and extracellular metabolome for this chapter, ESM 1-4, can be found online (doi:).

Acknowledgement

This work is dedicated to the late Prof. Brian McCarry (1946-2013). FF was supported by an Ontario Graduate Scholarship. MGS and DMEB are supported by the CIHR and hold Canada Research Chairs. Work in the Bowdish laboratory is supported by the McMaster Immunology Research Centre (MIRC) and work in the Bowdish and Surette labs are supported by the M.G. DeGroote Institute for Infectious Disease Research (IIDR). The authors would like to thank the Centre for Microbial Chemical Biology (CMCB) at McMaster for access to the LC-MS.

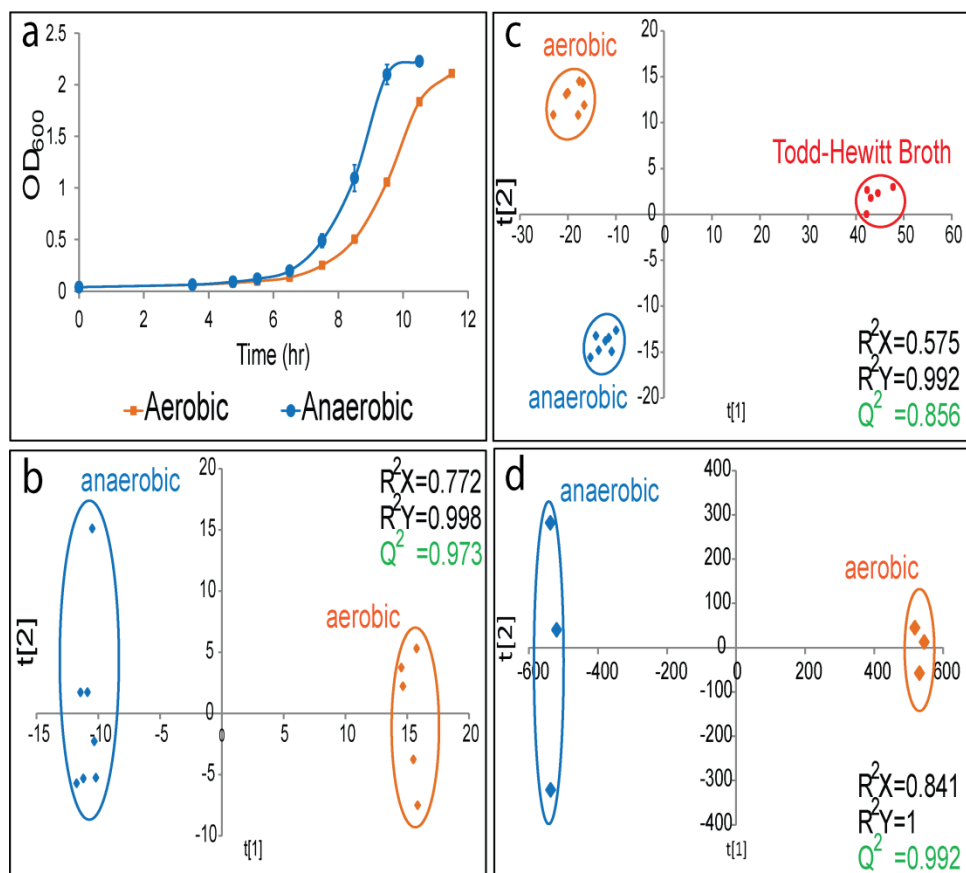


Figure 1 (a) The growth curves of *S. intermedius* cultured in either aerobic (in orange) or anaerobic (in blue) environment. The doubling time in aerobic condition was 54 min and the doubling time in anaerobic condition was 41 min. The optical density of the cells was measured every hour in triplicate. OPLS-DA score plots summarizing (b) 1885 metabolite features found in the intracellular extracts between aerobically and anaerobically cultured *S. intermedius*, (c) 3382 metabolite features found in extracellular medium of aerobically and anaerobically cultured *S. intermedius* and Todd Hewitt growth media, and (d) gene expression differences including 1815 transcripts obtained from RNA-seq of *S. intermedius* grown in either aerobic or anaerobic conditions. The samples belonging to the same treatment were highlighted with circles

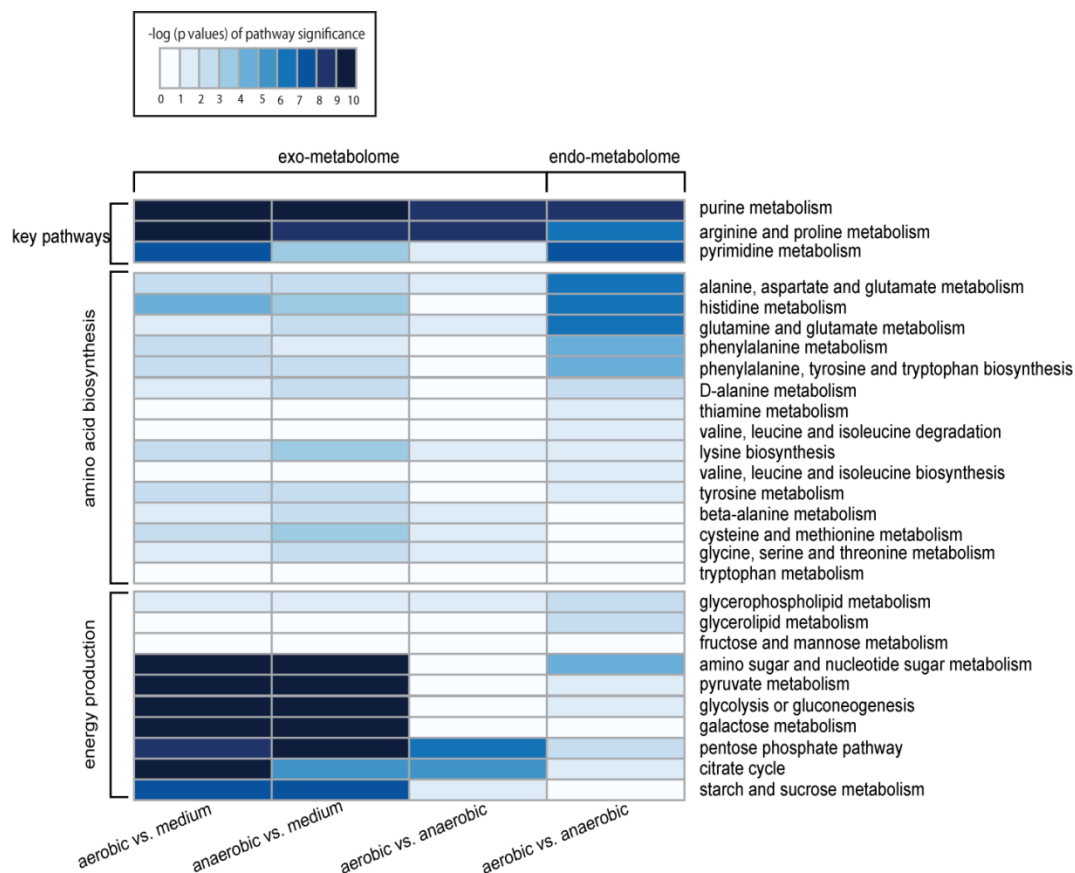


Figure 2 A heatmap of 28 pathways affected by aerobic or anaerobic growth conditions, based on identified metabolites in the intracellular and extracellular metabolomes of *S. intermedius*. The p value of each pathway was computed by pathway analysis using MetaboAnalyst 2.0 using Gram-positive *Staphylococcus aureus* as the model organism. Rows were metabolite pathways; columns were comparisons between treatments including aerobic and anaerobic growth conditions and the Todd Hewitt media blank. Comparisons were based on either intracellular *S. intermedius* cell extracts of 79 identified metabolites (exclude phospholipids) or extracellular supernatant of 92 identified metabolites. The color key indicates the $-\log_{10}$ of p values for pathway significance (refer to the color scale). The energy production pathways were calculated mostly based on C5, C6 monosaccharide and disaccharide abundances. A list of identified metabolites used for the analyses was included in the electronic supplementary material ESM 2. The citrate cycle (tricarboxylic acid (TCA) cycle) in *S. intermedius* is incomplete; however intermediates from the partial pathway are differentially produced.

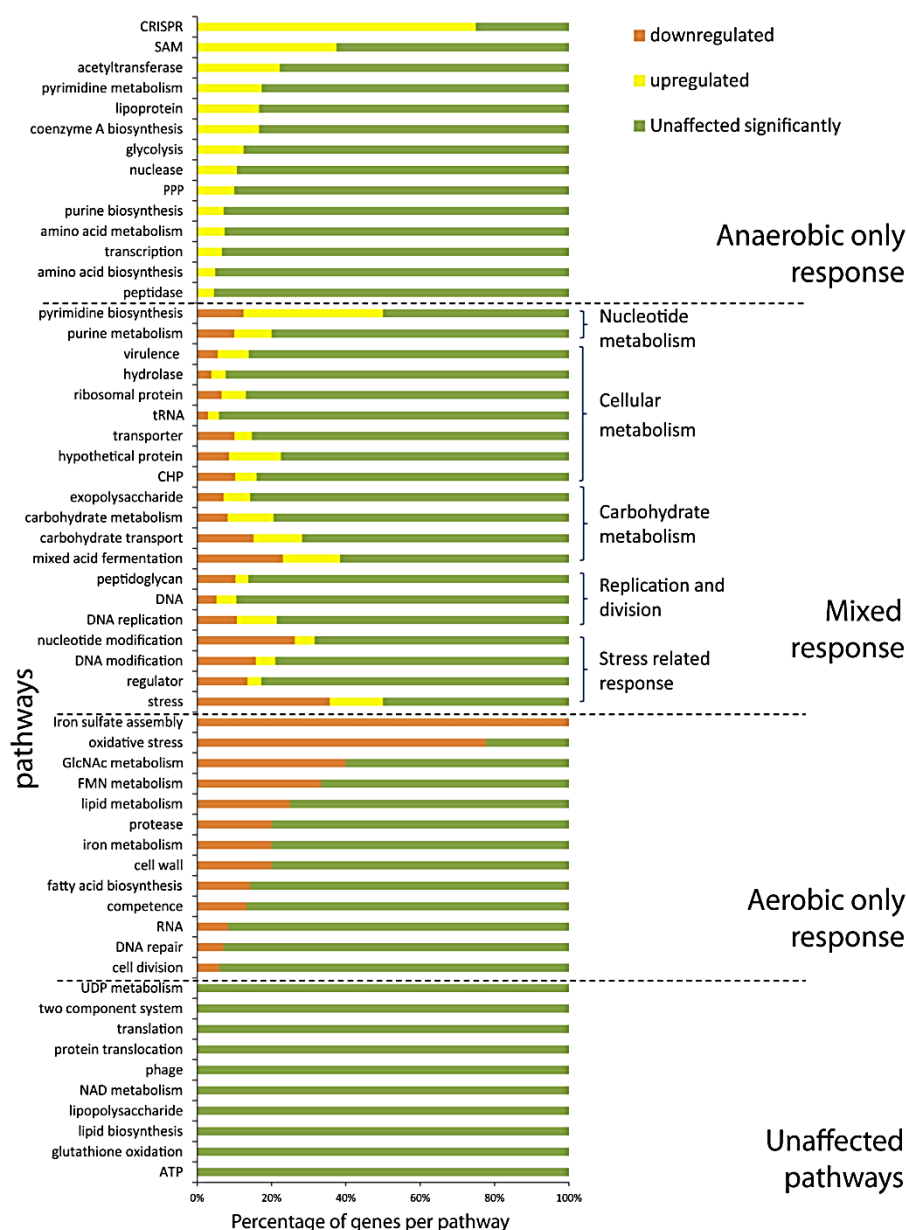


Figure 3 The effect of the presence of oxygen on gene expression of *S. intermedius*. Anaerobic gene expression was compared to aerobic and visualized with down-regulated, upregulated and unaffected genes coloured in orange, yellow and green respectively. Genes were classified into pathways and assembled into 4 groups based on response, namely, anaerobic response, mixed response, aerobic response and unaffected pathways. CRISPR: clustered regularly interspaced short palindromic repeats; SAM: S-adenosyl methionine; PPP: pentose phosphate pathway; CHP: conserved hypothetical protein; GlcNAc: N-acetylglucosamine; FMN: flavin mononucleotide; UDP: uridine diphosphate; NAD: nicotinamide adenine dinucleotide

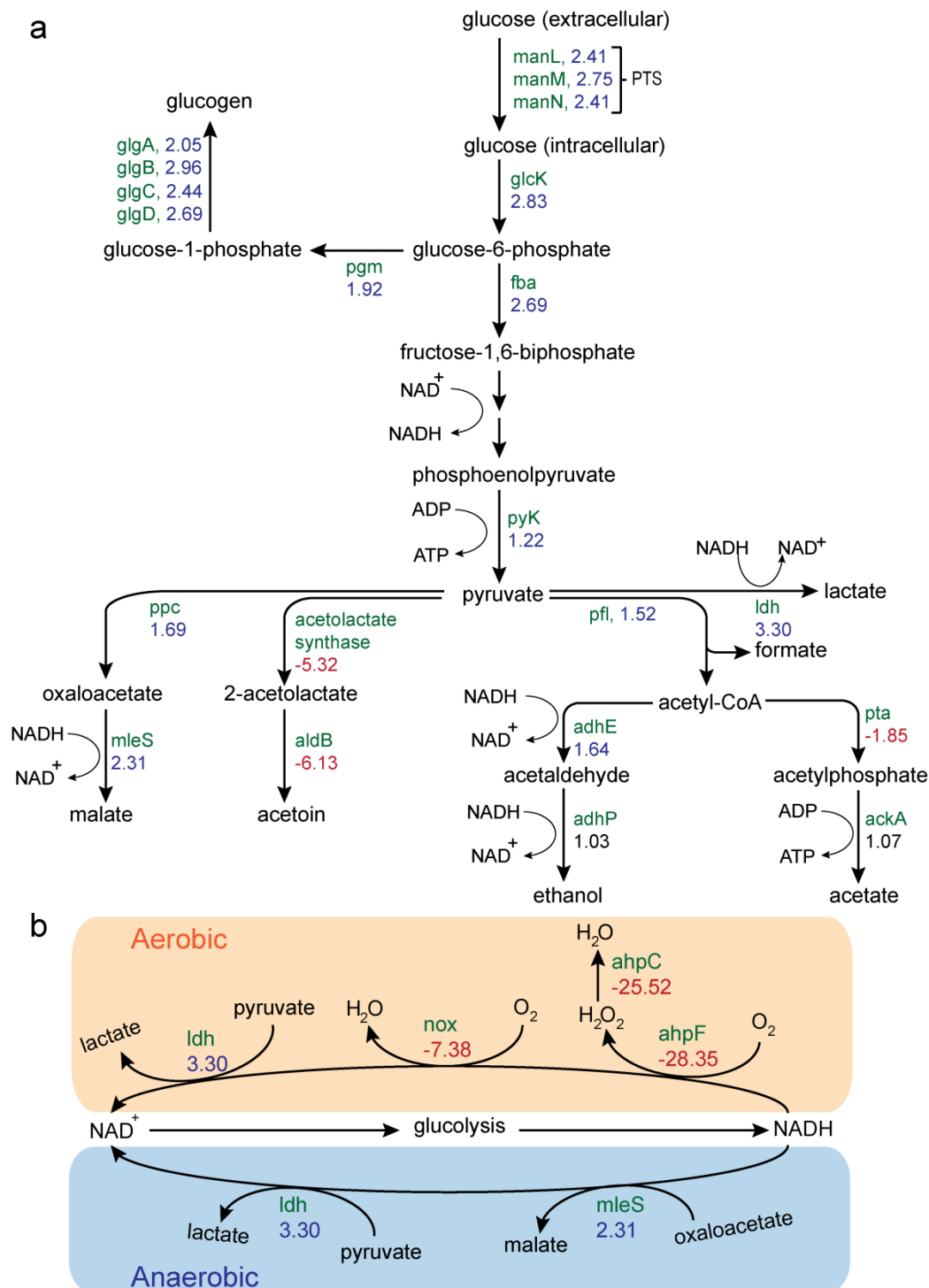


Figure 4 Overview of (a) glycolysis and mixed acid fermentation pathways and (b) NAD⁺/NADH cycling pathway of *S. intermedius*, which were affected by aerobic and anaerobic growth conditions. Genes that were up-regulated in anaerobic conditions were

indicated in blue; genes that were down-regulated in anaerobic condition were indicated in red; genes which expression was not statistically significant were indicated in black (Student's *t* test, $p > 0.05$). PTS, phosphotransferase system; *manL*, putative phosphotransferase system, mannose-specific EIIAB; *manM*, PTS system, mannose-specific IIC component; *manN*, PTS system, mannose-specific IID component; *glcK*, glucokinase putative; *fba*, fructose biphosphatealdolase; *pyK*, pyruvate kinase; *pgm*, putative phosphoglucomutase/phosphomannomutase; *glgA*, glycogen synthase, ADP-glucose type; *glgB*, glycogen branching enzyme; *glgC*, glucose-1-phosphate adenylyltransferase; *glgD*, glucose-1-phosphate adenylyltransferase, GlgD subunit; *ldh*, L-lactate dehydrogenase; *pfl*, formate C-acetyltransferase; *pta*, phosphate acetyl/butaryltransferase; *ackA*, acetate kinase; *adhE*, bifunctional acetaldehyde-CoA/alcohol dehydrogenase; *adhP*, alcohol dehydrogenase; *aldB*, alpha-acetolactate decarboxylase; *ppc*, phosphoenolpyruvate carboxylase; *mleS*, malate dehydrogenase; *nox*, NADH oxidase; *ahpC*, alkyl hydroperoxide reductase subunit C; *ahpF*, alkyl hydroperoxide reductase subunit F

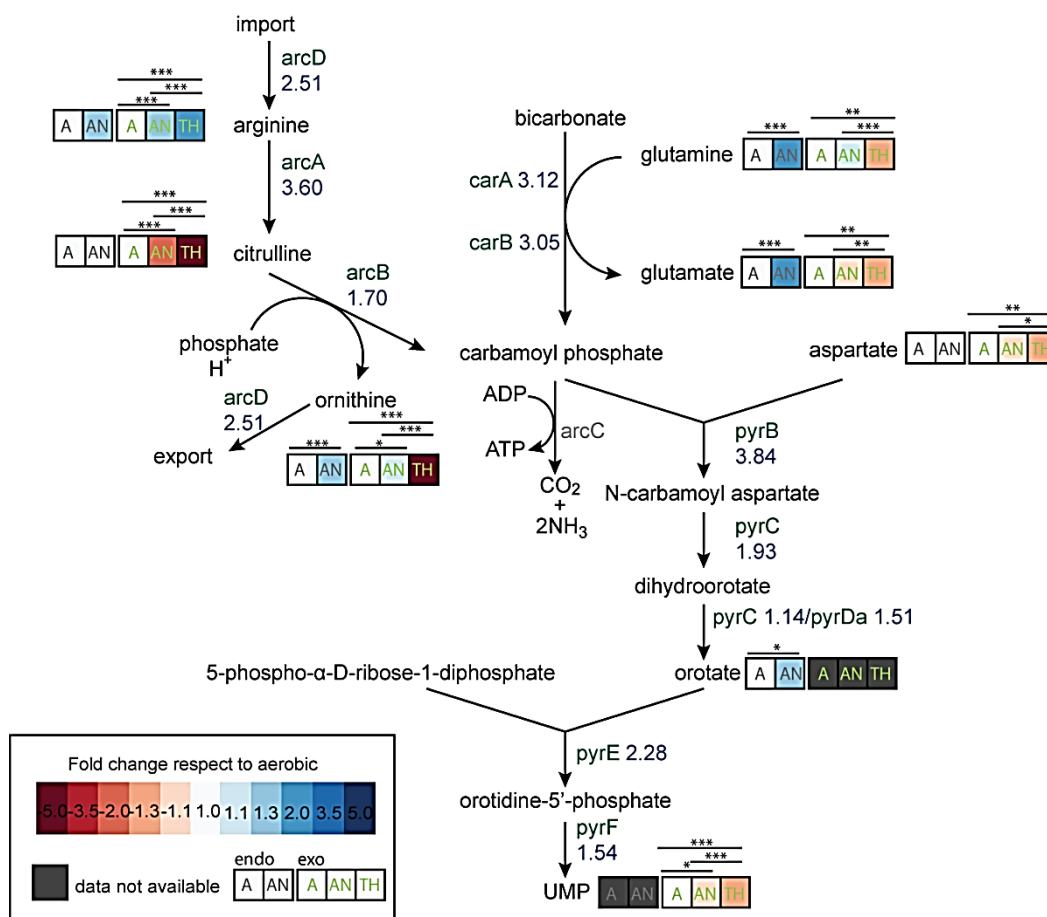


Figure 5 The arginine deiminase and pyrimidine de novo synthesis pathways. Pathway was constructed based on the BioCyc database for *S. intermedius* B196 and JTH08. The metabolite names were written in black and the gene names were written in green. The fold changes of metabolite expression were indicated in color scaled boxes for *S. intermedius* grown in aerobic (A) and anaerobic (AN) conditions and the Todd Hewitt media blank (TH). The endo-metabolome was colored in grey and the exo-metabolome was light green, fold changes in gene expressions were indicated by numerical values. The undetected metabolites were indicated with a black filled box. The fold changes in metabolite or gene levels were calculated respective to aerobic growth conditions for either endo- or exo-extracts, where an increase was shown in blue and a decrease was shown in red. n=7 except for intracellular cell extract in aerobic conditions and Todd Hewitt media. *p<0.05, **p<0.005, ***p<0.0001

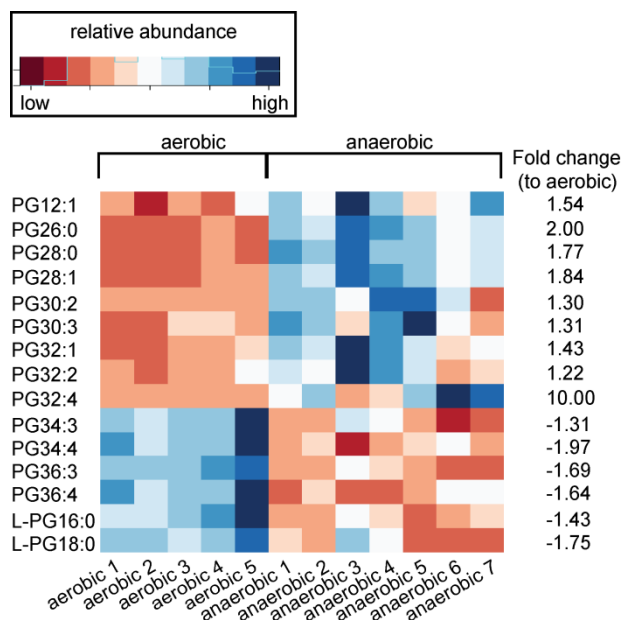


Figure 6 The heatmap of 15 statistically significantly affected phospholipids ($p < 0.05$) found in the intracellular metabolome of *S. intermedius* cultured in either aerobic or anaerobic conditions. The phospholipids were listed according to their alkyl chain length and saturation. The relative abundances of phospholipids were illustrated using a color scale, with blue indicating high abundances and red indicating low abundances. Note the chain length and saturation reported is cumulative for both fatty acid side chains.

References

- Ahn, S.-J., Ahn, S.-J., Browngardt, C.M., Burne, R.A. (2009). Changes in biochemical and phenotypic properties of *Streptococcus mutans* during growth with aeration. *Appl Environ Microbiol*, 75(8), 2517–2527.
- Ahn, S.-J., Wen, Z.T., Burne, R.A. (2007). Effects of oxygen on virulence traits of *Streptococcus mutans*. *J Bacteriol*, 189(23), 8519–8527. doi: 10.1128/JB.01180-07
- Asam, D., Spellerberg, B. (2014). Molecular pathogenicity of *Streptococcus anginosus*. *Mol Oral Microbiol*, 29(4), 145–155.
- Bassit, N., Boquien, C.-Y., Picque, D., Georges, C. (1993). Effect of initial oxygen concentration on diacetyl and acetoin production by *Lactococcus lactis* subsp. *lactis* biovar diacetylactis. *Appl Environ Microbiol*, 59(6), 1893–1897.
- Brittan, J.L., Buckeridge, T.J., Finn, A., Kadioglu, A., Jenkinson, H.F. (2012). Pneumococcal neuraminidase A: an essential upper airway colonization factor for *Streptococcus pneumoniae*. *Mol Oral Microbiol*, 27(4), 270–283.
- Broadhurst, D.I., Kell, D.B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, 2(4), 171–196.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Weerasinghe, D., Zhang, P., Karp, P.D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*, 42(D1), D459–D471.
- Choi, K.-H., Heath, R.J., Rock, C.O. (2000). β -ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J Bacteriol*, 182(2), 365–370.
- Coman, G., Pânzaru, C., Diculencu, D., Gotia, D., Cârlan, M., Dahorea, C., Butnaru, F. (1995). Pyogenic infections with different locations caused by *Streptococcus anginosus* alone or in association with anaerobic bacteria. *Rev Med Chir Soc Med Nat Iasi*, 99(3-4), 215–219.
- Cotter, P.D., Hill, C. (2003). Surviving the acid test : responses of gram-positive bacteria to low pH. *Microbiol Mol Biol Rev*, 67(3), 429–453.
- Creek, D.J., Dunn, W.B., Fiehn, O., Griffin, J.L., Hall, R.D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E.L., Sumner, L.W., Trengove, R., Wolfender, J.-L.

- (2014). Metabolite identification : are you sure ? And how do your peers gauge your confidence ? *Metabolomics*, 10(3), 350–353.
- Crow, V.L., Pritchard, G.G. (1977). Fructose 1,6-diphosphate-activated L-lactate dehydrogenase from *Streptococcus lactis*: kinetic properties and factors affecting activation. *J Bacteriol*, 131(1), 82–91.
- Cusumano, Z.T., Caparon, M.G. (2015). Citrulline protects *Streptococcus pyogenes* from acid stress using the arginine deiminase pathway and the F1F0-ATPase. *J Bacteriol*, 197(7), 1288–1296.
- Cusumano, Z.T., Watson, M.E., Caparon, M.G. (2014). *Streptococcus pyogenes* arginine and citrulline catabolism promotes infection and modulates innate immunity. *Infect Immun*, 82(1), 233–242.
- Derré-Bobillot, A., Cortes-Perez, N.G., Yamamoto, Y., Kharrat, P., Couvé, E., Da Cunha, V., Decker, P., Boissier, M.-C., Escartin, F., Cesselin, B., Langella, P., Bermúdez-Humarán, L.G., Gaudu, P. (2013). Nuclease A (Gbs0661), an extracellular nuclease of *Streptococcus agalactiae*, attacks the neutrophil extracellular traps and is needed for full virulence. *Mol Microbiol*, 89(3), 518–531.
- Fei, F., Bowdish, D.M.E., McCarry, B.E. (2014). Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal Bioanal Chem*, 406(15), 3723–3733.
- Filkins, L.M., Hampton, T.H., Gifford, A.H., Gross, M.J., Hogan, D.A., Sogin, M.L., Morrison, H.G., Paster, B.J., O'Toole, G.A. (2012). Prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. *J Bacteriol*, 194(17), 4709–4717.
- Fozo, E.M., Quivey, R.G. (2004). The fabM gene product of *Streptococcus mutans* is responsible for the synthesis of monounsaturated fatty acids and is necessary for survival at low pH. *J Bacteriol*, 186(13), 4152–4158.
- Gossling, J. (1988). Occurrence and pathogenicity of the *Streptococcus milleri* group. *Rev Infect Dis*, 10(2), 257–285.
- Gruening, P., Fulde, M., Valentin-Weigand, P., Goethe, R. (2006). Structure, regulation, and putative function of the arginine deiminase system of *Streptococcus suis*. *J Bacteriol*, 188(2), 361–369.
- Gupta, R., Yang, J., Dong, Y., Swiatlo, E., Zhang, J.-R., Metzger, D.W., Bai, G. (2013). Deletion of arcD in *Streptococcus pneumoniae* D39 impairs its capsule and attenuates virulence. *Infect Immun*, 81(10), 3903–3911.

- Higuchi, M. (1984). The Effect of oxygen on the growth and mannitol fermentation of *Streptococcus mutans*. *J Gen Microbiol*, 130(7), 1819–1826.
- Higuchi, M., Yamamoto, Y., Kamio, Y. (2000). Molecular biology of oxygen tolerance in lactic acid bacteria: functions of NADH oxidases and Dpr in oxidative stress. *J Biosci Bioeng*, 90(5), 484–493.
- Hirai, T., Kimura, S., Mori, N. (2005). Head and neck infections caused by *Streptococcus milleri* group: an analysis of 17 cases. *Auris Nasus Larynx*, 32(1), 55–58.
- Hocken, D.B., Dussek, J.E. (1985). *Streptococcus milleri* as a cause of pleural empyema. *Thorax*, 40(8), 626–628.
- Hytönen, J., Haataja, S., Finne, J. (2006). Use of flow cytometry for the adhesion analysis of *Streptococcus pyogenes* mutant strains to epithelial cells: investigation of the possible role of surface pullulanase and cysteine protease, and the transcriptional regulator Rgg. *BMC Microbiol*, 6(1), 18.
- Imlay, J.A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol*, 11(7), 443–454.
- Jakubovics, N.S., Smith, A.W., Jenkinson, H.F. (2002). Oxidative stress tolerance is manganese (Mn^{2+}) regulated in *Streptococcus gordonii*. *Microbiology*, 148(Pt10), 3255–3263.
- Kuhl, C., Tautenhahn, R., Neumann, S. (2010). LC-MS Peak Annotation and Identification with CAMERA.
- Laupland, K.B., Ross, T., Church, D.L., Gregson, D.B. (2006). Population-based surveillance of invasive pyogenic streptococcal infection in a large Canadian region. *Clin Microbiol Infect*, 12(3), 224–230.
- Lee, E.Y., Khatwa, U., McAdam, A.J., Bastos, M. d’Almeida, Mahmood, S.A., Ervoes, J.P., Boisselle, P.M. (2010). *Streptococcus milleri* group pleuropulmonary infection in children: computed tomographic findings and clinical features. *J Comput Assist Tomogr*, 34(6), 927–932.
- Marquis, R.E., Bender, G.R., Murray, D.R., Wong, A. (1987). Arginine deiminase system and bacterial adaptation to acid environments. *Appl Environ Microbiol*, 53(1), 198–200.
- Marraffini, L.A., Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev Genet*, 11(3), 181–90.

- Marrakchi, H., Zhang, Y.-M., Rock, C.O. (2002). Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem Soc Trans*, 30(part 6), 1050–1055.
- Midon, M., Schäfer, P., Pingoud, A., Ghosh, M., Moon, A.F., Cuneo, M.J., London, R.E., Meiss, G. (2011). Mutational and biochemical analysis of the DNA-entry nuclease EndA from *Streptococcus pneumoniae*. *Nucleic Acids Res*, 39(2), 623–634.
- Olson, A.B., Kent, H., Sibley, C.D., Grinwis, M.E., Mabon, P., Ouellette, C., Tyson, S., Graham, M., Tyler, S.D., Van Domselaar, G., Surette, M.G., Corbett, C.R. (2013). Phylogenetic relationship and virulence inference of *Streptococcus Anginosus* Group : curated annotation and whole-genome comparative analysis support distinct species designation. *BMC Genomics*, 14895.
- Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*, 37(18), e123.
- Parkins, M.D., Sibley, C.D., Surette, M.G., Rabin, H.R. (2008). The *Streptococcus milleri* group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatr Pulmonol*, 43(5), 490–497.
- Portela, C.A.F., Smart, K.F., Tumanov, S., Cook, G.M., Villas-Bôas, S.G. (2014). Global metabolic response of *Enterococcus faecalis* to oxygen. *J Bacteriol*, 196(11), 2012–2022.
- Prüb, B.M., Nelms, J.M., Park, C., Wolfe, A.J. (1994). Mutations in NADH:Ubiquinone Oxidoreductase. *J Bacteriol*, 176(8), 2143–2150.
- Ripley, R.T., Cothren, C.C., Moore, E.E., Long, J., Johnson, J.L., Haenel, J.B. (2006). *Streptococcus milleri* infections of the pleural space: operative management predominates. *Am J Surg*, 192(6), 817–821.
- Rogers, A.H., Zilm, P.S., Gully, N.J. (1987). Influence of arginine on the coexistence of *Streptococcus mutans* and *S. milleri* in glucose-limited mixed continuous culture. *Microb Ecol*, 14(3), 193–202.
- Ruoff, K.L. (1988). *Streptococcus anginosus* (“*Streptococcus milleri*”): the unrecognized pathogen. *Clin Microbiol Rev*, 1(1), 102–108.
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res*, 39(21), 9275–9282.

- Sezonov, G., Joseleau-Petit, D., D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol*, 189(23), 8746–8749.
- Shinzato, T., Saito, A. (1994). A mechanism of pathogenicity of “*Streptococcus milleri* group” in pulmonary infection: synergy with an anaerobe. *J Med Microbiol*, 40118–123.
- Shinzato, T., Saito, A. (1995). The *Streptococcus milleri* group as a cause of pulmonary infections. *Clin Infect Dis*, 21(Supplement 3), S238–S243.
- Sibley, C.D., Church, D.L., Surette, M.G., Dowd, S.E., Parkins, M.D. (2012). Pyrosequencing reveals the complex polymicrobial nature of invasive pyogenic infections: microbial constituents of empyema, liver abscess, and intracerebral abscess. *Eur J Clin Microbiol Infect Dis*, 31(10), 2679–2691.
- Sibley, C.D., Parkins, M.D., Rabin, H.R., Duan, K., Norgaard, J.C., Surette, M.G. (2008). A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci USA*, 105(39), 15070–15075.
- Siegman-Igra, Y., Azmon, Y., Schwartz, D. (2012). *Millieri group streptococcus*--a stepchild in the viridans family. *Eur J Clin Microbiol Infect Dis*, 31(9), 2453–2459.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*, 78(3), 779–787.
- Stadelmann, B., Hanevik, K., Andersson, M.K., Bruserud, O., Svärd, S.G. (2013). The role of arginine and arginine-metabolizing enzymes during *Giardia* – host cell interactions in vitro. *BMC Microbiol*, 13256.
- Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C. a, Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reilly, M.D., Thaden, J.J., Viant, M.R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7(3), 562–578.
- Trotter, E.W., Rolfe, M.D., Hounslow, A.M., Craven, C.J., Williamson, M.P., Sanguinetti, G., Poole, R.K., Green, J. (2011). Reprogramming of *Escherichia coli* K-12

metabolism during the initial phase of transition from an anaerobic to a micro-aerobic environment. *PloS One*, 6(9), e25501.

Van der Auwera, P. (1985). Clinical significance of *Streptococcus milleri*. *Eur J Clin Microbiol*, 4(4), 386–390.

Whiley, R.A. a, Beighton, D., Winstanley, T.G.G., Fraser, H.Y.Y., Hardie, J.M.M. (1992). *Streptococcus intermedius*, *Streptococcus constellatus*, and *Streptococcus anginosus* (the *Streptococcus milleri* group): association with different body sites and clinical infections. *J Clin Microbiol*, 30(1), 243–244.

Wong, C.A., Donald, F., Macfarlane, J.T. (1995). *Streptococcus milleri* pulmonary disease: a review and clinical description of 25 patients. *Thorax*, 50(10), 1093–1096.

Xia, J., Sinelnikov, I. V., Han, B., Wishart, D.S. (2015). MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res*, 43(W1), W251–W257.

Zheng, L., T'Kind, R., Decuypere, S., von Freyend, S.J., Coombs, G.H., Watson, D.G. (2010). Profiling of lipids in *Leishmania donovani* using hydrophilic interaction chromatography in combination with Fourier transform mass spectrometry. *Rapid Commun Mass Sp*, 242074–2082.

Zúñiga, M., Pérez, G., González-Candelas, F. (2002). Evolution of arginine deiminase (ADI) pathway genes. *Mol Phylogenet Evol*, 25(3), 429–444.

Supplementary Information

Growth Curve Measurement for *S. intermedius*

Optical density measurements

Samples were taken from the culture and used to measure the absorbance at 600 nm using cuvettes and blanking against the media. The Nanodrop 2000 (ThermoFisher Scientific, Waltham, USA) was used to measure the absorbance.

Colony forming units (CFU) measurements

Colony forming units were measured by taking a sample of the culture and serially diluting it before plating for colonies. A 1:10 serial dilution was used in volumes of 100 μ L. A 5 μ L volume of each dilution was plated and colonies counted to quantify the amount of bacteria.

RNA-seq Transcriptomics

RNA Purification

Three biological replicates were analyzed per condition tested (aerobic vs. anaerobic). The *S. intermedius* cell pellet was collected from samples with OD₆₀₀ 0.7. Broth cultures were centrifuged and pellets were resuspended in 700 μ L RNeasy Protect Bacteria reagent (Qiagen, Venlo, Netherlands) with 100 μ g/mL rifampicin and incubated for 10 minutes at room temperature before freezing at -80°C. Frozen samples were defrosted at room temperature and centrifuged for 20 minutes at 4°C. The cell pellet were resuspended in 700 μ L RNase free water with lysozyme (10 μ L of 100 mg/mL) and mutanolysin (5 μ L of 10 U/ μ L). The suspension was incubated at 37°C for 45 minutes. Cells were then centrifuged and treated with 1 ml TRIzol (Invitrogen, Carlsbad, CA, USA). The aqueous phase was collected and an equal volume of 70%v/v ethanol was added, and the RNeasy Mini Kit (Qiagen, Venlo, Netherlands) to isolate DNA-free total cellular RNA. Ribosomal RNA (rRNA) was depleted using Ribo-Zero rRNA removal Kit for Bacteria (Epicentre, Madison, WI, USA) according to the manufacture's protocol. Briefly, resuspended magnetic beads were washed and prepared. A 15 μ L volume of purified RNA was treated with 4 μ L RiboZero reaction buffer, 10 μ L of RiboZero RNA removal solution and 11 μ L of RNase free water and incubated at 68°C for ten minutes. The treated RNA was added to the magnetic beads and vortexed. It was then incubated at 50°C for 7 minutes before placing on the magnetic stand to separate the beads from the rRNA free supernatant. The RNA from the supernatant was purified using the Agencourt RNAClean XP Beads (Beckman Coulter, Brea, USA) as per directions. Agencourt RNAClean beads (180 μ L) were added to the supernatant (85 μ L). RNA was eluted from the beads with 32 μ L of RNase free water. An Experion RNA StdSens chip was used to confirm depletion of rRNA. This was followed with a DNase digestion using TURBO

DNase (Life Technologies, Carlsbad, USA). The reaction consisted of 30 μ L of RNA, 1.5 μ L of DNase and 3.5 μ L of 10x Buffer. The reaction was incubated for 30 minutes at 37°C. The RNA was purified again using the Agencourt RNAClean XP beads. To the DNase reaction (35 μ L), 70 μ L of magnetic beads were added and the final elution of RNA was done with 12 μ L of RNase free water.

cDNA Synthesis from RNA

The RNA was then converted to cDNA using the Superscript III first strand cDNA synthesis kit (Life Technologies, Carlsbad, CA, USA). Random hexamers were used as primers for cDNA synthesis. Single strand cDNA was purified using the Agencourt RNAClean XP beads (2x) and eluted with 22 μ L RNase free water. Strand specific RNA sequencing was carried out (Parkhomchuk *et al.* 2009). Complementary second strand of cDNA incorporated uridine instead of thymidine. To synthesize this, the purified single stranded cDNA (22 μ L) and 7.5 mM of each dNTP (dATP, dCTP, dGTP and dUTP) was treated with RnaseH and Klenow Fragment DNA polymerase (Invitrogen, Carlsbad, CA, USA) in a final volume of 40 μ L at 16°C for 2 hours. The double stranded cDNA was purified using the Agencourt AMPure XP beads (Beckman Coulter, Brea, USA) with 2x volume of beads added to the DNA.

cDNA Library Preparation

The cDNA was fragmented into $\sim$ 300 bp lengths using the covaris S220 ultrasonicator with 175 W peak power, 10% duty factor, 200 cycles/burst for 430 seconds (Covaris, Woburn, Massachusetts, USA). Fragmented cDNA ends were repaired using the NEBNext End Repair Module (New England Biolabs, Ipswich, Massachusetts, USA) as per directions. The AMPure XP beads were used to purify cDNA by adding 1x volume of beads and eluting with 32 μ L of RNase free water. To facilitate ligation of adaptors, dA-tailing of the cDNA fragments was conducted using the NEBNext dA-tailing Module (New England Biolabs, Ipswich, Massachusetts, USA). AMPure DNA XP bead purification was again conducted with 1x beads and cDNA eluted with 25 μ L RNase free water. The NEBNext Adaptor Ligation Module (New England Biolabs, Ipswich, Massachusetts, USA) was next used as per directions and followed with adding USER enzyme to the reaction. The USER enzyme (New England Biolabs, Ipswich, Massachusetts, USA) generates gaps where uracil is found in the cDNA. This was done for the adaptor as well as the second strand of cDNA, which consisted of uracils, in order for the second strand of cDNA to be degraded. Agencourt AMPure bead purification was then conducted with 1x beads and cDNA eluted with 20 μ L RNase free water. The final preparation step consisted of a Phusion High fidelity PCR (Life Technologies, Carlsbad, USA) with the primer index (specific for each biological replicate and condition) and the universal primer for 8 cycles using the directions for the NEBNext Kit.

DNA Sequencing

The libraries were submitted to the McMaster Genomics Facility (McMaster University) for quality control and sequencing. QC included assessment of fragment size on the BioAnalyzer and qPCR quantification. The libraries were then pooled in equimolar amounts, denatured, and diluted to 12 pM; sequencing was performed using 40% of one lane on the HiSeq 1000 with 101 bp paired end reads according to standard Illumina protocols. Following sequencing, the libraries were converted to FastQ format using Illumina's Casava software (version 1.8.2, San Diego, California, USA). No index mismatches were allowed during demultiplexing. Approximately 20 million reads were obtained per condition for each biological replicate, of which between 16 to 17 million reads mapped back to the genome.

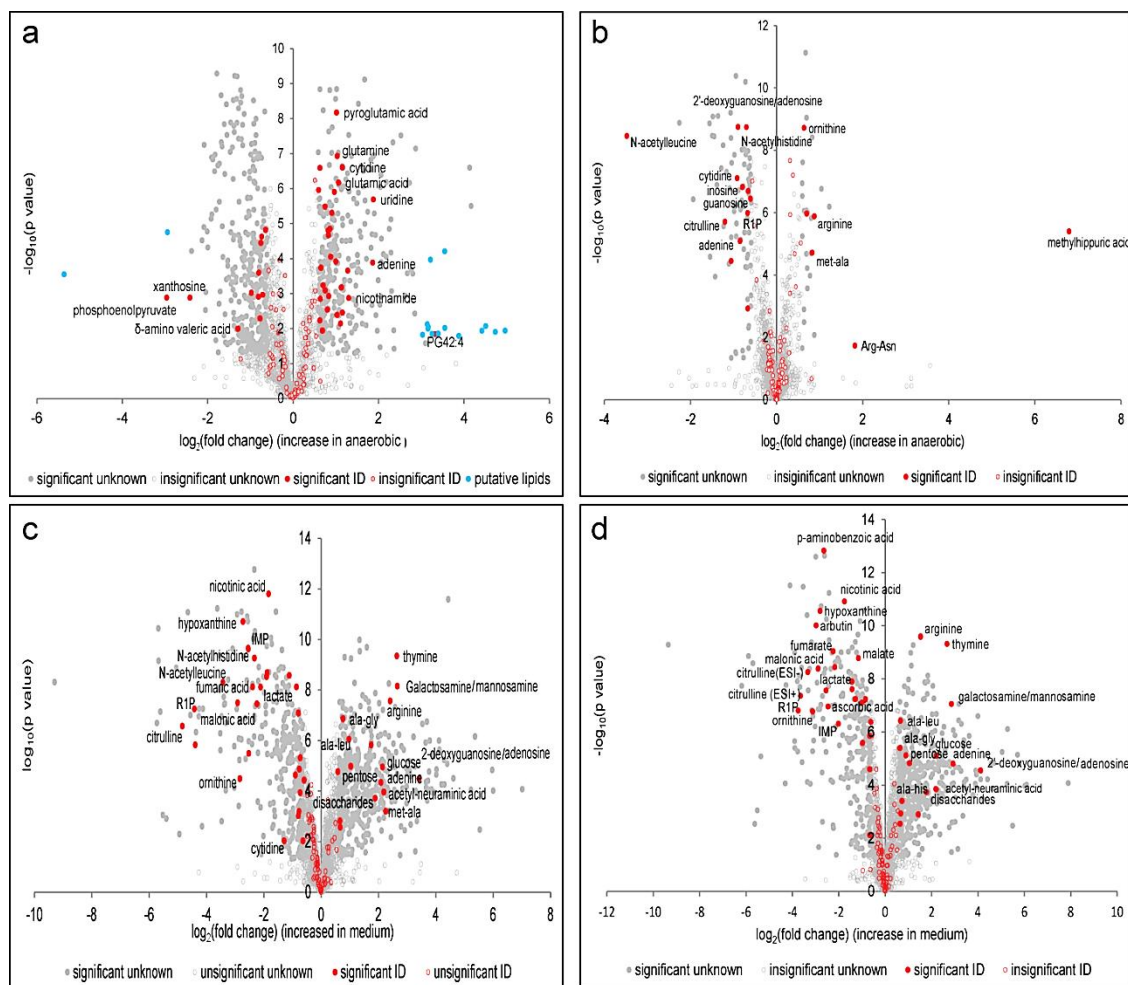


Figure S1 Global metabolomic differences between aerobically and anaerobically cultured *S. intermedius* shown by volcano plots, comparing (a) intracellular metabolome profiles of aerobic and anaerobic conditions, (b) extracellular metabolome profiles of aerobic and anaerobic conditions, (c) extracellular metabolome profiles of aerobic condition and Todd Hewitt media blank, and (d) extracellular metabolome profiles of anaerobic condition and Todd Hewitt media blank. Significant metabolite features with $p < 0.05$, and greater than 1.5 fold changes were indicated with solid circles, others were labelled with open circles. Identified metabolite features were colored in red, putative lipids were indicated in blue, and the unknowns were colored in grey. Some known significant metabolite features were labelled with their chemical names. Some metabolites appear multiple times in one volcano plot due to detection in both ESI- and ESI+ modes or as adduct ions

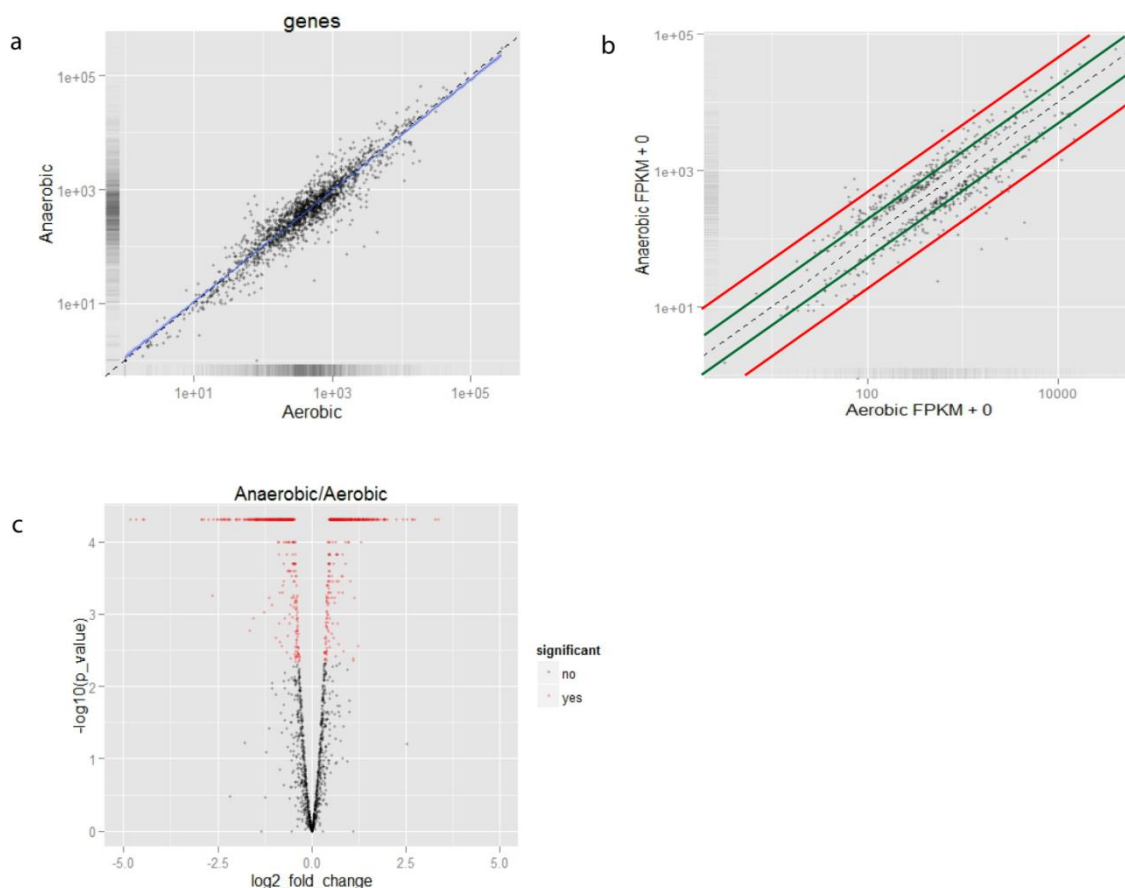


Figure S2 Overview of transcriptomic results for *S. intermedius* growth under aerobic and anaerobic conditions using CummeRbund. The scatter plot in (a) depicts the expression of all genes (1815 genes) under aerobic and anaerobic conditions. The theoretical correlation for equivalent expression under the two conditions (Blue) overlaps with the actual correlation. Scatter plot analysis of significant genes (625 genes) with p-value of 10^{-4.5} is shown in (b). The lines correlating to 2 fold and 4 fold changes in gene expression are shown. There are few genes upregulated above 4 fold, with the majority of genes being upregulated under aerobic conditions. The volcano plot in (c) highlights statistically significant genes in red. Figures were generated using CummeRbund (Trapnell *et al.* 2012)

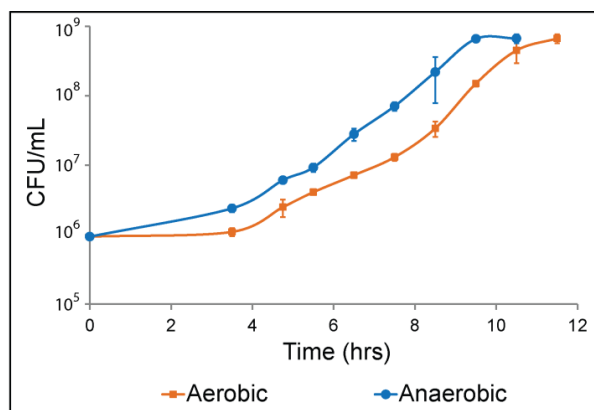


Figure S3 The growth curves (CFU/mL over time) of *S. intermedius* cultured in either aerobic (black) or anaerobic (blue) environment

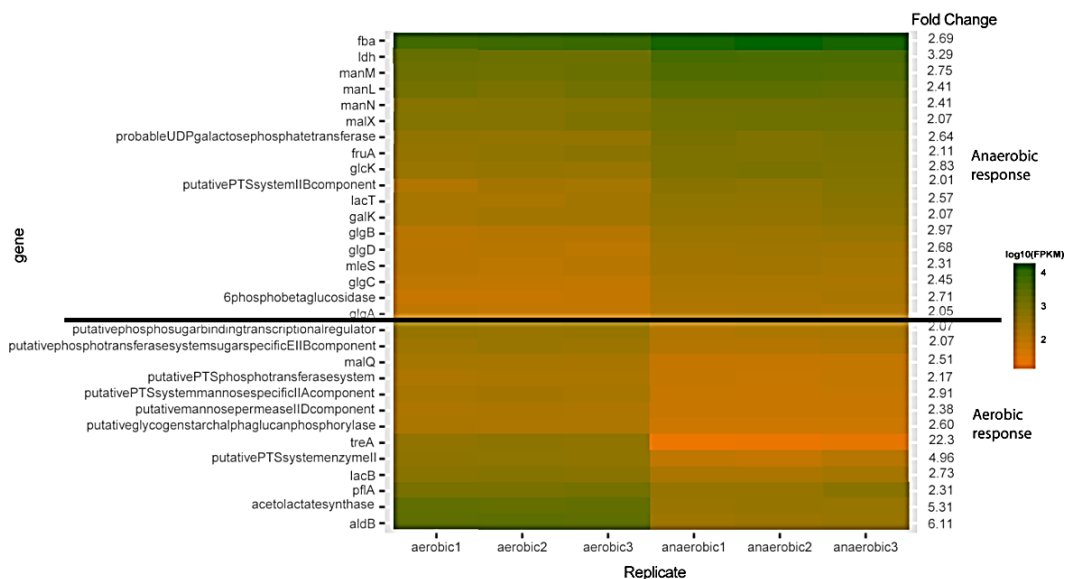


Figure S4 Variation in expression of genes involved in carbohydrate metabolism in the presence/absence of oxygen. Genes with statistically significant fold changes above 2 were classified either as an “Anaerobic response” or an “Aerobic response”, based on the condition wherein they are upregulated. Data was FPKM values generated from Cufflinks (Trapnell *et al.* 2012)

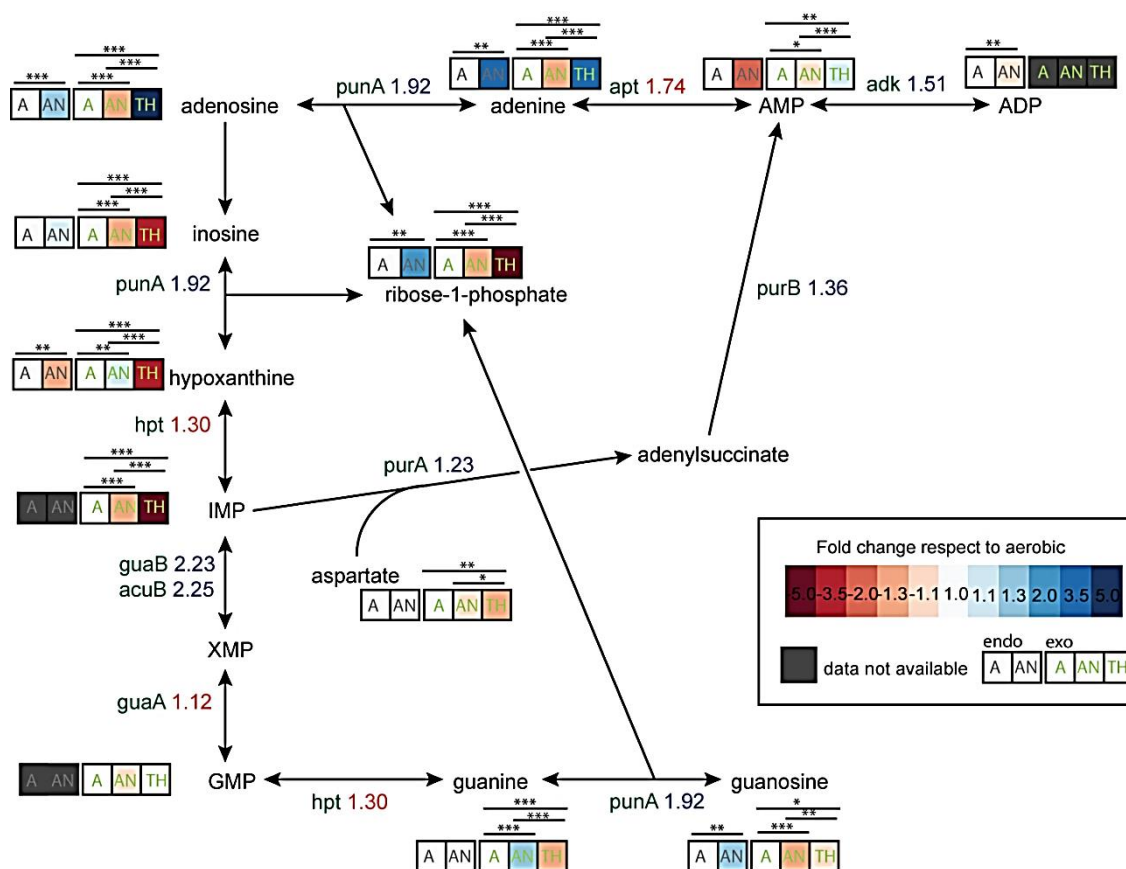


Figure S6 The purine metabolism pathway, constructed based on the BioCyc database for *S. intermedius* B196 and JTH08, as discussed in Fig. S5

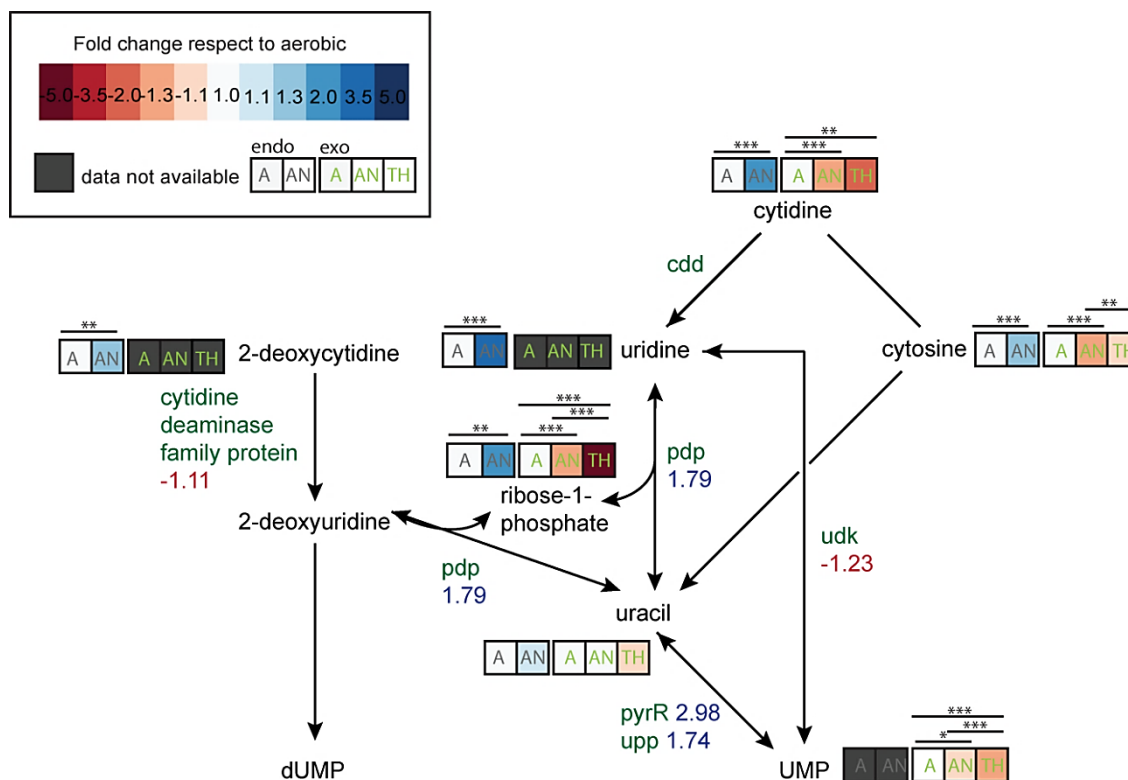


Figure S7 The pyrimidine salvage pathway, constructed based on the BioCyc database for *S. intermedius* B196 and JTH08, as discussed in Fig. S5

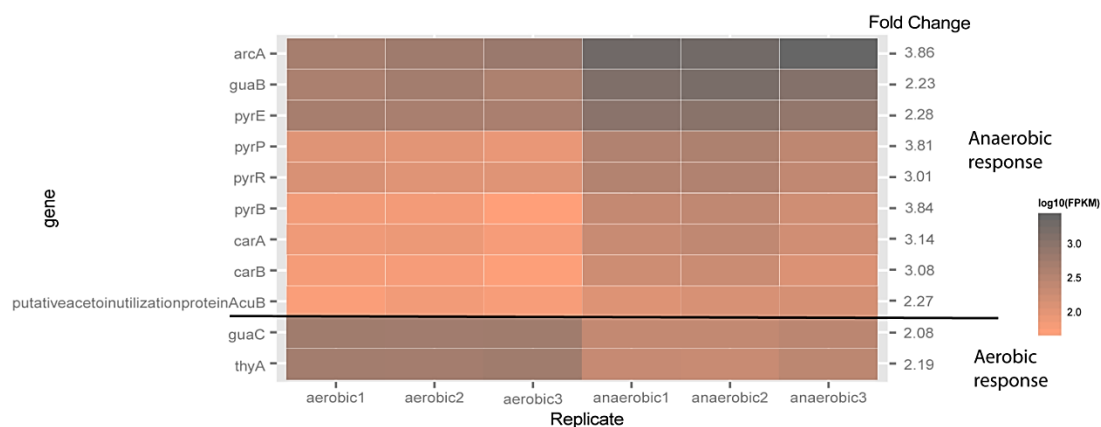


Figure S8 Genes involved in pyrimidine metabolism were affected by the presence/absence of oxygen. The response can be divided into “Anaerobic” or “Aerobic” based on the condition wherein they are up-regulated. Data was obtained using Cufflinks for RNAseq analysis (Trapnell *et al.* 2012).

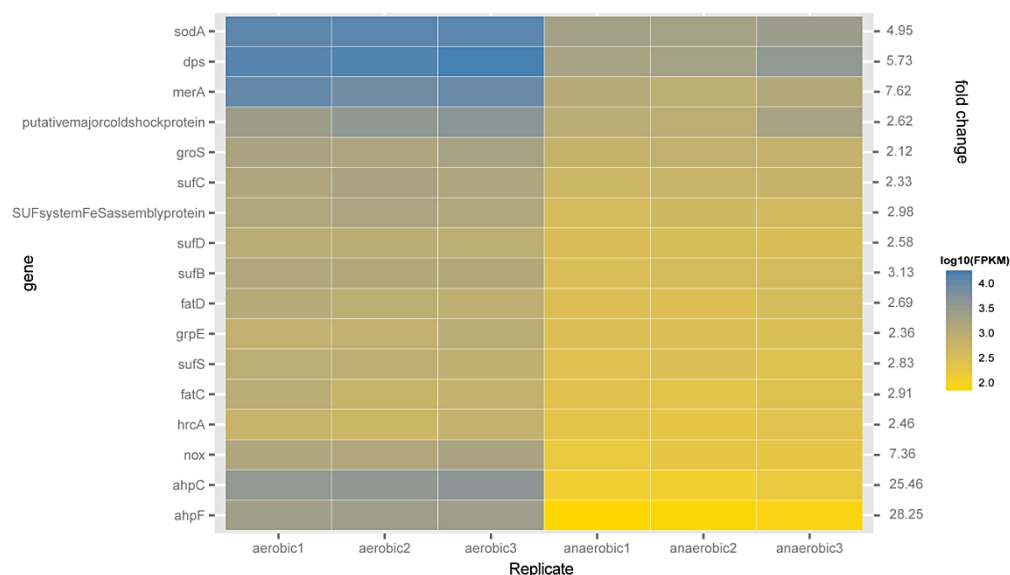


Figure S9 Induction of genes involved in oxidative stress under aerobic conditions. The heatmap includes genes that were induced aerobically above 2 fold. The up-regulated genes are involved in pathways including oxidative stress, iron metabolism and iron sulfur cluster assembly. The data is based on FPKM values generated from Cufflinks (Trapnell *et al.* 2012)

Table S1 Summary of metabolomic and transcriptomic data

	Intracellular metabolome		Extracellular metabolome			Transcriptome	
	aerobic	anaerobic	aerobic	anaerobic	THY	aerobic	anaerobic
Biological replicates	5	7	7	7	5	3	3
Percentage variance	22%	22%	14%	13%	14%	10.5%	13.7%
Number of metabolite/ gene features	1885		3382			1815	
Percentage of recovery ^a	81-105%		104-105%			----	
Identified features	124		116			----	
Identified metabolites	105		93			----	

^a determined based on tryptophan-d₅ in ESI positive mode

References

- Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*, 37(18), e123.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7(3), 562–578.

Chapter 4: Metabolic insights into the role of the multipartite genome of *Sinorhizobium meliloti* and its metabolic preferences in a nutritionally complex environment

Fan Fei^{1,2}, George C. diCenzo^{3*}, Dawn M. E. Bowdish², Brian E. McCarry¹, Turlough M. Finan^{3†}*

This chapter was rewritten based on the results published in: Fei, F., diCenzo, G. C., Bowdish, D. M. E., McCarry, B. E. & Finan, T. M. *Metabolomics*, It is reproduced here with permission.

Author Contributions

BEM and TMF conceived the study. FF, GCd, BEM, and TMF designed the study. GCd cultured strains. FF extracted cells, acquired and analyzed the metabolomic data, and generated the figures for the manuscript. **For the thesis chapter:** FF GCd and TMF interpreted the data. FF drafted the manuscript. FF and DMEB edited the manuscript.

Author Affiliations:

¹Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Canada L8S 4M1

²Department of Pathology and Molecular Medicine, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada L8N 3Z5

³Department of Biology, McMaster University, Hamilton, Canada L8S 4K1

*Both authors contributed equally

†Corresponding author: Email: finan@mcmaster.ca; Phone: (+1)905-525-9140 ext 22932; Fax: 905-522-6066

Abstract

Sinorhizobium meliloti is a ubiquitous nitrogen-fixing Gram negative soil bacterium that contributes to a symbiotic relationship with legume plants. The tripartite 6.7 Mb genome of *S. meliloti* consists of a 3.65 Mb chromosome, a 1.35 Mb pSymA megaplasmid and a 1.68 Mb chromosome-like pSymB chromid. Recent construction of mutant strains of *S. meliloti* lacking pSymA and/or pSymB provided the opportunity to study the global metabolic impact resulting from the loss of these replicons. Herein, the intra- and extracellular metabolomes of wild type, Δ pSymA, Δ pSymB and Δ pSymAB at various growth phases in multiple growth medium were measured using LC-HILIC-TOF-MS. There were 2008 extracellular metabolic features detected in cells cultured in rich LBmc media containing yeast extracts and protein hydrolysate. In contrast, 1474 intracellular metabolic features were detected in cells cultured in M9-sucrose minimal media. Metabolic differences were observed both across strains and across different growth phases. pSymA carries non-essential genes for cell survival, thus, relatively few metabolic deviations were observed between wild type and the Δ pSymA mutant. pSymB carries many transporter genes and the greatest metabolic differences were observed between wild type and strains lacking pSymB. However, Δ pSymB and Δ pSymAB had similar metabolic profiles. The inability to import, export, and utilize many essential metabolites due to lack of pSymB was the main contributor to the metabolic differences. Compared to wild type *S. meliloti*, increased intracellular sugars, amino acids, and nucleosides levels were observed in Δ pSymB and Δ pSymAB, and a loss of pSymB also impaired *S. meliloti*'s ability to catabolize exogenous amino acids. Although *S. meliloti* wild type, Δ pSymA, Δ pSymB, and Δ pSymAB were able to grow in LBmc and M9-sucrose media, metabolomics provides insights into the metabolic roles of the pSymA and pSymB replicons. Compared to pSymA, the pSymB replicon had a vital role in regulating *S. meliloti* metabolism, and this is consistent with the greater integration of this replicon within the *S. meliloti* genome.

Introduction

Historically, bacteria were thought to have their genome present in one circular chromosome. Nevertheless, some proteobacteria contain a multipartite genome where the genome is divided between multiple circular chromosomes known as replicons (Jumas-bilak, Michaux-charachon, Bourg, Ramuz, & Allardet-servent 1998). These proteobacteria include many pathogens and symbionts including *Agrobacterium*, *Vibrio*, *Burkholderia*, *Brucella*, *Rhizobium* and *Sinorhizobium* (Jumas-bilak *et al.* 1998). The metabolic and functional significance of these replicons and their contribution to the overall growth of the bacteria are almost unknown.

The multipartite genome of the Gram-negative α -proteobacterium *Sinorhizobium meliloti* was sequenced in 2001 (Galibert *et al.* 2001). *S. meliloti* serves as a N₂-fixing bacterium with leguminous plants and is used as a model organism for studying N₂-fixing symbiosis, carbon metabolism and the evolution and function of the multipartite genome (DiCenzo & Finan 2015; DiCenzo, MacLean, Milunovic, Golding, & Finan 2014; Galardini, Pini, Bazzicalupo, Biondi, & Mengoni 2013; Geddes & Oresnik 2012). Its 6.7 megabase (Mb) genome is divided into a 3.65 Mb chromosome, a 1.35 Mb pSymA megaplasmid and a 1.68 pSymB chromid (Barnett *et al.* 2001; Capela *et al.* 2001; Finan *et al.* 2001; Galibert *et al.* 2001). The sequence of pSymA is highly variable between wild-type *S. meliloti* isolates (Guo, Sun, Eardly, Finan, & Xu 2009) and is thought to be an accessory replicon mainly involved in symbiosis and adaptation (Barnett *et al.* 2001; DiCenzo *et al.* 2014; Galardini *et al.* 2013). pSymA is largely silent in free-living wildtype *S. meliloti* (Chen *et al.* 2000). Conversely, pSymB is much more ancient and conserved compared to pSymA (Galardini *et al.* 2013; Guo *et al.* 2009). pSymB is thought to have co-evolved with the chromosome (DiCenzo *et al.* 2014) and contains only copies of essential *tRNA^{arg}* and *minCDE* genes (Finan *et al.* 2001). It is also predicted to be involved in polysaccharide biosynthesis and metabolite/substrate transport via the ABC transporter system. Recently, a *S. meliloti* Rm2011 derivative lacking pSymB and/or pSymB replicons was constructed. These cured-strains provided a unique opportunity to study the role of pSymA and pSymB in *S. meliloti* metabolism.

The metabolic capacity of free-living *S. meliloti* has been studied using metabolomic approaches to examine the metabolic changes in respect to various carbon sources, growth rate and amino acid auxotrophy (Barsch, Patschkowski, & Niehaus 2004; Keum, Seo, Li, & Kim 2008; Ong & Lin 2003). Moreover, a metabolomic approach has also been used to examine the symbiosis between *S. meliloti* and legume plants both *in vitro* and *in silico* (Barsch, Tellström, Patschkowski, Küster, & Niehaus 2006; Gemperline, Jayaraman, Maeda, Ané, & Li 2015; Ye *et al.* 2013; Zhao, Li, Fang, Chen, & Wang 2012). However, the metabolic contributions of pSymA and pSymB replicons are unknown.

Here, we examined the intracellular and extracellular metabolic profiles of *S. meliloti* using HILIC-TOF-MS. This metabolomic strategy was successfully used to compare the metabolic capacity of wildtype, Δ pSymA, Δ pSymB and Δ pSymAB *S.*

meliloti and their substrate preference when cultured in a nutritionally complex environment. This study provided insights in understanding the metabolic contribution of the pSymA and pSymB replicons in *S. meliloti* metabolism.

Materials and Methods

Chemicals

HPLC-grade methanol, acetonitrile, chloroform, water and formic acid used for bacterial extraction and LCMS analysis were purchased from Caledon Laboratories (Georgetown, ON, Canada). Ammonium acetate was purchased from Fisher Scientific Company (Fairlawn, NJ, USA). L-methionine- d_3 (98%), L-tryptophan- d_5 (98%) as well as L-phenylalanine- d_8 (98%), diphenylalanine (phe-phe), glycine-phenylalanine (gly-phe) were purchased from Cambridge Isotope Laboratories (Andover, MA, USA) for recovery determination (RS) and peak intensity normalization (IS), respectively.

Strains, media and growth conditions

The four strains of *S. meliloti* used in this study were previously described (DiCenzo *et al.* 2014; Oresnik, Liu, Yost, & Hynes 2000), which included wild type Rm2011 (SU47 *str*-3), SmA818 (Δ pSymA), RmP3009 (Δ pSymB) and RmP2917 (Δ pSymAB). Δ pSymA, Δ pSymB and Δ pSymAB were derived from Rm2011 by the removal of pSymA, pSymB or both megaplasmids, respectively. Essential genes, *tRNA^{arg}* and *engA*, in pSymB were integrated into the chromosome.

LBmc (per liter: 10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl, 2.5 mM $MgSO_4$, 2.5 mM $CaCl_2$, 2 μ M $CoCl_2$) was used as the complex medium, and M9-sucrose medium (41 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 18.7 mM NH_4Cl , 10 mM sucrose, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.25 mM $CaCl_2$, 38 μ M $FeCl_3$, 5 μ M thiamine-HCl, 4.1 μ M biotin and 42 nM $CoCl_2$) was used as a minimal medium.

All four strains of *S. meliloti* were cultured to early stationary phase in 5 mL M9/LBmc at 30°C. At 3.0 OD_{600} , the bacterial culture was washed once by centrifugation with fresh media (M9 or LBmc as appropriate) and re-suspended in fresh media to a final OD_{600} of 0.05. These cultures were aliquoted into seven test tubes with 5 mL each and incubated at 30°C. Cells and supernatants from six of the test tube were collected for metabolomic analyses and the growth curve was measured using the seventh test tube.

Sample collection and extraction

For intracellular metabolic analysis, cells from M9-sucrose culture were collected in sextuplicate at approximately 0.3, 0.5, 0.9, 1.6 and 3.5 (except for Rm2011 at 4.8) OD_{600} . These samples were referred to as M1-5 with M1-4 referring to different stages of exponential growth phases and M5 representing the stationary phase (Fig. 1). Aliquots of 1000, 600, 300, 176, 100 μ L were taken at each time point, respectively, from the same 5 mL culture, so that a constant of 6×10^8 cells were collected in each sample. The cells

were centrifugated at 4°C, and washed once with saline (0.85% NaCl). The pellet was re-suspended and extracted using 2:2:1 MeOH/EtOH/H₂O as per a previously published protocol (Fei, Bowdish, & McCarry 2014) and stored at -80°C until LCMS analyses were conducted.

For extracellular metabolic analysis, 200 µL of the LBmc culture were collected in sextuplicate at mid-exponential phase (L1), early stationary phase (L2), and late stationary phase (L3) from the sample 5 mL culture (Fig. 1). L3 was taken 10 hours after L2. Cells were centrifuged at 4°C and 20 µL of the media supernatant were collected and extracted with 80 µL MeOH/EtOH (1:1) containing RS to remove precipitated protein. These extraction mixtures were vortexed for 2 min and centrifuged at 9500 x g for 3 min. The extracted supernatants were collected, diluted 2-fold with 100 µL MeOH and stored in -80°C until LCMS analyses were conducted.

Two sets of pooled samples were prepared for both the intracellular and extracellular metabolomic studies by combining 5 µL of all corresponding samples.

HILIC-TOF-MS analyses

The intracellular and extracellular extracts were analyzed in two separate batches using an Agilent Technologies 1200 RR Series II liquid chromatograph (LC) coupled to a Bruker MicrOTOF II Mass Spectrometer (MS) (Fei *et al.* 2014). An injection of 2 µL was separated on a 50 mm × 2.1 mm Kinetic 2.6 µm HILIC column of pore size of 100 Å (Phenomenex, CA, USA). The mobile phases were HPLC-grade acetonitrile (A) and 10 mM ammonium acetate in HPLC-grade water adjusted to pH 3 with formic acid (B) at a flow rate of 200 µL/min. The column temperature was maintained at 40 °C, and the autosampler storage tray was at 4°C. The mobile phase gradient eluted isocratically with 95% ACN for 0.5 min followed by a gradient to 35% ACN over 12 min. The gradient system maintained at 35% ACN for 0.5 min and returned to 95% ACN over 1 min. The gradient was then followed by a 10 min re-equilibration phase prior to the next injection. The total time for the HILIC gradient was 24 min for both ESI+ and ESI- modes. The positive ionization mode and the negative ionization mode were acquired separately. The MS setting was identical to those previous reported in Fei *et al.* (Fei *et al.* 2014).

A pooled sample was injected 7 times at the beginning of the analyses to condition the column and it was also injected after every five samples. A methanol blank and a standard mixture containing all IS and RS were also injected after every 10 samples. A total of 82 extracellular extracts and 115 intracellular extracts were analyzed in random order in both ESI- and ESI+ modes.

Data processing and metabolite identification

The data processing and analysis were modified from a previously published protocol (Fei *et al.* 2014). The LC-MS data files were converted to .mzXML format using Bruker CompassXport after internal calibration using intracellular sodium formate cluster ions by Bruker's DataAnalysis 4.0 SP4. The metabolic features were extracted and

aligned using open source XCMS with centWave algorithm (Smith, Want, O'Maille, Abagyan, & Siuzdak 2006). Adducts, isotopic ions, and in-source fragments were identified using CAMERA (Kuhl, Tautenhahn, & Neumann 2010).

Metabolite features with apparent retention factors k_{app} lower than 0.7 were removed as well as isotopic ions, features corresponding to IS, RS and sodium formate clusters. For the extracellular metabolome, regions of retention 7.0-7.8 min were excluded from the data matrix due to ion suppression. The peak areas of all metabolite features were normalized to IS and OD₆₀₀ for intracellular metabolic analysis; the peak areas of all metabolic features were only normalized to IS for extracellular metabolic analysis. Features with greater than 20% variance in the pooled sample were removed to obtain the final metabolite feature list (Dunn *et al.* 2011).

Metabolite features were identified by matching the m/z and retention values to those of the available authentic standards or matches to the tandem MS (MS/MS) fragment pattern on the METLIN database. MS/MS was performed on a Thermo Scientific Dionex Ultimate 3000 rapid separation LC coupled to a Bruker maXis 4G QTOF MS using a modified LC-HILIC-MS method, and detailed procedures can be found in supplementary information. Lists of identified intracellular and extracellular metabolites and their relative abundances between strains and across growth periods can be found in the Supplementary material ESM 1 and 2.

Statistical analyses

Both intracellular and extracellular metabolic data were analyzed using SIMCA-P+ 11 software (Umetrics, Kinnelon, NJ). Pareto scaling was applied prior to principal component analysis (PCA) and to orthogonal partial least-squares discriminative analysis (OPLS-DA). OPLS-DA was used to differentiate metabolite profiles between different strains and growth phases. The model validation parameters R^2X , R^2Y , and Q^2 were used to assess the fitness of the model. R^2X (R^2Y) indicated the fraction in which metabolite features (X) and treatment (Y) matrix was explained by the model. Briefly, a prediction statistic (Q^2) above 0.4 were indicative of a robust model, and Q^2 between 0.7-1.0 indicated the model was highly robust (Jones, Spurgeon, Svendsen, & Griffin 2008). Both R^2 and Q^2 followed an upward trend from 0 to 1. For an over fit model, R^2 approach 1, and Q^2 fell toward 0. Therefore, a valid and robust OPLS-DA model should have R^2X and R^2Y approaching 1 and Q^2 greater than 0.4.

Between subjects, two-way ANOVA and hierarchical cluster analysis (HCA) were computed and plotted using MetaboAnalyst 3.0 (Xia, Sinelnikov, Han, & Wishart 2015). Heat maps were plotted using R 2.12.2 and RStudio 0.98.501. HCA and heatmap were plotted based on Euclidean distances and complete clustering. Univariate analyses including two-tailed, unpaired heteroscedastic Student's *t* tests and non-parametric ANOVA with $p < 0.01$ with a Bonferroni correction were used to identify metabolite features that were significantly different between strains and growth phases.

Results

Growth profiles

In order to understand the contributions of pSymA and pSymB megaplasmid to *S. meliloti* metabolism throughout growth, we sampled wild type, Δ pSymA, Δ SymB and Δ pSymAB strains at five time points in M9-sucrose medium and three time points in LBmc medium (Fig. 1). The growth profiles of four strains cultured in M9-sucrose and LBmc media were consistent with previous work (DiCenzo *et al.* 2014). The growth kinetics of wild type and Δ pSymA strains were comparable in both M9-sucrose and LBmc media. Δ SymB had slower growth kinetics as compared to the wild type strain in both culture conditions, and Δ SymAB was even slower compared to Δ SymB. All four strains were able to reach 3.0-4.0 OD₆₀₀ at stationary phase when cultured in minimal M9-sucrose medium. However, when cultured in rich LBmc, the OD₆₀₀ at stationary phase was much different between strains. The final OD₆₀₀ was 6.6 for both wild type and Δ pSymA compared to 1.0-1.5 for Δ SymB and Δ pSymAB. Loss of pSymA has no apparent changes to *S. meliloti* growth. The loss pSymB led to slower growth and an approximately 6-fold drop in maximum cell concentration at stationary phase in LBmc. Overall, the pSymB chromid contributes to the growth of *S. meliloti* in either LBmc or M9 medium.

Metabolite analyses, features detection and quality control

The intracellular and extracellular metabolic profiles of all four strains at various growth phases were acquired using LC-HILIC-TOF-MS. A total of 3594 intracellular and 4081 extracellular metabolite features were identified using XCMS and CAMERA. Each feature had a unique retention time and m/z value. After data processing and reduction, there were 1474 intracellular and 2008 extracellular metabolite features in the final data set (Table S1), among which, 1237 and 820, respectively, were statistically significant between samples. Using authentic standards and MS/MS, 142 intracellular features were identified of which 49 were phospholipids and 73 were polar metabolites; 77 extracellular features were identified of which 2 were phospholipids and 66 were polar metabolites. These metabolic features were listed in the supplementary material 1 and 2.

The extraction efficiency was 73-85% and the biological variance of the sextuplicate samples was 18-29% for the intracellular metabolome and 12-26% for the extracellular metabolome (Table S2). OPLS-DA was performed to assess the variation of the pooled samples and instrumentation reproducibility. The pooled samples from intracellular or extracellular samples were tightly clustered on the score plots, thus indicating that the technical variability was minimal in comparison to the biological variance of experimental condition (Fig.S1, S2).

Intracellular metabolic profiles of *S. meliloti* strains cultured in minimal M9-sucrose medium

To understand the contribution of pSymA and pSymB to *S. meliloti* metabolism, the intracellular metabolomes of wild type, Δ pSymA, Δ pSymB and Δ pSymAB mutants, which were cultured in minimal M9-sucrose medium were measured at five different time points during growth. The global relationship between the four strains was investigated using multivariate OPLS-DA and HCA. This revealed significant metabolic differences between strains and across growth (Fig. 2, Table S3). The pSymA megaplasmid had little influence on *in vitro* *S. meliloti* metabolism. Similar metabolic profiles were observed for wild type and Δ pSymA during growth. Conversely, the pSymB chromid had a significant impact on *S. meliloti* metabolism. Compared to wild type, large metabolic disparities compared to wild type were observed in strains lacking pSymB (Δ pSymB and Δ pSymAB) (Fig. 2a). A progressive metabolic change was observed for each strain across each stages of growth (M1 to M5). Consistent with previous metabolic studies of yeast growth (Allen *et al.* 2003), the largest metabolic disparities were observed between M4 and M5 as wild type, Δ pSymA and Δ pSymB strains entered the stationary phase. However, the greatest metabolic difference for Δ pSymAB was observed between M1 and M2 at early exponential phase. Overall, the metabolic alteration resulting from growth was minor compared to the metabolic alteration resulting from the loss of pSymA or pSymB megaplasmids.

Metabolites contributing to the changes in intracellular metabolic profiles of *S. meliloti* strains

Student's t test and between subjects two-way ANOVA were performed to identify metabolic features or metabolites that differed between strains and between growth phases. The heat map consisting of the top 100 significantly changed metabolic features as ranked by ANOVA p values are illustrated in Figure S5. Although the majority of the features remaining unidentified, galactose, sucrose, glutamic acid, glutamyl-hydroxyproline and PG35:2 were among the top 100 significantly changed metabolites.

HCA and the heat map of 1237 significant metabolic features with a p value less than 0.01 (two-way ANOVA) are illustrated in Figure 2b-d. Three notable groups of metabolites were found to be significantly altered between strains and growth. Sugars such as sucrose, galactose, maltose, glucose and trehalose were higher levels in the Δ pSymB and Δ pSymAB strains during growth as compared to wild type and Δ pSymA. Except for trehalose, the levels of these sugar metabolites progressively decreased as growth progressed. Increasing cellular abundances of nucleosides including adenine, adenosine, inosine, 2-deoxyadenosine, guanine and deoxyinosine were also observed in Δ pSymB and Δ pSymAB strains. The purine degradation product, hypoxanthine, was progressively increased throughout growth in all four strains, however, its cellular level was much higher in strains lacking pSymB during growth phase M1-4 as compared to wild type and Δ pSymA at similar stages of growth. Thirdly, amino acids including tyrosine, arginine, lysine, leucine and glutamine were elevated in pSymB deficient strains, with the exception of serine and glutamine, which were increased in pSymA deficient

strains. Greater changes in the intracellular levels of polar metabolites were observed with the absence of pSymB chromid.

There were 49 phospholipids identified in the intracellular metabolome, among which were 12 phosphatidylcholines (PC), 13 phosphatidylglycerols (PG), 21 phosphatidylethanolamines (PE) and 2 phosphatidic acids (PA). From the list of lipids, there were 17 odd-chain phospholipids (cyclopropane-containing lipids) and 31 even-chain phospholipids. Δ pSymB and Δ pSymAB had greater levels of odd-chain PCs, PGs, PE33:0 and PE33:1 and even-short chain PCs including PC32:1 and PC34:1 compared to wildtype and Δ pSymA (Fig. 2d). PG34:0, PG34:1, PG36:0 and PG36:2 were more abundant in pSymA-cured strains. Levels of other phospholipids were not significantly different between strains and during growth. The relative abundances of the above mentioned metabolites can be found in electronic supplementary material ESM1.

Extracellular metabolic profiles of *S. meliloti* strains cultured in complex LBmc medium

The contribution of pSymA and pSymB megaplasms towards the utilization of substrates in a nutrient rich environment was examined by analyzing the metabolic composition of LBmc medium in the presence of each of the four *S. meliloti* strains at different growth phases. The blank LBmc, spent LBmc at mid-exponential phase, early and late stationary phase were analyzed, and the differences in LBmc substrate profiles after the growth of each strain were visualized using OPLS-DA, HCA and heat mapping (Fig. 4a-c). Incubating LBmc at 30°C for 45 hrs did not degrade or modify the substrate profiles of the LBmc medium, therefore, any changes in LBmc substrate level were result from the growth of *S. meliloti* (Figure 4c).

The substrate profiles of spent LBmc medium at mid-exponential phase for all four strains resemble that of the un-inoculated medium and major changes in the substrate profiles were observed as cells entered into stationary phase. The overall changes in the substrate profiles between wild type and Δ pSymA were similar during growth and this trend was distinctly different from the substrate profiles of Δ pSymB and Δ pSymAB. The changes in the LBmc substrate profiles during growth were less extensive in the pSymB-cured strains. Consistent with the differences in the intracellular metabolome between strains, pSymB had a greater role in modifying the LBmc substrate profiles as opposed to the pSymA.

Metabolites contributing to the changes in extracellular metabolic profiles of *S. meliloti* strains

Close to one half of the detected extracellular metabolic features (41%) were significantly different between the four *S. meliloti* strains and the un-inoculated LBmc blank (Fig. 4c). The top 100 significantly changed metabolic features were visualized in a heat map (Fig. S7). Amino acids and derivatives (ala-leu, met-ala, arginine, lysine and ornithine), nucleosides (adenine, adenosine, guanine and guanosine), maltose and 2-phenylglycine were among these top 100 features. The majority of the significantly

differentiated metabolic features were progressively depleted from the LBmc medium during the growth of wild type and Δ pSymA strains (Fig. 5). To a lesser extent, a similar trend was observed for strains lacking pSymB. Dipeptides such as met-ala, ala-his, ala-leu and nucleosides including guanosine, inosine, adenosine, adenine were found to decrease progressively from LBmc at different rates for wild type, Δ pSymA, Δ pSymB and Δ pSymAB strains. However, a gradual depletion of amino acids including arginine, ornithine, proline, glutamic acid, histidine and serine from LBmc during growth was only observed for wild type and Δ pSymA strains. No metabolic features were reduced during growth only in pSymA- and/or pSymB-cured strains. Conversely, increasing levels of some metabolic features in LBmc was noted during the growth of *S. meliloti* strains and the extent of these changes were significantly reduced with the removal of pSymA and/or pSymB from the genome. N-acetyl-glutamic acid, pantothenic acid and nicotinic acid were found elevated for all four strains. However, N-acetyl-phenylalanine, sedoheptulosan, N-acetyl-glucosamine (GlcNAc), and citrulline were only found to increase extracellularly for wild type and Δ pSymA strains. Guanine, allantoin, arabitol and malic acid accumulated in the spent media of Δ pSymB and Δ pSymAB strains, but were gradually depleted from spent media of the wild type and Δ pSymA. The relative abundances of above mentioned metabolites can be found in electronic supplementary material ESM2.

Discussion

Sinorhizobium meliloti is a Gram positive α -proteobacterium. It is ubiquitously found in soil environment as a free-living microorganism or a nitrogen-fixing symbiont within legume (e.g. alfalfa) root nodules. Its 6.7 Mb genome is one of the largest bacterial genomes consisting of a 3.65 Mb chromosome, a 1.35 Mb pSymA megaplasmid and a 1.68 pSymB chromid (Galibert *et al.* 2001). Over half of the protein-coding genes (3341/6204) in *S. meliloti* are located on the chromosome; pSymA and pSymB contains 1293 (20.8%) and 1570 (25.3%) protein-coding genes, respectively. Large fractions of genes on pSymA are involved in nitrogen fixation and nodulation (Barnett *et al.* 2001; Galibert *et al.* 2001). Seventeen percent of the genes on pSymB are ABC transport system and almost all are predicted to be involved in small molecule import while 12% of the pSymB genes are involved in polysaccharide biosynthesis (Capela *et al.* 2001; Finan *et al.* 2001). Both pSymA and pSymB included a significant number of functionally redundant genes, and 10-15% of the constructed chromosomal mutants were found to have duplicated genes located on megaplasmids (DiCenzo & Finan 2015). Despite the loss of large bulk of its genome, *S. meliloti* strains lacking pSymA and/or pSymB can be cultured in both rich and minimal media. This study examined the metabolic contributions of pSymA and pSymB and how the lack of these replicons influenced the intracellular and extracellular metabolome of *S. meliloti*.

In order to study the metabolic capacities of pSymA and pSymB, *S. meliloti* strains lacking pSymA and/or pSymB were grown in M9 or LBmc media. M9 medium consisted of only essential nutrients required to support the growth of *S. meliloti*, and 10 mM sucrose, which was the only carbon source in the medium. The minimal medium had

closer resemblance to the nutritional condition of bulk soil compared to LBmc (DiCenzo *et al.* 2014), and is thought to be better mimic of *S. meliloti* growth in its natural environment. By ensuring all strains used the same compounds to support growth, the contributions of pSymA and pSymB to cellular metabolism can be deduced via the study of the intracellular metabolome. Conversely, LBmc is a complex medium containing carbohydrates, amino acids, peptides, nucleotides, vitamins, casein hydrolysate and yeast extract. By comparing the extracellular metabolome of *S. meliloti* strains, we can examine the kinetics of substrate utilization in the LBmc medium as well as the metabolites excreted from *S. meliloti*. The differences of intracellular and extracellular metabolome among *S. meliloti* strains provide insights in metabolic roles of pSymA and pSymB.

Between pSymA and pSymA, the removal of pSymB from the *S. meliloti* genome had a much greater impact on *S. meliloti* growth and metabolism. Both Δ pSymB and Δ pSymAB showed significantly reduced growth and a 6-fold decreased in cell concentration at stationary phase compared to wild type and Δ pSymA when cultured in rich LBmc medium. The reduced growth capacity of strains pSymB-cured strains in LBmc was due to nutritional starvation. The principle carbon sources in LBmc medium were catabolizable amino acids from oligopeptides of tryptone, and the total concentration of sugar was less than 0.1 mM (Sezonov, Joseleau-Petit, & D'Ari 2007). Thus the growth of *S. meliloti* in LBmc depended on the utilization of those amino acids and oligopeptides. Over one half of the ABC transport system genes are located on the pSymB megaplasmid, and almost one half of the ABC transporter systems were involved in transporting sugar, amino acids, peptides, oligopeptides, choline and taurine (Finan *et al.* 2001). A microarray study had shown pSymB-cured strains were unable to use over one half of the 73 substrates as carbon source (DiCenzo *et al.* 2014). By profiling the spent LBmc media, the rate of utilization of many major metabolites including dipeptides, amino acids, nucleosides and taurine were significantly reduced as compared to the wild type. Presumably this is a result of loss or the reduction of the ABC transporter systems that occurs due to the loss of pSymB. Serine, aspartate, tryptophan, glutamate, glycine, proline, threonine and alanine were preferentially used by wild type *E. coli* K12 growing in tryptone broth medium (Sezonov *et al.* 2007). Unlike wild type and Δ pSymA, the inability to utilize proline, glutamate, threonine/homoserine and serine in pSymB-cured strains might contribute to their impaired growth capacity. Moreover, since one of three dipeptidase and three of five dipeptide transport were encoded on pSymB, the depletion of dipeptide from LBmc was also impaired in pSymB-cured strains, which might limit the availability of catabolizable amino acids for Δ pSymB and Δ pSymAB, thus hindering their growth. Overall, the inability to import peptides, amino acids and nucleosides limited the growth capacity of pSymB cured strains in LBmc.

The inability to utilize fermentable sugar efficiently from the growth media could also limit the growth of pSymB-cured strains as was observed in both LBmc and M9 media. Almost all ABC transporter system for sugars (102/106) were encoded by pSymB megaplasmid according to NCBI genome database. Close to one half of the predicted sugar kinases (10/22) are encoded by pSymB (Finan *et al.* 2001). Therefore, it was not surprising to observe that sugars such as sorbitol, glucose and maltose were not utilized

by Δ pSymB until the late stationary phase. However, in wild type strains, these sugars were utilized from the LBmc medium at a much faster rate and almost all depleted at early stationary phase. Since the growth capacity of *S. meliloti* in M9 was dependent on the use of sucrose and other essential nutrients, we did not observe a dramatic decrease in cell concentration in strains lacking pSymB. The build-up of intracellular sugar metabolites such as glucose, galactose, sucrose, trehalose and maltose were observed in pSymB-cured strains in M9 which was consistent with the slower growth rate observed for those strains. Since pSymB also carries duplicated copies of genes in glycolysis (DiCenzo & Finan 2015), loss of pSymB might also lead to decreased sugar catabolism and eventually slower growth.

From the metabolic profiles of depleted LBmc media of wild type strain, we observed a preferential usage of many substrates. Although none of the substrates was depleted completely from LBmc at mid-exponential phase, a significant amount of the substrates were completely utilized at early or late stationary phase, and some substrates were progressively used during growth but was not depleted at late stationary phase. The depletion of those substrates at L2 limited the growth of *S. meliloti*, inducing a switch to stationary phase from active proliferation. The ability of microorganisms to metabolize substrates in nutritional complex environment is commonly examined using phenotype microarray, such as those offered by Biolog Inc. (Bochner, Gadzinski, & Panomitros 2001). Such technology was used to examine ability of various microorganisms to use various substrates as carbon, nitrogen, phosphorus or sulfur source in cellular metabolism. Here, metabolomic studies using LC-HILIC-TOF-MS offered insights into utilization kinetics of substrates from a nutritionally complex environment to complement the current approach.

The loss of pSymB chromid also had a great impact on cellular metabolism of *S. meliloti* according to the large changes observed in intracellular metabolic profiles of pSymB-cured strains cross growth. In addition to the accumulation of intracellular sugar metabolites as discussed earlier, nucleosides and hypoxanthine in pSymB-cured strains were also elevated in the cell. Hypoxanthine is an intermediate in purine degradation and the salvage pathway (Caspi *et al.* 2014), and intracellular accumulation of hypoxanthine was reported as *E. coli* entered into stationary phase as a result of rRNA degradation (Rinas, Hellmuth, Kang, Seeger, & Schlieker 1995). In the wild type strain, accumulation of intracellular hypoxanthine was observed close to stationary phase. However, elevated levels of intracellular hypoxanthine were measured in the pSymB-cured strain throughout M1 to M5. Seventeen genes involved in purine/pyrimidine salvage/catabolic pathway were encoded on the 17 kb region of pSymB, and those included xanthine dehydrogenases and uracil/xanthine permeases (Finan *et al.* 2001). Xanthine dehydrogenases are involved in hypoxanthine degradation to xanthine and urate, deletion of these genes in pSymB-cured strains led to a buildup of the intracellular hypoxanthine level observed in our study. Guanine deaminases are encoded by pSymB (Caspi *et al.* 2014) and are involved in converting guanine to xanthine. In strains lacking pSymB, an intracellular accumulation of guanine was observed instead of the gradual decrease of guanine in the wild type strain. The utilization of purines from LBmc medium was also

limited and slowed over growth in pSymB-cured strains. The disruption of the purine degradation/salvage pathway in pSymB-cured strains might result in purine (adenosine, guanosine, adenine, guanine, and inosine) buildup through growth, and could also contribute to the reduced growth kinetics of pSymB-cured strains in M9 and LBmc media.

The loss of pSymB also influenced the cellular membrane composition in pSymB-cured strains. The lipid composition in the wild type strain was consistent with previous gas chromatography MS studies (Basconcillo, Zaheer, Finan, & McCarry 2009a). At lag phase (M1), PE36:2 was the most abundant phospholipid, which accounted for 62% of PC and 34% of the total detectable PLs. PC37:2 was the most abundant cyclopropanated-PL and accounted for 22% of PC and 12% of total PLs. The cyclopropanated-PL consisted 12-27% of the total detectable PLs. Since cyclopropane bond is more stable than a double bond, increases in cyclopropanated-PL are often noted when microorganisms are under stress. Increased cyclopropanated-PL in *S. meliloti* has been correlated with acid stress and nutrient limitations (Basconcillo, Zaheer, Finan, & McCarry 2009b); this was also observed in *E. coli* during acid stress and temperature shock (Zhang & Rock 2008). Genes involved in the synthesis of *S. meliloti* surface polysaccharides such as exopolysaccharides (EPSs), succinoglycan (EPS I), galactoglucan (EPS II), lipopolysaccharides (LPSs) and capsular polysaccharides (CPSs) were mapped to the pSymB chromid (Finan *et al.* 2001). The EPSs are thought to be crucial in protecting *S. meliloti* against plant defense during nodulation (K. M. Jones, Kobayashi, Davies, Taga, & Walker 2007). The absence of EPSs in pSymB-cured strains may result in increased cellular stress. As expected, cyclopropanated phospholipids (i.e. PCs except PC33:2, PGs, PE33:1 and PE33:2) were elevated in pSymB-cured strains. *S. meliloti* modifies membrane construct to adept to cellular defects as a result of pSymB deletion.

Unlike pSymB, the pSymA megaplasmid contributed very little to the metabolism of *S. meliloti* when cultured in *in vitro*. The function of pSymA in *S. meliloti* is specific to symbiosis and involved in adaptation to unique environment (DiCenzo *et al.* 2014; Galardini *et al.* 2013). Only a few proteins were identified to be uniquely associated to the pSymA, and pSymA is largely silent in the free-living cell (Chen *et al.* 2000). The removal of 1.35 Mb pSymA appeared to offset the metabolic burden on *S. meliloti*, and the growth kinetics of Δ pSymA in both LBmc and M9 were slightly faster compared to wild type strains. Similar intracellular and extracellular metabolic changes occurred in for both Δ pSymA and wild type strains during growth in M9 and LBmc media. The rate of substrate utilization from LBmc was comparable between Δ pSymA and wild type.

The *S. meliloti* genome in Δ pSymAB was reduced by 45% through the removal of both the pSymA megaplasmid and the pSymB chromid. The growth and metabolism of Δ pSymAB resembled those of Δ pSymB. In free-living *S. meliloti*, the more evolutionarily conserved pSymB chromid had a greater influence on the overall growth and metabolism, and pSymA megaplasmid was mostly silent. This metabolomic study was consistent with gene sequencing data (Barnett *et al.* 2001; Finan *et al.* 2001; Galibert *et al.* 2001), and

provided greater insights into the metabolic capacity of pSymA megaplasmid and pSymB chromid.

Electronic Supplementary Materials:

Electronic Supplementary datasets of intracellular and extracellular metabolomes for this chapter, ESM 1-2, can be found online (doi:). The raw LC-MS data was uploaded on Metabolomic Workbench.

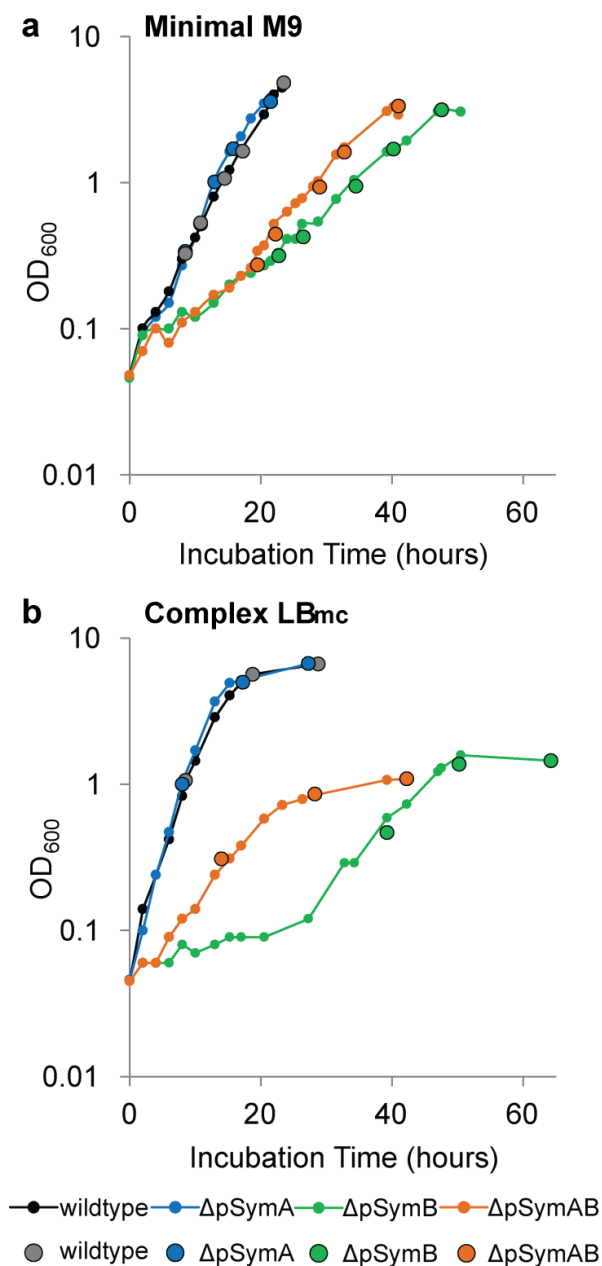


Figure 1 The growth curves of wild-type, ΔpSymA, ΔpSymB and ΔpSymAB *S. meliloti* strains. a, the growth of *S. meliloti* and its mutants were examined in M9 minimal medium and b, LBmc medium. The small solid circles represent OD₆₀₀ readings from a fixed sample of each strain throughout the entire growth curve. Samples were removed from culture tubes throughout growth and prepared for metabolomic analysis. The large solid circles with a black border indicate the average OD₆₀₀ readings of the quintuplicate or sextuplicate samples collected for intracellular and extracellular metabolomic analyses. Black—wild type; blue—ΔpSymA; green—ΔpSymB; orange—ΔpSymAB.

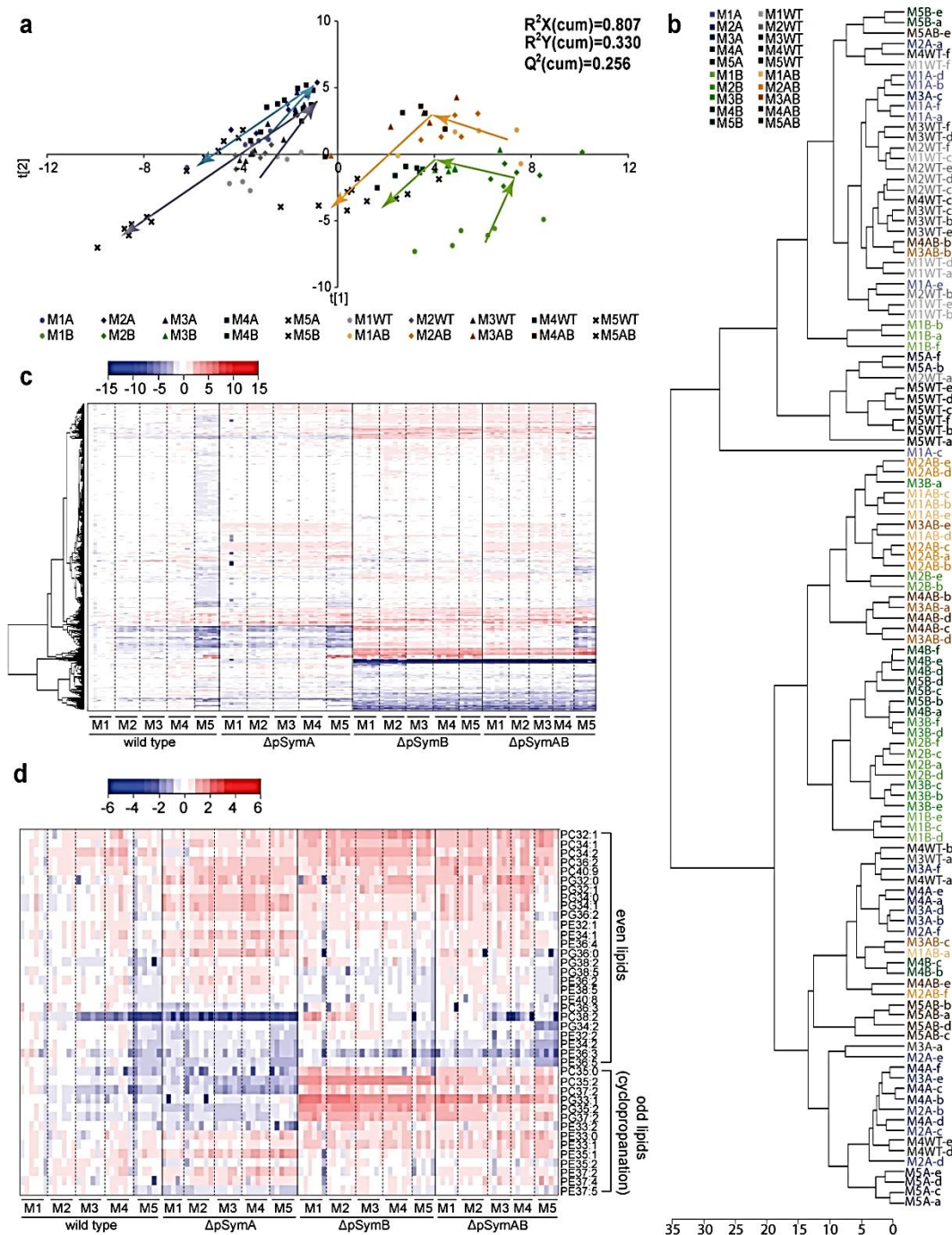


Figure 2 Statistical analyses of the intracellular metabolome of wild-type, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, and $\Delta p\text{SymAB}$ *S. meliloti* at various growth phases including exponential phase (M1-4) and stationary phase (M5). a, the OPLS-DA score plot (eight aligned and zero orthogonal components) showing the global relationship between the four *S. meliloti*

strains at various growth phases based on 1474 detected metabolite features. The progression of growth is indicated by the arrow direction. The predictive statistic for the model fell below 0.4 ($Q^2=0.256$) due to similar metabolic profiles between strains and growth phases. b, heatmap of 1237 significant metabolite features (between subject two-way ANOVA with Bonferroni correction, $p<0.01$) plotted based on $\log_2(\text{fold change})$ with respect to the average metabolite levels in M1WT. c, heatmap of 41 significant phospholipids (between subject two-way ANOVA with Bonferroni correction, $p<0.01$) based on $\log_2(\text{fold change})$ with respect to the average metabolite levels in M1WT. M1A sample c (M1A-c) was removed as an outlier. d, hierarchical clustering analysis of the total detectable features (1474 features) for all four strains at each growth stage. $\Delta p\text{SymAB}$. Black—wildtype; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$

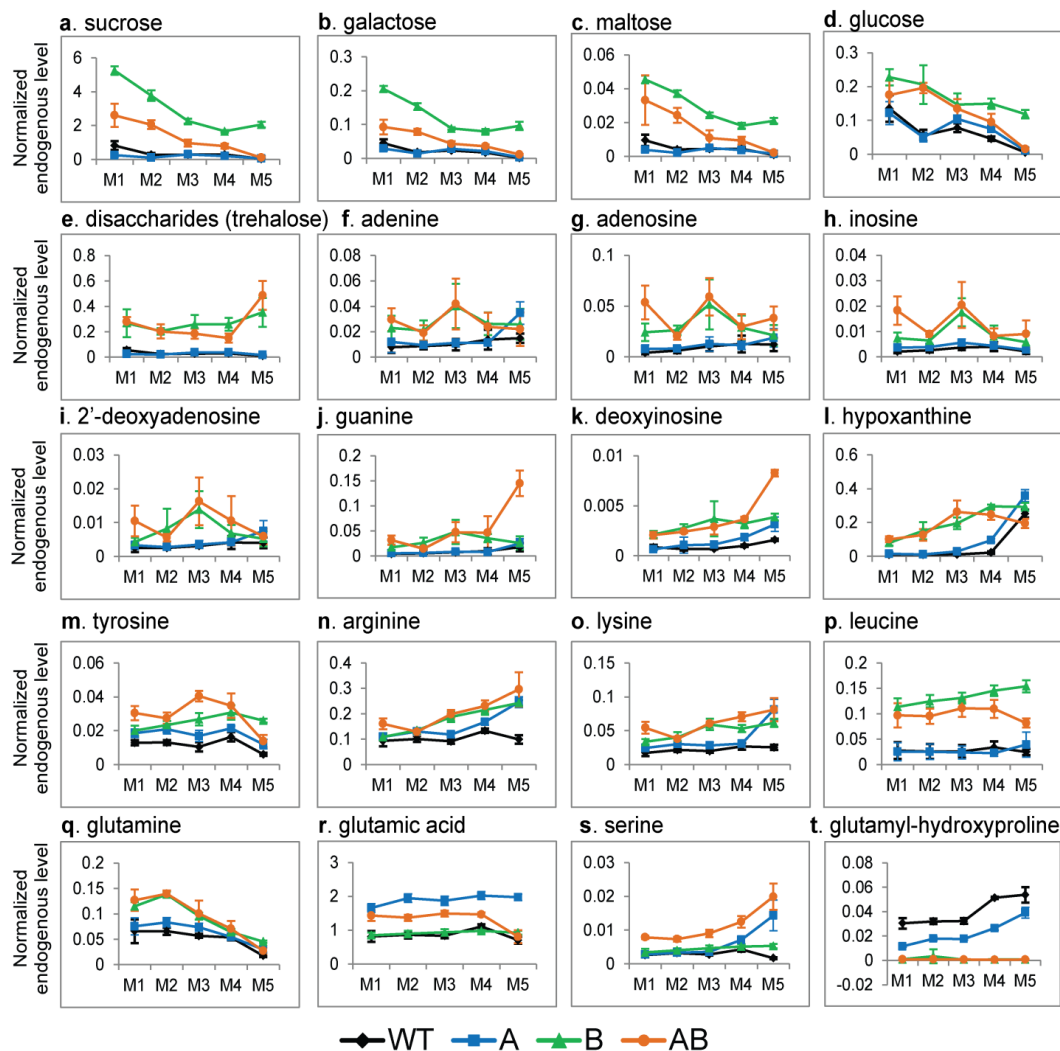


Figure 3 Intracellular levels of selected metabolites (a-t). Intracellular data were obtained from wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* strains cultured in minimal M9 medium at 5 time points. n=5-6. Error bar represents one standard deviation from the average. black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$. M1-4—exponential phase; M5—stationary phase

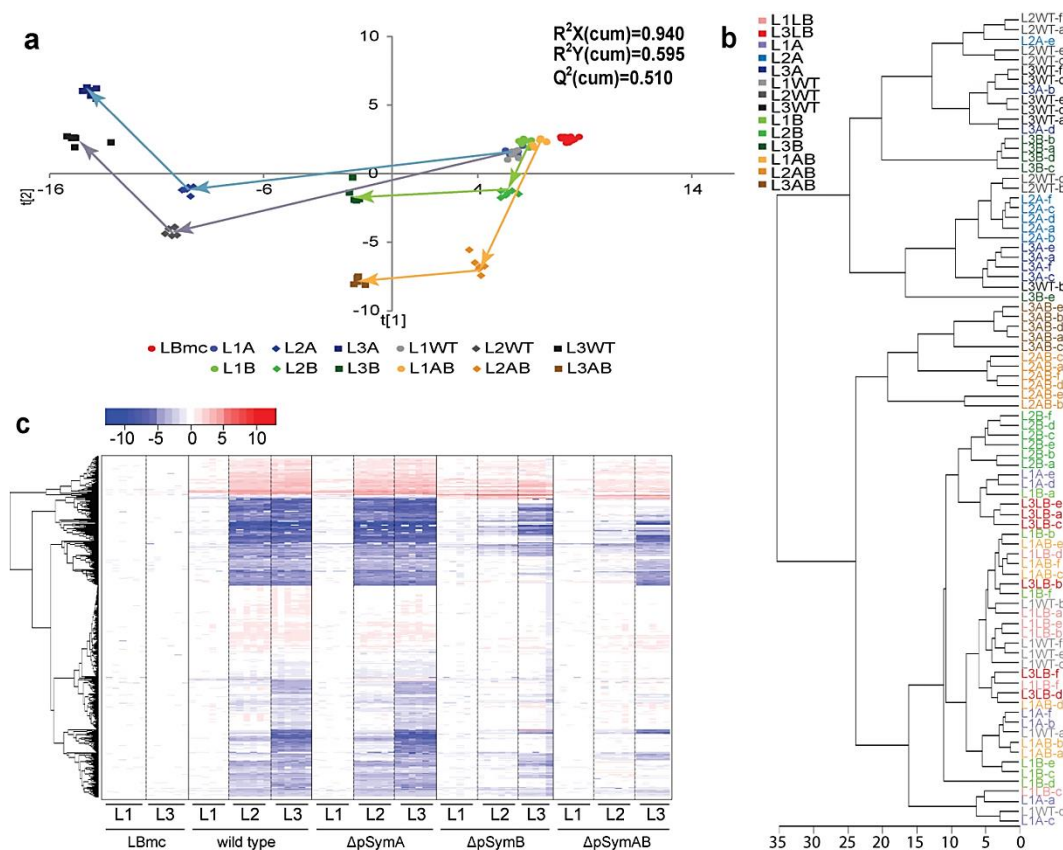


Figure 4 Statistical analyses of extracellular metabolome of wild-type, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, and $\Delta p\text{SymAB}$ *S. meliloti* at various growth phases including mid-exponential phase (L1), early stationary phase (L2) and late stationary phase (L3) cultured in LBmc. a, the OPLS-DA score plot (six aligned and eight orthogonal components) showing the global relationship between the four strains of *S. meliloti* at various growth phases based on 2008 detected metabolite features. The progression of growth is indicated by the arrow direction. b, heatmap of the 820 significant metabolite features (between subject two-way ANOVA with Bonferroni correction, $p < 0.01$) plotted based on $\log_2(\text{fold change})$ with respect to the average metabolite levels in LBmc. c, hierarchical clustering analysis of the total detectable features (2008 features) for all four strains at each growth stage. LBmc—red; black—wild type; blue— $\Delta p\text{SymA}$; green— $\Delta p\text{SymB}$; orange— $\Delta p\text{SymAB}$. L1—circle; L2—diamond; L3—square. WT—wild type; A— $\Delta p\text{SymA}$; B— $\Delta p\text{SymB}$; AB— $\Delta p\text{SymAB}$; LBmc—LBmc blank following 0 hr and 45 hr incubation at 30°C

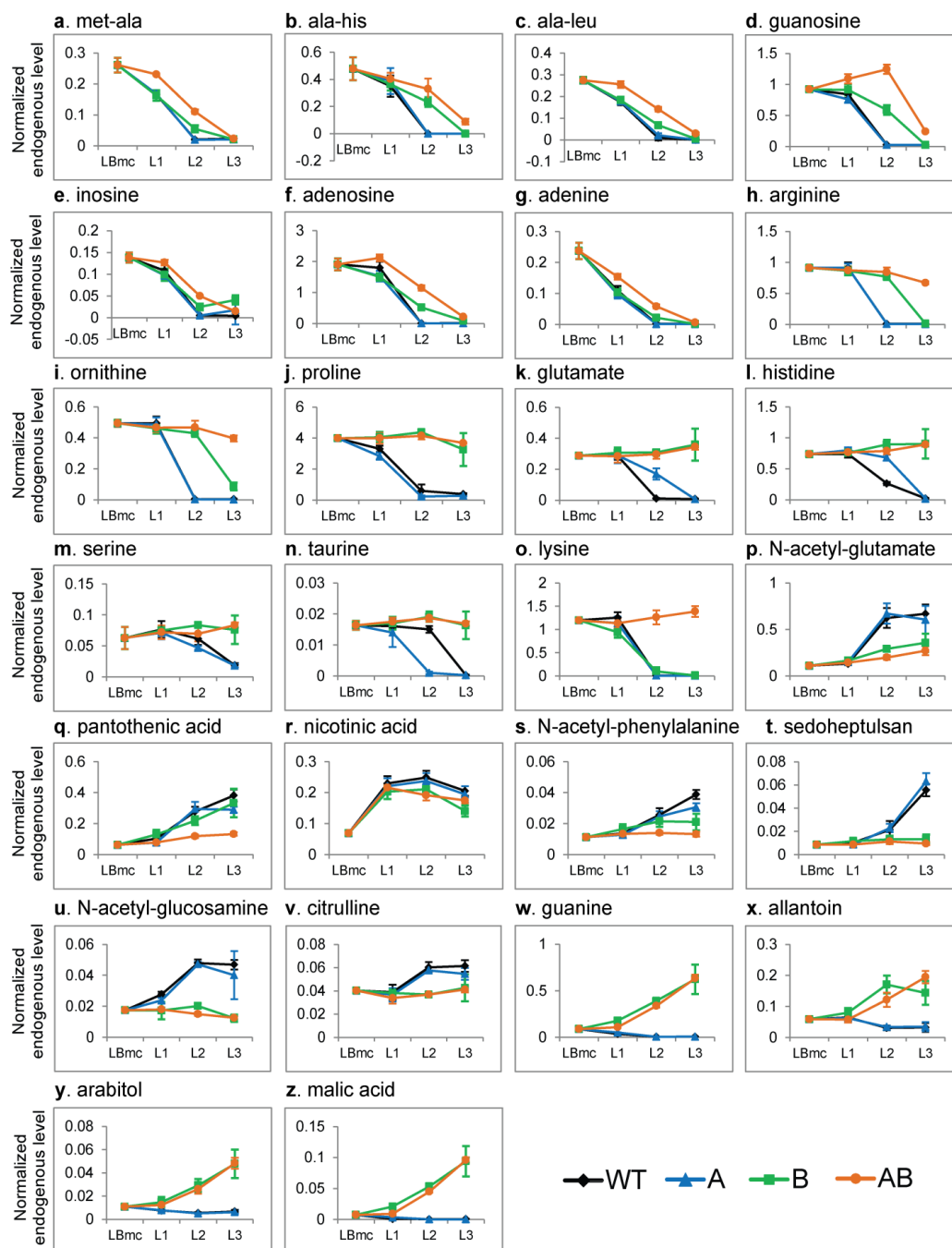


Figure 5 Extracellular data were obtained from various growth phases of wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* strains cultured in LBmc medium (a-z). Error bar represents one standard deviation from the average. black—wildtype; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$. L1—mid-exponential phase; L2—early stationary phase; L3—late stationary phase, LBmc—average of the readouts of 0 hr and 45 hr LBmc blank at 30°C.

References

- Allen, J., Davey, H. M., Broadhurst, D., Heald, J. K., Rowland, J. J., Oliver, S. G., & Kell, D. B. (2003). High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol*, 21(6), 692–696.
- Barnett, M. J., Fisher, R. F., Jones, T., Komp, C., Abola, A. P., Gurjal, M., ... Long, S. R. (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *PNAS*, 98(17), 9883–9888.
- Barsch, A., Patschkowski, T., & Niehaus, K. (2004). Comprehensive metabolite profiling of *Sinorhizobium meliloti* using gas chromatography-mass spectrometry. *Funct. Integr. Genomics*, 4(4), 219–230. doi:10.1007/s10142-004-0117-y
- Barsch, A., Tellström, V., Patschkowski, T., Küster, H., & Niehaus, K. (2006). Metabolite profiles of nodulated alfalfa plants indicate that distinct stages of nodule organogenesis are accompanied by global physiological adaptations. *MPMI*, 19(9), 998–1013.
- Basconcillo, L. S., Zaheer, R., Finan, T. M., & McCarry, B. E. (2009a). A shotgun lipidomics approach in *Sinorhizobium meliloti* as a tool in functional genomics. *J. Lipid Res.*, 50(6), 1120–1132.
- Basconcillo, L. S., Zaheer, R., Finan, T. M., & McCarry, B. E. (2009b). Cyclopropane fatty acyl synthase in *Sinorhizobium meliloti*. *Microbiology*, 155(Pt 2), 373–385.
- Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001). Phenotype MicroArrays for high-throughput phenotypic testing and assay of gene function. *Genome Res*, 11(7), 1246–1255.
- Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., ... Galibert, F. (2001). Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *PNAS*, 98(17), 9877–9882.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., ... Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, 42(D1), D459–D471.
- Chen, H., Higgins, J., Oresnik, I. J., Hynes, M. F., Natera, S., Djordjevic, M. A., ... Rolfe, B. G. (2000). Proteome analysis demonstrates complex replicon and luteolin interactions in pSymA-cured derivatives of *Sinorhizobium meliloti* strain 2011. *Electrophoresis*, 21, 3833–3842.

- DiCenzo, G. C., & Finan, T. M. (2015). Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome. *Mol. Genet. Genomics*, 290(4), 1345–1356.
- DiCenzo, G. C., MacLean, A. M., Milunovic, B., Golding, G. B., & Finan, T. M. (2014). Examination of prokaryotic multipartite genome evolution through experimental genome reduction. *PLoS Genet.*, 10(10), e1004742.
- Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-mcintyre, S., Anderson, N., ... The Human Serum Metabolome (HUSERMET) Consortium. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*, 6(7), 1060–1083.
- Fei, F., Bowdish, D. M. E., & McCarry, B. E. (2014). Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal. Bioanal. Chem.*, 406(15), 3723–3733.
- Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorho, F. J., ... Pühler, A. (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *PNAS*, 98(17), 9889–9894.
- Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E. G., & Mengoni, A. (2013). Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. *Genome Biol. Evol.*, 5(3), 542–558.
- Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., ... Boistard, P. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, 293(5530), 668–672.
- Geddes, B. A., & Oresnik, I. J. (2012). Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and L-arabitol in *Sinorhizobium meliloti*. *Microbiology*, 158(Pt 8), 2180–2191.
- Gemperline, E., Jayaraman, D., Maeda, J., Ané, J.-M., & Li, L. (2015). Multifaceted investigation of metabolites during nitrogen fixation in *Medicago* via high resolution MALDI-MS imaging and ESI-MS. *J. Am. Soc. Mass Spectrom.*, 26(1), 149–158.
- Guo, H., Sun, S., Eardly, B., Finan, T., & Xu, J. (2009). Genome variation in the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti*. *Genome*, 52(10), 862–875.
- Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., & Walker, G. C. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat. Rev. Microbiol.*, 5(8), 619–633.

- Jones, O. A. H., Spurgeon, D. J., Svendsen, C., & Griffin, J. L. (2008). A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm *Lumbricus rubellus*. *Chemosphere*, 71(3), 601–609.
- Jumas-bilak, E., Michaux-charachon, S., Bourg, G., Ramuz, M., & Allardet-servent, A. (1998). Unconventional genomic organization in the alpha subgroup of the proteobacteria. *J. Bacteriol.*, 180(10), 2749–2755.
- Keum, Y. S., Seo, J. S., Li, Q. X., & Kim, J. H. (2008). Comparative metabolomic analysis of *Sinorhizobium* sp. C4 during the degradation of phenanthrene. *Appl. Microbiol. Biotechnol.*, 80(5), 863–872.
- Kuhl, C., Tautenhahn, R., & Neumann, S. (2010). LC-MS Peak Annotation and Identification with CAMERA.
- Ong, L. C., & Lin, Y. (2003). Metabolite profiles and growth characteristics of *Rhizobium meliloti* cultivated at different specific growth rates. *Biotechnol. Prog.*, 19(3), 714–719.
- Oresnik, I. J., Liu, S.-L., Yost, C. K., & Hynes, M. F. (2000). Megaplasmid pRme2011a of *Sinorhizobium meliloti* ss not required for viability. *J. Bacteriol.*, 182(12), 3582–3586.
- Rinas, U., Hellmuth, K., Kang, R., Seeger, A., & Schlieker, H. (1995). Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.*, 61(12), 4147–4151.
- Sezonov, G., Joseleau-Petit, D., & D’Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.*, 189(23), 8746–8749.
- Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, 78(3), 779–787.
- Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.*, 43(W1), W251–W257.
- Ye, H., Gemperline, E., Venkateshwaran, M., Chen, R., Delaux, P.-M., Howes-Podoll, M., ... Li, L. (2013). MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the *Medicago truncatula*-*Sinorhizobium meliloti* symbiosis. *Plant J.*, 75(1), 130–145.
- Zhang, Y.-M., & Rock, C. O. (2008). Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.*, 6(3), 222–233.

Zhao, H., Li, M., Fang, K., Chen, W., & Wang, J. (2012). In silico insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PLoS One*, 7(2), e31287.

Supplementary information

HILIC-LC-MS/MS method

The HILIC-LC-MS/MS for unknown identification was a modified version of previously published method from HILIC-LC-MS used for metabolomic analyses in this study (Fei *et al.* 2014). The MS/MS spectra were acquired by analyzing extracts using an ThermoScientific Dionex Ultimate 3000 rapid separation LC coupled to Bruker maXis 4G QTOF MS. An injection of 2 μ L was separated using the same 50 mm $\times$ 2.1 mm Kinetex 2.6 μ m HILIC column of pore size of 100 Å (Phenomenex, CA, USA). Same mobile phases including HPLC grade acetonitrile (A) and 10 mM ammonium acetate in HPLC grade water adjusted to pH 3 with formic acid (B) were used at a flow rate of 250 μ L/min. The column temperature was maintained at 40 °C, and the auto sampler storage tray was set at 4°C. A linear LC gradient: 0-0.5 min, 95% A; 0.5-13.0 min, 40% A; 13.0-13.5 min, 40% A; 13.5-14.0 min, 95% A; 14.0-20.0 min, 95% A.

An identical ESI condition was used for both metabolomic and MS/MS methods: 3.0 bar nebulizer pressure; -150 V endplate offset; -3800 V or 4500 V capillary voltage; 6.0 L/min drying gas flow rate; 250°C dry gas temperature. The data were acquired in profile mode from 50 to 500 m/z at a scan rate of 1.0 Hz (computed using a rolling average value of 2). Auto MS/MS was set with target m/z ($\pm$ 0.20 m/z) and intensity threshold was set at 400 counts. The collision-induced dissociation (CID) energy was set to 20 or 30 V for optimum MS/MS spectra.

Table S1 Intracellular and extracellular metabolome data summary

No. features	Intracellular metabolome		Extracellular metabolome	
	ESI+	ESI-	ESI+	ESI-
Total (from XCMS)	1718	1876	2077	2004
Total	3594		4081	
k_{app}'<0.7	353	234	100	75
Isotopes	301	388	345	330
IS/RS, sodium formate	8	8	6	7
Ion suppression (7.0-7.8 min)	---	---	69	102
%CV>20% in pooled samples	828 ^a		1040	
Final data matrix	1474		2008	
Significant features^b	1237		820	

^a45 metabolic features that were identified as phospholipids had greater than 20% CV in pooled samples; however, they were not eliminated.

^bcomputed with between subjects two-way ANOVA with Bonferroni correction, p<0.01

Table S2 Biological variation of intracellular and extracellular metabolic profiles of the wild type, ΔpSymA, ΔpSymB, ΔpSymAB at various growth phases

	Intracellular metabolome					Extracellular metabolome		
	M1	M2	M3	M4	M5	L1	L2	L3
Wildtype	23%	20%	19%	20%	25%	15%	18%	22%
ΔpSymA	26%	20%	19%	18%	23%	15%	18%	26%
ΔpSymB	22%	26%	21%	19%	19%	17%	12%	13%
ΔpSymAB	26%	21%	23%	24%	29%	12%	14%	13%
LBmc	---	---	---	---	---	14%	---	12%

Table S3 Summary table of validation parameters (R^2X , R^2Y , Q^2) of the OPLS-DA model for the intracellular metabolome in Figures S3, S4. Model 1-4 corresponds to Figure S3 a-d, and model 5-9 corresponds to Figure S4 a-e

model	$R^2X(\text{cum})$	$R^2Y(\text{cum})$	$Q^2(\text{cum})$	Conditions
1	0.886	0.922	0.617	$\Delta p\text{SymA}$ (M1-5)
2	0.875	0.881	0.645	WT (M1-5)
3	0.808	0.831	0.677	$\Delta p\text{SymB}$ (M1-5)
4	0.798	0.842	0.631	$\Delta p\text{SymAB}$ (M1-5) M3 and M4 were considered as one group
5	0.915	0.982	0.925	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$ (M1)
6	0.879	0.962	0.857	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$ (M2)
7	0.888	0.985	0.882	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$ (M3)
8	0.856	0.976	0.872	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$ (M4)
9	0.889	0.98	0.956	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$ (M5)

Table S4 Summary table of validation parameters (R^2X , R^2Y , Q^2) of the OPLS-DA model for the extracellular metabolome in Figure S6. Model 1-7 corresponds to Figure S6 a-g

Model	$R^2X(\text{cum})$	$R^2Y(\text{cum})$	$Q^2(\text{cum})$	Conditions
1	0.903	0.987	0.981	$\Delta p\text{SymA}$ (L1-3)
2	0.917	0.99	0.979	WT (L1-3)
3	0.912	0.998	0.95	$\Delta p\text{SymB}$ (L1-3)
4	0.892	0.995	0.954	$\Delta p\text{SymAB}$ (L1-3)
5	0.885	0.835	0.657	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$, LBmc (L1)
6	0.925	0.988	0.947	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$, LBmc (L2)
7	0.938	0.98	0.914	WT, $\Delta p\text{SymA}$, $\Delta p\text{SymB}$, $\Delta p\text{SymAB}$, LBmc (L3)

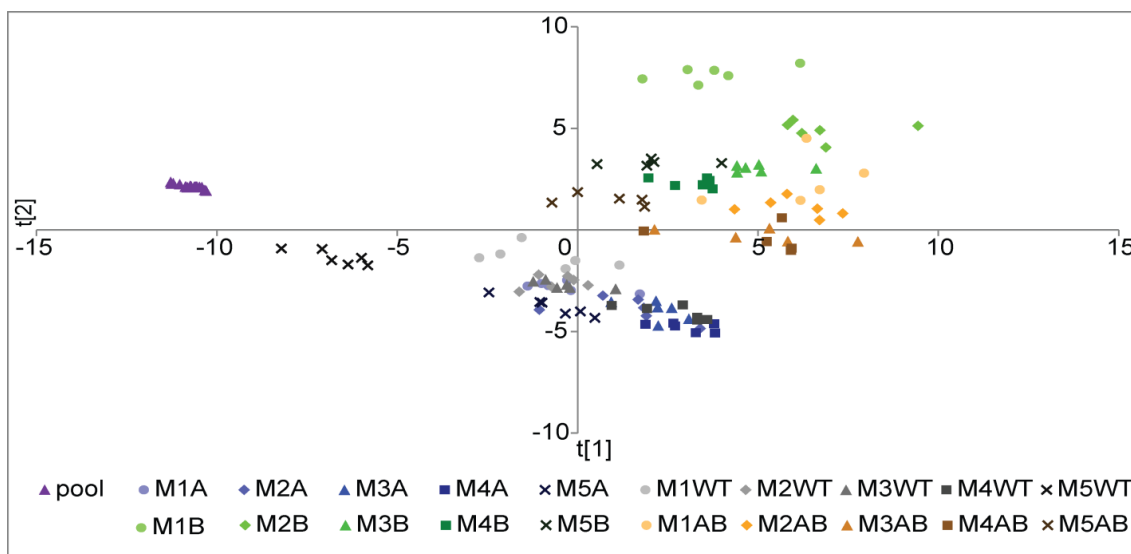


Figure S1 The OPLS-DA plot (eight aligned and one orthogonal components) of the intracellular metabolome of wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* strains cultured in M9-sucrose medium at multiple times points on the growth curve, as well as pooled samples. Metabolome variance between strains and during growth was summarized by x- and y-axes. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$, pooled samples—purple. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$

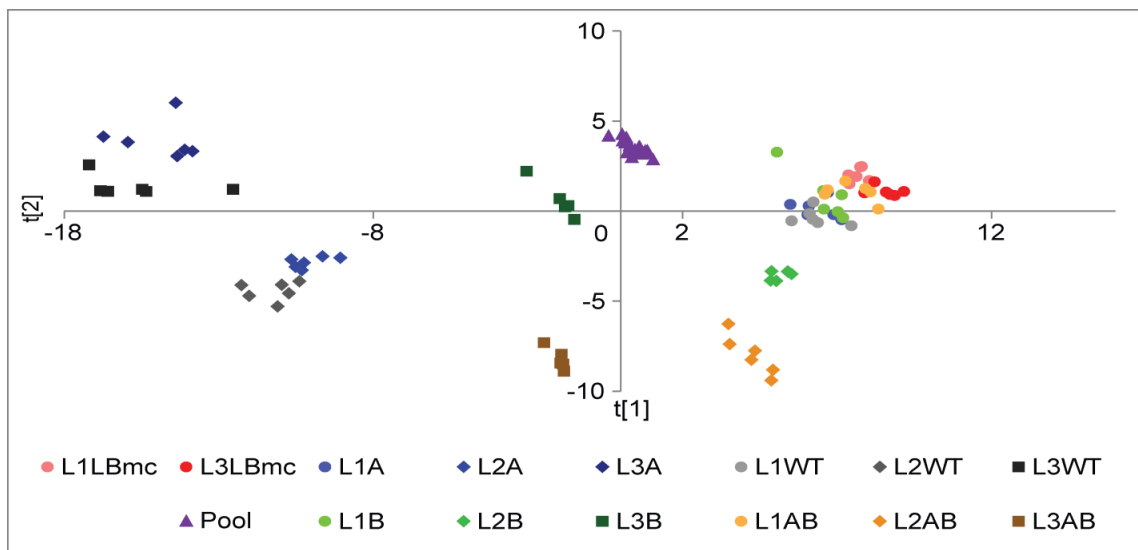


Figure S2 The OPLS-DA plot (six aligned and three orthogonal components) of the extracellular metabolome of wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* strains cultured in LBmc at multiple times points on the growth curve, as well as pooled samples. Metabolome variance between strains and during growth was summarized by x- and y-axes. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$. L1—circle; L2—diamond; L3—square. L1LBmc—fresh LBmc medium; L3LBmc—LBmc medium incubated at 30°C for 45hr; WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$

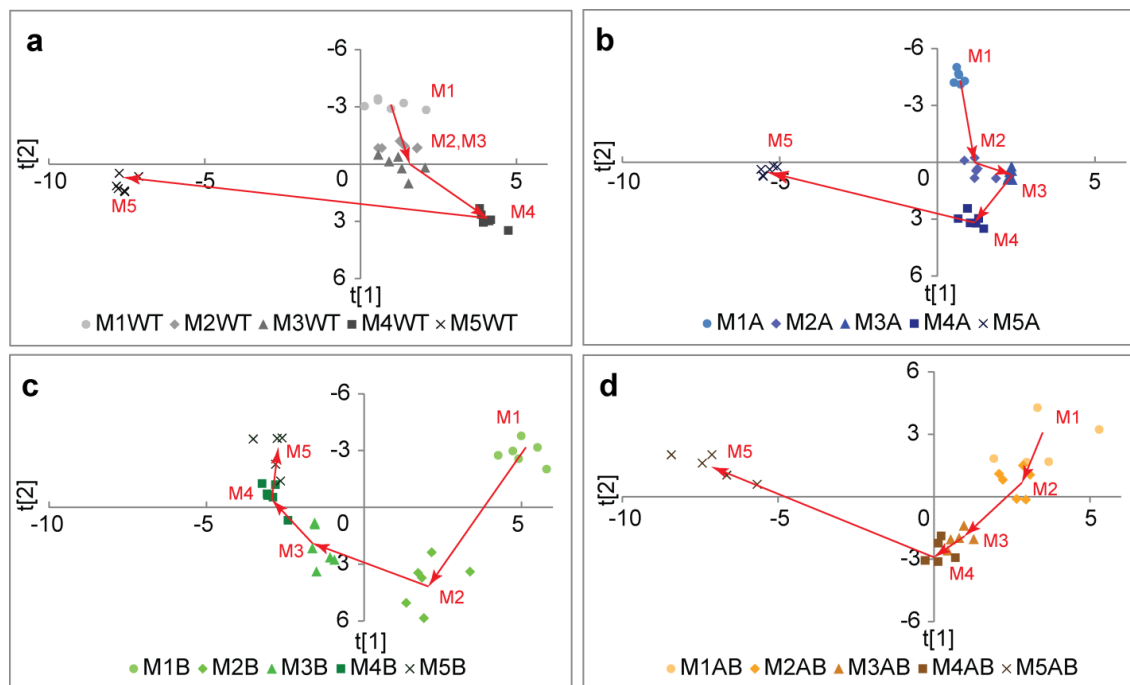


Figure S3 The OPLS-DA score plots of the intracellular metabolome of **a** wild-type, **b** $\Delta pSymA$, **c** $\Delta pSymB$, and **d** $\Delta pSymAB$ *S. meliloti* at various growth phases including exponential phase (M1-4) and stationary phase (M5). The progression of growth is indicated by the arrow direction. The metabolome variance for each strain during growth was summarized by x- and y-axes. The OPLS-DA models for wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ consist of four, four, four, three aligned and three, five, one, two orthogonal components, respectively. $\Delta pSymAB$. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$. M1—circle; M2—diamond; M3—triangle; M4—square; M5—cross. WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$

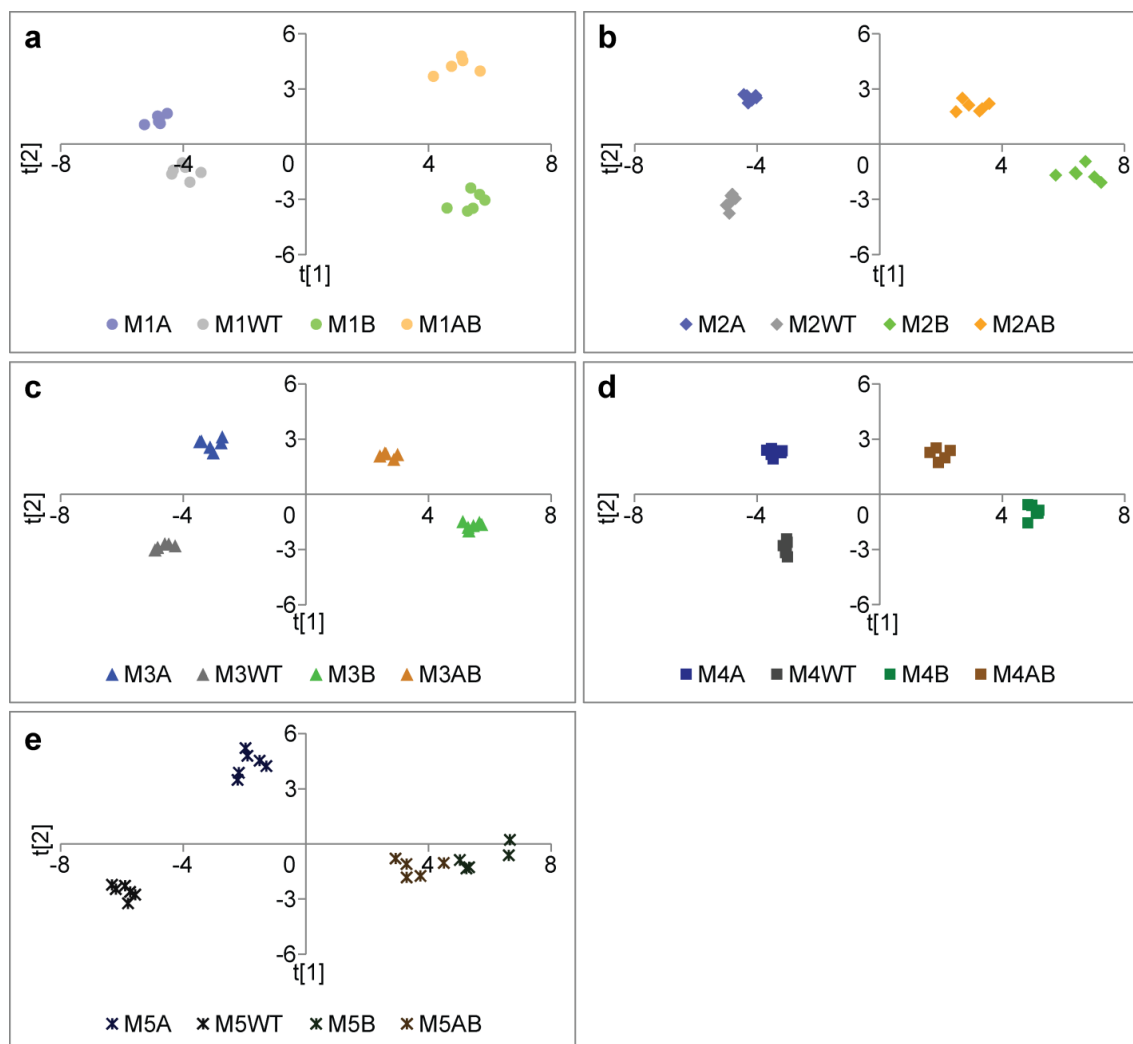


Figure S4 The intracellular metabolic phenotypes of $\Delta pSymA$, wild type, $\Delta pSymB$ and $\Delta pSymAB$ strains at exponential phase (a M1, b M2, c M3, d M4), and e stationary phase (M5) were summarized using OPLS-DA plots. Metabolome variance for each strain during growth was summarized by x- and y-axes. The OPLS-DA models for M1, M2, M3, M4, and M5 growth phases consist of three, three, three, three, three aligned and four, four, four, three, one orthogonal components, respectively. Black—wild type; blue— $\Delta pSymA$; green— $\Delta pSymB$; orange— $\Delta pSymAB$, pooled samples—purple. M1—circle; M2—diamond; M3—triangle; M4—square; M5—star. WT—wild type; A— $\Delta pSymA$; B— $\Delta pSymB$; AB— $\Delta pSymAB$

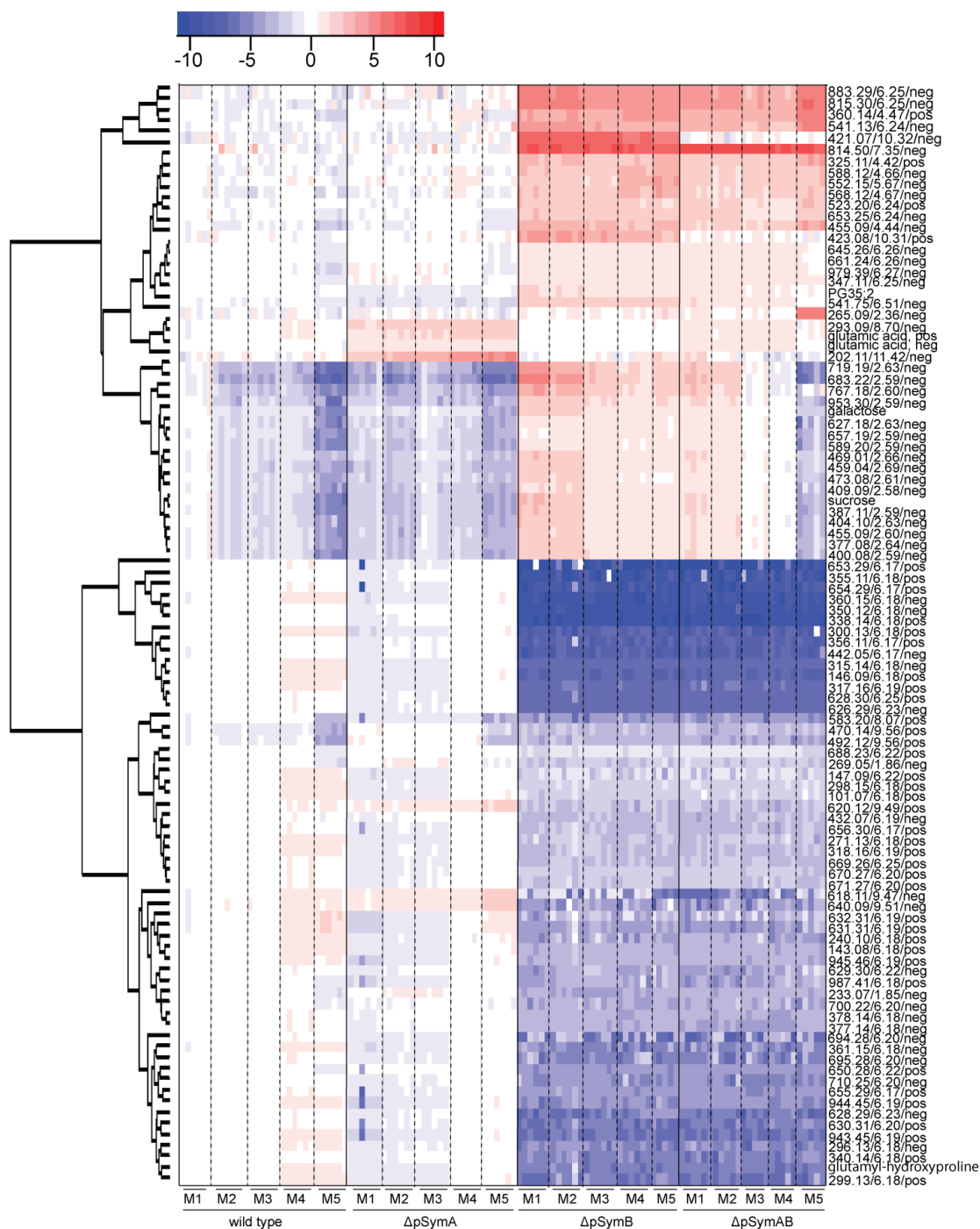


Figure S5 The top 100 intracellular metabolites with the most significant p_{overall} computed using between subjects two-way ANOVA ($p_{\text{overall}} = \min(p_{\text{time}}, p_{\text{strain}}, p_{\text{interaction}})$). The metabolites were labelled in the order of m/z, retention time (min), and electrospray ionization mode. M1-4—exponential phase; M5—stationary phase

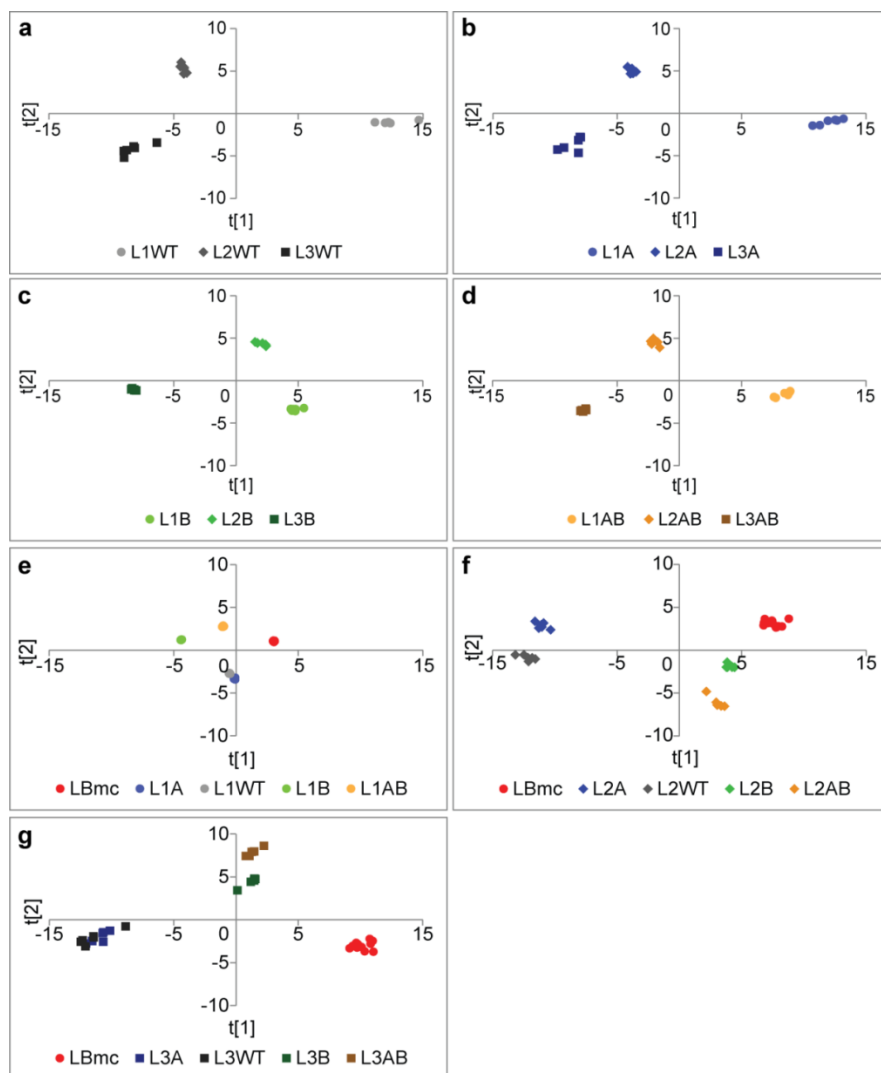


Figure S6 The OPLS-DA plots of the extracellular metabolome of wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ *S. meliloti* cultured in LBmc at various growth phases. The extracellular metabolic phenotypes of **a** wild type, **b** $\Delta pSymA$, **c** $\Delta pSymB$ and **d** $\Delta pSymAB$ at mid-exponential phase (L1), early stationary phase (L2) and late stationary phase (L3) were summarized using OPLS-DA plots. The OPLS-DA score plots of the extracellular metabolome of wildtype, $\Delta pSymA$, $\Delta pSymB$, $\Delta pSymAB$ and LBmc medium were compared at **e** L1, **f** L2, and **g** L3. For a-d, metabolome variance for each strain during growth was summarized by x- and y-axes. For e-g, metabolome variance between strains at each growth phase was summarized by x- and y-axes. The OPLS-DA models for wild-type, $\Delta pSymA$, $\Delta pSymB$, and $\Delta pSymAB$ consist of two, two, two, two aligned and one, one, three, two orthogonal components, respectively. The OPLS-DA models for L1, L2, and L3 growth phases consist of two, three, three aligned and two,

three, two orthogonal components, respectively. Black—wild type; blue— Δ pSymA; green— Δ pSymB; orange— Δ pSymAB. L1—circle; L2—diamond; L3—square. WT—wild type; A— Δ pSymA; B— Δ pSymB; AB— Δ pSymAB; LBmc—LBmc blank medium incubated for 0 or 45 hrs at 30°C

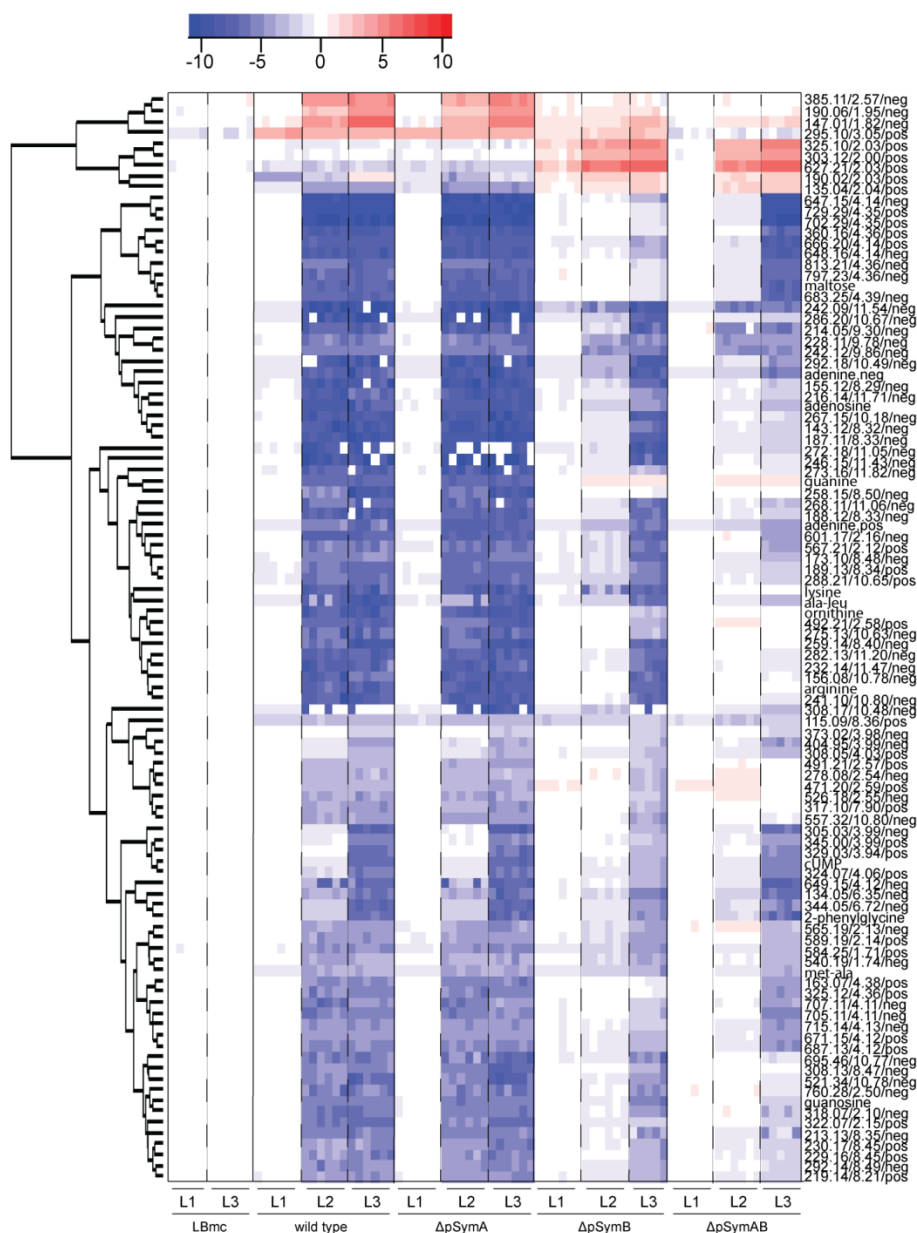


Figure S7 The top 100 extracellular metabolites with the most significant p_{overall} computed using between subjects two-way ANOVA ($p_{\text{overall}} = \min(p_{\text{time}}, p_{\text{strain}}, p_{\text{interaction}})$). The metabolites were labelled in the order of m/z, retention time (min), and electrospray

ionization mode. L1—mid exponential phase; L2—early stationary phase; L3—late stationary phase

References

Fei, F., Bowdish, D.M.E., McCarry, B.E. (2014). Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal Bioanal Chem*, 406(15), 3723–3733.

Chapter 5: Age-associated metabolic dysregulation in bone marrow-derived macrophages stimulated to lipopolysaccharide

Fan Fei^{1,2}, Keith M. Lee², Brian E. McCarry¹, Dawn M. E. Bowdish^{2},*

The work in this chapter has been submitted to Scientific Reports with minor revision requested.

Author Contributions

BEM and DMEB conceived the study. FF, BEM, and DMEB designed the study. KML cultured macrophages and performed gene expression experiment. FF extracted cells, design and acquired data from both comprehensive and targeted metabolomic experiments, and performed data analysis. FF and DMEB interpreted the data. FF and DMEB wrote the manuscript.

Author Affiliations:

¹ Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Canada L8S4M1

² Department of Pathology and Molecular Medicine, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada L8N3Z5

*Corresponding author: Email: bowdish@mcmaster.ca; Phone: (+1)905-525-9140 ext22313; Fax: (+1) 905-522-6750

Abstract

Macrophages are major contributors to age-associated inflammation. Inflammatory responses in macrophages are generally regulated by metabolic processes such as oxidative phosphorylation, glycolysis and the urea cycle. It is known that metabolic profiles change with age; therefore, we hypothesized that dysregulation of metabolic processes could contribute to age-associated macrophage dysfunction. We examined the intracellular metabolome of bone marrow-derived macrophages from young (6-8 wk) and old (18-22 mo) mice following lipopolysaccharide (LPS) stimulation and tolerance. We discovered known and novel metabolites that were associated with the LPS response of macrophages from young mice, which were not inducible in macrophages from old mice. Macrophages from old mice were largely non-responsive towards LPS stimulation, and we did not observe a shift from oxidative phosphorylation to glycolysis. The critical regulatory metabolites succinate, γ -aminobutyric acid, arginine, ornithine and adenosine were increased in LPS-stimulated macrophages from young mice, but not macrophages from old mice. A shift between glycolysis and oxidative phosphorylation was not observed during LPS tolerance in macrophages from either young or old mice. Metabolic bottlenecks may be one of the mechanisms that contribute to the dysregulation of inflammatory responses with age.

Introduction

Inflammation is an evolutionarily conserved response to infection and tissue injury, which triggers a complex cascade of metabolic and genomic responses (Nathan 2002). Both innate and adaptive immune function declines with age (Solana *et al.* 2006; Weng 2006; Gomez *et al.* 2005), and this contributes to decreased vaccine response (Weiskopf *et al.* 2009) and increased susceptibility to sepsis and inflammatory diseases (Ginaldi *et al.* 2001). Franceschi *et al.* proposed that macrophages play a central role in producing age-associated inflammation, which ultimately impairs the immune response (Franceschi *et al.* 2000). Macrophages are heterogeneous tissue-resident sentinel cells that are derived from hematopoietic progenitors (Gordon & Taylor 2005). They initiate inflammatory responses towards microbial pathogens and repair damaged tissues (Franceschi *et al.* 2000) by responding to their local cytokine environment and adapting to either pro-inflammatory (M1) or anti-inflammatory (M2) phenotypes (Murray *et al.* 2014). With age, macrophage functions, including phagocytosis, wound healing and polarization, are impaired (Plowden *et al.* 2004; Mahbub *et al.* 2012).

Bacterial lipopolysaccharide (LPS) is a potent inflammatory stimulant that is often used to study macrophage function. Upon repeated challenge with LPS, macrophages reduce inflammatory responsiveness and it can persist for 24-48 hrs after initial stimulation (Foster *et al.* 2007). Tolerance towards LPS is an essential immune-homeostatic response that protects against persistent infection (Cavaillon & Adib-Conquy 2006). Defects in LPS tolerance might contribute to septic and non-infectious systemic inflammatory response syndrome (SIRS) in humans (Cavaillon & Adib-Conquy 2006). Peritoneal macrophages of young mice develop LPS tolerance more effectively than macrophages from old mice at the transcriptional level (Sun *et al.* 2012). Whether failure to control inflammation resulting from chronic LPS exposure in old age contributes to increased susceptibility to inflammatory diseases is not known.

Inflammatory responses of macrophages can be regulated by intracellular and extracellular levels of metabolites. It is known that upon LPS stimulation, macrophages switch from oxidative phosphorylation to glycolysis as their primary energy source to sustain the increased energy demand during inflammation (Rodríguez-Prados *et al.* 2010; Sugimoto *et al.* 2011). Enhanced glycolytic function is measured by higher levels of intra- and extra-cellular lactate. Specific transcriptional responses promoting inflammation have been shown to be regulated by metabolites such as succinate and γ -aminobutyric acid (Tannahill *et al.* 2013). Additionally, M1/M2 polarization is regulated by increasing levels of urea cycle intermediates such as arginine, ornithine, citrulline (Mills *et al.* 2000). Moreover, increasing levels of adenosine as a result of inflammation

can regulate inflammatory responses and are protective against tissue damage (Haskó & Cronstein 2004). Metabolic changes have been noted in mice and humans as a result of aging (Houtkooper *et al.* 2011; Yu *et al.* 2012). Whether metabolic dysregulation can contribute to macrophage dysfunction with age is not known.

Here, for the first time, we identified age-specific metabolic dysregulation of LPS responses in bone marrow-derived macrophages. Additionally, we quantified the metabolic changes during LPS tolerance in both young and old macrophages. We discovered novel metabolites that are associated with LPS stimulation. We have found metabolic reprogramming of oxidative phosphorylation to glycolysis was suppressed in LPS stimulated macrophages from old mice. In addition, arginine metabolism, which is vital for macrophage polarization (Hibbs Jr *et al.* 1987; Mills *et al.* 2000), was also impaired in old macrophages. Our data indicate a possible metabolic bottleneck that prevents energy intensive inflammatory responses in old macrophages.

Results

In order to quantitate differences in macrophage metabolism during LPS stimulation and LPS tolerance, bone marrow derived macrophages from young and old mice were analyzed using both comprehensive and targeted metabolomic strategies (Fig. 1). Liquid chromatography-mass spectrometry (LC-MS) was used to create a comprehensive metabolomic profile, which was composed of 2125 metabolite features, of which 57 polar metabolites and 64 phospholipids were identified. Gas chromatography (GC)-MS was used for targeted metabolomic analysis, which included 25 intermediates in glycolysis, the citric acid cycle (TCA), the aspartate-argininosuccinate shunt, the γ -aminobutyric acid (GABA) shunt and the urea cycle pathways (Table S1).

Comprehensive analysis reveals novel metabolites associated with LPS responses

The metabolome of bone marrow derived macrophages from young mice were analyzed and compared at 0, 4 and 16 hr of LPS stimulation. To ensure any metabolic changes only resulted from LPS stimulation and were not a result of the 22 hr incubation, we examined the metabolic profiles of unstimulated macrophages at $t=0$ hr and $t=22$ hrs. Less than 1.3% of the metabolic features showed any significant change over the 22 hr period. Significant metabolic changes were observed for LPS stimulated macrophages after 4 hrs of stimulation, and a more dramatic change was noted after 16 hrs of LPS stimulation (Fig 2A). From the heatmap (Fig 2B), 4.5% (96/2125) of the metabolite features from young macrophages were differentially expressed after 4 hrs of LPS stimulation compared to the unstimulated control. After 16 hrs of LPS stimulation, over half of the metabolic features (1081/2125) were significantly altered in the young

macrophages. Of the differentially expressed features, 27.2% (579/2125) showed increased expression and 23.6% (502/2125) features were reduced compared to the unstimulated control. Metabolites that were found to increase in macrophages of young mice after 16 hrs of LPS stimulation included adenine, adenosine, ornithine, arginine (Fig. 3A-D), pantothenic acid, uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), N-acetyl-phenylalanine, taurine, hypotaurine, UDP-glucose (UDP-G), glucosamine-6-phosphate (GlcN6P), methyl-malonic acid, lysine, proline, glutamine, phosphatidylglycerols (PGs), phosphatidylethanolamine (PEs) and phosphatidylcholines (PCs). N-acetyl glutamic acid and N-acetyl-aspartic acid were reduced after 16 hrs of LPS exposure. Most metabolic features remained unidentified. The comprehensive metabolomic approach allows the discovery of novel metabolites associated to macrophage LPS responses (normalized levels of metabolites were the electronic supplementary material ESM 1).

Macrophage metabolism in response to LPS decreases with age

The metabolomes of bone marrow derived macrophages from both young and old mice were analyzed and compared prior to LPS stimulation ($t=0$ hr), and after 4 and 16 hrs of LPS stimulation. Only 0.4% of the metabolites were significantly different between macrophages derived from young and old mice in the absence of LPS indicating that there were virtually no detectable age-associated metabolic differences in unstimulated macrophages. The age-associated differences in metabolism after LPS stimulation were visualized using an OPLS-DA score plot (Fig. 2A). Old macrophages appeared to be non-responsive towards LPS stimulation as compared to the young. Age-associated metabolic changes after 4 hrs of LPS stimulation were apparent between young and old macrophages, and became more distinct after 16 hrs of LPS stimulation. Unlike LPS stimulated young macrophages where half of the metabolome was altered, only 2.2% (46/2125) and 13.3% (282/2125) of the features were altered in the old macrophages after 4 hrs and 16 hrs of LPS stimulation, respectively. 10.0% (211/2125) of the metabolic features that were significantly changed with similar magnitude in both young and old macrophages after 16 hrs of LPS stimulation. Metabolites that were found to be increased in both young and old macrophages after 16 hrs of LPS stimulation included pantothenic acid, UDP-GlcNAc, N-acetyl-phenylalanine, PEs and almost half of detectable PCs. Metabolic responses to LPS stimulation appears to decrease with age.

Macrophages from old mice have defects in core metabolism during LPS stimulation

Macrophages switch their core metabolism from oxidative phosphorylation to glycolysis when stimulated with LPS (Rodríguez-Prados *et al.* 2010; Sugimoto *et al.* 2011). To examine whether the core metabolism of activated macrophages was affected

by age, we designed a targeted metabolomics approach including in intermediates in glycolysis, the TCA cycle, the GABA shunt, the aspartate-argininosuccinate shunt and the urea cycle.

After 4 hrs of LPS stimulation, metabolites associated with the TCA cycle including oxaloacetate, malate, fumarate, succinate, α -ketoglutarate, citrate and glutamate were increased in macrophages from young mice (Fig. S2A, S3A). After 16 hrs of LPS stimulation, the above-named metabolites remained elevated, and in addition, fructose-6-phosphate, lactate, GABA, arginine and ornithine were also increased compared to the unstimulated control (Fig. 4A). In contrast, very few changes in this core metabolic pathway were noted in macrophages from old mice after LPS stimulation. Isocitrate, 2-phosphoglycerate (2PG) and 3PG were only decreased in the old macrophages after 4 hrs of LPS stimulation, and after 16 hrs of LPS stimulation, only arginine, malate, aspartate and GABA were increased (Fig. 5A, S2B, S3B). Decreased glucose-1-phosphate was observed in macrophages from both young and old mice after 4 and 16 hrs of LPS stimulation.

Metabolic changes during LPS tolerance

To examine the metabolic response associated with LPS tolerance in macrophages, we analyzed the metabolome of macrophages stimulated with secondary dose of LPS for 4 hrs (“tolerance”). As a control, after 16 hrs of LPS stimulation, the cells were washed and cultured in LPS-free medium for 6 hrs (“recovery”). As visualized by the OPLS-DA score plot, the “recovery” and “tolerance” metabolic profiles from young mice resembled the early stage of LPS stimulation (Fig. 1C). In contrast, these profiles from old macrophages were distinct from early and late stage LPS stimulation (Fig. 1D).

We confirmed the LPS tolerance by quantifying transcripts that were known to be associated with the LPS tolerance (Fig. S4). Some metabolic changes appeared to be reversible when LPS was removed, whereas others remained irreversible. Reversible metabolites were induced in LPS stimulation but were reversed when LPS were removed and were not inducible upon LPS re-stimulation; irreversible metabolites were induced in LPS stimulation and remained elevated in the LPS re-stimulation. There are also a unique set of metabolites that were only induced in the tolerant macrophages. For macrophages from young mice, 20.0% (424/2125) of the metabolic features were irreversible, 30.9% (657/2125) were reversible, and 9.4% (200/2125) of the features became differently expressed after secondary LPS stimulation. Metabolites that were increased in primary LPS stimulation and remained elevated in both the “recovery” and “tolerance” group included taurine, hypotaurine, UDP-G, GlcN6P, lysine, proline, ornithine, arginine, glutamine, PEs and PCs. PGs showed no evidence of tolerance, although the magnitude

of induction in the “tolerance” experiment was much lower compared to those after 16 hrs of LPS stimulation. Conversely, for macrophages from old mice, 6% (128/2125) of the metabolic features were irreversible, 7.2% (154/2125) were reversible, and 6.8% (145/2125) of the features were only significant in “tolerance” group as compared to the unstimulated macrophages from old mice. Interestingly, arginine, ornithine, UDP-GlcNAc, UDP-G, ornithine, glutamine and hypotaurine, which were only elevated in the young macrophages after 4 and 16 hrs of LPS stimulation, were found elevated in the “recovery” and “tolerance” groups of the old macrophages. It might suggest metabolic responses to LPS stimulation are delayed in old macrophages.

Similar trends were also observed for metabolites in macrophage core metabolism. In young macrophages, almost all of the metabolites that were elevated after 16 hrs of LPS stimulation were still elevated in the “recovery” and “tolerance” groups. These included citrate, oxaloacetate, malate, fumarate, succinate, α -ketoglutarate, arginine, ornithine, glutamate and glutamine. Only GABA and lactate decreased back to the levels observed in unstimulated macrophages. Interestingly, proline levels did not increase during LPS stimulation, but increased in the “recovery” group. Intermediates of the glycolysis pathway were induced only after secondary LPS exposure in the young macrophages, and those included isocitrate and glycolysis intermediates such as glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), 3PG and 2PG. Metabolic response to LPS was delayed in the old macrophages. Metabolites such as oxaloacetate, fumarate, glutamine, arginine and ornithine were only increased in old macrophages after 16 hrs of LPS stimulation and in both the “recovery” and “tolerance” macrophages.

Age-associated changes in arginine metabolism

Arginine metabolism via the urea cycle influences macrophage polarization (Mills *et al.* 2000). The urea cycle intermediates, arginine and ornithine, were increased in the young macrophages after 16 hrs of LPS stimulation (Fig. 4). The levels of arginine and ornithine remained elevated in the young macrophages when LPS was removed and also during LPS tolerance. In the old macrophages, LPS stimulation did not result in elevated levels of arginine and ornithine. Similar to changes in the core metabolism, there was a delay in arginine metabolism. Increasing levels of arginine and ornithine were only observed during the “recovery” and “tolerance” samples (Fig. 5).

The gene expression of *Arg1*, *iNOS* (*Nos2*) and cationic amino acid transport (*Slc7a2*) were measured to provide insights into arginine metabolism. The expression of *Arg1* and *iNOS* were elevated but not statistically different between LPS stimulated macrophages from young and old mice. Only *Slc7a2* was differentially expressed between young and old macrophages after 16 hrs of LPS exposure, for which the

expression of *Slc7a2* in young macrophages was 1.43-fold greater than the old (Fig. 3 E-H).

Discussion

LPS stimulation triggers a shift in macrophage core metabolism from oxidative phosphorylation to glycolysis (Rodríguez-Prados *et al.* 2010; Sugimoto *et al.* 2011). Glycolysis occurs in the cytoplasm and produces two ATPs per glucose. The end product of glycolysis, pyruvate, enters the mitochondria and initiates the TCA cycle and oxidative phosphorylation and results in the production of an additional 36 ATPs under aerobic conditions. Under anaerobic conditions, pyruvate is reduced to lactate in the cytoplasm and secreted (Zheng 2012). Although only 5% of the glucose's energy potential is taken advantage of during glycolysis, it can produce ATP at a much faster rate than oxidative phosphorylation. In addition to glucose metabolism, glutamine contributes to a third of the energy requirement of unstimulated macrophages via glutamine anaplerosis (Newsholme *et al.* 1999). The metabolic switch to glycolysis during LPS stimulation is a rapid way to accommodate the increased energy demand during inflammation. As macrophages shift from using oxidative phosphorylation to glycolysis, lactate and TCA intermediates such as malate, citrate and fumarate are increased intracellularly (Rodríguez-Prados *et al.* 2010; Sugimoto *et al.* 2011). Consistent with this, we have observed increasing levels of those metabolites in young macrophages after 16 hrs of LPS stimulation.

Macrophage specific adaptation to the TCA cycle is reported. Jha *et al.* have reported a M1 macrophage specific metabolic break-point in the TCA cycle (Jha *et al.* 2015). The metabolic flow between isocitrate and α -ketoglutarate is disrupted. Consistent with this result, in the young macrophages, we have observed a metabolic break in the TCA cycle between citrate and α -ketoglutarate where, despite a global increase of almost all the TCA cycle intermediates, isocitrate remained unchanged after 16 hrs by LPS stimulation. Since we also observed this in our model of inflammation, this break-point in the TCA cycle may not be specific to M1 phenotypes but rather an indication of energy metabolism during inflammation. Conversely, since the core metabolism of old macrophages did not change with LPS stimulation, we did not observe the citrate/ α -ketoglutarate metabolic break.

In the macrophages from old mice, the core metabolism was mostly unchanged to LPS stimulation, and this may suggest impaired mitochondrial function. Mitochondrial dysfunction during aging is well documented (Green *et al.* 2011). Although mitochondrial metabolic dysfunction has not been thoroughly studied in macrophages, several studies of mitochondria from muscle tissues have shown reduced rates of glycolysis and the TCA

cycle with age (Kaczor *et al.* 2006; Lee *et al.* 1994). We did not detect increasing levels of lactate, malate, fumarate and citrate, and this is consistent with mitochondrial dysfunction with age and an inability to shift from oxidative phosphorylation to glycolysis.

Interestingly, succinate was the only TCA cycle intermediate that was elevated in macrophages from both young and old mice in response to LPS stimulation (3.58- and 2.73- fold change at 16 hrs LPS respectively). Elevated succinate in macrophages in response to LPS stimulation has been shown to stabilize hypoxia-inducible factor-1 α (HIF-1 α), a transcription factor that is required for interleukin-1 β (IL-1 β) production (Tannahill *et al.* 2013). IL-1 β is a key inflammatory cytokine for macrophage activation during an immune response. The induction of glutamine anaplerosis and GABA shunt pathways were reported as principle source of succinate (Tannahill *et al.* 2013). In young macrophages, we have observed consistently elevated levels of GABA (2.37-fold), glutamate (2.38-fold) and glutamine (2.66-fold) that fed into succinate production. In contrast, only GABA (2.70-fold) was increased in old macrophages after LPS stimulation. Although there are conflicting reports as to whether old macrophages produce more or less inflammatory cytokines in response to LPS (Kohut *et al.* 2004; Mahbub *et al.* 2012), our data imply succinate and its biosynthetic pathway would not be a rate limiting factor in generating an inflammatory response.

Arginine is required for macrophage activation (Hibbs Jr *et al.* 1987). Elevated levels of arginine and ornithine were only observed in young macrophages following 16 hrs of LPS stimulation. Intracellular arginine is mostly imported from the extracellular environment via cationic amino acid transport (CAT) in both humans and mice (Bogle *et al.* 1992; Kurko *et al.* 2015). Of all the genes encoding the CAT, *Slc7a2* is the only gene whose expression is inducible and is required during both M1 and M2 macrophage polarization (Yeramian *et al.* 2006; Kurko *et al.* 2015). We have observed a 30% reduction in *Slc7a2* expression in the old macrophages compared to the young after 16 hrs of LPS stimulation, which likely contributes to the lower intracellular arginine level in the old macrophages. Arginine may also be synthesized via the aspartate-argininosuccinate shunt, which joins the TCA cycle with the urea cycle (Jha *et al.* 2015). Inhibition of the aspartate-argininosuccinate shunt inhibits M1 polarization with low *iNOS* expression and nitric oxide production (Jha *et al.* 2015). However, in our study, 16 hrs of LPS stimulation did not affect the aspartate-argininosuccinate shunt (i.e. aspartate, argininosuccinate) in macrophages from both young and old mice. Therefore, the aspartate-argininosuccinate shunt is unlikely to contribute to the increase of intracellular arginine in activated macrophages.

To further understand the role of the urea cycle in affecting macrophage dysfunction in old age, gene expression of *Arg1* and *iNOS* were measured. Metabolism of arginine via *Arg1* or *iNOS* in the urea cycle regulates macrophage polarization (Mills *et al.* 2000) (Fig. 3H). M1 macrophages express *iNOS* which metabolizes arginine to nitric oxide to prevent pathogen infection by limiting free arginine and producing nitric oxide, a powerful antimicrobial agent. M2 macrophages express *Arg1* which hydrolyzes arginine to ornithine to stimulate cell division and tissue repair through the production of polyamines and proline. LPS is known to induce both *Arg1* and *iNOS* (Murray *et al.* 2014), which we have also observed in LPS stimulated macrophages from young and old mice. Consistent with our data, there were no age-associated changes in *iNOS* or *Arg1* expression in bone marrow-derived macrophages; however, others have reported decreased expression in LPS stimulated splenic and peritoneal murine macrophages (Mahbub *et al.* 2012). Therefore, arginine metabolism is unlikely to be the rate limiting factor in the LPS response of old macrophages.

Dysregulation of LPS tolerance has been proposed as a possible explanation for the increased susceptibility of the elderly to sepsis and inflammatory disorders (Cavaillon & Adib-Conquy 2006). To test whether macrophages from old mice have defects in LPS tolerance, we measured a well-defined set of pro-inflammatory and anti-microbial genes associated with this phenomenon (Foster *et al.* 2007). Both the young and old macrophages experienced LPS tolerance equally at the transcriptional level. However, our data demonstrated that there were metabolic changes specific to LPS tolerance. In the young macrophages, lactate levels were increased after the initial LPS stimulation, but returned to baseline levels when LPS was removed and did not increase during the second LPS stimulation. This indicated that the core metabolism of young macrophages did not shift from oxidative phosphorylation to glycolysis during LPS tolerance. This inability to adjust macrophage core metabolism to during LPS tolerance may be compensated for by the up-regulation of glycolysis as indicated by the elevated levels of glycolysis intermediates. The inability to switch from glycolysis to oxidative phosphorylation may be a key metabolic break in LPS tolerance. In contrast to young macrophages, old macrophages did not shift from oxidative phosphorylation to glycolysis in either the first or second LPS stimulation. Moreover, elevated levels of arginine and TCA cycle intermediates were detected in both the first and second LPS stimulation in the young macrophages; however, those metabolites were only induced much later in the old macrophages. It appears as though the old macrophages had a delayed metabolic response to LPS stimulation. Whether this delay is because of age-associated deterioration in mitochondria function, and specifically ATP production from glucose, is worth future investigation.

Adenosine is a “retaliatory metabolite” whose intracellular level is amplified at sites of injury and inflammation, and also mediates the resolution of inflammation by limiting tissue destruction (Haskó & Cronstein 2004). Utilization of ATP during macrophage activation as a result of high metabolic activity leads to increased levels of intracellular adenosine (Ruiz-García *et al.* 2011). Moreover, elevated levels of adenosine are known to enhance glycolysis and ATP production that supports metabolism in activated macrophages (Ruiz-García *et al.* 2011). When adenosine is expressed in excess, it binds to adenosine receptors as an immune suppressor to preserve tissue homeostasis and prevent tissue damage (Haskó & Cronstein 2004). Preventing adenosine breakdown by inhibiting adenosine deaminase reduces systemic inflammation such as sepsis (Adanin *et al.* 2002). We observed increasing intracellular levels of adenosine and its precursor, adenine, after LPS stimulation in young macrophages. Young macrophages reduced the expression of adenosine and adenine to the baseline level of unstimulated macrophages during LPS tolerance. In contrast, levels of adenosine and adenine were not changed in condition in old macrophages. This might be a result of the decreased metabolic activity and reduced rate of glycolysis in old macrophages.

Overall, intricate age-associated metabolic dysfunction was observed in bone marrow-derived macrophages after LPS stimulation and during LPS tolerance. Inflammatory responses are energetically costly and result in high metabolic activity (Hotamisligil & Erbay 2008). For example, a 1°C to 4°C rise in core body temperature during fever helps to resolve bacterial and viral infections, and a 1°C increase in temperature demands a 10-12.5% increase in metabolic rate (Evans *et al.* 2015). However, fever responses are often absent or blunted in the elderly (Norman 1996), and rapid muscle wasting to sustain the high energy demand is common (Lochmiller & Deerenberg 2000). We have observed metabolic bottlenecks between the switch from oxidative phosphorylation to glycolysis, which might contribute to impaired inflammation in the elderly. Moreover, the elderly are more vulnerable to nutrient deficiencies, such as arginine (Brownie 2006). Low arginine can attribute to the retarded immune function (Hibbs Jr *et al.* 1987) and is associated to poor health outcomes in the elderly (Hurson *et al.* 1995). We found old macrophages are unable to adjust intracellular arginine level in response to LPS stimulation, but whether this is associated with arginine-related immune deficiencies is not known. In this study, metabolic bottlenecks were observed for macrophages from old mice during LPS stimulated inflammatory events, and metabolic dysregulation should be considered as a possible mechanism for declining immune function with age.

Experimental procedure

Bone marrow-derived macrophage culture

Bone marrow progenitors were collected from the femurs and tibia of young (6-8 wk) and old (18-22 mo) C57BL/6 mice (The Jackson Laboratory, Maine, USA). Progenitor cells were cultured for seven days in RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin and 15% L929 fibroblast cell conditioned medium on 150 mm Petri dish (Fisherbrand). On day 8, bone marrow-derived macrophages were sub-cultured and incubated overnight in 24-well tissue-culture-treated plates (Falcon). For metabolomic studies, three biological replicates were performed. For each biological replicate, 3×10^5 macrophages were seeded per well in 1 mL RPMI-1640 with two culture replicates per mouse ($n=6$ per treatment). For gene expression study, three biological replicates were performed. For each biological replicate, 1×10^6 macrophages were seeded per well in 3 mL RPMI-1640 ($n=3$ per treatment). All animal studies were approved by McMaster's Animal Research Ethics Board.

Macrophage LPS stimulation

Macrophages from both young and old mice were divided into six treatment groups. Groups 1 and 2 were stimulated with 100 ng/mL of bacterial lipopolysaccharide (LPS) in RPMI-1640 for 4 and 16 hrs respectively. Group 3 ("tolerance") was incubated for 16 hrs with an initial LPS challenge (100 ng/mL), washed in PBS, incubated in regular RPMI-1640 for 2 hrs, and then re-stimulated with 100 ng/mL of LPS for 4 hrs. Group 4 ("recovery") was incubated for 16 hrs with LPS (100 ng/mL), washed in PBS, incubated in regular RPMI-1640 for 6 hrs. Group 5 and 6 were controls that were cultured in regular RPMI-1640 for 0 and 22 hrs. The experimental protocol for LPS tolerance study was based on Foster *et al.* (Foster *et al.* 2007). All cells were washed once with cold phosphate-buffered saline (PBS) and collected for metabolomic and gene expression studies.

Macrophage extraction for metabolomic analyses

After macrophages were washed with 1 mL of cold PBS, the cells were detached from the 24-well plate using a cell lifter in the presence of 200 μ L cold extraction solvent mixture methanol/ethanol/H₂O (2:2:1) containing standards for recovery determination (98% L-methionine-d₃, 98% L-tryptophan-d₅) (Fei *et al.* 2014). The cell suspension was transferred into a 1.5 mL microtube (Diamed, ON, Canada) and vortex mixed for 2 min in the presence of two 2 mm ball bearings. After removal of the bearings, the mixture was centrifuged at $9500 \times g$ for 3 min at 4°C. The supernatant was collected and the precipitated pellet (containing DNA, RNA, and proteins) was re-extracted twice with 50 μ L cold methanol/ethanol/H₂O as with above. A total of 150 μ L cell extract was dried under nitrogen gas and re-solubilized in 50 μ L 60%v/v ACN/H₂O containing standards

(98% L-phenylalanine- d_8 , diphenylalanine, glycine-phenylalanine) for peak area normalization (IS) for LC-MS analysis. A quality control pooled sample was prepared by combining 5 μ L ACN/H₂O macrophage extracts from a total of 72 samples of all treatment groups.

For GC-MS analysis, 20 μ L of the ACN/H₂O macrophage extract or pooled samples was dried under nitrogen gas, and reconstituted in 25 μ L of 1%v/v chlorotrimethylsilane (TMCS) in N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) and 5 μ L anthracenemethanol (IS for GC-MS) in toluene (1.2 ng/ μ L). The samples were incubated at 60°C for 1 hr and analyzed by GC-MS immediately.

The entire sample preparation procedure was performed on ice or in a cold room. The sample extracts were stored in a -80°C freezer prior to analyses.

LC-HILIC-TOF-MS comprehensive analysis

Macrophage extracts were analyzed using a Agilent Technologies Model 1200RR series II liquid chromatograph coupled to a Bruker micrOTOF II Mass Analyzer as previously described (Fei *et al.* 2014). A Phenomenex Kinetix 2.6 μ m core shell HILIC column (2.1 x 50 mm, pore size 100 Å) was operated at 200 μ L/min using a linear gradient of acetonitrile (A) and 10 mM ammonium acetate, adjusted to pH 3 (B). The column temperature was maintained at 40 °C, and the auto sampler storage tray was set at 4°C. LC gradient: 0-0.5 min, 95% A; 0.5-12.5 min, 95% A to 35% A; hold at 35% A for 0.5 min; 35% A to 95% A over 1 min; re-equilibration at 95% A for 10 min prior to the next injection. A 2 μ L sample was injected to a total run time of 24 min for both positive and negative electrospray ionization (ESI) modes. The mass spectrometer setting was identical to those previously reported in Fei *et al.* (Fei *et al.* 2014).

GC-MS targeted analysis

GC-MS analyses were performed using an Agilent 6890N gas chromatograph (Santa Clara, CA, USA), equipped with a DB-17ht column (30 m $\times$ 0.25 mm i.d. $\times$ 0.15 μ m film, J & W Scientific) and a retention gap (deactivated fused silica, 5 m $\times$ 0.53 mm i.d.), and coupled to an Agilent 5973 MSD single quadrupole mass spectrometer. The autosampler storage tray was maintained below 5°C with a cooler system. The derivatized macrophage extract (1 μ L) was injected using Agilent 7683 autosampler in splitless mode. The injector temperature was 230 °C and carrier gas (helium) flow was 0.8 mL/min. The transfer line was 280°C and the MS source temperature was 230°C. The column temperature was set at 70°C for 0.1 min, raised to 225°C at 5°C/min, and then 310°C at 55°C/min and held there for 4 min. After a five minute solvent delay, mass spectra were acquired using electron ionization (EI) with a selected-ion-monitor (SIM) mode as in

Table S2. Metabolic intermediates of glycolysis, TCA, aspartate-argininosuccinate shunt, GABA shunt and urea cycle were included in this study (Table S1).

Quality Control

For both LC-MS and GC-MS metabolomic analyses, pooled samples were run 7 times to condition the column prior to sample analysis and also run after every 5th sample. MeOH/EtOH/H₂O blank and a standard mixture containing IS and RS were also run after every 10 samples. All samples were run in a randomized sequence.

Gene expression analysis

Modified from previously published protocol (Trapnell *et al.* 2012), the total RNA of macrophages was extracted and purified using TRIzol (Invitrogen, Carlsbad, CA, USA) and RNeasy Mini Kit (Qiagen, Venlo, Netherlands). Ribosomal RNA was depleted using the Human/mouse/Rat RiboZero Magnetic Kit (Epicentre, Madison, WI, USA), and verified using the Agilent RNA 6000 Nano Kit. DNases were removed using Turbo DNase (Invitrogen), and the sample was purified using RNAClean XP beads. The first strand of cDNA was synthesized using Superscript III (Invitrogen). Complimentary second strand cDNA was synthesized with RNase H and Klenow fragment of DNA polymerase I (Invitrogen). The cDNA was sonicated into 150 base pair fragments using a Covaris S220 Focused-ultrasonicator and deoxyadenosine monophosphate was incorporated to the cDNA fragment using NEBNext dA-Tailing Module (New England Biolabs). The cDNA library was sequenced using the Illumina HiSeq system.

Data Analyses and metabolite identification

The comprehensive LC-MS data were processed as in Fei *et al.* (Fei *et al.* 2014). The LC-MS spectra were converted to .mzXML format using Bruker CompassXport followed by internal mass calibration using sodium formate cluster in both ESI+ and ESI- modes by Bruker's DataAnalysis 4.0 SP4. The metabolite features were extracted and aligned using open source XCMS with centWave algorithm (minfrac = 0.8) (Smith *et al.* 2006); adducts, isotopic ions, and in-source fragments were identified using CAMERA (Kuhl *et al.* 2010). The metabolite features were normalized with IS eluted closest to their retention time (i.e. features eluted before 6.50 min were normalized by phe-phe; features eluted between 6.50 and 7.80 min were normalized by L-phenylalanine-d₈; features eluted after 7.80 min were normalized by gly-phe). After data reduction, the final data set was composed of 2125 metabolic features (Table S4).

The metabolite features were identified by matching the mass-to-charge (m/z) and the retention time of authentic standards or compound analogs (for phospholipid

identification only). There were 150 features identified to 121 metabolites in the final data set, 57 were polar metabolites and 64 were phospholipids. One metabolite could have multiple metabolic features resulting from adducts or in-source fragments.

For the GC-MS dataset, peak detection and spectrum deconvolution were processed using Agilent's Enhanced ChemStation. Multiple peaks generated from direct derivitization of a single metabolite were combined. The peak area of metabolites was normalized to anthracenemethanol. The final data sets for comprehensive LC-MS and targeted GC-MS analyses are included in electronic supplementary material ESM 1.

Statistical Analyses

The final data set of comprehensive LC-MS analysis was analyzed using multivariate analysis including principal component analysis (PCA) and OPLS-DA after pareto scaling by SIMCA-P+ 12.0.1 (Umetrics, Umeå, Sweden). The normality of both LC-MS and GC-MS data were analyzed using Kolmogorov-Smirnov test by SPSS 20 (SPSS, Chicago, IL, USA). Both LC-MS and GC-MS data were analyzed with univariate statistical tool including Student's t test (two-tailed, unpaired heteroscedastic) and one-way ANOVA by Microsoft Excel 2010 and MetaboAnalyst 3.0, respectively. Metabolic features or metabolites with p value less than 0.05 and fold change greater than 1.5 between treatment groups were considered significantly differentiated. Heatmap was generated with R Project 2.12.2 using gplots.

Electronic Supplementary Materials:

Electronic Supplementary datasets of intracellular metabolome and targeted metabolites for this chapter, ESM 1, can be found online (doi:).

Acknowledgements

This work was dedicated to late Prof. Brian McCarry (1946-2013). We like to thank Dr. M. Kirk Green for insightful discussion and Center for Microbial Chemical Biology (CMCB) for the access to the LC-MS instrumentation. DMEB is supported by the CIHR and hold Canada Research Chairs. Work in the Bowdish laboratory is supported by the McMaster Immunology Research Centre (MIRC) and M.G. DeGroot Institute for Infectious Disease Research (IIDR). FF is supported by Ontario Graduate Scholarship (OGS).

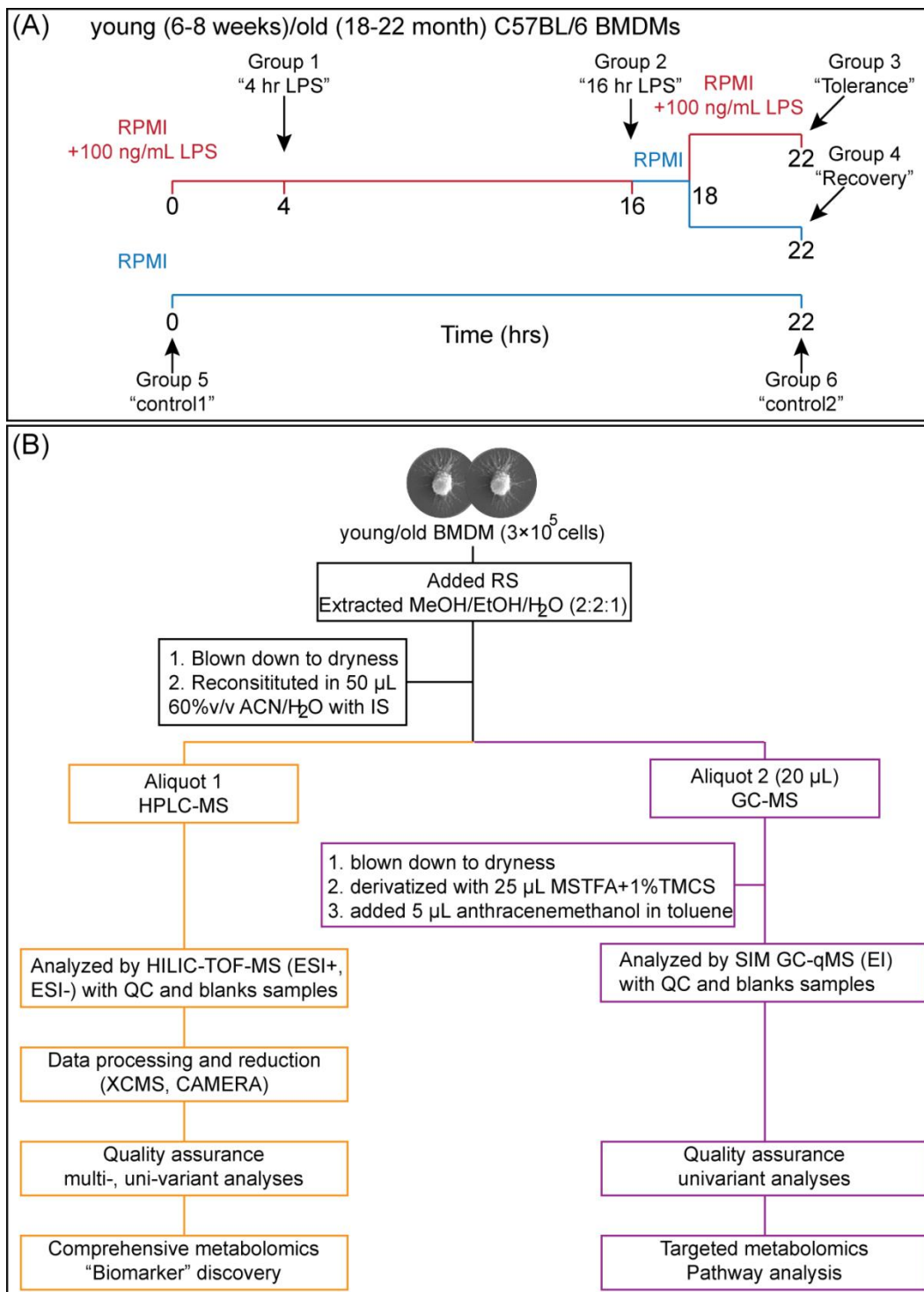


Figure 1 A) The experimental outline. B) The experimental workflow for analyzing macrophage extracts. From one macrophage culture, the sample extract was analyzed

separately with HILIC-TOF-MS and GC-qMS with distinct sample preparation, data acquisition, data processing, data analysis, and quality assurance.

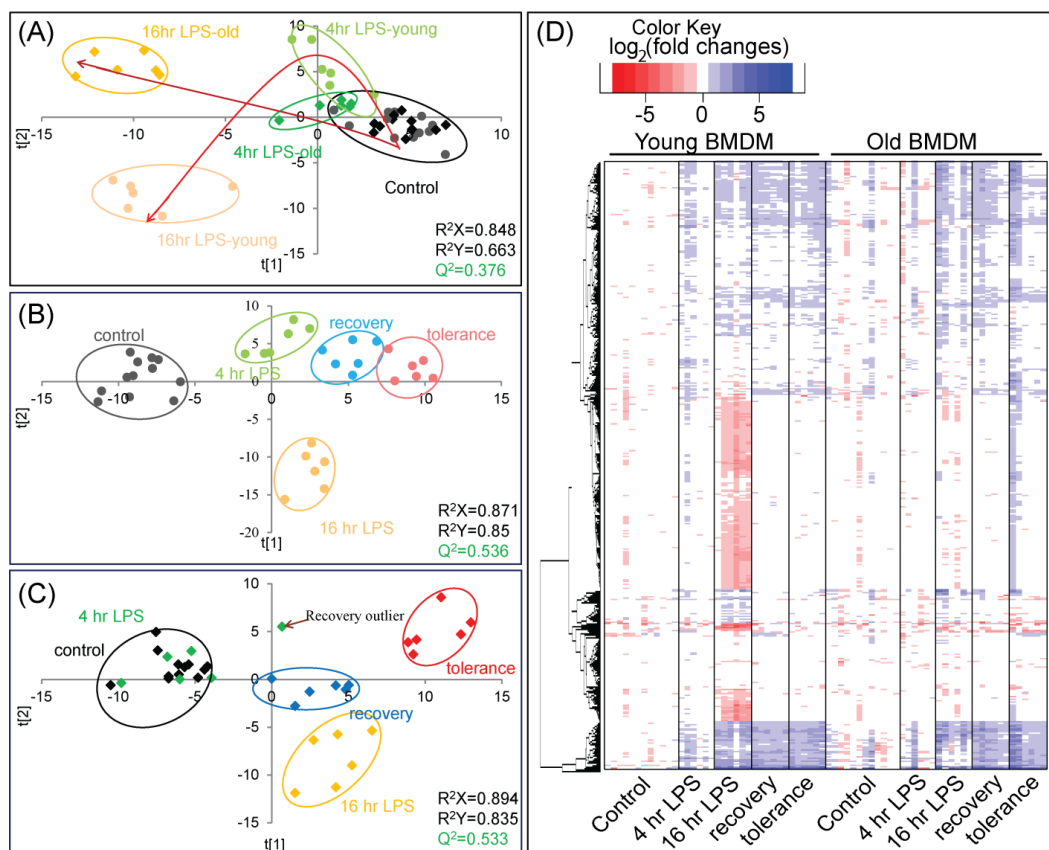


Figure 2 The comprehensive analyses of bone marrow-derived macrophage extracts acquired using HILIC-TOF-MS. The ionization responses of 2125 intracellular metabolite features were normalized using IS. Extracts were performed in sextuplicate with three biological replicates and two culture replicates. (A) OPLS-DA score plot comparing the metabolic profiles of control (0 hr), 4 hr and 16 hr LPS stimulated macrophage extracts from both young and old mice. (B) Heat map visualization of the intracellular metabolite changes of macrophages of young and old mice in response to LPS stimulation. The 920 significant metabolite features ($p < 0.05$ one-way ANOVA, fold change greater than 1.5 compared to control) are represented in rows, and the experimental conditions were listed in columns. The heat map is plotted based on $\log_2(\text{fold change})$ with respect to the average levels of each metabolite feature in the control of macrophages from young mice using Euclidean distance and complete-linkage clustering. OPLS-DA score plot comparing the metabolic profiles of control (0 hr), 4 hr, and 16 hr LPS stimulated macrophage extracts as well as “recovery” and “tolerance” treated macrophage extracts from (C) young or (D) old mice. Controls 1 and 2 obtained at the beginning ($t = 0$ hr) and the end of the experiments ($t = 22$ hrs) were considered as a single group. Samples belonging to the same treatment group were highlighted by open circles.

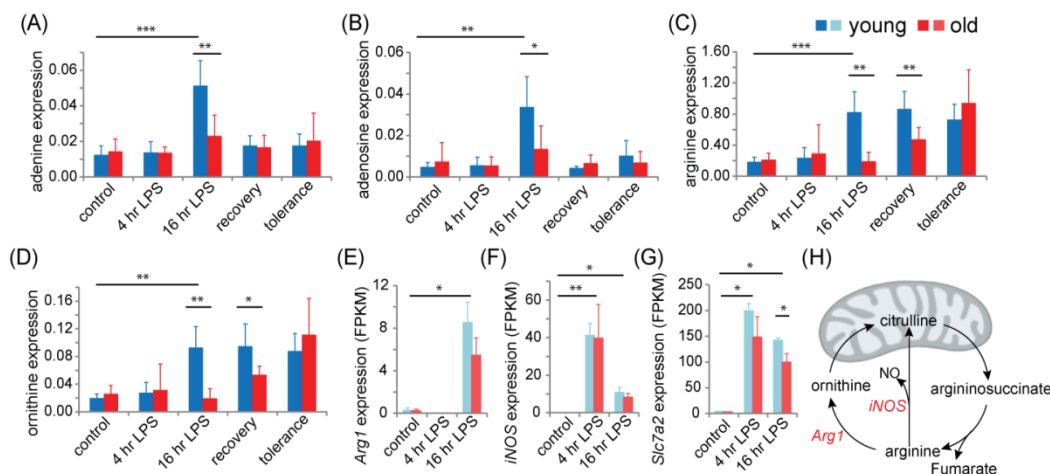


Figure 3 The relative intercellular levels and gene expression of selected metabolites and genes. The intracellular levels of adenine (A), adenosine (B), arginine (C), and ornithine (D) were acquired at control (0hr), 4 hr and 16 hr of LPS stimulation as well as at the “recovery” and “tolerance” conditions for macrophages of both young and old mice. The gene expressions of *Arg1* (E), *iNOS* (F), and *Slc7a2* (G) were acquired at control (0hr), 4 hr and 16 hr of LPS stimulation for macrophages of young and old mice. The urea pathway was illustrated in (H). Metabolomics and gene expression data of macrophages from young mice were labelled in shades of blue; those from old mice were labelled in shades of red. * p<0.05; ** p<0.01; *** p<0.001

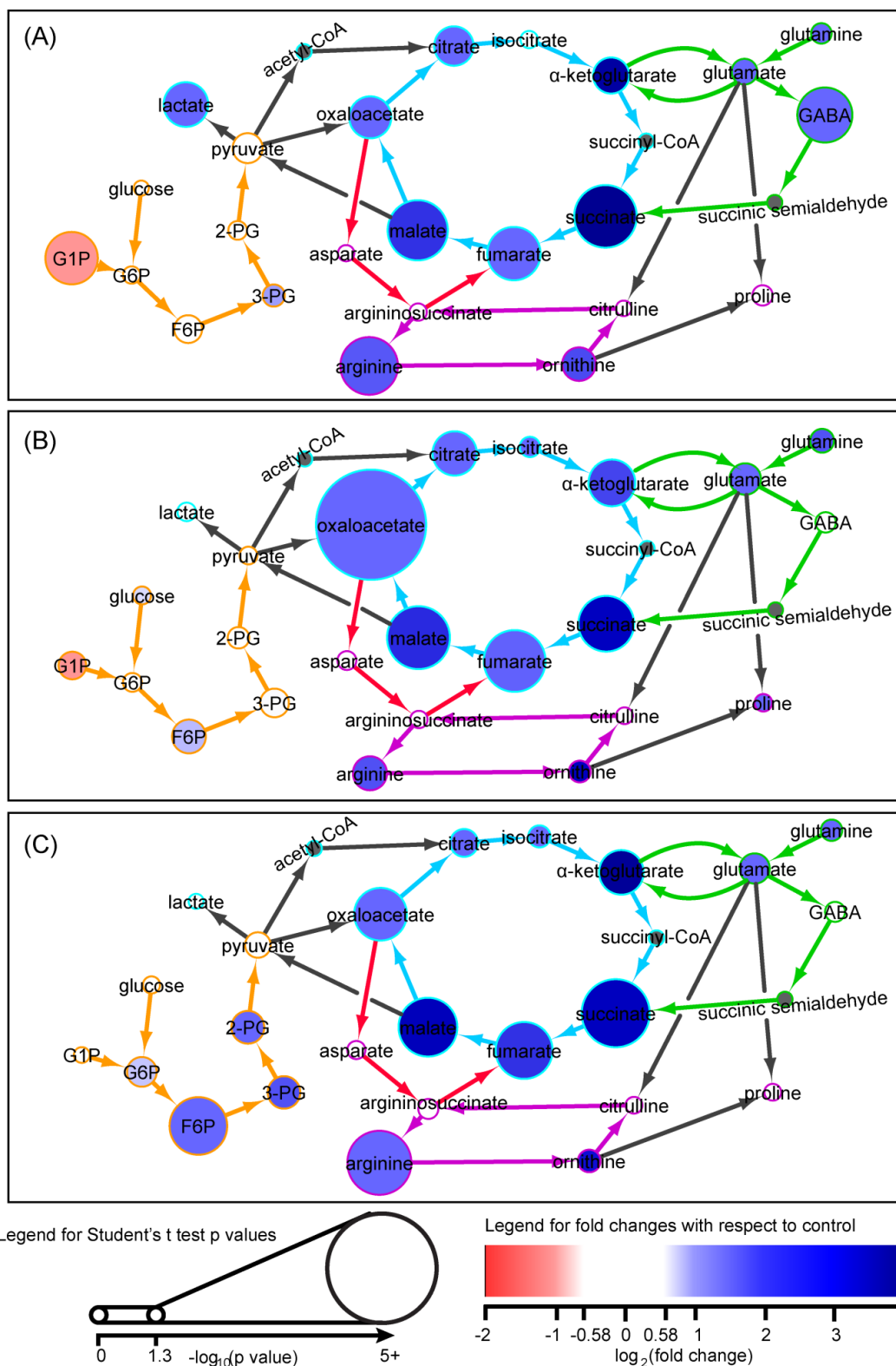


Figure 4 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates in bone marrow-derived macrophages from young mice in response to (A) 16 hr LPS stimulation, (B) “recovery”, and (C) “tolerance” as compared to unstimulated macrophages. The network of metabolite interactions was built based on the BioCyc database and pathway published from Jha *et al* (1). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2$ (fold changes) of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple.

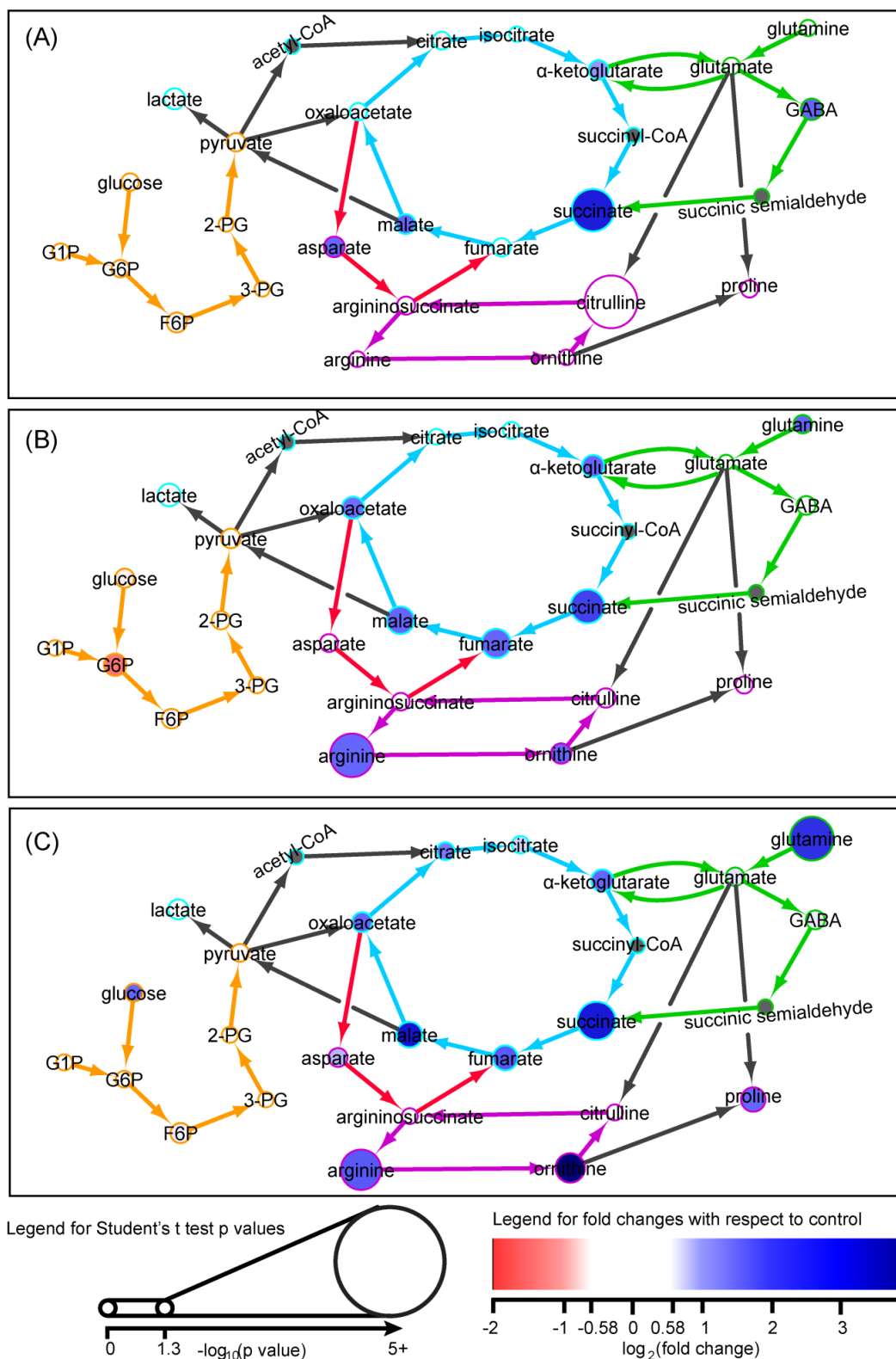


Figure 5 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates in bone marrow-derived macrophages from old mice in response to (A) 16 hr LPS stimulation, (B) “recovery”, and (C) “tolerance” as compared to unstimulated macrophages. The network of metabolite interactions was built based on the BioCyc database and pathway published from Jha *et al.* (1). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2$ (fold changes) of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple.

References

- Adanin S, Yalovetskiy I V, Nardulli BA, Sam II AD, Jonjev ŽS & Law WR (2002) Inhibiting adenosine deaminase modulates the systemic inflammatory response syndrome in endotoxemia and sepsis. *Am. J. Physiol. Regulatory Integrative Comp. Physiol* 282, R1324–R1332.
- Bogle RG, Baydoun AR, Pearson JD, Moncada S & Mann GE (1992) L-Arginine transport is increased in macrophages generating nitric oxide. *Biochem. J.* 284, 15–18.
- Brownie S (2006) Why are elderly individuals at risk of nutritional deficiency? *Int. J. Nurs. Pract.* 12, 110–118.
- Cavaillon J-M & Adib-Conquy M (2006) Bench-to-bedside review: Endotoxin tolerance as a model of leukocyte reprogramming in sepsis. *Crit. Care* 10, 233.
- Evans SS, Repasky E a & Fisher DT (2015) Fever and the thermal regulation of immunity: the immune system feels the heat. *Nat. Rev. Immunol.* 15, 335–349.
- Fei F, Bowdish DME & McCarry BE (2014) Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal. Bioanal. Chem.* 406, 3723–3733.
- Foster SL, Hargreaves DC & Medzhitov R (2007) Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* 447, 972–978.
- Franceschi C, Bonafè M, Valensin S, Olivieri F, De Luca M, Ottaviani E & De Benedictis G (2000) Inflamm-aging: An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* 908, 244–254.
- Ginaldi L, Loreto MF, Corsi MP, Modesti M & De Martinis M (2001) Immunosenescence and infectious diseases. *Microbes. Infect.* 3, 851–857.
- Gomez CR, Boehmer ED & Kovacs EJ (2005) The aging innate immune system. *Curr. Opin. Immunol.* 17, 457–462.
- Gordon S & Taylor PR (2005) Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* 5, 953–964.
- Green DR, Galluzzi L & Kroemer G (2011) Mitochondria and the Autophage-inflammation-cell death axis in organismal aging. *Science* 333, 1109–1112.

- Haskó G & Cronstein BN (2004) Adenosine: an endogenous regulator of innate immunity. *Trends Immunol.* 25, 33–39.
- Hibbs Jr JB, Vavrin Z & Taintor RR (1987) L-Arginine is required for expression of the activated macrophage effector mechanism causing selective metabolic inhibition in target cells. *J. Immunol.* 138, 550–565.
- Hotamisligil GS & Erbay E (2008) Nutrient sensing and inflammation in metabolic diseases. *Nat. Rev. Immunol.* 8, 923–934.
- Houtkooper RH, Argmann C, Houten SM, Cantó C, Jeninga EH, Andreux PA, Thomas C, Doenlen R, Schoonjans K & Auwerx J (2011) The metabolic footprint of aging in mice. *Sci. Rep.* 1, 134.
- Hurson M, Regan MC, Kirk SJ, Wasserkrug HL & Barbul A (1995) Metabolic effects of arginine in healthy elderly population. *JPEN J. Parenter. Enteral. Nutr.* 19, 227–230.
- Jha AK, Huang SC-C, Sergushichev A, Lampropoulou V, Ivanova Y, Loginicheva E, Chmielewski K, Stewart KM, Ashall J, Everts B, Pearce EJ, Driggers EM & Artyomov MN (2015) Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. *Immunity* 42, 419–430.
- Kaczor JJ, Ziolkowski W, Antosiewicz J, Hac S, Tarnopolsky MA & Popinigis J (2006) The effect of aging on anaerobic and aerobic enzyme activities in human skeletal muscle. *J. Gerontol. A. Biol. Sci. Med. Sci.* 61, 339–344.
- Kohut ML, Senchina DS, Madden KS, Martin AE, Felten DL & Moynihan JA (2004) Age effects on macrophage function vary by tissue site, nature of stimulant, and exercise behavior. *Exp. Gerontol.* 39, 1347–1360.
- Kuhl C, Tautenhahn R & Neumann S (2010) LC-MS Peak Annotation and Identification with CAMERA. , 1–14.
- Kurko J, Vähä-Mäkilä M, Tringham M, Tanner L, Paavanen-Huhtala S, Saarinen M, Nääntö-Salonen K, Simell O, Niinikoski H & Mykkänen J (2015) Dysfunction in macrophage toll-like receptor signaling caused by an inborn error of cationic amino acid transport. *Mol. Immunol.* 67, 416–425.
- Lee C-K, Klopp RG, Weindruch R & Prolla TA (1994) Gene expression profile of aging and its retardation by caloric restriction. *Science.* 3, 1290–1293.
- Lochmiller RL & Deerenberg C (2000) Trade-offs in evolutionary immunology : just what is the cost of immunity ? *OIKOS* 88, 87–98.

- Mahbub S, Deburghgraeve CR & Kovacs EJ (2012) Advanced Age Impairs Macrophage Polarization. *J. Interferon Cytokine Res.* 32, 18–26.
- Mills CD, Kincaid K, Alt JM, Heilman MJ & Hill AM (2000) M-1/M-2 Macrophages and the Th1/Th2 Paradigm 1. *J. Immunol.* 164, 6166–6173.
- Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdts S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T, Locati M, Mantovani A, Martinez FO, Mege J-L, Mosser DM, Natoli G, Saeij JP, Schultze JL, Shirey KA, Sica A, Suttles J, Udalova I, van Ginderachter JA, Vogel SN & Wynn TA (2014) Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* 41, 14–20.
- Nathan C (2002) Points of control in inflammation. *Nature* 420, 846–852.
- Newsholme P, Curi R, Curi TCP, Murphy CJ, Garcia C & de Melo MP (1999) Glutamine metabolism by lymphocytes, macrophages, and neutrophils: Its importance in health and disease. *J. Nutr. Biochem.* 10, 316–324.
- Norman DC (1996) Fever in the elderly. *Infect. Dis. Clin. North. Am.* 10, 148–151.
- Plowden J, Renshaw-Hoelscher M, Engleman C, Katz J & Sambhara S (2004) Innate immunity in aging: impact on macrophage function. *Aging cell* 3, 161–167.
- Rodríguez-Prados J-C, Través PG, Cuenca J, Rico D, Aragonés J, Martín-Sanz P, Cascante M & Boscá L (2010) Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* 185, 605–614.
- Ruiz-García A, Monsalve E, Novellademunt L, Navarro-Sabaté A, Manzano A, Rivero S, Castrillo A, Casado M, Laborda J, Bartrons R & Díaz-Guerra MJM (2011) Cooperation of adenosine with macrophage Toll-4 receptor agonists leads to increased glycolytic flux through the enhanced expression of PFKFB3 gene. *J. Biol. Chem.* 286, 19247–19258.
- Smith CA, Want EJ, O’Maille G, Abagyan R & Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787.
- Solana R, Pawelec G & Tarazona R (2006) Aging and innate immunity. *Immunity* 24, 491–494.
- Sugimoto M, Sakagami H, Yokote Y, Onuma H, Kaneko M, Mori M, Sakaguchi Y, Soga T & Tomita M (2011) Non-targeted metabolite profiling in activated macrophage secretion. *Metabolomics* 8, 624–633.

- Sun Y, Li H, Yang M-F, Shu W, Sun M-J & Xu Y (2012) Effects of aging on endotoxin tolerance induced by lipopolysaccharides derived from *Porphyromonas gingivalis* and *Escherichia coli*. *PLoS one* 7, e39224.
- Tannahill GM, Curtis AM, Adamik J, Palsson-McDermott EM, McGettrick F, Goel G, Frezza C, Bernard NJ, Kelly B, Foley NH, Zheng L, Gardet A, Tong Z, Jany SS, Corr SC, Haneklaus M, Caffrey BE, Pierce K, Walmsley S, Beasley FC, Cummins E, Nizet V, Whyte M, Taylor CT, Lin H, Masters SL, Gottlieb E, Kelly VP, Clish C, Auron PE, Xavier RJ & O'Neill LAJ (2013) Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature* 496, 238–242.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL & Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Weiskopf D, Weinberger B & Grubeck-Loebenstien B (2009) The aging of the immune system. *Transpl. Int.* 22, 1041–1050.
- Weng N-P (2006) Aging of the immune system: how much can the adaptive immune system adapt? *Immunity* 24, 495–499.
- Yeramian A, Martin L, Serrat N, Arpa L, Soler C, Bertran J, McLeod C, Palacín M, Modolell M, Lloberas J & Celada A (2006) Arginine transport via cationic amino acid transporter 2 plays a critical regulatory role in classical or alternative activation of macrophages. *J. Immunol.* 176, 5918–5924.
- Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstraß K, Wichmann H-E, Peters A, Suhre K, Li Y, Adamski J, Spector TD, Illig T & Wang-Sattler R (2012) Human serum metabolic profiles are age dependent. *Aging cell* 11, 960–967.
- Zheng J (2012) Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation. *Oncol. Lett.* 4, 1151–1157.

Supplementary information

GC-MS

For the targeted GC-MS analysis, metabolic intermediates of glycolysis, the citric acid cycle (TCA), the γ -aminobutyric acid (GABA) shunt and the urea cycle were included except acetyl-CoA, succinyl-CoA, succinyl-semialdehyde, ornithine and arginine (Table S1). The former three metabolites were excluded from the list due to limitations of GC-MS or lack of authentic standards. Ornithine and arginine were measured using LC-MS instead of GC-MS because arginine is partially converted to ornithine during trimethylsilylation step with MSTFA and 1% TMCS (Halket *et al.* 2005) (Fig. S1). The percentage of conversion is not linearly proportional to the arginine concentration. Therefore, it is important to analyze arginine and ornithine with LC-MS instead of GC-MS technique for better and reliable data quality.

TMCS is a silylation catalyst that increases the reactivity of other silylation reagents. 1% TMCS is important for trimethylsilylation derivation for GC-MS analysis for analyte with functional groups of low reactivity such as hydroxyl and amide groups. The reactivity of trimethylsilylation between functional groups and MSTFA follows the alcohol>phenol>carboxyl>amine>amide/hydroxyl order with primary amine/alcohol being more reactive than tertiary. Multiple adducts with different number of trimethylsilyl ester derivatives can be observed in the study for one metabolite (e.g. GABA). Multiple peaks can also be noted for glucose from direct derivitization due to varying structural forms such as cyclic or open-chain structure (Fiehn *et al.* 2000).

All targeted metabolites were found to undergo continuous derivitization while queued on the autosampler at ambient temperature. Cooling the autosampler to below 5°C quenched the ongoing derivitization for most targeted metabolites. Therefore, it is critical to analyze the macrophage extract by GC-MS immediately after derivitization. When not possible, the derivatized sample was stored in -20 or -80°C until GC-MS analysis, for which no further derivation was noted.

Anthracenemethanol was selected as internal standard for GC-qMS study because it can be detected in derivatized and non-derivatized form. In addition to be used for peak area normalization, it is used to monitor the completeness of the derivation procedure by monitoring the ratio between m/z 208 and m/z 280.

Table S1 Target Metabolites for GC-qMS analysis

No.	Metabolites	Retention time (min)	Quantifying ion (m/z)	Qualifying ion (m/z)
1	lactate	5.39	147	73, 117, 219
2	alanine	5.68	116	73, 147
3	pyruvate	5.80	217	73, 147
4	γ -aminobutyric acid (GABA-diTMS)	9.36	102	73, 147, 232
5	proline	9.36	142	73, 216
6	fumarate	10.21	245	73, 147
7	succinate	10.53	147	73, 247
8	GABA-triTMS	13.63	174	73, 147, 304
9	malate	13.69	233	73, 147, 335
10	methionine-d ₃ - monoTMS	14.09	107	73, 224
11	aspartate	14.31	232	73, 100, 147
12	Oxaloacetate-1	15.38	333	73, 147
13	Oxaloacetate-2	15.72	333	73, 147
14	ornithine-triTMS	15.46	142	70, 73, 348
15	glutamate	16.53	246	73, 128, 147, 156
16	α -ketoglutarate	16.78	347	73, 147, 318
17	phenylalanine-d ₈	17.39	219	73, 147
18	ornithine-tetraTMS	19.06	142	73, 174, 420
19	glutamine	20.69	156	73, 147, 245
20	citrate	20.74	273	73, 147, 465
21	Arginine ^a	20.86	256	73, 147, 157
22	glucose-1	20.96	204	73, 147, 191, 217
23	glucose-2	22.42	204	73, 147, 191, 217
24	2-Phosphoglyceric acid (2-PG)	21.14	299	73, 147, 387, 459
25	isocitrate	21.16	273	73, 147, 245, 465
26	3-Phosphoglyceric acid (3-PG)	21.61	299	73, 357, 387, 459
27	citrulline-tetraTMS	24.36	188	73, 100, 346
28	fructose-6-phosphate (F6P)	28.30	315	73, 387, 589
29	tryptophan-d ₅	29.70	207	73, 291
30	glucose-1-phosphate (G1P)	30.94	387	73, 204
31	Glucose-6-phosphate	31.14	387	73, 315, 589
32	anthracenemethanol	31.38	280	
33	argininosuccinate	35.94	149	73, 563, 578

Table S2 Single-ion-monitoring (SIM) method for GC-qMS

SIM region	Time (min)	Ions (m/z)
1	5.0-7.0	73, 116, 117, 147, 217, 219
2	7.0-10.0	73, 102, 142, 147, 216, 232
3	10.0-11.5	73, 147, 245, 247
4	11.5-14.9	73, 100, 107, 147, 174, 224, 232, 233, 304, 335
5	14.9-16.0	70, 73, 142, 147, 157, 173, 179, 218, 296, 333, 348
6	16.0-18.0	73, 128, 147, 156, 200, 219, 246, 318, 347
7	18.0-20.0	73, 142, 174, 420
8	20.0-23.5	73, 147, 156, 157, 191, 204, 217, 245, 256, 273, 299, 357, 387, 459, 465
9	23.5-26.0	70, 73, 100, 188, 346
10	26.0-30.5	70, 73, 100, 204, 207, 291, 315, 387, 589
11	30.5-33.0	73, 204, 208, 280, 387
12	33.0-40.0	73, 149, 563, 578

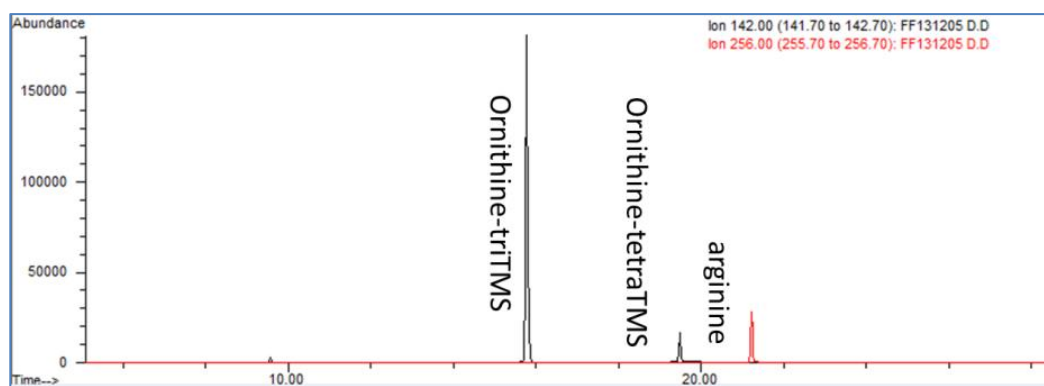


Figure S1 The GC-EI-MS chromatogram of arginine standard

Data quality control

The final comprehensive murine macrophage metabolic profile data set acquired by LC-MS consisted of 2125 metabolite features, of which 121 metabolites were identified. The extraction efficiency was 76-83% based on L-methionine-d₃, a standard for recovery determination (RS). The biological variance was 20-35% and 29-40% for macrophages derived from young or old mice, respectively (Table S3). The biological variance was consistently greater for macrophages derived from old mice for all treatment conditions. Post column addition using gly-phe had shown that there was no ion suppression noted in both ESI+ and ESI- modes except between 0.6-1.3 min and 7.1-7.8 min of the chromatogram. Features eluted between 0.6-1.3 were removed during data reduction process. Features eluted between 7.1-7.8 were normalized to L-phenylalanine-d₈, which

also eluted in the same region, to correct for ion suppression. The comprehensive data set was visualized by principle component analysis (PCA) to assess the LC-MS instrumentation reproducibility. The pooled samples were tightly clustered in the centre of the score plot, which indicating the technical variability was minimal in comparison to the biological variance and the pooled sample resembled the average of all samples from various treatment conditions. Of the comprehensive data set, 69.9% of the features were normally distributed; the metabolites included in the targeted data set were all normally distributed.

Table S3 Coefficient of variance of macrophage metabolites from young or old mice.

Treatment groups	Treatments	macrophages	
		Young mice	Old mice
Group 1	4 hr LPS	35%	40%
Group 2	16 hr LPS	35%	40%
Group 3	Tolerance	23%	43%
Group 4	Recovery	20%	29%
Group 5	Control 1	30%	40%
Group 6	Control 2	32%	43%

Table S4 Summary of total number of metabolite features extracted from LC-MS data after XCMS, CAMERA and data reduction

Total No. features	ESI +	ESI -
XCMS	2060	2224
Total	4284	
$k' < 0.7^a$	198	125
IS/RS	5	5
Salts (sodium formate)	14	14
Percentage variance >30% in pooled samples	596	
Finalized feature list	2125	

^a Retention factor $k' = \frac{t_R - t_0}{t_0}$, t_R is the retention time of metabolic feature and t_0 is the time required for mobile phase to pass through the column

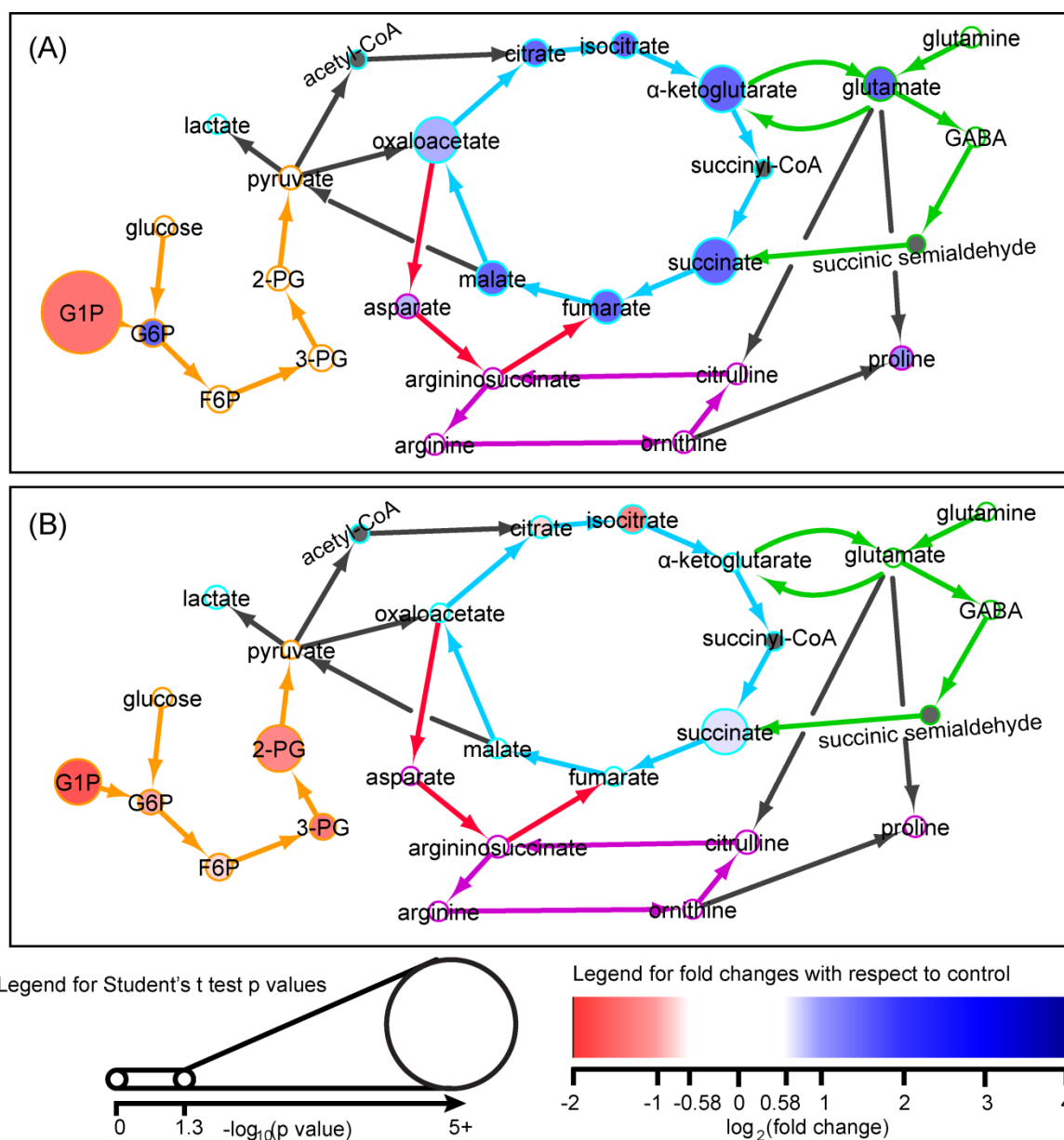


Figure S2 The enrichment map of glycolysis, the TCA cycle, the GABA shunt, and the urea cycle intermediates during 4 hrs of LPS stimulation in bone marrow-derived macrophages from (A) young mice or (B) old mice as compared to corresponding unstimulated macrophages. The network of metabolite interactions was built based on the BioCyc database and pathway published from Jha *et al.* (Jha *et al.* 2015). The node size is proportional to the significance of metabolite changes compared to the control. The colors of the nodes indicate the $\log_2(\text{fold change})$ of metabolite levels of each experimental condition compared to the control with a decrease colored in red and an increase colored in blue. Acetyl-CoA, succinyl-CoA and succinic semialdehyde are not detected and

therefore are labelled in grey. The glycolysis pathway is labelled in orange; the TCA cycle is labelled in blue; the GABA shunt is labelled in green; the glutamate-argininosuccinate shunt is labelled in red; the urea cycle is labelled in purple.

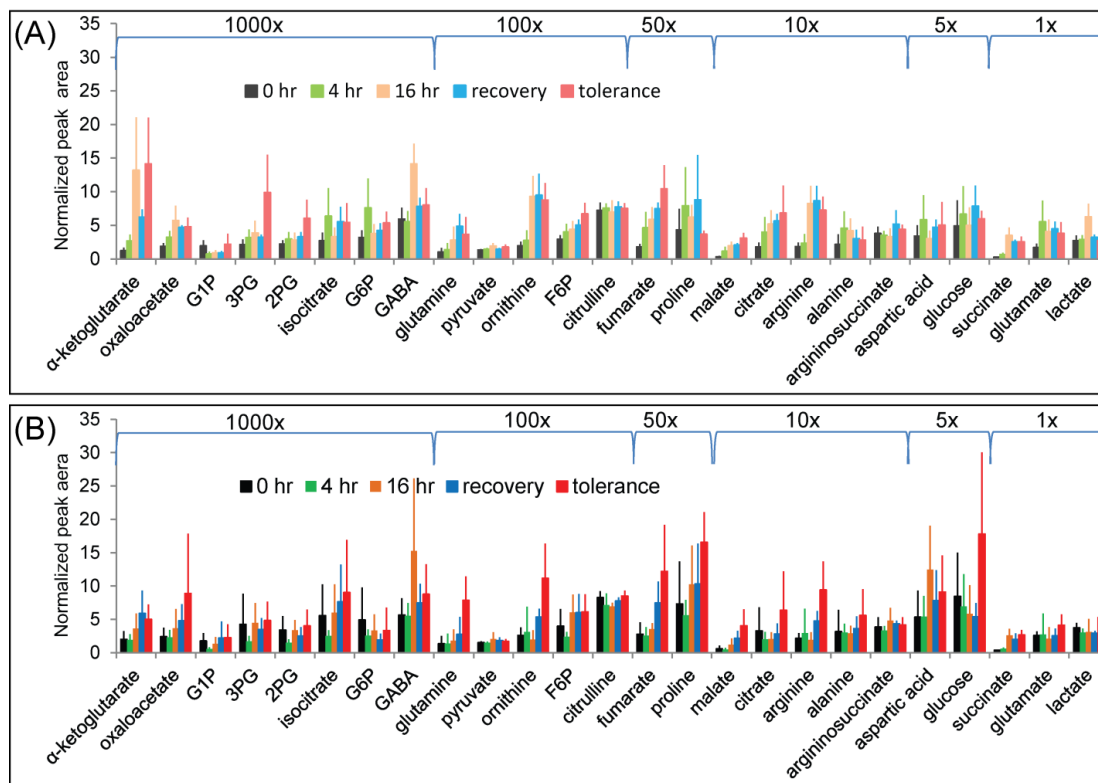


Figure S3 The relative abundances of the 25 targeted intracellular metabolites from macrophages of (A) young and (B) old mice from five different experimental conditions as acquired by GC-MS. The samples were performed in sextuplicate with three biological replicates and two culture replicates. The error bar represents one standard deviation. The relative abundances of some metabolites were amplified by 5, 10, 50, 100, or 1000 fold to scale to the figure. G1P, glucose-1-phosphate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; G6P, glucose-6-phosphate; GABA, γ -aminobutyric acid

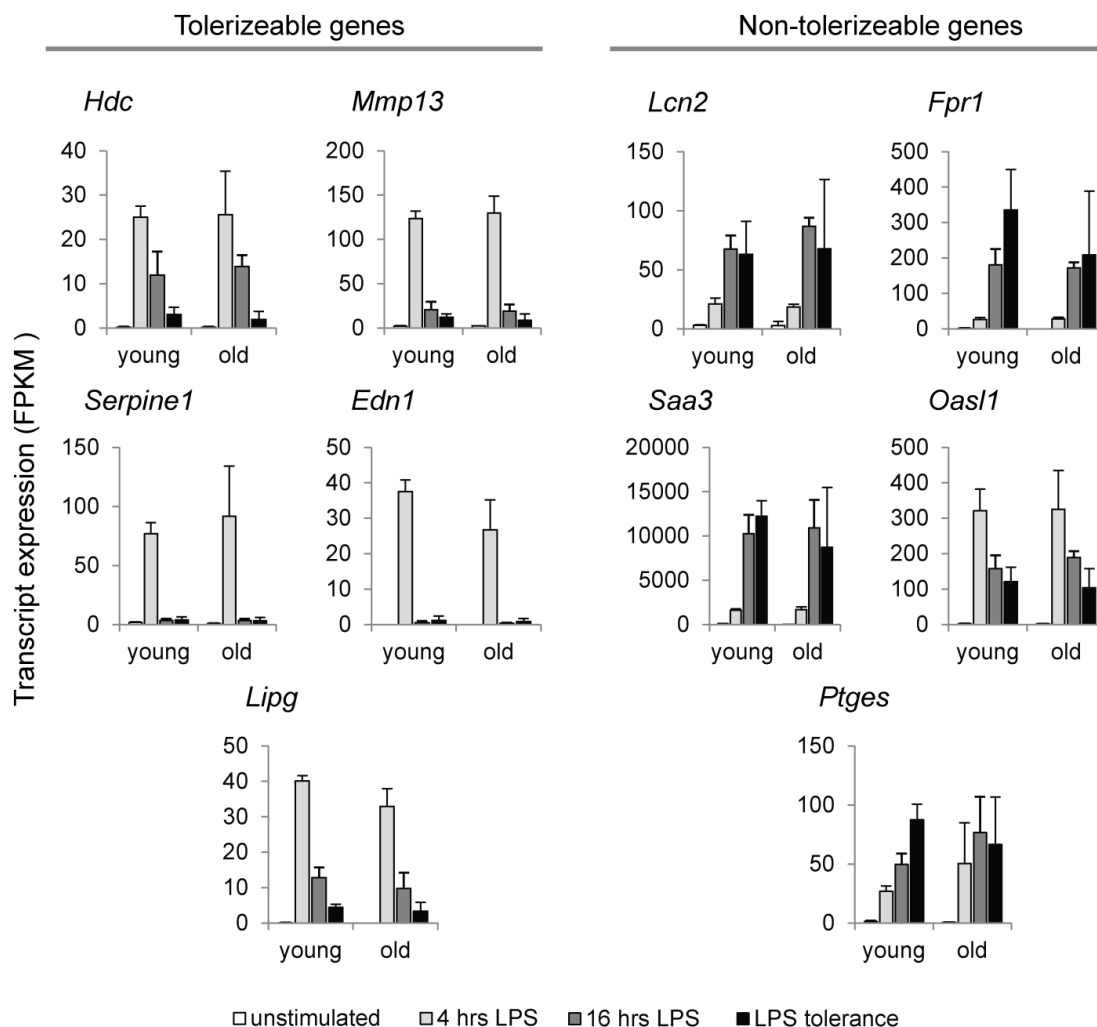


Figure S4 The gene expression of tolerance and non-tolerance associated genes identified by Foster *et al.* (Foster *et al.* 2007). The expressions of these genes were acquired from unstimulated (0 hr), 4 hrs and 16 hrs of LPS stimulation, and LPS tolerant macrophages from either young or old mice.

References

Fiehn O, Kopka J, Trethewey RN & Willmitzer L (2000) Identification of uncommon plant metabolites based on calculation of elemental compositions using Gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* 72, 3573–3580.

- Foster SL, Hargreaves DC & Medzhitov R (2007) Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* 447, 972–978.
- Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD & Bramley PM (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of experimental botany* 56, 219–43.
- Jha AK, Huang SC-C, Sergushichev A, Lampropoulou V, Ivanova Y, Loginicheva E, Chmielewski K, Stewart KM, Ashall J, Everts B, Pearce EJ, Driggers EM & Artyomov MN (2015) Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. *Immunity* 42, 419–430.

Chapter 6: Comprehensive Metabolomic Analysis of *Sinorhizobium meliloti* responses to sub-lethal PAH exposures

Fan Fei^{1,2}, *Elna D. Luckham*¹, *Dawn M. E. Bowdish*², *Brian E. McCarry*¹,

Author Contributions

BEM conceived the study. FF, EDL and BEM designed the study. FF and EDL cultured the bacterial strain, extracted cells, and acquired the metabolomic data. FF analyzed and interpreted the data, and drafted the manuscript. FF and DMEB edited the manuscript.

Author Affiliations:

¹ Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Canada L8S4M1

² Department of Pathology and Molecular Medicine, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada L8N3Z5

Abstracts

Polycyclic aromatic hydrocarbons (PAHs) are environmental pollutants that can lead to carcinogenic, mutagenic or dioxin-/furan-like effects in many animal species. The toxicities of single and binary mixtures of PAH have been studied at concentrations close to lethal levels and evaluated based on physiological changes in animal models. However, few studies have looked at the biological effects of PAH exposure at sub-lethal doses (i.e. at levels 10 to 100-fold less than lethal doses). In this study, we have exposed cultures of the soil bacterium, *Sinorhizobium meliloti*, to a single PAH and to a mixture of PAH derived from coal tar at sub-lethal concentrations in order to determine the metabolomic impacts of these exposures. This is the first comprehensive metabolomic analysis of the impacts of sub-lethal exposures of PAH in a bacterial model.

The ubiquitous Gram negative soil bacteria, *Sinorhizobium meliloti*, was exposed to fluorene and a low molecular weight PAH fraction (MW 128-202) derived from coal tar sample at two concentrations, 0.14 and 1.4 mg/L. The polar metabolome was extracted using the Bligh and Dyer method and analyzed by ZIC-HILIC-LC-MS. Although only the 1.4 mg/L PAH mixture impaired *S. meliloti* proliferation *in vitro*, changes in metabolic profiles were observed after exposure to 1.4 mg/L fluorene, and 0.14 mg/L and 1.4 mg/L PAH mixture exposures. Dose dependent changes in cellular metabolism were observed for both sub-lethal fluorene and PAH mixture exposure. Amplified metabolic responses were observed in PAH mixtures compared to exposures to an equivalent concentration of fluorene.

Introduction

Polycyclic aromatic hydrocarbons (PAH) are ubiquitous environmental contaminants formed by incomplete combustion of organic compounds from both natural and anthropogenic sources¹. Soil is a great reservoir for organic pollutant, and serves as a major reservoir for PAH². It is estimated that 90% of the 53,000 tons of total PAH were deposited in the soil in United Kingdom in 1995³. The accumulation of PAH in soil has a negative effect on the soil eco-system.

Coal tar is the by-product condensate from coke production, where coal is carbonized at high temperatures in a reductive environment¹. Coal tar contains PAHs, nitrogen containing polycyclic aromatic hydrocarbons (NPAHs) and polycyclic aromatic sulfur heterocycles (PASHs), many of which are known human carcinogens and bacterial mutagens⁴. The current US Environmental Protection Agency (EPA) approved method to evaluate PAH toxicity uses *Daphnia magna* based on mortality⁵. The sub-lethal effects of PAH and PAH mixtures on organisms are not well understood.

Much of the PAH toxicity work to date has been focused on the effect of a single compound or binary mixtures of compounds⁶. The biological effects of mixture exposures have been a long-standing research interest. The total effects of mixture depends on the mixture constituents, their concentrations and their composition ratio⁷. To date, the mixture effects of PAH have been studied in rats⁸, mice⁹, and *Daphnia*⁷ based primarily on transcriptomic, and physiological changes. Overall mixture effects included antagonistic, additive and synergistic effects (joint independent action) at lethal and sub-lethal concentrations.

Sinorhizobium meliloti is a Gram-negative, nitrogen-fixing bacterium that is found in soils. *S. meliloti* has been the subject of extensive genetic, biochemical, and targeted metabolic research. As a model organism, its genome has been sequenced in 2001¹⁰ and a novel reporter gene fusion library of about 2200 of the 6500 genes was constructed in 2006¹¹. Comprehensive metabolomic profiling has shown that *S. meliloti* undergoes oxidative stresses as a result of exposure to phenanthrene¹². The impacts of exposure to complex environmental mixtures using wild-type *S. meliloti* (RmP110) has not been explored.

As a ubiquitous soil organism, *S. meliloti* is exposed to PAH contamination in the soil. Here, we use *S. meliloti* to measure the toxicity between a single PAH compound and a low molecular weight PAH mixture derived from coal tar. With this study, we optimized an alumina chromatography method for obtaining NPAH free PAH fractions containing PAHs with MW 128-202. This coal tar derived PAH mixture had the same mode-of-action as a single PAH compound, fluorene, however, the PAH mixture demonstrated greater metabolomic disturbances compared to fluorene at equal concentrations.

Materials and Methods

Chemicals and growth media

GC-ResolvTM hexanes was purchased from Fisher Scientific (Fairlawn, New Jersey). HPLC grade methanol, toluene, acetonitrile, water, formic acid and distilled in glass dichloromethane were purchased from Caledon Laboratories (Georgetown, ON). PAH standards were received from various commercial sources. Pyrene-d₁₀ (98%) was purchased from Cambridge Isotope Laboratories (Andover, MA), which was used as internal standard for GC-MS analyses. The coal tar sample was kindly provided by a local steel company and was used as is. Ammonium acetate was purchased from Fisher Scientific Company (Fairlawn, NJ, USA). L-methionine-d₃ (98%), L-tryptophan-d₅ (98%) and L-phenylalanine-d₈ (98%), diphenylalanine (phe-phe), glycine-phenylalanine (gly-phe) were purchased from Cambridge Isotope Laboratories (Andover, MA, USA) for recovery determination (RS) and peak intensity normalization (IS), respectively. LBmc (per liter: 10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl, 2.5 mM MgSO₄, 2.5 mM CaCl₂, 2 µM CoCl₂) was used as the complex medium, and M9-sucrose (41 mM Na₂HPO₄, 22 mM KH₂PO₄, 18.7 mM NH₄Cl, 10 mM sucrose, 8.6 mM NaCl, 1 mM MgSO₄, 0.25 mM CaCl₂, 38 µM FeCl₃, 5 µM thiamine-HCl, 4.1 µM biotin, 42 nM CoCl₂) was used as minimal medium.

Coal tar fractionation using alumina chromatography

The alumina chromatography procedure was a modification of the procedure described by Later *et al.*¹³ which had been subsequently modified by Li¹⁴. The standard protocol used neutral alumina which had been activated at 170°C for at least 48 hours (Activity I). In order to obtain a fraction containing only low molecular weight PAH for this toxicity study, the alumina chromatography for coal tar clean-up was further optimized according to the experimental setup in Table 1.

Table 1 Alumina chromatographic optimization conditions for coal tar separation

Experiment number	Load ^a	Bed
C1	neutral, activity I (170°C)	neutral, activity I (170°C)
C2	neutral, activity IV	neutral, activity I (170°C)
C3	neutral, activity I (170°C)	pH4.5, activity I (170°C)
C4	neutral, activity IV	pH4.5, activity I (170°C)
C5	neutral, activity IV	pH3.0, activity I (185°C)

^a activity I alumina is made by incubating neutral or acidic alumina in 170 or 185°C oven for 48 hrs; activity IV alumina is made by adding 10%v/w HPLC grade H₂O to activity I alumina

Acidic alumina (pH 4.5) was prepared from neutral alumina by adding 19 drops (ca. 0.88 mL) of concentrated HCl dropwise to a slurry composed of 20 g of alumina and 200 mL of water. The slurry was stirred until reaching pH 4.5, and then evaporated to dryness using a rotary evaporator. Similarly, acidic alumina of pH 3.0 was made by adding 41 drops (ca. 1.9 mL) of concentrated HCl to the alumina slurry.

Activity I alumina was made by incubating neutral or acidic alumina (80-200 mesh, activity I, Anachemia) in a 170 or 185 °C oven for 48 hours. Activity IV alumina is made by adding 10%v/w water to the activity I alumina, the mixture was shaken until alumina was free flowing, and allowed to equilibrate in sealed containers for 30 min prior to use.

The crude coal tar was separated on alumina chromatography into four major fractions: A1, aliphatic hydrocarbons; A2, PAHs and PASHs; A3 and A4, NPAHs. To prepare the “sample load”, 1 g of crude coal tar was dissolved in ca. 10 mL dichloromethane and adsorbed onto 4 g of neutral activity I/IV alumina. The solvent was removed using rotary evaporator. The “sample load” was then packed on top of a 2 cm i.d. column with 20 g of neutral or acidic activity I alumina. The column was packed under hexanes flow, and a 1-1.5 cm sand layer was placed between the “sample load” and “column bed”. This procedure resulted in a ca. 12 cm long section of white alumina and a 3 cm long section of blackish, coal tar-coated alumina.

The coal tar sample was eluted with four solvents sequentially at a flow rate of ca. 5-8 mL/min: fraction A1, 80 mL hexanes; fraction A2, 250 mL toluene; fraction A3, 230 mL dichloromethane; fraction A4, 150 mL methanol. Sub-fractions of A2 was collected every 25 mL for the first 100 mL of toluene eluent and every 50 mL for 100-250 mL toluene eluent. Only for C5, the sub-fractions of A2 were collected according to Table 2. A1 fraction and A2 sub-fractions were diluted and analyzed by gas chromatography mass spectrometry (GC-MS). A3 and A4 fractions were evaporated to dryness under nitrogen gas, reconstituted in toluene and then analyzed with GC-MS. Pyrene-d₁₀ was added as an internal standard to each sample for a total concentration of 1 ng/μL.

Table 2 A2 toluene sub-fractions collected in optimized experiment C5

Sub-fractions	Volumes of eluent collected (mL)
A2-1	0-10
A2-2	11-15
A2-3	16-20
A2-4	21-25
A2-5	26-35
A2-6	36-50
A2-7	51-65
A2-8	66-80
A2-9	81-95
A2-10	96-120
A2-11	121-250

PAH profiling using GC-MS

PACs (polycyclic aromatic compounds i.e. PAHs, PASHs, NPAHs and corresponding methyl derivatives) in each of the coal tar fractions were analyzed on Agilent 6890N gas chromatograph (Agilent Technologies, Santa Clara, CA, USA), equipped with a DB-17ht column (30 m × 0.25 mm i.d. × 0.15 µm film, J & W Scientific) and a retention gap (deactivated fused silica, 5 m × 0.53 mm i.d.), coupled to an Agilent 5973 MSD single quadrupole mass spectrometer. The GC injector was a split-splitless injector operated at 230 °C in the splitless mode; one microliter volumes were injected using an Agilent 7683 autosampler. The carrier gas was helium with a flow velocity of 29 cm/s; the flow velocity was kept constant using electric pressure control. The temperature in the GC oven started at 90° C and increased linearly to 300°C at 2.5°C/min, followed by a hold at 300°C for 20 min; the total run time was 104 min. The same temperature program was used for the analysis of fraction A1 except that the initial oven temperature was 40°C. Full scan mass spectra were collected over with mass-to-charge (m/z) range from 100 to 350.

PAHs, PASHs, and NPAHs and their methyl-derivatives were identified based on comparison of their mass spectra to the mass spectra of reference standards. Table 3 provides a list of the generic ions used to identify and quantify PACs. In general, molecular ions were used as quantifying ions while major fragment ions were used as qualifying ions. Many PAH and PASH were identified based on the retention indexes of authentic standards.

Table 3 Qualifying and quantifying ions for PAH, PASH and NPAH, and their methylated derivatives using GC-MS. The molecular ions were used to quantify their levels.

Ions	PAH	Me-PAH	PASH	Me-PASH	NPAH	Me-NPAH
M	+++	+++	+++	+++	+++	+++
M-1		++		++		++
M-26 ^a	+					
M-27 ^b		+		+	+	
M-28					+	+
M-32 ^c			+			
M-33				+		
M-45 ^d			+			
M/2	++	+	+	+		
(M-1)/2					+	+

^a M-26 corresponds to loss of C₂H₂ or HCN

^b M-27 corresponds to loss of C₂H₂ and H[·] or HCN and H[·]

^c M-32 corresponds to loss of S

^d M-45 corresponds to loss of CHS[·]

The +, ++, +++ signs were referring to the relative ion intensities, with +++ being the greatest and + being the least intensive

***S. meliloti* culture and exposure to PAH**

Sinorhizobium meliloti RmP100 (wildtype strain) were used to examine the sub-lethal metabolomic impacts of single or complex mixture of PAH. *S. meliloti* were cultured to early stationary phase in 5 mL LBmc at 30°C overnight. The culture was pelleted, washed once with fresh saline (0.85% NaCl), and resuspended in 5 mL fresh M9 medium of 0.05 OD₆₀₀. 50 µL of fluorene and low molecular weight PAH mixture in methanol (1% v/v) were spiked in the *S. meliloti* culture so that the total PAH concentration was 0.14 or 1.4 mg/L. The PAH exposed cultures along with un-exposed controls were incubated for 24 hrs at 30°C. The final OD₆₀₀ were measured before cultures were harvested for metabolomic analysis (Table 4). Each treatment was repeated in six culture replicates. The low molecular weight PAH mixture was obtained from A2-2 sub-fraction of optimized C5 experiment, and the concentration of each PAH and PASH in this mixture were listed in Table 5.

Table 4 The OD₆₀₀ of control, 0.14 mg/L fluorene, 1.4 mg/L fluorene, 0.14 mg/L PAH mixture, and 1.4 mg/L PAH mixtures exposed *S. meliloti* culture at the time of the harvest

	PAH Exposure ^a	Replicates	OD ₆₀₀	Student's t test p value (treated compared to control)
Control	---	6	0.63 ± 0.06	---
0.14Flu	0.14 mg/L fluorene	6	0.64 ± 0.06	0.68
1.4Flu	1.4 mg/L fluorene	6	0.61 ± 0.10	0.76
0.14PAH	0.14 mg/L PAH mix	6	0.64 ± 0.04	0.63
1.4PAH	1.4 mg/L PAH mix	5	0.41 ± 0.04	0.0001

^a1% v/v methanol in all treatments

^bsee Table 5 for the concentration and composition of the PAH mix

Table 5 The concentrations of low molecular weight PAH mixtures used for *S. meliloti* metabolomics study. The sample was collected and diluted from C5 A2-2 fraction

	Sample 1(mg/L)	Sample 2 (mg/L)	Percentage weight
Total PAH	1.40	0.14	
naphthalene	0.0374	0.00374	2.4%
acenaphthylene	0.131	0.0131	8.5%
biphenyl	0.260	0.0260	17.0%
acenaphthene	0.0418	0.00418	2.7%
fluorene	0.226	0.0226	14.8%
phenanthrene	0.605	0.0605	39.6%
anthracene	0.0307	0.00307	2.0%
fluoranthene	0.0134	0.00134	0.9%

pyrene	0.0547	0.00547	3.6%
Me-PAH	0.0884	0.00884	5.8%
NPAH	0.00522	0.000522	0.3%
Me-NPAH	0	0	0
PASH	0.0357	0.00357	2.3%

***S. meliloti* extraction**

The polar metabolites of *S. meliloti* were extracted using the Bligh and Dyer method as outlined in Fei *et al.*¹⁵. Briefly, the bacterial cultures were pelleted and washed once in cold saline solution. The bacterial pellet was resuspended in 250 µL of 1:1 methanol to water (v/v) and 250 µL of chloroform in a 1 mL eppendorf tube. 10 µL of RS was added into the extracts with a 10 µL syringe. The bacterial extract was vortex mixed for 2 minutes and centrifuged at 4°C at 13000 rpm for 1 minute using an Eppendorf Refrigerated Microcentrifuge (Model 5415R). The methanol/water layer was collected; the protein film and chloroform layer was extracted twice with 250 µL of 1:1 methanol-water each time to make up a total of 750 µL of combined polar fraction. The combined polar fraction was back-extracted with 250 µL of chloroform twice to create a combined non-polar fraction of 750 µL. The non-polar fractions were blown down and stored in the -20°C freezer, and not analyzed in this experiment. A 10 µL IS was added to the combined polar fraction of 750 µL using a 10 µL syringe, which was stored in -80°C until metabolomic analysis using liquid chromatography mass spectrometry (LC-MS).

ZIC-HILIC-TOF-MS analysis

The polar metabolome of PAH exposed *S. meliloti* were analyzed using an Agilent Technologies 1200 RR Series II LC coupled to a BrukerMicroTOF II Mass Detector operated in the both positive (ESI+) and negative (ESI-) electrospray ionization modes. A 2 µL injection was separated on a 50 mm × 2.1 mm ZIC®3.5 µm HILIC Sulfonyletamine column with pore size of 100 Å (SeQuant, Umeå, Sweden). The mobile phases consisted of acetonitrile (A) and 10 mM ammonium acetate in water and titrated to pH 3 using formic acid (B) at a flow rate of 400 µL/min. The column temperature was maintained at 25 °C, and the auto sampler storage tray was set at 4°C. LC gradient: 0-3 min, 95% A to 90% A; 3-13 min, 90% A to 25% A; 13-16 min, hold at 25% A; 16-17 min, 25% A to 95% A; 17-22 min, equilibrate at 95% A. A pooled sample was prepared by combining equal aliquots of five samples, with one from each treatment group. Pooled samples were injected four times prior to the sample analysis, and also run after every 7-8 samples. The samples were run in a randomized sequence. The MS settings were identical to those previously reported in Fei *et al.*¹⁵.

Metabolomic data processing

The LC-MS data were processed as in Fei *et al.*¹⁵. The LC-MS spectra were converted to .mzXML format using Bruker CompassXport. The metabolic features were extracted and aligned using XCMS with centWave algorithm¹⁶; adducts, isotopic ions, and in-source fragments were identified using CAMERA¹⁷. The peak area of metabolic features was normalized with OD₆₀₀ and IS (features eluted before 6.40 min were normalized by phe-phe; features eluted between 6.40 and 7.40 min were normalized by L-phenylalanine-d₈; features eluted after 7.40 min were normalized by gly-phe). The data was reduced by removing metabolic features that eluted prior to 1.0 min (ion suppression region) and after 16 min and also isotopic ions, IS, RS and sodium formate ions. Metabolic features with greater than 30% variance in the pooled samples were also removed¹⁸. The final data set consisted of 2005 metabolic features.

Statistical analysis

The polar metabolome of PAH exposed *S. meliloti* was analyzed using both univariate and multivariate statistical analyses. The final data set was analyzed using principle component analysis (PCA) and orthogonal partial least-squares discriminative analysis (OPLS-DA) after pareto scaling using SIMCA-P+11 (Umetircs, Kinnelon, NJ). Heat maps were generated with R Project 2.12.2 using gplots. Metabolic features with two-tailed, unpaired heteroscedastic Student's t tests of $p < 0.05$ and a fold change greater than 1.5 compared to control were considered significantly expressed.

Results

Optimizing crude coal tar separation using alumina chromatography

The alumina chromatography for separation of crude coal tar was optimized in C1-4 experiments by adjusting activity and pH of the alumina (Fig 1, 2). The crude coal tar clean-up aimed to have a complete separation between PAHs and NPAHs, and also acquire a low molecular weight PAH fraction to study toxicity on the soil bacteria *S. meliloti*.

The alumina chromatography protocol for crude coal tar clean-up ascribed by Later *et al.*¹³ and Li¹⁴ was modified and repeated as in experiment C1. C1-A1 consisted of aliphatic hydrocarbons and 4.9% of the total detectable PAH with molecular weight of 128-202 Da. PASHs were completely eluted in C1-A2 fractions. Coeluting of the high molecular weight PAHs and low molecular weight NPAHs was noted in the C1-A2 fraction after 50 mL of toluene eluent (ca. 7 column volumes). The elution of residual PAHs was also observed in C1-A3 and C1-A4 fractions. However, C1-A3 and C1-A4 was dominated by NPAHs and methyl-NPAHs. In the C1 experiment, PAHs and NPAHs were not clearly separated between fractions.

In the C2 experiment, the activity of coal tar adsorbed alumina was reduced from activity I to activity IV. Almost all PAHs, methyl-PAHs and PASH were eluted in the C2-A2 fraction, with majority of the PAHs and PASHs eluted in the first 50 mL of toluene. NPAH and methyl-NPAH eluted 3.5 column volumes earlier in the C2-A2

fraction compared to the C1 experiment, which eluted after the first 25 mL of toluene. The use of activity IV alumina instead of activity I alumina in the “sample load” has sped up the eluting process of both PAHs and NPAHs.

Acidic alumina of pH 4.5 was used as the packed bed in experiment C3 and C4. A clear separation between PAHs and NPAHs were achieved. PAHs, methyl-PAHs and PASH were eluted in the A2 fractions; NPAHs and methyl-NPAHs were almost exclusively eluted in A3 and A4 fractions. Lowering the pH of alumina in the packed column bed retained NPAH until more polar eluents such as dichloromethane or methanol were used. Similar to the changes between the C1 and C2 experiments, reducing the activity of coal tar adsorbed alumina in C4, the majority of the PAHs, methyl-PAHs and PASHs were eluted in the first 50 mL of the toluene eluent. PAHs with molecular weight equal or less than 228 Da were mostly eluted in the C4-A2-1 sub-fraction within the first 25 mL of toluene, and those with molecular weight equal or greater than 252 were eluted in the C4-A2-2 fraction. Less bleeding of PAH was noted in the C4 experiment in the C4-A2 sub-fractions.

Optimized alumina chromatography for crude coal tar separation was conducted in experiment C5 (Fig 3a, b). Combining the usage of activity IV alumina in the “sample load” and acidic activity I alumina (pH 3.0), 99.7% of the PAH, 99.9% of methyl-PAH and 100% of PASHs were eluted in the C5-A2 fraction. No PAHs were eluted in the C5-A1 fraction. The C5-A2-3 sub-fraction (16-20 mL toluene) had the most PAHs and contributed 48.7% of the total detectable PAH mass in crude coal tar. Only 0.9% of the NPAH and 5.9% methyl-NPAH were eluted in the first nine column volume (65 mL toluene, sub-fraction A2-1 to A2-7) of C5-A2 fraction. 98.8% of the NPAHs eluted in the combined C5-A2-11 (121-250 mL toluene) and the more polar C5-A3 and C5-A4 fractions. The C5-A2-2 fraction contained 85.2% of total detectable low molecular weight PAHs (128-202 Da), 10.8% of methyl-PAHs, 3.41% of PASH, and 0.5% NPAH (Fig 3c). Additionally, phenanthrene was the most abundant PAHs followed by acenaphthene and fluorene, which contributed 40.6%, 13.9% and 9.1% of the mass in C5-A2-2, respectively. The C5-A2-2 sub-fraction was later used for the *S. meliloti* toxicity study.

Metabolomic analyses, features detection and quality control

The polar metabolome of PAH exposed *S. meliloti* were acquired using ZIC-HILIC-LC-MS. A total of 3861 features were identified using XCMS and CAMERA. Each metabolic feature was defined with unique combination of retention time and m/z values. With data reduction, 2005 metabolic features were included in the final data set. The extraction efficiency was 84-103% based on L-tryptophan-d₅ as recovery standard, and the biological variance of the sextuplicate samples were between 24-32%. The coefficient of variance for pooled samples was 15%. The pooled samples were tightly clustered in the center of the PCA score plot, indicating minimal technical variance as compared to the biological variance experienced with treated *S. meliloti* samples (Fig 4).

***S. meliloti* metabolic disturbance due to PAH exposure**

To compare the potency and the toxic mechanisms between the single PAH compound (fluorene) and a mixture of low molecular weight PAH fractionated from crude coal tar, we examined the growth and the polar metabolome of exposed *S. meliloti*. Fluorene and the PAH mixture were spiked into the growth medium of *S. meliloti* at 0.14 and 1.4 mg/L. A decrease in growth rate was only observed in the *S. meliloti* culture exposed to 1.4 mg/L PAH mixture (Table 4). As visualized by OPLS-DA score plot, no metabolic differences were observed between the un-exposed controls and *S. meliloti* exposed to 0.14 mg/L fluorene ($R^2X=0.61$, $R^2Y=0.27$, $Q^2=0.11$; data not shown). Only four metabolic features were significantly different between the 0.14 mg/L fluorene exposure samples to control. However, changes in metabolic profiles were observed in all other treatments, where *S. meliloti* were exposed to either 1.4 mg/L fluorene or the PAH mixture at both doses (Fig 5).

Both 0.14 mg/L of the PAH mixture and the 1.4 mg/L fluorene caused similar metabolic shifts in *S. meliloti*, and the metabolic profiles of these samples were different than that of the un-exposed controls. From the heatmap (Fig 5b), almost all of the biological samples with 1.4 mg/L fluorene were affected, however, only a few of the samples exposed to 0.14 mg/L PAH showed a metabolic shift. Subsequently, 464 metabolic features were significantly expressed between 1.4 mg/L fluorene exposed *S. meliloti* and the un-exposed controls, but only 77 metabolic features were differentially expressed between 0.14 mg/L PAH samples and the un-exposed controls. The intracellular levels of most of these significant features were elevated with the PAH exposure compared to the un-exposed control samples. The greatest metabolic shift was observed for *S. meliloti* exposed to 1.4 mg/L of the PAH mixture. Over 37% of the polar metabolome (750 metabolic features) were significantly different, and 84.8% of those significant features were increased due to PAH exposure, while only 15.5% of those features were decreased in response to exposure of 1.4 mg/L of the PAH mixture. Fluorene and PAH mixture resulted in similar changes in the *S. meliloti* polar metabolome.

Discussion

Various government agencies or health organizations have devoted their time in measuring the carcinogenicity and mutagenicity of many PACs^{19–23}. The mutagenicity and carcinogenicity of PAH increases with increasing molecular weight. Over 25% of the carcinogenic properties of coal tar arise from PAH with six or more rings (or MW above 300) (21). Studies of PAH modes-of-action are primarily focused on higher organisms including mammals and plants. Oxidative enzyme cytochrome P450 oxidizes PAH to hydroxyl-PAH, which forms adducts with DNA, and subsequently disrupts DNA function^{25,26}. PAH can also induce non-carcinogenic effects by binding to the aryl hydrocarbon (Ah) receptor, and altering gene expression²⁷. For microorganisms that lacking cytochrome P450 and an Ah receptor, PAH can result in swelling of the membrane bilayer, increasing membrane fluidity and causing dysregulation of the proton motive force, pH gradient and electrical potential across the membrane²⁸. In contrast to the carcinogenic and mutagenic properties of PAH measured in mammals, the inherent

toxicity of PAH towards microorganisms decreases with increasing molecular weight and greater solubility²⁹.

Heterocyclic PACs containing N, S, or O atoms in their ring structures (i.e. PASH, NPAH) contribute to 1-10% of the total PAC concentration in coal tar³⁰. The structures of NPAH and PASH were modified from PAH by replacing a six-membered aromatic ring in PAH with a five-membered pyrrole or thiophene ring. The electronegativity of nitrogen ($\chi = 3.04$) is much higher compared to carbon ($\chi = 2.55$) and sulfur ($\chi = 2.58$), which makes NPAH more polar than PAH and PASH. The bioavailability of NPAH to organisms is higher than both PAH and PASH because of their high water solubility. NPAH could also have a completely different mechanism in bacterial species due to its high polarity. For example, instead of being trapped inside the biological membrane of the bacteria and distorting its function, NPAH might be able to travel to the cytoplasm of the cell and cause damage. Carbon and sulfur shares similar electronegative properties, and are thought to have similar toxicity and modes-of-action on microorganisms.

Considering the varying toxicity and potentially different modes-of-action between PACs, a PAH fraction containing only low molecular weight PAHs and PASHs was obtained in the optimized coal tar clean-up experiment, C5. Alumina chromatography was chosen as the separation method for coal tar clean-up because it can be readily scaled up if toxicity tests on higher organisms (i.e. mice) is required. The PACs were separated on an open tubular normal phase alumina column by varying the eluent polarity. Using a small quantity of adsorbent with less adsorption activity IV alumina reduces the interaction between PACs with the stationary phase. It then subsequently reduces the elution time of analytes (in particularly PAHs), sample loss, and volume of hazardous eluents required for the coal tar clean-up. A partial overlap between mid to high molecular weight PAHs and low molecular NPAHs were observed in early optimization experiments (C1, C2). The electronegativity of nitrogen ($\chi = 3.04$) was much higher compared to carbon ($\chi = 2.55$), therefore, NPAHs were retained longer than PAHs. The separation between NPAHs and PAHs can be optimized by increasing the adsorbability of the stationary phase alumina. Lowering the pH of the alumina column and also increasing the alumina activity by incubating at higher temperature allowed a stronger interaction between basic NPAHs and the acidic stationary phase. In experiments C3-5, only more polar eluents such as dichloromethane and methanol were able to elute NPAHs. The electronegativity of carbon and sulfur ($\chi = 2.58$) were similar, therefore, simply using an open tubular normal phase alumina column was insufficient to separate PAH and PASH. Therefore, PAH and PASH were found to coelute despite optimization. With optimization, a single low molecular weight PAH mixture fraction containing only PAHs (MW 128-202) and PASHs were collected from crude coal tar using alumina chromatography.

Here, we evaluate the toxic metabolic impacts of a single PAH compound, fluorene, and a low molecular weight PAH mixture derived from coal tar on a sentinel microorganism, *S. meliloti*. The metabolic impacts of single PAH or mixtures of two to four PAHs have been reported^{6,31-35}, however, there has been few reports on the

metabolomic outcomes due to exposure to complex contaminant mixtures or mixture fractions. Although PAHs exhibit similar modes-of-action, the interactive carcinogenic effects of PAH in mixtures often result in either less than or more than the expected additive effects in tumor studies in mice or humans^{6,31,32}. The same might be expected regarding the metabolic changes induced by PAH mixture as compared to an individual PAH. In real life, microorganisms, plants, wildlife and humans are never exposed to a single chemical or well-defined mixture of compounds. Rather than predict metabolic outcomes based on data from individual PAHs, we decided to compare the metabolic changes of *S. meliloti* as a result of single PAH or PAH mixture exposure.

From the metabolomic study of fluorene or PAH mixture exposed *S. meliloti*, similar modes-of-action were observed for both fluorene and the PAH mixture at both doses. Although no *S. meliloti* growth defects resulted from exposure to 0.14 mg/L and 1.4 mg/L of fluorene, and 0.14 mg/L PAH mixture, metabolic disturbance was observed. Even though few metabolic features were significantly different as a result of fluorene exposure as compared to PAH mixture at equivalent doses, the overall trend was similar (Fig 5). In fact, the changes in the metabolic profile as a result of 1.4 mg/L fluorene exposure resembled the *S. meliloti* profiles change caused by exposure to the 0.14 mg/L PAH mixture. The metabolic changes due to both the fluorene and PAH mixture were dose dependent. At equal concentration, mixture of PAHs can cause greater metabolic toxicity in *S. meliloti* as compared to a single PAH.

Our study demonstrates that LC-MS metabolomics can discriminate responses of *S. meliloti* exposed to sub-lethal concentration of fluorene or coal tar derived PAH mixture. Current technique for assessing and comparing potencies of toxic compounds rely on the measure of LD₅₀ or LC₅₀. Metabolomics studies of toxin exposed microorganisms can be an asset in measuring potency of toxic compounds at sub-lethal dosages that might or might not cause any growth defects. Moreover, by examining the changes in metabolic profiles, we can assess and compare modes-of-action between toxins. With additional research in identifying significantly expressed metabolites, the modes-of-action of toxins can be deduced. Overall, metabolomics holds great promise for measuring potency and modes-of-action associated with toxins at sub-lethal concentrations.

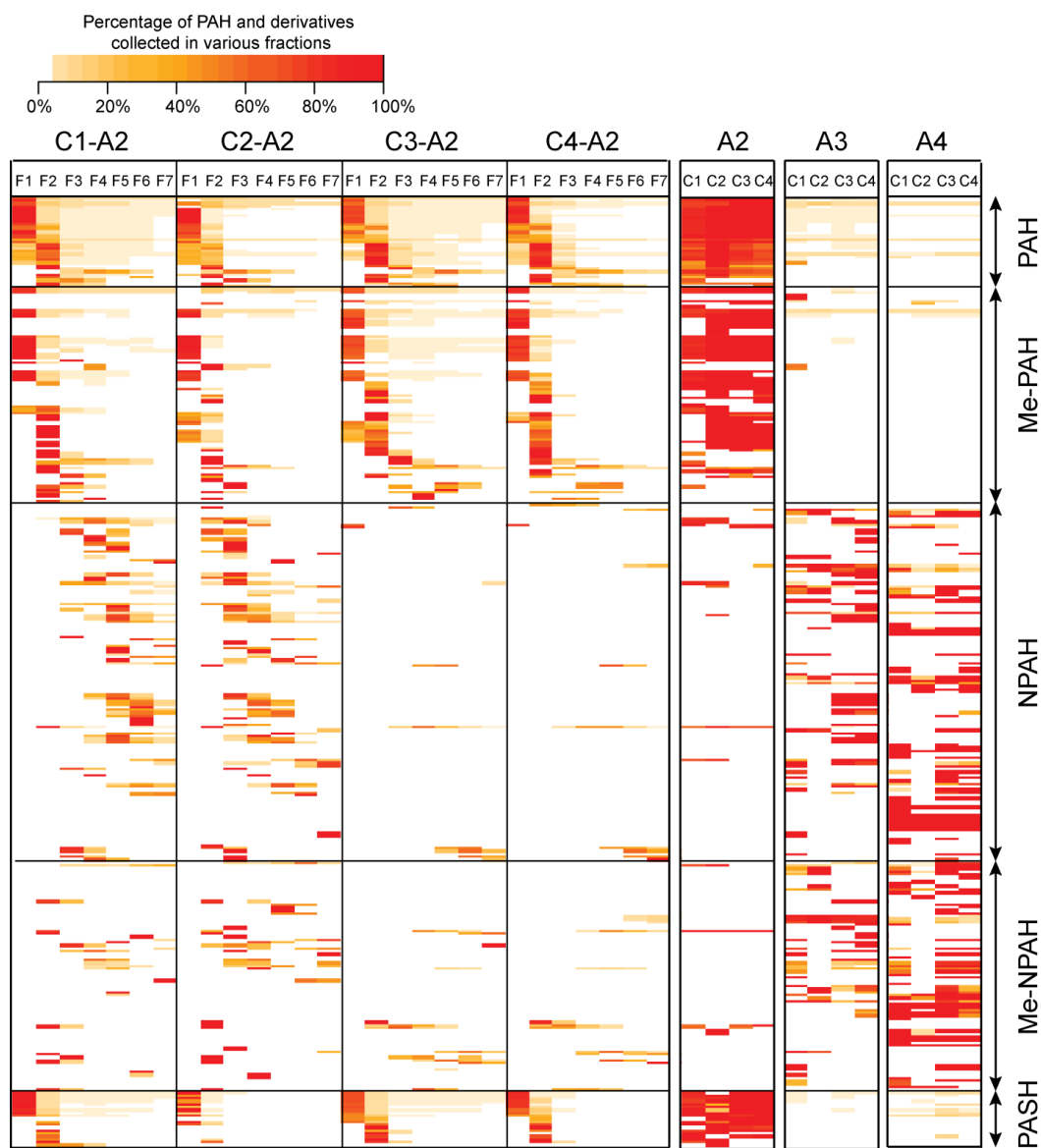


Figure 1 Optimization for coal tar clean-up using alumina chromatography. The heat map illustrated the extraction efficiency of individual PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs in A2 sub-fractions, A2, A3, and A4 fractions of experiment (a) C1, (b) C2, (c) C3 and (d) C4. The extraction efficiencies of PAHs and their derivatives in each fraction and sub-fraction were calculated by first normalizing the peak area of each metabolite with internal standards and then dividing the relative abundance of each individual PAHs at (sub-)fractions with its total recoverable sum in all A1-A4 fractions. Analytes in the PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs classes were ordered based on molecular weight.

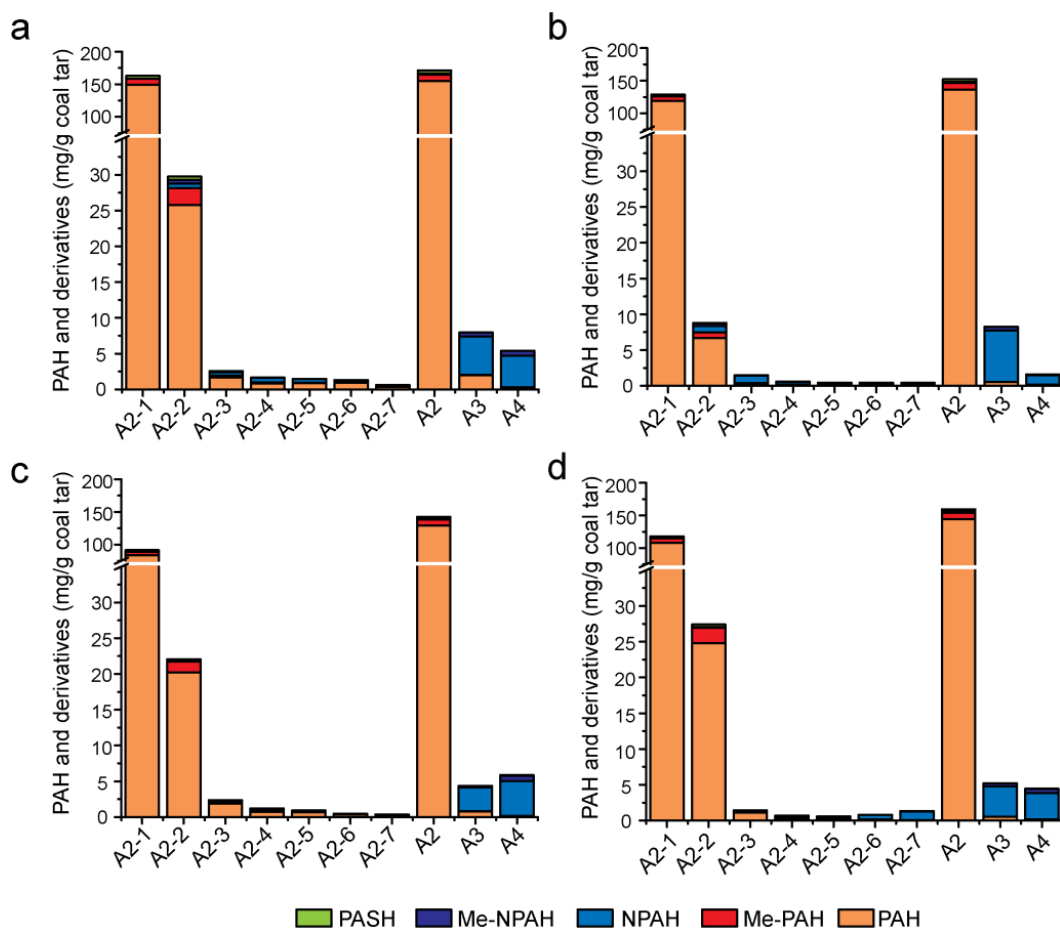


Figure 2 Optimization for coal tar clean-up using alumina chromatography. The chromatographic distributions of PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs in A2 subtraction fractions, A2, A3, and A4 fractions for experiment (a) C1, (b) C2, (c) C3 and (d) C4

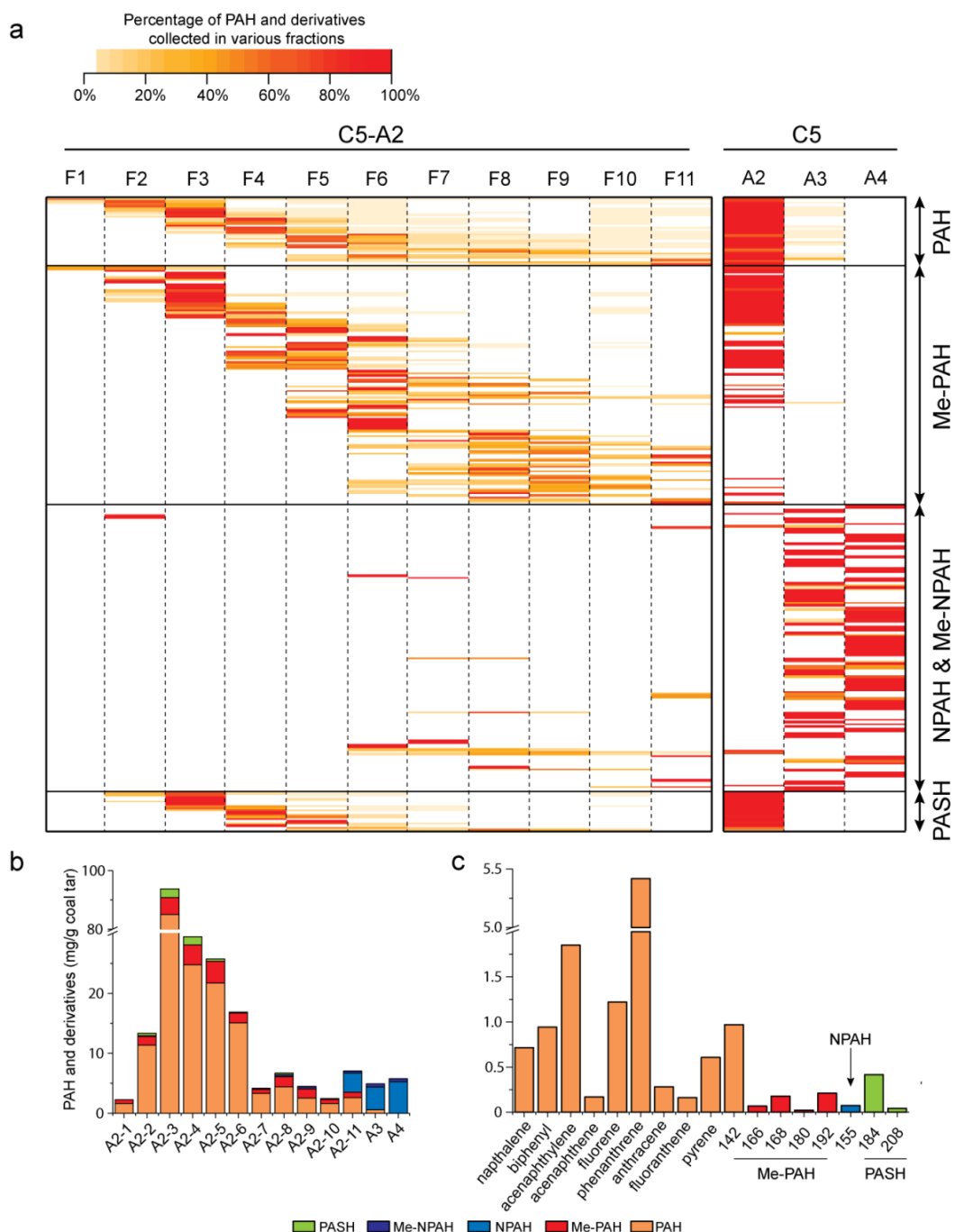


Figure 3 Elution profiles of PAH, Me-PAH, NPAH, Me-NPAH, PASH in A2 sub-fractions, A3 and A4 fractions in the optimized coal tar clean-up using alumina chromatography (experiment C5). (a) The heat map illustrate the elution pattern and extraction efficiencies of individual PAH or PAH derivatives in each fraction. Analytes in the PAHs, Me-PAHs, NPAHs, Me-NPAHs and PASHs classes were ordered based on

molecular weight. (b) The extractable concentration of each PAH groups at each fraction were illustrated. (c) The PAH concentration of A2-2 sub-fraction used for *S. meliloti* metabolomic study to investigate metabolic effects of sub-lethal PAH exposure

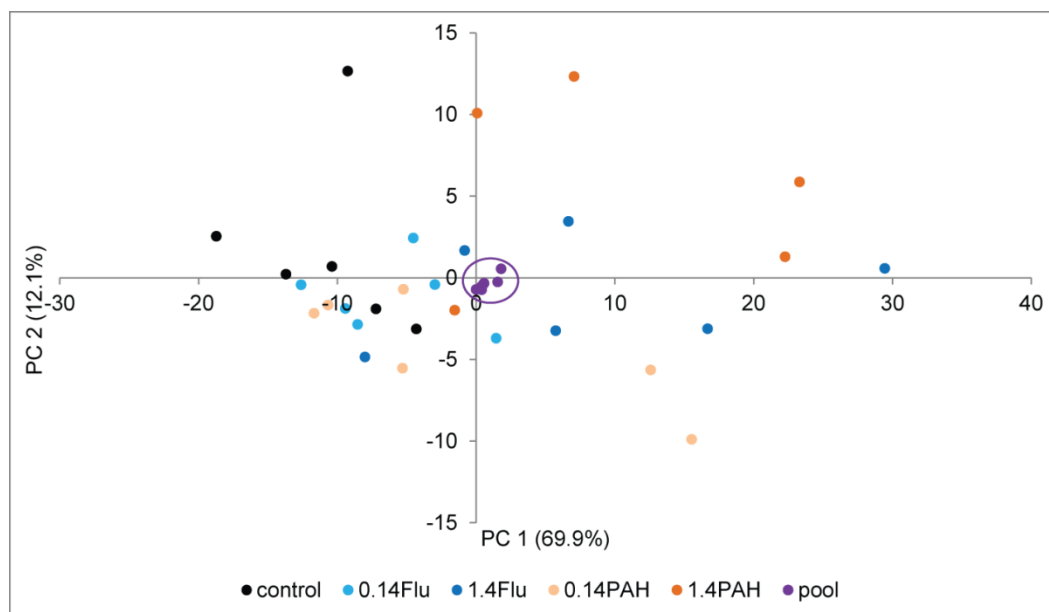


Figure 4 PCA score plot of *S. meliloti* polar metabolome to illustrate the reproducibility of the experiment. The pooled samples, colored in purple, were clustered in the center of the score plot indicating minimal technical variance. The polar metabolome consisted of 2005 metabolic features after data reduction

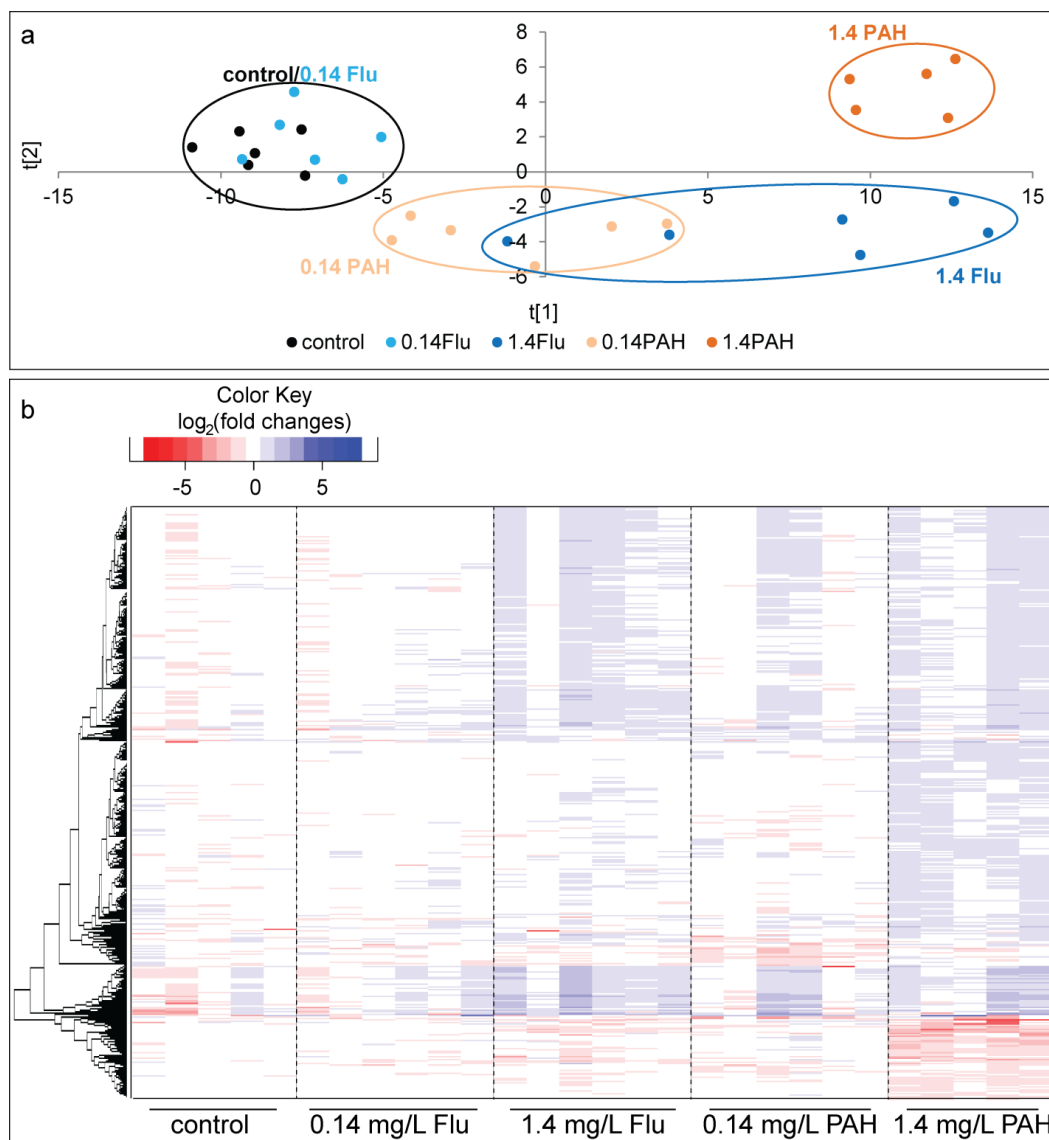


Figure 5 Comprehensive metabolomics of *S. meliloti* exposed to 0.14 mg/L and 1.4 mg/L fluorene or low molecular weight PAHs mixture. Exposure and metabolomic experiments were performed in sextuplicate. (a) Changes of metabolic profiles with regard to different PAH exposures were visualized in the OPLS-DA score plot ($R^2X=0.93$, $R^2Y=0.827$, $Q^2=0.409$). 2005 metabolic features were included in the analysis. No metabolic differences were observed between control samples and *S. meliloti* exposed to 0.14 mg/L fluorene. (b) Heat map illustrating individual metabolic feature changes with regards to the various PAH exposure. The 926 significantly expressed metabolic features (fold change >1.5 and $p<0.05$ with Student's *t* test between PAH exposed samples and control) were represented in rows, and various experimental conditions were listed in columns. The heat map was plotted based on the $\log_2(\text{fold change})$ with respect to the average

levels of each metabolic features in the control sample using Euclidean distance and complete-linkage clustering.

References

1. Liu, K., Han, W., Pan, W. P. & Riley, J. T. Polycyclic aromatic hydrocarbon (PAH) emissions from a coal-fired pilot FBC system. *J. Hazard Mater.* **84**, 175–188 (2001).
2. European Commission *Polycyclic aromatic hydrocarbons – occurrence in foods , dietary exposure and health effects. Management* (2002).
3. Wild, S. R. & Jones, K. C. Polynuclear aromatic hydrocarbons in the United Kingdom environment: A preliminary source inventory and budget. *Environ. Pollut.* **88**, 91–108 (1995).
4. Wise, S. A. *et al.* Polycyclic aromatic hydrocarbons (PAHs) in a coal tar standard reference material--SRM 1597a updated. *Anal. Bioanal. Chem.* **398**, 717–728 (2010).
5. US EPA *Guide to GLNPO assessment and remediation of contaminated sediments (ARCS) reports.* (Chicago, 1993).
6. Acros, J. C., Woo, Y. T. & Lai, L. Y. Database on binary combinations of effects of chemical carcinogens. *J. Environ. Sci. Health C6*, 1–150 (1988).
7. Olmstead, A. W. & Leblanc, G. A. Joint action of polycyclic aromatic hydrocarbons : Predictive modeling of sublethal toxicity. *Aquat. Toxicol.* **75**, 253–262 (2005).
8. Staal, Y. C. *et al.* Interactions between polycyclic aromatic hydrocarbons in binary mixtures: Effects on gene expression and DNA adduct formation in precision-cut rat liver slices. *Mutagenesis* **23**, 491–499 (2008).
9. Nesnow, S. *et al.* Lung tumorigenic interactions in strain A/J mice of five environmental polycyclic aromatic hydrocarbons. *Environ Health Perspect.* **106**, 1337–1346 (1998).
10. Galibert, F. *et al.* The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668–672 (2001).
11. Cowie, A. *et al.* An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. *Appl. Environ. Microbiol.* **72**, 7156–7167 (2006).

12. Keum, Y. S., Seo, J. S., Li, Q. X. & Kim, J. H. Comparative metabolomic analysis of *Sinorhizobium* sp. C4 during the degradation of phenanthrene. *Appl. Microbiol. Biotechnol.* **80**, 863–872 (2008).
13. Later, D. W., Lee, M. L., Bartle, K. D., Kong, R. C. & Vassilaros, D. L. Chemical class separation and characterization of organic compounds in synthetic fuels. *Anal. Chem.* **53**, 1612–1620 (1981).
14. Li, C.-L. Separation and characterization of thia-arenes and high mass polycyclic aromatic hydrocarbons in coal tar. 28–30 (1997).
15. Fei, F., Bowdish, D. M. E. & McCarry, B. E. Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. *Anal. Bioanal. Chem.* **406**, 3723–3733 (2014).
16. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
17. Kuhl, C., Tautenhahn, R. & Neumann, S. LC-MS Peak Annotation and Identification with CAMERA. 1–14 (2010).
18. Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **6**, 1060–1083 (2011).
19. US Department of Health and Human Services *Toxicological profile for polycyclic aromatic hydrocarbons, agency for toxic substances and disease registry.* (Atlanta, GA, 1995).
20. United Nations Environment Programme, International Labour Organization & World Health Organization *Selected non-heterocyclic polycyclic aromatic hydrocarbons: Environmental health criteria 202.* (Geneva, 1998).
21. Siemiatycki, J. *et al.* Listing occupational carcinogens. *Environ. Health Perspect* **112**, 1447–1459 (2004).
22. World Health Organization *IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Polynuclear Aromatic Compounds* (1985).
23. Poster, D. L., Schantz, M. M., Sander, L. C. & Wise, S. A. Analysis of polycyclic aromatic hydrocarbons (PAHs) in environmental samples: a critical review of gas chromatographic (GC) methods. *Anal. Bioanal. Chem.* **386**, 859–881 (2006).

24. Grimmer, G. *et al.* Relevance of polycyclic aromatic hydrocarbons as environmental carcinogens. *Fresenius J. Anal. Chem.* **339**, 792–795 (1991).
25. Baturay, N. Z. & Trombetta, L. D. Cocarcinogenic and tumor-promoting capabilities of anthralin. *Arch. Dermatol. Res.* **280**, 443–450 (1988).
26. Xue, W. & Warshawsky, D. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol. Appl. Pharmacol.* **206**, 73–93 (2005).
27. Hankinson, O. The Aryl hydrocarbon receptor complex. *Annu. Rev. Pharmacol. Toxicol.* **35**, 307–340 (1995).
28. Sikkema, J., de Bont, J. a & Poolman, B. Interactions of cyclic hydrocarbons with biological membranes. *J. Biol. Chem.* **269**, 8022–8028 (1994).
29. Calder, J. a & Lader, J. H. Effect of dissolved aromatic hydrocarbons on the growth of marine bacteria in batch culture. *Appl. Environ. Microb.* **32**, 95–101 (1976).
30. Giessel-Nielsen, G. & Nielsen, T. Phytotoxicity of acridine, an important representative of a group of tar and creosote contaminants, N-PAC compounds. *Polycycl. Aromat. Comp.* **8**, 243–249 (1996).
31. Schmahl, D., Schmidt, K. G. & Habs, M. Syncarcinogenic action of polycyclic aromatic hydrocarbons in automobile exhaust gas condensates. *IARC Sci. Publ.* **16**, 53–59 (1977).
32. Gibb, H. J. & Chen, C. Multistage model interpretation of additive and multiplicative carcinogenic effect. *Risk Anal.* **6**, 167–170 (1986).
33. Brown, S. A. E., McKelvie, J. R., Simpson, A. J. & Simpson, M. J. ¹H NMR metabolomics of earthworm exposure to sub-lethal concentrations of phenanthrene in soil. *Environ. Pollut.* **158**, 2117–2123 (2010).
34. McKelvie, J. R., Yuk, J., Xu, Y., Simpson, A. J. & Simpson, M. J. ¹H NMR and GC/MS metabolomics of earthworm responses to sub-lethal DDT and endosulfan exposure. *Metabolomics* **5**, 84–94 (2009).
35. Brown, S. A. E., Simpson, A. J. & Simpson, M. J. ¹H NMR metabolomics of earthworm responses to sub-lethal PAH exposure. *Environ. Chem.* **6**, 432–440 (2009).

Chapter 7: General discussion, future directions and conclusion

7.1 General discussion

In this final chapter, the advantage of the HILIC-LC-TOF-MS methods in Chapter 2 will be summarized. The versatility of this method and its use for advancing cell biology research will be discussed. Finally, future directions and general conclusions for metabolomic research in the field of cell biology are presented.

7.2 Implications of metabolomics

In this thesis, a comprehensive and simultaneous detection of intra- and extra-cellular phospholipids and polar metabolites using the HILIC-TOF-MS platform was presented. Additionally, a complimentary extraction method specifically designed for unicellular organisms with strong exterior cell walls was presented. Unlike the traditional approach, where polar and non-polar metabolites are extracted separately and analyzed separated using reverse phase (RP) and normal phase (NP) HPLC-MS. This approach is able to simultaneously analyse a broad range of polar and nonpolar metabolites in a single run. Moreover, the ability of the silica-HILIC column to retain lipids and to separate phospholipids by classes based on their polar head group is extremely helpful with regard to metabolite identification. Since each lipid class is eluted within a narrow time window on silica-HILIC, we could accurately identify the type of phospholipids and their chain length based on their retention time and m/z value without the use of authentic standards. This platform fills the demand for comprehensive metabolomics, which requires high throughput, high reproducibility and large metabolite coverage.

The versatility of the HILIC-TOF-MS platform has been successfully demonstrated with three different cell types: the Gram-positive bacterium *Streptococcus intermedius*, the Gram-negative bacterium *Sinorhizobium meliloti*, and eukaryotic macrophages. Additionally, the value of comprehensive metabolomics for *in vitro* cell metabolic research has been shown in various areas of application.

Chapter 3 provides insight on how *S. intermedius* metabolism adapts to various oxygen levels in the growth environment. Under anaerobic conditions, purine metabolism, pyrimidine *de novo* synthesis and pyrimidine salvage pathways were up-regulated at both metabolic and transcriptional levels to support the accelerated growth of *S. intermedius* in the absence of oxygen. As a facultative anaerobe, *S. intermedius* can colonize in both aerobic and anaerobic niches within human host, and infectious strains are found associated with obligate anaerobes. *S. intermedius* grows faster in anaerobic conditions in conjunction with upregulated metabolic activity, and this might be linked to its pathogenic associated with anaerobes in polymicrobial infection and disease progression

in the host. Moreover, understanding the metabolic regulation of *S. intermedius* under various oxygenated environment through comprehensive metabolomic studies may provide insights into adapting to various microenvironment within the human host.

Chapter 4 examines the metabolic contribution of pSymA megaplasmid and pSymB chromid in the multipartite *S. meliloti* genome through comprehensive intra- and extracellular metabolomic studies. The pSymA megaplasmid has a minor contribution to the cellular metabolism of free-living *S. meliloti*. In contrast, the pSymB chromid is highly integrated with the primary metabolic activity of *S. meliloti*. A lack of the pSymB chromid not only reduces the growth of *S. meliloti* in both rich LBmc and minimal M9-sucrose media, but also limits *S. meliloti* accessibility to extracellular nutrients, sugar metabolism, and the purine degradation/salvage pathway. The loss of pSymB also led to changes in various phospholipid levels in the cell membrane. The knowledge of the metabolic functions of pSymA and pSymB provides insight into the biology of *S. meliloti* and also the functional significance of secondary replicons in multipartite bacterial genomes. Moreover, Chapter 4 also demonstrated the potential of using comprehensive metabolomics to examine the metabolic capacity and preferences of a bacterial population grown in a nutritionally complex environment. The hierarchical substrate usage determined by examining the depletion rate of the substrate can be correlated with the preferred carbon, nitrogen, or phosphorus sources.

Chapter 5 reported the potential metabolic bottlenecks that can contribute to the dysregulation of inflammatory response with age. A lack of immediate energy output through the metabolic switch from oxidative phosphorylation to glycolysis was observed in the lipopolysaccharide (LPS) stimulated macrophages from old mice. This was correlated to the decreased expression of arginine and adenosine and urea cycle activity, which is also decreased in the LPS stimulated macrophages from old mice. The dysregulated immune response observed in the elderly could be linked to the age-associated deterioration in mitochondria function. In addition to the more popular proteomic (i.e. cytokines, NF- κ B, toll-like receptor), biochemical or physiological approach in studying age-associated decline in immune function and macrophage function, comprehensive and targeted metabolomics suggests that metabolic dysregulation as a possibly unexplored mechanism of immunosenescence.

Lastly, Chapter 6 uses a comprehensive metabolomic approach to examine the toxic effects of sub-lethal polycyclic aromatic hydrocarbons (PAHs). Sub-lethal concentrations of toxic chemicals that do not induce any physiological changes can be measured through metabolomic studies. Additionally, (dis)similar modes-of-action between different or

mixtures of different toxicants can also be determined through comprehensive metabolomics.

7.3 Future directions of research

Comprehensive metabolomics is a hypothesis generating research technique. Within the applications in this thesis, various biomarkers or significantly changed metabolites or key disturbed pathways has been discovered. Follow up studies have to be carried out in order to verify the causation between the metabolic response and the observed phenotype or function. For example:

- Whether the reduced metabolic activity results in the reduced growth rate for *S. intermedius* in aerobic growth or vice versa still requires investigation.
- The biosynthetic pathways associated to pSymA megaplasmid and pSymB chromid still need to be mapped. The metabolic models of *S. meliloti* could be developed and linked to its multipartite genome.
- There is an apparent lack of energy for LPS stimulated macrophages from old mice. How a lack of energy can interfere with macrophage function during inflammation needs to be addressed. The link between mitochondrial and macrophage dysfunction as a result of old age needs to be investigated.
- Specific modes-of-action of PAHs and also their effect on metabolic activity in *S. meliloti*, and whether such toxic effect of PAHs observed in *S. meliloti* can be generalized to all bacterial species still remained to be answered.

For the future of cell metabolomics, there is always demand for a faster and better quenching and extraction methodology that preserves the cellular integrity. An analytical platform that is able to cover more diverse metabolites with greater dynamic range is always desirable. The large number of unknown metabolites in the biological samples has been a key challenge for comprehensive metabolomics. The rapid development of MS/MS spectra database (i.e. HMDB, METLIN) allows identification of unknown metabolites. Ultimately, metabolic profiling extending to subcellular compartments (i.e. mitochondria, cytosol) could provide unique insights in organelle interactions and origins of metabolic activities in a single cell organism. Farré *et al.* could be the first group to venture into the cellular compartmentalization metabolomic study, where the metabolites in potato tubers were profiled using GC-MS method¹. Few other metabolomic studies have also focused on analyzing organelle fractions^{2,3}. Moreover, the progress made in the development of MS, microfluidics and capillary separation has made single cell analysis possible⁴. The ability to analyze metabolites in a single cell allows studies of biological variation, functional heterogeneity in a seemingly homogenous cell population.

7.4 General conclusion

In this thesis, the potentials of metabolomics in the field of *in vitro* cell biology have been illustrated in different studies with various cell types. The ability of analyzing phospholipids and polar metabolites makes the HILIC-TOF-MS desirable in the field of metabolomics. Pathways, significant features have been identified associated to various biological conditions using comprehensive metabolomics. Targeted metabolomics has led to the discovery of metabolic bottlenecks and a possible new mechanism for age-associated macrophage deterioration. However, it should be noted that, although metabolomics is a powerful tool for biomarker discovery or pathway analyses, relevant biological studies have to be performed to validate the findings.

7.5 References

1. Farré, E. M. *et al.* Analysis of the compartmentation of glycolytic intermediates , nucleotides , sugars , organic acids , amino acids , and sugar alcohols in potato tubers using a nonaqueous fractionation method. *Plant physiol.* **127**, 685–700 (2001).
2. Komatsu, S. *et al.* Comprehensive analysis of mitochondria in roots and hypocotyls of soybean under flooding stress using proteomics and metabolomics techniques. *J. Proteome Res.* **10**, 3993–4004 (2011).
3. Oikawa, A., Matsuda, F., Kikuyama, M., Mimura, T. & Saito, K. Metabolomics of a single vacuole reveals metabolic dynamism in an alga *Chara australis*. *Plant Physiol.* **157**, 544–551 (2011).
4. Rubakhin, S. S., Lanni, E. J. & Sweedler, J. V Progress toward single cell metabolomics. *Curr. Opin. Biotechnol.* **24**, 95–104 (2013).