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

*Fan Fei1\*, George C. diCenzo2\*, Dawn M. E. Bowdish3, Brian E. McCarry1, Turlough M. Finan2†*

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

2Department of Biology, McMaster University, Hamilton, Canada L8S 4K1

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

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

\*Both authors contributed equally

**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-d3(98%), L-tryptophan-d5 (98%)as well as 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.

**Strains, media and growth conditions**

The four strains of *S. meliloti* used in this study were previous described (DiCenzo, MacLean, Milunovic, Golding, & Finan, 2014; Oresnik, Liu, Yost, & Hynes, 2000), which included wild type Rm2011 (SU47 *str-3*), SmA818 (ΔpSymA), RmP3009 (ΔpSymB) and RmP2917 (ΔpSymAB). ΔpSymA, ΔpSymB and ΔpSymAB were derived from Rm2011by the removal of pSymA, pSymB or both megaplasmids, respectively. Essential genes, *tRNAarg* 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 MgSO4, 2.5 mM CaCl2, 2 μM CoCl2) was used as the complex medium, and M9-sucrose medium (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 and 42 nM CoCl2) was used as a minimal medium.

All four strains of *S. meliloti* were cultured to early stationary phase in 5 mL M9/LBmc at 30ºC. At 3.0 OD600, 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 OD600 of 0.05. These cultures were aliquoted into seven test tubes with 5 mL each and incubated at 30ºC. 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) OD600. 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 μL were taken at each time point, respectively, from the same 5 mL culture, so that a constant of 6108 cells were collected in each sample. The cells were centrifugated at 4ºC, 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 -80ºC 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 4ºC and 20 μL of the media supernatant were collected and extracted with 80 μL MeOH/EtOH (1:1) containing RS to remove precipitate 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 -80ºC 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 acetonitile (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 ºC, and the autosampler storage tray was at 4ºC. 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 kapp’ 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 material ESM1. 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 R2X, R2Y, and Q2 were used to assess the fitness of the model. R2X (R2Y) indicated the fraction in which metabolite features (X) and treatment (Y) matrix was explained by the model. Briefly, a prediction statistic (Q2) above 0.4 were indicative of a robust model, and Q2 between 0.7-1.0 indicated the model was highly robust (Jones, Spurgeon, Svendsen, & Griffin 2008). Both R2 and Q2 followed an upward trend from 0 to 1. For an over fit model, R2 approach 1, and Q2 fell toward 0. Therefore, a valid and robust OPLS-DA model should have R2X and R2Y approaching 1 and Q2 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.

**References**

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., Brown, M., et al. (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.

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.

Kuhl, C., Tautenhahn, R., & Neumann, S. (2010). LC-MS Peak Annotation and Identification with CAMERA.

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.

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.