

# Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

## ARTICLE TYPE

**Metabolomic analysis reveals functional overlapping of three signal transduction proteins in regulating ethanol tolerance in cyanobacterium *Synechocystis* sp. PCC 6803**

**Ye Zhu**<sup>1,2,3</sup>, **Guangsheng Pei**<sup>1,2,3</sup>, **Xiangfeng Niu**<sup>1,2,3</sup>, **Mengliang Shi**<sup>1,2,3</sup>, **Mingyang Zhang**<sup>1,2,3</sup>, **Lei Chen**<sup>1,2,3</sup>\*, **Weiwen Zhang**<sup>1,2,3</sup>\*

<sup>1</sup>Laboratory of Synthetic Microbiology, School of Chemical Engineering & Technology, Tianjin University, Tianjin 300072, P.R. China;

<sup>2</sup>Key Laboratory of Systems Bioengineering, Ministry of Education of China, Tianjin 300072, P.R. China;

<sup>3</sup>Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, P.R. China

\* To whom correspondence should be addressed:

Prof. Dr. Lei Chen or Prof. Dr. Weiwen Zhang

Laboratory of Synthetic Microbiology

<sup>15</sup>School of Chemical Engineering & Technology

Tianjin University, Tianjin 300072, P. R. China

Tel : 0086-22-2740-6364

Email: [lchen@tju.edu.cn](mailto:lchen@tju.edu.cn) (LC) or [wwzhang8@tju.edu.cn](mailto:wwzhang8@tju.edu.cn) (WZ)

**Abstract**

Low ethanol tolerance is a crucial factor that restricts the feasibility of bioethanol production in renewable cyanobacterial systems. Our previous studies showed that several transcriptional regulators were differentially regulated by exogenous ethanol in *Synechocystis*. In this study, by constructing knockout mutants of 34 *Synechocystis* putative transcriptional regulator-encoding genes and analyzing their phenotypes under ethanol stress, we found that three mutants of regulatory gene *sll1392*, *sll1712* and *slr1860* grew poorly in BG11 medium supplemented with ethanol when compared with the wild type in the same medium, suggesting that the genes may be involved in the regulation of ethanol tolerance. To decipher the regulatory mechanism, a targeted LC-MS and an untargeted GC-MS approach were employed to determine metabolic profiles of the three mutants and the wild type under both normal and ethanol stress conditions. The results were then subjected to PCA and WGCNA analyses to determine the responsive metabolites and metabolic modules related to ethanol tolerance. Interestingly, the results showed that there was a significant overlapping of the responsive metabolites and metabolic modules between three regulatory proteins, suggesting that a possible crosstalk between various regulatory proteins may be involved in combating against ethanol toxicity in *Synechocystis*. The study provided new insights to ethanol-tolerance regulation and knowledge important to rational tolerance engineering in *Synechocystis*.

**Keywords:** Regulatory proteins, Ethanol tolerance, Metabolomics, Overlapping, *Synechocystis*

## Introduction

Biofuel production directly from photosynthetic microorganisms, such as cyanobacteria, has attracted significant interests in recent years, mostly due to the exciting progresses in employing synthetic biology approaches to engineer cyanobacteria for biofuel production directly from solar energy and CO<sub>2</sub> [1-3]. As the most widely used biofuel, ethanol production was also achieved in engineered cyanobacteria by introducing ethanol-biosynthetic pathway from native ethanol-producing bacterium such as *Zymomonas mobilis* [4, 5]. Using systematic evaluation and selection of alcohol dehydrogenase (*adh*) genes from different sources and optimization of culturing conditions, a relatively high ethanol production rate of 212 mg/L-day was recently reported in an engineered *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) [6]. However, compared with native ethanol-producing microbes, such as yeast *Saccharomyces cerevisiae* and bacterium *Z. mobilis*, ethanol yield in the renewable cyanobacterial systems is still very low. Among various possible reasons that may be responsible for the low productivity, low tolerance to ethanol toxicity in cyanobacteria has been suggested as one of the crucial factors [7, 8], and needs to be addressed by tolerance engineering [9].

To determine tolerance mechanism of ethanol stress, we have previously applied quantitative iTRAQ LC-MS/MS proteomics and RNA-Seq transcriptomics of *Synechocystis* grown under exogenous ethanol stress, and the results showed that multiple transcriptional regulators were differentially regulated, suggesting that ethanol tolerance could be under regulation of different types of signal transduction proteins, including response regulators of two-component signal transduction system (TCS), transcriptional regulators (TR) and eukaryotic-like protein phosphatases (PP) in *Synechocystis* [10, 11]. Due to high complexities of the cells, cellular networks are typically organized into various functional modules that can be individually controlled by different regulatory proteins, and the finding of multiple regulators involved in ethanol tolerance thus suggested that multiple biological modules may be required to combat ethanol toxicity in *Synechocystis*, which is consistent with early conclusions from other biofuel-producing species that the microbial cells tend to employ a combination of multiple cellular changes as protection mechanisms against biofuel toxicity [7, 8]. In addition, the finding also raised a question of how different signal proteins are coordinating to assure the proper functionality of the different biological modules related to ethanol tolerance [12]. Although no direct answer available to the question, early studies have found that coordination or crosstalk of different signal proteins in respond to environmental stress are commonly observed in various microbes. For example, Antigueira *et al.* (2012) recently integrated transcriptional, protein-protein and allosteric or equivalent interactions to understand the regulatory dynamics of transcription factors (TFs) and their interplay with other cellular components, and they found that there is an extensive crosstalk between TFs and their target genes in *Escherichia coli* [13]. In another study, Hanna *et al.* (2013) found that a (p)ppGpp-dependent crosstalk between at least three stress responses (*i.e.*, nutrient, oxidative, and low-oxygen stress) played a central role for *Brucella suis* to adapt to growth-affecting stress conditions such as nutrient deficiency, in the host cells [14]. In cyanobacteria, although only several cases of crosstalk between regulatory proteins

have been reported and characterized [15, 16], it was proposed that the crosstalk between signaling cascades may be even more elaborate than is currently believed, playing important roles in coordinating cellular responses to a variety of stress conditions [17].

In addition to engineering individual gene or enzyme for better biofuels tolerance, more evidences suggested that manipulation of regulatory genes could be a better option as it provides a route to achieve complex phenotypes that are not readily accessible by targeting small number of metabolic genes [18, 19]. For example, global transcription machinery engineering (gTME) approach has been applied to *S. cerevisiae* for improved glucose/ethanol tolerance [20], and expression of a mutated global regulator gene *irrE* from an extremely radiation-resistant bacterium, *Deinococcus radiodurans*, has led to 10- to 100-fold enhancement of *E. coli* tolerances to ethanol or butanol in shock experiments [21]. However, currently little information is available regarding regulation of ethanol tolerance in cyanobacteria, which makes it challenging for further tolerance engineering.

*Synechocystis* genome contains a number of genes encoding putative transcriptional regulators (TR) [22] and eukaryotic-like Ser/Thr protein phosphatases (PP) [23]. So far only several genes were functionally characterized, and the results showed that they were involved in the regulation of a wide-range of physiological functions, such as survival under nitrogen stress [24], survival under inorganic carbon starvation and osmotic stress [25], low CO<sub>2</sub>-induced activation of the bicarbonate transporter [26], biogenesis of photosystem I [27], photosystem stoichiometry in response to high light [28], ferric uptake [29], and heat response [30]. However, so far only one TR gene, *sll1392*, was found related to ethanol tolerance in *Synechocystis* in our recent study [11], and no PP has ever been reported related to tolerance to solvents or biofuel products in *Synechocystis*. To decipher regulatory networks related to ethanol tolerance and identify potential targets for ethanol tolerance engineering, in this study, we first constructed 34 knockout mutants for genes encoding most of the TRs and PPs in the *Synechocystis* genome, and then conducted comparative analysis of their ethanol tolerance. The efforts led to discovery of one TR-encoding genes *sll1712* and one PP gene *slr1860*, whose disruptions caused increased sensitivity to ethanol in *Synechocystis*, and also further confirmed the involvement of TR *sll1392* in ethanol tolerance [11]. We then applied a metabolomics approach employing both gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) to determine the time-series metabolic changes in the wild type and three *Synechocystis* mutants grown under normal and ethanol stress conditions [31, 32]. Finally, a principal component analysis (PCA) and a weighted correlation network analysis (WGCNA) approaches were applied to the metabolomic data to reveal the responsive metabolites and metabolic modules associated with each of the mutants under ethanol stress [33]. The results suggested a significant functional overlapping in regulating ethanol tolerance by three different regulatory proteins in *Synechocystis*.

## Experimental

### Bacterial growth conditions

*Synechocystis* and the TR or PP knockout mutants constructed in this study were grown in BG11 medium (pH

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

## ARTICLE TYPE

7.5) under a light intensity of approximately 50  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  in an illuminating incubator of 130 rpm at 30°C (HNY-211B Illuminating Shaker, Honour, China) [10, 11]. Cell density was measured at OD<sub>730</sub> on a UV-1750 spectrophotometer (Shimadzu, Japan). For control growth and ethanol treatment, 10 mL fresh cells at OD<sub>730</sub> of 0.5 were collected by centrifugation and then inoculated into 50 mL BG11 liquid medium in a 250-mL flask. 1.9% ethanol (v/v) at a final concentration was added at the beginning of cultivation. Ethanol of analytical pure was purchased from Merck (Whitehouse Station, NJ, U.S.A). For each condition, three biological replicates were established independently, and each sample was measured in triplicates. Growth experiments were repeated at least three times to confirm the growth patterns.

**Construction and analysis of knockout mutants**

A fusion PCR based method was employed for the construction of gene knockout fragments [34]. Briefly, for the gene target selected, three sets of primers were designed to amplify a linear DNA fragment containing the chloramphenicol resistance cassette (amplified from a plasmid pACYC184) with two flanking arms of DNA upstream and downstream of the targeted gene. The linear fused PCR amplicon was used directly for transformation into *Synechocystis* by natural transformation. The chloramphenicol-resistant transformants were obtained and passed several times on fresh BG11 plates supplemented with 10  $\mu\text{g/mL}$  chloramphenicol to achieve full chromosome segregation (confirmed by PCR). The successful knockout mutants were confirmed by PCR and sequencing analysis, coupled with real-time RT-PCR. PCR primers for mutant construction and validation were listed in **Suppl. Table 1**. Similar to the cultivation conditions described above, comparative analysis of the wild type *Synechocystis* and the mutants were performed in 100-mL flasks each with 25 mL BG11 medium with or without ethanol of various concentration, and each condition was performed in biological triplicates. All the cultures were sampled and measured every 12 h.

**Real-time RT-PCR analysis**

Real-time RT-PCR analysis was performed as described previously [11]. Quantification of gene expression was determined according to standard process of RT-PCR which used serial dilutions of known concentration of chromosome DNA as template to make a standard curve. The 16s rRNA was used as an internal control. Three technical replicates were performed for each gene. Data analysis was carried out using the StepOnePlus analytical software (Applied Biosystems, Foster City, CA). Briefly, the amount of relative gene transcript was normalized by that of 16s rRNA in each sample (mutant or wild type), and the data presented were ratios of the amount of normalized transcript in the mutant compared with the wild type. RT-PCR primers for three genes were listed in **Suppl. Table 1**.

**Targeted LC-MS Based Metabolomics Analysis**

Due to its advantages toward chemically unstable metabolites, LC-MS based metabolomic analysis has

become increasingly popular method to investigate microbial metabolism recently [31, 35]. These unstable metabolites including the redox active nucleotides (NADPH, NADH) and the hydrolytically unstable nucleotides (ATP, GTP, cAMP, PEP) crucial for all major metabolic pathways in cells [36-38]. Most recently, LC-MS metabolomic analysis was also applied to characterize changes in the cyanobacterial primary metabolism under diverse environmental conditions or in defined mutants. The resulting identification of metabolites and their steady state concentrations have provided a better understanding of cyanobacterial metabolism [39]. A method of isolation and tandem LC-MS/MS quantification of a subset of targeted internal metabolites was previously established for *Synechococcus* sp. PCC 7002 [36]. Briefly, *i) Sample quenching, extraction, and preparation*: Chemicals used for LC-MS metabolomic isolation was purchased from Sigma-Aldrich (Taufkirchen, Germany). For metabolomic analysis, cells were collected from normal and ethanol-stressed cultures of the wild type and the mutants at 48 h and 72 h, respectively. For each sample, cells equivalent to 10<sup>6</sup> cells, were collected by centrifugation at 8,000 x g for 8 min at room temperature (Eppendorf 5430R, Hamburg, Germany). The cell samples were quenched and extracted rapidly with 900  $\mu\text{L}$  of 80:20 MeOH/H<sub>2</sub>O (-80°C) and then frozen in liquid nitrogen. The samples were then frozen-thawed three times to release metabolites from the cells. The supernatant was collected after centrifugation at 15,000 x g for 5 min at -4°C and then stored at -80°C. The remaining cell pellets were re-suspended in 500  $\mu\text{L}$  of 80:20 MeOH/H<sub>2</sub>O (-80°C) and the above extraction process was repeated. The supernatant from the second extraction was pooled with that from the first extraction and stored at -80°C until LC-MS analysis [36]; *ii) LC-MS analysis*: The chromatographic separation was achieved with a SYnergi Hydro-RP (C18) 150 mm x 2.0 mm I.D., 4  $\mu\text{m}$  80 Å particles column (Phenomenex, Torrance, CA, USA) at 40°C. Mobile phase A (MPA) is an aqueous 10 mM tributylamine solution with pH 4.95 adjusted with acetic acid and Mobile phase B (MPB) is 100% methanol of HPLC grade (Darmstadt, Germany). The optimized gradient profile was determined as follows: 0 min (0% B), 8 min (35% B), 18 min (35% B), 24 min (90% B), 28 min (90% B), 30 min (50% B), 31 min (0% B). A 14-minute post-time equilibration was employed, bringing total run-time to 45 min. Flow rate was set as a constant 0.2 mL/min [40]. LC-MS analysis was conducted on an Agilent 1260 series binary HPLC system (Agilent Technologies, Waldbronn, Germany) coupled to an Agilent 6410 triple quadrupole mass analyser equipped with an electrospray ionization (ESI) source. Injected sample volume for all cases was 10  $\mu\text{L}$ ; capillary voltage was 4000 V; and nebulizer gas flow rate and pressure were 10 L/min and 50 psi, respectively. Nitrogen nebulizer gas temperature was 300°C. The MS was operated in negative mode for multiple reaction monitoring (MRM) development, method optimization, and sample analysis. Data were acquired using Agilent Mass Hunter workstation LC/MS

285 acquisition software (version B.04.01) and chromatographic peaks were subsequently integrated via Agilent Qualitative Analysis software (version B.04.00); *iii*) *Targeted metabolite analysis*: a total of 24 metabolites were selected for LC-MS based targeted metabolite analysis in this study. 290 The abbreviations, molecular weights and MRM values determined and optimized for each of the 24 detected metabolites as well as the product ion formulas were provided in **Suppl. Table 2**. The standard compounds for these 24 metabolites were purchased from Sigma, and their 295 MS and MS/MS experimental parameters were optimized with the mix standard solution. The concentration of each standard metabolite used for analysis is 50  $\mu$ M. All metabolomics profile data was normalized by the internal control and the cell numbers of the samples, and then 300 subjected to principal component analysis (PCA) using software SIMCA-P 11.5 [41].

### GC-MS Based Metabolomics Analysis

Untargeted GC-MS based metabolomics is capable of 305 detecting a wide range of chemical metabolite classes in a single run, and achieving good coverage of polar metabolites, such as amino acids and organic acids, making it a powerful technique in deciphering metabolic response in cells [39]. Using a modified protocol established by Krall *et al.* (2009) [42], we previously applied the GC-MS 310 metabolomics to *Synechocystis* under butanol and salt stress, and the results have aided the exploration of the mechanism responsive to these stresses [43, 44]. All chemicals used for metabolome isolation and GC-MS analyses were obtained 315 from Sigma-Aldrich (Taufkirchen, Germany). The identical cells collected for LC-MS analysis were also used for GC-MS metabolomic analysis. The metabolomic analysis protocol included: *i*) *Metabolome extraction*: cells were re-suspended in 1.0 mL cold 10:3:1 (*v/v/v*) 320 methanol:chloroform:H<sub>2</sub>O solution (MCW), and frozen in liquid nitrogen and thawed for five times. Supernatants were collected by centrifugation at 15,000  $\times$  g for 3 min at 4°C (Eppendorf 5430R, Hamburg, Germany). To normalize variations across samples, an internal standard (IS) solution 325 (100  $\mu$ g/mL U-<sup>13</sup>C-sorbitol, 10  $\mu$ L) was added to 100  $\mu$ L supernatant in a 1.5-mL microtube before it was dried by vacuum centrifugation for 2-3 h (4°C). *ii*) *Sample derivatization*: derivatization was conducted according to the two-stage technique by Roessner *et al.* (2001) [45]. The 330 samples were dissolved in 10  $\mu$ L methoxyamine hydrochloride (40 mg/mL in pyridine) and shaken at 30°C for 90 min, then were added with 90  $\mu$ L N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and incubated at 37°C for 30 min to trimethylsilylate 335 the polar functional groups. The derivatized samples were collected by centrifugation at 15,000  $\times$  g for 3 min before GC/MS analysis. *iii*) *GC-MS analysis*: analysis was performed on a GC-MS system-GC 7890 coupled to an MSD 5975 (Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with a HP-5MS capillary column (30 m  $\times$  250 mm id). 2  $\mu$ L derivatized sample was injected in splitless mode at 230°C injector temperature. The GC was operated at constant flow of 1 mL/min helium. The temperature program started isocratic at 45°C for 2 min, 345 followed by temperature ramping of 5°C/min to a final temperature of 280°C, and then held constant for additional 2 min. The range of mass scan was *m/z* 38-650. *iv*) *Data processing and statistical analysis*: The mass fragmentation spectrum was analyzed using the Automated Mass Spectral 350 Deconvolution and Identification System (AMDIS) [46] to identify the compounds by matching the data with Fiehn

Library [47] and the mass spectral library of the National Institute of Standards and Technology (NIST). Peak areas of all identified metabolites were normalized against the 355 internal standard and the acquired relative abundances for each identified metabolite were used for future data analysis. All metabolomics profile data was normalized by the internal control and the cell numbers of the samples, and assessed by a PCA analysis using software SIMCA-P 11.5 360 [41]. To further reveal association between metabolite dynamics and the conditions/mutants, two-dimensional scatter plotting analysis for the conditions/mutants and metabolites was performed [48]. As both 365 conditions/mutants and metabolites were assumed to be of a unit variance, their projections on the plane reside within a circle of radius 1 centered at the origin. Variables with a strong relationship are projected in the same direction from the origin. Basically, the greater the distance from the origin, the stronger is the relationship. In this study, the third circle 370 with radius of 0.5 was selected as a cutoff for high associations between conditions/mutants and metabolites, as described in previous CCA analysis between genes and metabolites involved in *Escherichia coli* primary metabolism [48]. 375

### WGCNA Network Construction

Correlation network was created from the GC-MS metabolomic data, by calculating weighted *Pearson* correlation matrices corresponding to metabolite abundance, 380 and then by following the standard procedure of WGCNA to create the networks [32, 33, 49]. Briefly, weighted correlation matrices were transformed into matrices of connection strengths using a power function [33]. These connection strengths were then used to calculate topological 385 overlap (TO), a robust and biologically meaningful measurement that encapsulates the similarity of two metabolites' correlation relationships with all other metabolites in the network [33]. Hierarchical clustering based on TO was used to group metabolites with highly 390 similar correlation relationships into modules. Metabolite dendrograms were obtained by average linkage hierarchical clustering [32, 33, 49, 50], while the color row underneath the dendrogram showed the module assignment determined by the Dynamic Tree Cut of WGCNA. The network for 395 each module was generated with the minimum spanning tree with dissimilarity matrix from WGCNA. The modules with correlation  $r > 0.5$ , and *p*-value less than 0.05 were extracted for further investigation. Hub metabolites were screened by high connectivity with other metabolites ( $\geq 5$ ) 400 in the modules strongly associated with phenotype.

## Results and Discussion

### Ethanol tolerance analysis of the transcriptional regulator mutants

405 To identify possible TRs and PPs involved in ethanol tolerance, we constructed a mutant library including 34 putative TR and PP coding genes. Except for *sll1626*, *sll1423*, *sll0567* and *sll1712* genes where only partial segregation can be achieved after more than ten passages 410 (data not shown), chromosomal integration and full segregation were confirmed by PCR and sequencing for the other 30 regulatory genes. All mutants were screened for ethanol tolerance changes, first in 96-well cultivation plates and then confirmed in flask cultivation, in parallel with the 415 wild type *Synechocystis* in both normal BG11 medium and the BG11 medium supplemented with 1.9% (*v/v*) ethanol. The full list of TR mutants constructed in this study was provided in **Suppl. Table 1**. The comparative analysis

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

## ARTICLE TYPE

showed that two TR mutants,  $\Delta sll1392$  and partially segregated  $\Delta sll1712$ , and a PP mutant  $\Delta slr1860$ , grew poorly in the BG11 medium supplemented with ethanol, when compared with wild type, suggesting that they were more sensitive to ethanol and the genes may be involved in ethanol tolerance in *Synechocystis*. The involvement of  $sll1392$  in ethanol tolerance was reported previously and further confirmed in this study [11]. In addition, in the normal BG11 medium without ethanol, the mutants grew equally well as the wild type, suggesting the deletion of any of three genes had no negative effects on cell growth (Fig. 1). The fact that full segregation cannot be achieved for the  $\Delta sll1712$  mutant after more than 10 passages under selective pressure suggested that the gene might be essential for the growth under the test condition (Suppl. Fig. 1). To demonstrate the slow growth of the  $\Delta sll1712$  mutant under ethanol stress was due to down-regulation of  $sll1712$  gene, RT-qPCR assay was conducted, and the results demonstrated that the  $sll1712$  gene was down-regulated by about two and half folds in the partially segregated  $\Delta sll1712$  mutant, while expression of  $sll1392$  and  $slr1860$  dropped to undetectable levels (Suppl. Fig. 2). Cell morphologies of the wild-type *Synechocystis* and three mutants with or without ethanol stress were also compared using microscopic and flow cytometric analyses; however, no visible difference between the wild type and the mutants was observed (data not shown).

According to sequence analysis,  $sll1392$  and  $sll1712$  were determined as DNA-binding transcriptional regulators, while  $slr1860$  was one of the seven eukaryotic-like Ser/Thr protein phosphatases in the *Synechocystis* genome [23]. Early study found that Slr1860 exhibited divalent metal ion-dependent protein-serine phosphatase activity that catalyzes the dephosphorylation of phosphoprotein substrate Slr1856 *in vitro* [51]. Although this is the first time that these three genes were found functionally related to ethanol tolerance, they have been previously found involved in responses to various environmental stress in *Synechocystis*. The  $sll1392$  gene was also named as  $pfsR$ , an acronym for photosynthesis, iron homeostasis, and stress response [52]. In addition, the gene was identified as part of core transcriptional response in *Synechocystis* using large-scale expression profiles under various growth conditions [53], and found to be up-regulated following transfer from pH 7.5 to pH 10 [54]. The  $sll1712$  gene was repressed in response to UV-B and white light irradiation in *Synechocystis* [55], down-regulated by 3% and 0.3% inorganic phosphate (Pi) [56]; however, it was positively regulated by cadmium [57] and hexane [58]. The  $slr1860$  gene was also repressed in response to UV-B and white light irradiation in *Synechocystis* [55], induced by glucose or light deprivation, and required for cell growth under conditions of low concentration of inorganic carbon in the presence of glucose [59]; in addition, early study showed that the regulation of glucose catabolism by a histidine kinase Hik31 may be involved in  $slr1860$  phosphatase in *Synechocystis* [60].

#### Targeted LC-MS metabolomic analysis

Using the method with minor modifications, we eventually established reproducible analyses of 24 selected standard

metabolites for *Synechocystis*, including NADPH, NADP, NADH, acetyl coenzyme A (AcCoA),  $\alpha$ -nicotinamide adenine dinucleotide (NAD), adenosine-5'-diphosphoglucose (ADP-GCS), uridine 5'-diphosphoglucose (UDP-GCS), adenosine 5'-triphosphate (ATP), adenosine 5'-diphosphate (ADP), coenzyme A hydrate (COA), adenosine 5'-monophosphate (AMP), *D*-fructose 1,6-bisphosphate (FBP), *D*-ribulose 1,5-bisphosphate (RiBP), *D*-fructose 6-phosphate (F6P), *D*-glucose 6-phosphate (G6P), *D*-ribose 5-phosphate (R5P), *D*-(-)-3-phosphoglyceric acid (3PG), dihydroxyacetone (DHAP), *DL*-glyceraldehyde 3-phosphate (GHAP), phospho(enol)pyruvic acid (PEP), *L*-glutamic acid (GLU),  $\alpha$ -ketoglutaric acid (AKG), oxaloacetic acid (OXA), and sodium fumarate dibasicadenosine (FUM), most of which are unstable metabolites in central carbohydrate metabolism together with those involved in the cellular energy charge and redox poise in *Synechocystis* (Suppl. Table 2) [36]. Using pure metabolites as references, a semi-quantitative characterization of all 24 metabolites was achieved for all the cell samples.

Cell samples of the wild type,  $\Delta sll1392$ ,  $\Delta sll1712$  and  $\Delta slr1860$  mutants grown in BG11 media with or without ethanol were collected at two time points (*i.e.*, 48 and 72 h), which are corresponding to the middle-exponential phase and transition between exponential and stationary phases, respectively. Each condition analysis consisted of three biological replicates and two technical replicates. With the optimized LC-MS protocol, sixteen sets of the metabolomic profiles each with 24 metabolites detected were then obtained. The data was normalized by the internal control and the cell numbers (Suppl. Table 3) and then analyzed by PCA plots (Fig. 2A). The results showed that: *i*) clustering of the biological replicates for each sample was clearly observed for each condition, suggesting overall good quality of the LC-MS analysis; *ii*) the well separation of metabolomic profiles of the control and the ethanol-treated samples was clearly observed at both 48 and 72 h, suggesting that significant metabolic changes occurred upon ethanol stress in both wild type and mutants; in addition, the results also demonstrated that the LC-MS methodology we applied in the study is sensitive enough to distinguish the cellular responses upon stress; *iii*) in the control growth condition without ethanol, difference between three mutants and the wild type was found relatively small, consistent with the growth patterns observed in Fig. 1. The clustering of the wild type and the mutants was more obvious at 72 h, probably due to the fact that the cell aging became a dominating factor at this stage; *iv*) in the growth condition with ethanol stress, significantly different responses between the wild type and the mutants were observed at both 48 and 72 h, further confirming that these genes may be involved in response to ethanol stress; *v*) between the three mutants, the results showed that  $\Delta sll1392$  and  $\Delta slr1860$  seemed shared a slightly similar response to ethanol, while metabolic responses in  $\Delta sll1712$  were significantly different from the two other mutants, especially at 72 h after long time of ethanol stress, suggesting a

difference tolerance mechanism may be utilized in the *Δsll1712* mutant.

Heat maps of 24 metabolites in the wild type and the mutants at both 48 h and 72 h were generated (Fig. 2B). In the analysis, the ratio of a given metabolites was calculated between the concentration of the metabolite under a given condition and the average concentration of the metabolite in all samples. The similar approach has been successfully applied in transcriptomic analysis [61]. The results showed very significant up-regulation of a dozen metabolites under the ethanol stress conditions in the three mutants when compared with the wild type, including G6P, NAD, F6P, AMP, R5P, AcCoA, DHAP, NADP and OXA, suggesting an overall up-regulation of central carbohydrate metabolic pathway and its possible role in ethanol stress tolerance. In our previous proteomic analysis of *Synechocystis* under ethanol stress, we found that a number of proteins involved in core carbohydrate metabolism, such as pyruvate dehydrogenase, succinate dehydrogenase, 6-phosphofructokinase, fructose-1,6-bisphosphatase, and phosphoenolpyruvate carboxylase, were down-regulated [10]. The up-regulated metabolites in the mutants grown under ethanol stress thus suggested that the regulatory proteins may be functional as negative effectors for the central carbohydrate metabolism, although the details remain to be determined. Meanwhile, it is interesting that the intracellular abundances of these metabolites were regulated by all three regulatory genes, suggesting a possible functional overlapping between the networks regulated by three individual regulatory genes. Although none of the metabolites have been found previously related to ethanol tolerance, changes of metabolites such as F6P, G5P, R5P, OXA and NADP have been found related to environmental stress in various organisms [62-65]. In addition, the results also showed that NADPH, NADP, ADP and ATP were found up-regulated only in the *Δsll1712* mutant, consistent with its different metabolic status upon ethanol stress from other two mutants, as showed in the PCA plots (Fig. 2A). Moreover, a very similar pattern was observed between 48 h and 72 h in the heat maps, although the up-regulation patterns of the metabolites described above seemed becoming more significant at 72 h.

#### Untargeted GC-MS metabolomic analysis

In this study, following the same sampling and analytical strategy for LC-MS metabolomic analysis as described above, we collected cells of the wild type and three mutants (*i.e.* *Δsll1392*, *Δsll1712* and *Δslr1860*) grown in BG11 media with or without ethanol treatment at both 48 and 72 h. For each sample, three biological replicates were independently cultivated, metabolites-isolated, and analyzed by a GC-MS as described before [32, 43, 44]. Under the optimized analytical conditions, a good separation of intracellular metabolites was achieved on the GC column and the further MS analysis allowed chemical classification of a total 45 metabolites from all samples, including various amino acids, sugars and organic acids (Suppl. Table 4). Overall quality of the datasets was assessed first by a PCA analysis. In general, the score plots of the GC-MS metabolomic profiles revealed very similar patterns as we described above for the LC-MS metabolomic profiles, such as overall good reproducibility between biological replicates and a good separation between samples with or without ethanol stress. In addition, under ethanol stress condition, the relatively different metabolic status of the *Δsll1712* mutant from other two mutants (*i.e.* *Δsll1392* and *Δslr1860*)

was also observed in the GC-MS metabolomic datasets (Fig. 3). However, some differences from the LC-MS based analysis were also observed: for example, at 48 h, the *Δsll1392* mutant grown under ethanol stress were close to its growth without ethanol stress, suggesting that the metabolic changes in the *Δsll1392* mutant with or without ethanol, as reflected from the GC-MS metabolomic measurements, were not so significant, which may also highlight the values to utilize both LC-MS and GC-MS for a complete interpretation of metabolic status in cells.

As the major goal of this study is to determine response mechanisms mediated by each of the regulatory genes, we further employed two computational approaches, a two-dimensional scatter PCA analysis and a WGCNA network analysis, to the GC-MS metabolomic datasets in an attempt to determine the responsive metabolites and metabolic modules in each mutant and to establish links of metabolites, metabolic modules to the genes. The first approach, a two-dimensional scatter PCA plot that can be used to study associations between two sets of variables measured under the same experimental units [48], has been applied to integrate data originating from different 'omics' technologies [48, 66], such as clustering properties of metabolites and transcripts involved in primary metabolism in *E. coli* [48]. In this study, we assumed two variables as metabolites and the samples under growth conditions (*i.e.*, control or ethanol stress), which thus allowed determination of association between responsive metabolites and the stress condition in each of the mutants at each time point (*i.e.*, 48, 72). In each of the plots, we included the wild type and one of the mutants grown with or without ethanol stress, and the metabolomic measurements associated with these four samples (Fig. 4). The results showed that all four samples can be visibly separated in each of the plots; in addition, profiles of the wild type and the mutant grown under control conditions (*i.e.*, without ethanol) tended to be close in the plots, consistent with their similar growth patterns under control condition (Fig. 1). However, under ethanol stress condition, obvious separation between the mutant and the wild type can be observed, although the separation patterns could be different between three mutants (Fig. 4). The visualization of their correlations showed that for any given sample, a number of metabolites were found in close proximity to it, suggesting these metabolites may be highly associated with the cellular responses in the given sample. Using a cut off of 0.5 (*i.e.*, insider cycle), a dozen of metabolites that are important in differentiating the wild type and each of the mutants were identified for 48 and 72 h, respectively. At 48 h, we found *L*-serine clustered with *Δsll1392*, *L*-threonine, *L*-serine, succinic acid, myristic acid, *L*-pyroglutamic acid, palmitic acid, stearic acid, *L*(+) lactic acid, sucrose and 3-hydroxypyridine clustered with *Δsll1712*, *L*-threonine, glycerol, sucrose, methyl-beta-*D*-galactopyranoside, succinic acid, *L*-serine, urea and glyceric acid clustered with *Δslr1860*, respectively. Pie chart analysis showed that only *L*-serine was common among all three mutants at 48 h, where between *Δsll1712* and *Δslr1860*, succinic acid, sucrose and *L*-threonine were also commonly shared (Fig. 5A). While at 72 h, we found methyl palmitate, sucrose, methyl stearate, *L*-pyroglutamic acid, glycerol 1-phosphate, oleic acid and *D*-malic acid clustered with *Δsll1392*, phytol, heptadecanoic acid, palmitic acid, *D*-malic acid, glycerol 1-phosphate, glycine, linoleic acid, methyl palmitate, adenosine, sucrose clustered with *Δsll1712*, methyl palmitate, benzoic acid, stearic acid, sucrose, glycine, oleic acid, heptadecanoic acid clustered with *Δslr1860*, respectively. Pie chart analysis showed that

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

## ARTICLE TYPE

methyl palmitate and sucrose were common among all three mutants at 72 h (Fig. 5B). Meanwhile, the responsive metabolites shared by each pair of the mutants were presented in Fig. 5, suggesting there were also differences in terms of the constitute metabolites of the networks they mediated.

The roles of amino acid and sucrose in butanol tolerance were reported previously in *E. coli* [32, 67]. In a transcriptomic analysis of *E. coli* upon ethanol stress, results showed an increased expression level of genes related to amino acid metabolism (*i.e.*, tryptophan, histidine, valine, leucine, and isoleucine) [67]. While in a metabolomic and network analysis of *E. coli* upon butanol stress, we identified two butanol-associated modules, one containing serine, and another containing sucrose [32]. In addition, our previous metabolomics study showed that serine, sucrose and methyl palmitate were differentially regulated in *Synechocystis* under butanol stress [43], suggesting these metabolites could be involved in tolerance to biofuels. More importantly, the fact that these putative tolerance-related metabolites seemed regulated by three different regulatory proteins suggested a possible functional overlapping between three regulatory networks in *Synechocystis* upon ethanol stress.

In a second approach, we applied a WGCNA network analysis method to the GC-MS metabolomic datasets. WGCNA is a correlation-based and unsupervised computational method to describe and visualizes correlation patterns of data points [49, 50]. Previous comparative analysis demonstrated that the WGCNA method has overall advantages than other commonly used statistic methods such as PCA and batch learning self-organizing maps (BL-SOM) as it can additionally define “modules” of co-expressed metabolites explicitly and provide additional network statistics that described the systems properties of metabolic networks [68]. Following the standard protocol, we first constructed unsigned networks for 48 and 72 h, and then localized the correlated metabolites into various metabolic modules identified in the networks. In addition, the association of each distinguished metabolic module with sample conditions was determined, as highly associated modules indicated on the plots (Fig. 6). Setting a minimal number of metabolites in any module greater than 3, the WGCNA analysis identified 6 and 6 distinct metabolic modules within the metabolic networks of 48 and 72 h, respectively. Using a cutoff of correlation coefficients (*r* value) between module and sample condition greater than 0.5 and their statistical confidence (*p* value) less than 0.05, a total of 3 and 4 distinguished metabolic modules with high association with a given mutant or ethanol stress condition were identified from 48 h and 72 h, respectively. Metabolites included in each of the highly associated modules were indicated inside the plots. As shown in Fig. 6, except for the module M7 which was highly associated only with the gene knockout in the *Δslr1860* mutant at 72 h with *r*=0.53 and *p* value =0.008, all other six modules were found associated with positively ethanol stress. Among them the modules M1 and M5 were highly associated only with ethanol stress at 48h and 72 h, respectively, while four reminding modules (*i.e.*, module M2, M3, M4, and M6)

were associated with both ethanol stress and the gene knockout events, which may represent the key ethanol-induced responses in each of the mutants. Interestingly, while three modules (*i.e.*, M2, M3 and M6) were associated with *Δsl11712* and one module M4 associated *Δslr1860*, no module was identified associated with *Δsl11392*, which was consistent with the PCA analysis that the metabolic changes in the *Δsl11712* mutant under ethanol stress seemed significantly different from other two mutants.

Hub metabolites are those with high degree of connectivity in biological interaction networks and are thus supposed with high biological importance [69]. Assuming hub metabolites with connectivity greater than 5, from the WGCNA network we were able to identify seven hub metabolite, glutamic acid and pyroglutamic acid in the module M2 associated with both ethanol stress and *Δsl11712* at 48 h, sucrose and methyl palmitate in the module M4 associated with both ethanol stress and *Δslr1860*, linoleic acid and urea in the module M6 associated with both ethanol stress and *Δsl11712*, heptadecanoic acid in the module M7 associated with *Δslr1860* at 72 h (Fig. 7). Analysis of the hub-associated metabolites showed that the two networks associated with *Δsl11712* contains more amino acids (Fig. 7AC), while the two networks associated with *Δslr1860* contained more metabolites related to fatty acid metabolism (Fig. 7BD), which may reflect the different tolerance mechanism mediated by two regulatory proteins.

In a previous proteomic analysis of *Synechocystis* under various environmental perturbations, a cyanophycinase involved in the breakdown of cyanophycin, a storage molecule for excess carbon and nitrogen, into arginine and aspartic acid, was found moderately up-regulated [70]. Our previous metabolomic analysis confirmed this result by showing aspartic acid, succinic acid and *L*-glutamic acid were induced by butanol treatment [43], as arginine and aspartic acid can be further converted to glutamate and succinate, respectively [71]. The identification of glutamic acid as a hub metabolite in the *Δsl11712* mutant sensitive to ethanol further suggested that the cyanophycin degradation pathway may be involved in tolerance response to biofuels, and may worth further investigation. In addition, serine, methyl palmitate and sucrose identified by two-dimensional scatter PCA analysis were also found located in the module M2 associated with both ethanol stress and *Δsl11712* at 48 h, and the module M4 associated with both ethanol stress and *Δslr1860* at 72 h, respectively. Moreover, a good agreement was found between results from the WGCNA and the PCA analysis, both indicating that serine, sucrose and methyl palmitate could be important metabolites in combating ethanol toxicity, and their abundances were under regulation of three regulatory proteins, Sl11392, Sl11712 and Slr1860.

## Conclusions

In this study, by constructing and screening knockout

mutants for 34 regulator-encoding genes, two transcriptional regulators (TR) and one eukaryotic-like protein phosphatases (PP) were found involved in ethanol tolerance in *Synechocystis*. To decipher the regulation, we used a targeted LC-MS and an untargeted GC-MS approaches to determine metabolic profiles of the three mutants and the wild type under both normal and ethanol stress conditions. The metabolomic analysis along with network construction allowed identification of a few dozen stable and unstable metabolites, and seven metabolic modules related to ethanol tolerance in the *Synechocystis* mutants. Further comparative analysis of these responsive metabolites and metabolic modules showed an obvious functional overlapping between three networks mediated by the regulatory genes, suggesting a possible crosstalk between them. Although it is well established that microbial cells typically utilize a combination of multiple cellular changes as protection mechanisms to combat biofuel toxicity [7, 8, 10, 11, 37], this is the first report that the metabolic responses could be co-regulated by several different regulatory genes. The study provided new insights to the regulation of ethanol tolerance and useful knowledge key to tolerance engineering in *Synechocystis*.

### Abbreviations

**3PG**: D(-)-3-phosphoglyceric acid; **AcCoA**: acetyl coenzyme A; **ADP**: adenosine 5'-diphosphate; **ADP-GCS**: adenosine 5'-diphosphoglucose; **AKG**:  $\alpha$ -ketoglutaric acid; **AMDIS**: Automated Mass Spectral Deconvolution and Identification System; **AMP**: adenosine 5'-monophosphate; **ATP**: adenosine 5'-triphosphate; **BL-SOM**: batch learning self-organizing maps; **COA**: coenzyme A hydrate; **DHAP**: dihydroxyacetone; **ESI**: electrospray ionization; **F6P**: D-fructose 6-phosphate; **FBP**: D-fructose 1,6-bisphosphate; **FUM**: sodium fumarate dibasic; **G6P**: D-glucose 6-phosphate; **GAP**: DL-glyceraldehyde 3-phosphate; **GC-MS**: Gas Chromatography-Mass Spectrometry; **GLU**: L-glutamic acid; **gTME**: global transcription machinery engineering; **IS**: internal standard; **LC-MS**: Liquid Chromatography-Mass Spectrometry; **MPA**: mobile phrase A; **MPB**: mobile phrase B; **MRM**: multiple reaction monitoring; **MSTFA**: N-methyl-N-(trimethylsilyl) trifluoroacetamide; **NAD**:  $\alpha$ -nicotinamide adenine dinucleotide; **NIST**: National Institute of Standards and Technology; **OXA**: oxaloacetic acid; **PCA**: Principal Component Analysis; **PEP**: phospho(enol)pyruvic acid; **PP**: eukaryotic-like Ser/Thr protein phosphatases; **R5P**: D-ribose 5-phosphate; **RiBP**: D-ribulose 1,5-bisphosphate; **TCS**: two-component signal transduction system; **TFs**: transcription factors; **TO**: topological overlap; **TR**: transcriptional regulators; **UDP-GCS**: uridine 5'-diphosphoglucose; **WGCNA**: Weighted Correlation Network Analysis.

### Acknowledgments

The research was supported by grants from National Basic Research Program of China ("973" program, project No. 2011CBA00803 and No. 2014CB745101), National High-tech R&D Program ("863" program, project No. 2012AA02A707) and National Science Foundation of China (NSFC, project No. 31470217).

### References

1. R. A. Kerr, *Science*, 2005, **309**, 101-101.

2. N. Quintana, F. Van der Kooy, M. D. Van de Rhee, G. P. Voshol and R. Verpoorte, *Appl. Microbiol. Biot*, 2011, **91**, 471-490.
3. I. M. Machado and S. Atsumi, *J. Biotechnol*, 2012, **162**, 50-56.
4. M. D. Deng and J. R. Coleman, *Appl. Environ. Microb*, 1999, **65**, 523-528.
5. J. Dexter and P. Fu, *Energ. Environ. Sci*, 2009, **2**, 857-864.
6. Z. Gao, H. Zhao, Z. Li, X. Tan and X. Lu, *Energ. Environ. Sci*, 2012, **5**, 9857-9865.
7. S. A. Nicolaou, S. M. Gaida and E. T. Papoutsakis, *Metab. Eng*, 2010, **12**, 307-331.
8. M. J. Dunlop, *Biotechnol. Biofuels*, 2011, **4**, 32.
9. H. Jin, L. Chen, J. Wang and W. Zhang, *Biotechnol. Adv*, 2014, **32**, 541-548.
10. J. Qiao, J. Wang, L. Chen, X. Tian, S. Huang, X. Ren and W. Zhang, *J. Proteome Res*, 2012, **11**, 5286-5300.
11. J. Wang, L. Chen, S. Huang, J. Liu, X. Ren, X. Tian, J. Qiao and W. Zhang, *Biotechnol. Biofuels*, 2012, **5**, 89.
12. M. Y. Galperin, *Environ. Microbiol*, 2004, **6**, 552-567.
13. L. Antigueira, S. C. Janga and L. da Fontoura Costa, *Mol. Biosyst*, 2012, **8**, 3028-3035.
14. N. Hanna, S. Ouahrani-Bettache, K. L. Drake, L. G. Adams, S. Köhler and A. Occhialini, *BMC Genomics*, 2013, **14**, 459.
15. M. Asayama, S. Imamura, S. Yoshihara, A. Miyazaki, N. Yoshida, T. Sazuka, T. Kaneko, O. Ohara, S. Tabata and T. Osanai, *Biosci. Biotech. Bioch*, 2004, **68**, 477-487.
16. S. López-Gormollón, J. A. Hernandez, S. Pellicer, V. E. Angarica, M. L. Peleato and M. F. Fillat, *J. Mol. Biol*, 2007, **374**, 267-281.
17. R. Schwarz and K. Forchhammer, *Microbiology*, 2005, **151**, 2503-2514.
18. H. Alper and G. Stephanopoulos, *Metab. Eng*, 2007, **9**, 258-267.
19. D. Petranovic and G. N. Vemuri, *J. Biotechnol*, 2009, **144**, 204-211.
20. H. Alper, J. Moxley, E. Nevoigt, G. R. Fink and G. Stephanopoulos, *Science*, 2006, **314**, 1565-1568.
21. T. Chen, J. Wang, R. Yang, J. Li, M. Lin and Z. Lin, *PLoS One*, 2011, **6**, e16228.
22. T. Kaneko, Y. Nakamura, S. Sasamoto, A. Watanabe, M. Kohara, M. Matsumoto, S. Shimpo, M. Yamada and S. Tabata, *DNA Res*, 2003, **10**, 221-228.
23. C. C. Zhang, L. Gonzalez and V. Phalip, *Nucleic Acids. Res*, 1998, **26**, 3619-3625.
24. A. M. a. Muro-Pastor, A. Herrero and E. Flores, *J. Bacteriol*, 2001, **183**, 1090-1095.
25. R. Figge, C. Cassier-Chauvat, F. Chauvat and R. Cerff, *Mol. Microbiol*, 2001, **39**, 455-469.
26. T. Omata, S. Gohta, Y. Takahashi, Y. Harano and S.-i. Maeda, *J. Bacteriol*, 2001, **183**, 1891-1898.
27. J. Yu, G. Shen, T. Wang, D. A. Bryant, J. H. Golbeck and L. McIntosh, *J. Bacteriol*, 2003, **185**, 3878-3887.
28. T. Fujimori, M. Higuchi, H. Sato, H. Aiba, M. Muramatsu, Y. Hihara and K. Sonoike, *Plant Physiol*, 2005, **139**, 408-416.
29. M. Kobayashi, T. Ishizuka, M. Katayama, M. Kanehisa, M. Bhattacharyya-Pakrasi, H. B. Pakrasi and M. Ikeuchi, *Plant Cell Physiol*, 2004, **45**, 290-299.
30. S. K. Pilla, R. R. Balaga, S. Iwane, S. Sisinthy and S. Jogadhenu, *Biochem. J*, 2013, **449**, 751-760.
31. W. Zhang, F. Li and L. Nie, *Microbiology*, 2010, **156**, 287-301.
32. J. Wang, L. Chen, X. Tian, L. Gao, X. Niu, M. Shi and W. Zhang, *J. Proteome Res*, 2013, **12**, 5302-5312.
33. B. Zhang and S. Horvath, *Stat. Appl. Genet. Mol*, 2005, **4**, 1128.
34. H. L. Wang, B. L. Postier and R. L. Burnap, *Biotechniques*, 2002, **33**, 26, 28, 30 passim.
35. S. P. Putri, Y. Nakayama, F. Matsuda, T. Uchikata, S. Kobayashi, A. Matsubara and E. Fukusaki, *J. Biosci. Bioeng*, 2013, **115**, 579-589.
36. N. B. Bennette, J. F. Eng and G. C. Dismukes, *Anal. Chem*, 2011, **83**, 3808-3816.
37. A. Pradet and P. Raymond, *Annu. Rev. Plant Physiol*, 1983, **34**, 199-224.
38. N. Pollak, C. Dolle and M. Ziegler, *Biochem. J*, 2007, **402**, 205-218.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

## ARTICLE TYPE

39. D. Schwarz, I. Orf, J. Kopka and M. Hagemann, *Metabolites*, 2013, **3**, 72-100.
- 930 40. C. Park, Y. j. Lee, S. Y. Lee, H. B. Oh and J. Lee, *Bull Korean Chem Soc*, 2011, **32**, 524-530.
41. E. C. Laiakis, G. A. Morris, A. J. Fornace Jr and S. R. Howie, *PLoS One*, 2010, **5**, e12655.
- 935 42. L. Krall, J. Huege, G. Catchpole, D. Steinhauser and L. Willmitzer, *J. Chromatogr. B*, 2009, **877**, 2952-2960.
43. H. Zhu, X. Ren, J. Wang, Z. Song, M. Shi, J. Qiao, X. Tian, J. Liu, L. Chen and W. Zhang, *Biotechnol. Biofuels*, 2013, **6**, 106.
- 940 44. J. Wang, X. Zhang, M. Shi, L. Gao, X. Niu, R. Te, L. Chen and W. Zhang, *Funct. Integr. Genomic*, 2014, 1-10.
45. U. Roessner, A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer and A. R. Fernie, *Plant Cell*, 2001, **13**, 11-29.
46. S. E. Stein, *J. Am. Soc. Mass Spec.*, 1999, **10**, 770-781.
47. O. Fiehn, *Plant Mol. Biol.*, 2002, **48**, 155-171.
- 945 48. S. Jozefczuk, S. Klie, G. Catchpole, J. Szymanski, A. Cuadros-Inostroza, D. Steinhauser, J. Selbig and L. Willmitzer, *Mol. Syst. Biol.*, 2010, **6**.
49. P. Langfelder and S. Horvath, *BMC Bioinformatics*, 2008, **9**, 559.
- 950 50. J. Wang, G. Wu, L. Chen and W. Zhang, *BMC Genomics*, 2013, **14**, 112.
51. L. Shi, K. M. Bischoff and P. J. Kennelly, *J. Bacteriol.*, 1999, **181**, 4761-4767.
52. S. Jantaro, Q. Ali, S. Lone and Q. He, *J. Biol. Chem.*, 2006, **281**, 30865-30874.
- 955 53. A. K. Singh, T. Elvitigala, J. C. Cameron, B. K. Ghosh, M. Bhattacharyya-Pakrasi and H. B. Pakrasi, *BMC Syst. Biol.*, 2010, **4**, 105.
54. T. C. Summerfield and L. A. Sherman, *Appl. Environ. Microb.*, 2008, **74**, 5276-5284.
- 960 55. L. Huang, M. P. McCluskey, H. Ni and R. A. LaRossa, *J. Bacteriol.*, 2002, **184**, 6845-6858.
56. M. A. Fuszard, S. Y. Ow, C. S. Gan, J. Noirel, N. G. Teman, G. McMullan, C. A. Biggs, K. F. Reardon and P. C. Wright, 965 *Aquat. Biosyst.*, 2013, **9**, 5.
57. L. Houot, M. Floutier, B. Marteyn, M. Michaut, A. Picciocchi, P. Legrain, J. C. Aude, C. Cassier-Chauvat and F. Chauvat, *BMC Genomics*, 2007, **8**, 350.
58. J. Liu, L. Chen, J. Wang, J. Qiao and W. Zhang, *Biotechnol. Biofuels*, 2012, **5**, 68.
- 970 59. L. Beuf, S. Bedu, M. C. Durand and F. Joset, *Plant Mol. Biol.*, 1994, **25**, 855-864.
60. S. Kahlon, K. Beeri, H. Ohkawa, Y. Hihara, O. Murik, I. Suzuki, T. Ogawa and A. Kaplan, *Microbiology*, 2006, **152**, 647-655.
- 975 61. J. R. Waldbauer, S. Rodrigue, M. L. Coleman and S. W. Chisholm, *PLoS One*, 2012, **7**, e43432.
62. P. Desplats, E. Folco and G. L. Salemo, *Plant Physiol. Bioch.*, 2005, **43**, 133-138.
- 980 63. M. Igoillo-Esteve and J. J. Cazzulo, *Mol. Biochem. Parasit.*, 2006, **149**, 170-181.
64. L. Huan, X. Xie, Z. Zheng, F. Sun, S. Wu, M. Li, S. Gao, W. Gu and G. Wang, *Plant Cell Physiol.*, 2014, pku063.
65. D. Kurian, K. Phadwal and P. Mäenpää, *Proteomics*, 2006, **6**, 3614-3624.
- 985 66. K. A. Lê Cao, I. González and S. Déjean, *Bioinformatics*, 2009, **25**, 2855-2856.
67. T. Horinouchi, K. Tamaoka, C. Furusawa, N. Ono, S. Suzuki, T. Hirasawa, T. Yomo and H. Shimizu, *BMC Genomics*, 2010, 990 **11**, 579.
68. M. V. DiLeo, G. D. Strahan, M. den Bakker and O. A. Hoekenga, *PLoS One*, 2011, **6**, e26683.
69. J. D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. Walhout, M. E. Cusick and F. P. Roth, *Nature*, 2004, **430**, 88-93.
- 995 70. K. M. Wegener, A. K. Singh, J. M. Jacobs, T. Elvitigala, E. A. Welsh, N. Keren, M. A. Gritsenko, B. K. Ghosh, D. G. Camp and R. D. Smith, *Mol. Cell. Proteomics*, 2010, **9**, 2678-2689.
71. M. J. Quintero, A. M. Muro-Pastor, A. Herrero and E. Flores, *J. Bacteriol.*, 2000, **182**, 1008-1015.
- 1000

## Figure Legends:

**Fig. 1: Growth time courses of the wild type and the mutants in BG11 media with or without ethanol.** A) the wild type and the  $\Delta sll1392$  mutant; B) the wild type and the  $\Delta sll1712$  mutant; C) the wild type and the  $\Delta slr1860$  mutant. An ethanol concentration of 1.9% (v/v) was used.

1005

**Fig. 2: Targeted LC-MS metabolomic analysis.** A) PCA plots of the LC-MS metabolomic profiles at 48 h (above) and 72 h (below), respectively. B) Heat maps of the LC-MS metabolomic profiles at 48 h (above) and 72 h (below), respectively.

**Fig. 3: PCA plots of GC-MS metabolomic profiles.** GC-MS metabolomic profiles at 48 h (above) and 72 h (below).

**Fig. 4: Visualization of the PCA results of metabolites identified by GC-MS in each of the mutants under different conditions.** The wild type and three individual mutants under normal (-C) and ethanol-supplemented media (-E) were indicated beside the plots. In the plots, the metabolites classified were indicated by black triangles ( $\Delta$ ) and the clustering patterns were indicated by circles. The responsive metabolites were numbered and identified below. A) Wild type and  $\Delta sll1392$  at 48 h; B) Wild type and  $\Delta sll1712$  at 48 h; C) Wild type and  $\Delta slr1860$  at 48 h; D) Wild type and  $\Delta sll1392$  at 72 h; E) Wild type and  $\Delta sll1712$  at 72 h; F) Wild type and  $\Delta slr1860$  at 72 h.

1010

**Fig. 5: Overlapping of the responsive metabolites identified by GC-MS under ethanol stress condition in three mutants.** A) 48 h; and B) 72 h.

1015

**Fig. 6: WGCNA analysis of the GC-MS metabolic profiles of the wild type and the mutants under ethanol stress.** A) 48 h; B) 72 h. The distinct modules identified at each time point were indicated by color. The modules highly associated with a given condition/mutation are indicated, with the metabolites included in the modules indicated by parenthesis.

1020 **Fig. 7: The hub metabolites and their metabolic profiles as represented by the node and edge graph.** A) Glutamic acid and pyroglutamic acid in the module M2 at 48 h; B) Sucrose and methyl palmitate in the module M4 at 72 h; C) Linoleic acid and urea in the module M6 at 72 h; D) Heptadecanoic acid in the module M7 at 72 h. Only those nodes with high connectivity strength as displayed near the edges were shown.

1025

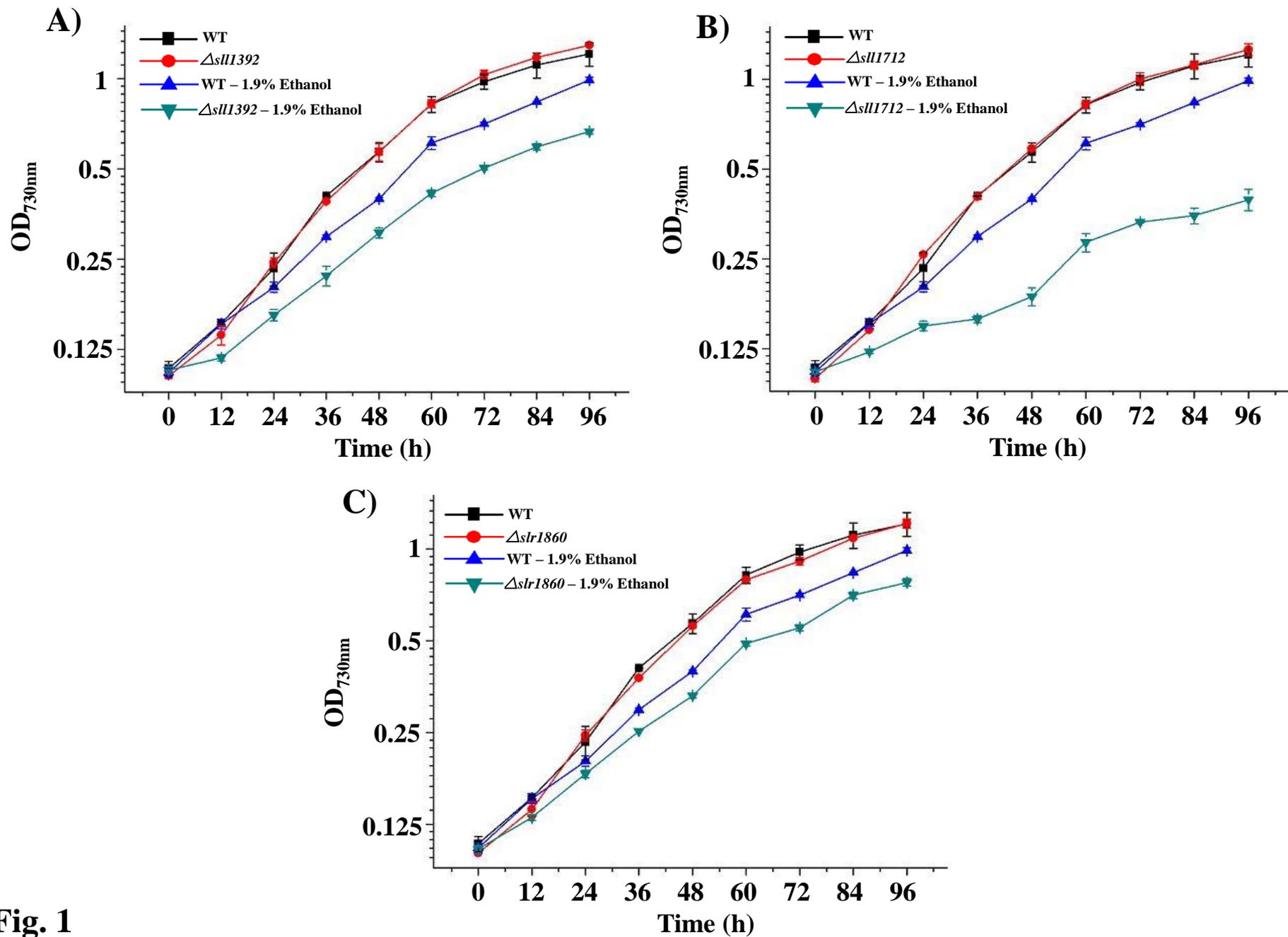
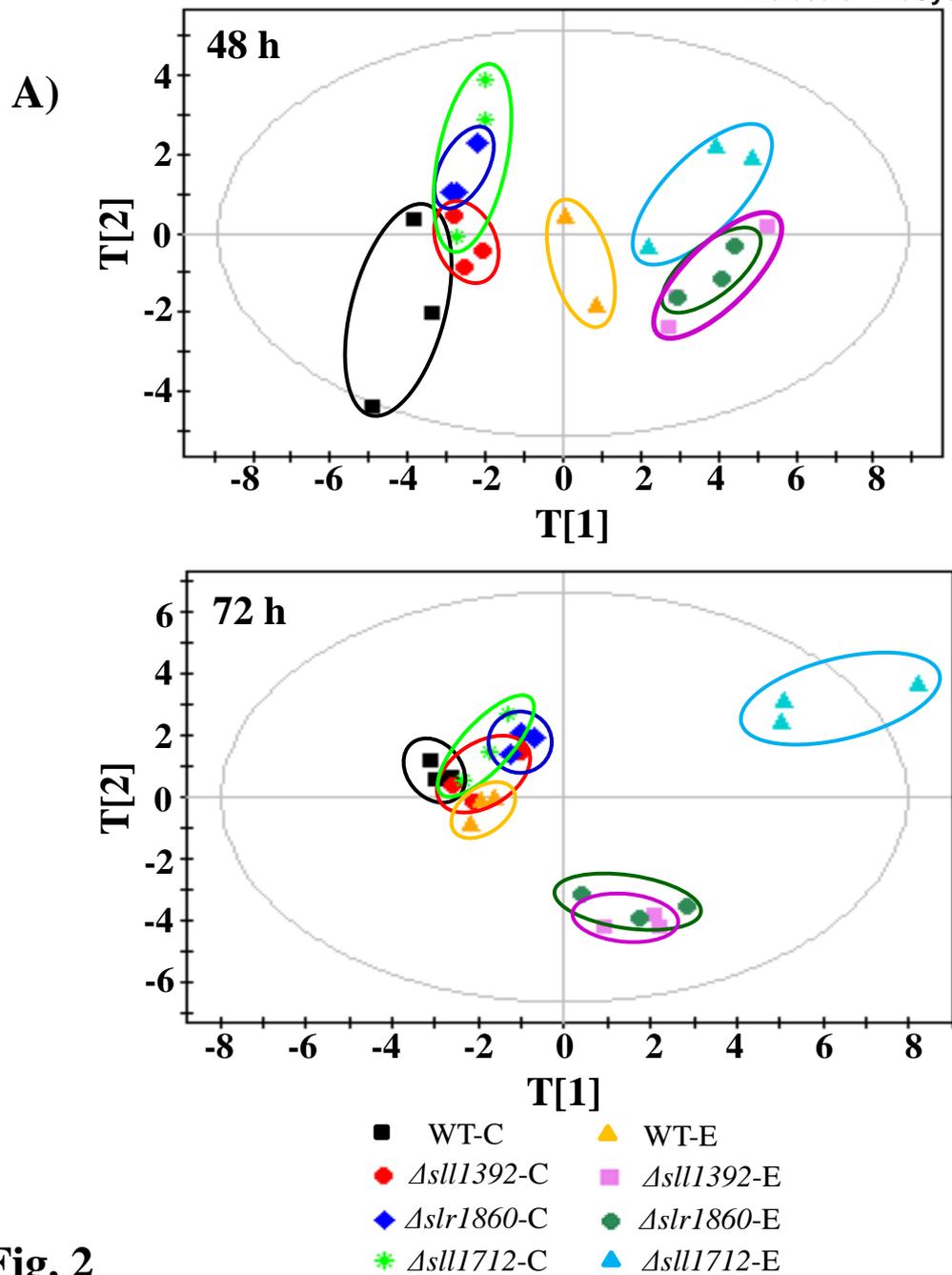
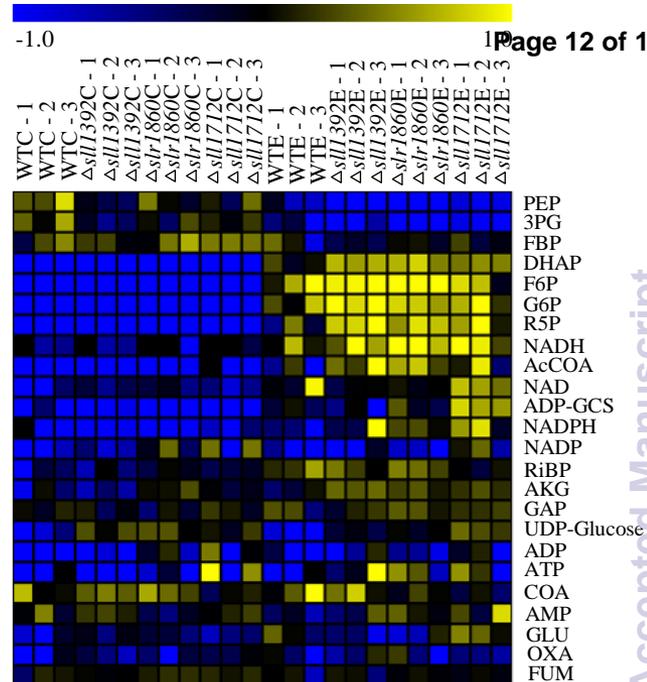
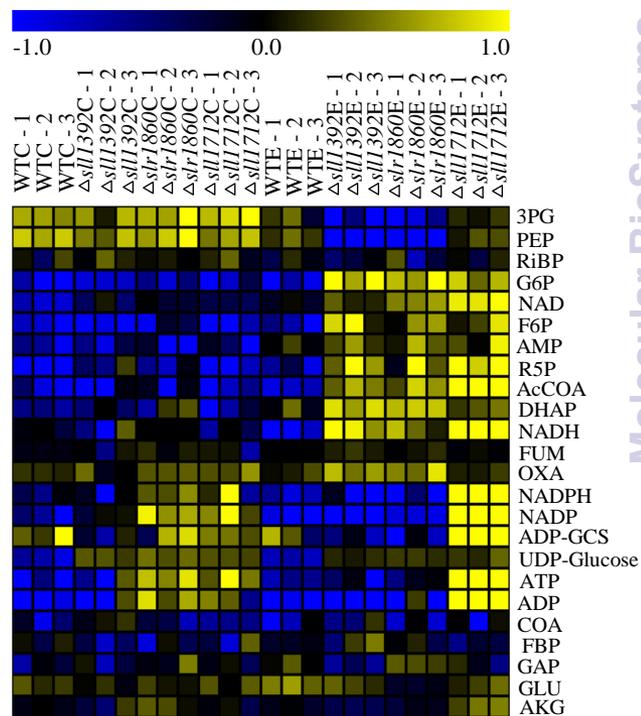
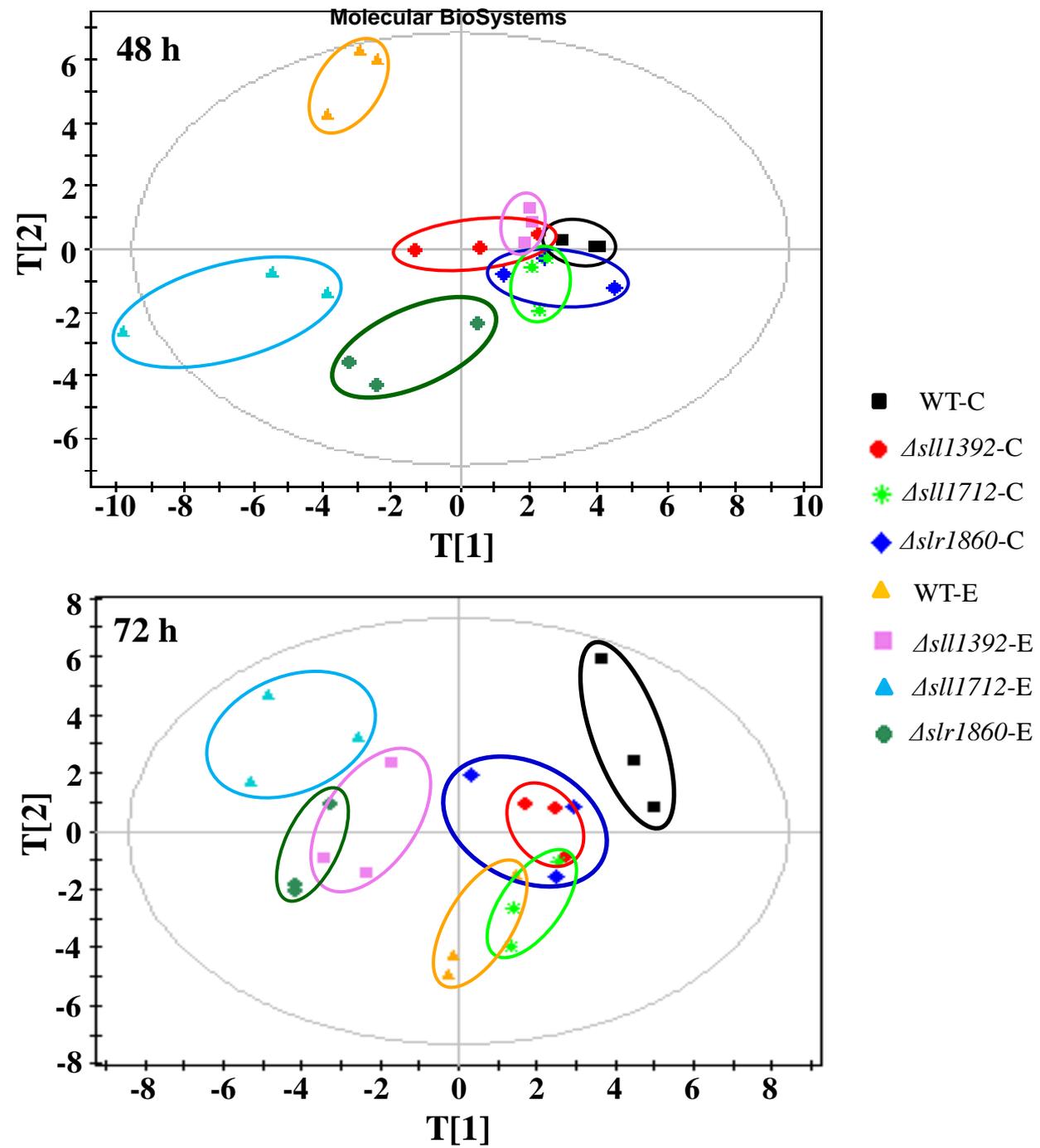
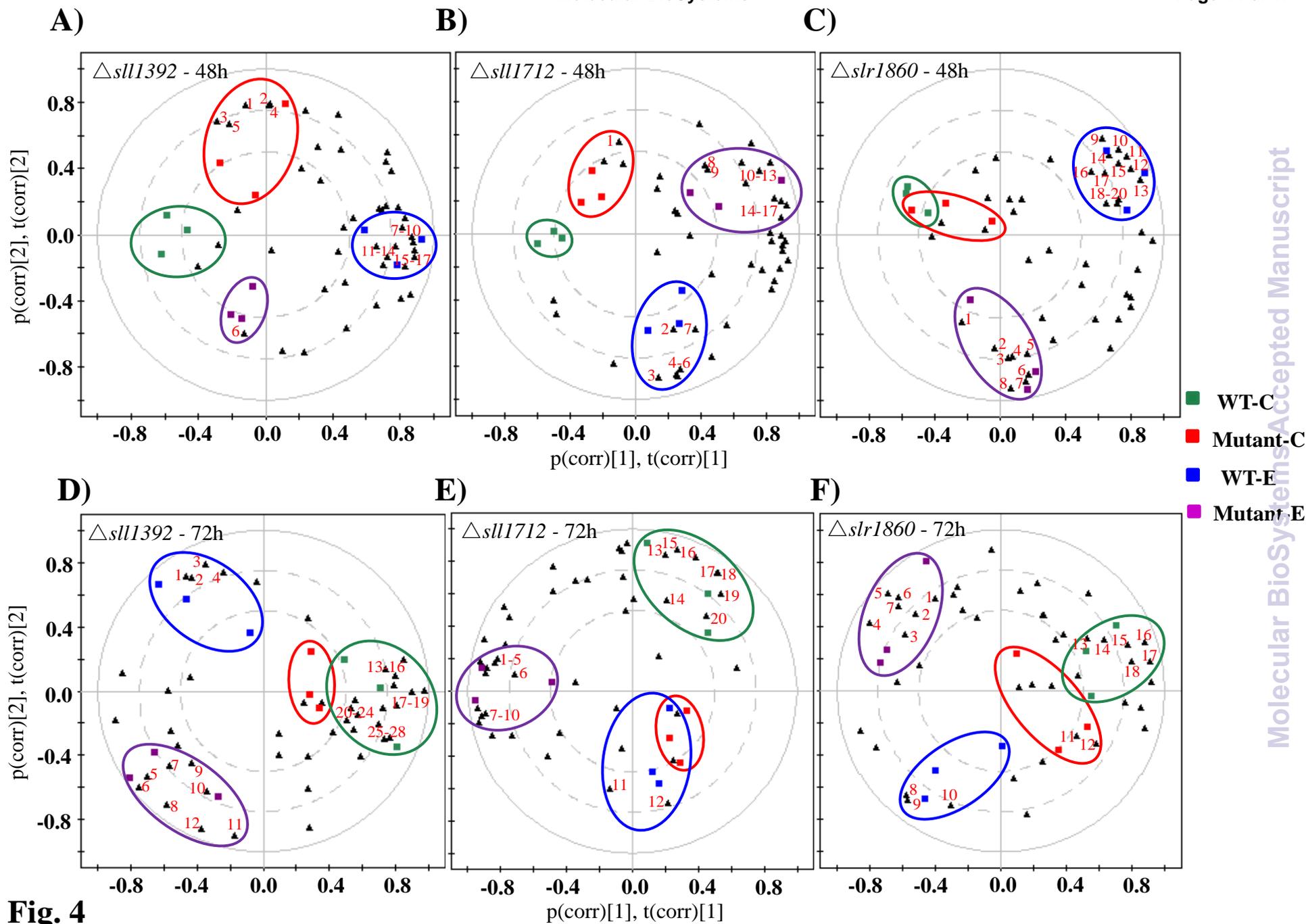


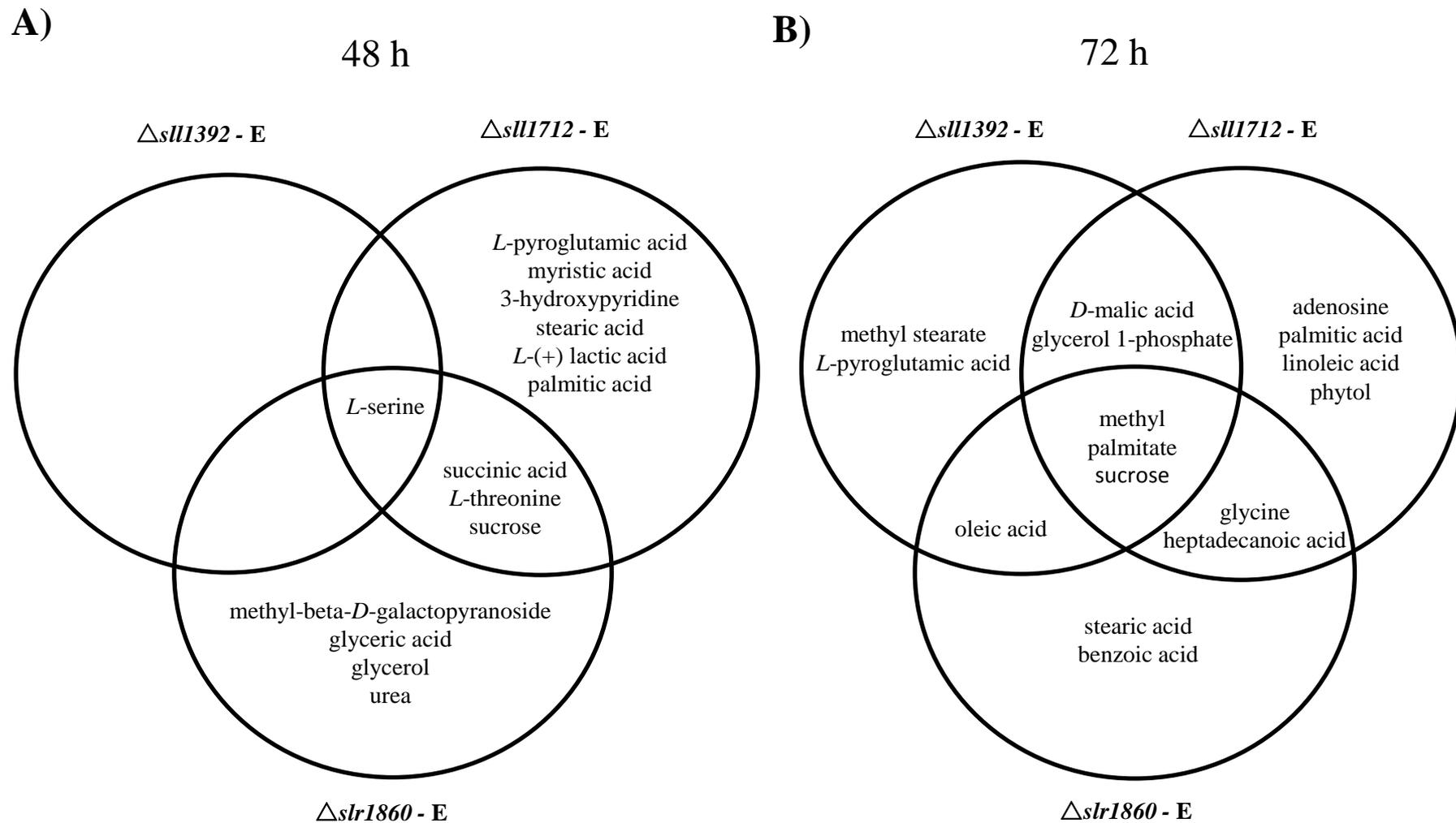
Fig. 1

**B)****48 h****72 h****Fig. 2**



**Fig. 3**

**Fig. 4**

**Fig. 5**

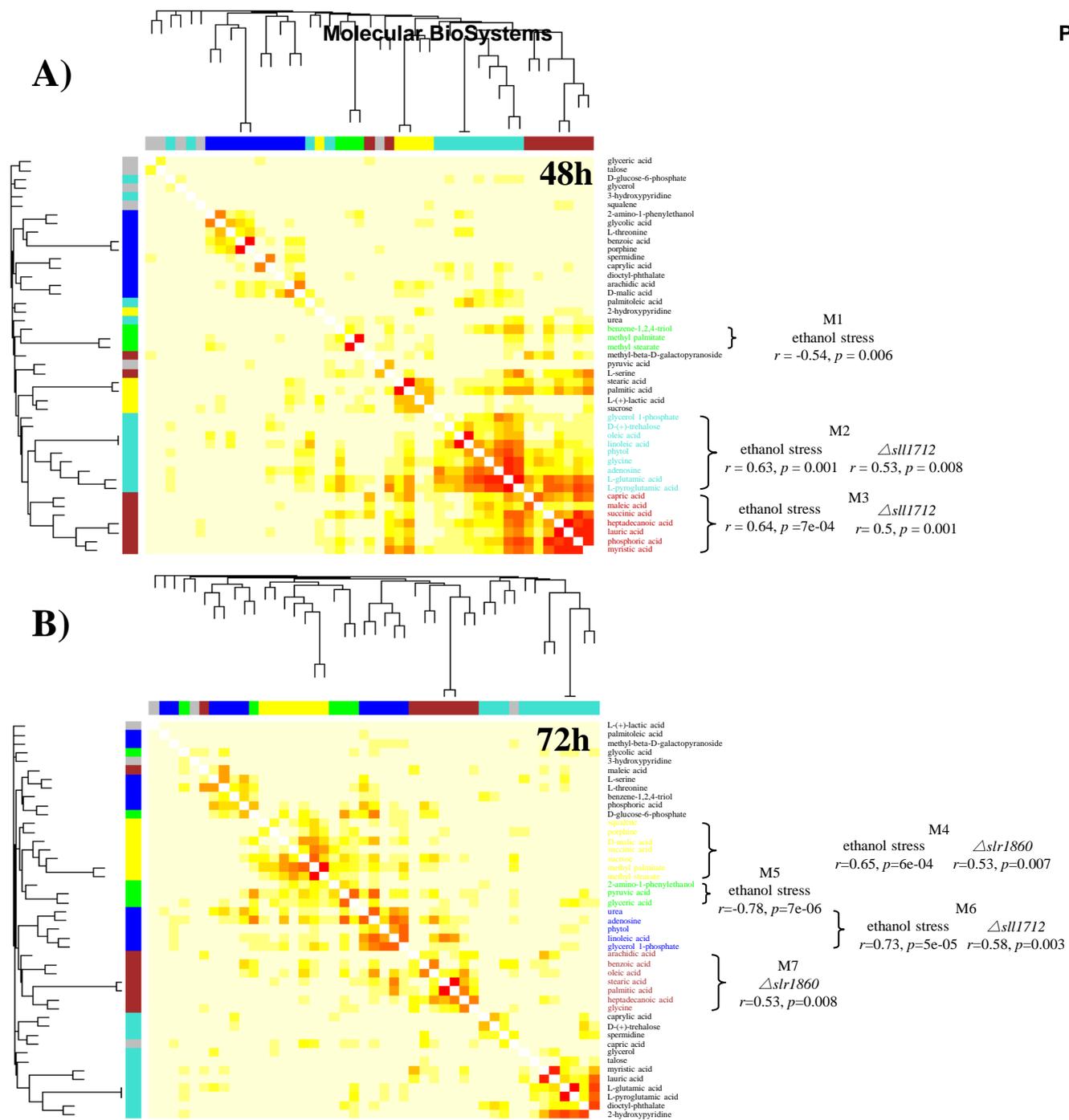


Fig. 6

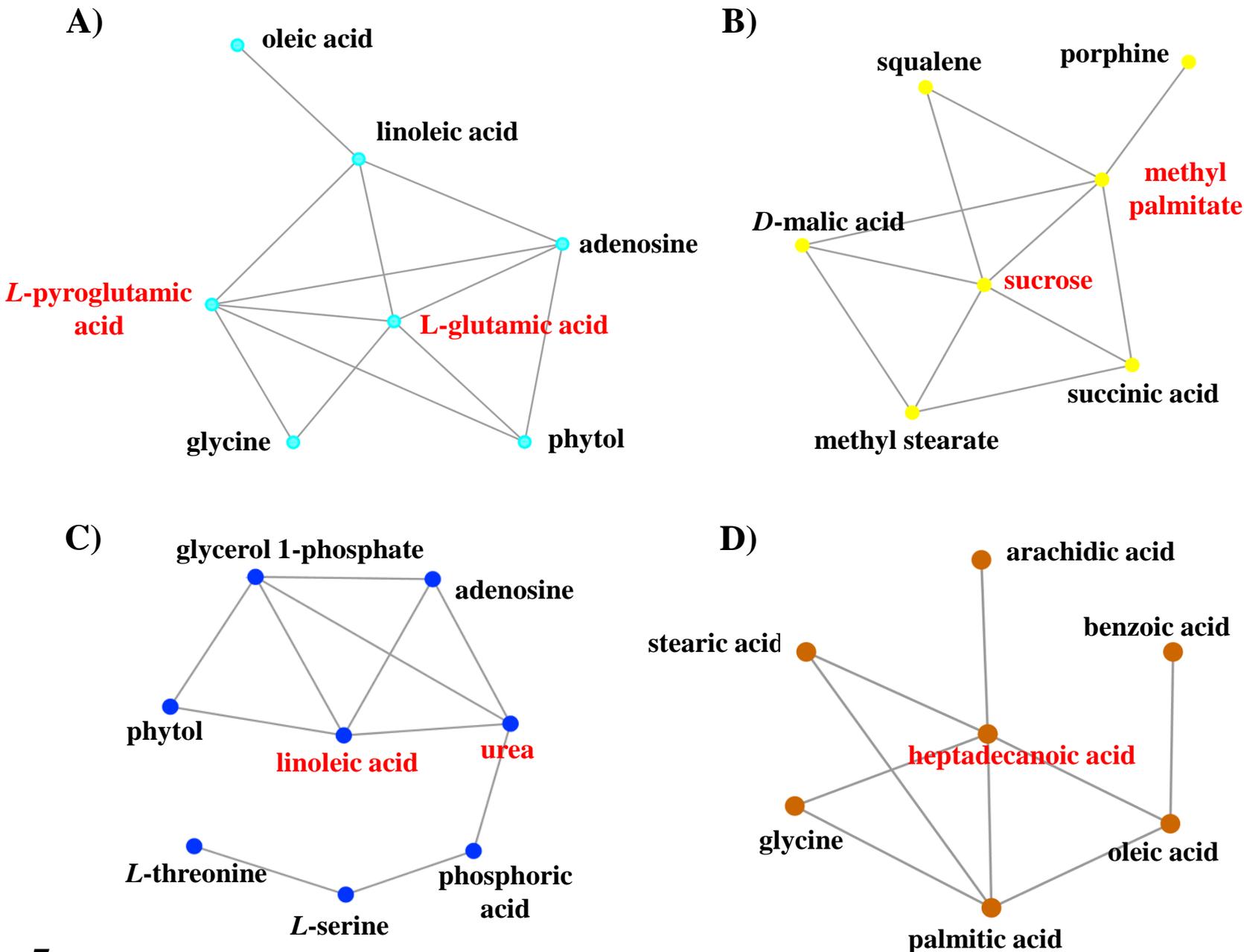


Fig. 7