Supplemental Information: **The influence of plant species, tissue type and temperature on the capacity of Shigatoxigenic *Escherichia coli* to colonise, grow and internalise into plants.**

1 Supplemental Methods

2 SM1 Primary Modelling

3 Growth curves were fitted using the full Baranyi model described by Baranyi and Roberts (3),

4 using DMFIT (2), as an Excel add on:

5 $$y(t) = Y_0 + \mu\, A(t) - \ln\left\{1 + \frac{\exp(\mu\, A(t)) - 1}{\exp(Y_{max} - Y_0)}\right\} \qquad (A)$$

6 $$A(t) = t + \frac{1}{\mu}\ln[\exp(-\mu\, t) + \exp(-h_0) - \exp(-\mu\, t - h_0)] \quad (B)$$

7 Y_0 is the starting concentration, Y_{max} the maximum natural logarithm of bacterial counts, μ the

8 maximum growth rate and h_0 the physiological state of the bacteria. (A) Cell concentration at

9 time $t > 0$. (B) Gradual delay in time. Equation was taken from McKellar and Lu (12). The

10 following inputs by default were used: mCurv = 10; nCurv = 0; lower bound = 0 and upper

11 bound = 9999 (4). Root mean square error (RMSE) was calculated as proposed by McKellar

12 and Lu (12) and the models ranked accordingly. The maximum growth rates μ, standard error

13 of the mean (SEM) were used directly from DMFIT output and plotted in Prism 7 for each

14 temperature individually. Statistical analysis was conducted in Prism using two way ANOVA,

15 Pearson correlation (16) and multiple comparison tests (9, 19) where necessary.

16 Due to the nature of the growth rates the formulas without lag phase have been chosen and

17 fitted. The package uses both models for the log based 10 instead of natural logarithm.

18 Baranyi without lag :   (C)

19 $$\log_{10}Y \sim \left(\log_{10}Y_{max} - \log_{10}\left(1 + \left(10^{\log_{10}Y_{max} - \log_{10}Y_0} - 1\right)\exp(-\mu\, t)\right)\right)$$

20

21   Buchanan without lag: (D)

22
$$\log_{10}Y \sim \log_{10}Y_0 + \left(t <= \left((\log_{10}Y_{max} - \log_{10}Y_0) * \frac{\log_{10}}{\mu} \right) \right) \mu \frac{t}{\log_{10}}$$

23
$$+ \left(t > \left((\log_{10}Y_{max} - \log_{10}Y_0) \frac{\log_{10}}{\mu} \right) \right) (\log_{10}Y_{max} - \log_{10}Y_0)$$

24   with $\log_{10}Y_{max}$ the maximum bacterial count, $\log_{10}Y_0$ the initial count, μ the maximal growth rate

25   and t the time (D) (5).

26   Models were run in pairs and either compared by (E) the corrected AIC (13) or the (F) root

27   mean square error as proposed by McKellar and Lu (12):

28
$$\Delta \text{AIC} = N * \ln\frac{SS2}{SS1} + 2\Delta DF \qquad\qquad (E)$$

29   SS represents the sum of squares for both models and ΔDF is the difference of degrees of

30   freedom.

31
$$\text{RMSE} = \sqrt{\frac{\Sigma(\hat{Y}-Y)^2}{df}} \qquad\qquad (F)$$

32   Ŷ is the fitted value, Y the sample value and df the degree of freedom. These criteria take into

33   account the variability of parameters amongst models, which would otherwise impact a

34   comparison by sum-of squares F test or adjusted R^2_{adj} (1, 6).

35   Models were tested for discrepancies of growth curve parameters if converted to the natural

36   logarithm (ln) or logarithm with the base 10 ($\log_{10}$) (Fig. S1) on the data set of *E. coli* JHI5025

37   at 18 °C.

SM2   Secondary modelling

To determine temperature-dependent correlations, secondary modelling of the growth data with the linear approximation of the Ratkowsky model as used by McKellar and Lu (12) was attempted. However, a linear regression could not be fitted to the data in plant extracts and the number of temperature points (n = 3) was insufficient for non-linear modelling with the full Ratkowsky model (14). This resulted in poor fittings after secondary modelling for all replicates and samples for bacterial growth curves for plant extracts. In contrast, the secondary modelling of the positive control in RDMG was successful and consistent throughout the data set with achieved good linear regression fittings with R^2 (0.996 to 1).

SM3 Apoplastic fluid collection

Apoplast extraction followed the protocol of Husted and Schjoerring (10), which was developed for rapeseed (*Brassica napus*) and adjusted for spinach and lettuce to minimize cytoplasmic contamination (11). Leaves from plants grown in compost were harvested using a sterilised scalpel and immediately stored on ice. They were submerged under SDW in a vacuum chamber using a low vacuum pump (Divac 2.4L, Leybold, Switzerland) (L = low: 40 l h^{-1}; polytetrafluoroethylene diaphragm) and infiltrated by three to five cycles (maximum 2 min) of vacuum and release. Leaves were then gently dried on paper towels and rolled without breaking to fit a 20 ml syringe (BD Plastipak™, Becton, Dickinson and Company, USA) with the plunger removed. The syringe was placed into a 50 ml Falcon tube (Falcon, Thermo Scientific, USA) and was subject to centrifugation (Megafuge 16R, Thermo Scientific, USA) at 200 rcf for spinach and 280 rcf for lettuce for 15 min. The resulting fluid, which contained the plant apoplast together with infiltrated water was collected and passed through a sterile 0.1 µm filter (Durapore, Merck, Germany). Extracts were stored at -20 °C until further use.

Contamination of apoplast with cellular cytoplasm was evaluated as described previously (15, 18). Fluid containing the apoplast was extracted using extraction buffer (50 mM potassium buffer, pH = 7.0; 0.05 % β-mercaptoethanol; 10 mM EDTA, pH = 8.0; protease inhibitor tablets (Sigma Aldrich, USA) instead of water. Infiltrated leaves were subject to centrifugation

65 (Megafuge 16R, Merck, Germany) at 200 rcf for spinach and 280 rcf for lettuce for 15 min and

66 the extraction buffer containing the apoplast was used immediately. Leaf lysates were

67 produced to allow relative quantification of cytoplasmic contamination, by grinding fresh leaves

68 with liquid nitrogen and mixing 1:1 w/v with extraction buffer. The dispersion was then

69 centrifuged (Megafuge 16R, Merck, Germany) at 5000 rcf for 15 min and supernatant was

70 used immediately for enzymatic assay. Cytoplasmic malate dehydrogenase (MDH) was used

71 for assessment since glucose-6-phosphate dehydrogenase (G6PDH) was undetectable, with

72 no enzymatic activity in leaf lysates over five minutes. Malate dehydrogenase (MDH) was

73 measured with 200 µl MOPS buffer (200 mM MOPS; 50 mM sodium acetate and 10 mM

74 EDTA); 50 µl 0.5 mM NADH; 10 µl plant extract; 540 µl dH_2O and 200 µl 2 mM oxaloacetic

75 acid at OD_{340} for 5 min at RT. Buffer only was used as a negative control. Validation was

76 carried out by spiking samples with a positive control containing 1 µl purified MDH. Enzymatic

77 assays were performed in triplicate. Cytoplasmic contamination, calculated as a ratio of

78 enzyme activity (OD 340 nm min^{-1}) between leave lysates and apoplast, was estimated as 1.8

79 % contamination in spinach and 5.0 % in lettuce.

80 **SM4   Metabolite analysis and Correction factor for GC-MS data**

81 10 ml plant extract samples were lyophilised (116 l h^{-1}) overnight and 40 mg extracted twice

82 in 1 ml ethanol (80 %) by heating at 80 °C for 30 min, cooled on ice, and clarified by

83 centrifugation (13000 RCF, 15 min) and then freeze dried under vacuum. Supernatants were

84 pooled for further analysis and re-suspended in 1 ml molecular biology grade H_2O. Glucose,

85 fructose and sucrose were measured in a 1:100 v/v dilution in a Dionex chromatography

86 machine fitted with a CarboPac® PA-100 column and 200 mM NaOH for 100 % of the eluent,

87 flow = 1.0 ml min^{-1} and column temperature at 30 °C for 15 min. Arabinose was measured with

88 a CarboPac® PA-20 column. Buffers were run on a gradient profile: SMBG H_2O; 200 mM

89 NaOH and 1 M NaOAc at 0.4 ml min^{-1} and column temperature at 30 °C, over 40 min. Standard

90 curves for glucose, fructose, sucrose and arabinose concentrations were created and fitted

91 using a linear regression, R^2 for 6 data points was: glucose = 0.999; fructose = 0.999; sucrose

92   = 0.999 and arabinose = 0.997. Concentrations were interpolated from the curve and

93   normalised to the dry weight in mg.

94   Polar fractions were prepared for GC-MS analysis as described by Shepherd, et al. (17):

95   glassware was pre-washed with dH_2O, methanol / dH_2O (3:1) and chloroform / methanol (2:1).

96   Eicosane, tertracosane, triacontane, tetratriacontane, octatriacontane (each 2 mg ml^{-1}),

97   undecane, tridecane (both 2.7 µg ml^{-1}), hexadecane (2.6 µg ml^{-1}) were used as retention

98   standards. 40 mg freeze dried plant extracts were suspended in 3 ml methanol with addition

99   of 100 µl ribitol (as internal standard), 0.75 ml dH_2O, 6 ml chloroform and water after each

100   interval of mixing at 1500 rpm at 30 °C for 30 min. Samples were collected from centrifugation

101   at 1200 rpm for 10 min and fractions separated into vials. The polar fraction was stored at -20

102   °C until analysis. 250 µl polar fraction was evaporated for 2 h with heat and derivatisation

103   conducted on day of analysis with addition of 20 mg methoxylamine hydrochloride (98 %) in 1

104   ml anhydrous pyridine. 50 µl of retention standard mixture and *N*-Methyl-*N*-

105   (trimethylsilyl)trifluoroacetamide (MSTFA) were added for derivatisation at 37 °C for 30 min.

106   Samples were analysed in a Quadrupole–mass spectrometry GC–MS system (Thermo

107   Finnigan Trace DSQ, USA) described previously (17): split was set at 80:1, interface

108   temperature was 250 °C, source temperature 200 °C and the full mass range over 35 – 900

109   atomic mass units was analysed. Raw data files from the GC-MS were analysed in Xcalibur™

110   v. 2.0.7 (Thermo Scientfic, USA). Ion characteristics were used to identify metabolites.

111   Selected ion chromatogram (SIC) was selected for each compound automatically, but was

112   reviewed and baseline corrected manually as required. The two SIC of oxoproline, threonic

113   acid, aspagarine, tryptophan and maltose were added together, to account for oximated and

114   unoximated derivatives. One SIC for aspartic acid was deleted and only the fully silyated

115   derivative used in further analyses. Correction factors were applied as described below. The

116   response ratio (RR) was calculated by comparing the SIC of each isolate against the SIC of

117   the internal standard. Blanks were deducted from the results. ANOVA and Pearson correlation

118   (16) was run in Prism 5 (Graphpad, USA) and GenStat 15 (VSN International, UK). A principal

119 components analysis (PCA) was calculated using GenStat 15 with an existing code (8). All

120 samples were run in triplicate, except lettuce apoplast, for which there was only one sample.

121 Correction factors were calculated for glucose and fructose to account for the ratio of

122 unoximated and oximated derivatives. This is a result of structural changes throughout the

123 derivatisation process prior to measurement. Correction factor was calculated by the ratio of

124 selected ion chromatogram / total ion chromatogram (TIC) of total sugar against SIC TIC^{-1} of

125 unoximated sugar and final multiplication with a system machine dependent factor (F_{System}),

126 either 1.361 (glucose) or 1.123 (fructose).

127
$$\left.\frac{\frac{SIC_{Total}}{TIC_{Total}}}{\frac{SIC_{Unoxi}}{TIC_{Unoxi}}}\right. * F_{System} \qquad (G)$$

128 where SIC is the selected ion chromatogram, TIC total ion chromatogram and F_{System} a

129 machine dependent factor for individual compounds.

130

131 **SM5 Conversion of absorption to viable counts**

132 Growth parameters used in mathematical modelling are usually required in log (cfu h^{-1}), which

133 requires conversion from optical density (OD) to viable counts. That was achieved by diluting

134 a PBS-washed overnight culture of *E. coli* isolate MG1655 to a range of OD_{600} nm and plating

135 onto MacConkey agar. Best fit was achieved with a linear regression to 10 data points (R^2 =

136 0.99), determined by the corrected AIC, which generated a slope for 4.5 * 10^8 cfu ml^{-1} (95 %

137 Confidence interval (CI): 4.1 * 10^8 to 4.8 * 10^8) with an intercept of -2.5 * 10^7 ± 1.4 * 10^7. The

138 measurements were validated from parallel absorption and viable counts for several time

139 points, which generated similar results: slope fitting with 8 linear data points (R^2 = 0.958) was

140 4.5 * 10^8 cfu ml^{-1} (CI: 3.6 * 10^8 to 5.5 * 10^8) with an intercept of -1.2 * 10^7 ± 2.0 * 10^7, over a

141 range OD 600 nm 0.005 to 2.000. However, reciprocal multiplication of the OD_{600} nm

142 measurements required an adjustment to 4.2 * 10^8 cfu ml^{-1}, which when compared against the

143   viable counts showed no significant differences (two-way ANOVA ($F_{(7, 70)}$ = 0.81; p = 0.58).

144   Therefore, the factor of 4.2 * 10^8 cfu ml^{-1} was used for all following growth experiments.

145

Supplemental Tables and Figures

147   **Supplemental Table 1**   Maximum growth rates of E. coli in plant extracts of

148   sprouts, lettuce and spinach

149   *E. coli* isolates were grown in a plate reader and the data converted to viable counts and

150   fitted with the Baranyi model in DMFIT to obtain the maximum growth rates. Temp =

151   Temperature in °C. Rates in log10 (cfu h-1); n = number of data points used for curve fitting;

152   R2_adj = Adjusted coefficient of determination.

153   **Supplemental Table 2**   Response ratio of assigned polar metabolites in

154   spinach and lettuce

155   Response ratio (RR) of metabolites determined against the internal standard ribitol. GC-MS

156   ion characteristics used for compound identification was validated by external standards. A

157   total of five plant species and tissues were examined: alfalfa, fenugreek, lettuce, spinach,

158   apoplast (AP), leaf lysates (LL) and root lysates (RL). Limits for colour are > 50 (black); > 20

159   (dark brown); > 10 (light brown); > 2 (red); > 1.5 (pink); > 1.0 (orange); > 0.8 (light orange); >

160   0.6 (light blue); > 0.4 (light green); > 0.2 (green); > 0.1 (dark green) and < 0.1 (blue). (n = 3,

161   except for LAP = 1). Individual metabolites that showed significant differences ($p < 0.05$)

162   between tissues are underlined.

163   **Supplemental Figure 1**   Manual correction of growth rate misfits in DMFIT.

164   Example of a correction with *E. coli* isolate JHI5039 grown in lettuce leaf lysate, 18 °C.

165   **A)** DMFIT could not fit a non-linear curve on data (n = 193) with a decrease in the

166   stationary phase (R^2_{adj} = 0.001). **B)** Data was cut off manually (n = 49) to achieve

167 better fits (R^2_{adj} = 0.996). A complete list of fits including data points are in

168 Supplemental Table 3.

169 **Supplemental Figure 2**   Simplified polar metabolic pathways in plants

170 Interaction between major polar pathways (colour coded) in green leafy plants.

171 Metabolism of carbohydrates degradation (green) is linked to amino acid degradation

172 (dark blue and purple), which feed into the TCA cycle (red). The arrows pointing

173 outside are entries into the non-polar fatty acid pathway. The glutamate group (orange)

174 leads into the urea cycle. The light blue cycle described the acyl chain synthesis.

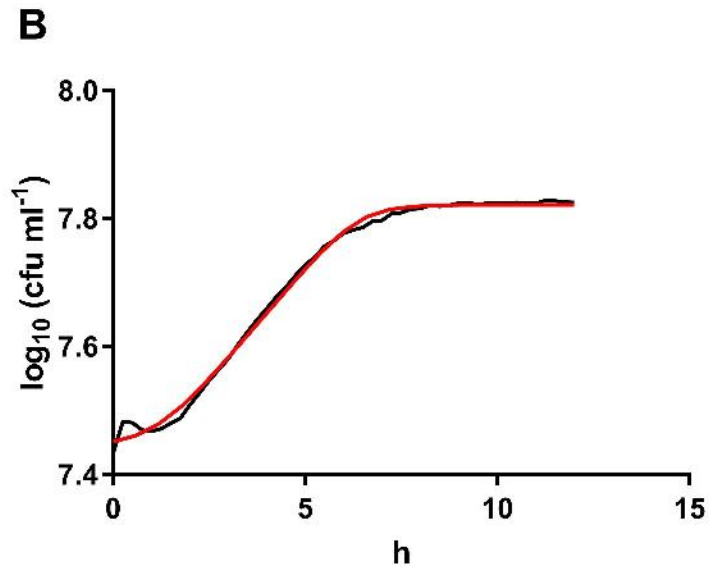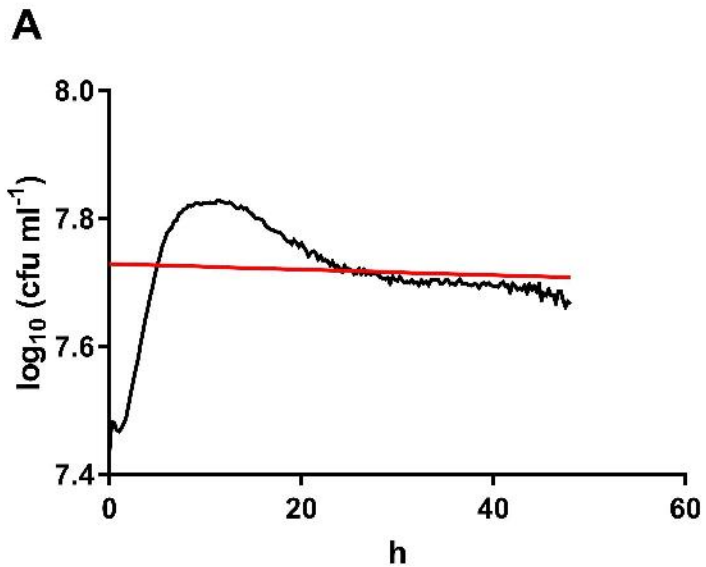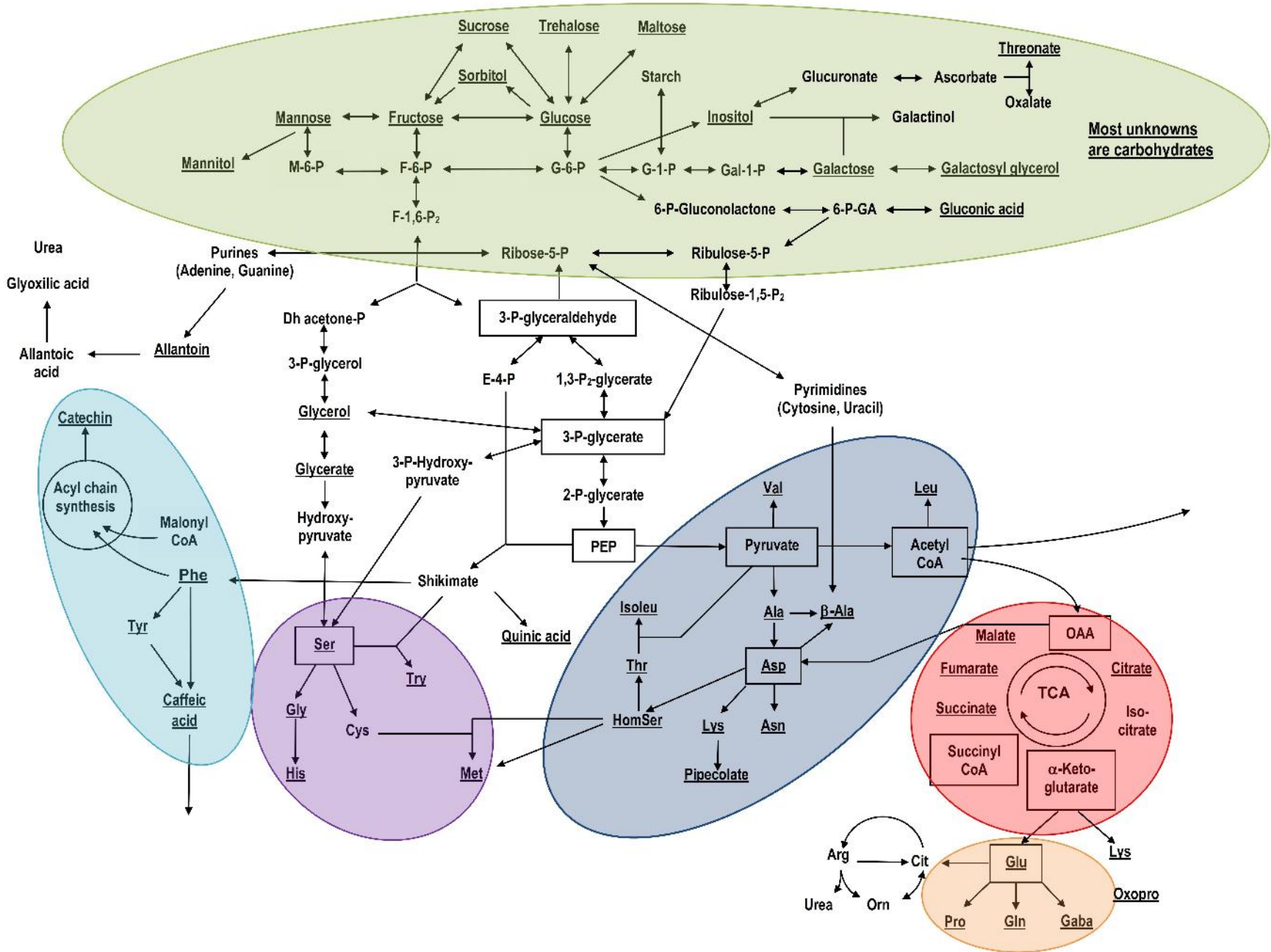175 Modified from the metabolomic pathway in *Solanum*, based on Dobson, et al. (7).

176

177 **References**

178 1.   **Akaike, H.** 1974. A new look at the statistical model identification. Ieee T Automat
179   Contr **19:**716-723.
180 2.   **Baranyi, J., and T. A. Roberts.** 1994. A dynamic approach to predicting bacterial
181   growth in food. Int J Food Microbiol **23:**277-94.
182 3.   **Baranyi, J., and T. A. Roberts.** 1995. Mathematics of predictive food microbiology.
183   Int. J. Food Microbiol. **26:**199-218.
184 4.   **Baranyi, J., and T. A. Roberts.** 2000. Principles and application of predictive
185   modelling of the effects of preservation factors on micro-organisms, p. 2024. *In* B.
186   Lund, T. C. Baird-Parker, and G. W. Gould (ed.), Microbiological safety and quality of
187   food, vol. 1. Asepn Publishers Inc, Maryland, USA.
188 5.   **Baty, F., C. Ritz, S. Charles, M. Brutsche, J. P. Flandrois, and M. L. Delignette-**
189   **Muller.** 2015. A toolbox for nonlinear regression in R: the package nlstools. J Stat
190   Softw **66:**1-21.
191 6.   **Burnham, K. P., and D. R. Anderson.** 2004. Multimodel inference. Sociol Method
192   Res **33:**261-304.
193 7.   **Dobson, G., T. Shepherd, R. Marshall, S. R. Verrall, S. Conner, D. W. Griffiths, J.**
194   **W. McNicol, D. Stewart, and H. V. Davies.** 2007. Dordrecht.
195 8.   **Dobson, G., T. Shepherd, S. R. Verrall, S. Conner, J. W. McNicol, G. Ramsay, L.**
196   **V. T. Shepherd, H. V. Davies, and D. Stewart.** 2008. Phytochemical diversity in

197 tubers of potato cultivars and landraces using a GC-MS metabolomics approach. J
198 Agri Food Chem **56:**10280-10291.

199 9. **Dunnett, C. W.** 1955. A multiple comparison procedure for comparing several
200 treatments with a control. Journal of the American Statistical Association **50:**1096-
201 1121.

202 10. **Husted, S., and J. K. Schjoerring.** 1995. Apoplastic pH and ammonium concentration
203 in leaves of *Brassica napus* L. Plant Physiol **109:**1453-1460.

204 11. **Lohaus, G., K. Pennewiss, B. Sattelmacher, M. Hussmann, and K. Hermann
205 Muehling.** 2001. Is the infiltration-centrifugation technique appropriate for the isolation
206 of apoplastic fluid? A critical evaluation with different plant species. Physiologia
207 Plantarum **111:**457-465.

208 12. **McKellar, R. C., and X. Lu.** 2004. Modeling microbial responses in food. CRC Press
209 LLC, Florida, USA.

210 13. **Motulsky, H. J.** 2017, posting date. GraphPad Curve Fitting Guide. GraphPad
211 Software. [Online.]

212 14. **Ratkowsky, D. A., R. K. Lowry, T. A. McMeekin, A. N. Stokes, and R. E. Chandler.**
213 1983. Model for bacterial culture growth rate throughout the entire biokinetic
214 temperature range. J Bacteriol **154:**1222-1226.

215 15. **Rico, A., and G. M. Preston.** 2008. *Pseudomonas syringae* pv. *tomato* DC3000 uses
216 constitutive and apoplast-induced nutrient assimilation pathways to catabolize
217 nutrients that are abundant in the tomato apoplast. Mol Plant Microbe Interact **21:**269-
218 282.

219 16. **Rodgers, J. L., and W. A. Nicewander.** 1988. Thirteen ways to look at the correlation
220 coefficient. The American Statistician **42:**59-66.

221 17. **Shepherd, T., G. Dobson, S. R. Verrall, S. Conner, D. W. Griffiths, J. W. McNicol,
222 H. V. Davies, and D. Stewart.** 2007. Potato metabolomics by GC–MS: what are the
223 limiting factors? Metabolomics **3:**475-488.

224 18. **Solomon, P. S., and R. P. Oliver.** 2001. The nitrogen content of the tomato leaf
225 apoplast increases during infection by *Cladosporium fulvum*. Planta **213:**241-9.

226 19. **Tukey, J. W.** 1949. Comparing individual means in the analysis of variance. Biometrics
227 **5:**99-114.

228

SFig. 1

SFig. 2